Reliability analysis for industrial devices based on data fusion

Kang, W.

DOI
10.4233/uuid:51fd940e-b77f-4ca6-826b-44f92e9dd734

Publication date
2025

Document Version
Final published version

Citation (APA)
Kang, W. (2025). *Reliability analysis for industrial devices based on data fusion*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:51fd940e-b77f-4ca6-826b-44f92e9dd734

RELIABILITY ANALYSIS FOR INDUSTRIAL DEVICES BASED ON DATA FUSION

RELIABILITY ANALYSIS FOR INDUSTRIAL DEVICES BASED ON DATA FUSION

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Monday 30 June 2025 at 10:00 o'clock

by

Wenda KANG

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. G. Jongbloed,	Delft University of Technology, *promotor*
Prof. dr. Y.B. Tian,	Beijing Institute of Technology, *promotor*
Dr. P. Chen,	TU Delft / Zhejiang University, China, *copromotor*
Independent members:	
Prof. dr. Y. Liu,	Norwegian University of Science and Technology
Prof. dr. H.M. Schuttelaars,	Delft University of Technology
Dr. ir. G.F. Nane,	Delft University of Technology
Dr. D. Zarouchas,	Delft University of Technology
Prof. dr. A.J. Cabo,	Delft University of Technology, *reserve member*

The doctoral research has been carried out in the context of an agreement on joint doctoral supervision between Beijing Institute of Technology, China and Delft University of Technology, the Netherlands.

Keywords: Data fusion, Degradation, Different data sources, Reliability analysis, Remaining useful life, Transfer learning.

ISBN 978-94-6518-085-4

An electronic copy of this dissertation is available at
https://repository.tudelft.nl/.

CONTENTS

SUMMARY

As technology advances, more industrial devices are achieving higher reliability and longer lifespans. However, challenges such as limited sample sizes of experimental data and the complexity of factors influencing device degradation are becoming increasingly prevalent. Simultaneously, abundant degradation information from other data sources, including data from other components, historical batches, and different experimental stress levels, is available. Thus, there is an urgent need to find ways to fully utilize these multi-source data for industrial device reliability analysis. Therefore, this thesis proposes several data fusion methods to perform the reliability analysis of industrial devices that collect degradation data from different sources. The research addresses three primary research objectives: developing a data fusion-based framework for predicting the remaining useful life (RUL) of industrial devices that collect multivariate sensor data, formulating reliability analysis methods for degradation data from different batches of industrial devices, and establishing a framework for analyzing degradation data under varying experimental stresses and stress levels.

For the first research objective, a novel feature-based degradation index is proposed, which automatically selects features and captures the nonlinear degradation trends of industrial devices that collect multivariate sensor data. A corresponding framework for predicting RUL is then developed. The effectiveness of this framework is validated through simulations and case studies on industrial induction motors, demonstrating superior predictive accuracy over existing methods.

For the second research objective, this thesis first focuses on improving the reliability evaluation of industrial devices by integrating current and historical batch degradation data through the consistency of failure mechanisms. Utilizing a Wiener process-based degradation model, the proposed method achieves accurate and stable reliability estimates, as evidenced by their application to simulation studies and

a case study on Metal-Oxide-Semiconductor Field-Effect Transistor degradation data. Additionally, motivated by the early cycle data from lithium-iron batteries, this thesis proposes a robust transfer learning method by employing a model average framework, where the weights are determined based on the distance between the source domain and the target domain. This method addresses the challenge of robustness and generalization of lifetime predictions for batteries from different batches due to the lack of diversity in training data.

For the third research objective, to model the intricate relationships between degradation patterns, various stress variables, and stress levels, this thesis introduces a functional data-driven framework for understanding such relationships. The superior performance and effectiveness of the proposed approach are demonstrated by simulation studies and a case study on accelerated degradation data of the waxy lubrication layer.

Overall, this thesis presents several data fusion-based methods for the reliability analysis of industrial devices using data from different sources. The proposed methods demonstrate effectiveness and performance, and their applicability is not limited to the presented case studies; they can also be extended to the reliability analysis of other industrial devices with similar data sources. This indicates the potential for these methods to address broader challenges in reliability engineering across various industrial applications.

SAMENVATTING

Naarmate de technologie vordert, bereiken meer industriële apparaten een hogere betrouwbaarheid en een langere levensduur. Uitdagingen zoals beperkte steekproefgroottes van experimentele gegevens en de complexiteit van factoren die de degradatie van apparaten beïnvloeden, worden echter steeds meer prominent. Tegelijkertijd is er een overvloed aan degradatie-informatie uit andere gegevensbronnen beschikbaar, waaronder gegevens van andere componenten, historische partijen en verschillende experimentele belastingniveaus. Daarom is er een dringende behoefte om manieren te vinden om deze gegevens uit verschillende bronnen volledig te benutten voor de betrouwbaarheidsanalyse van industriële apparaten. Dit proefschrift stelt daarom verschillende methoden van datafusie voor om de betrouwbaarheidsanalyse van industriële apparaten uit te voeren op basis van degradatiegegevens uit verschillende bronnen. Het onderzoek richt zich op drie primaire onderzoeksdoelen: het ontwikkelen van een op datafusie gebaseerd raamwerk voor het voorspellen van de resterende gebruiksduur (RUL) van industriële apparaten op basis van multivariate sensorgegevens verzamelen, het formuleren van betrouwbaarheidsanalysemethoden voor degradatiegegevens van verschillende industriële apparaten, en het opzetten van een raamwerk voor het analyseren van degradatiegegevens onder verschillende experimentele belastingniveaus.

Voor het eerste onderzoeksdoel wordt een nieuwe, op kenmerken gebaseerde, degradatie-index voorgesteld die automatisch kenmerken selecteert en de niet-lineaire degradatietrends vastlegt op basis van multivariate sensorgegevens. Vervolgens wordt een soortgelijk raamwerk ontwikkeld om de RUL te voorspellen. De effectiviteit van dit raamwerk wordt gevalideerd door simulaties en casestudies van industriële inductiemotoren, waarbij ten opzichte van bestaande methoden een uitstekende voorspellende nauwkeurigheid wordt geobserveerd.

Voor het tweede onderzoeksdoel richt dit proefschrift zich eerst op Voor het

tweede onderzoeksdoel richt dit proefschrift zich eerst op het verbeteren van de betrouwbaarheidsevaluatie van industriële apparaten door huidige- en historische partijdegradatiegegevens te integreren via de consistentie van faalmechanismen. Met behulp van een op het Wiener-proces gebaseerd degradatiemodel geeft de voorgestelde methode nauwkeurige en stabiele betrouwbaarheidsschattingen, zoals blijkt uit simulatiestudies en een casestudy over de degradatiegegevens van Metal-Oxide-Semiconductor Field-Effect Transistoren. Bovendien, gemotiveerd door de vroege cyclusgegevens van lithium-ijzerbatterijen, stelt dit proefschrift een robuuste transfer learning methode voor door een "raamwerk van modelgemiddelden" te gebruiken, waarbij de gewichten worden bepaald op basis van een afstand tussen het brondomein en het doeldomein. Deze methode adresseert de uitdaging van robuustheid en generalisatie van levensduurschattingen voor batterijen uit verschillende partijen bij gebrek aan diversiteit in trainingsgegevens.

Voor het derde onderzoeksdoel, om de ingewikkelde relaties tussen degradatiepatronen, verschillende belastingvariabelen en -niveaus te modelleren, introduceert dit proefschrift een functioneel data gestuurd raamwerk voor het begrijpen van dergelijke relaties. De uitstekende prestaties en effectiviteit van de voorgestelde benadering worden onderbouwd door simulatiestudies en een casestudy over versnelde degradatiegegevens van de zogenaamde was-achtige smeerlaag.

Concluderend, presenteert dit proefschrift verschillende op datafusie gebaseerde methoden voor de betrouwbaarheidsanalyse van industriële apparaten met behulp van gegevens uit verschillende bronnen. De voorgestelde methoden laten effectiviteit en goede prestaties zien. De toepasbaarheid van de methoden is niet beperkt tot de gepresenteerde casestudies; ze kunnen ook worden toegepast op andere industriële apparaten met vergelijkbare gegevensbronnen. Dit onderstreept de potentie van deze methoden om bredere uitdagingen op het gebied van betrouwbaarheidstheorie aan te pakken in andere industriële toepassingen.

List of Abbreviations

ADT	Accelerated Degradation Test
AI	Artificial Intelligence
CNN	Convolutional Neural Network
DI	Degradation Index
FDA	Functional Data Analysis
FPC	Functional Principal Component
FPCA	Functional Principal Component Analysis
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
MMD	Maximum Mean Discrepancy
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
MSE	Mean Squared Error
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
SoC	State-of-Charge
SoH	State-of-Health
TCA	Transfer Component Analysis
TL	Transfer Learning

1

INTRODUCTION

1

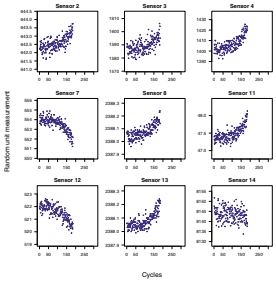
1.1. BACKGROUND

As technological advancements push the capabilities of industrial devices to new heights, achieving higher reliability and longer lifespans has become increasingly critical. However, this progress introduces significant challenges. One major issue are the limited sample sizes of experimental data, which complicate the accurate assessment of device reliability. Another challenge is the complexity of factors influencing device degradation, making it difficult to develop comprehensive models that accurately reflect real-world conditions.

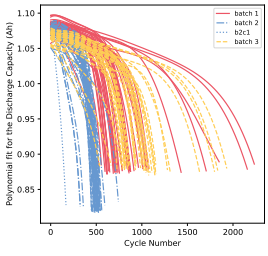
Despite these challenges, there is a wealth of degradation information available from different but related data sources, including data from different components [1], historical batches [2, 3], and varying experimental stress levels [4]. These diverse data present a valuable opportunity for enhancing reliability analysis through data fusion methods.

The examples in Figure 1.1 provide an intuitive understanding of the degradation data collected from various sources. Figure 1.1(a) shows an example of the degradation data collected from multivariate sensors in a NASA jet engine [1]. The plot displays signal measurements as a function of experimental cycles, where each subplot corresponds to a different sensor. While many sensor signals exhibit discernible trends over time, no single sensor signal can adequately represent the performance degradation process of the jet engine. Figure 1.1(b) presents degradation data from different batches of a kind of lithium-iron batteries [3], revealing that degradation trends and cycle lives vary across different batches. Figure 1.1(c) illustrates degradation data from different experimental stresses and stress levels (such as different temperature and pressure levels in this case) for a type of waxy lubrication layer, showing distinct variations in degradation trends across different stress levels.

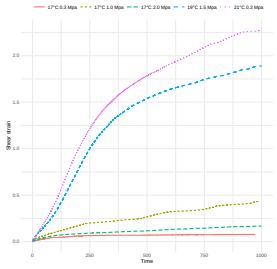
As mentioned above, degradation data of industrial devices collected from various sources can be integrated to enhance reliability analysis. To achieve this, this thesis aims to explore several data fusion methods for analyzing the reliability of industrial devices with different data sources. By leveraging multi-source degradation data, this thesis seeks to develop accurate models that improve the predictive accuracy and generalization of reliability assessments for industrial devices. These methods will not only improve predictive accuracy



(a) Example 1: Multi-sensor degradation data.



(b) Example 2: Degradation data from different batches.



(c) Example 3: Degradation data under various experimental stresses.

Figure 1.1: Illustrative examples of degradation data from various sources.

and robustness but also demonstrate flexibility in their application to various industrial devices, which will highlight the potential for these methods to be extended beyond the presented case studies, offering valuable tools for reliability analysis in a wide range of industrial applications.

1.2. RELATED CONCEPTS IN RELIABILITY ENGINEERING

THIS section introduces definitions of some concepts in reliability engineering related to this research.

- **Reliability:** Reliability is the probability that a system, vehicle, machine, device, and so on will perform its intended function under encountered operating conditions, for a specified period of time [5, 6].
- **Failure Time:** Failure time involves putting items into operation and observing them until they fail [7].
- **Degradation:** Degradation refers to the process by which a system or component deteriorates in performance or quality over time due to wear and tear, environmental factors, or operational use [6].
- **Accelerated Degradation Test:** An accelerated degradation test (ADT) is a testing methodology used to induce failures or accelerate the degradation

of a system or component under higher-than-normal stress levels to predict its lifespan or reliability under normal operating conditions [8].

- **Remaining Useful Life:** Remaining useful life (RUL) of a device or system is defined as the length from the current time to the time till failure [9].

1.3. SCOPE AND OBJECTIVES

THIS thesis aims to explore data fusion methods in the reliability analysis of industrial devices, where data fusion refers to the integration of data from multiple sources to enhance the accuracy and efficiency of the reliability analysis. The methods and models presented in this thesis are closely related to various data fusion techniques and are validated using real reliability data from several industrial devices. The results demonstrate the effectiveness and performance of the proposed methods. These methods are not limited to the case studies presented and can be directly extended to the reliability analysis of other industrial devices with similar data sources. The research is divided into the following three main research objectives:

- **Research Objective 1:** Propose a data fusion-based framework for RUL prediction of industrial devices that collect multi-channel sensor data.
- **Research Objective 2:** Develop data fusion-based reliability analysis methods for degradation data from different batches of industrial devices.
- **Research Objective 3:** Establish a data fusion-based framework for analyzing accelerated degradation data under different experimental stresses and stress levels.

1.4. OUTLINE OF THE THESIS

THE thesis consists of six chapters, focusing on the research of reliability assessment and RUL prediction by integrating information from different data sources. The overall framework of this thesis is illustrated in Figure 1.2, and the specific content of each chapter is summarized as follows:

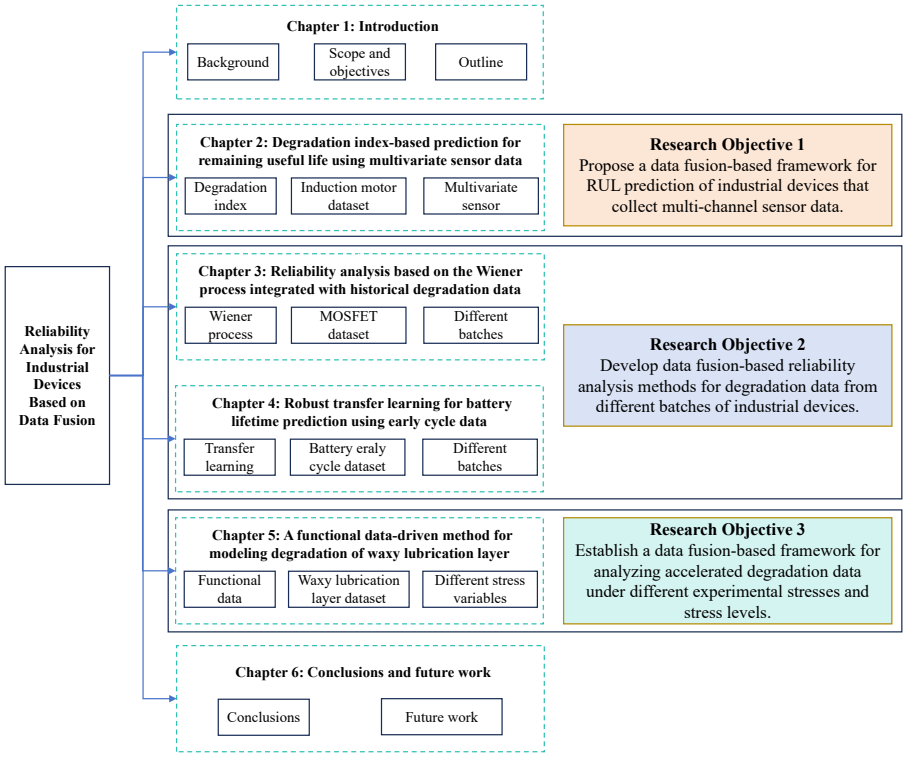


Figure 1.2: Overall framework of the PhD thesis.

Chapter 1 provides a brief introduction to the research background, motivated examples, scope and objectives, and outlines the structure of the thesis.

Chapter 2 addresses the fusion of commonly used multivariate sensor monitoring data [1, 10, 11] in industrial devices. It introduces a novel framework for representing multi-sensor data information, extracting key features, and predicting RUL online. The performance of the proposed framework is validated through simulation studies and a case study on three-phase industrial induction motors.

Chapter 3 focuses on integrating degradation information from different batches of products using the Wiener process. The goal is to improve the

reliability estimation accuracy for current batch products by leveraging abundant historical batch data. A new integration framework for historical and current degradation data is proposed, considering the consistency of failure mechanisms across different batches. The performance of the proposed method is validated through simulation studies and a case study on metal-oxide-semiconductor field-effect transistors (MOSFETs).

Chapter 4 focuses on RUL prediction for batteries using early cycle data [3] collected from different batches. A robust transfer learning [12, 13] method based on the model averaging framework is proposed to address the problem of different distributions between training and testing data. The effectiveness of the proposed method is demonstrated through a case study on lithium-iron phosphate/graphite cells.

Chapter 5 presents a functional data-driven reliability analysis framework that leverages dense observation properties in certain degradation phenomena, such as waxy degradation. The primary objective is to model the intricate relationships between degradation characteristics and two experimental stresses: temperature and pressure. This chapter introduces a novel model-free methodology for understanding these relationships in accelerated degradation tests. The proposed framework's performance is validated through simulation studies and a case study on a type of waxy lubrication layer.

Finally, Chapter 6 summarizes the thesis and suggests potential directions for future research.

2

DEGRADATION INDEX-BASED PREDICTION FOR REMAINING USEFUL LIFE USING MULTIVARIATE SENSOR DATA

Parts of this chapter have been published in *Quality and Reliability Engineering International* 40.7 (2024), pp. 3709–3728. DOI: 10.1002/qre.3615.

PLAIN SUMMARY

The prediction of RUL is a critical component of prognostic and health management for industrial systems. In recent decades, there has been a surge of interest in RUL prediction based on degradation data of a well-defined degradation index (DI). However, in many real-world applications, the DI may not be readily available and must be constructed from complex source data, rendering many existing methods inapplicable. Motivated by multivariate sensor data from industrial induction motors, this chapter proposes a novel prognostic framework that develops a nonlinear DI, serving as an ensemble of representative features, and employs a similarity-based method for RUL prediction. The proposed framework enables online prediction of RUL by dynamically updating information from the in-service unit. Simulation studies and a case study on three-phase industrial induction motors demonstrate that the proposed framework can effectively extract reliability information from various channels and predict RUL with high accuracy.

2.1. INTRODUCTION

2.1.1. BACKGROUND AND MOTIVATION

THE prediction of RUL is crucial for complex systems such as electrical systems and has become an increasingly popular research topic in recent years [10]. The RUL refers to the time remaining until a system can no longer perform its intended function, and accurate RUL prediction is essential for ensuring system safety and reliability, minimizing maintenance costs, and maximizing its lifespan. Modern sensor technology has facilitated the collection of multivariate sensor data, allowing real-time monitoring of a system's health status. These multivariate data provide valuable information on the system's performance, which can be used to predict its RUL. Therefore, developing effective methods for analyzing and processing multivariate sensor data is critical for accurate RUL prediction.

The major challenge in RUL prediction based on the multivariate sensor data is the inability of any single channel to fully capture the variation of RUL [1]. One example of such data is the multivariate three-phase industrial

induction motor data used in our case study [14]. Figure 2.1 shows an example of the motor, where 11 channels of raw signals, including current, voltage, and temperature, are presented as a function of the experimental period (i.e., the cycles shown in Figure 2.1), where the true value of RUL is also available. As seen, most signals do not exhibit a significant trend during the experiment, making them unsuitable for direct use in RUL prediction. This challenge demonstrates the need for effective techniques to extract relevant information from multiple sensor data for accurate RUL prediction.

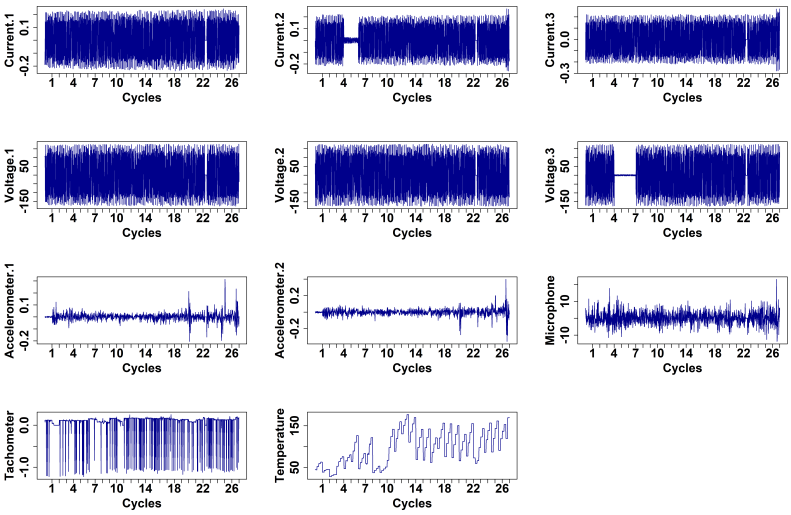


Figure 2.1: Example of multivariate sensor data for a three-phase induction motor.

In the following subsection, we provide a comprehensive review of the existing literature on RUL prediction based on multivariate sensor data, covering the three main areas of research: sensor fusion techniques, degradation index (DI) methods, and RUL prediction models.

2.1.2. LITERATURE REVIEW

SENSOR FUSION

In the field of degradation modeling for complex systems that collect multivariate sensor data, an effective fusion of the sensor data is a critical task. Existing fusion methods for multivariate sensor data can be broadly categorized into three groups: signal-level, feature-level, and decision-level [15–17].

Signal-level fusion involves the direct integration of all raw sensor signals. For instance, [18] directly fuses multisensor data using a 2-D convolutional neural network and applies several artificial intelligence (AI) methods to the fused data for fault detection and diagnosis of gearboxes. However, signal-level fusion requires caution since sensor recordings may have different acquisition, pre-filtering, and amplification settings, and raw data fusion often requires commensurate data as input [15].

Feature-level fusion predicts the health status by combining extracted features from the data of each raw sensor. This approach has been widely used due to its simplicity and effectiveness. For example, [19] proposes a RUL prediction method by performing a gated recurrent unit network on the extracted nonlinear features generated using kernel principal component analysis. [20] proposes an integrated deep multiscale feature fusion network for aero engine RUL prediction using multisensor data, and they integrate features extracted from the convolutional neural network and gated recurrent unit network.

The third category, decision-level fusion, involves integrating the decisions made from independent analyses of multivariate sensor data, such as fault diagnosis, RUL prediction, or other types of analysis tasks. For example, [21] develops a decision-level fusion method by combining the high-dimensional decisions transformed from low-dimensional decisions made based on individual sensor data. [22] proposes a decision-level method for multisensor fusion for collaborative fault diagnosis by using an enhanced voting fusion strategy. However, this approach is a post-processing technique that heavily depends on the quality of the raw data and is highly sensitive to the decision fusion rules, limiting its practical applications [23].

DEGRADATION INDEX

The previously reviewed multivariate sensor data fusion methods share a common drawback: the absence of a univariate index that credibly reflects the underlying degradation process. While some of these methods employ the raw sensor data or extracted features as input to different AI models, there is still no satisfactory fused indicator that meets requirements such as monotonicity, smoothness, and maximum range information [24]. As a result, these approaches tend to be less interpretable with respect to the underlying degradation process, making many existing statistical methods inapplicable. Consequently, the construction of an informative univariate index, or the DI as referred to in this chapter, is a crucial step towards describing the underlying degradation process based on multivariate sensor data [25, 26].

Several methods have been proposed for constructing the DI. For example, [25] proposes a method to construct the DI by fusing multi-sensor data at the signal level and using the resulting DI for the degradation modeling of an aircraft gas turbine engine. Subsequent work has been done by [24, 27–30]. In particular, [31] presents a DI building method for multivariate sensor data with censoring, which can automatically select informative sensor signals using the group LASSO penalty. However, these methods may not be suitable for all practical cases as they assume there should be a trend in some raw signals, which may not be the case where only extracted features show such trends. Furthermore, these methods are all focused on raw sensor data and may be time-consuming for high-dimensional feature spaces. In addition, [31] also notes that existing methods cannot perform automatic variable selection, and the DI and variable selection procedure in their own work lacked an explanation for the contribution of each sensor.

RUL PREDICTION

To accurately predict the RUL, it is necessary to establish a precise correlation between the constructed DI and RUL. Univariate DI-based RUL prediction methods typically fall into three categories: physical-based, data-driven, and hybrid approaches [11, 32]. Physical-based methods require a thorough understanding of the degradation behavior based on failure mechanisms, which

can be challenging to obtain for complex systems. Conversely, data-driven methods have gained attention in recent years due to their mechanism-agnostic approach, which infers the health status of products from monitored degradation signals. Hybrid methods combine both physical-based and data-driven methods, but their effectiveness may be limited by the difficulty of obtaining accurate failure mechanisms for complex systems.

Data-driven methods can be further classified into statistical and AI methods [33]. Statistical methods based on the Wiener process, Gamma process, and inverse Gaussian process have been widely used. Examples and applications can be found in [34–40], and references therein. However, these stochastic process methods have some strong assumptions such as the Markov property, and are also prone to model misspecification problems, which limit their application in engineering [41]. In contrast, AI methods are not affected by these limitations [42]. Among them, the similarity-based method is widely used for DI-based RUL prediction due to its intuitive and interpretable nature [43]. Further examples can be found in the review paper [44].

2.1.3. OBJECTIVE AND OVERVIEW

BASED on the literature review, the issues of existing methods can be summarized as follows. Although direct mapping of multivariate sensor data and RUL is possible, DI-based methods are often more intuitive and explainable. However, existing DI-based methods mainly focus on cases where the raw sensor data have significant trends, which is not suitable for many applications, as demonstrated in Figure 2.1. Despite the inclusion of feature engineering procedures, existing DI-based methods may still lack efficiency and applicability in high-dimensional feature spaces. Additionally, the accuracy of existing DI-based methods for RUL prediction heavily relies on the form of DI and the sample size of the training dataset, limiting their usefulness in certain applications.

Motivated by the above-mentioned issues, this chapter proposes DI-based prognostic frameworks for predicting RUL in complex systems. In contrast to existing DI-based methods, our constructed DI is feature-based and performs automatic feature selection, which is essential for accurately capturing the

underlying degradation trend of the system. Furthermore, our proposed DI incorporates a nonlinear relationship between the selected features and the degradation process, better reflecting the complex and nonlinear nature of practical engineering applications. Based on the constructed DI and similarity matching method, we have developed three frameworks for prognostic RUL prediction of complex systems that collect multivariate sensor data. These frameworks are designed to overcome the challenges of accurately predicting RUL. The basic principles of these frameworks are illustrated in Figure 2.2, highlighting the importance of data preprocessing, feature extraction and selection, DI construction, and RUL prediction.

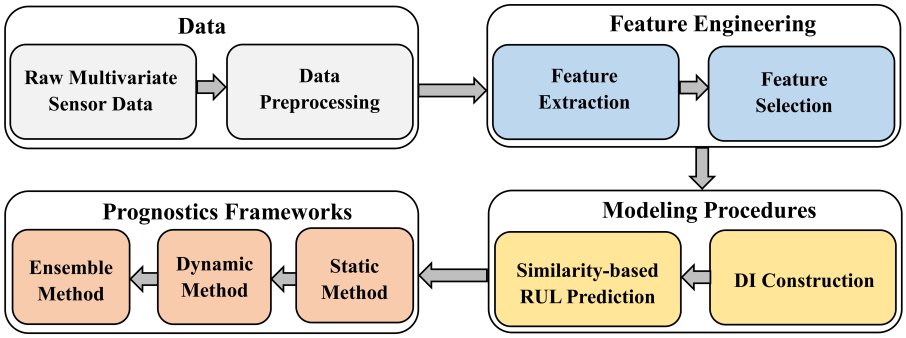


Figure 2.2: The basic procedures of the developed prognostics frameworks.

The main contributions of this chapter are summarized as follows:

(1) Introduction of feature-level prognostics frameworks for DI-based RUL prediction, which can also automatically select informative features.

(2) Development of a nonlinear form of DI to amalgamate representative features extracted from collected multivariate sensor data.

(3) Proposal of an ensemble approach that stably integrates common and individual features.

The remainder of this chapter is organized as follows. In Section 2.2, we provide details of the feature engineering process for multivariate sensor

data and the method used to construct a DI. The procedure for deriving similarity-based RUL and quantifying the uncertainty of predictions is presented in Section 2.3. We then illustrate the developed prognostics frameworks for DI-based RUL in Section 2.4. In Section 2.5, we conduct a simulation study to investigate the performance of the proposed frameworks. In Section 2.6, we provide a case study based on real-world induction motors degradation data. Finally, we give some concluding remarks and discussions in Section 2.7.

2.2. DEGRADATION INDEX CONSTRUCTION

2.2.1. FEATURE ENGINEERING

RAW sensor data typically consists of time series data with a fixed sampling frequency. However, analyzing the data at each time point can be computationally expensive and may not yield useful information. To address this challenge, feature extraction is commonly used to generate features from the raw time series that accurately describe the data while reducing computational costs [10, 11]. Additionally, feature selection can be employed to select the most informative subset of features, as not all extracted features may be useful. Therefore, feature extraction and feature selection techniques are crucial for exploring useful information and reducing computational costs.

In this study, we focus on investigating time-domain feature extraction techniques. Specifically, we employ the time domain features used in previous works such as [10] and [11]. The details of the extracted features are presented in Table 2.1, where h represents a time series with a length of k. Among all the features, (p_1, p_3, p_4, p_7) are used to capture the amplitude and energy of each signal, while the remaining ones reflect the distribution of the signal over the time domain. Note that the features listed in Table 2.1 differ for each signal and k denotes the total length of the signal.

Since noisy features can impede modeling accuracy, feature selection is often utilized to retain the most important subset of features. In many existing works on DI construction, Fisher's discriminant ratio is used as a criterion for feature

Table 2.1: Extracted features from raw sensor signals.

Feature	Description	Equation
p_1	Average amplitude	$\frac{1}{k}\sum_{i=1}^{k} h(i)$
p_2	Standard deviation	$(\frac{\sum_{i=1}^{k}(h(i)-p_1)^2}{k-1})^{1/2}$
p_3	Root mean square amplitude	$(\frac{1}{k}\sum_{i=1}^{k} h(i)^2)^{1/2}$
p_4	Squared mean rooted absolute amplitude	$(\frac{1}{k}\sum_{i=1}^{k} \lvert h(i)\rvert^{1/2})^{1/2}$
p_5	Kurtosis coefficient	$\frac{\sum_{i=1}^{k}(h(i)-p_1)^4}{(k-1)p_2^4}$
p_6	Skewness coefficient	$\frac{\sum_{i=1}^{k}(h(i)-p_1)^3}{(k-1)p_2^3}$
p_7	Peak value	$\max\lvert h(i)\rvert$
p_8	Peak factor	$\frac{p_7}{p_3}$
p_9	Margin factor	$\frac{p_7}{p_4}$
p_{10}	Waveform factor	$\frac{p_3}{\frac{1}{k}\sum_{i=1}^{k}\lvert h(i)\rvert}$
p_{11}	Impulse factor	$\frac{p_7}{\frac{1}{k}\sum_{i=1}^{k}\lvert h(i)\rvert}$

selection [41]. This ratio can be formulated as follows:

$$S_F(X_j) = \frac{(\mu_{j,1} - \mu_{j,2})^2}{\sigma_{j,1}^2 + \sigma_{j,2}^2}, \tag{2.1}$$

where $\mu_{j,c}$ and $\sigma_{j,c}^2$ are the mean and variance of feature X_j within the healthy ($c = 1$) or unhealthy ($c = 2$) class. To determine these two classes, we follow the approach used in [10] and [11], which assumes that the first few cycles are relatively healthy and the last few cycles become faulty. Specifically, we use the first 4 and the last 4 cycles and select top features with the highest Fisher's discriminant ratio to train the model, which reduces the number of features and improves computational efficiency [41].

Fisher's discriminant ratio method only identifies unit-specific informative features rather than general informative features across multiple reference units. To address this limitation, we propose a new feature selection method that selects the most common informative features across reference units. First, we calculate Fisher's discriminant ratios for all features in Table 2.1 of each unit

and sort them in descending order. We then select a certain percentage of features with the highest ratios. In practice, this percentage can be chosen by engineering background or cross-validation. In this chapter, the top 50% is used for a fair comparison, which is consistent with [10] and [11]. Next, we generate the most frequent features by selecting a certain percentage of the reference units (e.g., 5 out of 7 units in our case study) and taking the intersection of features from each subset. Finally, we select the union of these intersections as the set of selected features. This approach improves the robustness and generalization of our feature selection method. The entire feature engineering process is presented in Algorithm 2.1.

Algorithm 2.1: The overall process of feature engineering.

Input : Multivariate sensor signal data for all reference units after preprocessing.
Output: The general informative features of most reference units.

```
for each unit do
    Calculate the Fisher's discriminant ratios for all features according to Table 2.1;
    Sort these ratios in descending order;
    Take out a certain percentage of the top-ranked features;
Take a certain percentage of the reference units as subsets;
Find the intersection of features for each subset;
return The union of these intersections.
```

2.2.2. CONSTRUCTING DEGRADATION INDEX

AFTER feature engineering, the next step is to construct a suitable DI based on these selected features. Let p be the number of selected features from the raw sensor data, and $\boldsymbol{x}(s) = [x_1(s), x_2(s), \cdots, x_p(s)]$ be the corresponding features generated at operational time s.

To construct the DI, we employ the cumulative damage model [1, 31], which assumes that the degradation of a system accumulates over time and is widely

used in many engineering systems. Specifically, the DI $Z(t)$ is defined as follows:

$$Z(t) = \int_0^t u(s)ds, \tag{2.2}$$

where $u(s)$ is called the damage at the operational time point s which is a positive value and can be constructed by the selected features following the additive model

$$u(s) = \sum_{j=1}^{p} \beta_j f_j[x_j(s), \boldsymbol{\phi_j}], \tag{2.3}$$

where β_j is the parameter reflecting the contribution of the j-th feature, and $f_j[x_j(s), \boldsymbol{\phi_j}]$ is the corresponding effect function. To make $f_j[x_j(s), \boldsymbol{\phi_j}]$ flexible to capture potential nonlinear patterns of the features, we adopt a linear combination of spline basis

$$f_j[x_j(s), \boldsymbol{\phi_j}] = \sum_{l=1}^{L} \phi_{jl} bs_{jl}(s), \tag{2.4}$$

where j is the feature index, $bs_{jl}(s), l = 1, \cdots, L$ are the B-spline basis functions, L is the number of degrees of freedom, and ϕ_{jl} are the corresponding weight coefficients. The B-spline is used due to its simplicity of computation and wide applicability. More details can be found in [45].

Regarding the properties to construct the DI, we employ the following three widely used properties [25, 27–31]:

(1) *Monotonic degradation trend:* The trend of a constructed DI is assumed to be monotonic, showing a clear increasing or decreasing trend as the failure progresses during degradation. Without loss of generality, we assume that the DI is monotonically increasing in this work.

(2) *Consistent initial status:* In practice, the initial states of different units are often assumed to be the same, which can be achieved by setting the initial value of the constructed DI to 0.

(3) *Maximized range information:* The range information of the constructed DI starting from the initial to the failure time point should be maximum to guarantee a clear degradation trend.

The constructed $Z(t)$ naturally satisfies the first two properties, i.e., $Z(0) = 0$ and $Z(t)$ is monotonically increasing. As pointed out in [24] and [30], the maximum range information ensures that the range of the constructed DI (i.e., initial degradation level to the level of failure) is maximized, providing a clear degradation trend. To achieve this, we propose the following unconstrained optimization problem to determine the parameters $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$,

$$\min_{\boldsymbol{\beta}} (1-\lambda) R(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1, \tag{2.5}$$

where $R(\boldsymbol{\beta}) = 1/\min_i Z_i$ with Z_i being the DI value of the i-th reference unit at the failed time, $\lambda \in [0,1]$ is a tuning parameter which can be determined by cross-validation, and $\|\cdot\|_1$ is the L_1 norm. Note that minimizing $R(\boldsymbol{\beta})$ maximizes the overall range of the reference units. Moreover, by incorporating the lasso penalty, the proposed method automatically performs feature selection, which is a critical step often overlooked in DI-related studies. Because the objective function Eq. (2.5) is highly complex, it is recommended to use heuristic optimization algorithms [46] such as simulated annealing for optimization.

2.3. RUL PREDICTION

2.3.1. SIMILARITY-BASED RUL

WITH the constructed DI, we propose a similarity-based method for RUL prediction. The similarity-based prediction method is a popular data-driven approach that is widely used in RUL prediction because it does not require pre-knowledge of failure mechanisms or specific degradation models [43, 44]. The basic principle is to compare the DI of a test unit with those of reference units at specific time points. If they are similar, the RULs of the test and reference units should also be similar. The commonly used Euclidean distance is adopted in this chapter to measure the similarity between the DI of the test unit and the DIs of the reference units [43, 44].

Because the length of DIs for the test and reference units are often different, the distance cannot be calculated directly. Therefore, a reconstruction of the DIs is necessary to ensure that the test and reference units share the same length

of DI at the reconstructed segments. Assuming there are n failed reference units, the DI of the i-th reference unit is $\boldsymbol{Z}_i = (Z_{i,1}, \cdots, Z_{i,n_i})$, where $i = 1, 2, \cdots, n$ and $Z_{i,j}$ is the value of the DI for the i-th reference unit at time j. Note that the n_i values are integers, representing the cycle number at which the i-th reference unit failed. Let the DI of the test unit be $\boldsymbol{Z}_{T,t} = (Z_{T,1}, \cdots, Z_{T,t})$, where t is the current operating time, which is also an integer. Then, the DI of the i-th reference unit is reconstructed into $n_i - t + 1$ segments $\{\boldsymbol{Z}_{i,1}, \cdots, \boldsymbol{Z}_{i,j}, \cdots, \boldsymbol{Z}_{i,n_i-t+1}\}$, where $\boldsymbol{Z}_{i,j} = (Z_{i,j}, \cdots, Z_{i,j+t-1})$, $j = 1, \cdots, n_i - t + 1$. The Euclidean distance between $\boldsymbol{Z}_{T,t}$ and $\boldsymbol{Z}_{i,j}$ is calculated by

$$d_{i,j,t} = \| \boldsymbol{Z}_{T,t} - \boldsymbol{Z}_{i,j} \|_2, \tag{2.6}$$

where $\| \cdot \|_2$ is the L_2 norm.

The most similar segment from the i-th reference unit is then selected by using the smallest distance $d_{i,t} = \min_j d_{i,j,t}$. Let k be the index number of the most similar segment, and $\boldsymbol{Z}_{i,k} = (Z_{i,k}, \cdots, Z_{i,k+t-1})$. The corresponding predicted RUL at time t based on the i-th reference unit is derived as

$$\mathrm{RUL}_{i,t} = n_i - (k + t - 1). \tag{2.7}$$

Consequently, the similarity-based RUL of the test unit at time t is estimated by a weighted sum of the RULs derived from all the reference units, which can be expressed as

$$\mathrm{RUL}_t = \sum_{i=1}^{n} \omega_{i,t} \mathrm{RUL}_{i,t}, \tag{2.8}$$

where $\omega_{i,t} := \frac{1/d_{i,t}}{\sum_{i=1}^{n} 1/d_{i,t}}$ is the weight for the i-th reference unit at time t. Note that it is reasonable to use $\omega_{i,t}$ as the weight since a smaller value of $d_{i,t}$ indicates stronger similarity and $\sum_{i=1}^{n} \omega_{i,t} = 1$ [44].

2.3.2. PREDICTION INTERVAL FOR RUL

IN practice, the prediction interval is often more valuable and can be used to measure the uncertainty in prediction. To calculate the prediction interval of similarity-based RUL, we propose a method that applies the bootstrap method

after constructing the DIs of all units. The basic idea is to resample with replacement from the DIs of the reference units and then derive the prediction for the similarity-based RUL for the test unit using the procedures described in Section 2.3.1 based on the resampled data. This procedure is repeated M times to derive M predicted RULs. Finally, the corresponding prediction interval can be obtained by using the empirical percentiles of the M predicted RULs. For example, the 95% prediction interval can be calculated using Algorithm 2.2 by setting $\gamma = 0.05$.

Algorithm 2.2: Procedures of constructing the $100(1-\gamma)\%$ prediction interval for RUL.

Input : Constructed DIs of reference units and the DI of the test unit at time t.

Output: The $100(1-\gamma)\%$ prediction interval for the RUL of the test unit.

for $m \leftarrow 1$ **to** M **do**
  Resample with replacement from the construct DIs of reference units, and the sampling size is consistent with the number of reference units;
  Derive the most similar segments for the DI of the test unit based on the resampled DIs and Eq. (2.6);
  Determine the similarity-based RUL of the test unit at time t using Eq. (2.7) and Eq. (2.8);
Calculate the $\gamma/2$ and $1-\gamma/2$ quantiles of the M predicted RULs;
return The $\gamma/2$ and $1-\gamma/2$ quantiles of the M predicted RULs.

2.4. FRAMEWORKS FOR RUL PREDICTION

WITH the identified features, we can compute the similarity-based RUL for the test unit using the DIs constructed in Section 2.2 and the prediction procedure outlined in Section 2.3. In this section, we introduce three frameworks for RUL prediction. The first framework considers only the information from the reference units, the second framework dynamically incorporates information from both the test and reference units, and the third framework is an ensemble of the first two methods.

2.4.1. THE STATIC FRAMEWORK

FOR the collected multivariate sensor signal data, a straightforward approach is to train the parameter vector $\boldsymbol{\beta}$ in Eq. (2.5) based on the selected common features from the reference units and then use this parameter vector to predict the RUL for the test unit. We refer to this method as a static method since it relies solely on the information from reference units to derive the parameter vector $\boldsymbol{\beta}$.

To determine the DIs for the reference and test units, the common features are first selected from the raw sensor data of the reference units using the feature engineering method outlined in Section 2.2.1. Then, the contribution parameters $\beta_j, j = 1, \cdots, p$ in Eq. (2.3) can be derived by solving Eq. (2.5) using these data, denoted as $\boldsymbol{\beta_R}$. Thus, the DIs of reference and test units can be determined by substituting $\boldsymbol{\beta_R}$ and the corresponding feature data into Eq. (2.2). Using these DIs and the prediction method described in Section 2.3, the similarity-based RUL for the test unit can be derived. The flowchart of this method is illustrated in Figure 2.3.

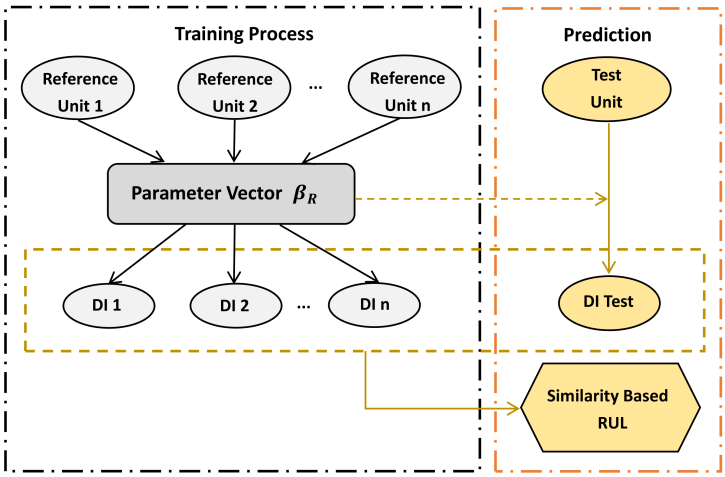


Figure 2.3: Flowchart of the static prognostics.

2.4.2. THE DYNAMIC FRAMEWORK

WHILE the static method presented relies solely on the common information shared by the reference units, it is important to note that each unit also has unique individual features that affect its degradation process. In order to account for both the common and individual features, we propose a dynamic prognostic method that integrates both. We divide the DI at time t in Eq. (2.2) into two parts, the common part $Z^{(1)}(t)$ and the individual part $Z^{(2)}(t)$, which can be expressed as

$$Z(t) = Z^{(1)}(t) + Z^{(2)}(t) = \int_0^t u_1(s)ds + \int_0^t u_2(s)ds, \tag{2.9}$$

and,

$$\begin{aligned} u_1(s) &= \sum_{j=1}^{p_1} \beta_{1_j} f_{1_j}[x_{1_j}(s), \boldsymbol{\phi_{1_j}}], \\ u_2(s) &= \sum_{j=1}^{p_2} \beta_{2_j} f_{2_j}[x_{2_j}(s), \boldsymbol{\phi_{2_j}}], \end{aligned} \tag{2.10}$$

where p_1 and p_2 are respectively the numbers of common and individual features, β_{1_j} and β_{2_j} are the corresponding contribution parameters, and $f_{1_j}[x_{1_j}(s), \boldsymbol{\phi_{1_j}}]$ and $f_{2_j}[x_{2_j}(s), \boldsymbol{\phi_{2_j}}]$ capture the effects of the corresponding features.

To obtain the DI of the test unit, the contribution parameters of the common part $\beta_{1_j}, j = 1, \cdots, p_1$ are assumed to be consistent with the reference units, while the individual parameters $\beta_{2_j}, j = 1, \cdots, p_2$ are allowed to vary based on the data collected from the test unit at operating time t. Using the common features of reference units and Eq. (2.5), $\beta_{1_j}, j = 1, \cdots, p_1$ can be derived by the static framework, denoted as $\boldsymbol{\beta}_{R_C}$. Each reference unit's individual contribution parameter, denoted as $\boldsymbol{\beta}_{R_I}$, is independently computed by solving Eq. (2.5). This computation focuses on the top informative individual features exclusive to each reference unit, excluding common features. It is important to note that there is no overlap between the sets of common and individual features, and the values of $\boldsymbol{\beta}_{R_I}$ vary among distinct reference units.

Up until this point, the proposed framework only utilizes the degradation data from the reference units and can be performed offline: the DIs of reference units can be determined by $\boldsymbol{\beta}_{R_C}$ and $\boldsymbol{\beta}_{R_I}$. As for the test unit, the individual contribution parameter $\boldsymbol{\beta}_{T_I}$ can be dynamically updated by solving Eq. (2.5) using the selected individual features at operating time t. Combining with the common parameter $\boldsymbol{\beta}_{R_C}$, we can dynamically obtain the DI of the test unit and calculate the corresponding similarity-based RUL at each operation time point. The flowchart of the dynamic method is illustrated in Figure 2.4.

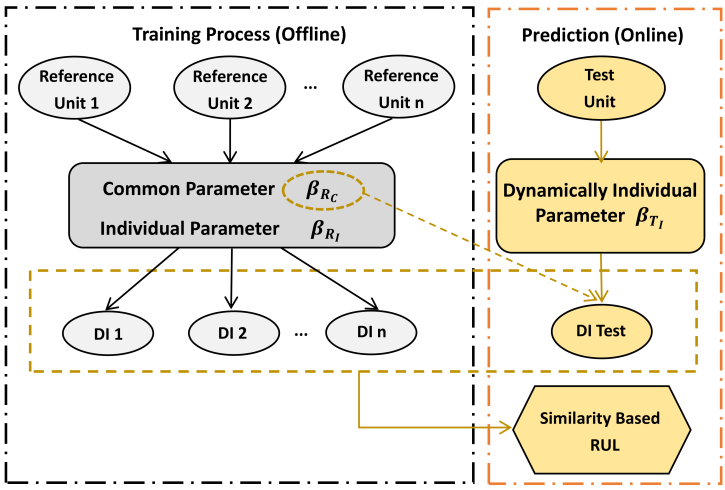


Figure 2.4: Flowchart of the dynamically updating prognostics.

2.4.3. THE ENSEMBLE FRAMEWORK

THE effectiveness of the proposed dynamic framework largely depends on the individual features selected for the test unit. This is because the data size of the test unit is typically smaller than that of the reference units, making it more susceptible to random errors during the dynamic updating process, particularly when the operating time t is short. To address this issue, the following ensemble framework is proposed to achieve a more stable prediction. The basic idea is to ensemble the predicted RULs from the static and dynamic

frameworks based on the fact that the RUL of a unit will not improve over time and will not experience a sudden big drop [10, 11]. Let $\text{RUL}_{S,t}$ denote the predicted RUL at the operating time t from the static framework, $\text{RUL}_{D,t+1}$ denote the predicted RUL at the operating time $t+1$ from the dynamic framework. Then $\text{RUL}_{S,t} - \text{RUL}_{D,t+1}$ can be defined as the ensemble condition. Specifically, if $\text{RUL}_{S,t} - \text{RUL}_{D,t+1}$ is smaller than a preset minimum drop, $\min_d$, then the prediction at time $t+1$ is $\text{RUL}_{S,t} - \min_d$. If $\text{RUL}_{S,t} - \text{RUL}_{D,t+1}$ is larger than a preset maximum drop, $\max_d$, then the prediction at time $t+1$ is $\text{RUL}_{S,t} - \max_d$. Otherwise, the prediction at time $t+1$ is $\text{RUL}_{D,t+1}$. The values of $\min_d$ and $\max_d$ can be determined through cross-validation. The flowchart of the ensemble method is illustrated in Figure 2.5.

2.5. SIMULATION STUDY

2.5.1. SIMULATED DATASET

In this section, a simulation study is conducted to evaluate the performance of the proposed frameworks. The settings are designed to be similar to those in the case study in Section 2.6. Specifically, there are 10 units, and for each unit, 10 signal data are collected. The failed cycles for the units are $(24, 26, 25, 23, 28, 22, 25, 24, 21, 19)$. Similar to [31], we assume each signal is a trend function of the experimental cycle with some noise, which can be formulated as $X_m(t) = g_m(t) + \varepsilon_m(t), m = 1, \cdots, 10$. Without loss of generality, the trend function $g_m(t)$ can be a constant, linear, power, or trigonometric function, and the noise $\varepsilon_m(t)$ follows a zero-mean normal distribution. The simulated dataset and corresponding functions $g_m(t)$ are shown in Figure 2.6.

2.5.2. FEATURE ENGINEERING AND DATA NORMALIZATION

Prior to applying the proposed frameworks, feature engineering is necessary as discussed in Section 2.2.1. Using the data in Figure 2.6, the features in Table 2.1 can be calculated. Each unit has a total of 110 features, with 10 signals and 11 features per signal. In the feature selection stage, we employ Fisher's discriminant ratio and Algorithm 2.1 to select useful features. In the simulated dataset, the number of subsets is 36, as we consider 7 out of 9 as the percentage

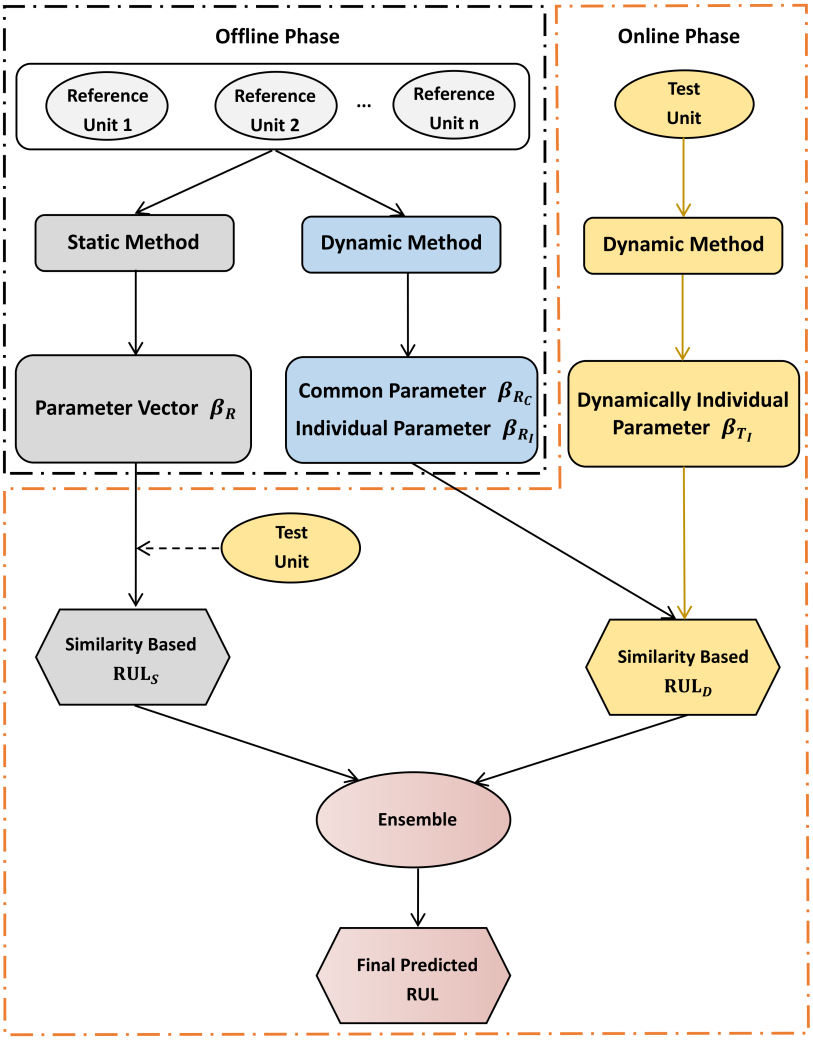


Figure 2.5: Flowchart of the ensemble prognostics.

mentioned in Algorithm 2.1. The number of selected common features for different units are as follows: (43, 43, 43, 48, 44, 48, 43, 43, 48, 45). Figure 2.7 illustrates a set of randomly selected features along with their corresponding

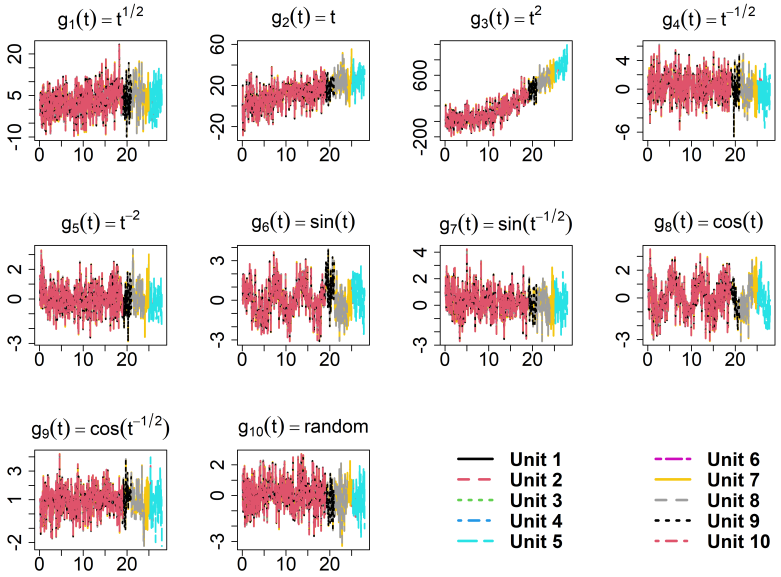


Figure 2.6: Simulated data for 10 units with 10 signals. Each panel represents a single signal. The horizontal axis shows experimental cycles and the vertical axis shows signal measurement. Different colors and line types represent different units.

coefficients, which were estimated by the static framework using simulated data. The visual representation demonstrates the successful generation and selection of informative features by the proposed framework. Additionally, it effectively captures both increasing and decreasing trends. The feature numbers (NO) represent the identifiers, while the coefficients reflect the contribution of each feature to the model.

To reduce the impact of varying data magnitudes, the min-max approach is used to normalize the selected features before model training [47]. For a selected feature, it can be formulated as

$$X' = \frac{X - \min(X)}{\max(X) - \min(X)}, \tag{2.11}$$

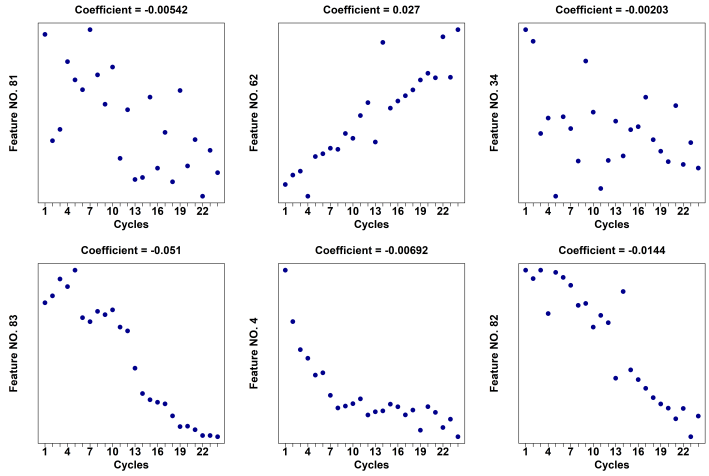


Figure 2.7: Example of the selected features and the corresponding coefficients in the simulated dataset.

where X denotes the original feature, $\max(X)$, $\min(X)$ are calculated over cycles, and X' refers to the normalized version.

2.5.3. PERFORMANCE EVALUATION

To evaluate the performance of the RUL prediction, the commonly used metric, the root mean square error (RMSE), is employed [10]. Let RUL and $\widehat{\text{RUL}}$ be the vector of true and predicted RUL*s* of one specific unit, respectively, and n_T be the corresponding failed cycle number. Then, the RMSE can be formulated as,

$$\text{RMSE} = \|\text{RUL} - \widehat{\text{RUL}}\|_2 / \sqrt{n_T}. \tag{2.12}$$

In addition to RMSE, it is also important to quantify the uncertainty associated with RUL predictions. To do this, a 95% prediction interval is widely used, which can be calculated using Algorithm 2.2 by setting $\gamma = 0.05$.

2.5.4. SIMULATION PERFORMANCE

To verify the feasibility and effectiveness of the proposed frameworks in this chapter, we utilize two additional methods from [10] and [11] as benchmarks since they also concentrate on the same dataset as in our case study.

These studies assume that the DI and RUL have a fixed linear [10] or nonlinear [11] relationship. To predict the RUL, they first build the model from the input features to the DIs using the feed-forward neural network with one hidden layer, and then they smooth the DI dynamically to improve the quality of the DI. Finally, with the fixed linear or nonlinear relationship, they predict the RUL based on the constructed DI. More details can be found in [10] and [11].

The leave-one-out approach is utilized to validate the performance and determine the reference units. For instance, if unit 1 is chosen as the test unit, the remaining 9 units are treated as reference units. RMSEs based on different methods are reported in Table 2.2, where M1 is the method in [10], M2 is the method in [11], M3 is the static method in Section 2.4.1, M4 is the dynamic method in Section 2.4.2, and M5 is the ensemble method in Section 2.4.3. The average value of the RMSE in predicting all the units is reported in the last row in the table and the minimum RMSE value for each unit is highlighted in bold. For uncertainty quantification, the results of 4 representative units are shown in Figure 2.8, where the solid black line is the true value of RUL.

Table 2.2 and Figure 2.8 suggest that the proposed methods (M3 and M5) generally outperform the existing methods (M1 and M2), as indicated by their smaller RMSE and narrower prediction intervals. In particular, the ensemble method (M5) outperforms the static method (M3) in most cases (6 out of 10 and the mean RMSE case), highlighting the effectiveness of integrating static and dynamic methods using the proposed ensemble framework. The dynamic method (M4) yields comparable results per unit to the existing methods, suggesting that it is able to extract useful information using dynamic updating. However, the inconsistent performance also underscores the necessity of the ensemble method (M5).

Note that M1 and M2 demonstrate similar performances, as evidenced by their comparable RMSEs. This could be due to the simulated dataset having a

Table 2.2: RMSE of RUL prediction for the simulated dataset.

Test Unit	M1	M2	M3	M4	M5
Unit 1	2.44	3.14	0.58	1.51	**0.42**
Unit 2	3.20	6.32	**1.84**	3.76	1.90
Unit 3	2.60	4.90	1.00	2.55	**0.87**
Unit 4	2.71	1.83	1.43	2.76	**1.28**
Unit 5	4.61	9.02	4.61	4.99	**4.40**
Unit 6	3.28	**1.39**	3.25	4.69	3.26
Unit 7	2.61	4.90	0.93	1.52	**0.80**
Unit 8	2.44	3.22	1.43	3.59	**1.17**
Unit 9	3.99	2.53	**0.83**	2.66	1.43
Unit 10	5.84	5.91	**1.96**	4.36	2.20
Mean	3.37	4.32	1.79	3.24	**1.77**

predominantly linear relationship between DI and RUL. As a result, M1 and M2 may have similar capabilities in capturing the underlying trend.

2.6. CASE STUDY

In this section, a case study on the degradation data of 8 three-phase industrial induction motors is presented to demonstrate the implementation of the proposed frameworks.

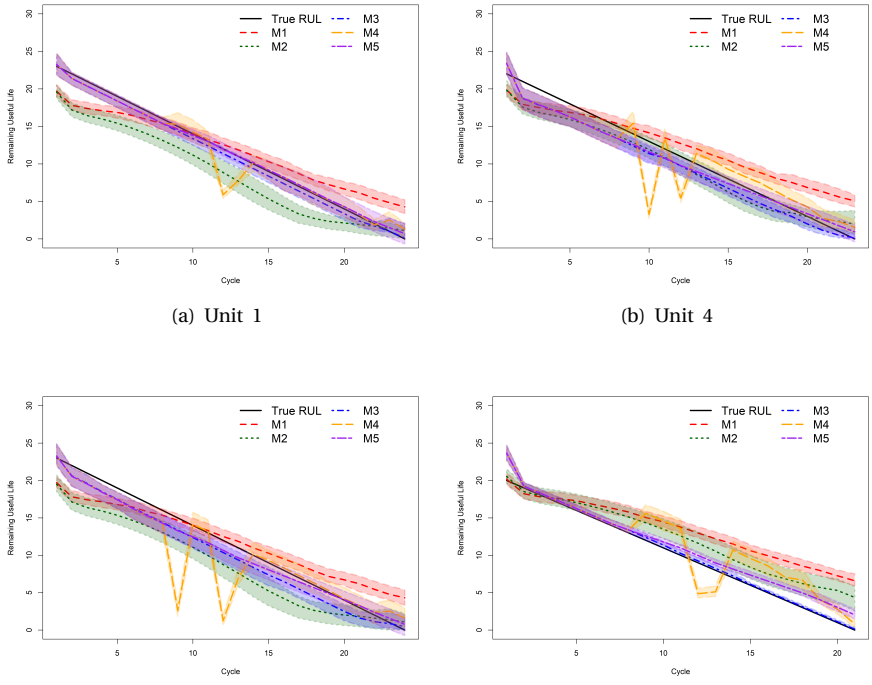
2.6.1. DATA OVERVIEW

The data was reported by [14], where ten 5-horsepower motors were used for the accelerated thermal aging process. As depicted in Figure 2.9(a), each thermal aging cycle lasts approximately one week. Further details of each cycle are provided below.

(1) Initial heating: Heat the motor in one of 3 identical EW-52402-91 ovens at 160°C (or 140°C) for 72 hours.

(2) Air cooling: Remove and allow to air cool for 6 hours.

(3) Quenching/humidity chamber: Quench in an enclosed shallow water pool for 15 minutes.



(a) Unit 1

(b) Unit 4

(c) Unit 8

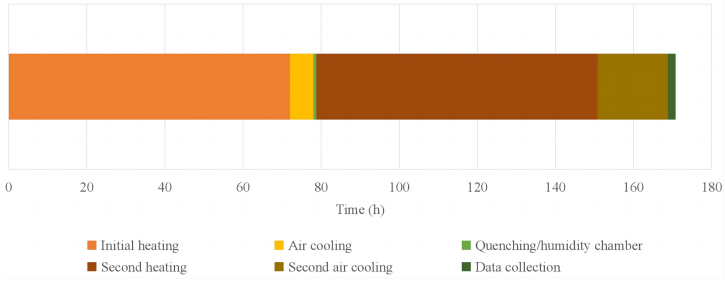
(d) Unit 9

Figure 2.8: RUL predictions and the 95% prediction interval for the simulated dataset.

(4) Second heating: Immediately place back in the oven and heat again for 72 hours.

(5) Second air cooling: Air cool for 18 hours before data collection.

In the data collection stage, as shown in Figure 2.9(b), each motor was connected to a Winco generator through an elastomeric coupling and instrumented with a data collection system. The steady-state data was collected for 2 seconds every 15 minutes at 10 kHz and 4 times per cycle for each motor [10, 11, 14]. The process of thermal aging and data collection was repeated



(a) Thermal aging process

(b) Data collection

Figure 2.9: Timeline of the accelerated thermal aging and data collection process.

until the motor failed to startup normally. The experimental device setup for accelerated thermal aging motor experiments is illustrated in Figure 2.10 [14].

In general, this experiment collected 13 channels of key signals including three-phase current (Current 1, 2, 3), three-phase voltage (Voltage 1, 2, 3), two directions of vibration (Accelerometer 1, 2), acoustic (Microphone), speed (Tachometer), temperature (Temperature), and load (output current and voltage) signals. The two channels of load signals were excluded because they were measured by connecting a motor to specific load equipment which is unavailable in practical systems [10, 11]. Since 2 out of the 10 motors experienced abnormal faults during the experiment, the data from 11 signal channels of the rest 8 motors were used in this chapter. The details of time to failure for each motor and the corresponding missing values are given in Table 2.3.

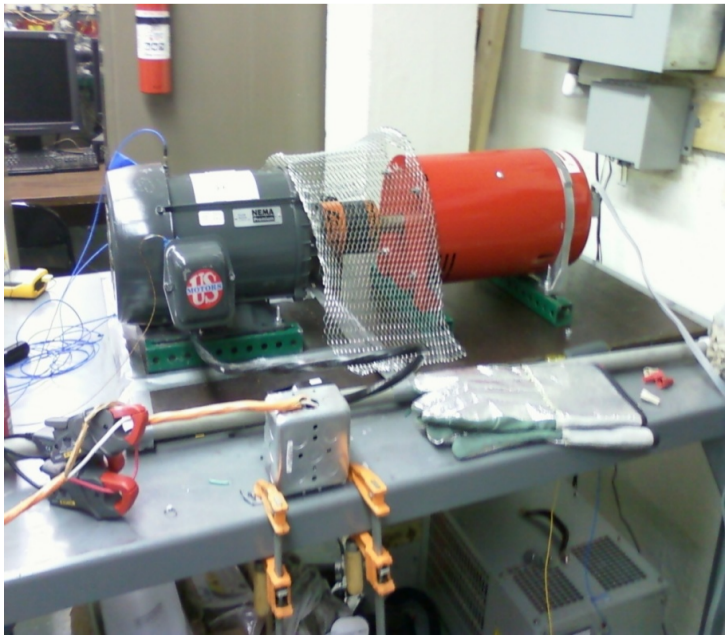


Figure 2.10: Experimental device for accelerated thermal aging of electric motor.

Table 2.3: Details of time to failure for each motor and the corresponding missing values.

Motor	Failed cycle	Missing values
1	18	Current 2: cycle 5 & 6; Voltage 3: cycle 5, 6 & 7
2	27	Current 2: cycle 5 & 6; Voltage 3: cycle 5, 6 & 7
3	26	Current 1: cycle 7; Current 2: cycle 5, 6 & 7; Current 3: cycle 7; Voltage 3: cycle 5, 6 & 7
4	29	Current 2: cycle 5 & 6; Voltage 3: cycle 5, 6 & 7
5	28	Current 2: cycle 2; Voltage 3: cycle 2 & 3
6	27	Current 2: cycle 5; Voltage 3: cycle 5, 6 & 7
7	27	Current 2: cycle 5 & 6; Voltage 3: cycle 5, 6 & 7
8	25	Current 2: cycle 5; Voltage 3: cycle 5 & 6

2.6.2. DATA PRE-PROCESSING AND FEATURE ENGINEERING

As shown in Table 2.3, some cycle signals have missing values due to human error in data acquisition or short-term damage to the data collection system. Following [10, 11], these missing values are replaced by the nearest historical values. Specifically, when a cycle of signals at a channel is missing, it is replaced with values from its previous cycle. For instance, for the missing values of Current 1 in cycle 7 for Motor 3, they are replaced by signals from cycle 6. Then, the features listed in Table 2.1 can be calculated. Each motor has 121 features in total, as there are 11 channel signals with 11 features per signal. The leave-one-out approach is employed to validate the performance and determine the reference motors, which is consistent with the simulation study. As highlighted in Section 2.2.1, we first select useful features using Fisher's discriminant ratio and Algorithm 2.1. For the proposed frameworks, the number of subsets is 21, as we consider 5 out of 7 as the percentage mentioned in Algorithm 2.1, and the number of selected common features for different test motors are $(50, 50, 50, 48, 47, 48, 50, 51)$.

Figure 2.11 displays an illustration of the selected features and their corresponding coefficients estimated by the static framework. The visual representation reveals that the proposed framework successfully generated and selected informative features, while also effectively capturing both increasing and decreasing trends.

2.6.3. PERFORMANCE OF THE PROPOSED FRAMEWORKS

SIMILAR to the simulation study, the RMSE in RUL prediction based on different methods of each motor is reported in Table 2.4. Recall that M1 refers to the method proposed in [10], M2 corresponds to the method presented in [11], M3 denotes the static method detailed in Section 2.4.1, M4 represents the dynamic method shown in Section 2.4.2, and finally, M5 indicates the ensemble method elaborated in Section 2.4.3. The last row in the table shows the mean value of the RMSE in predicting all motors, and the minimum RMSE value in each row is highlighted in bold for easy comparison of these methods. For uncertainty quantification, the results of 4 representative motors are shown in Figure 2.12, where the solid black line is the true value of RUL.

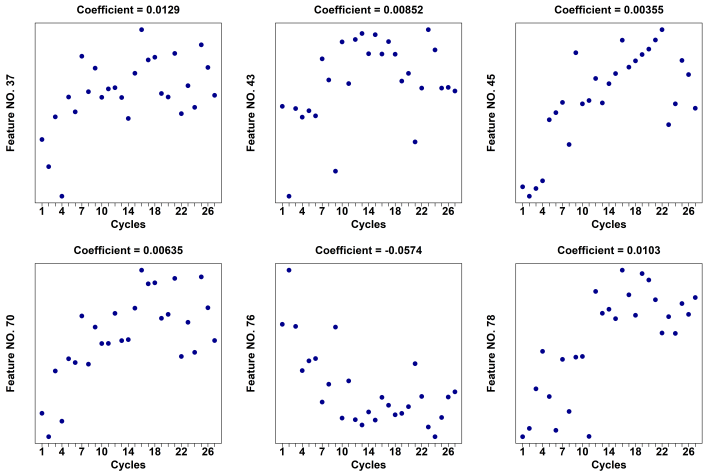
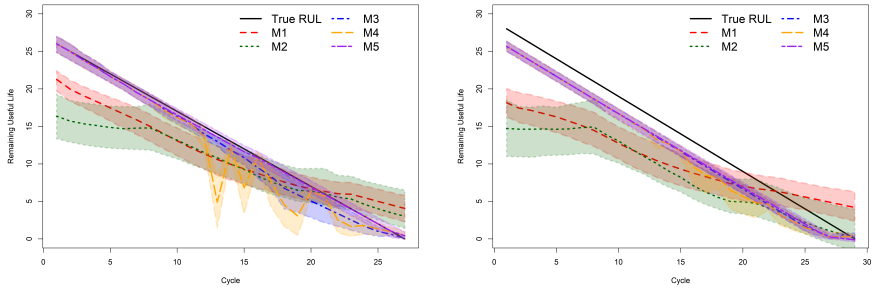


Figure 2.11: Example of the selected features and the corresponding coefficients in the case study.

Table 2.4: RMSE of RUL prediction for the case study.

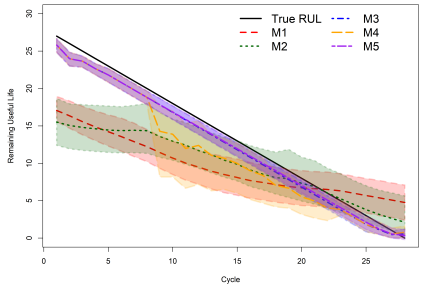
Test motor	M1	M2	M3	M4	M5
Motor 1	**5**.38	10.70	8.44	6.83	8.50
Motor 2	3.50	4.77	1.18	2.57	**0**.30
Motor 3	4.00	4.05	**1**.01	1.17	1.17
Motor 4	5.53	6.71	2.18	2.49	**2**.13
Motor 5	6.17	5.50	1.23	2.54	**1**.09
Motor 6	3.45	5.73	0.96	2.88	**0**.33
Motor 7	4.81	4.58	3.45	4.09	**1**.50
Motor 8	2.95	3.72	**1**.06	1.75	1.76
Mean	4.47	5.72	2.44	3.04	**2**.10

Table 2.4 and Figure 2.12 indicate that, in general, the proposed methods (M3 and M5) outperform the existing methods (M1 and M2), as evidenced by their lower RMSE values and narrower prediction intervals. While the dynamic method (M4) does not consistently improve prediction performance compared to the static method (M3), it still provides valuable information,

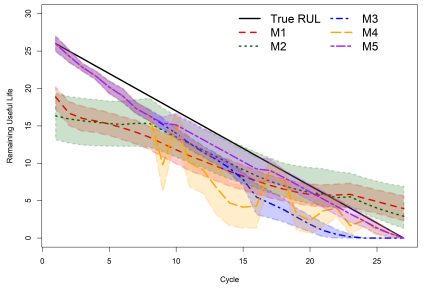


(a) Motor 2

(b) Motor 4



(c) Motor 5



(d) Motor 7

Figure 2.12: RUL predictions and the 95% prediction interval for the case study.

underscoring the importance of the ensemble method (M5). Notably, Table 2.4 and Figure 2.12 demonstrate that M5 achieves the best performance in terms of RMSE in most cases, indicating that the proposed ensemble framework effectively integrates the static and dynamic methods. Further comparisons of the RMSE performance of the proposed methods show that Motor 1 always got the worst RMSE performance compared to other motors, and this may be due to the much shorter failure time of Motor 1 (18 cycles vs. $25 \sim 29$ cycles).

2.7. CONCLUSIONS

THIS chapter presented novel DI-based prognostic frameworks for predicting RUL in complex systems that collect multivariate sensor data. The proposed DI, $Z(t)$ in Eq. (2.2), is feature-based and performs automatic feature selection, which is essential for capturing the underlying degradation trend of the system. Moreover, $Z(t)$ incorporates a nonlinear relationship between the selected features and the degradation process, which reflects the complex and nonlinear nature of the practical engineering applications. Based on the constructed DI, three frameworks were developed for prognostic RUL prediction utilizing various degradation sources. The proposed frameworks do not require prior knowledge of failure mechanisms or specific degradation models, making them applicable to a wide range of engineering systems. The performances of the proposed frameworks were validated through both simulation studies and a case study on the degradation data of 8 three-phase industrial induction motors. The numerical results demonstrate that the proposed prognostic frameworks outperform existing methods by a large margin in terms of predictive accuracy.

3

RELIABILITY ANALYSIS BASED ON THE WIENER PROCESS INTEGRATED WITH HISTORICAL DEGRADATION DATA

Parts of this chapter have been published in *Quality and Reliability Engineering International* 39.4 (2023), pp. 1376–1395. DOI: 10.1002/qre.3300.

PLAIN SUMMARY

For high-reliability and long-life electronic devices, reliability analyses based on degradation data with small sample sizes are important challenges in practice. Recently, increasing attention has been paid to leveraging abundant historical degradation data. Exploiting the information in these historical data efficiently and integrating them to benefit the current reliability analysis is an important issue. This study proposes a new integrating method for the historical and current degradation data based on the Wiener process by considering the consistency of failure mechanisms between the different batches of degradation data. Simulation studies show that the new method leads to superior reliability estimation, and is robust to the assumption of consistent failure mechanisms. Finally, the proposed method is used to analyze a real data set consisting of the metal-oxide-semiconductor field-effect transistor (MOSFET) degradation data.

3.1. INTRODUCTION

IN reliability engineering, it is oftentimes infeasible to analyze reliability based on the exact lifetimes of devices or systems due to their high reliability [8, 48, 49]. Some models based on ADT data have been proposed in the literature to overcome this challenge, including non-stochastic and stochastic processes [8, 50–55]. Compared with non-stochastic processes, the degradation models based on stochastic processes have better application prospects. For example, [56] utilizes the Wiener processes with random effects to model the degradation data, and applies the proposed method to bridge beam data. [57] explores the traditional Wiener process with positive drifts compounded with i.i.d. Gaussian noise and improved estimation efficiency. [58] proposes a general Wiener process-based degradation model to evaluate real-time reliability. In [59], a modified Wiener process is proposed to describe the dynamic, random, and non-linear degradation behavior of LED devices. The review paper [60] summarizes the recent methods for degradation data analysis based on the Wiener process, and introduces their applications in the field of prognostics and health management. Based on the Wiener and gamma processes, [61] models the degradation process under different types of thresholds. Additionally,

[62] employs the Wiener process to truncate normal distribution to model the metal-oxide-semiconductor field-effect transistor (MOSFET) data with both degradation and shock processes. [63] considers the measurement errors and proposes a two-stage degradation model in which the adaptation term also has the characteristics of the Wiener process.

In many applications, such as the experiments for MOSFET used in the intelligent power distribution system, it is hard to obtain enough degradation data from the current experimental batch due to long test cycles and high costs. Assessing reliability based on a small amount of degradation data is a significant challenge. However, additional historical degradation data are often available from the previous experimental batches. Some methods have been proposed in the literature to merge the historical and current data. [64] uses the Wiener process to model the degradation process and proposes a Bayesian method to integrate the historical ADT data from the laboratory with the failure data from the field. [65] considers the degradation data from historical units to determine the parameters in the prior distribution for the gamma process and applies their method to analyze the reliability of the computer numerical control machine tools.

Almost all of these methods assume that the current experimental units have the same degradation characteristics as those of the historical experimental units, which means that all batch degradation data have the same parameters. However, in practice, this assumption is not reasonable [66]. For example, Beijing Spacecrafts Manufacturing Factory uses the same manufacturing process for different batches of MOSFETs. However, in different batches, the raw materials and technicians would not be exactly the same, which will produce changes in degradation parameters. In this situation, the interest of engineers centers around whether the historical data can be employed to improve the accuracy of estimation of reliability for the current batch. This is also the main motivation for our study.

In the literature, there are few works to consider the problem of degradation heterogeneity [67–69] between different batches. In this chapter, a new method based on the Wiener process is proposed to integrate different batches of current and historical data by assuming consistent failure mechanisms. Inspired

by [70] and [71], this chapter uses the variance-to-mean ratio to quantify the consistency of failure mechanisms in the Wiener process. The newly proposed method includes two steps: (1) testing the assumption by using the likelihood ratio and (2) integrating the data if the assumption is accepted significantly. Some simulation studies and real data analyses are conducted to demonstrate the performance of the new method. The main contributions of this chapter are summarized as follows.

(1) Based on the ratio of the squared diffusion and drift parameters in the Wiener degradation process, a novel method is proposed to model different batches of data by considering the degradation heterogeneity.

(2) A failure mechanism consistency test for the Wiener process is proposed based on the likelihood ratio.

(3) A new integrated framework based on the Wiener process is proposed to fuse different batches of current and historical degradation data.

The remainder of this chapter is organized as follows. Section 3.2 shows the degradation model and parameter estimation based on the Wiener process. The likelihood ratio test for failure mechanism consistency is presented in Section 3.3. The newly integrated framework for reliability analysis based on the Wiener process is developed in Section 3.4. Simulation analyses demonstrating the feasibility and effectiveness of the proposed method are presented in Section 3.5. In Section 3.6, a case study based on real-world MOSFET degradation data is provided. Some conclusions and discussions are given in Section 3.7.

3.2. DEGRADATION MODEL AND PARAMETER ESTIMATION

IN the literature, the Wiener process is commonly used in the degradation models due to its mathematical properties and physical interpretations [57], which can be formulated as,

$$X(t_i) = X(t_1) + \mu\Lambda(t_i) + \sigma B(\Lambda(t_i)), i = 2, \cdots, n, \tag{3.1}$$

where n is the number of collected data points, $X(t_1)$ is the initial degradation value at the first time point, μ is the drift parameter, σ is the diffusion parameter, $B(\cdot)$ is standard Brownian motion, and $\Lambda(t)$ is a positive non-decreasing function known as the transformed time scale, which is commonly employed to capture potential curvature in the degradation path [72]. $\Lambda(t) = \Lambda(t, \boldsymbol{\theta})$ is a deterministic function with parameter θ, which is used to describe potential nonlinear degradation [34, 57]. For example, $\Lambda(t)$ can be linear $\Lambda(t) = t$ or a power-law form $\Lambda(t) = t^{\theta}, \theta \geq 0$.

Due to the independent increments of the Wiener process, it is easy to obtain,

$$\Delta X_i = X(t_{i+1}) - X(t_i) \overset{\text{i.i.d}}{\sim} N(\mu \Delta\Lambda_i, \sigma^2 \Delta\Lambda_i), \tag{3.2}$$

where $\Delta\Lambda_i = \Lambda(t_{i+1}) - \Lambda(t_i), i = 1, \cdots, n-1$ represents the time increment for the collected data. The maximum likelihood estimators (MLEs) $\hat{\mu}, \hat{\sigma}$ of parameters μ, σ are calculated as,

$$\hat{\mu} = \frac{\sum_{i=1}^{n-1} \Delta X_i}{\sum_{i=1}^{n-1} \Delta\Lambda_i}, \quad \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} \frac{(\Delta X_i - \hat{\mu}\Delta\Lambda_i)^2}{\Delta\Lambda_i}}, \tag{3.3}$$

by maximizing,

$$\begin{aligned} \ell(\mu, \sigma | \boldsymbol{\theta}) &= \ln \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2 \Delta\Lambda_i}} \exp\left(-\frac{(\Delta X_i - \mu\Delta\Lambda_i)^2}{2\sigma^2 \Delta\Lambda_i} \right) \\ &= \sum_{i=1}^{n-1} \left[-\frac{1}{2} \ln(2\pi\sigma^2 \Delta\Lambda_i) \ - \frac{(\Delta X_i - \mu\Delta\Lambda_i)^2}{2\sigma^2 \Delta\Lambda_i} \right]. \end{aligned} \tag{3.4}$$

For the case when $\boldsymbol{\theta}$ in $\Lambda(t, \boldsymbol{\theta})$ is unknown, the profile log-likelihood method [34] is employed to estimate the $\boldsymbol{\theta}$. Specifically, for any given $\boldsymbol{\theta}$, the estimators $(\hat{\mu}, \hat{\sigma})$ could be formulated as functions of $\boldsymbol{\theta}$, denoted as $(\hat{\mu}, \hat{\sigma}) = (\hat{\mu}(\boldsymbol{\theta}), \hat{\sigma}(\boldsymbol{\theta}))$. By substituting the estimators into Eq. (3.4), the log-likelihood function of $\boldsymbol{\theta}$ can be derived as $\ell(\boldsymbol{\theta}) = \ell((\hat{\mu}(\boldsymbol{\theta}), \hat{\sigma}(\boldsymbol{\theta}))|\boldsymbol{\theta})$. Then, the estimator for $\boldsymbol{\theta}$ can be calculated by maximizing the profile log-likelihood, which can be formulated as,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell((\hat{\mu}(\boldsymbol{\theta}), \hat{\sigma}(\boldsymbol{\theta}))|\boldsymbol{\theta}). \tag{3.5}$$

3.3. TEST FOR FAILURE MECHANISM CONSISTENCY

As the sample in the current batch of an experiment is sometimes very small, the MLEs of parameters could be inaccurate. To improve the accuracy of MLEs, the historical data, which is often sufficiently available in practice, are considered for data fusion. Usually, the degradation parameters are not the same in different batches of data, but the failure mechanisms could still be consistent. Here, we present a method to test the consistency of failure mechanisms between different batches. Then, a natural fusion method is proposed based on the assumption of consistent failure mechanisms.

According to [70] and [71], for Wiener process models, the consistent failure mechanism equivalents to a consistent ratio of the squared diffusion and drift parameters, i.e., ensuring the consistency of variance-to-mean-ratio in the increments of degradation data, $\frac{\sigma^2\Delta\Lambda_i}{\mu\Delta\Lambda_i}=\frac{\sigma^2}{\mu}$.

Let $\Delta\Lambda_{ci}$ represent the increment of time for the current data, where the subscript c denotes the current batch data, and let $X_c(t_{c_i})$ be the corresponding degradation observations, which follows Eq. (3.1) with drift parameter μ_c and diffusion parameter σ_c. Let $\Delta\Lambda_{hj}$ be the increment of time for the historical data and $X_h(t_{h_j}), j=1,\cdots,n_h$ be the historical degradation observations with different parameters μ_h, σ_h. Test the same failure mechanisms is equivalent to test $\frac{\sigma_c^2}{\mu_c}=\frac{\sigma_h^2}{\mu_h}=k$, which can be formulated as,

$$H_0: k_c=k_h=k \text{ versus } H_1: k_c\neq k_h \tag{3.6}$$

where $k_c=\frac{\sigma_c^2}{\mu_c}, k_h=\frac{\sigma_h^2}{\mu_h}$.

The log-likelihood function under H_0 and H_1 can be derived as

$$\begin{aligned}\ell(\mu_c,\sigma_c,\mu_h,\sigma_h) &= \ell(\mu_c,\sigma_c)+\ell(\mu_h,\sigma_h)\\ &= \sum_{i=1}^{n_c-1}\left[-\frac{1}{2}\ln(2\pi\sigma_c^2\Delta\Lambda_{ci})-\frac{(\Delta X_{ci}-\mu_c\Delta\Lambda_{ci})^2}{2\sigma_c^2\Delta\Lambda_{ci}}\right]\\ &\quad+\sum_{j=1}^{n_h-1}\left[-\frac{1}{2}\ln(2\pi\sigma_h^2\Delta\Lambda_{hj})-\frac{(\Delta X_{hj}-\mu_h\Delta\Lambda_{hj})^2}{2\sigma_h^2\Delta\Lambda_{hj}}\right],\end{aligned} \tag{3.7}$$

and

$$\begin{aligned}\ell(\mu_c,\mu_h,k) = &\sum_{i=1}^{n_c-1}\left[-\frac{1}{2}\ln(2\pi k\mu_c\Delta\Lambda_{ci}) - \frac{(\Delta X_{ci}-\mu_c\Delta\Lambda_{ci})^2}{2k\mu_c\Delta\Lambda_{ci}}\right] \\ &+\sum_{j=1}^{n_h-1}\left[-\frac{1}{2}\ln(2\pi k\mu_h\Delta\Lambda_{hj}) - \frac{(\Delta X_{hj}-\mu_h\Delta\Lambda_{hj})^2}{2k\mu_h\Delta\Lambda_{hj}}\right],\end{aligned} \tag{3.8}$$

respectively. Denote the parameter estimators based on Eqs. (3.7) and (3.8) as $\{\hat{\mu}_c,\hat{\sigma}_c,\hat{\mu}_h,\hat{\sigma}_h\}$ and $\{\tilde{\mu}_c,\tilde{\mu}_h,\tilde{k}_h\}$. Then, the corresponding maximum log-likelihood function $\hat{\ell}(\hat{\mu}_c,\hat{\sigma}_c,\hat{\mu}_h,\hat{\sigma}_h)$ and $\tilde{\ell}(\tilde{\mu}_c,\tilde{\mu}_h,\tilde{k})$ can be simplified as,

$$\begin{aligned}\hat{\ell}(\hat{\mu}_c,\hat{\sigma}_c,\hat{\mu}_h,\hat{\sigma}_h) = &\sum_{i=1}^{n_c-1}\left[-\frac{1}{2}\ln(2\pi\hat{\sigma}_c^2\Delta\Lambda_{ci}) - \frac{(\Delta X_{ci}-\hat{\mu}_c\Delta\Lambda_{ci})^2}{2\hat{\sigma}_c^2\Delta\Lambda_{ci}}\right] \\ &+\sum_{j=1}^{n_h-1}\left[-\frac{1}{2}\ln(2\pi\hat{\sigma}_h^2\Delta\Lambda_{hj}) - \frac{(\Delta X_{hj}-\hat{\mu}_h\Delta\Lambda_{hj})^2}{2\hat{\sigma}_h^2\Delta\Lambda_{hj}}\right],\end{aligned} \tag{3.9}$$

$$\begin{aligned}\tilde{\ell}(\tilde{\mu}_c,\tilde{\mu}_h,\tilde{k}) = &\sum_{i=1}^{n_c-1}\left[-\frac{1}{2}\ln(2\pi\tilde{k}\tilde{\mu}_c\Delta\Lambda_c) - \frac{(\Delta X_{ci}-\tilde{\mu}_c\Delta\Lambda_c)^2}{2\tilde{k}\tilde{\mu}_c\Delta\Lambda_c}\right] \\ &+\sum_{j=1}^{n_h-1}\left[-\frac{1}{2}\ln(2\pi\tilde{k}\tilde{\mu}_h\Delta\Lambda_h) - \frac{(\Delta X_{hj}-\tilde{\mu}_h\Delta\Lambda_h)^2}{2\tilde{k}\tilde{\mu}_h\Delta\Lambda_h}\right].\end{aligned} \tag{3.10}$$

Here, the likelihood ratio test method is employed and the corresponding statistic W can be derived as,

$$W = 2\left[\hat{\ell}(\hat{\mu}_c,\hat{\sigma}_c,\hat{\mu}_h,\hat{\sigma}_h) - \tilde{\ell}(\tilde{\mu}_c,\tilde{\mu}_h,\tilde{k}_h)\right]. \tag{3.11}$$

The following Theorem 3.1 shows that the asymptotic distribution of the test statistic W under H_0 is $\chi^2(1)$. Thus, the null hypothesis H_0 is not rejected when $W < \chi^2_{1-\alpha}(1)$, where $\chi^2_{1-\alpha}(1)$ is the $(1-\alpha)$ quantile of $\chi^2(1)$. Otherwise, the null hypothesis is rejected.

Theorem 3.1. *Assume the mild regularity conditions hold (see Lemma 3.1). Under $H_0: k_c = k_h = k$, the test statistic W in Eq. (3.11) satisfies $W \xrightarrow{d} \chi^2(1)$.*

Lemma 3.1. *Let $X_1, X_2, \cdots, X_n$ be i.i.d from $N(\mu, k\mu)$ and the corresponding probability density function is f_θ with respect to a σ-finite measure ν on $(\mathscr{R},\mathscr{B})$,*

where $\mathscr{R}$ is the real line, $\mathscr{B}$ is the Borel σ-field, $\theta = (\mu, k) \in \Theta$ and Θ is an open set in $\mathscr{R}^2$. Then, the following regularity conditions are satisfied.

(a) For every x, $f_\theta(x)$ is twice continuously differentiable in θ.

(b) Let $f_{\theta_i}(x), i = 1,2, \theta_1 = \mu, \theta_2 = k$ satisfy

$$\frac{\partial}{\partial \theta_i} \int \psi_{\theta_i}(x) d\nu = \int \frac{\partial}{\partial \theta_i} \psi_{\theta_i}(x) d\nu$$

for $\psi_{\theta_i}(x) = f_{\theta_i}(x)$ and $= \partial f_{\theta_i}(x)/\partial \theta_i$, where ν is a σ-finite measure.

(c) The Fisher information matrix

$$I_1(\theta) = E\left\{ \frac{\partial}{\partial \theta} \ln f_\theta(X_1) \left[\frac{\partial}{\partial \theta} \ln f_\theta(X_1) \right]^T \right\}$$

is positive definite.

(d) For any given $\theta \in \Theta$, there exists a positive number c_θ and a positive function h_θ such that $E[h_\theta(X_1)] < +\infty$ and

$$\sup_{\gamma : \|\gamma - \theta\| < c_\theta} \left\| \frac{\partial^2 \ln f_\gamma(x)}{\partial \gamma^2} \right\| \le h_\theta(x)$$

for all x, where $\|A\| = \sqrt{\mathrm{tr}(A^T A)}$ for any matrix A.

The proof of Theorem 3.1 can directly follow Theorem 6.5 in [73] when the mild regularity conditions in Lemma 3.1 hold. The proof of Lemma 3.1 can be found in APPENDIX A.1.

3.4. RELIABILITY ANALYSIS BASED ON INTEGRATED DATA

ACCORDING to Section 3.3, for the degradation increments of the current and the historical degradation data, we have,

$$\Delta X_{ci} \sim N(\mu_c \Delta\Lambda_{ci}, \sigma_c^2 \Delta\Lambda_{ci}), \Delta X_{hj} \sim N(\mu_h \Delta\Lambda_{hj}, \sigma_h^2 \Delta\Lambda_{hj})).$$

If only the current data are considered, the estimators can be derived as,

$$\hat{\mu}_c = \frac{\sum_{i=1}^{n_c-1} \Delta X_{ci}}{\sum_{i=1}^{n_c-1} \Delta \Lambda_{ci}}, \hat{\sigma}_c = \sqrt{\frac{1}{n_c-1} \sum_{i=1}^{n_c-1} \frac{(\Delta X_{ci} - \hat{\mu}_c \Delta \Lambda_{ci})^2}{\Delta \Lambda_{ci}}}. \tag{3.12}$$

However, if the assumption that the failure mechanisms of the current and historical data are consistent has not been rejected, the estimators $\tilde{\mu}$ and $\tilde{k}$ can be derived by maximizing Eq. (3.8) and then we have $\tilde{\sigma} = \sqrt{\tilde{k}\tilde{\mu}}$.

Thus, the final estimators can be described as

$$\begin{aligned} \breve{\mu}_c &= I(W)\tilde{\mu}_c + (1 - I(W))\hat{\mu}_c, \\ \breve{\sigma}_c &= I(W)\tilde{\sigma}_c + (1 - I(W))\hat{\sigma}_c, \end{aligned}$$

where $I(W)$ is an indicator function for the test statistic W, satisfies

$$I(W) = \begin{cases} 0 & W > \chi^2_{1-\alpha}(1), \\ 1 & W \le \chi^2_{1-\alpha}(1). \end{cases} \tag{3.13}$$

In practice, $\tilde{\mu}_c, \tilde{\sigma}_c$ are used as the estimators when engineers confirm that the failure mechanisms are consistent for different batches of degradation data, which is a special case of $\breve{\mu}_c, \breve{\sigma}_c$ with $I(W) \equiv 1$.

Given the relative failure threshold Q, the time-to-failure T can be described as,

$$T = \inf(t | X(t) \ge Q), \tag{3.14}$$

and the corresponding reliability function can be formulated as,

$$R(t) = 1 - P(T < t) = 1 - \Phi\left(\frac{\mu\Lambda(t) - Q}{\sigma\sqrt{\Lambda(t)}}\right) - \exp\left(\frac{2\mu Q}{\sigma^2}\right)\Phi\left(-\frac{\mu\Lambda(t) + Q}{\sigma\sqrt{\Lambda(t)}}\right). \tag{3.15}$$

Then, we can estimate the reliability function as,

$$\breve{R}(t) = 1 - \Phi\left(\frac{\breve{\mu}_c\Lambda(t) - Q}{\breve{\sigma}_c\sqrt{\Lambda(t)}}\right) - \exp\left(\frac{2\breve{\mu}Q}{\breve{\sigma}_c^2}\right)\Phi\left(-\frac{\breve{\mu}_c\Lambda(t) + Q}{\breve{\sigma}_c\sqrt{\Lambda(t)}}\right). \tag{3.16}$$

Algorithm 3.1 describes the steps for the analyses of reliability following the

proposed integrated method.

Algorithm 3.1: Framework of the proposed integration method

Input: The current and historical degradation data

1: Model the current and historical data with the Wiener process, and derive the estimated parameters $\{\hat{\mu}_c, \hat{\sigma}_c, \hat{\mu}_h, \hat{\sigma}_h\}$ and $\{\tilde{\mu}_c, \tilde{\mu}_h, \tilde{k}\}$ from Eqs. (3.7) and (3.8).
2: Calculate $\tilde{\sigma}_c^2 = \tilde{\mu}_c \times \tilde{k}$, $\tilde{\sigma}_h^2 = \tilde{\mu}_h \times \tilde{k}$ and $W = 2\left[\hat{\ell}(\hat{\mu}_c, \hat{\sigma}_c, \hat{\mu}_h, \hat{\sigma}_h) - \tilde{\ell}(\tilde{\mu}_c, \tilde{\mu}_h, \tilde{k})\right]$.
3: **if** the failure mechanism consistency is confirmed **then**
4: Using $\tilde{\mu}_c, \tilde{\sigma}_c$ as the estimators;
5: **else if** $W > \chi^2_{1-\alpha}(1)$, where α is the given significance level, **then**
6: Using $\hat{\mu}_c, \hat{\sigma}_c$ as the estimators;
7: **else**
8: Using $\tilde{\mu}_c, \tilde{\sigma}_c$ as the estimators.
9: **end if**
10: Calculate the reliability function of the current data with the derived drift and diffusion parameters.

Output: Estimated reliability function for the current data.

3.5. SIMULATION STUDIES

IN this section, several simulation studies are conducted to investigate the performance of the proposed integrated method.

During the simulation studies, the performances of the proposed integrated method under the following two assumptions are included: 1) the linear Wiener model, where $\Lambda(t) = t$, and 2) the Wiener model with a widely used transformed time scale [34, 57, 74], where $\Lambda(t) = \Lambda(t, \boldsymbol{\theta}) = t^{\theta}$ with $\theta = \frac{1}{2}$.

The following four methods are used to estimate the drift and diffusion parameters for the current data.

Firstly, the estimators $\hat{\mu}_c$, $\hat{\sigma}_c$ in Eq. (3.12) can be computed when only the current degradation data are considered, and we denote this method as M1. When considering both the current and historical degradation data, it is often assumed that these data follow the same distribution [58, 65, 75], and the estimators using these data are denoted as M2. This study proposes two methods to integrate historical and current data. The method directly based on

the assumption of failure mechanism consistency is denoted as M3, and the method considering both the failure mechanism consistency test and M3 is denoted as M4.

3.5.1. PERFORMANCE OF THE LINEAR WIENER MODEL

LET $X_h(0) = X_c(0) = 1$, $\Lambda(t) = t$, $Q = 1.2$, $t_h = 0 \cdot \Delta\Lambda_h$, $1 \cdot \Delta\Lambda_h, \cdots, (n_h - 1) \cdot \Delta\Lambda_h$, $t_c = 0 \cdot \Delta\Lambda_c$, $1 \cdot \Delta\Lambda_c, \cdots, (n_c - 1) \cdot \Delta\Lambda_c$, $n_h = 100$, $n_c = \{10, 20, \cdots, 100\}$, respectively. The significance level α is set to be 0.1 in these simulations.

In these simulations, the mean squared error (MSE) between estimated and true values of reliability functions is considered as the metric. The MSE can be formulated as,

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{R}(t_i) - R(t_i) \right)^2, \tag{3.17}$$

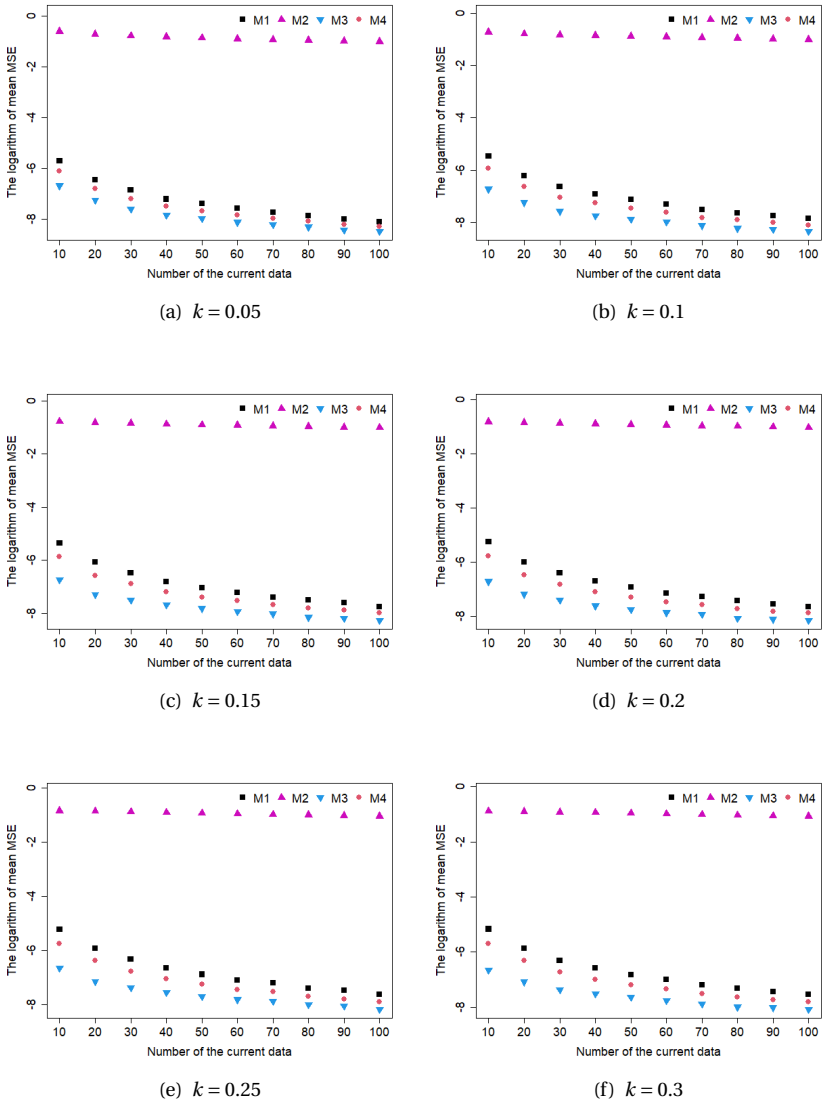
where $\hat{R}(t_i), R(t_i)$ are the estimated and the true reliability function at the time point t_i, $i = 1, 2, \cdots, n$, respectively.

Table 3.1 shows the different setting parameters μ_h, σ_h, μ_c and σ_c used in the simulations. To evaluate the accuracy of the estimated reliability using the above four methods, the simulated results based on 5000 repetitions are shown in Figure 3.1 and Figure 3.2.

Table 3.1: Different settings for k.

Setting	μ_h	σ_h^2	μ_c	σ_c^2	k
1	1.8	0.09	0.2	0.01	0.05
2	0.9	0.09	0.1	0.01	0.1
3	0.6	0.09	0.067	0.01	0.15
4	0.45	0.09	0.05	0.01	0.2
5	0.36	0.09	0.04	0.01	0.25
6	0.3	0.09	0.033	0.01	0.3

For better visualization of the results, we use the log scale to present. The mean value of 5000 MSEs is shown in Figure 3.1 on the log scale under different settings according to Table 3.1. In Figure 3.1, the black square, the purple triangle point-up, the blue point-down, and the red circle are the mean values

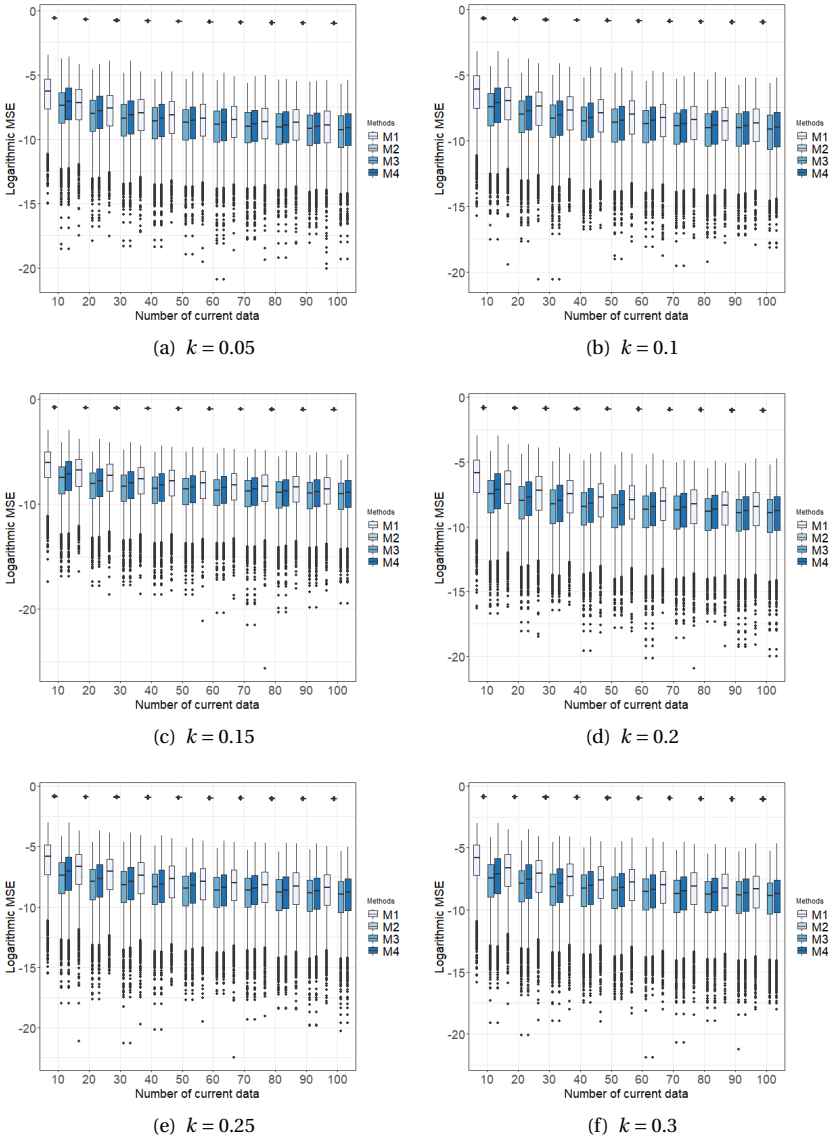


(a) $k = 0.05$ (b) $k = 0.1$

(c) $k = 0.15$ (d) $k = 0.2$

(e) $k = 0.25$ (f) $k = 0.3$

Figure 3.1: Mean value of MSEs on the log scale of different models under the setting parameters in Table 3.1, where (a) $k = 0.05$; (b) $k = 0.1$; (c) $k = 0.15$; (d) $k = 0.2$; (e) $k = 0.25$; (f) $k = 0.3$.



(a) $k = 0.05$ (b) $k = 0.1$

(c) $k = 0.15$ (d) $k = 0.2$

(e) $k = 0.25$ (f) $k = 0.3$

Figure 3.2: The box plots of logarithmic MSE of different models under the settings parameters in Table 3.1, where (a) $k = 0.05$; (b) $k = 0.1$; (c) $k = 0.15$; (d) $k = 0.2$; (e) $k = 0.25$; (f) $k = 0.3$.

of 5000 MSEs on the log scale derived from M1, M2, M3, and M4, respectively. The box plots of 5000 ln(MSE) is given in Figure 3.2.

Figure 3.1 and Figure 3.2 suggest that, in these simulations, the MSE of M2 is significantly larger than the other three methods when $k_c = k_h = k$, which also confirms that it is not suitable to assume that the current and historical degradation data follow the same distribution. From these figures, we also find that, for different amounts of current data, the proposed M3 and M4 perform better than M1 and M2 consistently due to the smaller mean values and the narrower boxes. As the number of current data decreases, this trend is clearer.

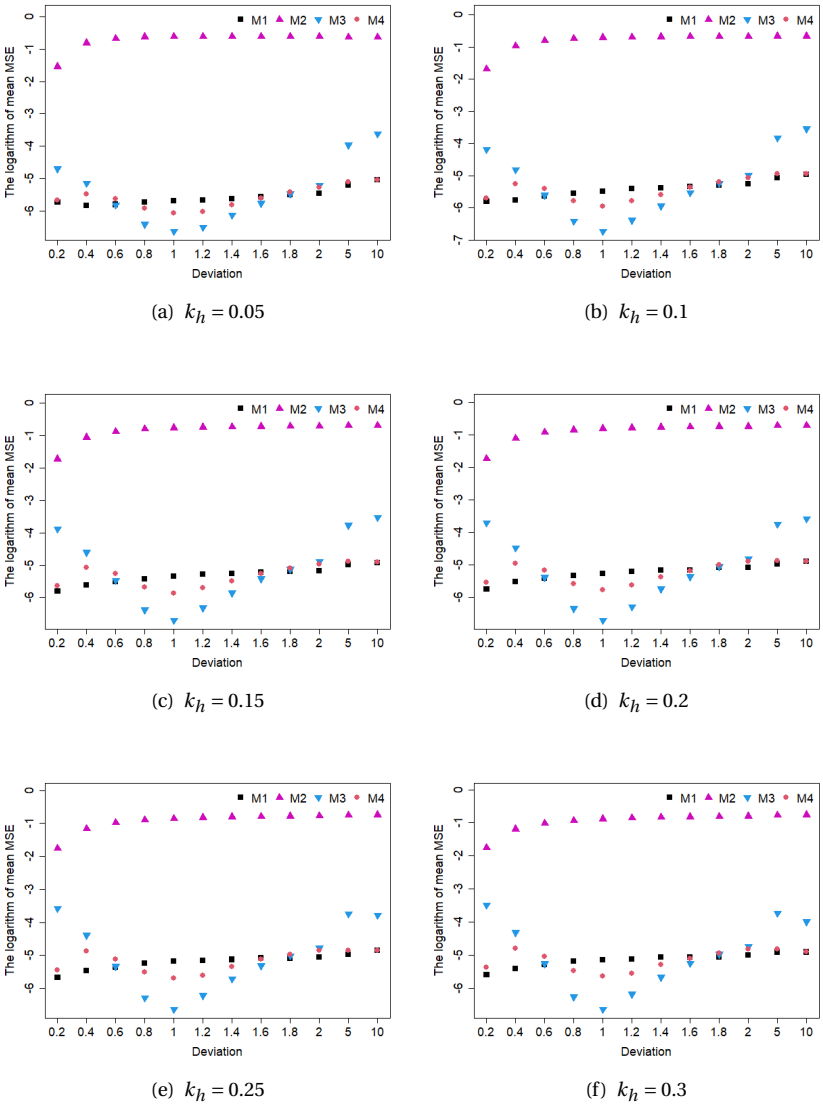
The following simulations are conducted to demonstrate the necessity of the proposed failure mechanism consistency test. Table 3.2 shows the settings for parameters used in this simulation, and the significance level $\alpha = 0.1$. For each parameter combination, 5000 repetitions are used to calculate the MSE metric. Considering the case where there are only 10 current data as an example, Figure 3.3 and Figure 3.4, where the x-axis represents the deviation $\frac{k_c}{k_h}$, show the simulated results.

Table 3.2: Different settings for $\frac{k_c}{k_h}$ and k_h.

Setting	σ_h^2	σ_c^2	$\frac{k_c}{k_h}$	k_h
1	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.05
2	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.1
3	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.15
4	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.2
5	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.25
6	0.09	0.01	{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 5, 10}	0.3

The mean value of 5000 MSEs is shown in Figure 3.3 on the log scale under different settings according to Table 3.2. In Figure 3.3, the black square, the purple triangle point-up, the blue point-down, and the red circle are the mean values of 5000 MSEs on the log scale derived from M1, M2, M3, and M4, respectively. The box plots of 5000 ln(MSE) is given in Figure 3.4.

Figure 3.3 and Figure 3.4 also suggest that the MSE of M2 is significantly larger than the other three methods when $k_c \neq k_h$. From these figures, we also

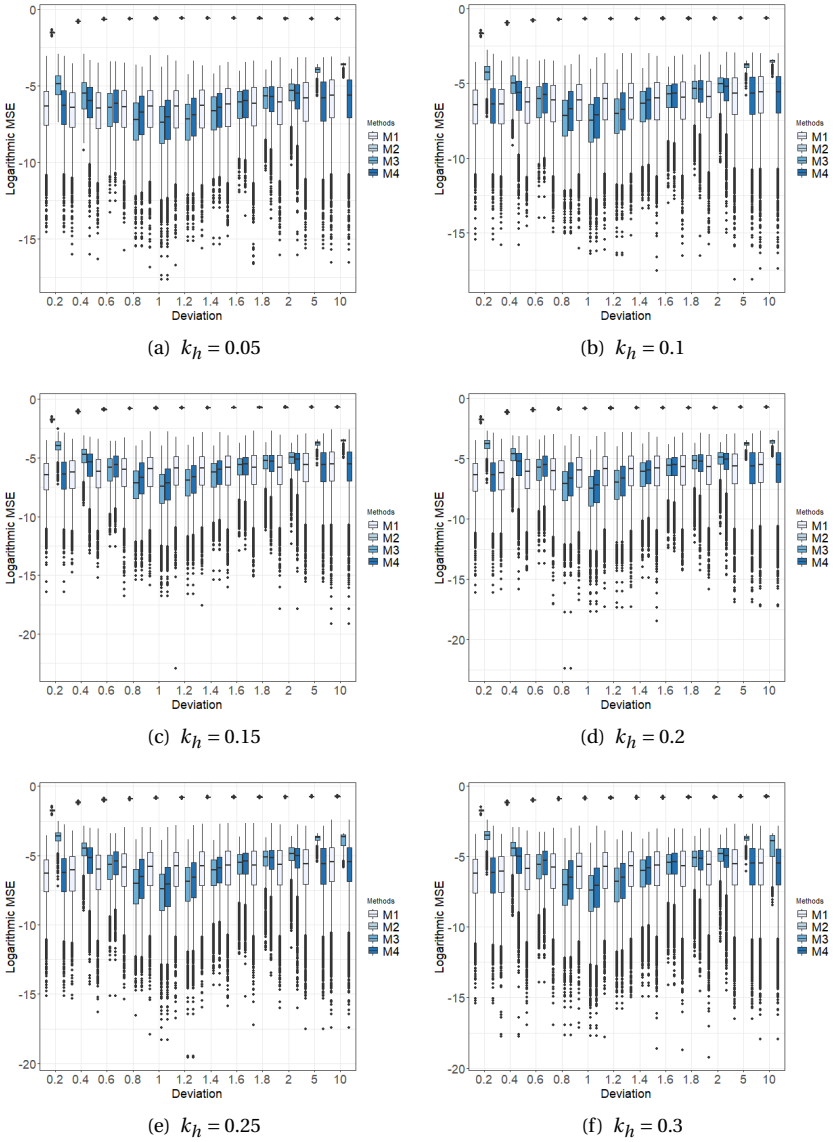


(a) $k_h = 0.05$ (b) $k_h = 0.1$

(c) $k_h = 0.15$ (d) $k_h = 0.2$

(e) $k_h = 0.25$ (f) $k_h = 0.3$

Figure 3.3: Mean value of MSEs on the log scale of different models under the setting parameters in Table 3.2, where (a) $k_h = 0.05$; (b) $k_h = 0.1$; (c) $k_h = 0.15$; (d) $k_h = 0.2$; (e) $k_h = 0.25$; (f) $k_h = 0.3$.



(a) $k_h = 0.05$ (b) $k_h = 0.1$

(c) $k_h = 0.15$ (d) $k_h = 0.2$

(e) $k_h = 0.25$ (f) $k_h = 0.3$

Figure 3.4: The box plots logarithmic MSEs of different models under the setting parameters in Table 3.2, where (a) $k_h = 0.05$; (b) $k_h = 0.1$; (c) $k_h = 0.15$; (d) $k_h = 0.2$; (e) $k_h = 0.25$; (f) $k_h = 0.3$.

find that, in general, M3 will have a catastrophic effect on the estimation accuracy of reliability when the true value of k_c is far away from k_h. Fortunately, in these extreme cases, M4 improves M3 quite a lot, which also confirms the validity and necessity of the proposed failure mechanism consistency test.

3.5.2. PERFORMANCE OF A TRANSFORMED TIME-SCALE WIENER MODEL

IN this section, we construct several simulations on the Wiener process with the transformed time scale, i.e., $\Lambda(t) = \Lambda(t, \boldsymbol{\theta}) = t^{\theta}, \theta = \frac{1}{2}$. The rest of the parameter settings are consistent with Section 3.5.1. The results are similar with Section 3.5.1, which confirms that the proposed methods can also be applied to the transformed time scale Wiener process. Corresponding results Figure B1 to Figure B4 are provided in APPENDIX A.2.

Combined with the results in Section 3.5.1, these simulation studies show that the proposed integrated method outperforms existing methods in both linear and transformed time-scale Wiener models. These results also imply that the proposed integrated methods M3 and M4 could improve the accuracy and stability of the reliability estimation. Besides, our methods show superior performance when the current data are of a small sample size.

Specifically, M2 performs poorly in all cases because the drift and diffusion parameters in these simulations are quite different. It is not reasonable to directly assume that these data follow the same distribution. When the difference between k_c and k_h is small, M3 and M4 outperform M1, which means that the proposed integrated method is effective. In these cases, M3 is slightly better than M4 due to the type I error [76] in the failure mechanism consistency test, but M4 is still comparable. However, M3 can be disastrous when k_c and k_h are significantly different, and in extreme cases, M4 and M1 are more accurate and have comparable results.

In practice, the M4 is recommended when there is no prior knowledge about the differences in failure mechanisms between different batches. Otherwise, the M3 should be selected.

3.6. CASE STUDY

A real-world case study on MOSFET is presented to demonstrate the implementation of the proposed method.

The data used in this section are the ADT data from different batches of a type of MOSFET used in the design of Chinese Tiangong aircraft. The data are provided by the Beijing Spacecrafts Manufacturing Factory. The engineers wanted to know whether the degradation data from the historical MOSFETs explored earlier using the different batches of raw materials and by different technicians could be used to analyze the reliability of the new ones.

In this chapter, the on-resistance is adopted as the degradation indicator follows [77] and [78], and the linear Wiener process model is considered in this case. The failure threshold Q is 1.2 times the initial degradation value. Table 3.3 shows the historical and current data, where T_h, X_h, T_c, X_c are the testing hours and corresponding degradation data for historical and current batches. The test temperatures for different batches are all 150°C.

The drift and diffusion parameters are estimated from the data in Table 3.3 by Eq. (3.3), and the estimators are denoted as $\hat{\mu}_{c0}, \hat{\sigma}_{c0}, \hat{\mu}_{h0}, \hat{\sigma}_{h0}$. The estimated values are $\hat{\mu}_{c0} = 9.66 \times 10^{-5}, \hat{\sigma}_{c0} = 1.08 \times 10^{-3}, \hat{\mu}_{h0} = 4.21 \times 10^{-5}$, and $\hat{\sigma}_{h0} = 1.92 \times 10^{-3}$. Then, the value of test statistic W in Eq. (3.11) is calculated as $2.24 < 2.71 = \chi^2_{0.9}(1)$, which means that the data in Table 3.3 do not reject the null hypothesis test H_0 in Eq. (3.6) at the significance level $\alpha = 0.1$. Thus, these data can be fused according to the proposed integrated method.

Table 3.4 shows the estimated results (based on MLE), where $\hat{\mu}_{c1}$, $\hat{\sigma}_{c1}$ are the estimated values using M1, $\hat{\mu}_{c2}$, $\hat{\sigma}_{c2}$ are the estimated values using M2, and $\hat{\mu}_{c3}$, $\hat{\sigma}_{c3}$ are the estimated values using M4. In Table 3.4, we also report the standard deviations (SD), which are obtained by normal asymptotics of maximum likelihood estimators.

Then, we obtain the reliability functions in Eq. (3.18) according to Table 3.4 and Eq. (3.15), where $\hat{R}_{C_1}(t)$, $\hat{R}_{C_2}(t)$ and $\hat{R}_{C_3}(t)$ are the reliability functions derived using M1, M2, and M4 in Section 3.5.1, respectively. In practice, the lower bound of the estimated reliability function is more concerned. All of the unknown parameters are estimated by the MLE. Therefore, the confidence interval for the reliability function can be obtained through normal asymptotics

Table 3.3: The case study data.

T_h (Hours)	X_h	T_h (Hours)	X_h	T_c (Hours)	X_c
0	38.713	1440	40.233	0	38.241
144	39.422	1584	41.551	144	38.855
288	39.220	1728	41.956	288	39.672
432	40.537	1872	41.449	432	39.877
576	39.929	2016	41.247	576	41.104
720	39.828	2160	41.956	720	40.899
864	40.334	2304	40.537		
1008	39.321	2448	42.260		
1152	39.524	2592	41.551		
1296	39.118	2736	43.172		

Table 3.4: The estimated results.

	MLE	SD
$\hat{\mu}_{c_1}$	9.66×10^{-5}	4.01×10^{-5}
$\hat{\sigma}_{c_1}$	1.08×10^{-3}	3.41×10^{-4}
$\hat{\mu}_{c_2}$	5.34×10^{-5}	3.06×10^{-5}
$\hat{\sigma}_{c_2}$	1.80×10^{-3}	2.60×10^{-4}
$\hat{\mu}_{c_3}$	3.80×10^{-5}	3.31×10^{-5}
$\hat{\sigma}_{c_3}$	1.51×10^{-3}	2.51×10^{-6}

and the delta method [34, 73]. The corresponding reliability curves are shown in Figure 3.5, where the dashed red lines and shadow are the estimated mean and 95% lower bound of $\hat{R}_{C_1}(t)$, the dot-dash green lines and shadow are the results of $\hat{R}_{C_2}(t)$, and the solid blue lines and shadow are the results of $\hat{R}_{C_3}(t)$. Figure 3.5 suggests that, compared with the proposed integrated method M4 in this study, M1 and M2 underestimate the reliability of the current degradation data.

$$
\begin{aligned}
\hat{R}_{C_1}(t) &= 1-\Phi\left(\frac{\hat{\mu}_{c_1}t-1.2}{\hat{\sigma}_{c_1}\sqrt{t}}\right)-\exp\left(\frac{2\times\hat{\mu}_{c_1}\times 1.2}{\hat{\sigma}^2_{c_1}}\right)\Phi\left(-\frac{\hat{\mu}_{c_1}t+1.2}{\hat{\sigma}_{c_1}\sqrt{t}}\right),\\
\hat{R}_{C_2}(t) &= 1-\Phi\left(\frac{\hat{\mu}_{c_2}t-1.2}{\hat{\sigma}_{c_2}\sqrt{t}}\right)-\exp\left(\frac{2\times\hat{\mu}_{c_2}\times 1.2}{\hat{\sigma}^2_{c_2}}\right)\Phi\left(-\frac{\hat{\mu}_{c_2}t+1.2}{\hat{\sigma}_{c_2}\sqrt{t}}\right),\\
\hat{R}_{C_3}(t) &= 1-\Phi\left(\frac{\hat{\mu}_{c_3}t-1.2}{\hat{\sigma}_{c_3}\sqrt{t}}\right)-\exp\left(\frac{2\times\hat{\mu}_{c_3}\times 1.2}{\hat{\sigma}^2_{c_3}}\right)\Phi\left(-\frac{\hat{\mu}_{c_3}t+1.2}{\hat{\sigma}_{c_3}\sqrt{t}}\right).
\end{aligned}
\tag{3.18}
$$

With the estimators in Table 3.4, we can also visualize the fitness between the built model and the real data. In this chapter, the bootstrap method is used to visualize fitness, and the fitted results from 5000 repetitions are shown in Figure 3.6, where the dashed red lines and shadow are the estimated mean and 95% confidence interval by using M1, the dot-dash green lines and shadow are the results of M2, and the solid blue lines and shadow are the results of M4, and the points are the first difference of the data in Table 3.3. From Figure 3.6 we can find that the proposed method does integrate historical and current data since M1 only fits well with current data and M2 is too broad.

3.7. CONCLUSIONS

THIS study examines the reliability evaluation for high-reliable and long-life devices with limited degradation data from current experimental units and abundant degradation data from historical experimental units. The Wiener process is used to construct the degradation model, and this study proposes two integrating methods for current and historical degradation data to improve the estimated accuracy for reliability of current data. This study also provides a likelihood ratio test to check the consistency of the failure mechanisms

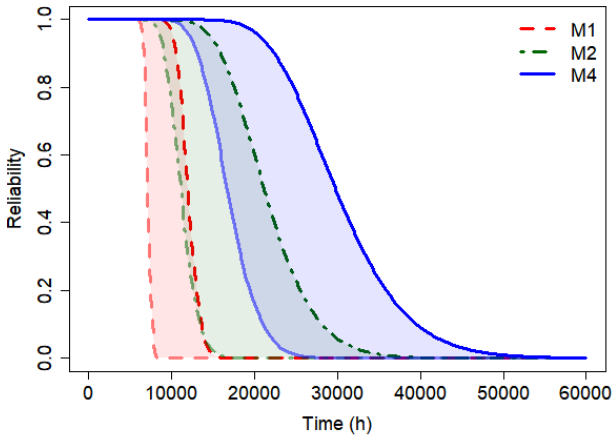


Figure 3.5: Reliability Curves.

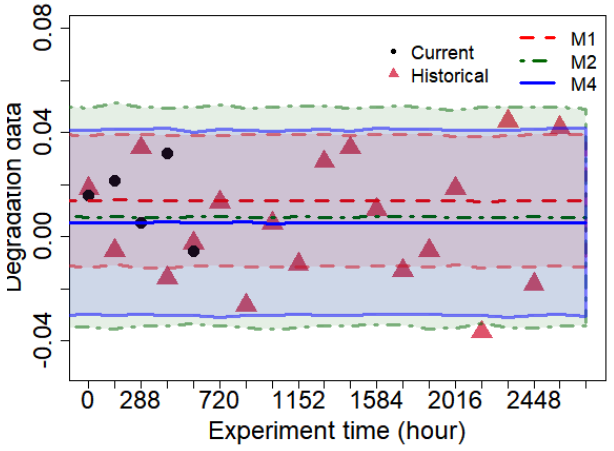


Figure 3.6: Visualization of models fit to the data.

between different batches of devices. Simulation studies show that the proposed integrated method provides accurate and stable estimation results in both linear and transformed time-scale Wiener models. Finally, we have applied the proposed method to real-world degradation data of a type of MOSFET with wide applicability in practical applications like power supply systems and solid-state power controllers.

APPENDIX A

APPENDIX A.1 THE PROOF OF LEMMA 3.1

Let $f_\theta(x)$ be the probability density function of $N(\mu, k\mu)$, which means,

$$\begin{aligned} f_\theta(x) &= \frac{1}{\sqrt{2\pi k\mu}}\exp\left(-\frac{(x-\mu)^2}{2k\mu}\right) \\ &= \exp\left\{\begin{pmatrix} \frac{1}{k} & -\frac{1}{2k\mu} \end{pmatrix}\begin{pmatrix} x \\ x^2 \end{pmatrix} - \left(\frac{\mu}{2k} - \frac{1}{2}\ln(2\pi k\mu)\right)\right\}. \end{aligned} \tag{A.1}$$

Let $\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{k} \\ -\frac{1}{2k\mu} \end{pmatrix}$, $\varphi(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$ and $A(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\ln\left(\frac{\pi}{\eta_2}\right)$, then Eq. (A.1) turns to,

$$f_\eta(x) = \exp\left\{\eta_1 x + \eta_2 x^2 - \left(-\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\ln\left(\frac{\pi}{\eta_2}\right)\right)\right\} = \exp\left(\eta^T\varphi(x) - A(\eta)\right). \tag{A.2}$$

Therefore, $X_i, i = 1, \cdots, n$ has distribution in a natural exponential family and the proof is equivalent to proving that these regularity conditions hold for $f_\eta(x)$.

(a) $f_\eta(x)$ is twice continuously differentiable in η because functions in Eq. (A.3) are continuous functions of η.

$$\begin{cases} \frac{\partial f_\eta(x)}{\partial \eta_1} = f_\eta(x)\left(x - \frac{\partial A(\eta)}{\partial \eta_1}\right) = f_\eta(x)\left(x + \frac{\eta_1}{2\eta_2}\right) \\ \frac{\partial f_\eta(x)}{\partial \eta_2} = f_\eta(x)\left(x^2 - \frac{\partial A(\eta)}{\partial \eta_2}\right) = f_\eta(x)\left(x^2 - \frac{\eta_1^2+2\eta_2}{4\eta_2^2}\right) \\ \frac{\partial^2 f_\eta(x)}{\partial \eta_1^2} = \frac{\partial f_\eta(x)}{\partial \eta_1}\left(x + \frac{\eta_1}{2\eta_2}\right) + f_\eta(x)\frac{1}{2\eta_2} \\ \frac{\partial^2 f_\eta(x)}{\partial \eta_1 \partial \eta_2} = \frac{\partial f_\eta(x)}{\partial \eta_2}\left(x + \frac{\eta_1}{2\eta_2}\right) - f_\eta(x)\frac{\eta_1}{2\eta_2^2} \\ \frac{\partial^2 f_\eta(x)}{\partial \eta_2 \partial \eta_1} = \frac{\partial f_\eta(x)}{\partial \eta_1}\left(x^2 - \frac{\eta_1^2+2\eta_2}{4\eta_2^2}\right) - f_\eta(x)\frac{\eta_1}{2\eta_2^2} \\ \frac{\partial^2 f_\eta(x)}{\partial \eta_2^2} = \frac{\partial f_\eta(x)}{\partial \eta_2}\left(x^2 - \frac{\eta_1^2+2\eta_2}{4\eta_2^2}\right) + f_\eta(x)\frac{\eta_1^2+\eta_2}{2\eta_2^3} \end{cases} \tag{A.3}$$

(b) This is a direct consequence of Theorem 2.7.1 in [79], because $f_\eta(x)$ belongs to a natural exponential family.

(c) According to Proposition 3.2 in [73], $E\left(\varphi(x)\right) = \frac{\partial A(\eta)}{\partial \eta}$.

Then,

$$I_1(\eta) = E\left\{\frac{\partial}{\partial\eta}\ln f_\eta(x)\left[\frac{\partial}{\partial\eta}\ln f_\eta(x)\right]^T\right\} = \mathrm{Var}\left(\varphi(x)\right),$$

as

$$\frac{\partial}{\partial\eta}\ln f_\eta(x) = \varphi(x) - \frac{\partial A(\eta)}{\partial\eta}.$$

Therefore, the Fisher information matrix is positive definite.

(d) For any given η_0,

$$\sup_{\eta:\|\eta-\eta_0\|<c_{\eta_0}}\left\|\frac{\partial^2\ln f_\eta(x)}{\partial\eta^2}\right\| = \sup_{\eta:\|\eta-\eta_0\|<c_{\eta_0}}\left\|-\frac{\partial^2 A(\eta)}{\partial\eta^2}\right\|.$$

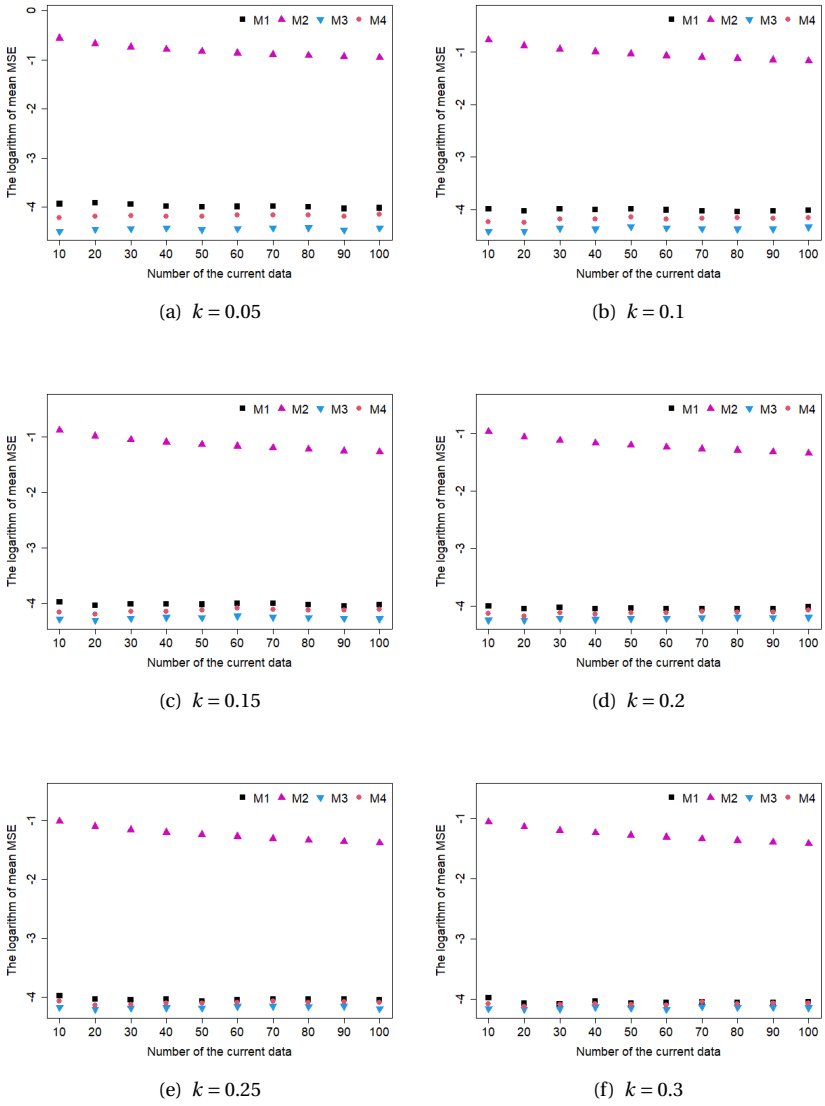
Therefore, this condition is satisfied as $E\left[\sup_{\eta:\|\eta-\eta_0\|<c_{\eta_0}}\left\|-\frac{\partial^2 A(\eta)}{\partial\eta^2}\right\|\right] < +\infty$ holds for any given η_0.

APPENDIX A.2 SIMULATION RESULTS IN SECTION 3.5.2

We provide the simulation results for the employed transformed time-scale Wiener model, i.e., $\Lambda(t) = \Lambda(t,\boldsymbol{\theta})$, $\Lambda(t) = t^\theta, \theta = \frac{1}{2}$. The parameter settings are consistent with Section 3.5.1.

Then, similar to Section 3.5.1, we first construct 5000 repetitions to evaluate the accuracy of the estimated reliability. Figure B1 to Figure B4 show the corresponding simulated results, where the meanings of the legends are consistent with the results in Section 3.5.1.

In summary, the results are similar with Section 3.5.1, which confirms that the proposed methods can also be applied to the transformed time scale Wiener process.

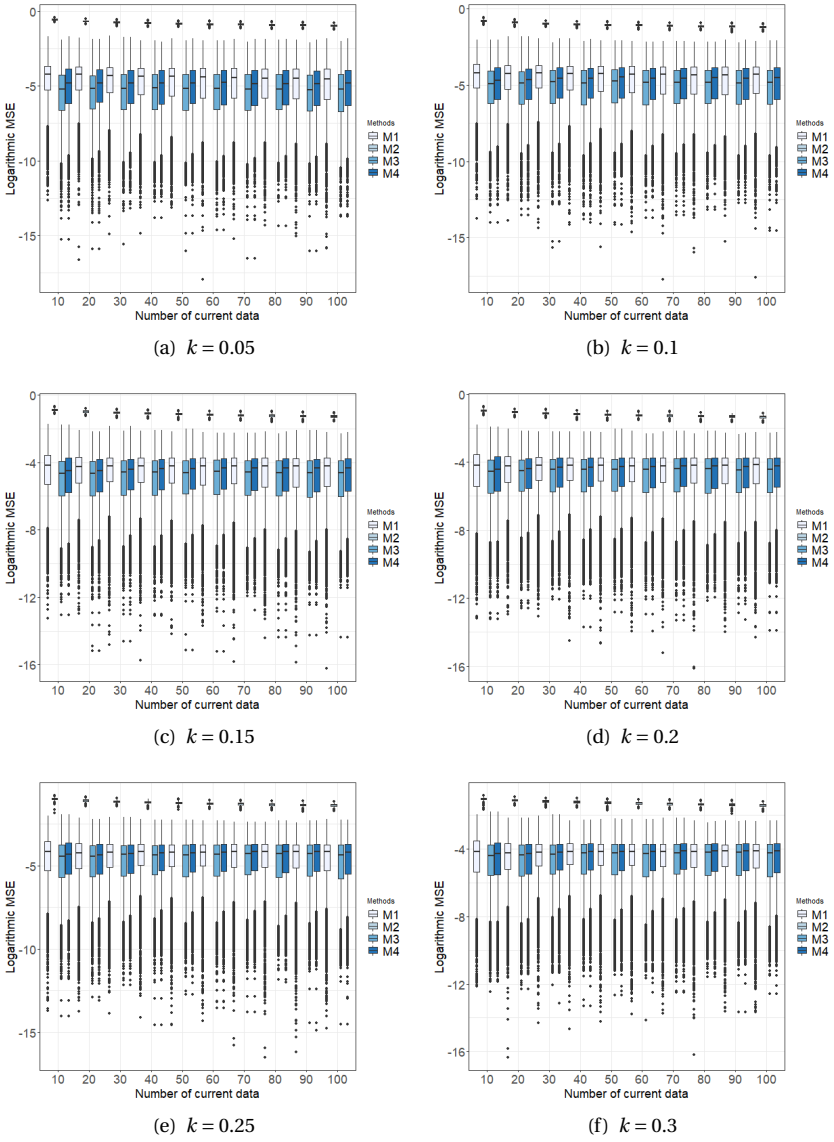


(a) $k = 0.05$ (b) $k = 0.1$

(c) $k = 0.15$ (d) $k = 0.2$

(e) $k = 0.25$ (f) $k = 0.3$

Figure B1: Mean value of MSEs of different models under the setting parameters in Table 3.1, where (a) $k = 0.05$; (b) $k = 0.1$; (c) $k = 0.15$; (d) $k = 0.2$; (e) $k = 0.25$; (f) $k = 0.3$.

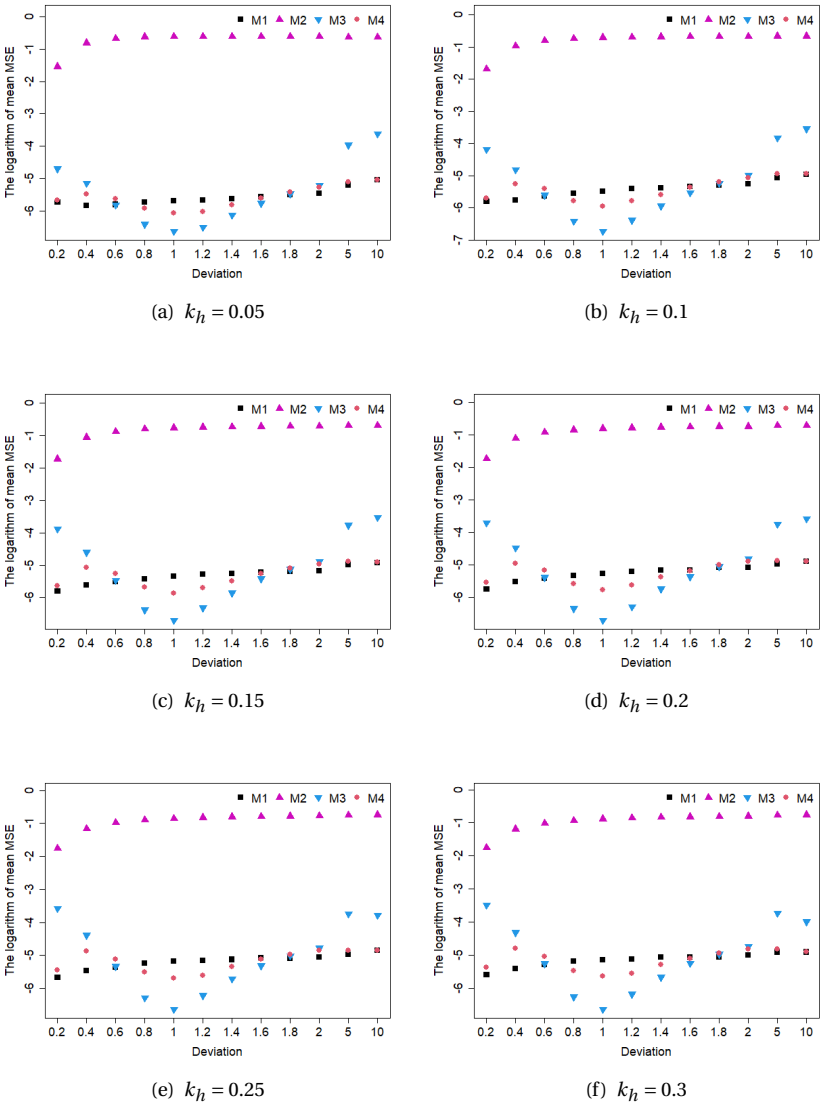


(a) $k = 0.05$ (b) $k = 0.1$

(c) $k = 0.15$ (d) $k = 0.2$

(e) $k = 0.25$ (f) $k = 0.3$

Figure B2: The box plots of logarithmic MSEs of different models under the setting parameters in Table 3.1, where (a) $k = 0.05$; (b) $k = 0.1$; (c) $k = 0.15$; (d) $k = 0.2$; (e) $k = 0.25$; (f) $k = 0.3$.

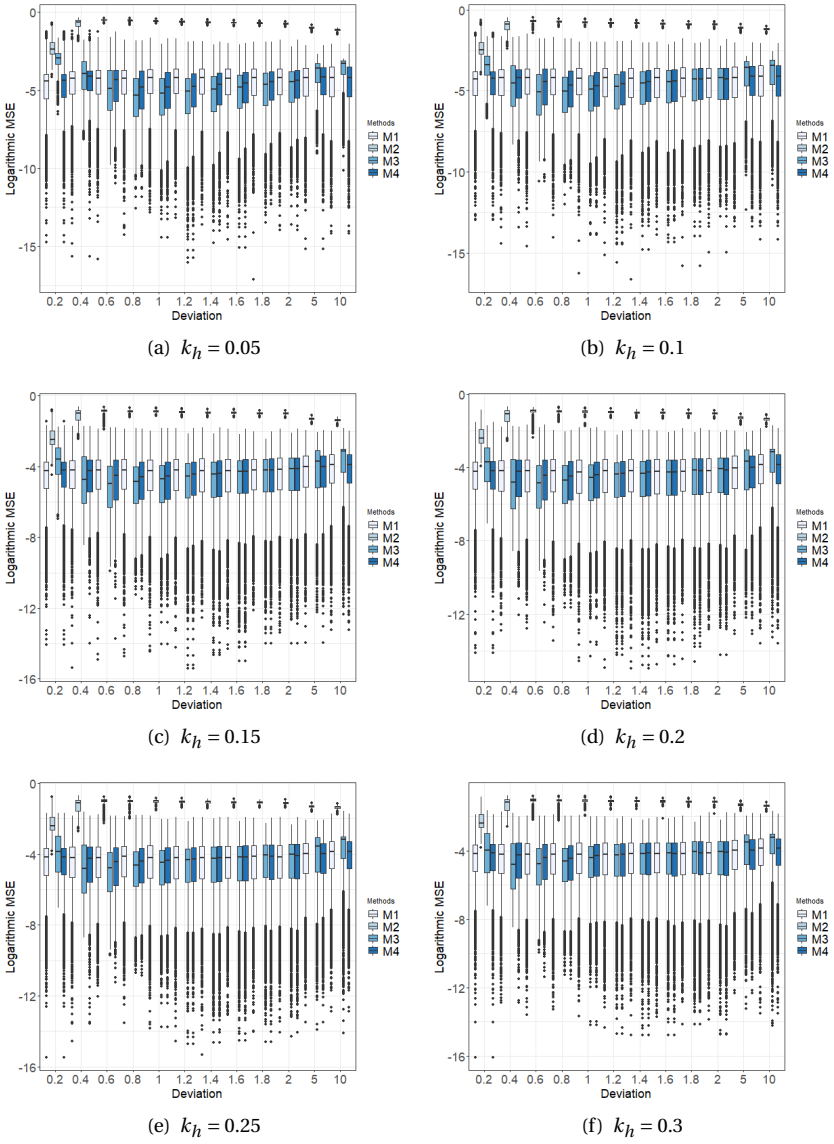


(a) $k_h = 0.05$ (b) $k_h = 0.1$

(c) $k_h = 0.15$ (d) $k_h = 0.2$

(e) $k_h = 0.25$ (f) $k_h = 0.3$

Figure B3: Mean value of MSEs of different models under the setting parameters in Table 3.2, where (a) $k_h = 0.05$; (b) $k_h = 0.1$; (c) $k_h = 0.15$; (d) $k_h = 0.2$; (e) $k_h = 0.25$; (f) $k_h = 0.3$.



(a) $k_h = 0.05$ (b) $k_h = 0.1$

(c) $k_h = 0.15$ (d) $k_h = 0.2$

(e) $k_h = 0.25$ (f) $k_h = 0.3$

Figure B4: The box plots of logarithmic MSEs of different models under the setting parameters in Table 3.2, where (a) $k_h = 0.05$; (b) $k_h = 0.1$; (c) $k_h = 0.15$; (d) $k_h = 0.2$; (e) $k_h = 0.25$; (f) $k_h = 0.3$.

4

ROBUST TRANSFER LEARNING FOR BATTERY LIFETIME PREDICTION USING EARLY CYCLE DATA

PLAIN SUMMARY

Battery lifetime prediction is crucial in industrial applications. However, the lack of diversity in training data often poses challenges regarding the robustness and generalization of lifetime predictions for batteries from different batches. Motivated by the early cycle data from lithium-iron batteries, this chapter proposes a robust transfer learning method by employing a model average framework, where the weights are determined based on the distance between the source domain and the target domain. Kernel regression is used to build the prediction of battery lifetime using early cycle data, and transfer component analysis is utilized to transfer knowledge between different domains. The case study on lithium-iron phosphate/graphite cells demonstrates that the proposed method can mitigate the impact of negative transfer and has superior performance compared to traditional methods.

4.1. INTRODUCTION

LARGE battery storage systems have been widely used to stabilize energy systems like the electricity grid and offer various benefits [80]. However, the costliness and uneconomical nature of batteries pose significant challenges, necessitating the efficient utilization of their limited resources [81]. Therefore, enhancing energy density and extending the battery's lifespan are crucial objectives. To achieve these goals, accurately predicting battery lifetime through modeling is essential [82].

Battery cell aging is affected by both the duration of use and various usage conditions. The aging process is typically divided into three stages: early, stable, and decline stages [83]. Various approaches in the literature are available for modeling battery aging, including physics-based, semi-empirical, and data-driven methods. Generally, physics-based and semi-empirical methods require a thorough understanding of degradation behavior based on failure mechanisms, which may limit their applicability [84].

On the other hand, data-driven methods, such as machine learning models, have gained significant attention in recent years due to their ability to operate without explicit knowledge of failure mechanisms [85]. However, the machine

learning methods heavily rely on the assumption that the training and test data are sampled from the same distribution. It is very common that this assumption is not satisfied in practice. Consider the task of predicting battery lifetime, although there is a lot of historical data collected from different batches, the data from different batches may not follow the same distribution due to factors such as battery type, operating conditions, and manufacturing variations that can impact battery performance [85].

Transfer learning (TL) can help overcome this issue by transferring knowledge from a source problem (training data) to a target problem (test data) with similar characteristics but different underlying distributions [12]. There are many papers focusing on using TL to predict batteries' lifetime in the literature. For example, [86] presents a new method for state-of-charge (SoC) estimation by exploiting the temporal dynamics of the measurements and the ability to transfer consistent estimates across different temperatures. [87] proposes a transferable multistage state-of-health (SoH) estimation model to perform TL across batteries in the same degradation stage. More references can be found in Section 4.2.

However, these existing methods suffer from two common drawbacks. Firstly, they often rely on data where battery capacity degradation has already occurred, necessitating a certain number of cycles. As emphasized in [3, 83], accurate predictions based on early-stage data—referred to as early-cycle data in this paper—are critical. They provide valuable insights into long-term performance without the need for extended testing, allowing for the rapid detection of early failures, manufacturing defects, or deviations from expected performance. This accelerates battery development and design optimization, lowers production costs, and enables quick assessments of battery quality and performance.

Regarding the prediction models utilizing early-cycle data, the two primary categories [83] are models based on long short-term memory (LSTM) [88, 89] or the convolutional neural network (CNN) [90, 91]. Additional references can be found in the review paper [83].

An example of early cycle battery data is depicted in Figure 4.1, where Figure 4.1(a) and Figure 4.1(b) exhibit the capacity curves for the full life cycles and the corresponding curves for early cycles (the first 100 cycles in this case).

From Figure 4.1(b), it is apparent that there is a negligible degradation trend for capacity curves in the early cycles, underscoring the challenge of accurately predicting battery lifetime using such data. However, as highlighted in [3], certain features extracted from early cycle data demonstrate strong correlations with cycle life. An example of this is shown in Figure 4.1(c) and Figure 4.1(d), where the later stages of degradation are reflected in the discharge voltage curve and the features extracted from it. More details will be introduced in Section 4.4.1.

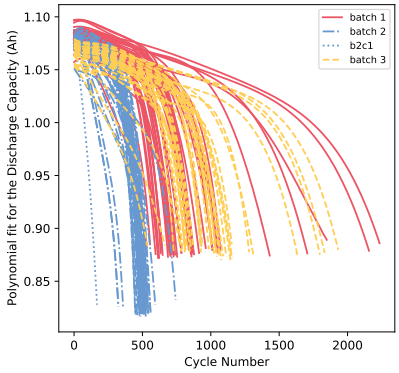
Furthermore, existing work primarily focuses on what part of data should be transferred and how to utilize the TL methods, but without addressing the challenge of reducing the impact of negative transfer. Negative transfer occurs when the knowledge gained in the source domain negatively affects the performance of a model when applied to the target domain, and this phenomenon can be attributed to factors such as domain discrepancies, task misalignment, overfitting to the source domain, etc. To our knowledge, there is limited literature exploring the utilization of TL for battery lifetime prediction based on early cycle data and discussing strategies to mitigate the effects of negative transfer.

Motivated by the aforementioned challenges and the early cycle dataset of lithium-iron batteries in [3], this chapter presents a novel robust TL framework for battery lifetime prediction using early cycle data. The new framework addresses three key issues: determining the relevant information (*what*) for effective transfer, devising an appropriate transfer strategy (*how*) to leverage knowledge from source to target data, and providing a robust TL framework by choosing appropriate weights, although we did not explicitly identify the transferable conditions (*when*).

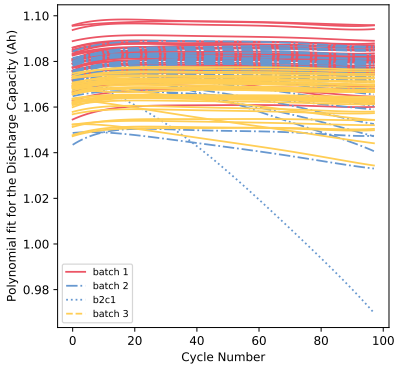
The main contributions of this chapter can be summarized as follows:

(1) Introduce a novel robust TL framework for predicting battery lifetime using early cycle data.

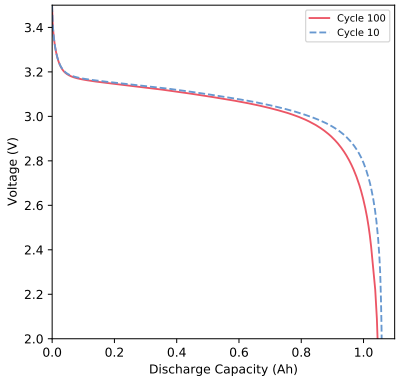
(2) Present a method for determining the weight in the proposed robust TL framework based on the distributional difference, which can reduce the effects of negative transfer.



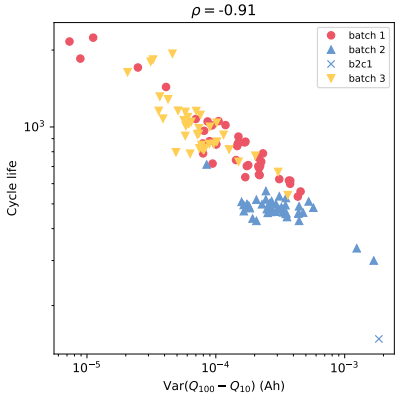
(a) Discharge capacity curves after preprocessing



(b) Discharge capacity curves over the first 100 cycles



(c) An example of the discharge voltage curves for the 100th and 10th cycle



(d) Variance of $\Delta Q_{100-10}(V)$ against battery cycle life for all cells

Figure 4.1: Overview and examples of features in the dataset.

(3) Validate the performance of the method using a substantial dataset of lithium-iron batteries.

The remainder of this chapter is organized as follows. Section 4.2 reviews some related work, including *what, how, when* to transfer. Section 4.3 provides a detailed explanation of the methods employed and details of the proposed robust TL framework. In Section 4.4, a case study based on a substantial dataset of lithium-iron batteries is conducted to validate the performance of the proposed method. Finally, Section 4.5 concludes the chapter with a discussion.

4.2. RELATED WORK

THIS section provides a review of related work concerning the three fundamental questions in TL: *what, how,* and *when* to transfer.

4.2.1. WHAT TO TRANSFER?

WHAT to transfer involves selecting the relevant information or data to transfer. Generally, most of the TL methods for battery lifetime prediction are feature-based. According to the way of extracting features, there are two commonly used methods for transferring information: transferring the physical significance features or data-driven features [92, 93]. In practice, the selection of feature type should be determined based on the specific engineering context and requirements.

PHYSICAL SIGNIFICANCE FEATURES

In the case of physical significance features, the feature extractor remains fixed, wherein the extracted features possess specific physical significance. Subsequently, the transferred data can be integrated into a data-driven predictor.

For example, [94] introduces a TL-based method for predicting target SoH values using degradation data from other batteries by employing transfer component analysis (TCA) with an extreme learning machine algorithm. [95] proposes a framework to monitor SoH of batteries by integrating maximum mean discrepancy (MMD), semi-supervised TCA, and mutual information, and

validates the effectiveness by using a real dataset containing four batteries operating under different conditions. [96] proposes a TL-based framework for battery RUL prediction, which utilizes features generated from electrochemical theory, capacity-differential voltage curves, and electrochemical impedance spectroscopy curves.

DATA-DRIVEN FEATURES

Another commonly used method is to transfer the data-driven features generated from an adaptive feature extractor, where the feature extractor itself is also a data-driven model. LSTM and neural networks are the most popular data-driven feature extractors. For example, in the work by [97], a convolutional neural network-based feature extractor is utilized to extract deep features from the collected data, which include the current, voltage, and capacity characteristics of the battery during the charging/discharging cycle. [98] adopts an LSTM network model to extract deep features from the collected capacity curves, and proposes a TL-based battery RUL prediction model by integrating a particle filter model.

4.2.2. HOW TO TRANSFER?

THE question *how* to transfer concerns selecting the appropriate TL method and hyperparameters. There are three main approaches to performing TL, namely based on the fine-tuning strategy, the discrepancy metric, or the domain adversarial [92, 93].

FINE-TUNING-BASED

The strategy to transfer through fine-tuning is widely used in the lifetime prediction of batteries. The basic idea of the fine-tuning strategy is to involve retraining a pre-trained data-driven model for better performance on specific tasks, which uses only a small amount of data from the target dataset [92]. Specifically, in battery life prediction, fine-tuning involves first building a pre-trained data-driven model using source data and then retraining the model using a small amount of target data, thereby improving accuracy.

For example, [99] proposes a TL-based framework for battery SoH prediction, which combines an LSTM network with adjustable fine-tuning-based fully connected layers. [100] merges degradation pattern recognition with a fine-tuning-based LSTM approach to present a TL-based SoH prediction model, which demonstrates superior performance compared to existing methods. [101] proposes a TL-based framework for SoH prediction of lithium-ion batteries, which employs a deep neural network architecture that integrates equivalent circuit simulated layers and a fine-tuning network hierarchy. More references can be found in review papers [92, 93].

As mentioned earlier, the fine-tuning strategy typically necessitates a small amount of labeled target data. However, in practical scenarios, such labeled testing data for training is often unavailable, thereby constraining the applicability of the fine-tuning strategy.

DISCREPANCY METRIC-BASED

The methods based on discrepancy metrics are widely used in TL. The literature applying TCA for transfer generally offers advantages in computational efficiency. Despite these approaches, another common category within discrepancy metric-based methods involves incorporating the discrepancy metric into the loss function during model training. For example, [102] presents a novel deep learning framework that addresses this challenge by combining a deep LSTM network to model the nonlinear mapping from monitored data (e.g., terminal voltage and current) to battery capacity, and a domain adaptation layer with MMD to align degradation features between source and target batteries. [103] combines the MMD with a gated recurrent unit recurrent neural network to reduce the domain discrepancy for battery SoH prediction. [104] integrates the MMD loss with a convolutional neural network to predict the battery SoH.

DOMAIN ADVERSARIAL-BASED

The domain adversarial-based approach is another effective TL method aimed at learning a domain-invariant feature space where the domain classifier cannot distinguish the domain of the input data. Although adversarial adaptation methods are effective, there are only a limited number of references

using them as TL strategies at this stage. For example, [105] introduces a temperature-adaptive transfer network that employs adversarial adaptation and MMD to minimize domain divergence for estimating the SoC in batteries. In [106], domain adversarial training and TL methods are integrated into Bayesian deep learning to propose an innovative RUL prediction framework capable of handling diverse machines with limited data.

While domain-adversarial-based TL solutions have demonstrated success, a notable drawback is the resource-intensive nature of the adversarial training process. This computational demand can somewhat restrict their applicability in intricate industrial scenarios. Therefore, further research and development efforts are necessary to align these methods with the practical requirements of the industry [92, 93].

4.2.3. WHEN TO TRANSFER?

COMBINING the methods in *what* (Section 4.2.1) and *how* (Section 4.2.2), various TL-based methods can be generated. Despite the effectiveness of these transfer methods, they can encounter challenges in complex and demanding environments, potentially resulting in negative transfer, as introduced in Section 4.1. Hence, there is a need to develop methods for quantifying transferability or assessing negative transfer, which is essentially a question of *when* to transfer.

However, there is limited literature focusing on the *when* to transfer issue in battery health management. For example, [107] proposes a Gaussian process regression model to forecast the capacity of batteries and uses hyperparameters of the kernel function in the covariance matrix to control the negative transfer. [108] introduces the product of MMD and a tuning parameter as the penalty term in their loss function to reduce the likelihood of negative transfer in battery SoC estimation. Nonetheless, these methods either fail to predict battery lifetime or require the capacity to exhibit a degradation trend, which is inconsistent with our task of predicting battery lifetime based on early cycle data shown in Figure 4.1. Additionally, they require labeled data in the target domain, which is difficult to obtain in practice.

4.3. ROBUST TRANSFER LEARNING FRAMEWORK

IN this section, the proposed robust TL framework is introduced, which is designed to address the three questions outlined in Section 4.2.

Regarding the *what* to transfer issue, we prioritize the use of features with physical significance due to their interpretability. Specifically, in battery lifetime prediction research involving early cycle data, domain knowledge pertaining to lithium-ion batteries is commonly leveraged. Key features, such as initial discharge capacity, charge time, and cell temperature are widely employed. For further details regarding these features, refer to Section 4.4 and [3].

4.3.1. HOW TO TRANSFER: TRANSFER COMPONENT ANALYSIS AND THE KERNEL REGRESSION

AS discussed in Section 4.2.2, there are three commonly employed approaches to tackle the *how to transfer* issue. In this chapter, we employ discrepancy metric strategies for information transfer because they do not require labeled data or substantial computing resources in the target domain, unlike fine-tuning and domain adversarial approaches.

The procedures of discrepancy metric strategies can be further divided into three steps: choosing the discrepancy metric, domain adaptation, and developing the predictor. These steps are discussed in detail as follows.

DISCREPANCY METRIC

Discrepancy metrics play a pivotal role in the discrepancy metric-based strategy, and among the various metrics available, MMD is particularly popular.

Consider a labeled source domain dataset $(X_S, Y_S) = (X_{S,1}, Y_{S,1}), \cdots, (X_{S,n_S}, Y_{S,n_S})$, where $X_{S,i}$, $i = 1, \cdots, n_S$, represents the i-th input of the source domain dataset, and $Y_{S,i}$ denotes its corresponding output. Additionally, there exists an unlabeled target domain dataset $X_T = X_{T,1}, \ldots, X_{T,n_T}$, where $X_{T,j}$, $j = 1, \ldots, n_T$, signifies the j-th input of the target dataset. The total size of the source and target datasets is represented by $n = n_S + n_T$. Let $\mathscr{P}(X_S)$ and $\mathscr{Q}(X_T)$ (or $\mathscr{P}$ and $\mathscr{Q}$ in short) denote the marginal distributions of inputs from the source and target datasets, respectively. The task is to predict the output of the target domain,

$Y_{T,j}, j = 1, \cdots, n_T$. We assume that $\mathscr{P} \neq \mathscr{Q}$ while there exists a mapping function $\phi(\cdot)$ such that $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$.

The key issue in discrepancy metric-based TL is mapping different domain instances onto a common space. Specifically, MMD employs a function $\phi(\cdot)$ to map each instance to the reproduced Hilbert space $\mathscr{H}$ associated with the kernel $k : \chi \times \chi \to \mathbb{R}$, where $k(X_{S,i}, X_{T,j}) = \phi(X_{S,i})^\top \phi(X_{T,j})$, and χ is the feature space of the source and target domains.

The MMD can be defined as the distance between two different projections of the means. By using a kernel trick, the squared MMD distance can be re-formulated as

$$\begin{aligned} \mathrm{MMD}^2(X_S, X_T) &= \| \frac{1}{n_S} \sum_{i=1}^{n_S} \phi(X_{S,i}) - \frac{1}{n_T} \sum_{j=1}^{n_T} \phi(X_{T,j}) \|^2_{\mathscr{H}} \\ &= Tr(KL), \end{aligned} \tag{4.1}$$

where $K = \begin{bmatrix} K_{X_S,X_S} & K_{X_S,X_T} \\ K_{X_T,X_S} & K_{X_T,X_T} \end{bmatrix} \in \mathbb{R}^{n \times n}$ is a composite kernel matrix, with K_{X_S,X_S}, K_{X_S,X_T}, K_{X_T,X_S}, and K_{X_T,X_T} being the kernel matrices defined by k based on the data in the source domain, the target domain, and across domains, respectively. L is a matrix with the (i, j)-th entry L_{ij} defined as

$$L_{ij} = \begin{cases} 1/n_S^2 & x_{S,i}, x_{S,j} \in X_S \\ 1/n_T^2 & x_{T,i}, x_{T,j} \in X_T \\ -1/n_S n_T & \text{other.} \end{cases}$$

MMD is a kernel-based distance metric that plays a crucial role in identifying underlying patterns in data and recognizing distribution differences. The choice of an appropriate kernel is essential for the effective utilization of the metric. However, the task of determining the most suitable kernel for MMD can be intricate and remains an ongoing research area.

DOMAIN ADAPTION

Once the discrepancy metric is selected, the subsequent step focuses on minimizing the chosen metric using the data from the source and target domains. In this chapter, we employ a widely used domain adaptation method

known as TCA. As stated in [12], TCA aims to minimize the distance between the marginal distributions by utilizing a kernel feature extraction framework. Using MMD, TCA can be achieved by solving the following kernel learning problem [12],

$$\begin{aligned} \min_{W} \quad & Tr(W^\top KLKW) + \lambda Tr(W^\top W) \\ \text{s.t.} \quad & W^\top KHKW = I_m, \end{aligned} \tag{4.2}$$

where W is a matrix that transforms the kernel map features to a m-dimensional space, $H = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$ is the centering matrix, with $I_n \in \mathbb{R}^{n\times n}$, $I_m \in \mathbb{R}^{m\times m}$ being identity matrices and $\mathbf{1}_n \in \mathbb{R}^{n\times 1}$ being the column vector with all ones. Subsequently, the matrix W is obtained by identifying the first m smallest eigenvalues of $(I_m + \lambda KLK)^{-1}KHK$, where λ is a hyperparameter determined through cross-validation. More details and explanations of TCA can be found in [12].

THE PREDICTOR

Following the acquisition of the most closely matched transferred data, the subsequent phase involves constructing a predictor. This is achieved by training a model using the transferred source data. In practical applications, the choice of a specific predictor depends on the context. For example, in the case of the motivated battery dataset described in [3], the authors employed a regularization model known as the elastic net, which combines the Lasso and Ridge regression methods for making predictions. The underlying assumption of the elastic net is a linear relationship between the input variables and the corresponding response. However, such a parametric assumption may not be maintained in the transferred data after TCA. Therefore, we adopt the widely used nonparametric approach, the kernel regression [109], as our predictor.

Let $(X_{S,1}, Y_{S,1}), \cdots, (X_{S,n_S}, Y_{S,n_S})$ be the given source data set, the estimated result of the i-th sample based on the kernel regression can be expressed as,

$$\hat{Y}_{S,i} = \frac{\sum_{j\neq i} Y_{S,j} \times k(X_{S,i}, X_{S,j})}{\sum_{j\neq i} k(X_{S,i}, X_{S,j})}. \tag{4.3}$$

For the target data set $\{X_{T,1}, \cdots, X_{T,n_T}\}$, the prediction for the i-th sample can

be derived as,

$$\hat{Y}_{T,i} = \frac{\sum_{j=1}^{n_S} Y_{S,j} \times k(X_{T,i}, X_{S,j})}{\sum_{j=1}^{n_S} k(X_{T,i}, X_{S,j})}, \tag{4.4}$$

More details and explanations of kernel regression can be found in [109].

4.3.2. WHEN TO TRANSFER: A ROBUST TRANSFER LEARNING METHOD

THE problem about *when to transfer* is important because inappropriate TL may lead to negative transfer. It is also difficult to verify performance since there is no labeled data in the target domain [92].

THE RELATIVE CHANGE IN MMD IS NOT RELIABLE

A straightforward idea to solve this issue is using the relative change in MMD to assess the success of TL since they are specifically designed to measure the difference between the source and target domains. However, the subsequent simulation study reveals that the reduction in MMD values does not consistently correspond to increased similarities. Additionally, the effectiveness of TCA is observed to be contingent on the initial MMD value.

Note that the transferred knowledge in this paper refers to the data from the source and target domains that have been transformed using the transformation matrix W. Successful transfer indicates that the distributions of the transferred source and target domain data are significantly closer to each other compared to the distributions of the original source and target domains. To illustrate the behavior of traditional TCA, we present three simulated examples in Figure 4.2. The first scenario involves two samples with a small initial MMD value, the second involves two samples with an initial MMD within an appropriate range, and the third involves two samples with a large initial MMD value. We utilize the widely used t-SNE [110] for visualizing the distributional changes in the high-dimensional feature space resulting from TCA, providing an intuitive representation essential for assessing domain adaptation success and understanding the alignment between source and target domains.

Specifically, in Figure 4.2(a), we observe that when the initial MMD value is relatively small, applying TCA, while successful in reducing MMD (from

0.1022 to 0.0576), may introduce alterations to the distribution shape. In such cases, the prediction without TL, denoted as NoTL, could be preferred, as the distribution shapes of the raw source and target domains are more similar than those after TCA. Turning to Figure 4.2(b), we focus on scenarios where the initial MMD falls within a suitable range. Here, TCA significantly enhances the similarity between the source and target distributions, demonstrating its effectiveness (reducing the MMD value from 0.5177 to 0.0002). In Figure 4.2(c), we explore the case with a large initial MMD value. Despite TCA's ability to reduce the MMD metric (from 1.2345 to 0.2711), the transferred data still exhibits considerable dissimilarity, emphasizing the limitations of TCA when faced with large MMD values. Simultaneously, the smaller MMD values may still offer valuable insights, given the distinct dissimilarities in distribution shapes observed in both the raw source and target domains, as well as the transferred source and target domains. Figure 4.2(a) and Figure 4.2(c) illustrate that although the MMD values decreased after applying TCA, indicating improved global alignment, the increased divergence observed in the t-SNE visualization suggests that local differences may have become larger. This highlights that, in some cases, the relative change in MMD may be less reliable.

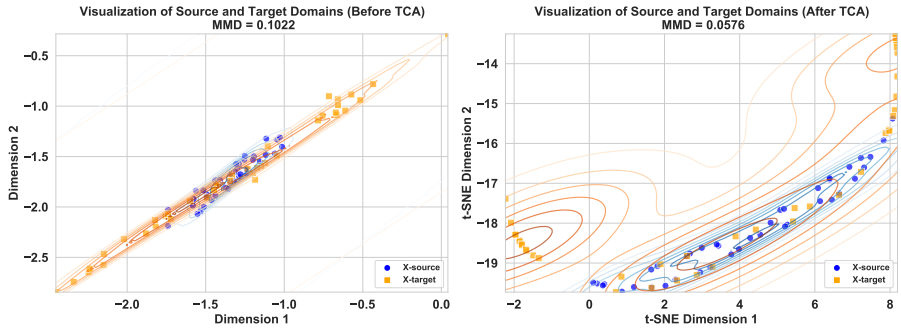
The insights from Figure 4.2 suggest that the initial value of MMD, indicating the similarity between the original source and target domain, highly influences the performance of TCA. Thus, a robust transfer learning framework is essential to strike a balance between NoTL and those with TL.

A ROBUST TRANSFER LEARNING METHOD BASED ON MODEL AVERAGING

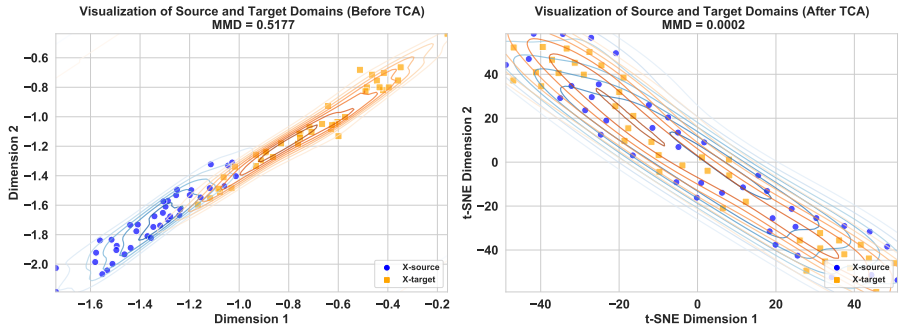
To mitigate the impact of negative TL, this chapter adopts the idea of model averaging [111] to propose a more robust TL method.

Let $\hat{\boldsymbol{Y}}_{T,1}$ be the predicted result using NoTL, and $\hat{\boldsymbol{Y}}_{T,2}$ be the predicted result using TL. Then, the final result $\hat{\boldsymbol{Y}}_T = (1-w) \times \hat{\boldsymbol{Y}}_{T,1} + w \times \hat{\boldsymbol{Y}}_{T,2}$, where $w \in [0,1]$ is the weight to balance NoTL and TL. Thus, the issue turns to finding a suitable form of w.

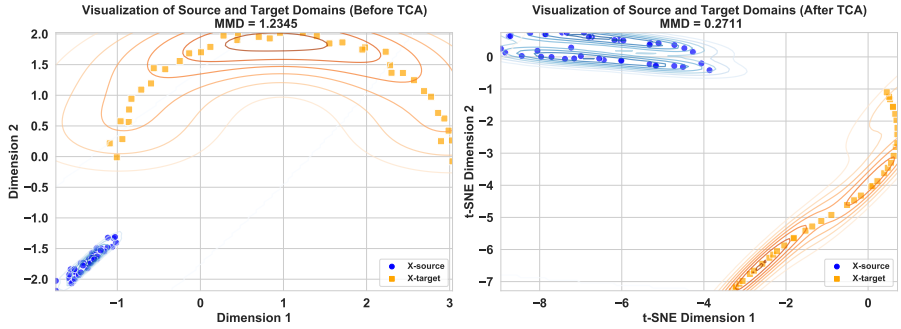
Considering the similarity between the raw source data $\mathscr{P}$ and target data $\mathscr{Q}$, if $\mathscr{P}$ and $\mathscr{Q}$ yield a small initial MMD value, indicating their similarity in this case, then w should be set to 0. This concern is affirmed by the results shown



(a) Example 1: The impact of a small initial MMD value on distribution shape.



(b) Example 2: Improving distribution similarity with an appropriate initial MMD value.



(c) Example 3: Non-negligible differences in distribution shape after transfer still exist for a larger initial MMD value.

Figure 4.2: Illustrative examples of TCA and distribution preservation with varying initial MMD values.

in Figure 4.2(a), where TCA increased the dissimilarity of the distribution shapes when $\mathscr{P}$ and $\mathscr{Q}$ yielded a small initial MMD value. Additionally, w should also be smaller when there is no similarity between the transferred source data $\mathscr{P}_{new}$ and target data $\mathscr{Q}_{new}$ as this case indicates that after TL, the source and target domain are still very different, as shown in Figure 4.2(c). As significant distributional differences persist after transfer in such cases, the effectiveness of TCA becomes uncertain. Choosing traditional machine learning algorithms is preferable in these instances, as they do not rely on the assumptions associated with transfer learning.

To quantify the degree of similarity, a threshold is necessary. In this chapter, we utilize the empirical distribution of the MMD metric to determine such a threshold, which can be obtained through a permutation approach [112]. Specifically, the empirical distribution function of the MMD is derived by repeatedly shuffling the datasets and recalculating the MMD value. Subsequently, the $1-\alpha$ quantile of the derived empirical distribution is employed as the threshold, denoted as T. If the original MMD value, denoted as MMD_0, exceeds T, we conclude that the two distributions are different; otherwise, we state that the two distributions are similar enough. Naturally, the probability of obtaining results at least as extreme as the result actually observed, denoted as p, can be chosen as a weight. Thus, the final form of the weight term w can be expressed as,

$$w = p \times \mathbb{I}_{\{\text{MMD}_0 > T\}}, \tag{4.5}$$

where $\mathbb{I}_{\{\text{MMD}_0 > T\}}$ is the indicator function, whose value is equal to 1 when the MMD_0 calculated by $\mathscr{P}$ and $\mathscr{Q}$ exceeds T, and 0 otherwise.

The weight form in Eq. (4.5) satisfies the aforementioned properties we desired. When the raw source and target distributions $\mathscr{P}$ and $\mathscr{Q}$ are similar enough, the calculated MMD_0 should be located within the threshold T, resulting in the value of w being equal to 0. Otherwise, TL should be performed, and the corresponding value of w is equal to p, which is derived from the transferred distributions $\mathscr{P}_{new}$ and $\mathscr{Q}_{new}$.

4.3.3. ROBUST TRANSFER LEARNING FRAMEWORK

THE proposed robust TL framework is presented in Algorithm 4.1, and the fundamental structure of the approach is illustrated in Figure 4.3. Following data pre-processing, we extract selected features from both the source and target domains. Subsequently, the weight term w in Eq. (4.5) and the hyperparameter λ in Eq. (4.2) can be determined following the procedure introduced in Section 4.3.2. The weight term w is then utilized to balance the predictions $\hat{\boldsymbol{Y}}_{T,1}$ and $\hat{\boldsymbol{Y}}_{T,2}$.

Algorithm 4.1: The proposed robust transfer learning framework.

Input: Source domain $\boldsymbol{X}_S$ and corresponding response $\boldsymbol{Y}_S$, target domain $\boldsymbol{X}_T$, kernel function type, λ;

Output: Predicted response for target domain $\hat{\boldsymbol{Y}}_T$.

1: Data preprocessing, encompassing techniques such as smoothing and outlier handling;
2: Generate features according to the preprocessed data;
3: Calculate the weight term w following the procedure introduced in Section 4.3.2;
4: Derive the prediction $\hat{\boldsymbol{Y}}_{T,1}$ using NoTL method;
5: If $w \neq 0$, derive the prediction $\hat{\boldsymbol{Y}}_{T,2}$ using TL method, otherwise, set $\hat{\boldsymbol{Y}}_{T,2} = 0$;
6: Derive the final prediction $\hat{\boldsymbol{Y}}_T = (1-w) \times \hat{\boldsymbol{Y}}_{T,1} + w \times \hat{\boldsymbol{Y}}_{T,2}$.

4.4. CASE STUDY

IN this section, a case study on the lifetime data of 124 lithium-ion batteries is presented to demonstrate the implementation of the proposed framework.

4.4.1. DATA OVERVIEW

THE dataset used for the case study is taken from Severson et al. [3], which comprises 3 batches consisting of a total of 124 lithium-ion phosphate/graphite batteries (41/43/40 in batch 1/2/3 respectively). The batches can be considered as separate experiments, with batch 3 conducted nearly a year after batches 1 and 2. For each battery, the following data is attached,

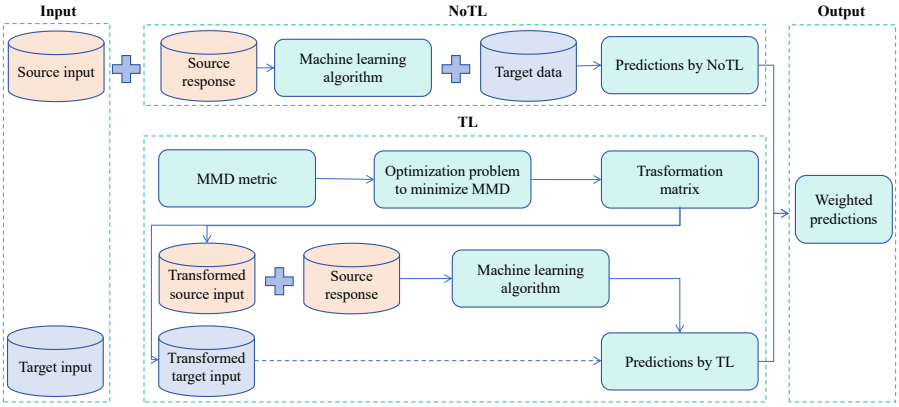


Figure 4.3: The basic structure of the proposed robust transfer learning approach.

(1) Cycle Life: The number of cycles until the battery's capacity has decreased below 80% (and for batch 2, 75%) of its nominal capacity.

(2) Charge Policy: All batteries are charged according to different fast-charging conditions in a temperature-controlled environment, see [3] for more details.

(3) Summary Data: The summary data contains information for each cycle, including the cycle number, discharge capacity, charge capacity, internal resistance, maximum temperature, average temperature, minimum temperature, and charging time.

(4) Cycle Data: The information contained within a cycle includes the time, charge capacity, current, voltage, temperature, and discharge capacity. Additionally, calculated variables such as the discharge rate, the discharge capacity interpolated linearly, and temperature interpolated linearly are also included.

Note that the dataset is the largest publicly available for nominally identical commercial lithium-ion batteries cycled under controlled conditions.

4.4.2. PREDICTION MODELS AND ANALYTICAL SCENARIOS

FOR the prediction model, the original paper [3] uses a feature-based method. After pre-processing the curves and fitting a polynomial to the discharge capacity as a function of voltage in the 10th and 100th cycle $\Delta Q_{100-10}(V)$, several features can be extracted from these values, such as mean, variance, minimum, skewness, and kurtosis. The logarithmic values of the summary statistics exhibit a strong correlation with the logarithmic cycle life, while the discharge capacities of only the first 100 cycles are weakly correlated with the cycle life. The capacity curves and an example of the discharge voltage curves are shown in Figure 4.1. The battery cell 'b2c1', which has a relatively low lifetime, will be excluded from the dataset to prevent its negative impact on the results.

Since the features based on $\Delta Q_{100-10}(V)$ have high predictive power, three different models proposed by the paper are summarized in Table 4.1.

Table 4.1: Various models containing different feature sets which are included in the case study results

Model	**Description**
Variance	A univariate model, which uses $\log(\mathrm{Var}(\lvert\Delta Q_{100-10}(V)\rvert))$ to predict log cycle lives.
Discharge	Includes summary statistics of $\Delta Q_{100-10}(V)$ such as minimum, mean, variance, skewness, kurtosis and $\Delta Q(V=2)$, in addition to other candidate features obtained during discharge, such as the slope and intercept of the linear fit for the capacity fade curve cycle 2 to 100 and cycle 91 to 100, discharge capacity at cycle 2 ($Q(n=2)$) and 100 ($Q(n=100)$), and the difference between the maximal discharge capacity for a cell and cycle 2 ($\max_n(Q(n)) - Q(n=2)$).
Full	Next to the features included in the variance and discharge model, additional features are included from different data streams, such as temperature and internal resistance.

To fully utilize the dataset, we consider seven different scenarios for both the source and target data, allowing for comprehensive comparisons. A summary

of these scenarios is presented in Table 4.2, providing a clear overview of the variations studied. Both the RMSE and the mean absolute percentage error (MAPE) are used as prediction performance metrics.

Table 4.2: Different scenarios for source and target data

Scenario	Source data	Target data
1	Batch 1	Batch 2
2	Batch 1	Batch 3
3	Batch 2	Batch 3
4	Batch 1&2	Batch 3
5	Original train set (from [3])	Original test set (from [3])
6	Original train set (from [3])	Batch 3
7	Original test set (from [3])	Batch 3

4.4.3. PREDICTION PERFORMANCE

THE performance of kernel regression and TL is affected by the hyperparameter λ in Eq. (4.2) and the choice of the kernel type. In this chapter, we adopt four widely used kernel types: linear, polynomial, Rbf, and Laplace kernel. To ensure fair comparisons across different scenarios, the hyperparameter λ is determined through cross-validation, and the hyperparameters of the kernel functions remain consistent across all scenarios. Taking scenario 2 as an example, RMSEs for different models across various kernel types are displayed in Figure 4.4, where the red, green, and blue boxes represent the results derived from NoTL, TL, and the proposed robust TL methods, respectively. Figure 4.4 indicates that the proposed TL framework does indeed possess the desired ability to balance NoTL and TL methods for different models. Similar results for MAPEs are observed but not displayed to conserve space.

Taking the linear kernel as an example, in the variance model case, the NoTL approach outperforms TL, indicating that TL introduces a negative effect. In this scenario, the proposed method favors NoTL, resulting in predictions that align closely with NoTL, thereby minimizing the negative impact caused by TL.

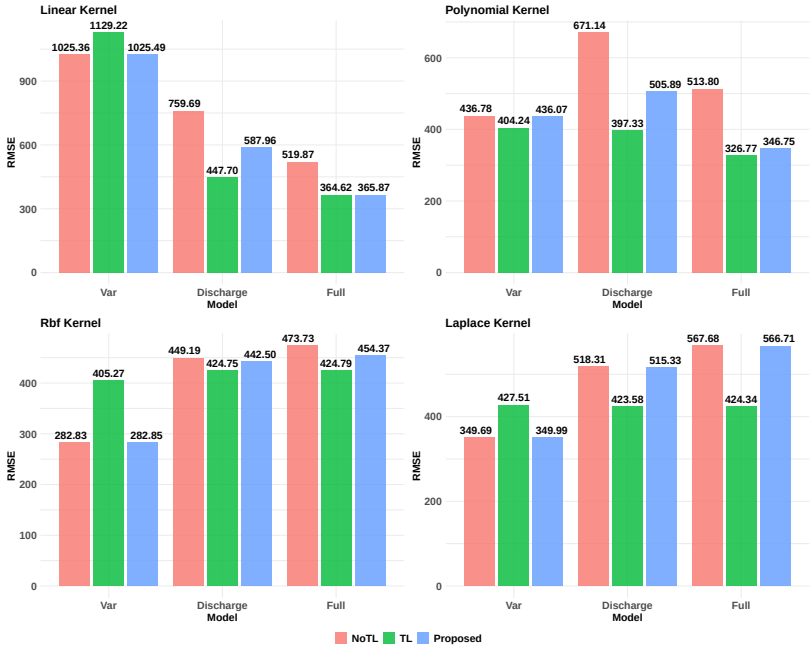


Figure 4.4: An example of RMSEs for different methods via the tuning parameter λ and kernel types.

Conversely, for the discharge and full models, TL significantly outperforms NoTL. In these cases, the proposed approach prioritizes TL, with final predictions derived from a weighted combination of TL and NoTL using weight w, thereby leveraging the benefits of TL.

Regarding the RMSEs of lifetime predictions under all scenarios, kernel types, and models, the results are reported in Table 4.3. Cases where the proposed robust TL framework successfully mitigates the impact of negative TL in direct TL are highlighted in bold for enhanced visibility.

In terms of computational cost, the TCA-based approach took 0.02 seconds for data transformation. The experiments were conducted on the following system specifications:

Operating System: Windows 10 (Version 10.0.22621)

CPU: Intel 16-core processor (Intel64 Family 6 Model 141)

Memory: 32 GB.

The results presented in Figure 4.4 and Table 4.3 demonstrate that the proposed robust TL framework effectively balances NoTL and TL in general. Specifically, the proposed method performs well in extracting positive outcomes from TL when it significantly improves prediction results. Conversely, when TL yields unfavorable results, the proposed method favors the NoTL approach, aligning the results with NoTL for comparability. These observations are particularly noticeable in scenarios 2, 3, 4, 6, and 7.

However, exceptions exist, particularly in scenarios 1 and 5. The performance in scenario 1 exhibits instability. Upon investigating the root cause of this instability, we identify a potential factor: a violation of the fundamental assumption of TL, namely, the sharing of the same distribution between the responses of the source and target domains. As the performance of the proposed framework is influenced by kernel types, a kernel two-sample test introduced in [112] is employed. The p-values resulting from hypothesis tests across scenarios and kernel types, assessing whether the responses of the source and target domains share the same distribution, are presented in Table 4.4. Taking scenario 2 as an example, the p-values across different kernel types are all larger than 0.01, indicating that we cannot reject the null hypothesis. This suggests that, in scenario 2, the response distribution in scenario 2 is not statistically different and likely shares the same distribution. Notably, for various kernel types, scenario 1 consistently rejects the null hypothesis that the responses of source and target domains share the same distribution.

For scenario 5, where the initial MMD value of the raw source and target data is already smaller enough, the results should align with the NoTL method, as $w = 0$ in this case. The results presented in Table 4.3 reflect the performance without considering this setting, confirming the necessity of the indicator function in Eq. (4.5). Moreover, results for scenario 5 indicate that when the raw source and target data derive a small initial MMD value, TL may provide unstable results, potentially worsening overall predictions.

To further demonstrate the robustness of the proposed approach, we conducted additional experiments based on the *full* model, integrating the

framework with LSTM and CNN-based techniques. The results, presented in Table 4.5, further highlight the robustness of the proposed framework. Specifically, '-KR', '-LSTM', and '-CNN' represent the corresponding results based on kernel regression, LSTM, and CNN, respectively. Regardless of the method used, the results in Table 4.5 indicate that when TL is significantly impacted by negative transfer, our method tends to favor the NoTL approach, effectively minimizing the negative effects.

In terms of computational cost, the approach based on kernel regression took approximately 0.89 seconds per scenario, while the LSTM-based approach took around 5.59 seconds and the CNN-based approach around 4.05 seconds. This level of computational efficiency suggests that the proposed method is feasible for use in real-world battery management systems.

4.4.4. SENSITIVITY ANALYSIS

ACCORDING to the results, both the hyperparameter λ in Eq. (4.2) and the choice of kernel type significantly impact the performance. Given that selecting suitable kernel function types remains an ongoing area of research, we specifically focus on exploring the sensitivity associated with the hyperparameter λ. This uncertainty can influence the performance of both the TL and the proposed robust TL framework.

Taking the full model in scenario 1 as an example, we find that the optimal λs for four commonly used kernel types in the proposed robust TL framework fall within the range $[0.001, 0.01]$. Consequently, we present the RMSEs for different methods across various kernel types, with lambda values set to 0.001, 0.003, 0.005, 0.007, and 0.01, in Figure 4.5.

The results presented in Figure 4.5 demonstrate that the proposed robust TL framework consistently achieves a balance between the NoTL and TL methods, irrespective of the kernel types. Notably, this balancing function remains unaffected by specific values of λ.

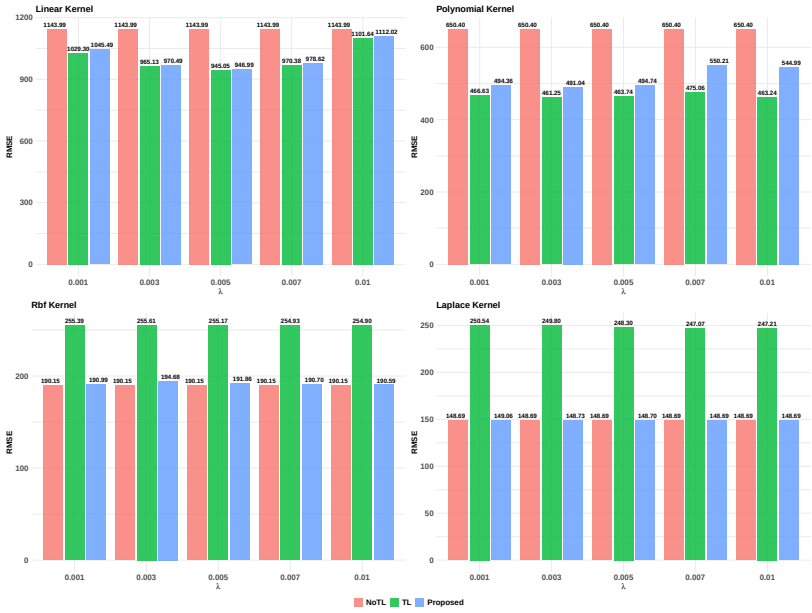


Figure 4.5: An example of RMSEs for different methods via the tuning parameter λ and kernel types

4.5. CONCLUSIONS

THIS study introduces a robust TL framework that balances the outcomes of NoTL and TL through a weight form associated with the MMD metric. The strength of the proposed framework lies in its flexibility, enabling the integration of various traditional machine learning algorithms within the transfer learning process. The case analysis results, conducted on early cycle battery data for predicting lifespan, demonstrate that the proposed method tends to favor TL results when TL performs well and leans towards NoTL results when TL performance is suboptimal. This robustness is crucial in practical applications as TL can not guarantee consistently better performance.

Table 4.3: RMSEs via different scenarios, kernel types, and models.

	Scenarios		1	2	3	4	5	6	7
Variance	Linear	NoTL	7525	1025	1037	1018	1050	1026	1019
		TL	17247	1129	1030	1015	1512	1017	1014
		Weight	0	0.0034	0	0	0.3368	0	0
		Proposal	**7526**	**1025**	1037	1018	1171	1026	1019
	Polynomial	NoTL	270	437	895	459	484	641	568
		TL	250	404	779	458	483	527	462
		Weight	0	0.0209	0	0	0.3997	0	0
		Proposal	270	436	895	459	483	641	568
	Rbf	NoTL	138	283	814	333	439	491	399
		TL	238	405	703	502	434	565	537
		Weight	0	0.0003	0	0	0.7194	0	0
		Proposal	**138**	**283**	814	**333**	435	491	**400**
	Laplace	NoTL	139	350	895	362	459	551	476
		TL	256	428	694	534	436	586	560
		Weight	0.1067	0.0041	0.0018	0.0004	0.8797	0.0005	0.0004
		Proposal	**139**	**350**	894	**362**	439	551	**476**
Discharge	Linear	NoTL	545	760	918	725	560	792	759
		TL	924	448	809	517	586	606	524
		Weight	0.9623	0.5315	0.2098	0.4951	0.9880	0.3529	0.4458
		Proposal	905	588	899	630	585	736	666
	Polynomial	NoTL	372	671	845	598	427	711	691
		TL	428	397	662	507	494	548	517
		Weight	0.8974	0.6159	0.5358	0.4432	0.9854	0.2546	0.4704
		Proposal	414	506	761	559	493	675	616
	Rbf	NoTL	197	449	721	458	427	561	533
		TL	254	425	687	531	429	581	557
		Weight	0.6902	0.2069	0.1471	0.5403	0.9272	0.0964	0.4891
		Proposal	219	442	716	498	429	**563**	544
	Laplace	NoTL	187	518	771	483	436	605	571
		TL	249	424	689	530	428	580	556
		Weight	0.0514	0.0318	0.0583	0.0205	0.9965	0.1143	0.0658
		Proposal	**186**	515	767	**484**	428	602	570
Full	Linear	NoTL	1144	520	846	497	496	583	580
		TL	945	365	660	431	484	486	461
		Weight	0.9892	0.9847	0.9806	0.6028	1.0000	0.9781	0.9635
		Proposal	947	366	665	455	484	488	466
	Polynomial	NoTL	650	514	815	526	428	634	617
		TL	467	327	651	491	466	509	489
		Weight	0.7882	0.8852	0.4078	0.5665	0.9958	0.6382	0.8614
		Proposal	494	347	757	506	466	557	508
	Rbf	NoTL	190	474	741	516	417	596	572
		TL	256	425	688	533	429	582	557
		Weight	0.1624	0.3549	0.0310	0.5318	0.9234	0.1696	0.2266
		Proposal	**195**	454	740	524	428	594	569
	Laplace	NoTL	149	568	804	565	419	649	631
		TL	251	424	690	531	426	582	557
		Weight	0.0190	0.0073	0	0	0.9993	0	0
		Proposal	**149**	567	804	565	426	649	631

Table 4.4: P-values for hypothesis test on responses distribution homogeneity across scenarios and kernel types

Scenario	Linear	Polynomial	Rbf	Laplace
1	0.0001	0.0001	0.0001	0.0001
2	0.1751	0.8388	0.1192	0.0967
3	0.0001	0.0001	0.0001	0.0001
4	0.0001	0.0308	0.0094	0.0028
5	0.5480	0.5828	0.1637	0.2371
6	0.0001	0.0153	0.0158	0.0067
7	0.0002	0.1548	0.0229	0.0107

Table 4.5: RMSEs via different scenarios using different approaches under the full model.

Scenarios		1	2	3	4	5	6	7
Linear	NoTL-KR	1144	520	846	497	496	583	580
	TL-KR	945	365	660	431	484	486	461
	Weight-KR	0.9892	0.9847	0.9806	0.6028	1.0000	0.9781	0.9635
	Proposal-KR	**947**	**366**	**665**	**455**	**484**	**488**	**466**
	NoTL-LSTM	331	202	646	216	124	227	236
	TL-LSTM	276	197	574	187	103	365	210
	Weight-LSTM	0.7174	0.8332	0.8297	0.3681	1.0000	0.0776	0.4477
	Proposal-LSTM	**290**	195	**587**	**185**	**103**	**234**	**213**
	NoTL-CNN	415	366	675	342	396	532	426
	TL-CNN	517	329	609	440	399	443	498
	Weight-CNN	0.3667	0.9824	0.9816	0.0754	0.9999	0.9757	0.1655
	Proposal-CNN	**439**	**328**	**610**	**342**	399	**445**	**431**
Polynomial	NoTL-KR	650	514	815	526	428	634	617
	TL-KR	467	327	651	491	466	509	489
	Weight-KR	0.7882	0.8852	0.4078	0.5665	0.9958	0.6382	0.8614
	Proposal-KR	**494**	**347**	**757**	**506**	466	**557**	**508**
	NoTL-LSTM	331	202	646	216	124	227	236
	TL-LSTM	459	256	634	388	99	333	336
	Weight-LSTM	0.1945	0.0062	0.2361	0.0035	0.6787	0.0279	0.0042
	Proposal-LSTM	**344**	**202**	643	**217**	**98**	**229**	**237**
	NoTL-CNN	415	366	675	342	396	532	426
	TL-CNN	321	404	639	468	399	506	546
	Weight-CNN	0.4718	0.0395	0.4130	0.0138	0.5025	0.6614	0.0193
	Proposal-CNN	**362**	**367**	660	**342**	402	**514**	**427**
Rbf	NoTL-KR	190	474	741	516	417	596	572
	TL-KR	256	425	688	533	429	582	557
	Weight-KR	0.1624	0.3549	0.0310	0.5318	0.9234	0.1696	0.2266
	Proposal-KR	**195**	**454**	740	524	428	594	569
	NoTL-LSTM	331	202	646	216	124	227	236
	TL-LSTM	402	374	617	502	191	394	566
	Weight-LSTM	0.0242	0.0220	0.0335	0	0.9205	0.0022	0
	Proposal-LSTM	**331**	**203**	645	**216**	185	**227**	**237**
	NoTL-CNN	415	366	675	342	396	532	426
	TL-CNN	304	411	623	477	394	577	533
	Weight-CNN	0.1584	0.3509	0.0076	0.0949	0.7795	0.0142	0.0746
	Proposal-CNN	**393**	379	674	**344**	393	**532**	**429**
Laplace	NoTL-KR	149	568	804	565	419	649	631
	TL-KR	251	424	690	531	426	582	557
	Weight-KR	0.0190	0.0073	0	0	0.9993	0	0
	Proposal-KR	**149**	567	804	565	426	649	631
	NoTL-LSTM	331	202	646	216	124	227	236
	TL-LSTM	477	400	643	573	193	547	499
	Weight-LSTM	0.0172	0.0085	0	0	0.9993	0	0
	Proposal-LSTM	331	**203**	646	**216**	193	**227**	**237**
	NoTL-CNN	415	366	675	342	396	532	426
	TL-CNN	308	405	649	514	398	516	530
	Weight-CNN	0.0168	0.0091	0	0	0.9938	0	0.0002
	Proposal-CNN	412	**367**	675	**342**	398	532	**426**

5

A FUNCTIONAL DATA-DRIVEN METHOD FOR MODELING DEGRADATION OF WAXY LUBRICATION LAYER

Parts of this chapter have been published in *Quality and Reliability Engineering International* 40.7 (2024), pp. 3751–3773. DOI: 10.1002/qre.3622.

PLAIN SUMMARY

Wax is a prevalent lubrication material extensively employed in various engineering applications. Understanding the degradation characteristics of the waxy lubrication layer under diverse stress variables and levels is crucial for ensuring system security and reliability. Due to the unclear mechanism governing the degradation of the waxy lubrication layer under different stress variables, existing degradation models are unsuitable for modeling waxy lubrication layer degradation data. To address this challenge, we propose a functional data-driven method leveraging dense observations of waxy degradation. Through extensive simulations and a case study, we demonstrate the superior performance and effectiveness of the proposed approach.

5.1. INTRODUCTION

THE safe storage and transportation of explosives has always been a major concern, with significant implications for public safety and environmental protection. In this context, wax serves as an effective lubrication layer that helps reduce the sensitivity of explosives to external factors [113–116]. Besides its economical and practical benefits, wax stands out from other desensitizing materials due to its ability to spread easily across crystal surfaces. This property ensures that energetic crystals are thoroughly coated, enhancing their safety and stability [117, 118]. However, the challenges posed during the transportation and storage of explosives arise from the inherent characteristics of wax. High temperature and pressure can induce the migration of the wax layer, thereby reducing the thickness of the lubrication interface. This migration not only compromises the effectiveness of desensitization but also amplifies the safety risks associated with explosives [119]. This migration highlights the necessity of understanding how temperature and pressure affect the reduction in the thickness of the waxy lubrication layer. In this work, this reduction is referred to as degradation behavior or creep behavior, and its measurements can be utilized as a degradation or creep index.

Moreover, the relatively low softening point of wax makes it prone to plastic deformation when exposed to both temperature and pressure. Compared

to isolated stressors, the combined effects of thermal and mechanical stress significantly worsen the degradation of the waxy lubrication layer surrounding explosive crystals. In Figure 5.1, we present four examples illustrating the degradation curves of the waxy lubrication layer under various temperatures and pressures. The shear strain exhibits distinct variations across different stress levels in these examples. Thus, it is essential to examine the degradation behavior of the waxy layer in terms of thickness across a range of temperature and stress conditions.

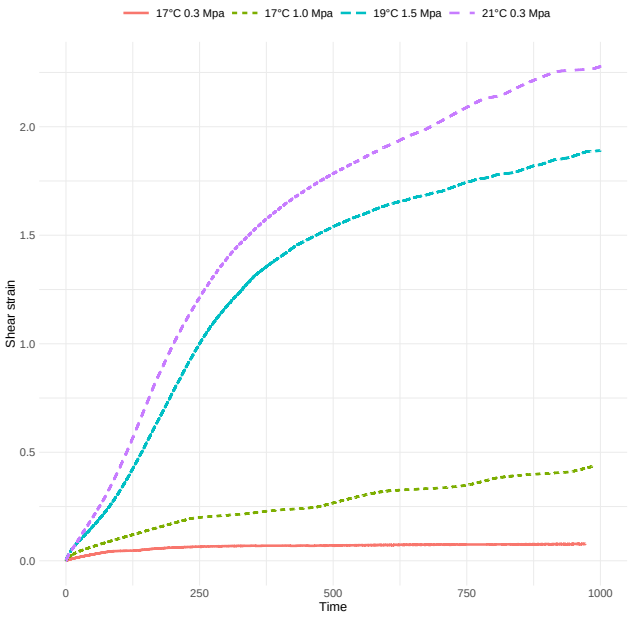


Figure 5.1: Examples of wax degradation data at various stress levels.

The distinctive degradation trends under varying stress variables and levels, as shown in Figure 5.1, make fitting the degradation data with parametric models quite challenging. Furthermore, this complexity poses a challenge in the extrapolation of degradation curves from high-stress levels to lower-stress levels, which is critical in engineering applications.

When subjected to deformation, the wax lubrication layer exhibits a

viscoelastic creep, followed by structural fracture and breakdown [120]. A comprehensive understanding of the waxy lubrication layer's behavior requires consideration of both creep and yield processes[121–123].

Recent research focuses on exploring the complexities of the creep and yield behavior of the waxy lubrication layer. It is worth noting that conventional models, such as the Burger mechanical-analogy model, which combines a Maxwell model with a Kelvin-Voigt model in series, have certain limitations when applied to describe the nonlinear creep phenomena in the waxy lubrication layer. While effective in certain contexts, these models may fall short of representing the entirety of the creep and yield processes [124].

Considering modeling the creep process of the wax, [125] conducts experimental investigations to analyze the creep behavior of the wax and proposes a mechanical-analogical creep equation based on their findings. As this model does not fully capture the accelerated creep phase that occurs following the yielding process, a more comprehensive nonlinear creep damage model has been proposed by [126], which is capable of representing the entire creep and yield process.

However, these models primarily focused on fitting wax creep at specific stress levels, overlooking the interaction of various experimental stress variables and neglecting the extrapolation of the fitted model to normal stress levels. To the best of our knowledge, there is limited research emphasizing stress-based extrapolation methods for wax creep models, especially in the context of simultaneously considering temperature and pressure stress variables. Extrapolation is crucial in practical applications, enabling the use of wax creep data collected under high-stress conditions to predict and extrapolate creep behavior at normal working stress levels. The necessity for this extrapolation is common in accelerated degradation experiments; for additional details, we refer to [49]. Such extrapolation not only holds the potential to enhance our comprehension of how the waxy lubrication layer responds to diverse stress variables and levels but also carries significant practical implications. This is especially relevant in industries where precise predictions of creep behavior are crucial for upholding the safety of equipment.

Additionally, existing approaches for fitting wax creep require a comprehensive

understanding of the physical mechanisms involved, thereby limiting their practical applications. For instance, when considering the creep behavior illustrated in Figure 5.1, our understanding of the mechanisms governing this particular type of waxy lubrication layer under both temperature and pressure stress variables is inadequate. However, the densely collected data provides an opportunity for the application of effective nonparametric approaches.

The aforementioned concerns have motivated this work. Specifically, it is crucial to investigate the relationship between the creep characteristics of the waxy lubrication layer and two experimental stresses: temperature and pressure. However, there are certain challenges to resolving this task due to uncertainty regarding the physical mechanism of wax lubrication layer degradation and the potential interaction between experimental stresses. Additionally, the shape and value range of the data under different stress levels show substantial inconsistencies, as depicted in Figure 5.1. Furthermore, while there is only one sample per stress level, advancements in modern sensor technology have enabled the collection of dense data for each sample. This suggests that nonparametric methods as developed in the field of functional data analysis (FDA), would offer better fitting performance. Therefore, this work introduces an innovative functional data-driven method for modeling the degradation data characterized by both the absence of a clear physical mechanism and the densely collected, highly flexible nature of the degradation curves.

In contrast to traditional approaches in the literature, the proposed method is entirely model-free and does not rely heavily on the choice of a physical mechanism. To extrapolate the degradation behavior to normal stress levels, a critical aspect in accelerated degradation tests, a quadratic model derived from features extracted based on functional data is proposed. The basic principles of this chapter are shown in Figure 5.2, underscoring the significance of nonparametric modeling in accelerated degradation data analysis. The main contributions of this work can be summarized as follows:

(1) Addressing the challenge of modeling wax creep behavior under the combined influence of temperature and pressure stress variables, this study provides a comprehensive analysis of this complex phenomenon.

(2) Utilizing a nonparametric approach based on FDA, this study offers a fresh perspective to fitting degradation data, diverging from traditional parametric methods.

(3) Proposing a pioneering data-driven framework, this research aims to model the intricate relationships between degradation patterns, various stress variables, and stress levels, introducing a novel model-free methodology for understanding such relationships in accelerated degradation tests.

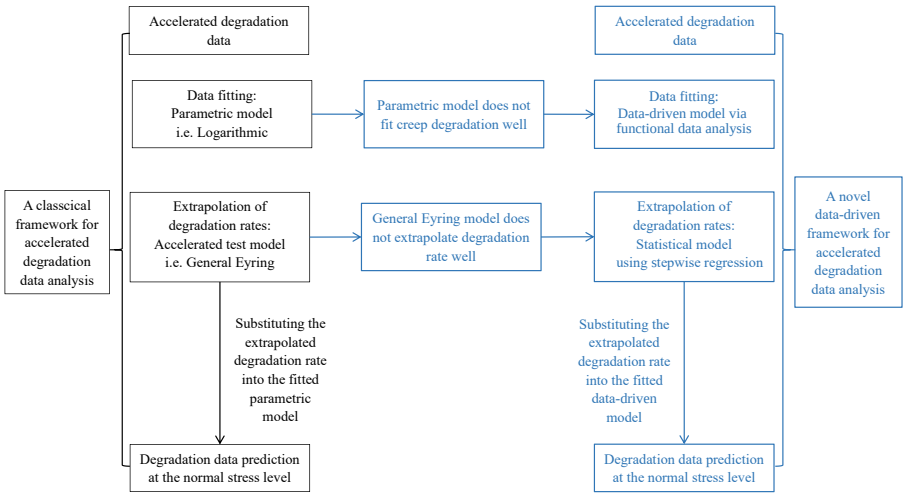


Figure 5.2: The basic principles for the proposed functional data-driven framework for modeling degradation of waxy lubrication layer.

The remainder of this chapter is structured as follows. In Section 5.2, we briefly review the details of traditional accelerated degradation data analysis methods and the fundamentals of FDA. The proposed functional data-driven framework is presented in Section 5.3. Section 5.4 presents simulation studies illustrating the performance of the proposed framework. Subsequently, detailed comparisons are conducted by using the waxy lubrication layer dataset in Section 5.5, where the limitations of the traditional data analysis framework are also underscored. Finally, concluding remarks are given in Section 5.6.

5.2. RELATED WORKS

To describe the innovation of the proposed method, here we briefly review the traditional models for degradation data in the reliability field and the methods for fitting the functional data in the literature.

5.2.1. TRADITIONAL METHODS FOR DEGRADATION DATA ANALYSIS

In the context of the accelerated degradation data, the data analysis process can be divided into two phases: data fitting and stress extrapolation, both of which are essential for practical application [34, 127]. The fundamental goal of data fitting is to construct a mathematical model that accurately describes the observed behavior of the system under accelerated degradation conditions. On the other hand, extrapolation becomes indispensable to extend our comprehension beyond the tested stress conditions, enabling predictions of the system's performance under normal or diverse stress levels.

When the physical mechanism is unknown, some approaches are available, which can be broadly divided into two categories: stochastic processes or the general path. The widely used stochastic processes, such as the Wiener[2, 39, 128, 129], gamma[72, 130], inverse Gaussian processes[69, 131], and their various variations [35, 40, 46, 132–136], have been extensively applied in various studies. However, stochastic process methods are often susceptible to issues of model misspecification, thereby limiting their application in engineering. For the general path models, which are widely used due to their flexibility in modeling the unit-to-unit variation, more references can be found in the recent representative work [137] and references therein. The configuration of the general path is usually shaped by the requirements of the specific application problem, taking into account the unique characteristics of the degradation paths to ensure a tailored and effective solution [138]. However, the assumption that all units share the same regular functional form for degradation paths [139] may not be met in practice, thus limiting their practical applications.

After data fitting, the subsequent step, as previously mentioned, involves extrapolating the degradation model to the normal stress levels. Regarding the extrapolation model, commonly utilized ones include the Arrhenius model, the inverse power law model, and the exponential model [140].

Most studies focus on integrating the aforementioned fitting approaches with these extrapolation models. However, these works primarily concentrate on cases involving a single stress variable, which does not align with the wax creep described in Section 5.1. There are few works dedicated to considering the problem with two stress variables. For example, [4] examines positive increments for a degradation process and employs the gamma process and the generalized Eyring model (GEM) to analyze data collected from a two-type constant stress accelerated degradation test. [141] introduces an analytical approach to design the experiment and describes the accelerated degradation process using a linear mixed-effects model. However, these methods are not available for modeling degradation data of the waxy lubrication layer, particularly given the unclear understanding of its underlying physical mechanisms.

In the case of wax creep, as elaborated in Section 5.1, there exist established physical models for describing creep behavior. Therefore, in this chapter, we opt for the utilization of a commonly employed model in the analysis of wax creep. This model primarily focuses on characterizing the instantaneous elastic strain, deceleration creep stage, and steady-state creep stage at a fixed temperature, while not explicitly accounting for yield stage considerations, as described in [126]. This model is based on the Riemann Liouville theory [126] and can be formulated as Eq. (5.1) when the temperature stress is fixed and only pressure stress is considered.

$$y_T(t,P) = \frac{P}{\eta}\frac{t^{\beta}}{\Gamma(1+\beta)}, \tag{5.1}$$

where $y_T(t,P)$ represents the reduction in the thickness of the wax at a specific temperature stress level T concerning time t and pressure stress level P, $\Gamma(\cdot)$ represents the gamma function, $\beta \in [0,1]$ is a tuning parameter, and η is the viscosity coefficient. Both β and η are chosen through cross-validation. For a detailed explanation of the physical meaning of Eq. (5.1) and its internal parameters, we refer to [126].

The conventional model outlined in Eq. (5.1) effectively serves its purpose in data fitting but comes with inherent limitations. It is restricted to fitting data at specific temperatures and falls short of comprehensively accounting for the interactions between temperature and pressure stress. Furthermore, this

model is inherently confined to data fitting and lacks the capability for stress extrapolation.

Considering the aforementioned issues in existing works, there is a crucial demand for a more robust and model-free degradation data analysis framework for wax creep. Fortunately, advancements in data collection and storage technologies have provided access to substantial real-time datasets, exemplified by the wax creep data depicted in Figure 5.1, collected through sensor technology. Such datasets, characterized by dense sampling or continuous monitoring, enable the application of nonparametric methods for data fitting, with FDA emerging as a representative and comprehensive nonparametric approach. A brief review of the FDA will be presented in the following subsection.

5.2.2. FUNCTIONAL DATA ANALYSIS

FUNCTIONAL data, as introduced by [142], refers to data where the observations are entire functions or curves rather than individual data points. Traditional statistical methods struggle when dealing with the complexities of these intricate data structures. Functional data often involve large amounts of high-frequency, real-time, and streaming data, which can be represented as curves [143] or images [144]. The densely sampled or observed characteristics of such data align with the wax creep behavior shown in Figure 5.1. The high dimension of this data presents significant challenges both in theory and computational capability. Recognizing this challenge, [145] developed specialized analytical methods tailored to functional data. After its initiation, FDA has found applications in various scientific fields, such as econometrics [146] and automated plant phenotyping [147]. For additional references, please see [148].

Denote functional data as $y(\cdot)$, which is defined over a domain $\mathcal{T}$. Furthermore, they can also be conceptualized as realizations of a square-integrable stochastic process, which is characterized by a mean function $\mu(t) = \mathbb{E}\big(y(t)\big)$, and a covariance function $\Sigma(s,t) = \mathrm{Cov}\big(y(s), y(t)\big)$. In this chapter, observed data are represented as $y_i(\cdot)$, with the sample population assumed to consist of N independent subjects. Let $\mathcal{T} = [0, T_0]$ represent the range of values for the time index, where 0 and T_0 denote the start and end of the data collection period,

respectively. To simplify the analysis, we assume uniform sampling schedules, designated as $0 = t_1 < t_2 < \cdots < t_{M-1} < t_M = T_0$ for every subject. Consequently, the corresponding observations are structured as $\boldsymbol{y}_i = [y_{i,1}, y_{i,2}, \cdots, y_{i,M}]$, where $y_{i,j} = y_i(t_j) + e_{i,j},\ i = 1,2,\cdots,N,\ j = 1,2,\cdots,M,\ y_i(t_j)$ is the true value of the i-th subject at the j-th time point, and random noise terms $e_{i,j}$ satisfying the conditions $E(e_{i,j}) = 0$, $\mathrm{Var}(e_{i,j}) = \sigma_e^2$. These random errors are assumed to be independent across subjects i and time points j, which are often interpreted as measurement errors.

The mean and the combination of $\Sigma(s,t)$, the covariance function, and σ_e^2, the variance of the noise term, can then be empirically estimated at the measurement times using the sample mean and sample covariance. Note that these two estimated functions are point-wise, which are given by $\hat{\mu}(t_j) = \frac{1}{N}\sum_{i=1}^{N} y_{i,j}$ and $\frac{1}{N}\sum_{i=1}^{N}(y_{i,k} - \hat{\mu}(t_k))(y_{i,l} - \hat{\mu}(t_l))$ for $k \neq l$. Then, continuous estimates of the mean and covariance functions over $\mathcal{T}$ can be approximated by smooth interpolation.

Smoothing is the first step in FDA, and its purpose is to convert raw discrete data points into a smoothly varying function. This emphasizes patterns in the data by minimizing short-term deviations due to observational errors, such as measurement errors or inherent system noise. There are various smoothing methods that are useful for the FDA, including local least squares and spline smoothing, for which many excellent references exist [149–151]. As mentioned by [152], it is essential to explicitly mention the chosen smoothing approach. In this chapter, we utilize B-spline smoothing, a widely adopted method known for its simplicity and widespread application. Specifically, to fit the discrete observations, a basis function expansion for $y_i(t_j)$ is used as

$$y_i(t_j) \approx \sum_{l=1}^{L} c_{i,l}\phi_l(t_j). \tag{5.2}$$

where L is the total number of B-splines, $c_{i,l}, \phi_l(\cdot), l = 1,\cdots,L$ are the corresponding unknown coefficients and B-splines, respectively.

B-splines can be defined by dividing the interval over which $y(\cdot)$ is approximated into $m+1$ sub-intervals with a sequence of knots. On each interval, polynomial functions of a specified order k (with a degree of $k-1$) are

used to fit the data.

5.3. THE FUNCTIONAL DATA-DRIVEN APPROACH

IN this section, we will introduce the details of the proposed functional data-driven framework. This encompasses aspects such as data fitting, extrapolation of model parameters, predictions, and an overview providing insights into the structure of the proposed framework.

5.3.1. DATA FITTING AND EXTRAPOLATION

IT is worth noting that observations of wax creep can not be negative. Thus, we directly approximate the logarithm of the original observed data $X(t)$, which is denoted as $y(t)$. Then we have

$$y_{i,j} = y_i(t_j) + e_{i,j} = \sum_{l=1}^{L} c_{i,l}\phi_l(t_j) + e_{i,j} = \boldsymbol{c}_i^\top \boldsymbol{\phi}(t) + e_{i,j}, \tag{5.3}$$

where $\boldsymbol{c}_i = (c_{i,1}, \cdots, c_{i,L})^\top$, $\boldsymbol{\phi}(t) = [\phi_1(t), \cdots, \phi_L(t)]^\top$. $\boldsymbol{c}_i$ can be estimated from the data via the unconstrained least squares method.

Specifically,

$$\hat{\boldsymbol{c}}_i = (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \boldsymbol{y}_{i,M}, \tag{5.4}$$

where $\boldsymbol{b} = [\mathbf{b}_1, \cdots, \mathbf{b}_L]$, $\mathbf{b}_l = [\phi_l(t_1), \cdots, \phi_l(t_M)]^\top$, and $\boldsymbol{y}_{i,M} = [y_{i,1}, \cdots, y_{i,M}]^\top$.

Thus, the estimator $\hat{X}_i(\cdot)$ is given by

$$\hat{X}_i(t) = \exp(\hat{y}_i(t)) = \exp(\sum_{l=1}^{L} \hat{c}_{i,l}\phi_l(t)). \tag{5.5}$$

Additionally, defining $||\phi||^2 = \mathbb{E}[\phi(x)^2]$, the rate convergence, as well as the corresponding asymptotic distribution of our estimator, are shown in the following theorem. Assumptions and proofs are presented in the APPENDIX B.

Theorem 5.1. *Under the assumptions stated in APPENDIX B, it follows that*

$$\frac{\hat{y}_i(t) - y_i(t)}{\sigma(t)} \rightsquigarrow N(0,1), \ \textit{with } M \to +\infty, \tag{5.6}$$

$$||\hat{y}_i - y_i|| = O_p(m^{-k} + \sqrt{m/M}), \tag{5.7}$$

for $i = 1, \cdots, N$, *where*

$$\sigma^2(t) = \sigma_e^2 \boldsymbol{\phi}(t)^\top (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{\phi}(t), \tag{5.8}$$

$\boldsymbol{\phi}(t) = [\phi_1(t), \cdots, \phi_L(t)]^\top$, $\boldsymbol{b} = [\mathbf{b}_1, \cdots, \mathbf{b}_L]$, $\mathbf{b}_l = [\phi_l(t_1), \cdots, \phi_l(t_M)]^\top$, σ_e^2 *is the variance of measurement error, and* $\rightsquigarrow$ *denotes converge in distribution.*

Obviously, the fitting performance of Eq. (5.3) relies on the number of B-splines, L, and a larger value of L is necessary for better fitting quality. Consequently, this leads to challenges in interpreting the degradation behavior based on these coefficients, especially in the case of accelerated degradation. For better interpretability, functional principal component analysis (FPCA) could be utilized to extract the degradation trend. FPCA stands out as a prominent technique for extracting valuable insights from functional data. The objective of FPCA is to discern the primary sources of variation within a collection of realizations stemming from a stochastic process, preserving the maximum amount of total variation possible. FPCA leverages a linear combination of functional principal components (FPCs), which are both uncorrelated and systematically ordered, to provide an accurate approximation of an infinite-dimensional function or curve. Meanwhile, it selects the top-performing FPCs for dimensionality reduction, which captures and retains the majority of the inherent variation. FPCA has taken off to become the most prevalent tool in the FDA, some applications include neuroimaging study [153], acceleration rate curves [154], and so on.

For FPCA, we aim to find the first FPC $\beta_1(t)$ to maximize

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^{N} \xi_{i,1}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\int_{\mathcal{T}} \beta_1(t) y_i(t)\, dt\right]^2 \\ &\text{s.t.} \quad \|\beta_1(t)\|^2 = \int_{\mathcal{T}} \left(\beta_1(t)\right)^2 dt = 1, \end{aligned} \tag{5.9}$$

where $\xi_{i,1}, i = 1, \cdots, N$ is the score of the i-th subject on the first principal component. Sequentially, j-th FPCs can be obtained by solving the following

constrained optimization problem

$$\begin{aligned} \max \frac{1}{N}\sum_{i=1}^{N}\xi_{i,j}^2 &= \frac{1}{N}\sum_{i=1}^{N}\left[\int_{\mathscr{T}}\beta_j(t)y_i(t)dt\right]^2 \\ \text{s.t.} \quad &\int_{\mathscr{T}}\left(\beta_j(t)\right)^2 dt = 1 \\ &\int_{\mathscr{T}}\beta_j(t)\beta_k(t)dt = 0, k = 1,\cdots,j-1. \end{aligned} \tag{5.10}$$

There is another characterization of FPCA introduced in [155], which is in terms of the eigenanalysis of the variance-covariance function or operator. Take the first FPC as an example, denote it as $\beta(t)$ for simplification. Define the estimated covariance function as

$$v(s,t) = \frac{1}{N}\sum_{i=1}^{N}(y_i(s)-\bar{y}(s))(y_i(t)-\bar{y}(t)), \tag{5.11}$$

where $\bar{y}(s)$ and $\bar{y}(t)$ denote the subject averages of the signals at times s and t. Then define

$$V\beta(s) = \int_{\mathscr{T}} v(s,t)\beta(t)dt = \lambda\beta(s), \tag{5.12}$$

where V is the covariance operator.

The cumulative variance contribution rate $\sum_{i=1}^{K}\lambda_i/\sum_{i=1}^{N-1}\lambda_i$ can be applied to indicate the amount of total variance accounted for by K FPCs.

Then, via smoothing mentioned in the previous section, we have some fitted curves as $\hat{y}_i(t) = \sum_{l=1}^{L}\hat{c}_{i,l}\phi_l(t)$, $i = 1,\cdots,N$, where $\phi_l(t)$ are B-spline functions for $l = 1,\cdots,L$. Denote the parameter matrix as $\mathbf{C} = (\hat{c}_{i,l})_{N\times L}$ Furthermore, without loss of generality, assume $\mu(t) = 0$. Thus, the covariance function is

$$v(s,t) = \frac{1}{N}\boldsymbol{\phi}(s)\mathbf{C}^\top\mathbf{C}\boldsymbol{\phi}(t)^\top. \tag{5.13}$$

A common way to estimate FPCs is basis function expansion [155], suppose that $\beta(t)$ has an expansion via the same B-spline functions

$$\beta(t) = \sum_{l=1}^{L}a_l\phi_l(t) = \boldsymbol{\phi}(t)^\top\boldsymbol{a}. \tag{5.14}$$

Then, we have

$$\frac{1}{N}\boldsymbol{\phi}(t)\mathbf{C}^\top\mathbf{C}\mathbf{W}\boldsymbol{a} = \lambda\boldsymbol{\phi}(t)^\top\boldsymbol{a}, \tag{5.15}$$

where $\mathbf{W} = \int \boldsymbol{\phi}\boldsymbol{\phi}^\top$. It then implies a purely matrix equation

$$\begin{aligned} &\frac{1}{N}\mathbf{C}^\top\mathbf{C}\mathbf{W}\boldsymbol{a} = \lambda\boldsymbol{a}, \\ &\text{s.t.} \quad \boldsymbol{a}^\top\mathbf{W}\boldsymbol{a} = 1,\ \boldsymbol{a}_j^\top\mathbf{W}\boldsymbol{a}_k = 0,\ k \neq j. \end{aligned} \tag{5.16}$$

To get the required FPCs, we introduce the Cholesky decomposition $\mathbf{W} = \mathbf{L}^\top\mathbf{L}$, define $\boldsymbol{u} = \mathbf{L}\boldsymbol{a}$, solve the equivalent eigenvalue problem

$$\frac{1}{N}\mathbf{L}\mathbf{C}^\top\mathbf{C}\mathbf{L}^\top\boldsymbol{u} = \lambda\boldsymbol{u}, \tag{5.17}$$

and compute $\boldsymbol{a} = \mathbf{L}^{-1}\boldsymbol{u}$ for each eigenvector. It is worth noting that if basis functions are orthonormal, FPCA will reduce to traditional multivariate PCA of $\mathbf{C}$.

After the FPCA, we have fitted curves as

$$X_i(t) = \exp\left(\sum_{j=1}^{K}\xi_{i,j}\beta_j(t)\right), \tag{5.18}$$

where $\beta_j(t)$ are FPCs for $j = 1,\cdots,K$, obtained by a linear combination of B-spline functions. Note that K could be much smaller than L in Eq. (5.5), extracting the main information among these curves.

In the extrapolation of degradation curves from high-stress levels to lower-stress levels, it is reasonable to assume that the coefficients $\xi_{i,j}$ in Eq. (5.18) depend on the stress variables and levels, as their values correspondingly vary with different combinations of these variables and levels. Regression models can then be employed to model the relationship between the coefficients and combinations of these variables and levels, particularly in cases where a clear understanding of the physical mechanism is lacking. As mentioned in Section 5.2.1, we consider temperature T and pressure P as the experimental stress variables. To accurately capture the variations of $\xi_{i,j}$ concerning T and P, we employ the full quadratic model, incorporating T, P, P^2, T^2, TP,

$1/T$, $1/P$,$1/T^2$, $1/P^2$, T/P, and P/T as independent variables. This approach allows us to account for both main effects and interactions. Subsequently, stepwise regression, which is an easily calculated and widely utilized method, is employed to select the significant variables in the regression model and estimate the corresponding coefficients. Take $\xi_{i,1}$ as an example, the estimate based on stepwise regression is the solution to

$$\begin{aligned}\min \sum_{i=1}^{n} \Big(\xi_{i,1} - \big(\eta_1 T + \eta_2 T^2 + \eta_3 P + \eta_4 P^2 + \eta_5 TP + \\ \frac{\eta_6}{T} + \frac{\eta_7}{P} + \frac{\eta_8}{T^2} + \frac{\eta_9}{P^2} + \eta_{10}\frac{T}{P} + \eta_{11}\frac{P}{T}\big)\Big)^2.\end{aligned} \tag{5.19}$$

The aforementioned process encompasses the entire data fitting procedure, establishing the relationship between the estimated coefficients and the experimental stress levels on the training curves.

5

5.3.2. PREDICTION

AFTER solving Eq. (5.19), the predicted curve under normal stress T^* and P^* is as follows,

$$\hat{X}^*(t, T^*, P^*) = \exp\left(\sum_{j=1}^{K} \hat{\xi}_j(T^*, P^*)\beta_j(t)\right), \tag{5.20}$$

where $\hat{\xi}_j$ can be obtained by substituting T^* and P^* into the results of the fitted regression model.

However, unlike the training phase, it is not guaranteed that the predicted curve meets the non-decreasing trend of wax creep. To address that, we apply the FDA again to acquire a monotonic curve. We first derive a data set $\{(t_0, \hat{X}^*(t_0)), \cdots, (t_M, \hat{X}^*(t_M))\}$. And the FDA is used here to smooth these points. As introduced previously, we assume a basis function expansion Eq. (5.2) and estimate coefficients via the data set. Meanwhile, a constraint is imposed to ensure a non-decreasing prediction. According to [156], a sufficient condition for $\sum_{l=1}^{L} c_l\phi_l(t)$ to be non-decreasing is $c_{l-1} \le c_l$ for $l = 2, \cdots, L$. Denote $\mathscr{C}^L$ as the set of all $(c_1, \cdots, c_L) \in \mathscr{R}^L$ satisfying $c_{l-1} \le c_l$, the estimated coefficients can

then be obtained by

$$\tilde{\boldsymbol{c}} = \underset{\boldsymbol{c}\in\mathscr{C}^L}{\arg\min} \frac{1}{M}\sum_{i=1}^{M}\left(\ln \hat{X}^*(t_i, T^*, P^*) - \sum_{l=1}^{L} c_l \phi_l(t_i)\right)^2. \tag{5.21}$$

Finally, the non-negative and non-decreasing prediction $\tilde{X}^*(\cdot)$ under temperature T^* and pressure P^* is given by

$$\tilde{X}^*(t, T^*, P^*) = \exp\left(\sum_{l=1}^{L} \tilde{c}_l \phi_l(t)\right). \tag{5.22}$$

Consequently, the proposed functional data-driven framework for degradation analysis is established. Regarding degradation paths as curves, FDA is employed to represent them via some preset B-spline functions. Further, we apply FPCA to extract the main variation and use stepwise regression to model the corresponding principle component scores. The complete method is outlined in Algorithm 5.1.

Algorithm 5.1: The proposed functional data-driven framework for degradation data analysis.

Input: $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_N$: training data set; (T^*, P^*): testing conditions;
Output: $\tilde{X}^*(t)$: predicted degradation curve.

$y_i(t) = \ln(\boldsymbol{X}_i), i = 1, \cdots, N \leftarrow$ Logarithmic transformation;
$\phi_l(t), l = 1, \cdots, L \leftarrow$ B-spline functions with the preset L;
$\hat{\boldsymbol{c}}_i \leftarrow$ estimated expansion coefficients via Eq. (5.4);
$K \leftarrow$ minimal number of FPC to capture 99.99% of the total variation;
$\beta_1(t), \cdots, \beta_K(t) \leftarrow K$ functional principal components based on Eq. (5.10);
$\xi_{i,j}$ for $j = 1, \cdots, K$ and $i = 1, \cdots, N \leftarrow$ corresponding FPC scores;
$y_i(t) = \sum_{j=1}^{K} \xi_{i,j}\beta_j(t),\ i = 1, \cdots, N$;
for $j \leftarrow 1 : K$ **do**
  $\boldsymbol{\xi}_j \leftarrow [\xi_{1,j}, \cdots, \xi_{i,j}]$;
  $\hat{\xi}_j(\cdot) \leftarrow$ create a stepwise regression model with training conditions;
end for
$\hat{\xi}_j(T^*, P^*) \leftarrow$ predictions at new condition T^* and P^* for $j = 1, \cdots, K$;
$\hat{X}^*(t, T^*, P^*) = \exp\left(\sum_{j=1}^{K} \hat{\xi}_j(T^*, P^*)\beta_j(t)\right)$;
$\tilde{X}^*(t, T^*, P^*) \leftarrow$ monotonic estimator based on $\hat{X}^*(t, T^*, P^*)$.

5.4. SIMULATION STUDY

IN this section, two simulation studies are conducted to demonstrate the effectiveness of the proposed functional data-driven framework. Given that the physical model shown in Eq. (5.1) is not suitable for extrapolation, we exclude it from consideration in this section. For performance metrics, the goodness of fit (R^2) and root mean square error (RMSE) are employed.

The first simulation aims to demonstrate that when both the data fitting and extrapolation models are accurate, the proposed functional data-driven framework can still accomplish comparable results. To achieve this, an extrapolable parametric model is needed. Without loss of generality, two stress variables, denoted as S_1 and S_2, are considered in this simulation. Based on engineers' advice and the statistical characteristics of wax creep, as illustrated in Figure 5.1, we employ a logarithmic model, which can be formulated as

$$y(t, S_1, S_2) = \psi(S_1, S_2)\ln(t) + \varepsilon, \tag{5.23}$$

where $y(t, S_1, S_2)$ represents the value of creep with respect to time t, the first stress variable S_1, and second stress variable S_2, $\psi(S_1, S_2)$ represents the degradation rate under S_1, S_2, and ε represents the noise term.

Considering the temperature and pressure stress variables in wax creep, the well-established GEM is then employed in the stress extrapolation phase to describe the relationship between degradation rates and the experimental stress variables in wax creep. Let $\{(S_{1,k_1}, S_{2,k_2}), k_1 = 1, 2, \cdots, K_1, k_2 = 1, 2, \cdots, K_2\}$ be the combinations of two stress variables, where K_1 and K_2 denote the number of the first and second stress variable levels respectively. Then, the GEM is expressed as [4]

$$\psi(S_{1,k_1}, S_{2,k_2}) = \exp(\alpha_0 + \alpha_1 S_{1,k_1} + \alpha_2 S_{2,k_2} + \alpha_3 S_{1,k_1} S_{2,k_2}), \tag{5.24}$$

where $k_1 = 1, 2, \cdots, K_1, k_2 = 1, 2, \cdots, K_2, \psi(S_{1,k_1}, S_{2,k_2})$ is the degradation rate at (S_{1,k_1}, S_{2,k_2}). Subsequently, the logarithm of Eq. (5.24), which can be solved by the least squares estimation method, is derived as

$$\ln\psi_{(S_{1,k_1}, S_{2,k_2})} = \alpha_0 + \alpha_1 S_{1,k_1} + \alpha_2 S_{2,k_2} + \alpha_3 S_{1,k_1} S_{2,k_2}. \tag{5.25}$$

Specifically, in the first simulation, we employ the following model to generate the simulated data

$$y(t) = (\psi(S_1, S_2) + \varepsilon_1)\ln(t) + \varepsilon_2, \tag{5.26}$$

where $\varepsilon_i, i = 1,2$ are independent white noise terms following normal distributions. The following GEM is employed as the underlying extrapolation model

$$\psi(S_1, S_2) = \exp(\alpha_0 + \alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_1 S_2), \tag{5.27}$$

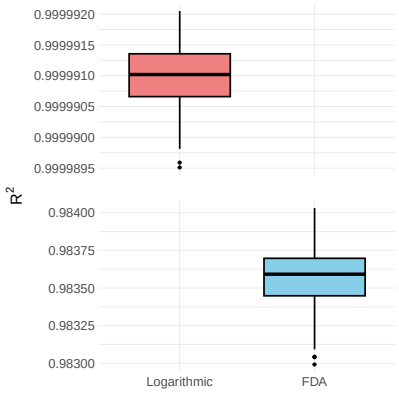
where $\alpha_j, j = 0,1,2,3$, represents the fitting parameters. In this simulation, $S_1 = S_2 = \{1,2,3,4\}, \alpha_0 = -1, \alpha_1 = \alpha_2 = 0.5, \alpha_3 = 0.2, \varepsilon_1 \sim N(0,0.1), \varepsilon_2 \sim N(0,0.1)$.

The results of this simulation are visually presented in Figure 5.3. Figure 5.3(a) and Figure 5.3(b) show the performance in the data fitting stage, from which we can find that the proposed functional data-driven framework provides comparable R^2 and smaller RMSE to the traditional approach. Note that the vertical axis has been separated in Figure 5.3(a) and Figure 5.3(b) for better visualization. Negligible differences in terms of R^2 and RMSE in the extrapolation stage are observed between the traditional and proposed functional data-driven methods, as shown in Figure 5.3(c) and Figure 5.3(d). Our findings indicate that when the underlying fitting model is logarithmic and the extrapolation model is GEM, our proposed functional data-driven framework performs comparably to the traditional approach in both the fitting and extrapolation stages.

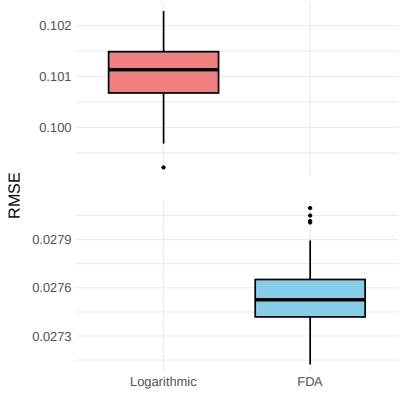
However, the accuracy of the model presented in Eq. (5.26) relies heavily on a thorough understanding of physical mechanisms, which is challenging to obtain in practice. A misspecified data-generating mechanism could lead to substantial inaccuracies, emphasizing the need for a model-free fitting method in such cases. To underscore the effectiveness of the proposed functional data-driven framework in the absence of clear physical mechanisms, the second simulation employs a data-generating mechanism based on B-spline functions. The model can be formulated as

$$y(t) = \sum_{i=1}^{4} (\psi_i(S_1, S_2) + \varepsilon_i)\phi_i(t) + \varepsilon_5, \tag{5.28}$$

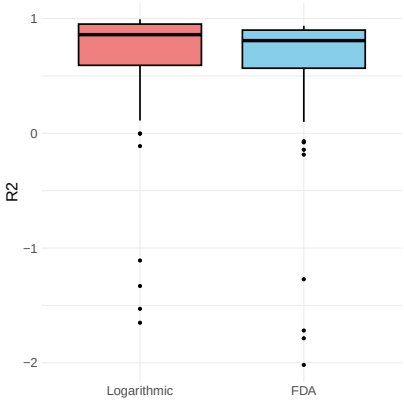
where c_i and $\phi_i(t)$ denote the coefficients and B-spline of the i-th term,



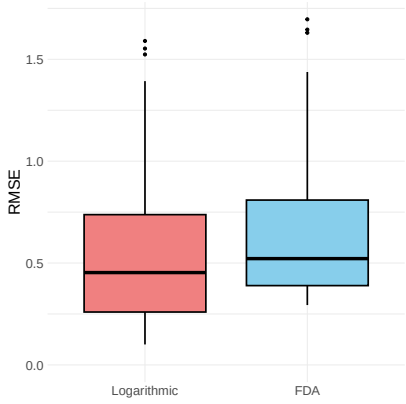
(a) R^2 for simulation 1 in the fitting stage.



(b) RMSE for simulation 1 in the fitting stage.



(c) R^2 for simulation 1 in the extrapolation stage.



(d) RMSE for simulation 1 in the extrapolation stage.

Figure 5.3: Performance evaluation for simulation 1.

$i = 1, \cdots, 4$, $\varepsilon_j, j = 1, \cdots, 5$, are white noise terms that follow normal distributions.

In this setting, we still use the GEM as the underlying extrapolation model, as shown in Eq. (5.29):

$$\psi_i(S_1, S_2) = \exp(\alpha_{0i} + \alpha_{1i} S_1 + \alpha_{2i} S_2 + \alpha_{3i} S_1 S_2), i = 1, \cdots, 4, \tag{5.29}$$

where $(\alpha_{j_1}, \cdots, \alpha_{j_4}) = (0.1, 0.2, 0.3, 0.3)$, $j = 0, 1, 2, 3$, to maintain the increasing degradation trend.

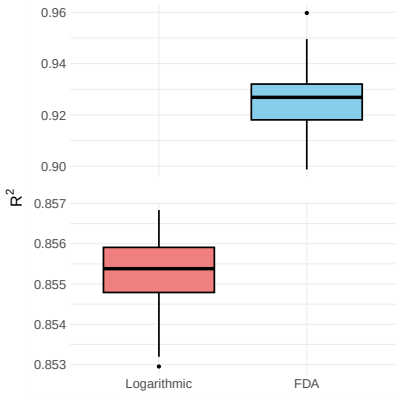
After conducting 100 repetitions, the results of this simulation are visually presented in Figure 5.4. Notably, Figure 5.4 suggests that in both fitting and extrapolation stages, when the underlying degradation model is unknown, even if the extrapolation model remains GEM, our proposed functional data-driven framework consistently outperforms the traditional method, as our approach provides larger R^2 and smaller RMSE. Similar to the first simulation, the vertical axis has also been separated in Figure 5.4(a) and Figure 5.4(b) for better visualization.

To enhance the comparison between the proposed functional data-driven framework and the traditional approach, the mean values of R^2 and RMSE based on 100 repetitions in both the fitting and extrapolation stages of the above two simulations are also reported. These results are summarized in Table 5.1.

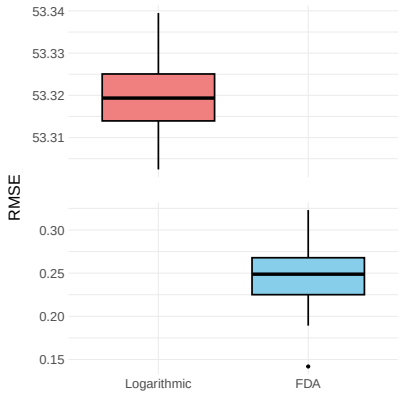
Table 5.1: Mean values based on 100 repetitions.

		Simulation 1		Simulation 2	
		Logarithmic	FDA	Logarithmic	FDA
Fitting	RMSE	0.1010506	0.0275388	53.31943	0.2484148
	R^2	0.999991	0.9835755	0.8552554	0.9256569
Extrapolation	RMSE	0.5415627	0.550598	0.4408871	0.2584508
	R^2	0.688867	0.6659219	0.4172232	0.7655868

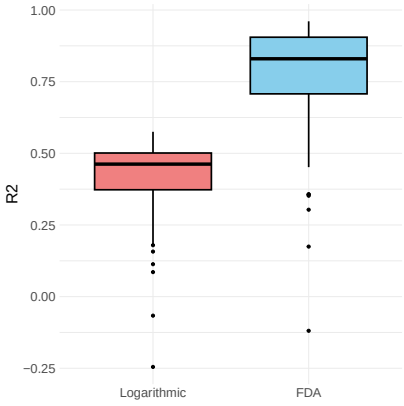
Based on the results from the two simulation studies presented in Figure 5.3, Figure 5.4, and Table 5.1, several key conclusions can be drawn. Firstly, when the underlying degradation model is from a known model, our proposed functional data-driven framework demonstrates performance comparable to that of the traditional approach in both the fitting and extrapolation stages. It is important to note that in our proposed framework, there is no assumption for the extrapolation model, and the comparable results confirm the feasibility of our approach in exploring the relationship between experimental stress levels and the fitted degradation model. Moreover, as expected, our proposed framework outperforms the traditional approach in both the fitting and extrapolation



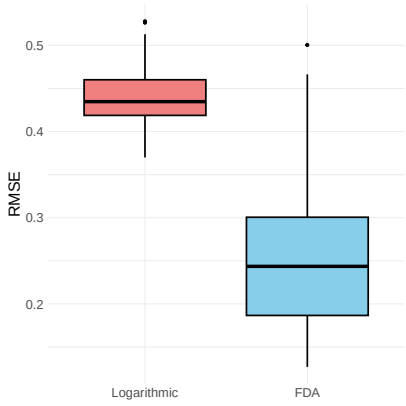
(a) R^2 for simulation 2 in the fitting stage.



(b) RMSE for simulation 2 in the fitting stage.



(c) R^2 for simulation 2 in the extrapolation stage.



(d) RMSE for simulation 2 in the extrapolation stage.

Figure 5.4: Performance evaluation for simulation 2.

stages when the underlying degradation model is unknown. These findings underscore the robustness of our proposed functional data-driven framework in accommodating different types of degradation and extrapolation models, showcasing its versatility and effectiveness in data analysis.

5.5. CASE STUDY

In this section, a case study based on real accelerated degradation data for wax lubrication layers is conducted to show the advantages of the proposed functional data-driven framework.

5.5.1. DATA OVERVIEW

A comprehensive set of 20 different scenarios (as outlined in Table 5.2) were meticulously executed. To capture the nuanced creep patterns exhibited by the lubrication layer under diverse accelerated aging tests, specialized sensors were employed, continuously monitoring and transmitting data throughout the experimentation, culminating in an overall recording duration of approximately 1,000 minutes. It is important to note that the sensor system used herein offers real-time and efficient tracking of the lubrication layer's creep behavior. Nevertheless, it is imperative to acknowledge that the frequency of sensor transmissions was not uniform across various test sets, resulting in disparities in time intervals and observation periods among the 20 datasets. Specifically, the data points ranged from a minimum of nearly 30,000 to a maximum of 230,000, aggregating to a dense data set.

Table 5.2: 20 different groups of all combinations of temperature and pressure where data are collected.

Scenario (NO.)	Temperature ($^{\circ}C$)	Pressure (MPa)	Scenario (NO.)	Temperature ($^{\circ}C$)	Pressure (MPa)
1	17	0.3	9	21	0.3
2	17	0.5	10	21	0.5
3	17	1	11	21	1
4	17	1.5	12	21	1.5
5	17	2	13	21	2
\	19	0.3	14	25	0.3
\	19	0.5	15	25	0.5
6	19	1	16	25	1
7	19	1.5	17	25	1.5
8	19	2	18	25	2

The experiments were conducted through a type of multi-axis automated thermal aging test system which functions as an aging box that can load temperature and pressure stress variables while monitoring the corresponding

creep. The schematic diagram of the experimental setup is shown in Figure 5.5.

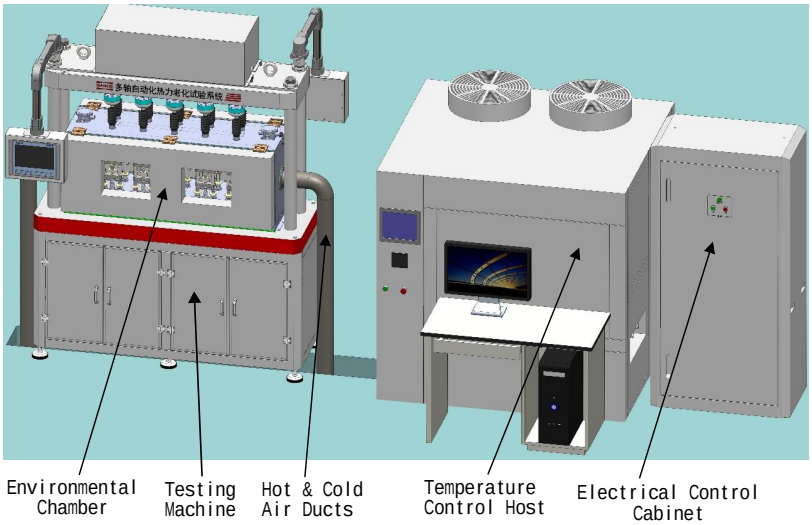


Figure 5.5: Schematic diagram of the wax degradation experimental setup.

Both the traditional and the proposed functional data-driven framework are individually used to analyze and extrapolate the observed degradation curves under accelerated stress conditions. During the preprocessing, the data collected at 19 $^{\circ}C$, 0.3 Mpa, as well as at 19 $^{\circ}C$, 0.5 Mpa, are removed based on the engineers' assessment, as the samples under these two operating conditions experienced abnormal faults during the experiment. Additionally, the data with $T^* = 17^{\circ}C$ and $P^* = 0.3$ Mpa are viewed as testing data, and others are training data. Note that the data is initialized by subtracting the initial value at $t = 0$ and for temperature, the absolute temperature is used.

5.5.2. FITTING THE ACCELERATED DEGRADATION DATA

We begin by evaluating the effectiveness of three distinct data fitting methods: the physical model described in Eq. (5.1), the logarithmic model defined in Eq. (5.23), and our proposed FDA model as presented in Eq. (5.3). The results of these fitting procedures are illustrated in Figure 5.6.

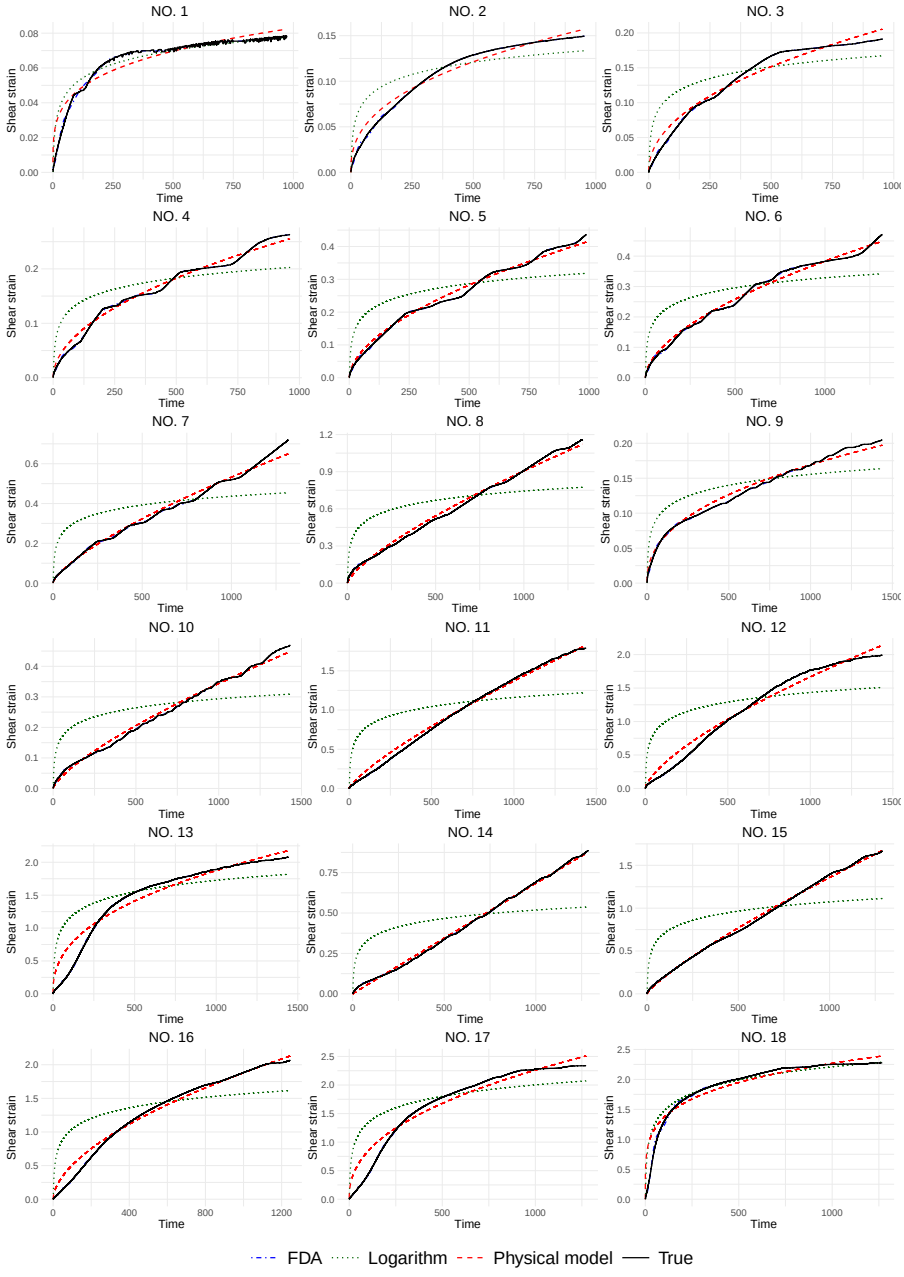


Figure 5.6: 18 true and fitted curves via different models.

Upon closer examination of Figure 5.6, it becomes evident that the functional data-driven framework consistently outperforms the other two methods, with the physical model coming in as the second-best performer. However, it is essential to note that, as discussed in Section 5.2.1, the physical model Eq. (5.1) lacks the ability to extrapolate degradation rates under working stress conditions, which diminishes its overall utility. Therefore, it becomes apparent that the FDA method stands as the optimal choice.

The logarithmic model, while demonstrating respectable performance under certain circumstances where the underlying creep trend closely aligns with a logarithmic pattern, falls short in other scenarios. This underscores the necessity and effectiveness of our proposed functional data-driven framework in data fitting, as it consistently outperforms both the physical and logarithmic models, making it the superior choice for this application.

Furthermore, an observation from the logarithmic model indicates that the current wax degradation data exhibits deviations from the assumptions of a purely logarithmic model in certain cases. Therefore, it becomes essential to categorize the data based on their specific characteristics. In this regard, the use of FDA, as demonstrated in our framework, offers an effective clustering approach, allowing us to account for variations in the underlying creep behavior and enabling more accurate and reliable data analysis. The approach is detailed as follows.

The cluster analysis plays a pivotal role in statistical research, automatically grouping observational data samples based on their inherent characteristics and proximity, all without prior knowledge. This process enhances our understanding of the structural characteristics within groups, where strong similarity exists among individuals, while weaker similarity characterizes individuals between groups. Such a method can examine if there are different degradation paths during all testing data. However, unlike traditional data, what we are studying here are curves. Therefore, functional data cluster analysis serves as an important tool for clustering degradation paths. Some attempts at functional data clustering methods include clustering time series [157], analyzing trajectory data [158], model-based clustering techniques [159] and categorizing by amplitude variation or phase variation [160], and so on. For an extensive

exploration of functional cluster methods, refer to the review paper [161].

The key point of the cluster approach is how to measure the distance between curves. From two different curves, $X_i = \left(X_{i,1}, \cdots, X_{i,M}\right)$ and $X_j = \left(X_{j,1}, \cdots, X_{j,M}\right)$, the difference can be defined directly as the L_p norm. Although such distance is easy to understand and can be directly connected with the classic clustering algorithm, it has two obvious flaws. One is that the observed data must be a noise-free realization at the same time point, which lacks universality in practical applications; the other is that dynamic features are not involved. Begin with the expansion Eq. (5.2), there are two common distances for the cluster, which are

$$\begin{aligned} d(X_i, X_j) &= \Big(\sum_{l=1}^{L} (c_{i,l} - c_{j,l})^p\Big)^{1/p}, \\ d_\theta(X_i, X_j) &= \Big(\int_{\mathscr{T}} (X_i^{(\theta)}(t) - X_j^{(\theta)}(t))^p dt\Big)^{1/p}, \end{aligned} \tag{5.30}$$

with $p \in \mathbb{R}$ and $\theta \in \mathbb{Z}$. These nonparametric clustering methods are based on the shape of the curve and its derivative function, providing multi-perspective clustering information. However, the cluster is sensitive to the choice of p and θ, which still needs further study. Then, adaptive model-based clustering is proposed by assuming that coefficients follow some specific distributions. In this chapter, we use such a procedure called funHDDC [162]. The foundation of this approach lies in a functional latent mixture model designed to accommodate functional data within group-specific functional subspaces. Through the imposition of constraints on model parameters both within and across groups, a versatile family of parsimonious models emerges, enabling their application to diverse scenarios. More specifically, a latent variable $\boldsymbol{Z} = (Z_1, \cdots, Z_K) \in \{0,1\}^K$ is introduced to indicate if X belongs to the k-th group ($Z_k = 1$) or not ($Z_k = 0$). Then, the aim is to predict the value $\boldsymbol{z}_i$ for each curve $X_i(t)$ with coefficients $\boldsymbol{c}_i$ that follow the Gaussian mixture distributions,

$$\pi(\boldsymbol{c}) = \sum_{k=1}^{K} \pi_k \phi_g(\boldsymbol{c}; \mu_k, \Sigma_k), \tag{5.31}$$

where $\phi_g(\cdot)$ is the standard Gaussian density function. All parameters can be estimated given coefficients by traditionally maximizing the likelihood through

the EM algorithm, for more details see [162].

5.5.3. EXTRAPOLATION TO NORMAL STRESS

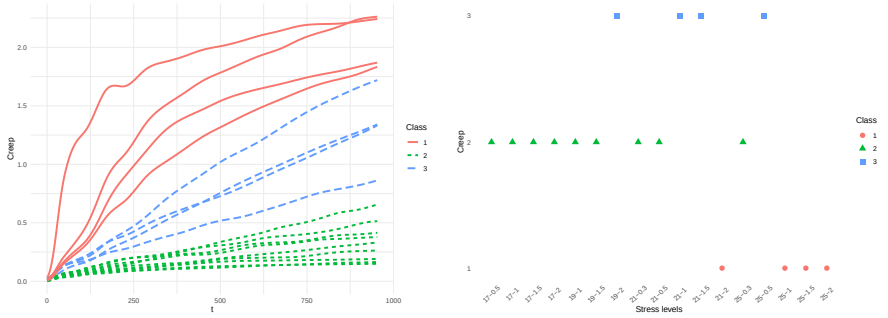
THE benchmark model used in the traditional approach is the logarithmic model shown in Eq. (5.23). This choice is made because the physical model in Eq. (5.1) is not suitable for extrapolation when changes in temperature and pressure are considered simultaneously. To establish the GEM for extrapolation after fitting the data with the traditional approach, stress analysis is conducted through analysis of variance (ANOVA) utilizing the estimated coefficients from Eq. (5.23). The results are detailed in Table 5.3, with a significance level set at 0.1. From the findings in Table 5.3, it is evident that the interaction effect between temperature and pressure is not statistically significant. This implies that the coefficient of the interaction term in the GEM Eq. (5.25), denoted as α_3, can be set to 0. The GEM demonstrates a well-fitted result with an R^2 of 0.8156, supported by a corresponding p-value of 0.006269, which is less than the significance level of 0.1. This statistical significance indicates that the GEM effectively captures the degradation rates of wax. Subsequently, we proceed to forecast wax creep under normal stress levels (T^*, P^*) by integrating these values into the fitted GEM.

Table 5.3: ANOVA results.

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Temperature	1	0.2135	0.2135	5.547	0.06514
Pressure	1	0.8705	0.8705	22.617	0.00508
Temperature : Pressure	1	0.0526	0.0526	1.367	0.29502
Residuals	5	0.1924	0.0385		

For the proposed functional data-driven framework, as discussed in Section 5.5.2, cluster analysis using FDA is needed before extrapolation. In this chapter, we use a general procedure for clustering, which is proposed by [162]. It employs a functional latent mixture model with some constraints on model parameters within and between groups. As seen from Figure 5.7(a) and Figure 5.7(b), there are three classes of curves, suggesting that training data

involve three different degradation patterns. Such three kinds of curves are different in location while red ones also have different fluctuations. According to three classes in Figure 5.7(a) and Figure 5.7(b), it seems that conditions with similar levels share similar creep curves, which coincides with the real case. Thus, we choose curves with lower stress levels (green dashed lines and triangles in Figure 5.7(a) and Figure 5.7(b)) as training data.



(a) FDA fitting curves for different classes. (b) Three classes with different stress levels.

Figure 5.7: Results for cluster analysis via FDA.

Via FPCA, the mean and variance functions can be calculated. As seen in Figure 5.8, the trend of degradation is increasing, coinciding with the real case. Additionally, the variance function, involving variances of $X_i(t)$ and the error term, also grows. Since the variance of e_{ij} is a constant, the trend shown in the variance function suggests that curves are similar at the beginning and differ from each other as time goes on. After the FPCA, four FPCs are obtained, as illustrated in Figure 5.9, from which we can find that the first three FPCs are representative enough to maintain the information as the total percentage of variability is more than 98.1%.

In the extrapolation phase, the stepwise regression results reveal that for the first FPC, the significant coefficients correspond to elements T^2, TP, and $\frac{P}{T}$. For the second FPC, the significant coefficients relate to elements T^2, TP, P^2, $\frac{1}{T^2}$, $\frac{P}{T}$, and $\frac{1}{P^2}$.These results suggest that temperature, pressure, and their interaction significantly influence the creep behavior of the waxy lubrication layer. This

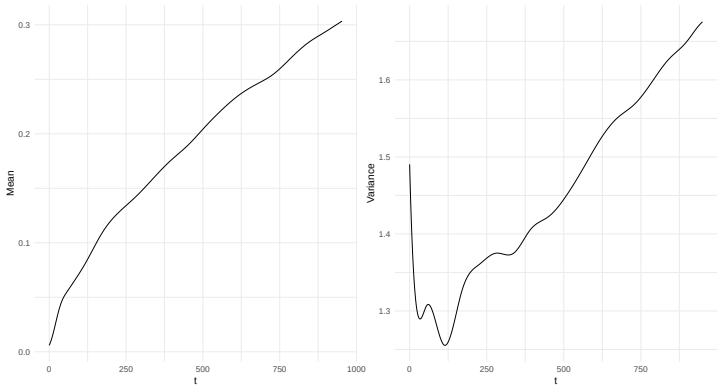


Figure 5.8: The functional mean (left) and variance (right).

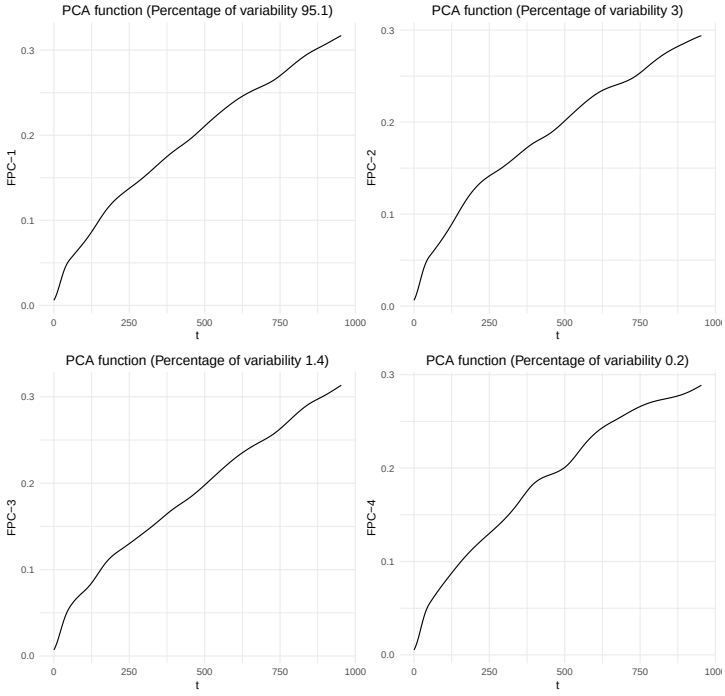


Figure 5.9: Four FPCs obtained via FPCA.

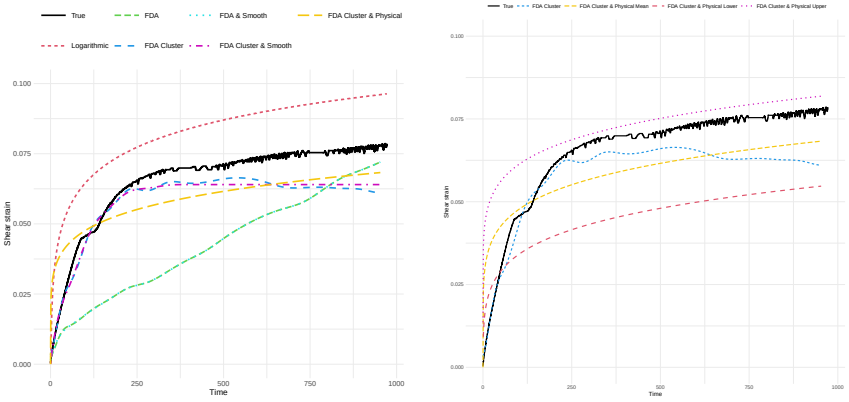
finding differs from the results in Table 5.3 based on the traditional approach. The outcomes from the proposed functional data-driven framework indicate that the interaction of temperature and pressure does influence creep behavior, offering valuable insights for practical applications. For example, it can provide engineers with guidelines to understand the physical mechanisms of the waxy lubrication layer under temperature and pressure stress variables. Engineers can conduct corresponding experiments to explore the exact influence of the interaction of temperature and pressure on the creep behavior.

To enhance visualization, Figure 5.10(a) displays various predictive outcomes. The solid black line represents the true values, the red dashed line corresponds to the predictions made by the logarithmic model, the green dashed line illustrates predictions without clustering in the FDA, the light-blue dotted line represents smoothed values (closely aligned with the non-smoothed ones in this case), the blue dashed line shows the results of the FDA considering clustering, and the purple dot-dashed line corresponds to the smoothed values, contributing significantly to their smoothness.

Notably, Figure 5.10(a) emphasizes the results of the FDA considering clustering as the closest match to the true values. However, it is essential to recognize that the trend of this prediction in the tail phase deviates from the expected increasing trend. Therefore, we also provide an alternative in the form of a fitted value based on the physical model in Eq. (5.1) and the prediction based on FDA, denoted by the yellow dashed line in Figure 5.10(a). This alternative not only demonstrates the effective prediction of creep but also offers a means for time extrapolation. Additionally, we also provide a quantitative assessment of the associated uncertainty by computing the 95% confidence interval, as depicted in Figure 5.10(b).

5.6. CONCLUSIONS

IN this study, we presented a novel dataset of accelerated creep behavior for waxy lubrication layers subjected to both temperature and pressure stress variables. Moreover, we introduced a functional data-driven framework for modeling the degradation data. This approach offers a novel perspective



(a) Prediction results based on different strategies.

(b) Prediction interval based on the FDA considering clustering and the physical model.

Figure 5.10: Prediction performance for the case study.

on fitting degradation data, departing from traditional modeling approaches. Through extensive simulations and a practical case study on waxy lubrication layers, our results have demonstrated the superior performance of the proposed functional data-driven approach in comparison to traditional models in terms of data fitting. Additionally, our framework exhibited remarkable capabilities in the extrapolation of degradation parameters. In a broader context, the functional data-driven method showcases enhanced robustness and broader applicability than existing methodologies, indicating its potential to address more complex real-world scenarios. Overall, this research offers valuable insights into the characterization and analysis of degradation data under diverse stress variables, opening avenues for more comprehensive and accurate assessments in engineering applications.

APPENDIX B

APPENDIX B.1 NOTATIONS

We assume that $\mathcal{T}$ is a compact subset of some Euclidean space. In what follows, via the sequence of knots τ_m, B-splines with degree $k-1$ forms a linear space with dimension $L = m + k$, denoted as $\mathcal{G}_m^k = \mathcal{G}(k, \tau_m)$. For any $g_1, g_2 \in \mathcal{G}_m^k$, set $\langle g_1, g_2 \rangle = \mathbb{E}[g_1(x)g_2(x)]$ and $\langle g_1, g_2 \rangle_M = (1/M)\sum_{i=1}^{M}[g_1(t_i)g_2(t_i)]$. For simplification, we use y instead of y_i during the proof.

APPENDIX B.2 ASSUMPTIONS

We need the following technical assumptions for our results.

(1) The function y is k times continuously differentiable for some $k \geq 2$, and has bounded $k-1$-th derivative.

(2) The density of t is bounded away from zero and infinity on $\mathcal{T}$.

(3) The knot sequence $\tau_m = \{s_0 < s_1 \cdots < s_{m+1}\}$ has a bounded mesh ratio. That is, there exists a constant c such that

$$\frac{\max\limits_{0 \leq i \leq m} s_{i+1} - s_i}{\min\limits_{0 \leq i \leq m} s_{i+1} - s_i} \leq c. \tag{5.32}$$

And $m \geq CM^{1/(2k+1)}$ for some constant $C > 0$.

(4) $m \to \infty$ and as $M \to \infty$, $m/M \to 0$, $M^{1/(2k+1)}/L \to 0$, and $L \log M/M \to 0$.

APPENDIX B.3 PROOF

Proof of Theorem 5.1. As noted by [163], asymptotic properties are the same whether the constraints of monotone non-decreasing are applied or not. Therefore, we consider an unconstrained estimator here, which is also denoted as $\hat{y}$. Then, we have

$$\hat{\boldsymbol{c}} = (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \boldsymbol{y}_M, \tag{5.33}$$

where $\boldsymbol{b} = [\mathbf{b}_1, \cdots, \mathbf{b}_L]$, $\mathbf{b}_l = [\phi_l(t_1), \cdots, \phi_l(t_M)]^\top$, and $\boldsymbol{y}_M = [y_1, \cdots, y_M]^\top$.

First we consider the asymptotic distributions, the estimator $\hat{y}(\cdot)$ is given by

$$\hat{y}(t) = \sum_{l=1}^{L} \hat{c}_l \phi_l(t) = \hat{\boldsymbol{c}}^\top \boldsymbol{\phi}(t). \tag{5.34}$$

Denote the true coefficients as $\boldsymbol{c} = [c_1, \cdots, c_L]^\top$, we have

$$\begin{aligned}
\mathbb{E}[\hat{y}(t) - y(t)] &= \mathbb{E}[\hat{\boldsymbol{c}}^\top \boldsymbol{\phi}(t) - \boldsymbol{c}^\top \boldsymbol{\phi}(t)] \\
&= \mathbb{E}[(\hat{\boldsymbol{c}} - \boldsymbol{c})^\top \boldsymbol{\phi}(t)] \\
&= ((\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \mathbb{E}[\boldsymbol{y}_M] - \boldsymbol{c})^\top \boldsymbol{\phi}(t) \\
&= ((\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \boldsymbol{b}\boldsymbol{c} - \boldsymbol{c})^\top \boldsymbol{\phi}(t) \\
&= 0.
\end{aligned} \tag{5.35}$$

Thus, $\hat{y}(\cdot)$ is an unbiased estimator. It then follows that

$$\begin{aligned}
\mathrm{Var}[\hat{y}(t) - y(t)] &= \mathrm{Var}[\hat{y}(t)] \\
&= \boldsymbol{\phi}(t)^\top \mathrm{Var}[(\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \boldsymbol{y}_M] \boldsymbol{\phi}(t) \\
&= \boldsymbol{\phi}(t)^\top (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \mathrm{Var}[\boldsymbol{y}_M] \boldsymbol{b} (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{\phi}(t) \\
&= \sigma_e^2 \boldsymbol{\phi}(t)^\top (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{b}^\top \boldsymbol{b} (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{\phi}(t) \\
&= \sigma_e^2 \boldsymbol{\phi}(t)^\top (\boldsymbol{b}^\top \boldsymbol{b})^{-1} \boldsymbol{\phi}(t),
\end{aligned} \tag{5.36}$$

where σ_e^2 is the variance of measurement error. Under some assumptions mentioned before, Theorem 3 and 4 in [164] indicate that the first conclusion follows.

Secondly, consider the linear space $\mathscr{G}_m^k$ and define $\tilde{y} = \mathbb{E}(\hat{y}|t)$, we have the decomposition

$$\hat{y} - y = [\hat{y} - \tilde{y}] + [\tilde{y} - y] \tag{5.37}$$

where $\hat{y} - \tilde{y}$ and $\tilde{y} - y$ are referred to as the approximation and estimation errors, respectively. For $F \in C^k(\mathscr{T})$, [149] indicates that there exists a $g \in \mathscr{G}_m^k$ and a constant c such that

$$||g - y||_\infty \leq c m^{-k}. \tag{5.38}$$

According to Lemma 5.1 and Theorem 5.1 in [165], it then follows that

$$||\tilde{y} - y||_\infty \leq \inf_{g \in \mathcal{G}_m^k} ||g - y||_\infty \leq cm^{-k}. \tag{5.39}$$

Therefore, $||\tilde{y} - y||_\infty = O_p(m^{-k})$. Based on Corollary 3.1 in [165], we have $\hat{y}(t) - \tilde{y}(t) = O_p(\sqrt{m/M})$ uniformly in $t \in \mathcal{T}$. Therefore, $\hat{y} - y = O_p(m^{-k} + \sqrt{m/M})$. The result on the L_2 norm follows directly from the one on the supremum norm. □

6

CONCLUSIONS AND FUTURE WORK

6.1. CONCLUSIONS

THE overall research objective of this thesis is to explore data fusion methods in the reliability analysis of industrial devices, particularly focusing on fusing data from different sources. The proposed methods demonstrate effectiveness and performance in the context of different types of industrial devices, and their applicability is not limited to the presented case studies but can also be extended to the reliability analysis of other industrial devices with similar data sources. As mentioned in Chapter 1, the overall research objective is divided into three specific research objectives addressed through four published or submitted journal papers. The main conclusions corresponding to each research objective are summarized as follows:

Research Objective 1: Propose a data fusion-based framework for RUL prediction of industrial devices that collect multi-channel sensor data.

Chapter 2 presents novel DI-based prognostic frameworks for predicting RUL in industrial devices collecting multivariate sensor data. The frameworks utilize a feature-based DI that performs automatic feature selection and captures the nonlinear nature of degradation. Validation through simulation studies and a case study on industrial induction motors shows that the proposed frameworks significantly outperform existing methods in predictive accuracy.

Research Objective 2: Develop data fusion-based reliability analysis methods for degradation data from different batches of industrial devices.

Chapter 3 addresses the reliability evaluation for high-reliability, long-life devices with limited current experimental data and abundant historical data. By integrating current and historical degradation data using a Wiener process-based model, the study provides accurate and stable reliability estimates. The consistency of failure mechanisms across different batches is verified using a likelihood ratio test, with practical applications demonstrated on MOSFET degradation data.

Chapter 4 introduces a robust TL framework that balances NoTL and TL outcomes using the MMD metric. The case study on early cycle battery data shows the framework's robustness in favoring TL results when TL is effective and favoring NoTL when the negative transfer exists, enhancing practical applicability.

Research Objective 3: Establish a data fusion-based framework for analyzing accelerated degradation data under different experimental stresses and stress levels.

Chapter 5 presents a novel dataset on the accelerated creep behavior of waxy lubrication layers under temperature and pressure stress. A functional data-driven framework is proposed for modeling degradation data, demonstrating superior performance in data fitting and parameter extrapolation compared to traditional methods. The framework's robustness and broader applicability suggest potential for addressing complex real-world scenarios.

In summary, this thesis provides several data fusion-based methods for reliability analysis across various industrial applications. However, certain challenges remain unaddressed and require further research, which are stated in the following section.

6.2. FUTURE WORK

BUILDING on the findings of this research, several avenues for future work related to the proposed three research objectives have been identified:

- **Research Objective 1** (Related to the work presented in Chapter 2.)
  - **Enhancing Optimization Algorithms:** The computational efficiency of the proposed prognostic frameworks can be further improved by exploring more efficient optimization algorithms or parallel computation techniques. This is particularly important for large datasets involving numerous reference units or experimental cycles.
  - **Leveraging Additional Data Sources:** Future research could develop methods to incorporate information from additional sources, such as data from different experimental settings. Advanced techniques like TL could be explored to enhance the accuracy and applicability of the proposed frameworks.
- **Research Objective 2** (Related to the work presented in Chapter 3 and Chapter 4.)

  - **Expanding Data Integration Methods:** Applying more historical data from various batches can improve the reliability estimation for newly developed devices. Constructing corresponding hypothesis testing problems will further validate the consistency of failure mechanisms across different batches. (Chapter 3.)
  - **Exploring Alternative Weights and Algorithms:** Investigating alternative forms of weights and TL algorithms could optimize the balance between NoTL and TL outcomes. Future research should also consider different kernel function types to tailor methods for specific problems. (Chapter 4.)

- **Research Objective 3** (Related to the work presented in Chapter 5.)
  - **Incorporating Monotonic Constraints:** Integrating monotonic constraints into principal components could improve the framework's applicability to degradation data exhibiting monotonic trends.
  - **Alternative Time Extrapolation Methods:** Exploring different time extrapolation methods can extend the versatility of the proposed framework. Integrating these methods with the FDA approach may offer more comprehensive and accurate assessments of degradation data under diverse stress variables.

BIBLIOGRAPHY

[1] Y. Hong, M. Zhang, and W. Q. Meeker. "Big data and reliability applications: The complexity dimension". In: *Journal of Quality Technology* 50.2 (2018), pp. 135–149. DOI: 10.1080/00224065.2018.1438007.

[2] W. Kang, Y. Tian, H. Xu, D. Wang, H. Zheng, M. Zhang, and H. Mu. "Reliability analysis based on the Wiener process integrated with historical degradation data". In: *Quality and Reliability Engineering International* 39.4 (2023), pp. 1376–1395. DOI: 10.1002/qre.3300.

[3] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, *et al.* "Data-driven prediction of battery cycle life before capacity degradation". In: *Nature Energy* 4.5 (2019), pp. 383–391. DOI: 10.1038/s41560-019-0356-8.

[4] T.-R. Tsai, W.-Y. Sung, Y. Lio, S. I. Chang, and J.-C. Lu. "Optimal two-variable accelerated degradation test plan for gamma degradation processes". In: *IEEE Transactions on Reliability* 65.1 (2015), pp. 459–468. DOI: 10.1109/tr.2015.2435774.

[5] C. E. Ebeling. *An introduction to reliability and maintainability engineering*. Waveland Press, 2019. URL: https://www.waveland.com/browse.php?t=392.

[6] W. Q. Meeker, L. A. Escobar, and F. G. Pascual. *Statistical methods for reliability data*. John Wiley & Sons, 2022. URL: https://www.wiley.com/go/meeker/reliability2e.

[7] J. F. Lawless. *Statistical models and methods for lifetime data*. John Wiley & Sons, 2011. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118033005.

[8] W. Q. Meeker, L. A. Escobar, and C. J. Lu. "Accelerated degradation tests: Modeling and analysis". In: *Technometrics* 40.2 (1998), pp. 89–99. DOI: 10.2307/1270643.

[9] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou. "Remaining useful life estimation–a review on the statistical data driven approaches". In: *European Journal of Operational Research* 213.1 (2011), pp. 1–14. DOI: 10.1016/j.ejor.2010.11.018.

[10] F. Yang, M. S. Habibullah, T. Zhang, Z. Xu, P. Lim, and S. Nadarajan. "Health index-based prognostics for remaining useful life predictions in electrical machines". In: *IEEE Transactions Industrial Electronics* 63.4 (2016), pp. 2633–2644. DOI: 10.1109/tie.2016.2515054.

[11] F. Yang, M. S. Habibullah, and Y. Shen. "Remaining useful life prediction of induction motors using nonlinear degradation of health index". In: *Mechanical Systems and Signal Processing* 148 (2021), p. 107183. DOI: 10.1016/j.ymssp.2020.107183.

[12] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. "Domain adaptation via transfer component analysis". In: *IEEE Transactions on Neural Networks* 22.2 (2010), pp. 199–210. DOI: 10.1109/tnn.2010.2091281.

[13] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan. *Transfer learning*. Cambridge University Press, 2020. DOI: 10.1017/9781139061773.

[14] M. E. Sharp. "Prognostic approaches using transient monitoring methods". PhD thesis. University of Tennessee, 2012. URL: https://trace.tennessee.edu/utk_graddiss/1431/.

[15] N. Eleftheroglou, D. Zarouchas, T. Loutas, R. Alderliesten, and R. Benedictus. "Structural health monitoring data fusion for in-situ life prognosis of composite structures". In: *Reliability Engineering & System Safety* 178 (2018), pp. 40–54. DOI: 10.1016/j.ress.2018.04.031.

[16] G. Niu, B.-S. Yang, and M. Pecht. "Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance". In: *Reliability Engineering & System Safety* 95.7 (2010), pp. 786–796. DOI: 10.1016/j.ress.2010.02.016.

[17] A. A. Broer, R. Benedictus, and D. Zarouchas. "The need for multi-sensor data fusion in structural health monitoring of composite aircraft structures". In: *Aerospace* 9.4 (2022), p. 183. DOI: 10.3390/aerospace9040183.

[18] M. Azamfar, J. Singh, I. Bravo-Imaz, and J. Lee. "Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis". In: *Mechanical Systems and Signal Processing* 144 (2020), p. 106861. DOI: 10.1016/j.ymssp.2020.106861.

[19] J. Chen, H. Jing, Y. Chang, and Q. Liu. "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process". In: *Reliability Engineering & System Safety* 185 (2019), pp. 372–382. DOI: 10.1016/j.ress.2019.01.006.

[20] X. Li, H. Jiang, Y. Liu, T. Wang, and Z. Li. "An integrated deep multiscale feature fusion network for aeroengine remaining useful life prediction with multisensor data". In: *Knowledge-Based Systems* 235 (2022), p. 107652. DOI: 10.1016/j.knosys.2021.107652.

[21] Y. Wei, D. Wu, and J. Terpenny. "Decision-level data fusion in quality control and predictive maintenance". In: *IEEE Transactions Automation Science and Engineering* 18.1 (2020), pp. 184–194. DOI: 10.1109/tase.2020.2964998.

[22] H. Shao, J. Lin, L. Zhang, D. Galar, and U. Kumar. "A novel approach of multisensory fusion to collaborative fault diagnosis in maintenance". In: *Information Fusion* 74 (2021), pp. 65–76. DOI: 10.1016/j.inffus.2021.03.008.

[23] M. Kim, C. Song, and K. Liu. "A generic health index approach for multisensor degradation modeling and sensor selection". In: *IEEE Transactions Automation Science and Engineering* 16.3 (2019), pp. 1426–1437. DOI: 10.1109/tase.2018.2890608.

[24] 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". In: *IEEE Transactions Automation Science and Engineering* 14.3 (2015), pp. 1504–1514. DOI: 10.1109/tase.2015.2446752.

[25] K. Liu, N. Z. Gebraeel, and J. Shi. "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis". In: *IEEE Transactions Automation Science and Engineering* 10.3 (2013), pp. 652–664. DOI: 10.1109/tase.2013.2250282.

[26] L. Gu, R. Zheng, Y. Zhou, Z. Zhang, and K. Zhao. "Remaining useful life prediction using composite health index and hybrid LSTM-SVR model". In: *Quality and Reliability Engineering International* 38.7 (2022), pp. 3559–3578. DOI: 10.1002/qre.3151.

[27] C. Song, K. Liu, and X. Zhang. "Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis". In: *IEEE Transactions Reliability* 67.2 (2017), pp. 640–650. DOI: 10.1109/tr.2017.2715180.

[28] A. Chehade, C. Song, K. Liu, A. Saxena, and X. Zhang. "A data-level fusion approach for degradation modeling and prognostic analysis under multiple failure modes". In: *Journal of Quality Technology* 50.2 (2018), pp. 150–165. DOI: 10.1080/00224065.2018.1436829.

[29] C. Song and K. Liu. "Statistical degradation modeling and prognostics of multiple sensor signals via data fusion: A composite health index approach". In: *IISE Transactions* 50.10 (2018), pp. 853–867. DOI: 10.1080/24725854.2018.1440673.

[30] D. Wang and K. Liu. "An integrated deep learning-based data fusion and degradation modeling method for improving prognostics". In: *IEEE Transactions Automation Science and Engineering* (2023), pp. 1–14. DOI: 10.1109/tase.2023.3242355.

[31] Y. Wang, I.-C. Lee, Y. Hong, and X. Deng. "Building degradation index with variable selection for multivariate sensory data". In: *Reliability Engineering & System Safety* 227 (2022), p. 108704. DOI: 10.1016/j.ress.2022.108704.

[32] W. Peng, Z. Wei, C.-G. Huang, G. Feng, and J. Li. "A hybrid health prognostics method for proton exchange membrane fuel cells with internal health recovery". In: *IEEE Transactions Transportation Electrification* 9.3 (2023), pp. 4406–4417. DOI: 10.1109/tte.2023.3243788.

[33] J. Zhu, N. Chen, and W. Peng. "Estimation of bearing remaining useful life based on multiscale convolutional neural network". In: *IEEE Transactions on Industrial Electronics* 66.4 (2018), pp. 3208–3216. DOI: 10.1109/tie.2018.2844856.

[34] Q. Zhai, P. Chen, L. Hong, and L. Shen. "A random-effects Wiener degradation model based on accelerated failure time". In: *Reliability Engineering & System Safety* 180 (2018), pp. 94–103. DOI: 10.1016/j.ress.2018.07.003.

[35] Z. Wang, Q. Zhai, and P. Chen. "Degradation modeling considering unit-to-unit heterogeneity-A general model and comparative study". In: *Reliability Engineering & System Safety* 216 (2021), p. 107897. DOI: 10.1016/j.ress.2021.107897.

[36] C. Luo, L. Shen, and A. Xu. "Modelling and estimation of system reliability under dynamic operating environments and lifetime ordering constraints". In: *Reliability Engineering & System Safety* 218 (2022), p. 108136. DOI: 10.1016/j.ress.2021.108136.

[37] S. Zhang, Q. Zhai, and Y. Li. "Degradation modeling and RUL prediction with Wiener process considering measurable and unobservable external impacts". In: *Reliability Engineering & System Safety* 231 (2023), p. 109021. DOI: 10.1016/j.ress.2022.109021.

[38] Q. Guan, X. Wei, H. Zhang, and L. Jia. "Remaining useful life prediction for degradation processes based on the Wiener process considering parameter dependence". In: *Quality and Reliability Engineering International* 40.3 (2024), pp. 1221–1245. DOI: 10.1002/qre.3461.

[39] A. Xu, B. Wang, D. Zhu, J. Pang, and X. Lian. "Bayesian reliability assessment of permanent magnet brake under small sample size". In: *IEEE Transactions on Reliability* (2024). DOI: 10.1109/tr.2024.3381072.

[40] Q. Zhai, Z. Ye, C. Li, M. Revie, and D. B. Dunson. "Modeling recurrent failures on large directed networks". In: *Journal of the American Statistical Association* (2024). DOI: 10.1080/01621459.2024.2319897.

[41] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin. "Machinery health prognostics: A systematic review from data acquisition to RUL prediction". In: *Mechanical Systems and Signal Processing* 104 (2018), pp. 799–834. DOI: 10.1016/j.ymssp.2017.11.016.

[42] C.-G. Huang, H.-Z. Huang, and Y.-F. Li. "A bidirectional LSTM prognostics method under multiple operational conditions". In: *IEEE Transactions Industrial Electronics* 66.11 (2019), pp. 8792–8802. DOI: 10.1109/tie.2019.2891463.

[43] Y. Liu, X. Hu, and W. Zhang. "Remaining useful life prediction based on health index similarity". In: *Reliability Engineering & System Safety* 185 (2019), pp. 502–510. DOI: 10.1016/j.ress.2019.02.002.

[44] B. Xue, H. Xu, X. Huang, K. Zhu, Z. Xu, and H. Pei. "Similarity-based prediction method for machinery remaining useful life: A review". In: *The International Journal of Advanced Manufacturing Technology* 121.3 (2022), pp. 1501–1531. DOI: 10.1007/s00170-022-09280-3.

[45] S. Jahani, R. Kontar, S. Zhou, and D. Veeramani. "Remaining useful life prediction based on degradation signals using monotonic B-splines with infinite support". In: *IISE Transactions* 52.5 (2020), pp. 537–554. DOI: 10.1080/24725854.2019.1630868.

[46] C. Luo, P. Chen, and P. Jaillet. "Portfolio optimization based on almost second-degree stochastic dominance". In: *Management Science* (2024). DOI: 10.1287/mnsc.2022.01092.

[47] L. Zhuang, A. Xu, and X.-L. Wang. "A prognostic driven predictive maintenance framework based on Bayesian deep learning". In: *Reliability Engineering & System Safety* 234 (2023), p. 109181. DOI: 10.1016/j.ress.2023.109181.

[48] C. J. Lu and W. Q. Meeker. "Using degradation measures to estimate a time-to-failure distribution". In: *Technometrics* 35.2 (1993), pp. 161–174. DOI: 10.2307/1269661.

[49] L. A. Escobar and W. Q. Meeker. "A review of accelerated test models". In: *Statistical Science* (2006), pp. 552–577. DOI: 10.1214/088342306000000321.

[50] V. Bagdonavičius, A. Bikelis, and V. Kazakevičius. "Statistical analysis of linear degradation and failure time data with multiple failure modes". In: *Lifetime Data Analysis* 10.1 (2004), pp. 65–81. DOI: 10.1023/b:lida.0000019256.59372.63.

[51] J. I. Park and S. J. Bae. "Direct prediction methods on lifetime distribution of organic light-emitting diodes from accelerated degradation tests". In: *IEEE Transactions on Reliability* 59.1 (2010), pp. 74–90. DOI: 10.1109/tr.2010.2040761.

[52] X.-S. Si and D. Zhou. "A generalized result for degradation model-based reliability estimation". In: *IEEE Transactions on Automation Science and Engineering* 11.2 (2013), pp. 632–637. DOI: 10.1109/tase.2013.2260740.

[53] Z. Wang, J. Li, X. Ma, Y. Zhang, H. Fu, and S. Krishnaswamy. "A generalized Wiener process degradation model with two transformed time scales". In: *Quality and Reliability Engineering International* 33.4 (2017), pp. 693–708. DOI: 10.1002/qre.2049.

[54] X. Wang, B. X. Wang, W. Wu, and Y. Hong. "Reliability analysis for accelerated degradation data based on the Wiener process with random effects". In: *Quality and Reliability Engineering International* 36.6 (2020), pp. 1969–1981. DOI: 10.1002/qre.2668.

[55] Q. Dong and L. Cui. "Reliability analysis of a system with two-stage degradation using Wiener processes with piecewise linear drift". In: *IMA Journal of Management Mathematics* 32.1 (2021), pp. 3–29. DOI: 10.1093/imaman/dpaa009.

[56] X. Wang. "Wiener processes with random effects for degradation data". In: *Journal of Multivariate Analysis* 101.2 (2010), pp. 340–351. DOI: 10.1016/j.jmva.2008.12.007.

[57] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht. "Degradation data analysis using Wiener processes with measurement errors". In: *IEEE Transactions on Reliability* 62.4 (2013), pp. 772–780. DOI: 10.1109/tr.2013.2284733.

[58] X. Wang, P. Jiang, B. Guo, and Z. Cheng. "Real-time reliability evaluation with a general Wiener process-based degradation model". In: *Quality and Reliability Engineering International* 30.2 (2014), pp. 205–220. DOI: 10.1002/qre.1489.

[59] J. Huang, D. S. Golubović, S. Koh, D. Yang, X. Li, X. Fan, and G. Zhang. "Degradation modeling of mid-power white-light LEDs by using Wiener process". In: *Optics Express* 23.15 (2015), A966–A978. DOI: 10.1364/oe.23.00a966.

[60] Z. Zhang, X. Si, C. Hu, and Y. Lei. "Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods". In: *European Journal of Operational Research* 271.3 (2018), pp. 775–796. DOI: 10.1016/j.ejor.2018.02.033.

[61] Q. Dong and L. Cui. "A study on stochastic degradation process models under different types of failure thresholds". In: *Reliability Engineering & System Safety* 181 (2019), pp. 202–212. DOI: 10.1016/j.ress.2018.10.002.

[62] H. Zheng and H. Xu. "Reliability analysis for degradation and shock process based on truncated normal distribution". In: *Communications in Statistics-Simulation and Computation* 51.8 (2022), pp. 4241–4256. DOI: 10.1080/03610918.2020.1740264.

[63] Q. Guan, X. Wei, W. Bai, and L. Jia. "Two-stage degradation modeling for remaining useful life prediction based on the Wiener process with measurement errors". In: *Quality and Reliability Engineering International* (2022), pp. 1–28. DOI: 10.1002/qre.3147.

[64] L. Wang, R. Pan, X. Li, and T. Jiang. "A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information". In: *Reliability Engineering & System Safety* 112 (2013), pp. 38–47. DOI: 10.1016/j.ress.2012.09.015.

[65] J. Guo, Y.-F. Li, B. Zheng, and H.-Z. Huang. "Bayesian degradation assessment of CNC machine tools considering unit non-homogeneity". In: *Journal of Mechanical Science and Technology* 32.6 (2018), pp. 2479–2485. DOI: 10.1007/s12206-018-0505-1.

[66] W. Nelson. "Analysis of accelerated life test data - Part I: The Arrhenius model and graphical methods". In: *IEEE Transactions on Dielectrics and Electrical Insulation* EI-6.4 (1971), pp. 165–181. DOI: 10.1109/tei.1971.299172.

[67] Y. Zhu, S. Liu, K. Wei, H. Zuo, R. Du, and X. Shu. "A novel based-performance degradation Wiener process model for real-time reliability evaluation of lithium-ion battery". In: *Journal of Energy Storage* 50 (2022), p. 104313. DOI: 10.1016/j.est.2022.104313.

[68] J. Guo, Y.-F. Li, W. Peng, and H.-Z. Huang. "Bayesian information fusion method for reliability analysis with failure-time data and degradation data". In: *Quality and Reliability Engineering International* 38.4 (2022), pp. 1944–1956. DOI: 10.1002/qre.3065.

[69] H. Zheng, J. Yang, H. Xu, and Y. Zhao. "Reliability acceptance sampling plan for degraded products subject to Wiener process with unit heterogeneity". In: *Reliability Engineering & System Safety* 229 (2023), p. 108877. DOI: 10.1016/j.ress.2022.108877.

[70] H. Wang, H. Liao, X. Ma, R. Bao, and Y. Zhao. "A new class of mechanism-equivalence-based Wiener process models for reliability analysis". In: *IISE Transactions* (2021), pp. 1–18. DOI: 10.1080/24725854.2021.2000075.

[71] K. Song and L. Cui. "Fiducial inference-based failure mechanism consistency analysis for accelerated life and degradation tests". In: *Applied Mathematical Modelling* 105 (2022), pp. 340–354. DOI: 10.1016/j.apm.2021.12.048.

[72] P. Chen and Z.-S. Ye. "Uncertainty quantification for monotone stochastic degradation models". In: *Journal of Quality Technology* 50.2 (2018), pp. 207–219. DOI: 10.1080/00224065.2018.1436839.

[73] J. Shao. *Mathematical statistics*. Springer Science & Business Media, 2003. DOI: 10.1007/b97553.

[74] Z. Chen, T. Xia, Y. Li, and E. Pan. "Random-effect models for degradation analysis based on nonlinear Tweedie exponential-dispersion processes". In: *IEEE Transactions on Reliability* 71.1 (2021), pp. 47–62. DOI: 10.1109/tr.2021.3107050.

[75] Y. Wen, J. Wu, D. Das, and T.-L. B. Tseng. "Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity". In: *Reliability Engineering & System Safety* 176 (2018), pp. 113–124. DOI: 10.1016/j.ress.2018.04.005.

[76] J. Li, Y. Tian, and D. Wang. "Change-point detection of failure mechanism for electronic devices based on Arrhenius model". In: *Applied Mathematical Modelling* 83 (2020), pp. 46–58. DOI: 10.1016/j.apm.2020.02.011.

[77] X. Cai, Y. Tian, and W. Ning. "Change-point analysis of the failure mechanisms based on accelerated life tests". In: *Reliability Engineering & System Safety* 188 (2019), pp. 515–522. DOI: 10.1016/j.ress.2019.04.002.

[78] H. Zheng, X. Kong, H. Xu, and J. Yang. "Reliability analysis of products based on proportional hazard model with degradation trend and environmental factor". In: *Reliability Engineering & System Safety* 216 (2021), p. 107964. DOI: 10.1016/j.ress.2021.107964.

[79] E. L. Lehmann, J. P. Romano, and G. Casella. *Testing statistical hypotheses*. Vol. 3. Springer, 2005. DOI: 10.1007/978-3-030-70578-7.

[80] B. Dunn, H. Kamath, and J.-M. Tarascon. "Electrical energy storage for the grid: A battery of choices". In: *Science* 334.6058 (2011), pp. 928–935. DOI: 10.1126/science.1212741.

[81] R. Schmuch, R. Wagner, G. Hörpel, T. Placke, and M. Winter. "Performance and cost of materials for lithium-based rechargeable automotive batteries". In: *Nature Energy* 3.4 (2018), pp. 267–278. DOI: 10.1038/s41560-018-0107-2.

[82] W. Vermeer, G. R. C. Mouli, and P. Bauer. "A comprehensive review on the characteristics and modeling of lithium-ion battery aging". In: *IEEE Transactions on Transportation Electrification* 8.2 (2021), pp. 2205–2232. DOI: 10.1109/tte.2021.3138357.

[83] M. Yang, X. Sun, R. Liu, L. Wang, F. Zhao, and X. Mei. "Predict the lifetime of lithium-ion batteries using early cycles: A review". In: *Applied Energy* 376 (2024), p. 124171. DOI: 10.1016/j.apenergy.2024.124171.

[84] X. Hu, D. Cao, and B. Egardt. "Condition monitoring in advanced battery management systems: Moving horizon estimation using a reduced electrochemical model". In: *IEEE/ASME Transactions on Mechatronics* 23.1 (2017), pp. 167–178. DOI: 10.1109/tmech.2017.2675920.

[85] D. P. Finegan and S. J. Cooper. "Battery safety: Data-driven prediction of failure". In: *Joule* 3.11 (2019), pp. 2599–2601. DOI: 10.1016/j.joule.2019.10.013.

[86] Y. Qin, S. Adams, and C. Yuen. "Transfer learning-based state of charge estimation for lithium-ion battery at varying ambient temperatures". In: *IEEE Transactions on Industrial Informatics* 17.11 (2021), pp. 7304–7315. DOI: 10.1109/tii.2021.3051048.

[87] Y. Qin, C. Yuen, X. Yin, and B. Huang. “A transferable multistage model with cycling discrepancy learning for lithium-ion battery state of health estimation”. In: *IEEE Transactions on Industrial Informatics* 19.2 (2022), pp. 1933–1946. DOI: 10.1109/tii.2022.3205942.

[88] Z. Tong, J. Miao, S. Tong, and Y. Lu. “Early prediction of remaining useful life for Lithium-ion batteries based on a hybrid machine learning method”. In: *Journal of Cleaner Production* 317 (2021), p. 128265. DOI: 10.1016/j.jclepro.2021.128265.

[89] X. Pang, W. Yang, C. Wang, H. Fan, L. Wang, J. Li, S. Zhong, W. Zheng, H. Zou, S. Chen, *et al.* “A novel hybrid model for lithium-ion batteries lifespan prediction with high accuracy and interpretability”. In: *Journal of Energy Storage* 61 (2023), p. 106728. DOI: 10.1016/j.est.2023.106728.

[90] Y. Yang. “A machine-learning prediction method of lithium-ion battery life based on charge process for different applications”. In: *Applied Energy* 292 (2021), p. 116897. DOI: 10.1016/j.apenergy.2021.116897.

[91] S. Kim, H. Jung, M. Lee, Y. Y. Choi, and J.-I. Choi. “Model-free reconstruction of capacity degradation trajectory of lithium-ion batteries using early cycle data”. In: *ETransportation* 17 (2023), p. 100243. DOI: 10.1016/j.etran.2023.100243.

[92] J. Chen, R. Huang, Z. Chen, W. Mao, and W. Li. “Transfer learning algorithms for bearing remaining useful life prediction: A comprehensive review from an industrial application perspective”. In: *Mechanical Systems and Signal Processing* 193 (2023), p. 110239. DOI: 10.1016/j.ymssp.2023.110239.

[93] L. Shen, J. Li, L. Meng, L. Zhu, and H. T. Shen. “Transfer learning-based state of charge and state of health estimation for li-ion batteries: A review”. In: *IEEE Transactions on Transportation Electrification* 10.1 (2024), pp. 1465–1481. DOI: 10.1109/tte.2023.3293551.

[94] B. Jia, Y. Guan, and L. Wu. “A state of health estimation framework for lithium-ion batteries using transfer components analysis”. In: *Energies* 12.13 (2019), p. 2524. DOI: 10.3390/en12132524.

[95] Y. Li, H. Sheng, Y. Cheng, D.-I. Stroe, and R. Teodorescu. “State-of-health estimation of lithium-ion batteries based on semi-supervised transfer component analysis”. In: *Applied Energy* 277 (2020), p. 115504. DOI: 10.1016/j.apenergy.2020.115504.

[96] Z. Jiao, H. Wang, J. Xing, Q. Yang, M. Yang, Y. Zhou, and J. Zhao. “LightGBM-based framework for lithium-ion battery remaining useful life prediction under driving conditions”. In: *IEEE Transactions on Industrial Informatics* 19.11 (2023), pp. 11353–11362. DOI: 10.1109/tii.2023.3246124.

[97] S. Su, W. Li, J. Mou, A. Garg, L. Gao, and J. Liu. “A hybrid battery equivalent circuit model, deep learning, and transfer learning for battery state monitoring”. In: *IEEE Transactions on Transportation Electrification* 9.1 (2022), pp. 1113–1127. DOI: 10.1109/tte.2022.3204843.

[98] D. Pan, H. Li, and S. Wang. “Transfer learning-based hybrid remaining useful life prediction for lithium-ion batteries under different stresses”. In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–10. DOI: 10.1109/TIM.2022.3142757.

[99] Y. Tan and G. Zhao. “Transfer learning with long short-term memory network for state-of-health prediction of lithium-ion batteries”. In: *IEEE Transactions on Industrial Electronics* 67.10 (2019), pp. 8723–8731. DOI: 10.1109/tie.2019.2946551.

[100] Z. Deng, X. Lin, J. Cai, and X. Hu. “Battery health estimation with degradation pattern recognition and transfer learning”. In: *Journal of Power Sources* 525 (2022), p. 231027. DOI: 10.1016/j.jpowsour.2022.231027.

[101] C. Dai Nguyen and S. J. Bae. “Equivalent circuit simulated deep network architecture and transfer learning for remaining useful life prediction of lithium-ion batteries”. In: *Journal of Energy Storage* 71 (2023), p. 108042. DOI: 10.1016/j.est.2023.108042.

[102] T. Han, Z. Wang, and H. Meng. “End-to-end capacity estimation of Lithium-ion batteries with an enhanced long short-term memory network considering domain adaptation”. In: *Journal of Power Sources* 520 (2022), p. 230823. DOI: 10.1016/j.jpowsour.2021.230823.

[103] Z. Ye and J. Yu. “State-of-health estimation for lithium-ion batteries using domain adversarial transfer learning”. In: *IEEE Transactions on Power Electronics* 37.3 (2021), pp. 3528–3543. DOI: 10.1109/tpel.2021.3117788.

[104] G. Ma, S. Xu, T. Yang, Z. Du, L. Zhu, H. Ding, and Y. Yuan. “A transfer learning-based method for personalized state of health estimation of lithium-ion batteries”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022). DOI: 10.1109/tnnls.2022.3176925.

[105] L. Shen, J. Li, J. Liu, L. Zhu, and H. T. Shen. “Temperature adaptive transfer network for cross-domain state-of-charge estimation of li-ion batteries”. In: *IEEE Transactions on Power Electronics* 38.3 (2022), pp. 3857–3869. DOI: 10.1109/tpel.2022.3220760.

[106] R. Zhu, W. Peng, D. Wang, and C.-G. Huang. “Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data”. In: *Mechanical Systems and Signal Processing* 183 (2023), p. 109628. DOI: 10.1016/j.ymssp.2022.109628.

[107] A. A. Chehade and A. A. Hussein. "A collaborative Gaussian process regression model for transfer learning of capacity trends between li-ion battery cells". In: *IEEE Transactions on Vehicular Technology* 69.9 (2020), pp. 9542–9552. DOI: 10.1109/tvt.2020.3000970.

[108] I. Oyewole, A. Chehade, and Y. Kim. "A controllable deep transfer learning network with multiple domain adaptation for battery state-of-charge estimation". In: *Applied Energy* 312 (2022), p. 118726. DOI: 10.1016/j.apenergy.2022.118726.

[109] Y. Ding, M. Jia, Q. Miao, and P. Huang. "Remaining useful life estimation using deep metric transfer learning for kernel regression". In: *Reliability Engineering & System Safety* 212 (2021), p. 107583. DOI: 10.1016/j.ress.2021.107583.

[110] L. van der Maaten and G. Hinton. "Visualizing data using t-SNE". In: *Journal of Machine Learning Research* 9.86 (2008), pp. 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.

[111] T. Ando and K.-C. Li. "A model-averaging approach for high-dimensional regression". In: *Journal of the American Statistical Association* 109.505 (2014), pp. 254–265. DOI: 10.1080/01621459.2013.838168.

[112] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. "A kernel two-sample test". In: *Journal of Machine Learning Research* 13.1 (2012), pp. 723–773. URL: http://jmlr.org/papers/v13/gretton12a.html.

[113] C.-m. Lin, S.-j. Liu, Y.-s. Wen, J.-h. Liu, G.-s. He, X. Zhao, Z.-j. Yang, L. Ding, L.-p. Pan, J. Li, *et al.* "Sandwich-like interfacial structured polydopamine (PDA)/Wax/PDA: A novel design for simultaneously improving the safety and mechanical properties of highly explosive-filled polymer composites". In: *Energetic Materials Frontiers* 3.4 (2022), pp. 189–198. DOI: 10.1016/j.enmf.2022.03.003.

[114] H. Sinapour, S. Damiri, and H. Pouretedal. "The study of RDX impurity and wax effects on the thermal decomposition kinetics of HMX explosive using DSC/TG and accelerated aging methods". In: *Journal of Thermal Analysis and Calorimetry* 129.1 (2017), pp. 271–279. DOI: 10.1007/s10973-017-6118-6.

[115] D.-X. Wang, S.-S. Chen, S.-H. Jin, Q.-H. Shu, Z.-M. Jiang, F.-Q. Shang, and J.-X. Li. "Investigation into the coating and desensitization effect on HNIW of paraffin wax/stearic acid composite system". In: *Journal of Energetic Materials* 34.1 (2016), pp. 26–37. DOI: 10.1080/07370652.2014.993049.

[116] Ø. Hammer Johansen, J. Digre Kristiansen, R. Gjersøe, A. Berg, T. Halvorsen, K.-T. Smith, and G. O. Nevstad. "RDX and HMX with reduced sensitivity towards shock initiation–RS-RDX and RS-HMX". In: *Propellants, Explosives, Pyrotechnics: An International Journal Dealing with Scientific and Technological Aspects of Energetic Materials* 33.1 (2008), pp. 20–24. DOI: 10.1002/prep.200800203.

[117] C. Zhang, X. Cao, and B. Xiang. "Understanding the desensitizing mechanism of olefin in explosives: Shear slide of mixed HMX-olefin systems". In: *Journal of Molecular Modeling* 18 (2012), pp. 1503–1512. DOI: 10.1007/s00894-011-1174-5.

[118] L. Yang. "Surface polarity of β-HMX crystal and the related adhesive forces with estane binder". In: *Langmuir* 24.23 (2008), pp. 13477–13482. DOI: 10.1021/la802494b.

[119] C. Zhang. "Understanding the desensitizing mechanism of olefin in explosives versus external mechanical stimuli". In: *The Journal of Physical Chemistry C* 114.11 (2010), pp. 5068–5072. DOI: 10.1021/jp910883x.

[120] L. Wardhaugh and D. Boger. "The measurement and description of the yielding behavior of waxy crude oil". In: *Journal of Rheology* 35.6 (1991), pp. 1121–1156. DOI: 10.1122/1.550168.

[121] J. Zhang, B. Yu, H. Li, and Q. Huang. "Advances in rheology and flow assurance studies of waxy crude". In: *Petroleum Science* 10 (2013), pp. 538–547. DOI: 10.1007/s12182-013-0305-2.

[122] L. F. Dalla, E. J. Soares, and R. N. Siqueira. "Start-up of waxy crude oils in pipelines". In: *Journal of Non-Newtonian Fluid Mechanics* 263 (2019), pp. 61–68. DOI: 10.1016/j.jnnfm.2018.11.008.

[123] A. Fakroun and H. Benkreira. "Rheology of waxy crude oils in relation to restart of gelled pipelines". In: *Chemical Engineering Science* 211 (2020), p. 115212. DOI: 10.1016/j.ces.2019.115212.

[124] G. Sun, J. Zhang, and H. Li. "Structural behaviors of waxy crude oil emulsion gels". In: *Energy & Fuels* 28.6 (2014), pp. 3718–3729. DOI: 10.1021/ef500534r.

[125] L. Hou and J. Zhang. "A study on creep behavior of gelled Daqing crude oil". In: *Petroleum Science and Technology* 28.7 (2010), pp. 690–699. DOI: 10.1080/10916460902804648.

[126] H.-Y. Li, Q.-B. Xie, H. Sun, W. Guo, F. Yan, Q. Miao, C.-F. Nie, Y. Zhuang, Q. Huang, and J.-J. Zhang. "A nonlinear creep damage model for gelled waxy crude". In: *Petroleum Science* 19.4 (2022), pp. 1803–1811. DOI: 10.1016/j.petsci.2021.12.027.

[127] X. Zhao, P. Chen, S. Lv, and Z. He. "Reliability testing for product return prediction". In: *European Journal of Operational Research* 304.3 (2023), pp. 1349–1363. DOI: 10.1016/j.ejor.2022.05.012.

[128] C. Wang, J. Liu, Q. Yang, Q. Hu, and D. Yu. "Recoverability effects on reliability assessment for accelerated degradation testing". In: *IISE Transactions* 55.7 (2023), pp. 698–710. DOI: 10.1080/24725854.2022.2089784.

[129] H. Zheng, J. Yang, and Y. Zhao. "Reliability analysis of multi-stage degradation with stage-varying noises based on the nonlinear Wiener process". In: *Applied Mathematical Modelling* 125 (2024), pp. 445–467. DOI: 10.1016/j.apm.2023.09.007.

[130] Q. Sun, P. Chen, X. Wang, and Z.-S. Ye. "Robust condition-based production and maintenance planning for degradation management". In: *Production and Operations Management* 32.12 (2023), pp. 3951–3967. DOI: 10.1111/poms.14071.

[131] S. Hao, J. Yang, and C. Berenguer. "Degradation analysis based on an extended inverse Gaussian process model with skew-normal random effects and measurement errors". In: *Reliability Engineering & System Safety* 189 (2019), pp. 261–270. DOI: 10.1016/j.ress.2019.04.031.

[132] P. Chen, Z.-S. Ye, and X. Xiao. "Pairwise model discrimination with applications in lifetime distributions and degradation processes". In: *Naval Research Logistics (NRL)* 66.8 (2019), pp. 675–686. DOI: 10.1002/nav.21875.

[133] X. Zhao, P. Chen, O. Gaudoin, and L. Doyen. "Accelerated degradation tests with inspection effects". In: *European Journal of Operational Research* 292.3 (2021), pp. 1099–1114. DOI: 10.1016/j.ejor.2020.11.041.

[134] Q. Zhai, A. Xu, J. Yang, and Y. Zhou. "Statistical modeling and reliability analysis for degradation processes indexed by two scales". In: *IEEE Transactions on Industrial Informatics* (2023). DOI: 10.1109/tii.2023.3313668.

[135] S. Chen, Z. Lu, Q. Hu, M. Xie, and D. Yu. "Bayesian Analysis of Lifetime Delayed Degradation Process for Destructive/Nondestructive Inspection". In: *IEEE Transactions on Reliability* 73.2 (2024), pp. 990–1004. DOI: 10.1109/tr.2023.3330288.

[136] Q. Liu, L. Jin, H. K. T. Ng, Q. Hu, and D. Yu. "Multivariate t Degradation Processes for Dependent Multivariate Degradation Data". In: *IEEE Transactions on Reliability* (2024). DOI: 10.1109/tr.2024.3398652.

[137] L. Lu, B. Wang, Y. Hong, and Z. Ye. "General path models for degradation data with multiple characteristics and covariates". In: *Technometrics* 63.3 (2021), pp. 354–369. DOI: 10.1080/00401706.2020.1796814.

[138] G. A. Veloso and R. H. Loschi. "Dynamic linear degradation model: Dealing with heterogeneity in degradation paths". In: *Reliability Engineering & System Safety* 210 (2021), p. 107446. DOI: 10.1016/j.ress.2021.107446.

[139] Z.-S. Ye and M. Xie. "Stochastic modelling and analysis of degradation for highly reliable products". In: *Applied Stochastic Models in Business and Industry* 31.1 (2015), pp. 16–32. DOI: 10.1002/asmb.2063.

[140] Z.-S. Ye, L.-P. Chen, L. C. Tang, and M. Xie. "Accelerated degradation test planning using the inverse Gaussian process". In: *IEEE Transactions on Reliability* 63.3 (2014), pp. 750–763. DOI: 10.1109/tr.2014.2315773.

[141] G. Fang, R. Pan, and J. Stufken. "Optimal setting of test conditions and allocation of test units for accelerated degradation tests with two stress variables". In: *IEEE Transactions on Reliability* 70.3 (2020), pp. 1096–1111. DOI: 10.1109/tr.2020.2995333.

[142] J. O. Ramsay. "When the data are functions". In: *Psychometrika* 47 (1982), pp. 379–396. DOI: 10.1007/bf02293704.

[143] J. O. Ramsay and B. W. Silverman. *Fitting differential equations to functional data: Principal differential analysis*. Springer, 2005. DOI: 10.1007/0-387-22751-2_19.

[144] N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, K. Cohen, G. Boente, R. Fraiman, B. Brumback, C. Croux, *et al.* "Robust principal component analysis for functional data". In: *Test* 8 (1999), pp. 1–73. DOI: 10.1007/bf02595862.

[145] J. O. Ramsay and C. Dalzell. "Some tools for functional data analysis". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 53.3 (1991), pp. 539–561. DOI: 10.1111/j.2517-6161.1991.tb01844.x.

[146] P. Kokoszka, H. Oja, B. Park, and L. Sangalli. "Special issue on functional data analysis". In: *Econometrics and Statistics* 1.C (2017), pp. 99–100. DOI: 10.1016/j.ecosta.2016.11.003.

[147] Y. Xu, Y. Qiu, and J. C. Schnable. "Functional modeling of plant growth dynamics". In: *The Plant Phenome Journal* 1.1 (2018), pp. 1–10. DOI: 10.2135/tppj2017.09.0007.

[148] S. Ullah and C. F. Finch. "Functional data modelling approach for analysing and predicting trends in incidence rates-an application to falls injury". In: *Osteoporosis International* 21 (2010), pp. 2125–2134. DOI: 10.1007/s00198-010-1189-2.

[149] C. De Boor. *A practical guide to splines*. Vol. 27. springer-verlag New York, 1978. DOI: 10.1007/978-1-4612-6333-3.

[150] R. L. Eubank. *Nonparametric regression and spline smoothing*. CRC Press, 1999. DOI: 10.1201/9781482273144.

[151] J. Fan. *Local polynomial modelling and its applications: Monographs on statistics and applied probability 66*. Routledge, 2018. DOI: 10.1201/9780203748725.

[152] S. Ullah and C. F. Finch. "Applications of functional data analysis: A systematic review". In: *BMC Medical Research Methodology* 13 (2013), pp. 1–12. DOI: 10.1186/1471-2288-13-43.

[153] C. Happ and S. Greven. "Multivariate functional principal component analysis for data observed on different (dimensional) domains". In: *Journal of the American Statistical Association* 113.522 (2018), pp. 649–659. DOI: 10.1080/01621459.2016.1273115.

[154] Y. Nie and J. Cao. "Sparse functional principal component analysis in a new regression framework". In: *Computational Statistics & Data Analysis* 152 (2020), p. 107016. DOI: 10.1016/j.csda.2020.107016.

[155] J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer, 2005. DOI: 10.1007/b98888.

[156] L. Schumaker. *Spline functions: Basic theory*. Cambridge University Press, 2007. DOI: 10.1017/cbo9780511618994.

[157] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah. "Time-series clustering–a decade review". In: *Information Systems* 53 (2015), pp. 16–38. DOI: 10.1016/j.is.2015.04.007.

[158] Y. Zheng. "Trajectory data mining: An overview". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 6.3 (2015), pp. 1–41. DOI: 10.1145/2743025.

[159] F. Chamroukhi and H. D. Nguyen. "Model-based clustering and classification of functional data". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9.4 (2019), e1298. DOI: 10.1002/widm.1298.

[160] A. S. Cheam and M. Fredette. "On the importance of similarity characteristics of curve clustering and its applications". In: *Pattern Recognition Letters* 135 (2020), pp. 360–367. DOI: 10.1016/j.patrec.2020.04.024.

[161] M. Zhang and A. Parnell. "Review of clustering methods for functional data". In: *ACM Transactions on Knowledge Discovery from Data* 17.7 (2023), pp. 1–34. DOI: 10.1145/3581789.

[162] B. Charles and J. Julien. "Model-based clustering of time series in group-specific functional subspaces". In: *Advances in Data Analysis and Classification* 5 (2011), pp. 281–300. DOI: 10.1007/s11634-011-0095-6.

[163] L. Xue and J. Wang. "Distribution function estimation by constrained polynomial spline regression". In: *Journal of Nonparametric Statistics* 22.4 (2010), pp. 443–457. DOI: 10.1080/10485250903336802.

[164] X. Shen, D. Wolfe, and S. Zhou. "Local asymptotics for regression splines and confidence regions". In: *The Annals of Statistics* 26.5 (1998), pp. 1760–1782. DOI: 10.1214/aos/1024691356.

[165] J. Z. Huang. "Local asymptotics for polynomial spline regression". In: *The Annals of Statistics* 31.5 (2003), pp. 1600–1635. DOI: 10.1214/aos/1065705120.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who has supported and guided me throughout my PhD journey. This dissertation is the result of collaborative efforts and unwavering support from many individuals.

First and foremost, I am profoundly grateful to my supervisors at Delft University of Technology (TU Delft) and Beijing Institute of Technology (BIT) for their guidance and encouragement.

At TU Delft, I extend my heartfelt thanks to my promotor, Prof. dr. ir. Geurt Jongbloed, for his insightful advice and continuous support. The "Goedemorgen allemaal" he often sang in the morning always energized me during my early mornings at TU Delft. My sincere appreciation also goes to Dr. Piao Chen, my daily supervisor at TU Delft, whose invaluable feedback and mentorship were crucial for my research progress. He was always there when I needed guidance, providing important advice and suggestions in both life and scientific research.

At BIT, I am deeply indebted to my promotor, Prof. dr. Yubin Tian, for her unwavering support and guidance. I also wish to thank Dr. Dianpeng Wang, my daily supervisor at BIT, for his constant encouragement and practical advice. They introduced me to the path of scientific research and provided me with comprehensive training in thinking and writing, which greatly benefited me. Thanks also to Prof. dr. Houbao Xu at BIT, who helped me a lot when I was lost at the beginning of my PhD journey.

A warm thanks to Ardjen, Chenxi, Dan, Dominique, Elli, Femke, Huiling, Jeffrey, Joffrey, Laura, Marc, Máté, Naqi, Roniet, Shirong, Shixuan, Xiwei, Yanyan, and all other dear colleagues at Delft Institute of Applied Mathematics, TU Delft, and the School of Mathematics and Statistics, BIT. I am deeply thankful for all your kind help and the happy times we shared together. Special thanks to my dear friend Shixiang, with whom I have shared numerous memorable experiences throughout this journey. Your friendship has been a source of strength and motivation.

Finally, I would like to express my deepest gratitude to my family. Their unconditional love and sacrifices have been the foundation upon which I have built my academic career. Their support has made all of this possible.

CURRICULUM VITAE

Wenda KANG

12-09-1996	Born in Bozhou, Anhui province, China.

EDUCATION

2014–2018	Bachelor of Science in Mathematics and Applied Mathematics School of Mathematical Science Ocean University of China Qingdao, China
2018-2024	PhD. & Statistics School of Mathematics and Statistics Beijing Institute of Technology (2018-2021, 2023-2024) Beijing, China Delft Institute of Applied Mathematics Delft Institute of Technology (2021-2023) Delft, Netherlands
Thesis:	Reliability analysis for industrial devices based on data fusion
Promotor:	Prof. dr. ir. G. Jongbloed Prof. dr. Y.B. Tian Dr. P. Chen

AWARDS

2022	Best paper award at the 4th International Conference on System Reliability and Safety Engineering (SRSE), December 15-18, 2022.

List of Publications

1 **W. Kang**, D. Wang, G. Jongbloed, J. Hu, and P. Chen. "Robust transfer learning for battery lifetime prediction". *IEEE Transactions on Industrial Informatics* (2025). DOI: 10.1109/tii.2025.3545079.

2 **W. Kang**, S. Li, Y. Tian, Y. Yin, H. Sui, and D. Wang. "A functional data-driven method for modelling degradation of waxy lubrication layer". In: *Quality and Reliability Engineering International* 40.7 (2024), pp. 3751–3773. DOI: 10.1002/qre.3622.

3 **W. Kang**, G. Jongbloed, Y. Tian, and P. Chen. "Degradation index-based prediction for remaining useful life using multivariate sensor data". In: *Quality and Reliability Engineering International* 40.7 (2024), pp. 3709–3728. DOI: 10.1002/qre.3615.

4 **W. Kang**, Y. Tian, H. Xu, D. Wang, H. Zheng, M. Zhang, and H. Mu. "Reliability analysis based on the Wiener process integrated with historical degradation data". In: *Quality and Reliability Engineering International* 39.4 (2023), pp. 1376–1395. DOI: 10.1002/qre.3300.

5 H. Zheng, J. Yang, **W. Kang**, and Y. Zhao. "Accelerated degradation data analysis based on inverse Gaussian process with unit heterogeneity". In: *Applied Mathematical Modelling* 126 (2024), pp. 420–438. DOI: 10.1016/j.apm.2023.11.003.