2018

Prognostic methodologies for repeated measurement data of multiple units

Jian Guo
Western New England University

Prognostic Methodologies for Repeated Measurement Data of Multiple Units

by

Jian Guo

A dissertation submitted to the Faculty of
Western New England University
in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in Engineering Management

Springfield, MA

May 20, 2018

Approved by

Dr. Zhaojun Li, Chair, Assistant Professor	Date
Dr. Thomas Keyser, Co-chair, Professor	Date
Dr. Seyed Ahmad Niknam, Committee Member, Assistant Professor	Date
Dr. Ruolin Zhou, Committee Member, Assistant Professor	Date
Dr. Thomas Keyser, EMGT Program Director	Date
Dr. S. Hossein Cheraghi, Dean, College of Engineering	Date

Abstract

This research investigates prognostics modeling methods on repeated measurement data of multiple units in order to improve the prediction accuracy and facilitate effective predictive maintenance. Engineering systems deteriorate in performance over time and are subject to the stresses in operation. Maintenance is carried out to assure satisfactory levels of reliability during the life of systems. Predictive maintenance (PdM) is one of the most effective maintenance policies, where maintenance actions are planned based on the actual system performance. Remaining useful life (RUL) prediction is the keystone of PdM. Prognostics and health management (PHM) is involved in PdM due to its strength in the system's RUL prediction and related health management. Prognostics is the core process of PdM and PHM, that aims to predict the RUL based on the available performance data. Multiple uncertainties, such as input uncertainties and model uncertainties, undermine the prediction performance of prognostics models. To characterize the inherent variability in the degradation process, repeated measurement design is exercised, where repeated measurement data of multiple units is obtained. This type of data is used to detect multiple-source variability and requires advanced modeling techniques due to its complex structures. This research aims to develop adequate prognostics models to quantify multiple-source variability in this type of data and develop robust algorithms for complex data structure with unbalanced/missing data.

Based on the way of modeling the multiple-source variability, four prediction methods are proposed to model the prognostics process based on repeated measurement data of multiple units. General mixed-effect models (GMM), containing fixed and random effects, are widely used to account multiple sources of variability in repeated measures. The combination of fixed and random effects illustrates the variability in a stochastic process. Fixed effects describe the characteristics of the population average over units and the random effects demonstrate the variation of units. Because of the difficulty of GMM dealing with

unbalanced data, a joint modeling method (JMM) is proposed where the degradation processes of each unit is interpreted as multivariate normal distributions. The concept of joint modeling is that the mean and covariance are decomposed firstly and then unknown parameters of the mean function and covariance matrix are estimated jointly. In the proposed method, mean, variance, and correlation of measurements are firstly decomposed based on Cholesky decomposition. Trigonometric functions are used to parameterize the correlation matrix. A penalized maximum likelihood estimation is proposed for parameter estimation in JMM. The expensive computation in GMM and JMM due to the high dimension covariance matrix necessitates the dimension reduction techniques. For this purpose, functional principal component analysis (FPCA) is deployed in this research. Functional data refers to data where each observation is modeled as a curve, a surface, or a hypersurface. FPCA applies the concept of functional data analysis in principal component analysis to reduce computation complexity. Finally a general spatio-temporal model is proposed based on the aforementioned methods, where spatial, temporal trends and their dependency will be quantified. Spatial trends can be analogized as the difference between units, while temporal trends illustrate the degradation process.

To reduce the model error, physical understanding is incorporated into the models. Covariate selection for all the proposed methods is done based on physics-based model. With the degradation model, the distribution of time to failure (TTF) can be estimated through simple numerical simulations. This research aims to apply and validate the proposed methods in battery capacity degradation to provide accurate prediction on cycle to failure and elucidate the mechanism of capacity fade.

Keywords: prognostics, degradation modeling, remaining useful life, time to failure, prediction, repeated measurement data, multiple units, multiple-source variability

Acknowledgments

I want to thank my committee members for serving on my dissertation committee, sharing their expertise with me during my studies and for their valuable suggestions. I am so honored and grateful to have worked with my advisor Dr. Zhaojun Li during my PhD. I could not have imagined having a better advisor and mentor for my PhD study. His vision, knowledge, and encouragement always inspired me and helped me stay focused and pursue my research. Dr. Thomas Keyser is also highly deserving of my gratitude for his extensive support during my PhD study and class teaching. I would like to thank Dr. Ruolin Zhou for her insightful comments and guidance. Her encouragement widens my research from various perspectives. I am also hugely appreciative of Dr. Seyed Ahmad Niknam, especially for sharing his knowledge and expertise so willingly, and for being so dedicated to his role as my committee member.

I would like to thank faculty at Western New England University whom I have worked with over the last four years for showing me what it means to be a dedicated and the substantial influence that their courses have had on my research. Moreover, I gratefully thank Dr. Julie Dryzmalski for her huge support at the beginning of my PhD.

Finally, but by no means least, thanks go to my lovely wife Dr. Wanying Shi and my family for their extreme supports. They are the most important people in my world, and I dedicate this thesis to them.

Table of Contents

Page

Page

List of Tables

List of Figures

List of Abbreviations

PdM	Predictive maintenance
RUL	Remaining useful life
TTF	Time to failure
PHM	Prognostics and health management
PoF	Physics of failure
GMM	General mixed model
JMM	Joint modeling method
FPCA	Functional principle component analysis
ANOVA	Analysis of variance
BIC	Bayesian information criteria

1. INTRODUCTION

This chapter presents the statement of problem with motivation, background of the overall area, research goals, and proposed research tasks.

1.1 Statement of problem

Engineering systems deteriorate in performance over time and are subject to the stresses in operation. Therefore, maintenance should be implemented to assure the satisfactory level of reliability during the life of systems. Predictive maintenance (PdM) can minimize life cycle cost and effectively avoid catastrophic failures. Prognostics and health management (PHM) is involved in PdM and is highly desired due to its strength in performance prediction and health management. For both PdM and PHM, prognostics is the key process, which implements the remaining useful life (RUL) prediction. Uncertainties in prognostics, such as input uncertainties and model uncertainties affect the prediction performance of prognostics models. Repeated measurement studies aim to characterize the input uncertainties, i.e., statistical variability, and also bring challenges for prognostics modelers. Firstly, for repeated measurements of multiple units, multiple sources of variability exists. Secondly, complex repeated measurement data requires more robust algorithms. Thirdly, expensive computation is involved due to large volumes variety of repeated measurement data. This research aims to develop adequate prognostics methodologies to tackle these challenges.

1.1.1 Prognostics and Health Management

Predictive maintenance (PdM) has been popular in industries due to its advantages in minimizing life cycle cost. It estimates the future degradation/damage state and then take maintenance actions before a failure occurs. Figure 1.1 illustrates the comparison of the

traditional maintenance and predictive maintenance. Predictive maintenance is an extended condition-based maintenance policy with a predictive capability in the degradation process [1]. Performance data monitoring is continuously required in preventive maintenance so that dynamic maintenance can be scheduled. The advantages of predictive maintenance can be seen through the increase of system/component availability, extension of operational life, the reduced process downtime, and the improved the personnel and environment safety. The implementation of predictive maintenance includes the deployment of hardware and software that monitor, detect, and predict system performance and the major interest is to estimate the remaining useful life of degrading equipment.

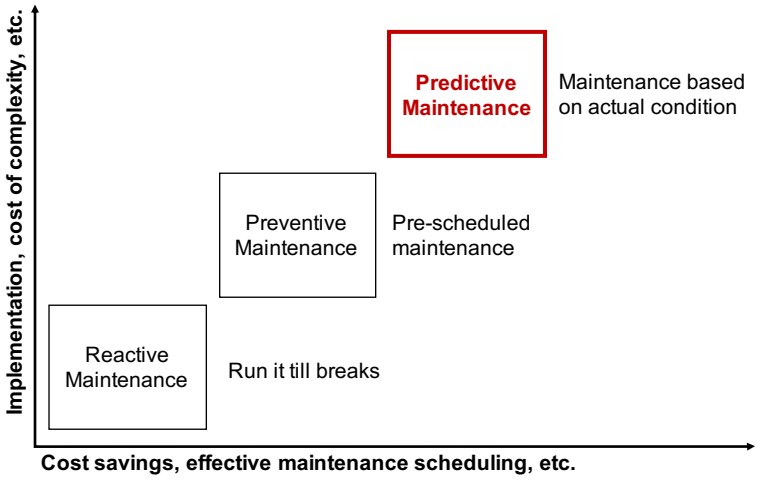

Figure 1.1: Comparison of traditional maintenance policies and predictive maintenance

The concept of Prognostics and Health Management (PHM) was first mentioned by Department of Defense in the Joint Strike Fighter Program [2] and has been highly desired in predictive maintenance since then. It is considered as the key factor to definitely promote a qualitative jump toward intelligent maintenance [3]. PHM is an emerging discipline which integrates the failure mechanism analysis and life cycle management [4]. Its major objectives are real-time health assessment, future degradation prediction, and decision management. PHM involves various fields including sensing technologies, physics of failure (PoF), data science, reliability engineering, and management science. Figure 1.2

presents the architecture of PHM cycle [5]. The implementation can be outlined as five steps: data acquisition, data processing, feature extraction, diagnostics, prognostics, and decision management. These steps are linked as a continuous improvement cycle. The first step is data acquisition, where the performance measurement data is collected from sensors. The second step is data processing, where the data cleaning, noise filtering, and feature extraction are done. The third step is prognostics where the future degradation is predicted based on the available data. This step is the core process of the whole PHM cycle. Techniques in data science and knowledge of failure mechanisms are incorporated to predict future state. Diagnosis is considered as the paralleled third step, where the fault detection and identification are implemented. Physical understandings of the degradation process are attained. The last step is decision management where the optimal maintenance scheduling and logistics support are determined. PHM are separated into prognostics and

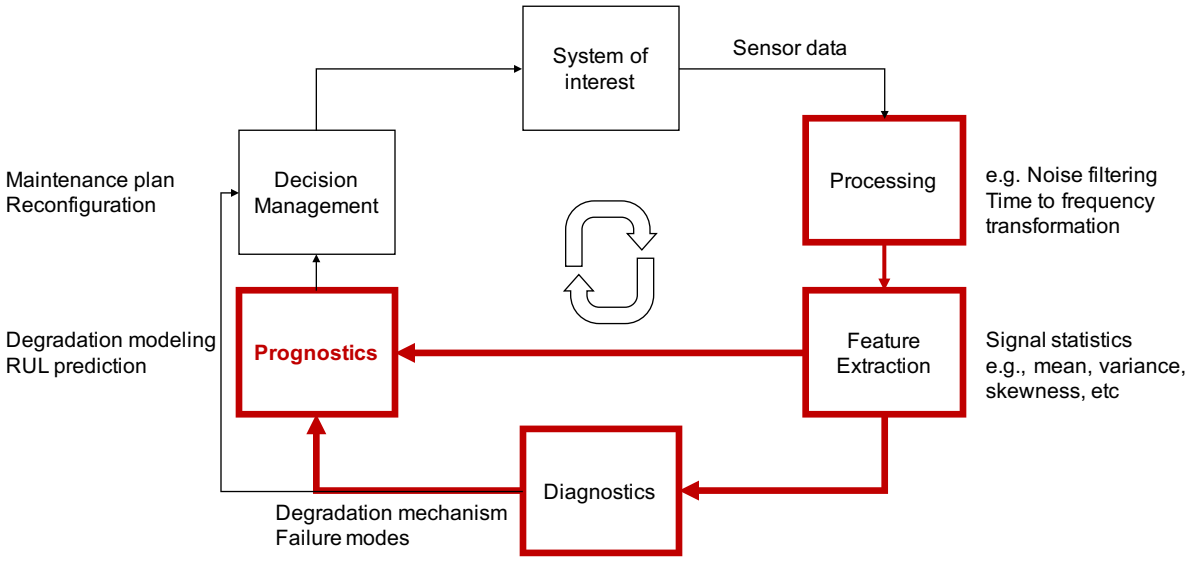

Figure 1.2: Schematic of PHM

health management [6, 7]. Prognostics is considered as the key process that provides current and predicts the future performance for the decisions related to maintenance planning and life management. This research aims to develop adequate prognostics methods using

observed repeated measurements and physical understandings of the degradation process, which is emphasized in Figure 1.2.

1.1.2 Prognostics

The term "prognostic" is defined in Merriam-Webster dictionary as something that foretells. In prognostics and health management, prognostics is a process to predict future degradation and the remaining useful life (RUL) of the system based on the available degradation data. There are several definitions of interest for prognostics.

- Prognostics is the process of estimating the remaining life of the component [8].
- Prognostics is an estimation of time to failure (TTF) and risk for one or more existing and future failure mode [9].
- Prognostics is the process of predicting the future reliability of a product of assessing the extent of deviation or degradation of the product from its expected norm operating conditions. It is the prediction of the future state of health based on current and historical health conditions [10].
- Prognostics is to predict future damage/degradation and the remaining useful life of in-service systems based on the measurement damage data [11].
- Prognostics addresses the use of automated methods to detect, diagnose, and analyze the degradation of physical system performance, calculating the remaining life in acceptable operating state before failure or unacceptable degradation of performance occurs [12].

Across all these definitions, performance/measurement data and RUL prediction are the most frequently mentioned terms. Based on the above information, prognostics can be defined as the process to model the degradation process and predict RUL based on the available data. Figure 1.3 presents the schematic of prognostics, where prognostics approaches are expected to provide accurate RUL prediction, to be robust, and comprehensive.

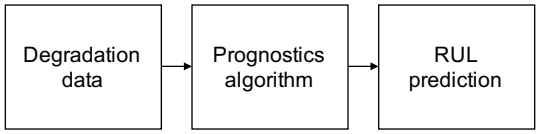

Figure 1.3: A generic procedure of prognostics

1.2 Motivation

Multiple sources of uncertainties, such as input uncertainties and model uncertainties, exist in prognostics. These uncertainties deteriorate the accuracy of RUL prediction. Given the availability of the complex repeated measurement data of multiple units, these uncertainties incur more challenges in prognostics. This research focuses on developing robust prognostics methods to tackle these challenges associated with these uncertainties.

1.2.1 Uncertainties in prognostics

A degradation process involves multiple sources of uncertainties which can all cause the inaccurate prediction of RUL. From the perspective of statistical uncertainty, the degradation process varies from unit to unit that are manufactured under the same process and condition. In a certain operation environment, there would be different degradation trajectories among units. For example, lithium ion batteries tested under the same profile and environment would form different capacity degradation processes. Meanwhile, uncertainties from the operating condition and environmental condition, such as future loads and circumstance changes, exist in the degradation modeling. Modern prognostics approaches aim to effectively quantify these aforementioned uncertainties so that an accurate RUL prediction can be obtained. Advanced techniques such as machine learning algorithms and artificial network, are used to model the degradation process. However, model errors still exist, which might be caused by misspecified models, missing failure modes, and unmodeled phenomena. It can be considered as biased understandings of the degradation process of interest. Comprehensive prognostics models intend to incorporate physical understandings of the degradation process so that model errors can be reduced. For example, the incorporation of

gas-path models in the degradation modeling of turbine engines has been investigated [13]. However, physical understandings of the degradation mechanism are usually limited and incomplete for complex systems. Therefore, uncertainties from model errors can be rarely eliminated even though with improved methods and investigations. Recently, uncertainties caused by measurement devices attract attentions since the measurement data is taken as the input of prognostics algorithm. Biased measurement/performance data might lead to total invalid prognostics. As a result, sensor failure, noises, sensor architecture, and signal fusion have been frequently investigated.

These uncertainties can be classified into four categories [5, 14, 15].

- Input uncertainties are related to inherent variability for any process, such as initial state estimation, material property, geometric characteristics, manufacturing variability, etc. These types of uncertainties are "born" with the degradation process and can not be eliminated. Design of multiple experiment runs, such as repeated measurement design, is the most commonly used method to characterize these uncertainties.

- Model uncertainties are related to model errors, such as misspecified methods, unexplained features, unmodeled phenomenons, etc. Advanced methods including data science and testing techniques have been developed to reduce these types of uncertainties. With the development of advanced sensing technique, large volumes of measurement data are available. The "big data" challenge might cause more model uncertainties.

- Operational uncertainties are related to during operation and involving environmental conditions. These types of uncertainties are similar with model uncertainties, both of which stem from the limited understanding of the degradation process. Deeper investigation, especially under various operating and environmental conditions, can reduce these uncertainties.

- Measurement uncertainties are related to uncertainties in measurement data, such as sensor noise, filter error, etc. Improved sensing techniques and advanced methods in data collection and processing can reduced this type of uncertainties.

Adequate prognostics methods are expected to handle all of these uncertainties. Specifically, prognostics models should have 1) capability in representing variety and veracity in data, 2) comprehensive physical basis so that model uncertainties can be highly reduced, and 3) robustness and computational effectiveness. This research is dedicated to developing comprehensive and robust prognostics methods that can sufficiently support decision management in PHM.

1.2.2 Repeated measurement data of multiple units

Repeated measurement data of multiple units is the type of data collected in a process where multiple responses of each experimental unit in multiple conditions and occasions and multiple units are observed [16, 17]. This type of data was firstly investigated in the medical and health care field. Repeated measurements studies are motivated by the desire of quantifying statistical uncertainties in a process. Repeated measurement data of multiple units can be used to detect within-unit and between unit over the measurement factor. The advantages of the repeated measurements studies over cross-sectional studies are displayed through 1) higher statistical power, 2) fewer required subjects, 3) various experimental conditions being tested, 4) variability within and between units being separated, and 5) the possibility of investigating individual change [18]. With the development of sensing and computation techniques, repeated measurement data of multiple units are also available in PHM. For example, NASA published several data repositories over the degradation experiments of algae raceway, milling, bearing, and lithium ion battery, where all degradation data are shown in the form of repeated measurements of multiple units [19–22]. By running multiple experiments, repeated measurement data can provide great insights of input uncertainties mentioned in the last section. In the meantime, more challenges for prognostics are also emerging and advanced modeling techniques are required. Figure 1.4 shows the prognostics for repeated measurement data of multiple units. Challenges in prognostics modeling with repeated measurement data of multiple units can be summarized as follows.

- **HOW** to quantify multiple sources of variability. Between-and-within unit variability describes the variability of a particular value for individual unit and that across units. Between-units variability stems from the difference over units, and within-unit variability is from the uncertainties of individual trajectory.

- **HOW** to deal with complex input data structure. The uncontrollable conditions, such as failed sensor and external hazard, result in unbalanced or missing data.

- **HOW** to solve computational issue in the high dimension repeated measurement data. Repeated measurement data often involves high volume and verity. Reducing dimension becomes one of important parts of prognostics modeling with this type of data.

- **HOW** to incorporate physical understanding. Multiple degradation trajectories complicates the incorporation of the physical understanding. For example, space state function in Kalman filter can not illustrate the dynamic degradation process of multiple units.

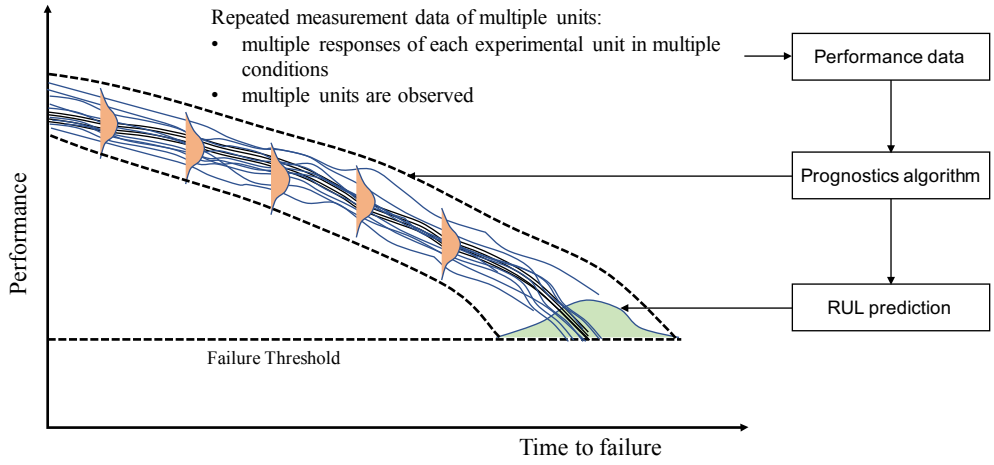

Figure 1.4: Repeated measurement data in prognostics

To tackle these above challenges, adequate models are proposed in this research to obtain the accurate prediction on remaining useful life and explicate the degradation mechanism of

interest. This research aims to develop adequate and robust methodologies that can tackle the challenges: 1) multiple sources of variability, 2) complex input data structure, 3) computational issue, and 4) difficulties to incorporate physical understandings into degradation modeling. Three research objectives are proposed:

1. to quantify multiple sources of variability,
2. to devise prognostics model for complex data structure, e.g., unbalanced data sets with missing data,
3. to explore general spatio-temporal model for repeated measurement data.

1.2.3 Significance of this research

This research aims to investigate prognostics modeling using repeated measurement data of multiple units. By achieving the three aforementioned research objectives, this research can 1) benefit the interpretation and quantification of inherent uncertainties of the degradation process and aid design improvement through examining and comparing individual across the time; 2) enhance the adaptability of prognostics models in complex data structures especially unbalanced data and improve the understanding of variations and correlation of degradation process via modeling both the mean and covariance as functions of time-varying factors; 3) extend the repeated measurement design analysis to complex system prognostics in reach of the spatio-temporal model.

1.3 Organization of the thesis

The remainder of this dissertation is organized as follows. Chapter 2 presents the literature review and research gaps. More specifically, the existing prognostics approaches and application in repeated measurement data are summarized. Chapter 3 proposes a general mixed-effect regression model to quantify between-unit uncertainties in prognostics given repeated measurement data of multiple units by random effects. To improve the prediction accuracy and enhance model interpretation, covariates selection is based on the

physical understanding of the degradation process. Chapter 4 introduces the joint modeling method-based prognostics, where the degradation process is considered as a Gaussian process. The mean and covariance matrix of the degradation process are decomposed and modeled as functions of time-varying factors. The covariance matrix is parameterized through trigonometric functions. Penalized maximum likelihood estimation (PMLE) is employed to estimate parameters in the mean and covariance matrix, where measurements of each object are considered to be independent so that the scenario of unbalanced data can be well handled. Chapter 5 presents two other prognostics modeling methods to target the challenges of intensive computation caused by high dimension and spatial-temporal uncertainties. Applications of proposed prognostics approaches in battery capacity prognostics is used for the purpose of model validation. Moreover, performance evaluation of the above proposed methods is discussed. Chapter 6 concludes prognostics methodologies in repeated measurements studies and discusses the future work.

2. BACKGROUND

This chapter reviews fundamental work upon prognostics modeling. Studies on prognostics have been growing fast due to its significance. A vast of amount of research has been done on the review of prognostics approaches and their applications, such as [12, 23–30]. These prognostics approaches are often classified into three categories, i.e., data-driven, physics-based, and hybrid approaches, based on the usage of physical knowledge of the degradation mechanism (see Figure 2.1). This chapter extends the discussion to prognostics approaches for repeated measurements studies and identifies the research gaps.

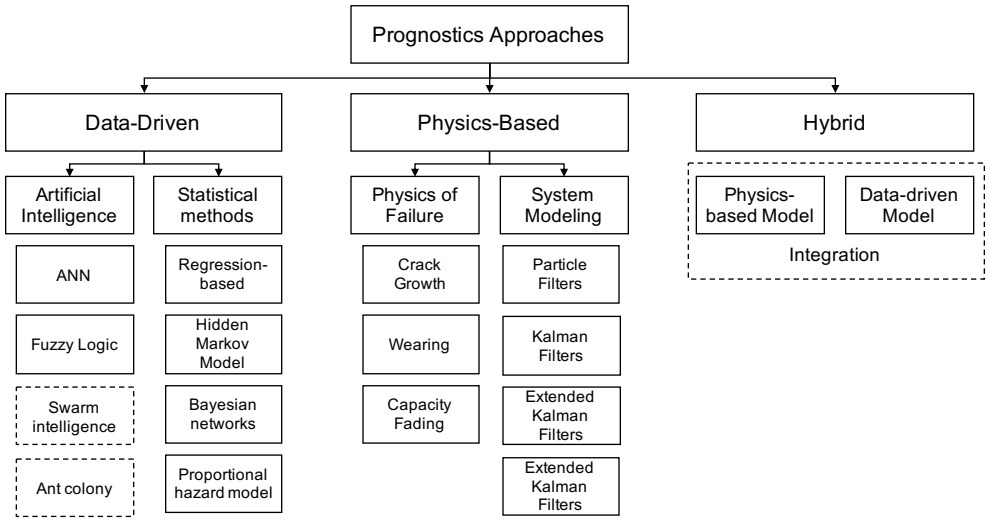

Figure 2.1: Summary of prognostics approaches

2.1 Data-driven prognostics

2.1.1 Overview

Data driven approaches are used to identify the characteristics of the current damage state and to predict the future damage using available historical data (observed data) when

rare physical understandings are available. The specialty of data-driven approaches is to derive models directly from the available data. Gouriveau et al. [1] took the analogy of a black box to describe data-driven prognostics. Data-driven approaches rely on the statistical characteristics [31]. Therefore, identification of these characteristics is the key process to develop data-driven methods. There are two major advantages of these methods with the firt one being low cost. These models can simply quantify the relationship the time and degradation states without or with rare knowledge of the system and so they come with a low implementation cost [32]. The development of computation techniques facilitates the implementation of data-driven methods. The other advantage is improved performance. With powerful machine learning techniques such as learning algorithms, data-driven approaches can largely reduce the model errors. Sometimes these models can broaden the cognition of a complex degradation process. However, it is clear that there are some disadvantages as well. These models are hard to interpret due to lacking system knowledge. Without physical understandings, it is difficult to illustrate model parameters. Sufficient data is the most important ingredient and lacks of training data holds back the implementation of these methods. Along with large amount of data, computational issues could exist. At last, data-driven methods are data-specified and so the obtained model could not be reused in another data.

There are two major types of data-driven approaches. Pecht, Eker, and Javed [10,30,33] divided data-driven approaches into machine learning and statistical models. Jardine et al., Dragomir et al., Peng et al., Elatter, and Schwabacher et al. [12, 23, 24, 29, 34] classified data-driven prognostics into artificial intelligence and statistical techniques. Data-driven approaches aim to automatically model the degradation behavior using historical data. On the other hand, machine learning is considered as a sub-field of artificial intelligence (AI) which uses historical data to automatically learn a model of the degradation process. Therefore, it is more reasonable to categorize data-driven approaches into AI and statistical techniques. Categories are not strictly separated and integrated methods are continuously emerging.

As listed in Table 2.1, artificial intelligence prognostics approaches include artificial neural network, fuzzy logic, decision trees, Bayesian network, and learning vector quantification. Statistical approaches include Gaussian process regression, least square regression, maximum likelihood estimation, statistical filters, hidden Markov model, mixed effects regression, and multivariate statistics. Merits and limitations of these data-driven approaches are also summarized (see Table 2.1).

Table 2.1: Summary of data-driven approaches

Methods	Merits	limitations
Artificial Neural Network (ANN) [35–39]	powerful approximation capability	non-transparent
Fuzzy logic [7, 40–43]	good at condition classification	missing the prediction
Decision tree [44–46]	transparent	unstable and complex
Bayesian network [47–51]	ready for sparse data and causality relationship	expensive computation
Learning vectors [52–54]	ready for high dimensional data	instability and non-transparent
Statistical methods		
Gaussian process [31, 55, 56]	good at multivariate analysis	strong assumption
Mixed effects modeling [57–59]	multiple sources of variability	computation
Stochastic filters [27, 60]	transparent	expensive computation

Artificial neural network (ANN) is the most popular approach and it is widely applied in prognostics. ANN is a powerful method to model complex degradation processes and modern statistical packages have developed mature ANN function which broaden its application. However, ANN has its own drawback which is lacks of transparency. It is very challenging to intuitively explain the outputs [12, 34]. Computation issue is raised when the size of the network grows. The philosophy of fuzzy logic can effectively explain the

conditional uncertainties, but the lack of prediction of time to failure limits its application [25, 61]. Learning algorithms, such as decision tree and learning vector quantification share merits and limitations, transparent but unstable [54]. For statistical methods, computation issue is the common limitation. Because statistical methods aim to quantify the variability in data, their implementation is not as easy as AI approaches.

2.1.2 Data-driven prognostics with repeated measurement data

Data-driven prognostics can be easily used to analyze repeated measurement data. The existing prognostics methods that analyze repeated measurement data are mainly Gaussian process [55, 62] and mixed effects regression [58, 59, 63]. Multiple sources of variability quantification is one of the objectives of these methods. For the Gaussian process, covariance matrix modeling is used to explain the individual variability and stochastic behaviors of parameters to illustrate the between-unit variability. Expensive computation limited its application in the large volumes of repeated data. Mixed-effect/random-coefficient regression, usually used in longitudinal data analysis, is a very adequate tool to analyze repeated measurement data. The fixed effect quantify the temporal variability and the random effects describes the difference between individuals. The implementation of mixed effects regression is usually met with computation challenges due to the stochastic behavior of its parameters. Both of these methods involve Bayesian inference through which model parameters are estimated. Statistical methods can provide effective models based on the performance data, but the parameter interpretation is still a challenge. Moreover, modeling complex repeated measurement data with unbalanced or missing data requires robust algorithms.

2.2 Physics-based approaches

2.2.1 Overview

Physics-based or model based prognostics is to incorporate the physics of failure and to quantify characteristics of the degradation process under various loads and operation conditions. These approaches can be categorized as physics of failure (PoF) approach and system modeling approaches. Generally PoF models are system/component specific, such as crack-growth model, rotation machine model, wearing model, gas-path model, and electrochemical model. To obtain these physics of failure models, several steps of failure modes, mechanisms and effects analysis (FMMEA), feature extraction, and RUL estimation are needed (see Figure 2.2). This approach is effective and descriptive due to incorporating physical understandings of the degradation process. System modeling approaches assume the system can be described by a model that can illustrate the stochastic behavior of the system over degradation. For example, state space model can transform a physical system into a set of input, output, and state variable related to the first-order differential equations. One of the advantages of the physics-based prognostics is to successfully incorporate physical understanding of the system. This eases the interpretation of parameters in prognostics models. More importantly, with physical understanding, model errors caused by unexplained features or un-modeled phenomena can be significantly reduced. Therefore, physics-based prognostics can be used to improve the accuracy of prognostics modeling. The limitation of physics-based approaches mainly lies in its high implementation cost. physics-based models are considered as the most expensive approaches among prognostics approaches [32]. To obtain high fidelity physical model, sufficient experiments are required. Intensive computation might cause model fail. Physics-based model are usually system specific and then the reusability is limited. Table 2.2 summarizes the most commonly used physics-based approaches.

The benefit of physics-based model is the ability to incorporate physical understanding of the system and improve the accuracy of prognostics modeling. Physical laws, such as crack growth, fatigue, and wearing, can provide a “long-term” prediction under various

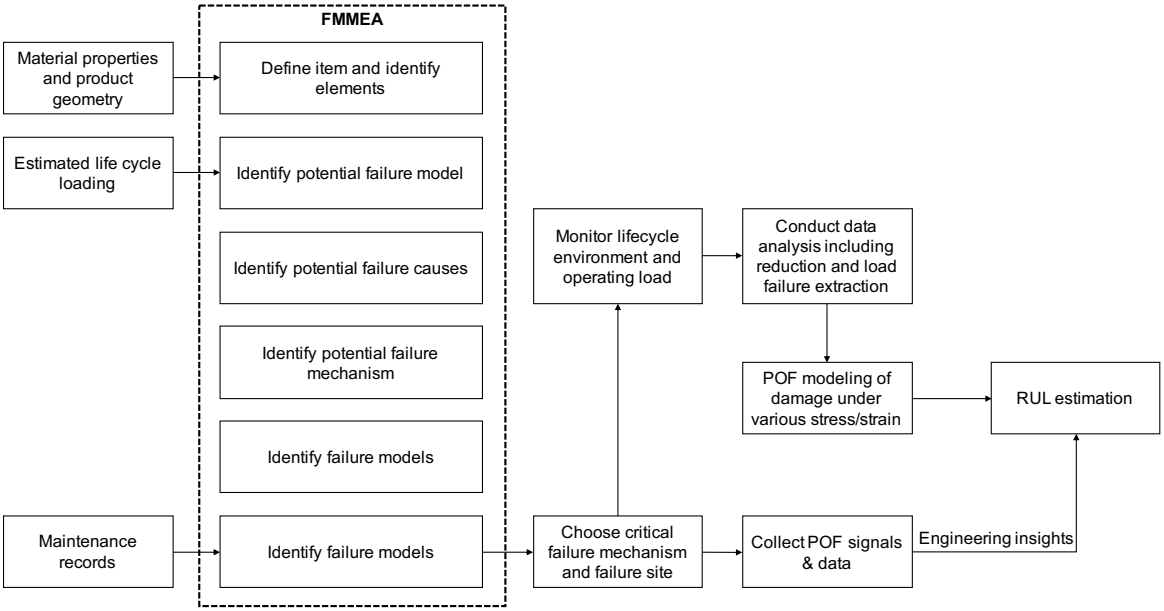

Figure 2.2: A procedure of physics-based prognostics (adapted from [10])

Table 2.2: Summary of physics-based approaches

Models	Application
Physics of Failure	
Paris's law crack	[64–67]
Bearing	[68, 69]
Fatigue	[70, 71]
Stiffness-based rule	[65, 72]
Electronic	[22, 73]
System Modeling	
Statistical filter	[74]

loads. Statistical filters used to estimate the state space are considered as an integrated method that combines physics-based and data-driven prognostics. Physics-based models,

such Paris' law, are usually deterministic functions, which are limited in analyzing variability of the degradation process.

2.2.2 Physics-based prognostics with repeated measurement data

Repeated measurement data on multiple units illustrates multiple degradation trajectories. The characteristic of system/component specific limits the application of physics-based model in repeated measurements studies. However, the physical understanding from physics-based models can be used to improve the interpretation and accuracy in repeated measurements studies. The incorporation of system knowledge in multiple degradation trajectories is a challenging task, since there might be multiple measurements at the same measurement factor, i.e., measurement time.

2.3 Hybrid approaches

2.3.1 Overview

Hybrid approaches is the integration of both data-driven and physics-based prognostics. It is an intuitive idea to leverage the strength of both aforementioned data-driven and physics-based prognostics to improve the prediction performance. Hybrid approaches, in one side, can broaden the usage of physics-based models and also improve the interpretation of data-driven models. Figure 2.3, adapted from [32], illustrates the the application range and the prediction accuracy of data-driven and physics-based models. Hybrid models "inherit" the wide application range from data-driven approaches and high fidelity from physics-based approaches. Extensive work of hybrid approaches has been done. For the implementation procedure, Kumar et al. [75] and Cheng et al. [76] proposed 9 steps for fusing prognostics method and emphasized the importance of understanding physics of failure (see Figure 2.5). Based on the major objective of diagnostics, Sankavaram et al. [77] designed a six-step integrated diagnostic and prognostic process to implement the hybrid approaches. These framework can be used for any system but not the only way. Five key

steps required for a hybrid method include data acquisition, feature extraction, diagnostics, degradation modeling, and RUL prediction. Diagnostics is to identify failure modes and degradation mechanism, which can be used to provide the physical understanding for prognostics modeling. Based on methods of degradation modeling and RUL prediction, Liao et al. [78] classifies the hybrid approaches into four categories:

- use a data-driven model to infer a prognostics model and use a physics-based model to predict RUL, such as [76, 79, 80],
- use a data-driven model to replace the system model of physics-based model, such as [81],
- use a data-driven model to prediction future degradation state and use physics-based for RUL prediction, such as [82],
- use both data-driven and physics-based model for RUL prediction and fuse their results, such as [83–85].

The first three categories can be understood as data-driven or physics-based model serve as priori for each other, i.e., these two methods work in series. The forth category is to launch data-driven and physics-based models simultaneously and fuse their results, i.e., two integrated methods work in parallel.

2.3.2 Hybrid prognostics with repeated measurement data

Hybrid models are widely used for prognostics due to their strength inheriting from both data-driven and physics-based models. However, existing hybrid methods used in repeated measurement data, mainly regarding the development of data-driven techniques and physical understanding are rarely incorporated. This research aims to develop adequate hybrid prognostics models. To maintain the model accuracy for the purpose of prediction, the incorporation of the physical understanding in this research is done through covariate identification and functional forms selection for the models.

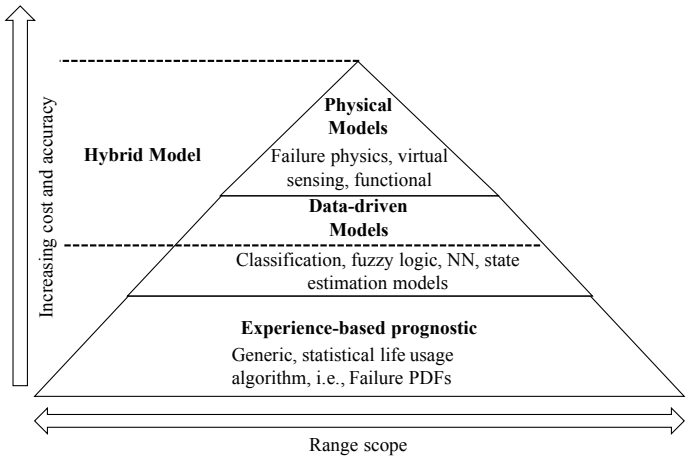

Figure 2.3: Application and cost of various prognostics approaches

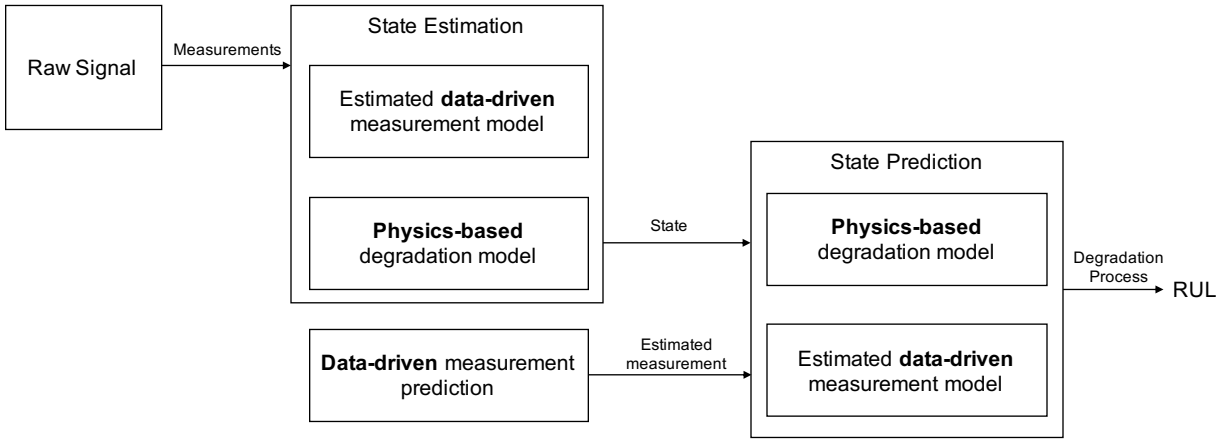

Figure 2.4: Framework of hybrid approaches (adapted from [78])

2.4 Summary of prognostics approaches

Prognostics with repeated measurement data on multiple units confront challenges from the multiple sources of variability, complex data structure, and expensive computation. As mentioned in Table 1, data-driven techniques, such as mixed effects model, can be used to quantify multiple sources of variability. As well, techniques from data science are desired to reduce risks of uncertainties from the input data. Physics-based model can improve the

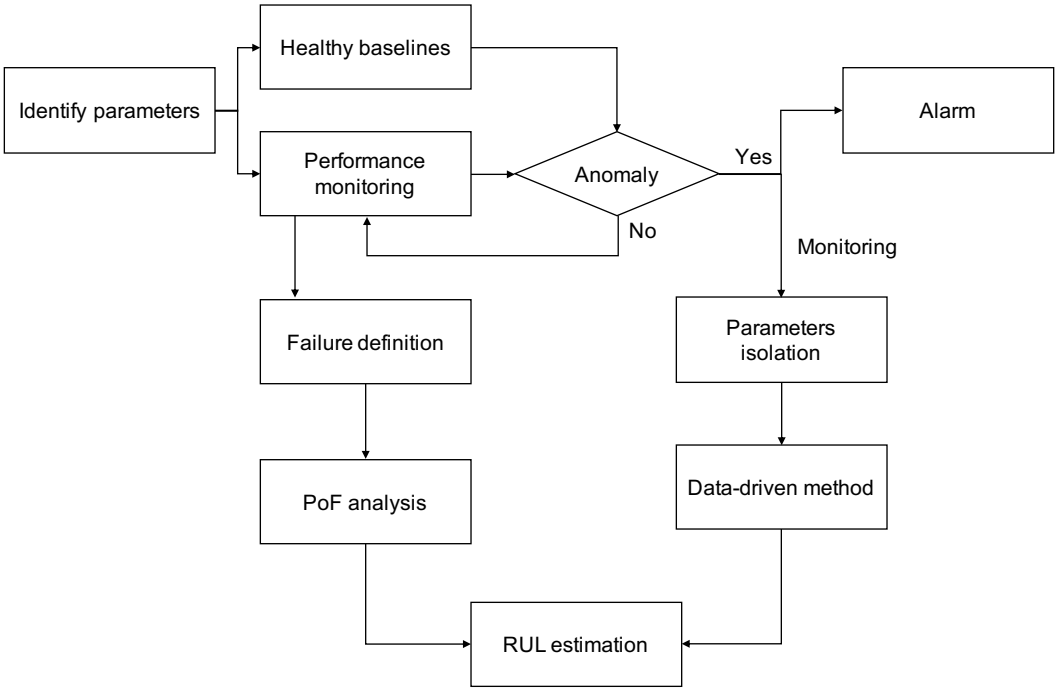

Figure 2.5: Procedure of hybrid approaches (adapted from [10])

prediction performance and keep the prognostics parsimony. For prognostics with repeated measurement data, hybrid methods are desired to combine the strengths of these two types of approaches. However, prognostics models for repeated measurement data are not well investigated. Specifically, few work is proposed to

1. quantify multiple sources of variability
2. model for complex repeated measurements data
3. assess temporal and spatial variability.

To fill these research gaps, this research proposes the general mixed-effect model, joint modeling method, and functional principal component analysis-based prognostics. Eventually a general spatio-temporal model is proposed based on the aforementioned methods, where spatial, temporal trends and their dependency will be quantified, analog to as the variation between units , temporal variability, and their correlation.

3. PROGNOSTICS BASED ON GENERAL MIXED-EFFECT MODEL

This chapter presents the application of general mixed-effect model (GMM) in prognostics modeling and the case study of lithium ion battery capacity degradation is used to demonstrate the proposed method.

GMMs containing fixed and random effects are widely used to account multiple sources of variability in repeated measures and longitudinal data. The fixed effects can be interpreted as the characteristics of the population average over measuremenets. The random effects show the variation of units. The combination of fixed and random effects illustrates the variability in a stochastic process [86]. General mixed-effect models can be used to quantify between-unit variability coming from the systematic difference among units in repeated measurement data and within the degradation process from the temporal stochastic. This research proposes a general linear mixed-effect model-based prognostics methodology to investigate the multiple sources of variability within the repeated measurement data. Covariate identification, model selection, and time to failure prediction are included in the proposed method. Figure 3.1 presents the procedure of prognostics based on the general mixed-effect model.

3.1 General mixed-effect model

Let y_{ij}, $i = 1, \ldots, n, j = 1, \ldots, m_i$ be the jth measurement on the ith unit, where n is the number of units and m_i is the number of measurements on the ith unit. The mixed-effect model has a two-level representation, i.e., within-unit and between-unit. For a certain j, $j = 1, \ldots, m_i$, the within-unit regression model is given by

$$y_{ij} = z_{ij}\boldsymbol{\beta_i} + e_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i, \tag{3.1}$$

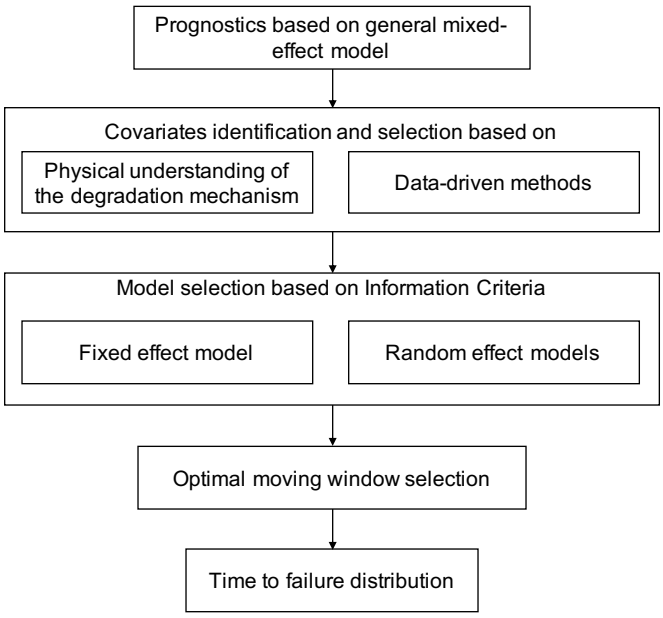

Figure 3.1: Schematic of prognostics based on mixed effects model [59]

where $\boldsymbol{\beta_i}$ is a $q \times 1$ vector of regression coefficient that is specific for the ith unit. z_{ij} is the jth row of $\boldsymbol{Z_i}$, a $m_i \times q$ matrices of covariates. $e_{ij} \sim N(0, \sigma_{ij})$ are zero mean residuals. In the between-unit stage, $\boldsymbol{\beta}_i$, $i = 1, \ldots, n$ are considered as q-dimensional random vector with mean of $E(\boldsymbol{\beta_i}) = A_i\boldsymbol{\alpha}$ and variance var$(\boldsymbol{\beta}_i) = \boldsymbol{D}$. A_i is a diagonal matrix with diagonal entities which are factors that vary from unit to unit and $\boldsymbol{\alpha}$ is the regression coefficients depends only the population average as shown in Equation (3.2). That is, $\boldsymbol{\beta}_i$ can be interpreted as a special case of the population average coefficient $\boldsymbol{\alpha}$. Let $\boldsymbol{b}_i$, $i = 1, \ldots, n$, be a random variable with the mean of zero and variance $\boldsymbol{D}$. Therefore, $\boldsymbol{\beta}_i$ can be rewritten as Equation (3.3).

$$\boldsymbol{A_i} = \begin{bmatrix} a_{li} & 0 & 0 & \cdots & 0 \\ 0 & a_{2i} & 0 & \cdots & 0 \\ 0 & 0 & a_{3i} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{qi} \end{bmatrix}, \boldsymbol{\alpha} = (\alpha_1^T, \ldots, \alpha_q^T). \tag{3.2}$$

$$\boldsymbol{\beta_i} = \boldsymbol{A_i}\boldsymbol{\alpha} + \boldsymbol{b_i} \tag{3.3}$$

Substituting Equation (3.3) in the matrix of Equation (3.1), between-unit regression model can be written as Equation (3.4).

$$\begin{aligned} \boldsymbol{y_i} &= \boldsymbol{Z_i}\boldsymbol{\beta}_i + \boldsymbol{e_i}, \\ \boldsymbol{y}_i &= \boldsymbol{Z}_i\boldsymbol{A}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \\ \boldsymbol{y_i} &= \boldsymbol{X_i}\boldsymbol{\alpha} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{e_i}, \end{aligned} \tag{3.4}$$

where $\boldsymbol{e}_i$ is a zero mean multivariate normal random variable, i.e., $\boldsymbol{e}_i \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma_e})$. The dimension of $\boldsymbol{\Sigma_e}$ is determined by the number of measurements of each unit. By assuming that there are the same number of measurements of each unit, i.e., balanced data, $\boldsymbol{\Sigma}_e$ can be written as $\boldsymbol{\sigma}_i^2\boldsymbol{I}$. To make this mixed model more flexible, a generic covariate vector $\boldsymbol{X}_i$ is used to replace $\boldsymbol{Z}_i\boldsymbol{A}_i$. $\boldsymbol{\alpha}$ and $\boldsymbol{b}_i$ are called fixed and random effects coefficients, respectively. Assuming that $\boldsymbol{b}_i$ are independently distributed as multivariate normal distribution, i.e., $\boldsymbol{b}_i \sim MVN(0, \boldsymbol{D})$. The response $\boldsymbol{y}_i$ can be illustrated in the form of a multivariate normal distribution with the mean $\boldsymbol{X}_i\boldsymbol{\alpha}$ and the covariance $\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i + \boldsymbol{\sigma}_i^2\boldsymbol{I}$ as shown in Equation (3.5).

$$\boldsymbol{y}_i \sim MVN(\boldsymbol{X}_i\boldsymbol{\alpha}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i + \boldsymbol{\sigma}_i^2\boldsymbol{I}). \tag{3.5}$$

According to Equation (3.5), the maximum likelihood estimation (MLE) is an intuitive way to estimate the parameters $\boldsymbol{\alpha}$, $\boldsymbol{D}$, and $\boldsymbol{\sigma}_i$. Cnaan et al. [87] discussed the basic MLE of $\boldsymbol{\alpha}$ and the variance $\boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i + \boldsymbol{\sigma}_i^2\boldsymbol{I}$, as shown in Equation (3.6).

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \sum \boldsymbol{X}_i^T\hat{V}_i\boldsymbol{X}_i \sum \boldsymbol{X}_i^T\hat{V}_i\boldsymbol{y}_i, \\ \hat{V}_i &= \boldsymbol{Z}_i\hat{\boldsymbol{D}}\boldsymbol{Z}_i^T + \hat{\boldsymbol{\sigma}}_i^2\boldsymbol{I}. \end{aligned} \tag{3.6}$$

A ***R*** package developed by Bates et al. [88], ***lme4*** in ***R*** [89], is used in this research to estimate the fixed and random effects. The main idea of this function is based on the penalized weighted residual sum-of-squares on the $\boldsymbol{\alpha}$, $\boldsymbol{D}$ and $\boldsymbol{\sigma}_i$.

3.1.1 Model selection

Model selection is utilized to select parsimonious models with desirable properties. The existence of fixed and random effects complicates the model selection in mixed-effect models. There are two major steps in model selection for a mixed effects model, the first one

being covariates identification. Predictive covariates are used to illustrate the degradation mechanism. It is an intutive idea to use physics-based models to select the candidates of covariates. For example, the measurement factor, such as measurement time t, is usually the first choice of covariates for repeated measurement data. Physical understanding is needed to specify the function form of the measurement time, e.g., t, t^2, or $\sqrt{t}$. Especially when the number of fixed and random effects is not known, a subset of covariates determined based on physics-based models can be helpful to decrease the cost of the mixed model. When there are large number of candidates, statistical tools, such as analysis of variance (ANONA) and L_1 norm penalized least square (LASSO), can be used to find the optimal subset [90]. This research applies design of experiment and ANOVA to select the identified candidate covariates through physics-based models [91]. The second step is to determine the fixed and random effects. The selected model is expected to be a well-fitting and also in an intuitive interpretation form. Information criteria including Akaike's Information (AIC) and Bayesian Information Criteria (BIC) is the most commonly used techinique for model selection [92]. Due to the high robustness of BIC in multi-variance structure mixed-effect models, this reserach uses BIC to select the optimal model. BIC is the log-likelihood-based metric for model selection that overcomes the over parameterization issue by penalizing the number of model parameters as well as the data size [93]. Equation (3.7) illustrates the calculation of BIC. A smaller BIC value indicates a better model fit.

$$\begin{aligned} \text{BIC} &= -2\ell + (p-2)\log(n), \\ \ell &= -\frac{n}{2}(\log 2\pi + \log \sum_{i=1}^{n} (\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i)^2 + 1), \end{aligned} \tag{3.7}$$

where p is the number of parameters in the model and n is the total number of observed measurements.

3.1.2 Performance prediction and time to failure distribution

3.1.2.1 Performance prediction

This research investigates two types of data input schemes, i.e., the cumulative data input and moving window data input. Cumulative data input is to incorporate all the available historical data to predict future values. It is obvious that the strength of the "impact" that one measurement has on another measurement will on average decay as a function of their corresponding time lag in the time sequence. For a relatively slow degradation process, such as lithium ion battery capacity degradation, the previous data has a weak effect on the current observation and might distort the prediction. The concept of moving-window is to use information of the most recent observations to predict the future values. Models based on a large window become the less specific information along the perturbation direction, while those based on a small window are more effective in obtaining the pertinent information of perturbation direction. But the small moving window might not be robust with high noise levels. The optimal length of moving window can be determined by the prediction performance.

The prediction performance of moving window input and cumulative input schemes can be assessed through the deviation between the predicted and actual value. There are two major prediction strategies, one step ahead prediction and multi-step ahead prediction. In multi-step ahead prediction, the variable of more than two steps ahead is estimated based on all or some previous observations. One step ahead prediction is to estimate the variable of next step immediately. Uncertainties, such as an accumulative errors, arise from various sources in multi-step ahead prediction. To avoid such uncertainties and obtain the "best" model parameter, one step ahead prediction is used to determine the optimal length of the moving window and then compare the performance of the aforementioned two data input schemes. The mean absolute deviation (MAD) as shown in Equation (3.8) used in this research as the prediction performance of the aforementioned two data input schemes.

Lower the MAD is, more accurate the prediction is. Figure 3.2 presents the moving window and cumulative data input strategies for one-step ahead prediction.

$$\begin{aligned} \text{MAD} &= \frac{\sum_{i=1}^{n} |\Delta_i - \bar{\Delta}|}{n}, \\ \Delta_i &= y_{observed} - y_{predicted}, \\ \bar{\Delta} &= \frac{\sum_{i=1}^{n} \Delta_i}{n}. \end{aligned} \tag{3.8}$$

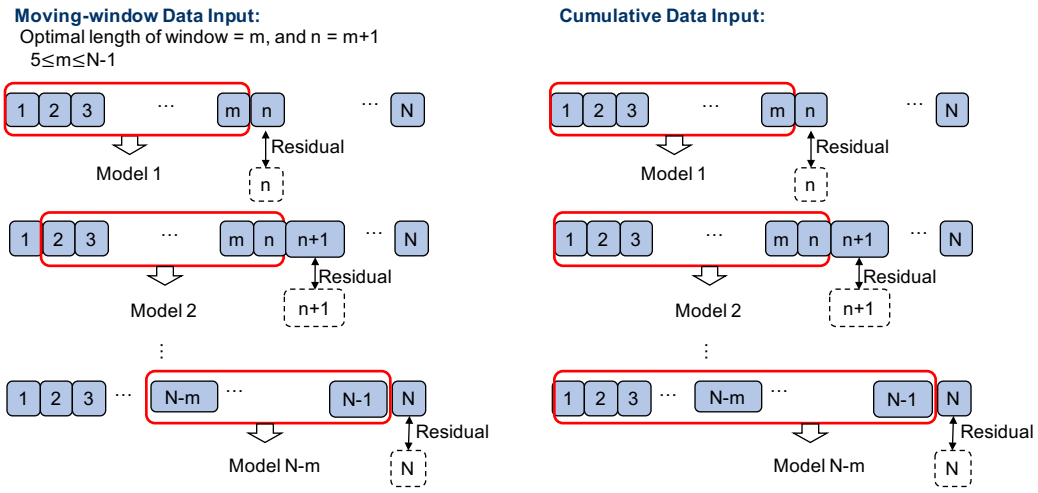

Figure 3.2: Moving window and cumulative data input scheme for performance prediction [59]

3.1.2.2 Time to failure distribution assessment

Based on the selected prediction model, time to failure (TTF) can be estimated with the given the failure performance threshold. Due to random effects, random effects coefficients are estimated and given in the form of statistical distributions. That is, coefficients are random variables. Based on the basic idea of the numerical simulation, random samples of coefficients can be obtained from their statistical distributions. For each sample, a point of time to failure can easily be calculated based on the degradation model. With a certain number of simulated TTF points, such as 10,000, the distribution of TTF can be evaluated.

3.2 A Bayesian approach for Li-Ion battery capacity fade modeling and cycles to failure prognostics

This research investigates the application of general mixed-effect regression model in battery capacity fade modeling over repetitive cycles by considering both within-battery and between-battery variations. Physics-based covariates are integrated with functional forms for modeling the capacity fade. A systematic approach based on co-variate identification, model selection, and a strategy for prognostics data. selection is presented. The proposed Bayesian method is capable of quantifying the uncertainties in predicting battery capacity / power fade and end-of-life cycles to failure distribution under various operating conditions.

3.2.1 State of the art

Lithium Ion batteries attract increasing attentions due to their high energy density and long lifetime. Battery capacity is defined as the maximum amount of electric charge that a fully charged battery can deliver, which is one of three major performance indicators along with the internal resistance and self-discharge. The lifetime of a battery is deemed to end when its capacity reaches 80% of the initial capacity in general. The loss of active Lithium ions during the micro-electrochemical reaction inside the battery during charge/discharge process causes the capacity fade. In order to extend the lifetime of batteries, extensive research has been conducted to explore new battery materials for electrodes and electrolytes, designing new electrode structures, and investigating various battery aging mechanisms [94–107]. Capacity fade prediction over repetitive charge/discharge cycling based on quantitative mathematical models is another important research topic. Accurate estimation and prediction of battery failure time provides the decision information for timely replacement of degraded batteries before the batteries reach the end of their useful life.

Battery capacity fade is mainly attributable to electrochemical reactions, and the associated factors causing such reactions can be leveraged for accurate performance degradation prediction [108]. These factors include charge-discharge protocol, temperature,

state of charge (SOC), and materials of electrode and electrolyte. For different types of batteries, capacity fade mechanisms are usually dominated by unique aging mechanisms [109–111]. For example, it has been demonstrated that active lithium loss due to electrolyte decomposition forming the solid electrolyte interphase (SEI) in the carbon negative is the dominant aging mechanism for $LiFePO_4$ batteries [110]. It has also been shown that a substantial decrease in the surface conductivity in lithium nickel cobalt aluminum oxide ($LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$) cells dominate the cells' aging processes [111]. In addition, Broussely et al. [109] observed that the anode's structural instability due to volumetric change induced stress during repeated charge/discharge cycling can also lead to accelerated capacity fade. Smith et al. [112] interpreted the degradation mechanism of Li-ion cells using a lithium inventory model to account for the unwanted parasitic reactions during the battery charge and discharge cycling process. Parasitic processes and reactions, including the solid electrolyte interphase (SEI) growth and repair at the negative electrode, the electrolyte oxidation at the positive electrode, the dissolution of transition metal ions, and the rate of positive electrode damage, are incorporated in the proposed lithium inventory model.

The uncertainties occurring during the chemical reactions result in the difficulties of the estimation of capacity fade under cycling usage. A variety of approaches to capture the randomness of the capacity fade are reported in recent literature, which can be categorized mainly into data-driven and hybrid methods [113]. Data-driven prognostic techniques incorporate the available and historical information to statistically and probabilistically derive decisions, estimates, and predicts the health and reliability [10]. Many works of machine learning algorithm have been adopted for modeling batteries performance, such as Support vector machines (SVMs) [52, 114–116], Relevance vector machines (RVMs) [117, 118], k-nearest neighbor (kNN) [119], and Artificial neural network (ANN) [120–126]. The concept of entropy from the information theory is also applied to model the battery capacity, which is proved to be very effective to quantify the capacity variation information with non-standard charge/discharge cycling [127, 128]. Data-driven methods can provide parsimonious models with accurate prediction performance. However, it can become very

difficult to provide meaningful physical understandings to covariates in these models. Hybrid methods incorporate the physics of failure into data-driven methods. From the electrical engineering point of view, filter-based methods are developed to evaluate the state of charge based on the change of the open-circuit voltage and estimate the capacity of Lithium Ion batteries within a closed circuit [129–139]. From the electrochemical point of view, the formation, growth, and repair of solid electrolyte interface (SEI) are the main reasons for losing active lithium ions [93,140,141]. Analytic models of capacity fading are explored through analyzing the chemical process of SEI's formation and growth [142,143]. Some literature, such as [56, 144–148], focus on the temporal uncertainty in the capacity degradation process.

Most of the aforementioned literature, however, fails to quantify multiple sources of uncertainties in the battery degradation process modeling, such as within-unit and between-unit variations. To fill the mentioned gap, this paper aims to investigate the multiple sources of variations in repeated measurement design involving multiple units. Repeated measurement design is motivated by the desire of quantifying statistical uncertainties in a process. Multiple responses of each experimental unit in multiple conditions and occasions and multiple units are observed so that variation of within-unit and between-unit over the measurement factors can be detected. Guo et al. [59] proposed a mixed-effect model with the moving window to investigate between-and-with unit uncertainties in the capacity degradation using balanced data.

The novel Bayesian method based on the general mixed-effect model, discussed in Section 3.1, is used to investigate battery capacity fade over repetitive cycles by considering both within-battery and between-battery variations. Physics-based covariates are integrated with functional forms for modeling the capacity fade. A systematic approach based on covariate identification, model selection, and a strategy for prognostics data selection is presented. The proposed Bayesian method is capable of quantifying the uncertainties in predicting battery capacity fade and end-of-life cycles to failure distributions under various operating conditions. The schematics for battery capacity fade modeling and prognostics is shown in Figure 3.3.

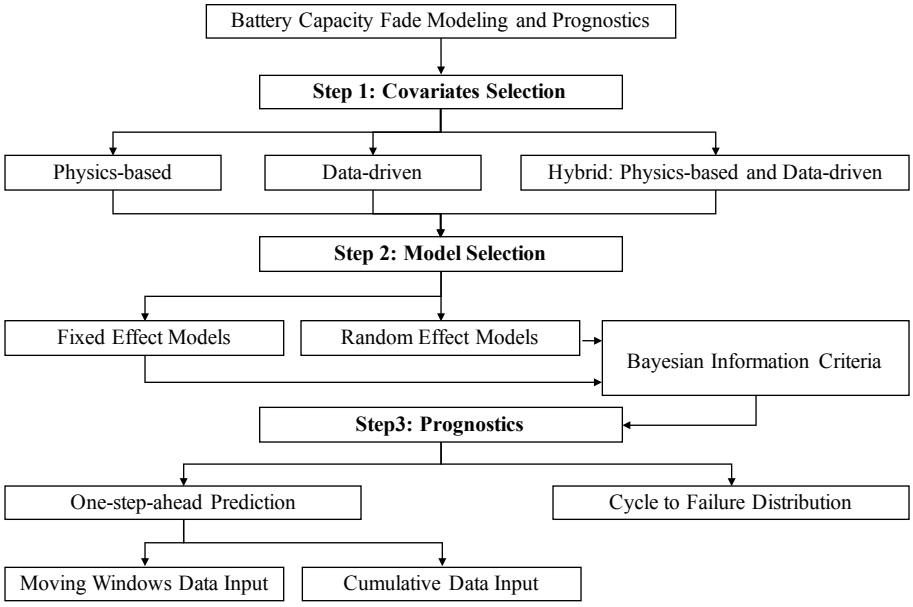

Figure 3.3: Schematic for battery capacity fade modeling and prognostics

3.2.2 Covariate identification and selection

To accurately model the capacity fade of Li-Ion batteries, appropriate predictors or covariates need to be identified such that a specific functional prediction model can be established using the identified predictors. The functional prediction models also need to consider the physics-of-failure mechanisms based on the major failure modes of Li-Ion batteries, such as the formation, repair, and restoration of the solid electrolyte interphase (SEI). SEI formation can be related to battery type, charge/discharge cycles, charge rate, and environmental factors. Among these candidates, designed experiments are used to select the major significant covariates associated with capacity fade so that adequate predictive models can be built.

3.2.2.1 Predictive covariate identification based on physics-of-failure analysis

Battery capacity fade over time is usually attributed to the internal electrochemical reactions of battery cells. Capacity fade also depends on varying external environment all

conditions, operating conditions, and battery materials. In the literature, SEI formation and increases in anode impedance are regarded at the main failure modes of Li-ion batteries, and are related to the charge-discharge rate, cycle number, temperature, and anode material particle size. We studied the capacity fade effects related to battery types, discharge rates, and the number of charge/discharge cycles.

3.2.2.2 Covariate selection based on designed experiments

The aforementioned aging factors of battery capacity fade were investigated using full factorial design of experiments. Eight batteries from two types of $LiCoO_2$ batteries with different structural configurations, denoted as *CS2* and *CX2*, were tested at two levels of discharge rates ($0.5C$ and $1C$) [91]. The basic specification description of these batteries is shown in Table 3.1.

Table 3.1: Specifications of battery *CS2* and battery*CX2*

	Rating capacity(Ah)	Cathode	Anode	Weight(g)	Dimensions(mm)
CS2	1.10	$LiCoO_2$	graphite	21.1	$5.4 \times 33.6 \times 50.6$
CX2	1.35	$LiCoO_2$	graphite	28.0	$6.6 \times 33.8 \times 50.0$

All batteries were tested under the same charge-discharge profile: each battery was first charged at a constant current rate until the voltage reached a cutoff value, then a constant voltage charging was sustained until the charge current dropped to a threshold value. The battery was discharged at a constant current until the discharging voltage dropped to a cutoff value. The current and voltage change profiles are shown in Figure 3.4. The cutoff voltage was set as 4.2V (charging cutoff value) and 2.7V (discharging cutoff value) in these tests. Of the eight batteries of two types, four of them, i.e.,*CS2_33*, *CS2_34*,*CS2_35*, and *CS2_36*, were tested under the profile in which the discharge rate was $0.5C$, and the other four, i.e.,*CX2_33*,*CX2_34*, *CX2_35*,and *CX2_36* were tested under a $1C$ discharge rate. The experimental settings and battery specifications are summarized in Table 3.2. Table 3.3

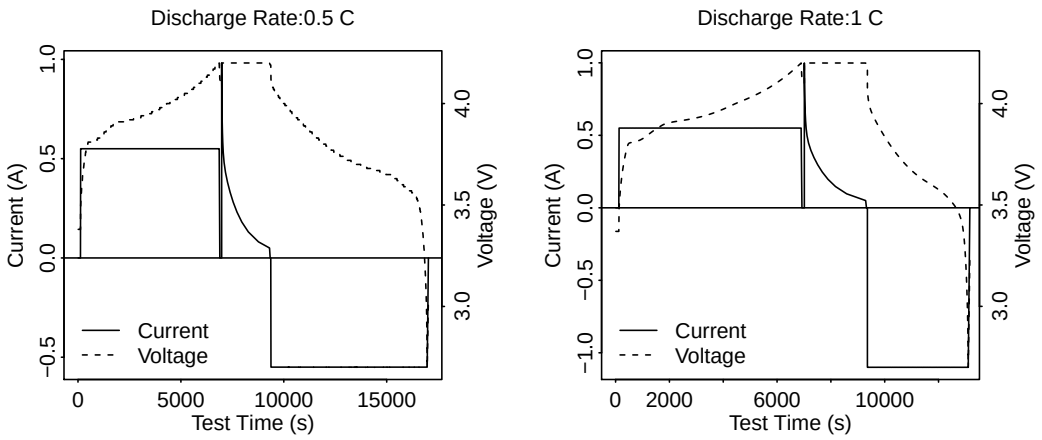

Figure 3.4: Current and voltage profiles under two levels of discharge rates

shows the test results of both capacity and power fade data of the eight batteries. Figure 3.5 presents the individual battery capacity and power fade over the charge and discharge cycles.

Table 3.2: Summary of experiment data

	Sample	Discharge rate	Cycles number	Test time (h)
CS2	*CS2_33*	$0.5C$	862	3120.1
	CS2_34	$0.5C$	774	5055.7
	CS2_35	$1C$	930	2664.6
	CS2_36	$1C$	970	2702.6
CX2	*CX2_33*	$0.5C$	1071	4411.8
	CX2_35	$0.5C$	1747	6908.1
	CX2_34	$1C$	1724	5158.0
	CX2_36	$1C$	1958	5926.0

Based on the collected battery performance degradation data and pre-set experiment conditions, a full factorial design with two replications was implemented to evaluate the capacity fade over prolonged usage life. Number of cycles, discharge rate, and battery type

Table 3.3: Test results for *CS2* and *CX2* batteries

Sample	Cycle to failure	Starting capacity(Ah)	Ending capacity(Ah)	Starting discharge energy(Wh)	Ending discharge energy(Wh)
CS2_33	615	1.16	0.06	4.34	0.56
CS2_34	564	1.15	0.52	4.28	1.83
CS2_35	645	1.14	0.30	4.16	1.02
CS2_36	549	1.16	0.37	4.18	0.58
CX2_33	704	1.29	1.01	4.76	3.65
CX2_34	703	1.36	0.56	4.86	1.88
CX2_35	912	1.35	0.56	5.04	1.95
CX2_36	762	1.36	0.54	4.08	1.81

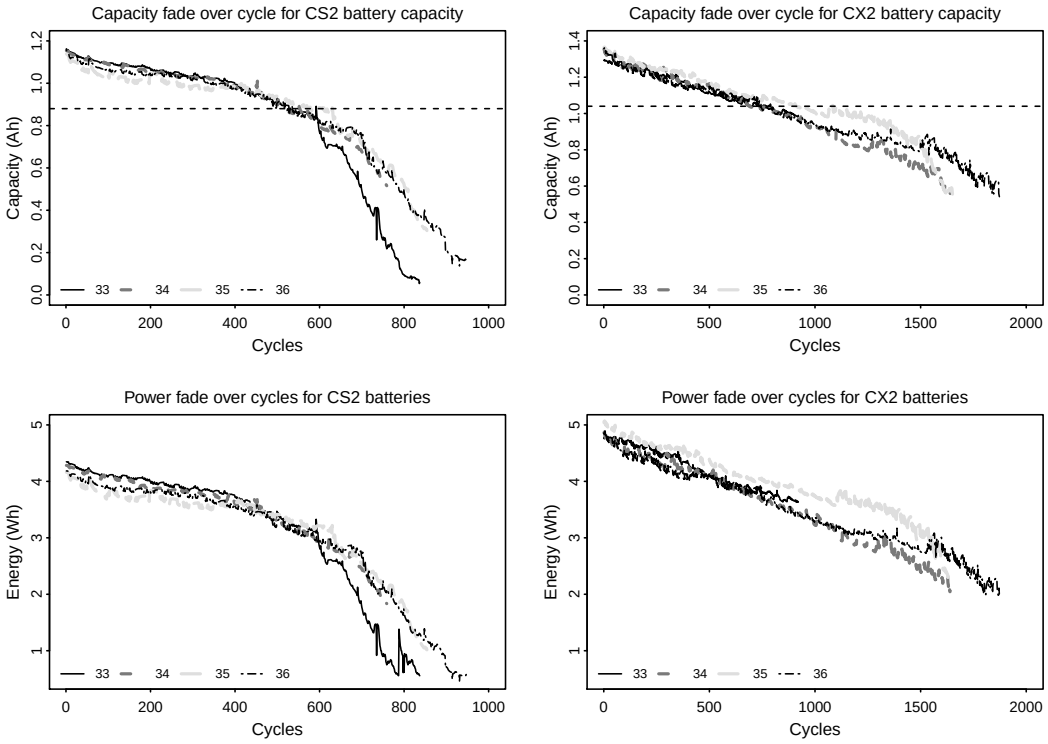

Figure 3.5: Capacity and power fade over cycles for battery types *CS2* and *CX2*

were considered in this experimental design, and the levels of each factor were designed as in Table 3.4. The ANOVA results of the full factorial experiment are shown in Table 3.5. At $\alpha = 0.05$ significance level, the number of cycles and battery type are significant with respect to capacity fade for individual effects and interaction effects. The discharge

rate and its interactions with the other two factors were not significant at the two selected discharge rates. As a result, different types of batteries need to adopt different capacity fade models.

Based on both physics of failure analysis and design of experiment analysis, the number of cycles is important covariate for the battery capacity fade. Moreover, The thickness or growth of SEI has been physically modeled to be proportional to the square root of the number of cycles [110]. Ecker et al. [149] investigated the calendar cycle life of Li-ion batteries and developed capacity fade models with variables of time, square root of time and logarithm time. Bloom et al. [150] demonstrated that the area-specific impedance rise and power fade both follow the square root of time kinetics, which explains the SEI layer growth. In literature, capacity has also been modeled as a function of the logarithm of the number of cycles [111]. Therefore, it can be concluded that the square root of cycles and logarithm of cycles are the candidate functional forms for capacity fade modeling.

Table 3.4: Design factors and factors levels

Factor	Levels	Values			
Cycle	4	0-300	500-600	600-900	> 900
Discharge rate	2	$0.5C$	$1C$		
Battery type	2	*CS2*	*CX2*		

3.2.3 Capacity fade modeling using linear mixed effects models

Given identified predictive covariates and its candidate functional forms, seven models with various combinations of fixed effects and random effects are built to quantify uncertainties due to both between-battery and within-battery variations. The explored models are evaluated and compared using Bayesian Information Criterion (BIC), and an optimal model is selected for modeling battery capacity fade and health prognostics. Equation (3.9)-(3.15) present these seven models.

Table 3.5: Analysis of variance for capacity, using adjusted Sum of Squares (SS) for tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Cycle	3	1.13546	1.31547	0.43849	164.37	0.000
Discharge Rate	1	0.00783	0.00002	0.00002	0.01	0.937
Battery Type	1	1.2651	1.19625	1.19625	448.41	0.000
Cycle * Discharge Rate	3	0.00858	0.00165	0.00055	0.21	0.891
Cycle * Battery Type	3	0.34180	0.35201	0.11734	43.98	0.000
Discharge Rate * Battery Type	1	0.00389	0.00490	0.00490	1.84	0.195
Cycle * Discharge Rate * Battery Type	3	0.00928	0.00928	0.00309	1.16	0.358
Error	15	0.04002	0.04002	0.00267		
Total	30	2.67336				
S=0.00516505, R-Sq=98.50%, R-Sq(adj)=97.01%						

1. Capacity fade models with square root of cycle as a covariate

Model 1: Fixed effect intercept and slope model:

$$\begin{aligned} y_{j[i]} &= \alpha + \beta_1 \times \sqrt{t_{j[i]}} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2), i = 1, 2, \cdots, n_J \end{aligned} \tag{3.9}$$

Model 2: Random effect intercept and slope model:

$$\begin{aligned} y_{j[i]} &= \alpha_{j[i]} + \beta_{1j[i]} \times \sqrt{t_{j[i]}} + \varepsilon_i \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \beta_{1j} \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2) \\ (\alpha_i, \beta_{1j}) &\sim MVN((\mu_\alpha, \mu_{\beta_1})^T, \Sigma), j = 1, 2, \cdots, J, i = 1, 2, \cdots, n_J. \end{aligned} \tag{3.10}$$

2. Capacity fade model with logarithm of cycle as a covariate

Model 3: Fixed effect intercept and slope model:

$$\begin{aligned} y_{j[i]} &= \alpha + \beta_2 \times \log t_{j[i]} + \varepsilon_i \\ \varepsilon_i &\sim N(0,\sigma^2), j = 1, 2, \cdots, J, i = 1, 2, \cdots, n_J. \end{aligned} \tag{3.11}$$

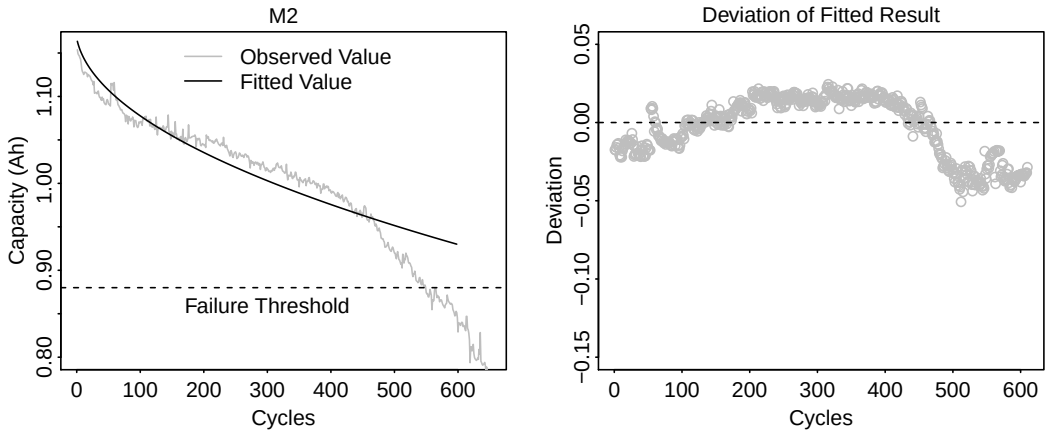

Figure 3.6: Fitted values of Model 2 compared to observed values

Model 4: Random effect intercept and random slope model:

$$
\begin{aligned}
y_{j[i]} &= \alpha_{j[i]} + \beta_{2j[i]} \times \log t_{j[i]} + \varepsilon_i \\
\alpha_j \sim N(\mu_\alpha,&\sigma_\alpha^2), \beta_{2j} \sim N(\mu_{\beta_2}, \sigma_{\beta_2}^2), \varepsilon \sim N(0, \sigma^2) \\
(\alpha_j, \beta_{2j}) &\sim MVN((\mu_\alpha, \mu_{\beta_2})^T, \Sigma), j = 1, 2, \cdots, J, i = 1, 2, \cdots, n_J.
\end{aligned}
\tag{3.12}
$$

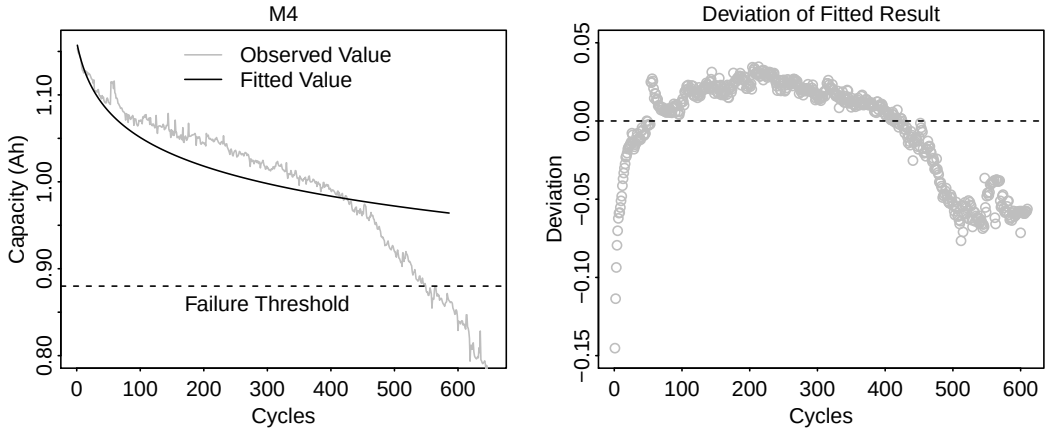

Figure 3.7: Fitted values of M4 compared to observed values

3. Capacity fade and model with logarithm of cycle and square root cycle as covariates

Model 5: Fixed effect intercept and slope for the square root cycle and logarithm cycle models:

$$\begin{aligned} y_{j[i]} &= \alpha + \beta_1 \times \sqrt{t_{j[i]}} + \beta_2 \times \log t_{j[i]} + \varepsilon_i \\ \varepsilon_i &\sim N(0, \sigma^2), j = 1, 2, \cdots, J, i = 1, 2, \cdots, n_J \end{aligned} \tag{3.13}$$

Model 6: Random effect intercept and slope mode for both the square root cycle and the logarithm cycle models:

$$\begin{aligned} y_{j[i]} &= \alpha_{j[i]} + \beta_{1j[i]} \times \sqrt{t_{j[i]}} + \beta_{2j[i]} \times \log t_{j[i]} + \varepsilon_i \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \beta_{1j} \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2), \beta_{2j} \sim N(\mu_{\beta_2}, \sigma_{\beta_2}^2) \\ (\alpha_j, \beta_{1j}, \beta_{2j}) &\sim MVN((\mu_\alpha, \mu_{\beta_1}, \mu_{\beta_2})^T, \Sigma), j = 1, 2, \cdots, J, i = 1, 2, \cdots, n_J. \end{aligned} \tag{3.14}$$

Model 7: Varying intercept and varying slope models for both the square root cycle and the logarithm cycle with an interaction term:

$$\begin{aligned} y_{j[i]} &= \alpha_{j[i]} + \beta_{1j[i]} \times \log t_{j[i]} + \beta_{2j[i]} \times \sqrt{t_{j[i]}} + \beta_{3j[i]} \times \sqrt{t_{j[i]}} \times \log t_{j[i]} + \varepsilon_i \\ \alpha_j &\sim N(\mu_\alpha, \sigma_\alpha^2), \beta_{1j} \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2), \beta_{2j} \sim N(\mu_{\beta_2}, \sigma_{\beta_2}^2), \beta_{3j} \sim N(\mu_{\beta_3}, \sigma_{\beta_3}^2) \\ \varepsilon_i &\sim N(0, \sigma^2), (\alpha_j, \beta_{1j}, \beta_{2j}, \beta_{3j}) \sim MVN((\mu_\alpha, \mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3})^T, \Sigma) \\ j &= 1, 2, \cdots, J, i = 1, 2, \cdots, n_J. \end{aligned} \tag{3.15}$$

where $y_{j[i]}$ is the ith observation of the jth unit, $t_{j[i]}$ is the measurement time, i.e., the number of cycles in this case. α, β, α_j, and β_j are the mixed effects coefficients.

3.2.4 Model selection using BIC

The optimal model from the seven candidate capacity fade models can be selected based on the Bayesian Information Criterion (BIC). BIC, as shown in Equation (3.7), is a log-likelihood-based metric for model selection that overcomes the over parameterization issue by penalizing the number of model parameters as well as the data size [93]. A smaller BIC value indicates a better model fit. The parameters of the random effects and BIC values of all the models have been estimated using *R* "*lme4*" package, as shown in Table 3.6. It was observed that all the models with random effects have smaller BIC values than those with

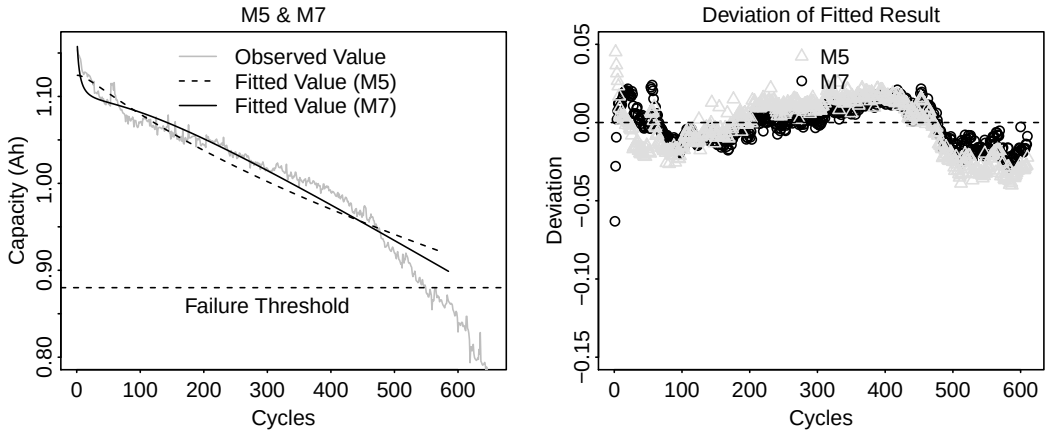

Figure 3.8: Fitted values of Models 5 and 7 compared to observed values

fixed effects, which indicates that between-battery variations cannot be ignored. Models with the square root of time can explain capacity fade better than those with logarithm time according to the comparison of Model 1 with Model 3 as well as Model 2 with Model 4. Model 7, which includes both the square root of time and logarithm time as well as their interaction effect, has the smallest BIC value of all seven models. As a result, Model 7 can be selected for battery capacity fade modeling and health prognostics. In addition, the mean and standard deviation of Model 7 are the smallest for the capacity estimate using the last 100 cycles of data, which indicates that Model 7 has the optimal fitting result (Table 3.6 and 3.7). The above model selection is based on the *CS2*-type battery data. Using the capacity fade data from the other type of *CX2* cells, Model 7 also shows a good fitting result, as seen from Figure 3.9. The deviations between the predicted results and observed values fall randomly around zero. Power fade is the other way of measuring Li-ion battery performance. Given that power fade shares similar causes with capacity fade, such as the formation of SEI, it is reasonable to apply the similar functional form of cycles used in capacity fade modeling for power fade modeling. Figure 3.10 shows the fitting results based on Model 7 for power fade prediction using the *CS2* battery data. The fitting results indicate that Model 7 is a good candidate for both capacity and power fade prediction.

Table 3.6: Parameter estimates for Model 1-7

Model	Model type	α	β_1	β_2	β_3	σ	BIC
M1	Fixed	1.182	-0.010	-	-	-	-9620.9
M2	Random	$N(1.18, 0.03^2)$	$N(-0.01, 0.001^2)$	-	-	0.021	-10508.2
M3	Fixed	1.302	-	-0.053	-	-	-8221.2
M4	Random	$N(1.299, 0.028^2)$	-	$N(-0.052, 0.002^2)$	-	0.032	-8663.9
M5	Fixed	1.106	-0.015	0.028	-	-	-9870.8
M6	Random	$N(1.107, 0.006^2)$	$N(-0.015, 0.003^2)$	$N(0.028, 0.013^2)$	-	0.019	-10952.3
M7	Random	$N(1.096, 0.017^2)$	$N(0.120, 0.025^2)$	$N(-1.107, 0.031^2)$	$N(-0.016, 0.003^2)$	0.015	-11967.2

Table 3.7: Mean and standard deviation for the deviations

Model	Average deviation of last 100 cycles	BIC
M1	-0.021	-9620.9
M2	-0.021	-10508.2
M3	-0.041	-8221.2
M4	-0.041	-8663.9
M5	-0.014	-9870.8
M6	-0.015	-10952.3
M7	-0.006	-11967.2

3.2.5 One-step-ahead capacity prognostics and end-of-life cycle distribution

With the selected capacity fade model (Model 7), in this section, we proceed with determining an optimal data input strategy for accurate one- step-ahead capacity fade prognostics. Two types of data input schemes are investigated, i.e., the cumulative data input and moving window data input. The prognostics performance of these two data input schemes was assessed using the mean absolute deviation (MAD) metric based on predicted capacity and actual observed capacity values, see Figure 3.11. We also investigated the end-of-life cycles to failure distributions, which can be further utilized for evaluating prediction uncertainty.

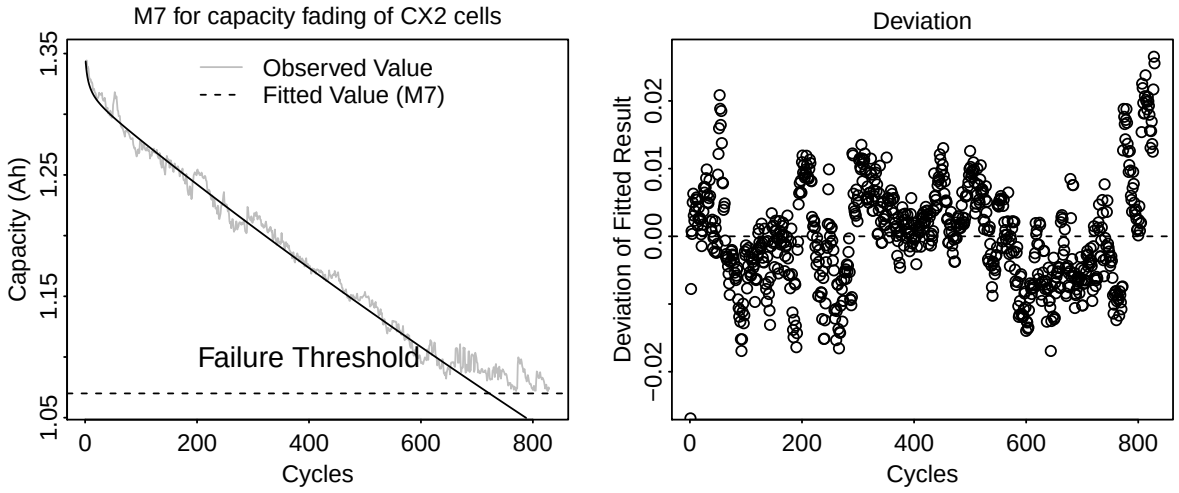

Figure 3.9: Model 7 for capacity fading of CX2 cells

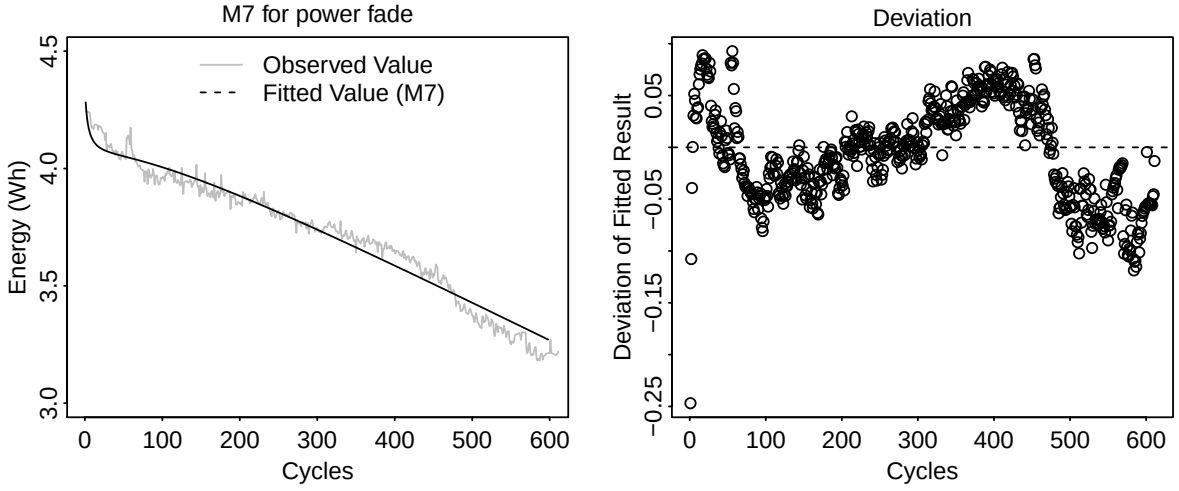

Figure 3.10: Model 7 for power fade of CS2 cells

3.2.5.1 One-step-ahead capacity fade prognostics

Data input strategies proposed to implement one-step-ahead capacity fade prognostics over the repetitive cycling process are described as follows. The first method uses all of the

cumulative data to update the capacity fade model and predict the capacity at a given future number of cycles. The cumulative data input method includes all historical information across all tested battery cells for future capacity prediction. However, the prediction of next cycle capacity using all previous capacity data may decrease the accuracy of the prediction due to obsolete data inputs. The second method, which uses the data from a given length of moving window, is studied, and its performance is compared with the cumulative data input capacity fade prediction. The moving window data input method has the risk of missing useful information by using a too short window length. Thus, determining the length of the moving window becomes a critical step when this data input strategy is used.

Deviations of model-fitted capacity and observed capacity using two types of data input methods are compared. In Figure 3.8, comparisons of moving window data input and cumulative data input are shown. As an example, a moving window with a length of five is used to demonstrate the basic logic of the moving window data input: the first five observed data form the first window to predict the capacity at the sixth cycle based on parameters estimated using the data in this window. Then, the five data points in the second window (the second to the sixth one) are used to predict the capacity at the seventh cycle. Similarly, the sth five data points (the sth to the $s+4$th data points) in the sth moving window can be used to predict the $s+5$th cycle's capacity, where $1 \leq s \leq n-5$, and n is the total number of observed data points. Differences between predicted and observed capacity values, as well as the mean of these differences, are calculated. Based on the same logic, set the window length as $l(5 \leq l \leq n-1)$, for the lth window,$l \leq s \leq n-l$, l of data points in this window are used to predict the $l+s$th cycle's capacity, and Δ_s is set as the absolute value of the difference between observation and prediction. The mean absolute deviation, MAD_l , of each window is averaged over all Δ_i. In summary, the MAD calculation can be done in the following four steps:

(1) Set the length of moving window l, $5 \leq l \leq n$, where n is total number of observed data points.

(2) Estimate the model parameters, α, β_1, β_2, and β_3, for the selected moving window length.

(3) Predict the $l+s$th cycle's capacity for the sth window; the difference between the predicted value and observed value δ_s is calculated, where $l \leq s \leq n-l$.

(4) Calculate the mean absolute deviation of the moving window l, MAD_l:

$$MAD_l = \frac{\sum |\Delta_s - \bar{\Delta}|}{n_l}, \Delta_s = y_{os} - y_{ps}, \bar{\Delta} = \frac{\sum_{s=1}^{n-l} \Delta_s}{n_l} \tag{3.16}$$

where $n_l, 1 \leq n_l \leq n-l$, is the number of moving windows for a window length l.

The deviation calculation of the cumulative data input method is relatively simpler. For consistency, MAD_l is set as the deviation for the $l, 5 \leq l \leq n-l$ the cycle's capacity prediction, which equals the difference between the prediction and observation. From Figure 3.11, it can be seen that the MAD_l of using the cumulative data input are all greater than those of using the moving window data input method. Thus, the moving window data input method should be used for battery fade modeling and health prognostics.

The two bottom panels in Figure 3.11 show that the moving window data input method generally performed better than the cumulative data input method for one-step ahead capacity prediction over the whole observed number of cycles. However, battery health prognostics near the end of life are usually more critical for failure prevention. Thus, we investigate the optimal moving window length for capacity prediction of the last 100 cycles using historical capacity fade data. As shown in Figure 3.12, the minimum MAD value for predicting the last 100 cycle capacity is achieved by choosing a moving window length of 195. To have a better understanding of the cycles to failure estimation near the end of the battery life, the moving window data input method with an optimal window length of 195 is used to update the parameters estimate for model 7 for the last 15 experimental cycles. The mixed effects coefficients are randomly generated based on their distribution. The end-of-life probability distributions are evaluated using the moving window data input method with an optimal window length of 195 for updating the random effect parameters in Model 7.

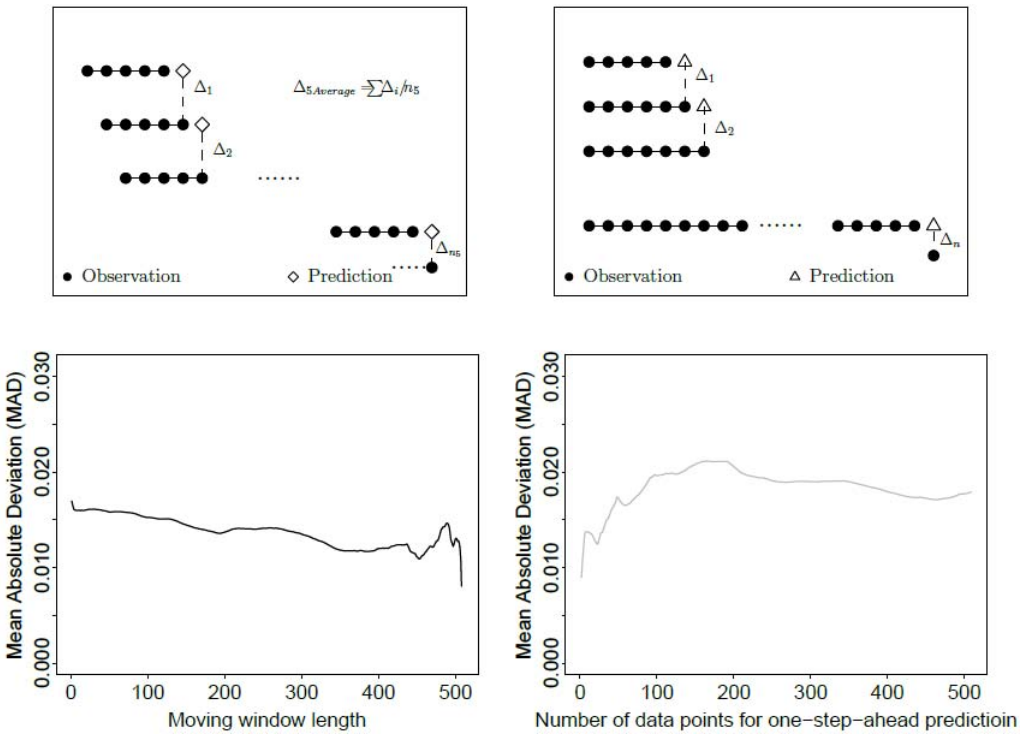

Figure 3.11: Two types of data input schemes for capacity prognostics

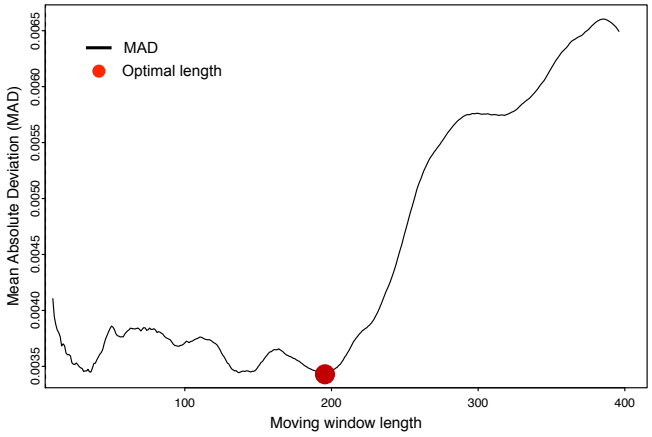

Figure 3.12: Determination of optimal moving window length

3.2.5.2 End-of-life cycles-to-failure distribution

In addition to the capacity prediction at a given number of cycles or a point estimate for the number of cycles at the end-of-life, i.e., 80% of the initial rated capacity, the cycles-

to-failure distribution near the end of battery life can provide further prediction uncertainty information. In this section, we first determine the appropriate probability distribution forms for the cycles-to-failure distribution using the best-fitted random effect model, then study the cycles-to-failure prediction uncertainty close to the end of life using the optimal moving window length, as discussed in the previous section. The random effect model (Model 7) is used to simulate the number of cycles to failure. As discussed in Section 3.2.3, the random effect parameters in Model 7 follow a multivariate normal distribution. The *mvnorm* function within the MASS package in *R* is used to simulate 10,000 samples of each of the four parameters. The cycles to failure can then be evaluated using the battery capacity failure threshold value of 0.88 Ah.

$$
\begin{aligned}
&(\alpha_j, \beta_{1j}, \beta_{2j}, \beta_{3j}) \sim MVN((\mu_\alpha, \mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3})^T, \Sigma) \\
&(\mu_\alpha, \mu_{\beta_1}, \mu_{\beta_2}, \mu_{\beta_3})^T = \begin{pmatrix} 1.097 \\ 0.120 \\ -0.107 \\ -0.016 \end{pmatrix} \\
&\Sigma = \begin{pmatrix} 8.26e-05 & -1.04e-05 & 1.02e-04 & 1.18e-05 \\ -1.04e-05 & 1.17e-04 & -1.65e-05 & -1.97e-05 \\ 1.02e-05 & -1.66e-05 & 2.50e-04 & 1.54e-06 \\ 1.18e-05 & -1.97e-05 & 1.54e-05 & 2.41e-05 \end{pmatrix}
\end{aligned}
\tag{3.17}
$$

A goodness-of-fit test for the simulated cycles to failures was conducted using two typical lifetime probability distributions, i.e., the Weibull distribution and Normal distribution. Under the normal goodness-of-fit testing, a value close to 0.05 indicates that the normal distribution is a good model for the cycle-to-failure distributions. The p-value for the Weibull goodness-of-fit test is much less than 0.05, indicating that the Weibull model is not an appropriate for modeling Li-ion batteries' cycles to failure. This result is consistent with existing literature findings based on physics-of-failure analysis. Figure 3.13 shows the cycles-to-failure distribution under both Normal and Weibull probability distributions.

To have a better understanding of the cycles to failure estimation near the end battery life, the moving window data input method with an optimal window length of 195 is used

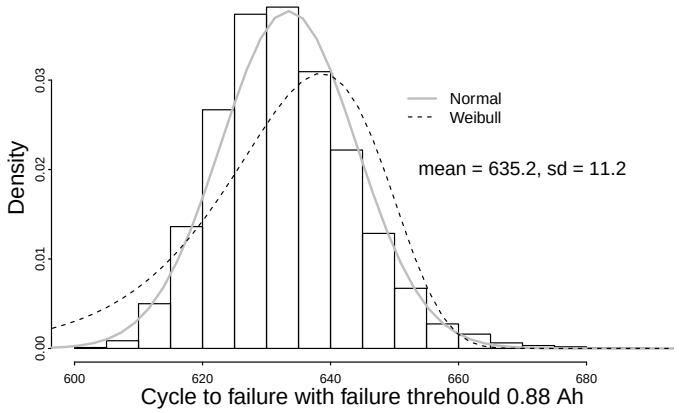

Figure 3.13: Histogram and density function of cycles to failure and fitted distributions

to update the parameters estimate for model 7 for the last 15 experimental cycles. For example, at cycle 500, data from cycle 305 and cycle 499 are used to estimate parameters: α, β_1, β_2, and β_3. Figure 3.14 shows the differences between the point estimate of capacity prediction results and the observed actual capacity for the last 15 charge/discharge cycles.

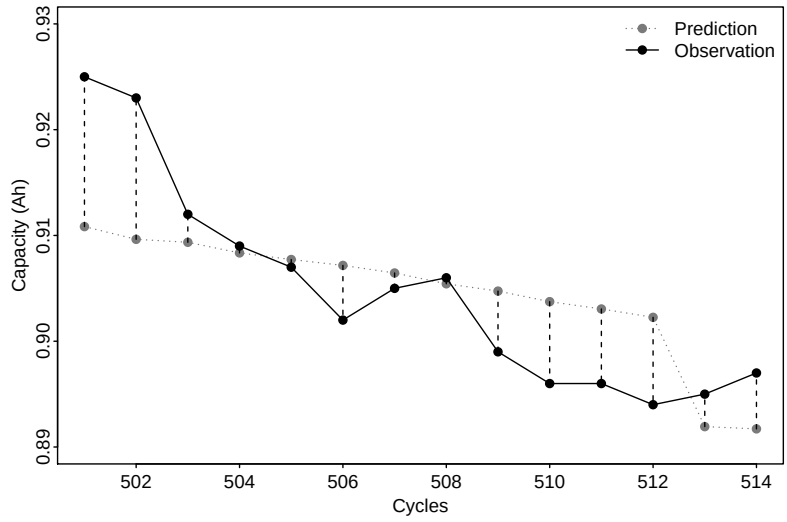

Figure 3.14: Differences between the actual and predicted capacity in late-stage

In addition, the end-of-life probability distributions are evaluated using the moving window data input method with an optimal window length of 195 for updating the random

effect parameters in Model 7. Figure 3.15 shows the fitted normal probability distributions for the end-of-life cycles to failure. The means of these normal distributions range from 547.8 to 551.3, and the standard deviations range from 8.9 to 10.4. With the more recent data inputs for parameter estimation, the mean of end-of-life prediction gets closer to 551 cycles from the initial 547 cycles.

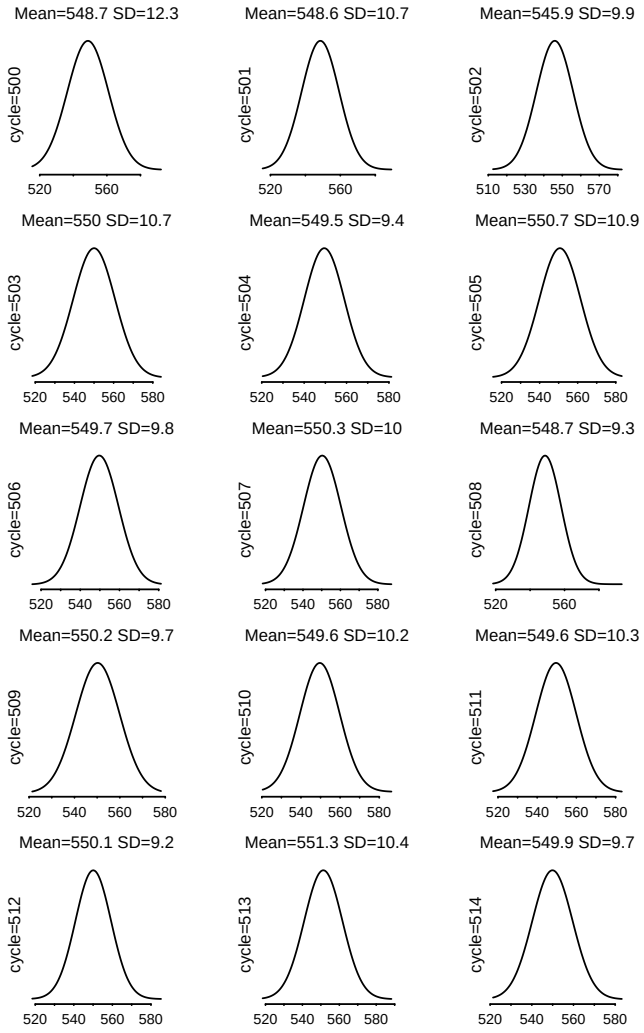

Figure 3.15: The distribution of cycles to failure in late stage

3.2.6 Conclusion and future work

Predicting variables were first identified using designed experiments and existing literature results. A series of Bayesian fixed and random effects models using these predicting variables was evaluated and compared using BIC for the best model selection. In addition to using BIC as a model selection metric, the absolute mean deviation was also assessed for model comparisons. The comparison results showed that the Bayesian random effect model with the square root of cycle, logarithm of cycle, and their interaction as predictors has the best fitting outcomes based on one type of battery data. The effectiveness of the selected model for capacity fade modeling and prediction is further verified using the other type of battery data. In addition to capacity fade modeling, power fade modeling is investigated under the proposed method. We observed that the capacity fade models are also appropriate for power fade modeling using the available testing data. To obtain better capacity fade and end-of-life cycles to failure prediction results, two data input methods for model parameter estimation and updating were investigated. We concluded that the moving data input method provided better prediction results for lithium-ion battery health prognostics. In this research, we also confirmed that the normal cycles to failure distribution is a better probability distribution model than the Weibull distribution model. Even though the random effect models perform better than the traditional fixed effect models with the between-battery variations better quantified, the residuals from the best fitted random effects model still show a slight pattern, which indicates that certain variations are not captured in the model. In our future research, we will explore more advanced random effect Bayesian models to capture such auto-correlation patterns. We also plan to test more batteries to increase sample size and include other stress factors in the modeling method. Moreover, the internal resistance, another main basic parameter determining the performance of lithium ion batteries, limits the battery specific power and energy efficiency. High internal resistance, especially in heavy loads such as power tools and electric powertrains, can cause the battery to heat up and then the voltage to drop under load, leading to early failure. Therefore, some degradation studies intend to investigate the battery

capacity degradation mechanism based on both capacity and internal resistance [151]. The assumption of balanced data might not be true in many cases. There are two difficulties in unbalanced data – one is to obtain the population average and the other one is to estimate parameter in GMM. In Chapter 4, a prognostics model based on joint modeling method is proposed to solve the aforementioned challenges.

4. PROGNOSTICS BASED ON JOINT MODELING METHOD

In the repeated measurements analysis, the mixed effects model can be simplified as Equation (4.1), where y_i represent the measurements of the ith unit, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are the covariates related to the fixed and random effects, respectively, and the errors are $\boldsymbol{e}_i$ following $N(0, \boldsymbol{\sigma}^2\boldsymbol{I})$. $\boldsymbol{\alpha}$ and $\boldsymbol{b}_i$ are the fixed and random effects parameters, respectively. Under the assumption of $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$, Equation (4.1) can be expressed in the form of a multivariate normal distribution, i.e., Equation (3.5). By setting $\boldsymbol{I}$ as the same over all units, the assumption of balanced data is made. With $\boldsymbol{I}_i$, different number of measurements of each unit, we can obtain

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{X}_i\boldsymbol{\alpha} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i \\ \boldsymbol{y}_i &\sim MVN(\boldsymbol{X}_i\boldsymbol{\alpha}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\sigma}_i^2\boldsymbol{I}_i) \\ \boldsymbol{y}_i &\sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \end{aligned} \tag{4.1}$$

Given $\boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, the mean $\boldsymbol{\mu}_i$, a $m_i \times 1$ vector, describes the general trend of the degradation process. The variance and correlation in the covariance matrix $\boldsymbol{\Sigma}_i$, a $m_i \times q$ matrix, indicate the the uncertainties and relationship between degradation levels at different time point. For a multivariate normal distribution, it is a natural and feasible idea to model the mean $\boldsymbol{\mu}_i$ and covariance $\boldsymbol{\Sigma}_i$ individualy [152]. The concept of joint modeling is introduced, where the mean and covariance are decomposed and then unknown parameters of the mean function and covariance matrix are estimated jointly. This research proposes a joint modeling method (JMM)-based prognostics. Without the assumption that all $\boldsymbol{I}_i$, $i = 1, \ldots, N$, are the same over all units, JMM can well address the issue of unbalanced data. In the proposed method, mean, variance, and correlation of $\boldsymbol{y}_i$ are firstly decomposed based on Cholesky decomposition. Trigonometric functions then are used to parameterize the correlation matrix so that the correlation matrix can be easily modeled with unconstrained parameters. To improve the prediction performance, covariates in JMM are also selected based on the physical understanding of the degradation process of interest. Pe-

nalized maximum likelihood estimation (PMLE) is proposed to estimate parameters of the mean, variance, and correlation function so that high-fidelity and parsimonious model can be obtained. For the purpose of comparison, the mean-covariance decomposition method is also investigated, where the parameters of mean and covariance function are estimated independently. The Gaussian process regression is used for prediction and the time to failure can be estimated through the numerical simulation. Being modeled as functions of time-varying factors, the mean function indicates the longitudinal trend and quantifies the temporal variation, while the variance function and correlation function illustrate the time-varying random effects between-unit. Figure 4.1 presents the basic procedure of prognostics based on joint modeling method.

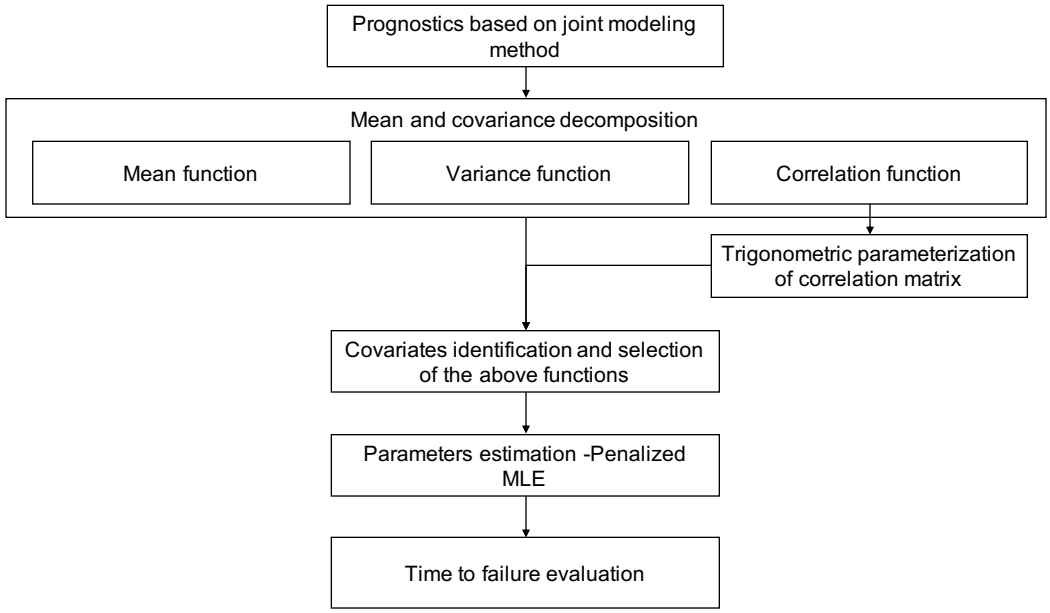

Figure 4.1: Schematic of prognostics based on joint modeling method

4.1 Mean-variance-correlation decomposition and representation

This section presents the decomposition of the mean and covariance matrix based on Chelosky decomposition.

4.1.1 Modeling mean and covariance matrix

Given $\boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, the concept of joint modeling method is introduced, where the mean and covariance are decomposed and then unknown parameters of the mean function and covariance matrix are estimated jointly. Being modeled as functions of time-varying factors, the mean function indicates the longitudinal trend and quantifies the temporal variation, while the variance and correlation functions illustrate the time-varying between-unit random effects.

The commonly used modeling approaches of the mean $\boldsymbol{\mu}_i$ are generalized regression models. For modeling the covariance matrix $\boldsymbol{\Sigma}_i$, there are two major challenges: high dimensionality and positive definiteness. Unconstrained parameterization methods are used to decompose the covariance matrix into unconstrained parameters to avoid high dimensionality and ensure the positive definiteness. Pinheiro et al. [153] summarized five unconstrained parameterizations of the covariance matrix, among which Cholesky decomposition methods attract the most interests due to its high computation efficiency and easy interpretation of entities of the decomposed matrix. The Cholesky decomposition-based methods attract the most interest due to its high computation efficiency and easy interpretation of entities of the decomposed matrix. Pourahmadi [154] proposed the joint modeling of mean-variance-correlation based on Cholesky decomposition of the covariance matrix. The entities of matrix $\boldsymbol{T}$ are interpreted as the auto-regression coefficients, where $\boldsymbol{\Sigma}_i = \boldsymbol{T}_i \boldsymbol{D}_i \boldsymbol{T}_i'$ and $\boldsymbol{D}_i = \text{diag}(\sigma_{i1}^2, \ldots, \sigma_{im_i}^2)$. Based on Pourahmadi's work, approaches to improve the interpretation of parameterization of covariance matrix with various forms of Cholesky decompositions are explored. Smith et al. [155] and Chen et al. [156] interpreted the entities of the decomposed matrix as one-step predictive coefficients and random effects coefficients, respectively. In the framework of Cholesky decomposition, covariance matrix of $\boldsymbol{y_i}$ can be decomposed as $\boldsymbol{\Sigma}_i = \boldsymbol{D}_i \boldsymbol{R}_i \boldsymbol{D}_i$, where $\boldsymbol{D}_i$ represents a diagonal matrix with entities of variances σ_{ij}, $1 \leq j \leq m_i$ and $\boldsymbol{R}_i$ with entities of $(\rho_{ijk})_{j,k=1}^{m_i}$ is the corresponding correlation matrix. Zhang et al. [157] characterized the entities of the decomposition matrix as moving average parameters. Most currently, the geometric interpretation based

on the trigonometric parameterization of the correlation matrix in [158] emerges due to its advantages in computation and model interpretation [159, 160]. Entities of the correlation matrix in trigonometric parameterization are represented as trigonometric functions which can be modeled as functions of time-varying factors. In the framework of Cholesky decomposition, a general joint model of the mean, variance, and correlation model can be summarized as Equation (4.2).

$$\begin{aligned} \boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \boldsymbol{\Sigma}_i = \boldsymbol{T}_i \boldsymbol{D}_i \boldsymbol{T}_i', \\ \boldsymbol{\mu}_i = m(\boldsymbol{x}_i, \boldsymbol{\alpha}\boldsymbol{\Sigma}_i), \log \boldsymbol{\sigma}_i^2 = v(\boldsymbol{z}_i, \boldsymbol{\beta}), \boldsymbol{\varphi}_{ijk} = d(\boldsymbol{w}_{\boldsymbol{ijk}}, \boldsymbol{\gamma}), \end{aligned} \tag{4.2}$$

where $\boldsymbol{\mu}_i$ and $\log \boldsymbol{\sigma}_i^2$ are mean and logarithm of variance of the observations of the ith unit, while $\boldsymbol{\varphi}_{\boldsymbol{ijk}}$ are entities of $\boldsymbol{T}_i$. $\boldsymbol{x}_i$, $\boldsymbol{z}_i$, and $\boldsymbol{w}_{ijk}$ represent the covariates, while $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ signify the parameters of the mean, variance, and correlation function, respectively. This research extends the application of the trigonometric parameterization in the correlation matrix and applies JMM in prognostics modeling.

4.1.2 Trigonometric parameterization of correlation matrix

Since the correlation matrix is positive definite and symmetric with off-diagonal entities between -1 and 1, $\boldsymbol{R}_i$ can be decomposed as $\boldsymbol{R}_i = \boldsymbol{L}_i \boldsymbol{L}_i'$, where $\boldsymbol{L}_i$ is the lower triangular decomposed matrix. The trigonometric parameterization is based on the Givens rotation as shown in Equation (4.3). $\mathbf{G}(i, j; \phi)$ is different from the m-dimensional identity matrix with $\mathbf{G}_{ii} = \cos(\phi)$, $\mathbf{G}_{i,j} = -\sin(\phi)$, $\mathbf{G}_{j,i} = \sin(\phi)$, and $\mathbf{G}_{j,j} = \cos(\phi)$. For a vector $\boldsymbol{l}$, $\mathbf{G}(i, j; \phi)\boldsymbol{l}$ rotates vector $\boldsymbol{l}$ anti-clockwise by the angle ϕ in $\{\mathbf{e}_i, \mathbf{e}_j\}$. For $\boldsymbol{l}_{ij}$, the jth column of $\boldsymbol{L}_i'$, there are $j-1$ rotations from $\boldsymbol{e}_1$. That is, $\boldsymbol{l}_{ij} = \prod_{k=1}^{j-1} \mathbf{G}(k, k+1; \phi_{jk})\boldsymbol{e}_1$. $\phi_{ijk} \in [0, \pi)$ is the angle between the $\mathbb{P}_k \boldsymbol{l}_{ij}$ and $\boldsymbol{e}_k$, where $\mathbb{P}_{ik} = \{\underbrace{0, \ldots, 0}_{k-1}, 1, \ldots, 1, \underbrace{0, \ldots, 0}_{m_i - j + 1}\}$ and $\boldsymbol{e}_k = \{\underbrace{0, \ldots, 0}_{k-1}, 1, \underbrace{0, \ldots, 0}_{m_i - k}\}$ and ϕ_{ijk} must make $\langle \boldsymbol{l}_{ij}, \boldsymbol{l}_{ik} \rangle = \rho_{ijk}$. Since the first column of $\boldsymbol{L}_i'$ is $\{1, \underbrace{0, \ldots, 0}_{m_i - 1}\}$, the other elements of $\boldsymbol{L}_i$ can be parameterized as functions of angles. For example, given $\langle \boldsymbol{l}_{i1}, \boldsymbol{l}_{i2} \rangle = \cos \phi_{i21} = \rho_{i21}$, $\boldsymbol{l}_{i2} = \{\cos \phi_{i21}, \sin \phi_{i21}, \underbrace{0, \ldots, 0}_{m_i - 2}\}$ with

$\phi_{i21} = \arccos \rho_{i21}$. And given $\langle \boldsymbol{l}_{i1}, \boldsymbol{l}_{i3} \rangle = \rho_{31}$ and $\langle \boldsymbol{l}_{i2}, \boldsymbol{l}_{i3} \rangle = \rho_{32}$, $\boldsymbol{l}_{i3} = \{\cos\phi_{i31}, \cos\phi_{i32}\sin\phi_{i21}, \sin\phi_{i32}\sin\phi_{i31}, \underbrace{0, \ldots, 0}_{m_i-3}\}$ with $\phi_{i31} = \arccos\rho_{i31}$ and $\phi_{i32} = \arccos\left[\frac{\rho_{i32} - \rho_{i21}\rho_{i31}}{\sin(\arccos\rho_{i21})\sin(\arccos\rho_{i31})}\right]$. With the same logic, we can represent elements of $\boldsymbol{L}_i$ using trigonometric functions, as shown in Equation (4.4). With $\rho_{ijk} = \sum_{l=1}^{k} l_{ijl} l_{ikl}$, the relationship between correlation coefficients and angles can also be obtained, as shown in Equation (4.5).

$$\mathbf{G}(i,j;\phi) = \begin{Bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cos(\phi) & \cdots & -\sin(\phi) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sin(\phi) & \cdots & \cos(\phi) & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{Bmatrix}. \tag{4.3}$$

$$\begin{aligned} & l_{i11} = 1; l_{ij1} = \cos(\phi_{ij1}), 1 \le j \le m_i; \\ & l_{ijk} = \begin{cases} \cos(\phi_{ijk}) \prod_{l=1}^{k-1} \sin(\phi_{ijl}), & 2 \le k < j \le m_i; \\ \prod_{l=1}^{k-1} \sin(\phi_{ijl}), & k = j; j = 2, \cdots, m_i. \end{cases} \end{aligned} \tag{4.4}$$

$$\begin{aligned} \rho_{ij1} &= \cos\phi_{ij1}, \rho_{ijj} = 1 \\ \rho_{ijk} &= \sum_{l=1}^{k-1} [\cos(\phi_{ijl})\cos(\phi_{ikl}) \prod_{t=1}^{l-1} \sin(\phi_{ijt})\sin(\phi_{ikt})] \\ &\quad + \cos(\phi_{ijk}) \prod_{l=1}^{k-1} \sin(\phi_{ijl})\sin(\phi_{ikl}), 2 \le k < j \le m_i \\ \phi_{ijk} &= \arccos\left[l_{ijk} / \prod_{l=1}^{k-1} \sin(\phi_{ijl})\right], 1 \le k < j \le m_i \end{aligned} \tag{4.5}$$

Equaiton (4.5) maps from a general correlation matrix $\boldsymbol{R}_i$ to the angles ϕ_{ijk}. Angles ϕ_{ijk} are unconstrained in the range of $[0, \pi)$ and can be easily driven by the time-varying factors. Then the entities of the correlation matrix can be modeled as a function of time-varying

factors. Modeling ϕ_{ijk} can also reduce the number of time-varying factors when the dimensionality of $\boldsymbol{y}_i$ increases [159]. For a degradation process, we can have a joint model of mean-variance-correlation can be rewritten as Equation (4.6).

$$\boldsymbol{\mu}_i = m(\boldsymbol{X}_i, \boldsymbol{\alpha}), \log \boldsymbol{\sigma}_i^2 = v(\boldsymbol{Z}_i, \boldsymbol{\beta}), \phi_{ijk} = d(\boldsymbol{W}_{ijk}, \boldsymbol{\gamma}) \tag{4.6}$$

where $\boldsymbol{\mu}_i$, $\log \boldsymbol{\sigma}_i^2$, and ϕ_{ijk} are the mean, log-variance, and angles. Functions $m(\cdot)$, $v(\cdot)$, and $d(\cdot)$ are the mean, log-variance, and angles functions. Specific cases of linear models of $m(\cdot)$, $v(\cdot)$, and $d(\cdot)$, such as polynomial models, are practical interest. Covariates $\boldsymbol{X}_i$, $\boldsymbol{Z}_i$, and $\boldsymbol{W}_{ijk}$ can be determined according to the characteristics of the degradation process of interest. $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are the unknown parameters.

4.1.3 Covariate identification and selection

The selection of covariates in the joint model mainly depends on the physical understanding of the degradation process. In Equation (4.6), $\boldsymbol{x}_i$ are the covariates that are used to model the mean of the measurements. The selection of $\boldsymbol{x}_i$ might mainly depend on the analytic model of the degradation process, i.e., the physics-based model. For example, in the battery capacity degradation, it is a natural idea that the covariates could be selected from the factors affecting SEI formation during the electro-chemical reaction, such as the number of cycles, charge rate and discharge rate, temperature, etc. These factors are also good candidates of $\boldsymbol{z}_i$, the covariates of the logarithm square variance of the jth observations. The correlation between measurements $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ decay over the time difference between the measurements. Thus $\boldsymbol{w}_{ijk}$ should include time varying covariates that may depend on measurement time t_{ik} and t_{ij} as it is capturing the correlation between the responses at these two measurement time.

4.1.4 Mean-covariance decomposition-based method

Based on the mean, log-variance, and angles in Equation (4.6), the mean-covariance decomposition (MCD)-based method is investigated, where the three functions are mod-

eled and their parameters are estimated independently. The motivation of the MCD-based method is to explore the characateristics of the mean, log-variance, and angles. Results of the MCD-based method can be used as the initial solution of the joint modeling method.

4.1.4.1 Moving-window scheme

Considering of large scale of variance and angles for correlations, the concept of moving window is introduced to improve the prediction accuracy and computational complexity. It is evident that the strength of the "impact" that one measurement has on another measurement will depend on average decay as a function of their corresponding time lag in the time sequence. For a relatively slow degradation process, such as lithium ion battery capacity fading, the previous observations have a weak effect on the current observation and might distort the prediction. The moving-window method is to use information from the most recent observations to predict the future values. Models based on a large window provide less accurate information along the perturbation direction, while those based on a small window are effective to obtain the pertinent information of perturbation direction. However, the small moving-window might not be robust with high noise level [161]. The optimal length of a moving window in this paper is determined using the prognostics performance in the term of the mean absolute deviation (*MAD*) based on one-step-ahead predicted value and actual observed value. The core of calculating *MAD* of one moving window is to use a constant size of historical data to predict the next one value and obtain the residual. The details of the four-steps of determining moving window length can be found in [59]. Moreover, the moving-window scheme with the constant length simplifies the modeling of the correlation which is the function of the difference between the measured time being the same for a certain length of the moving window. For each moving window, parameters in Equation (4.4) are estimated through balancing the measurement fit and the model complexity. The basic form of functions of mean, log-variance, and angles for correlations can be obtained from the physical understanding and empirical functions. For example, the chemical analytic model from [93, 142] can be referred to as the basic

model of mean, which is discussed in Chapter 3. Due to the lack of physical understanding of log-variance and correlation of the capacity degradation process, the log-variance $\log \boldsymbol{\sigma}_i^2$ and angles ϕ_{ijk} are estimated through non-parametric methods. This research uses the support vector machine (SVM) regression with ϵ-insensitive loss function in the consideration of its advantages in efficiency, accuracy, and robustness.

4.1.4.2 Log-variance and correlation modeling using SVM

SVMs were developed to solve the classification and regression problem through support vector methods. The kernel function is a powerful tool to avoid the computation of high dimensional inner product. As a kernel based method, SVM can project the original low dimension data space to high dimension feature space [162, 163]. Support vectors are trained in SVM regression so that the data points lie in between the two borders of the margin which is maximized under suitable conditions to avoid outliers inclusion. SVM regression is formulated as controlling the model complexity and margins through the squared norm of the parameters vector and loss function respectively. Take a general regression model $y = \boldsymbol{w}^T g(\boldsymbol{x}) + w_0$ for example. In this model, $g(\boldsymbol{x})$ is a function mapping each input $\boldsymbol{x}$ to a higher dimensional space, $\boldsymbol{w} = (w_1, w_2, \ldots, w_m)^T$ denotes a set of linear weights connecting the feature space $g(\boldsymbol{x})$ to the output y, and w_0 is the constant. Parameters $\boldsymbol{w}$ and w_0 can be estimated through minimizing $_{\boldsymbol{w},w_0,\boldsymbol{\xi},\boldsymbol{\xi}^*} \frac{1}{2} \parallel \boldsymbol{w} \parallel^2 + C \sum_{i=1}^{m} (\xi_i + \xi_i^*)$ with constraints of $\boldsymbol{w}^T g(x_i) + w_0 - y_i \leq \epsilon + \xi_i, y_i - \boldsymbol{w}^T g(x_i) - w_0 \leq \epsilon + \xi_i^*, \xi_i, \xi_i^* \geq 0$, where slack variables ξ_i and ξ_i^* penalize predictions out of the ϵ-intensive tube, and the penalty parameter $C > 0$ determines the trade off between the flatness of the function and the amount up to which deviations larger than ϵ are tolerated.

The above optimization problem can be transformed into the dual problem as $\sum_{i=1}^{n_{sv}} (a_i - a_i^*) g(x_i)^T g(x) + w_0$ with the constraint of $0 \leq a_i, a_i^* \leq C$ where $a_i, a_i^*, i = 1, \ldots, n_{SV}$ are the Lagrange multipliers and n_{sv} is the number of support vectors. The inner product $g(x_i)^T g(x)$ can be integrated as a kernel function $K(\cdot)$. Karatzoglou et al. [164] summarized the methods, such as sequential minimization optimization (SMO), chunking, and

simple SVM, which are used to solve the quadratic optimization problem. Chang and Lin [165] developed the library ***libsvm*** providing a very efficient SVM implementation based on SMO, which is used in this paper. Moreover, in the aforementioned soft margin optimization problem, there are three undetermined parameters, i.e., the kernel function $\langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$, the penalty parameter C, and ϵ. The choice of the kernel function depends on the characteristics of the data. The Gaussian kernel $k(\mu, \nu) = \exp(\gamma \parallel \mu - \nu \parallel^2)$ and polynomial kernel $k(\mu, \nu) = (\alpha \mu^T \nu + c)^d$ are often used for general purposes. The parameters for Gaussian kernel parameter γ or that for polynomial kernel α and c are given for the optimization problem. Usually the penalty C and γ are given on the intervals $[2^{-5}, 2^{15}]$ and $[2^{-15}, 2^3]$. Choosing the "best" parameters is needed for the selection of the best model. Chang and Lin [165] proposed a grid-search method in the *tune* function of ***libsvm***, where various pairs of these given parameters tested over cross-validation randomized samples. However, the grid search is very inefficient when the large scale data set is used as input or small grid is required. This paper uses random search where multiple random sampled candidates of C and γ are used for cross-validation based on the ***R*** function of SVM [166]. One of the major advantages of random search is to decrease the computation time so that a large range of C and γ can be tested [167]. To model log-variances and angles in MCD, their covariates, such as the number of cycles and cycle lags, are considered as inputs of the kernel functions.

4.2 Joint modeling method

The aforementioned MCD method models the mean, log-variance, and angle functions and estimate their parameters independently. Results of MCD method can be used as the initial understanding the longitudinal trends and uncertainties in the degradation process. However, With the assumption that $\boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, maximum likelihood estimation (MLE) is one of the most effective candidate tools to estimate parameters in the joint model. Given lack of analytic models of log-variance and angles for the corrlation, this research

proposes a penalized MLE where a penalty function is added to the maximum likelihood function to obtain the high-fidelity and parsimonious models.

4.2.1 Penalized maximum likelihood estimation

With the assumption that $\boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, penalized maximum likelihood estimation (PMLE) is one of the most effective candidate tools to estimate parameters in the joint model. Due to lack of analytic models of log-variance and angles, this paper employs PLME where a penalty function is added to the maximum likelihood estimate to determine parameters and select covariates simultaneously. For irreversible degradation processes, each unit degrades variously. Therefore, there is only one observation for the multivariate normal distribution of each degradation trajectory. Then the minus twice log-likelihood function without the constant is

$$-2\log l(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log|\boldsymbol{\Sigma}_i| + \sum_{i=1}^{n} (\mathbf{y}_i - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i), \tag{4.7}$$

where $\boldsymbol{\theta}$ is the parameter vector of the mean, log-variance, and angle functions. $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})^T$ is a $m_i \times 1$ vector. $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean and covariance matrix of multivariate normal distributions for the tested units.

There exists an unique lower triangular matrix $\boldsymbol{Q}_i$ satisfying $\boldsymbol{\Sigma}_i = \boldsymbol{Q}_i^{-1} \boldsymbol{D}_i^2 \boldsymbol{Q}_i'^{-1}$. $|\boldsymbol{\Sigma}_i| = |\boldsymbol{D}_i^2| = \sum_{j=1}^{m_i} \log \sigma_{ij}^2$ can be obtained.

$$\log|\boldsymbol{\Sigma}_i| = \sum_{j=1}^{m_i} \log \sigma_{ij}^2 \tag{4.8}$$

By $\boldsymbol{\Sigma}_i = \boldsymbol{D}_i \boldsymbol{L}_i \boldsymbol{L}_i' \boldsymbol{D}_i$, $\boldsymbol{\Sigma}_i^{-1} = \boldsymbol{D}_i^{-1} \boldsymbol{L}_i'^{-1} \boldsymbol{L}_i^{-1} \boldsymbol{D}_i^{-1}$,

$$\begin{aligned} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha}) &= [\boldsymbol{L}_i^{-1} \boldsymbol{D}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})]' [\boldsymbol{L}_i^{-1} \boldsymbol{D}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\alpha})] \\ &= \sum_{j=1}^{m_i} \Delta_{ij}^2 \end{aligned} \tag{4.9}$$

where Δ_{ij} is the jth element of $\boldsymbol{\Delta}_i$ with $\boldsymbol{\Delta}_i = \boldsymbol{L}_i^{-1}\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$. As $\boldsymbol{\Delta}_i = \boldsymbol{L}_i^{-1}\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$, $\Delta_{ij} = \sum_{k=1}^{j} \frac{v_{ijk}}{\sigma_{ij}}(y_{ij} - \mu_{ij})$, where v_{ijk} is the (j,k)th element of $\boldsymbol{L}_i^{-1}$. Based on Equation (4.8) and (4.9), the twice minus log-likelihood estimate can be written as

$$-2l(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}[\log \sigma_{ij}^2 + \Delta_{ij}^2] \tag{4.10}$$

By minimizing Equation (4.7), parameters can maximize the likelihood function given the sample of observations and provide a parsimonious model. However, computation infeasibility and instability come along with the maximum likelihood estimation [168,169]. A penalty function $p(\boldsymbol{\theta})$ can be added to Equation (4.10) to select the significant covariates and estimate coefficients,

$$\arg\min_{\boldsymbol{\theta}} PL(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}[\log \sigma_{ij}^2 + \Delta_{ij}^2] + \lambda p(\boldsymbol{\theta}), \tag{4.11}$$

where λ is the tuning parameter. The penalty function $p(\boldsymbol{\theta})$ is selected to reduce the loss in the estimation. Given the proposed joint model, penalties on parameters of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ can be used to reduce the model error and complexity. The LASSO method, $p(\boldsymbol{\theta}) = \| \boldsymbol{\theta} \|_1$, is used in this paper due to its advantages in the case of highly noisy observations. The penalty differential of $p(\boldsymbol{\theta})$ is

$$D_\lambda = \begin{cases} -\lambda & \text{if } \theta < 0; \\ [-\lambda, \lambda] & \theta = 0; \\ \lambda & \theta > 0. \end{cases} . \tag{4.12}$$

To determine the tuning parameter λ, the Bayesian information criterion (BIC) is used. For a given λ,

$$BIC(\lambda) = N\log[\boldsymbol{\sigma}^2(\lambda)] + k(\lambda)\log(N), \tag{4.13}$$

where $\boldsymbol{\sigma}^2(\lambda) = N^{-1}[\boldsymbol{y} - \hat{\boldsymbol{y}}(\lambda)]^T\boldsymbol{\Sigma}(\lambda)^{-1}(\boldsymbol{y} - \hat{\boldsymbol{y}})$. $k(\lambda)$ is the number of nonzero regression coefficients. The λ that minimizes $BIC(\lambda)$ is selected.

4.2.2 Score functions and Fisher matrices

A Newton-Raphson method is used to obtain the penalized maximum likelihood (PMLE). The relevant score functions and Fisher information matrices can be obtained.

$$
\begin{aligned}
S_1(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\alpha}) + D_\lambda, \\
S_2(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{i=1}^{n} \boldsymbol{Z}_i' (\mathbf{1}_{m_i} - \boldsymbol{T}_i) + D_\lambda \\
S_3(\boldsymbol{\theta}) &= -2 \sum_{i=1}^{n} \sum_{j=1}^{m_i} \Big[\frac{\partial \log l_{ijj}}{\partial \boldsymbol{\gamma}} \Delta_{ij}^2 + \Delta_{ij} \sum_{k=1}^{j-1} \boldsymbol{u}_{ijk} \Delta_{ik} \Big] + D_\lambda \\
\boldsymbol{I}_{11} &= \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i, \boldsymbol{I}_{22} = \frac{1}{4} \sum_{i=1}^{n} \boldsymbol{Z}_i' [\boldsymbol{I}_{m_i} + \boldsymbol{R}_i^{-1} \circ \boldsymbol{R}_i] \boldsymbol{Z}_i \\
\boldsymbol{I}_{23} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \Big(\frac{\partial \log l_{ijj}}{\partial \boldsymbol{\gamma}} + \frac{1}{2} \sum_{k=1}^{j-1} \boldsymbol{u}_{ijk} \sum_{l=k}^{j} l_{ilk} v_{ijk} \boldsymbol{z}_{il}' \Big), \\
\boldsymbol{I}_{33} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \Big(4 \frac{\partial \log l_{ijj}}{\partial \boldsymbol{\gamma}} \frac{\partial \log l_{ijj}}{\partial \boldsymbol{\gamma}'} + \sum_{k=1}^{j-1} \boldsymbol{u}_{ijk} \boldsymbol{u}_{ijk}' \Big) \\
\boldsymbol{I}_{12} &= \boldsymbol{I}_{21}^T = 0, \boldsymbol{I}_{13} = \boldsymbol{I}_{31}^T = 0
\end{aligned}
\tag{4.14}
$$

where $\boldsymbol{T}_i = diag\{\boldsymbol{R}_i^{-1} \boldsymbol{D}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)' \boldsymbol{D}_i^{-1}\}$, $\mathbf{1}_{m_i}$ is a vector with elements of 1, $\boldsymbol{u}_{ijk} = \sum_{l=k}^{j} \frac{\partial l_{ijk}}{\partial \boldsymbol{\gamma}} v_{ijl}$, and v_{ijl} is the (j,l) element of $\boldsymbol{L}_i^{-1}$.

The derivation of the score and Fisher matrices is as follows.

$S_1(\boldsymbol{\theta})$ and $\boldsymbol{I}_{11}$ can be easily obtained.

$$
\begin{aligned}
S_1(\boldsymbol{\theta}) &= \frac{\partial PL(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} = \boldsymbol{X}_i' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\alpha}) + D_\lambda \\
\boldsymbol{I}_{11} &= -E\Big(\frac{\partial S_1(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}'} \Big) = \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i
\end{aligned}
\tag{4.15}
$$

By $\log \sigma_{ij}^2 = \boldsymbol{z}_{ij} \boldsymbol{\beta}$, the score is

$$
\begin{aligned}
S_2(\boldsymbol{\theta}) &= \frac{\partial PL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (\boldsymbol{z}_{ij} + 2 \frac{\partial \Delta_{ij}}{\partial \boldsymbol{\beta}} \Delta_{ij}) + D_\lambda \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m_i} [\boldsymbol{z}_{ij} - \sum_{k=1}^{j} \frac{v_{ijk}}{\sigma_{ik}} (y_{ik} - \mu_{ik}) \boldsymbol{z}_{ik} \sum_{k=1}^{j} \frac{v_{ijk}}{\sigma_{ik}} (y_{ik} - \mu_{ik})] + D_\lambda \\
&= \sum_{i=1}^{n} \boldsymbol{Z}_i' (\mathbf{1}_{m_i} - \boldsymbol{T}_i)
\end{aligned}
\tag{4.16}
$$

where $\boldsymbol{T}_i = diag\{\boldsymbol{R}_i^{-1}\boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'\boldsymbol{D}_i^{-1}\}$ and $\mathbf{1}_{m_i}$ is a vector with elements of 1.

$$\begin{aligned}\boldsymbol{I}_{22} &= -E\Big(\frac{\partial S_2(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}'}\Big) \\ &= \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{m_i}[\sum_{k=1}^{j}\frac{v_{ijk}}{\sigma_{ik}}(y_{ik}-\mu_{ik})\boldsymbol{z}_{ik}\frac{v_{ijk}}{\sigma_{ik}}(y_{ik}-\mu_{ik})\boldsymbol{z}_{ik}\sum_{k=1}^{j}\frac{v_{ijk}}{\sigma_{ik}}(y_{ik}-\mu_{ik})\boldsymbol{z}_{ik}] \\ &= \frac{1}{4}\sum_{i=1}^{n}\boldsymbol{Z}_i'[\boldsymbol{I}_{m_i} + \boldsymbol{R}_i^{-1}\circ\boldsymbol{R}_i]\boldsymbol{Z}_i\end{aligned} \tag{4.17}$$

$$\boldsymbol{I}_{23} = -E\Big(\frac{\partial S_2(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma}'}\Big) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big[\frac{\partial \log l_{ijj}}{\partial \boldsymbol{\gamma}}\boldsymbol{z}_{ij}' + \frac{1}{2}\sum_{k=1}^{j-1}\boldsymbol{u}_{ijk}\sum_{k=1}^{j}l_{ilk}v_{ijk}\boldsymbol{z}_{il}'\Big] \tag{4.18}$$

By $\boldsymbol{L}_i\boldsymbol{\Delta}_i = \boldsymbol{D}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)$, $\Delta_{ij} = \frac{1}{l_{ijj}}\Big(\frac{y_{ij}-\mu_{ij}}{\sigma_{ij}} - \sum_{k=1}^{j-1}l_{ijk}\Delta_{ik}\Big)$. Therefore,

$$\begin{aligned}S_3(\boldsymbol{\theta}) &= -\sum_{i=1}^{n}\sum_{j=1}^{m_i}2\Delta_{ij}\frac{\partial\Delta_{ij}}{\partial\boldsymbol{\gamma}} + D_\lambda \\ &= -2\sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big[\frac{\partial\log l_{ijj}}{\partial\boldsymbol{\gamma}}\Delta_{ij}^2 + \Delta_{ij}\sum_{k=1}^{j-1}\boldsymbol{u}_{ijk}\Delta_{ik}\Big] + D_\lambda\end{aligned} \tag{4.19}$$

With $\frac{\partial\Delta_{ij}}{\partial\boldsymbol{\gamma}} = -\frac{\partial\log l_{ijj}}{\partial\boldsymbol{\gamma}}\Delta_{ij} - \sum_{k=1}^{j-1}\boldsymbol{u}_{ijk}\Delta_{ik}$,

$$\begin{aligned}\boldsymbol{I}_{33} &= -\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\Big(\frac{\partial^2\log l_{ijj}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} + \Delta_{ij}\frac{\partial^2\Delta_{ij}}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} + \frac{\partial\Delta_{ij}}{\partial\boldsymbol{\gamma}}\frac{\partial\Delta_{ij}}{\partial\boldsymbol{\gamma}'}\Big) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m_i}\Big(4\frac{\partial\log l_{ijj}}{\partial\boldsymbol{\gamma}}\frac{\partial\log l_{ijj}}{\partial\boldsymbol{\gamma}'} + \sum_{k=1}^{j-1}\boldsymbol{u}_{ijk}\boldsymbol{u}_{ijk}'\Big)\end{aligned} \tag{4.20}$$

By Equation (4.5),

$$\frac{\partial l_{ijk}}{\partial\boldsymbol{\gamma}} = \begin{cases} l_{ijk}\sum_{l=1}^{k-1}W_{ijl}/\tan\phi_{ijl}, & k = j; \\ -l_{ijk}w_{ijk}\tan\phi_{ijk} + \sum_{l=1}^{k-1}\frac{W_{ijl}}{\tan\phi_{ijl}}, & k < j \end{cases}. \tag{4.21}$$

For a fixed λ, an iterative Fisher's scoring algorithm is as follows:

1. Select initial value $\hat{\boldsymbol{\alpha}}^{(0)}$, $\hat{\boldsymbol{\beta}}^{(0)}$, and $\hat{\boldsymbol{\gamma}}^{(0)}$ and set $k = 0$. The convenient initial values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are their LASSO estimates using the balanced data

2. Update the current $\hat{\boldsymbol{\beta}}^{(k)}$ and $\hat{\boldsymbol{\gamma}}^{(k)}$ using

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k+1)} \\ \hat{\boldsymbol{\gamma}}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k)} \\ \hat{\boldsymbol{\gamma}}^{(k)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{I}_{22}(\hat{\boldsymbol{\theta}}^{(k)}) & \boldsymbol{I}_{23}(\hat{\boldsymbol{\theta}}^{(k)}) \\ \boldsymbol{I}_{32}(\hat{\boldsymbol{\theta}}^{(k)}) & \boldsymbol{I}_{33}(\hat{\boldsymbol{\theta}}^{(k)}) \end{pmatrix}^{-1} \begin{pmatrix} S_2(\hat{\boldsymbol{\theta}}^{(k)}) \\ S_3(\hat{\boldsymbol{\theta}}^{(k)}) \end{pmatrix}$$

3. Calculate $\hat{\boldsymbol{\Sigma}}^{(k)}$ based on Equation (4.5)

4. Update the current $\hat{\boldsymbol{\alpha}}^{(k)}$ using

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \hat{\boldsymbol{\alpha}}^{(k)} + \boldsymbol{I}_{11}(\hat{\boldsymbol{\theta}})^{-1} S_1(\hat{\boldsymbol{\theta}}^{(k)})$$

5. Repeat steps 2-4 until the difference between $PL(\hat{\boldsymbol{\theta}}^{(k+1)})$ and $PL(\hat{\boldsymbol{\theta}}^{(k)})$ is no greater than ϵ

The initial values of parameters, $\hat{\boldsymbol{\alpha}}^{(\mathbf{0})}$, $\hat{\boldsymbol{\beta}}^{(\mathbf{0})}$ and $\hat{\boldsymbol{\gamma}}^{(\mathbf{0})}$, determine the computation efficiency of a Newton-Raphson algorithm. A randomly given initial value could significantly slow the search of the "optimal" result. Inspired by the decomposition of mean, variance, and correlation, this research estimates $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ independently. That is, mean, variance, and angle functions are assumed to be independent from one another and their parameters are estimated separately.

4.3 Prediction and time to failure evaluation

Assume a prediction y^* at a new input t^*. The degradation process can be evaluated at $y^* = \mu(t^*) + S(t^*)$ where $\mu(t^*)$ is given based on the degradation trend and $S(t^*)$ is the covariance of the input t^*. It is reasonable to assume that $y^* \sim N(\mu^*, \Sigma^*)$. For a joint distribution, the simple kriging predictor is a good choice for unbiased linear predictors. Assuming X and Y are two random variables with their Gaussian joint distribution of

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix} \right)$$

then,

$$Y|X \sim N(\mu_Y + C_{YX}C_{XX}^{-1}(X - \mu_X), C_{YY} - C_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}). \tag{4.22}$$

For a degradation process, the joint distribution of $S = (s_1(T), s_2(T), \ldots, s_M(T))'$ and the prediction $S^* = (s^*(t))$ can be written as

$$\begin{pmatrix} S \\ S^* \end{pmatrix} = N\left(0, \begin{pmatrix} C(T,T) & C(T,t^*) \\ C(t^*,T) & C(t^*,t^*) \end{pmatrix}\right) \tag{4.23}$$

where $C(T, t^*)$ is the covariance matrix between the model output $\hat{y}$ and y^*, $C(T, T)$ and $C(t^*, t^*)$ are the covariance of y and y^*, respectively. The conditional distribution of S^* given S is

$$S^*|S \sim N(C(t^*,T)C(T,T)^{-1}S, C(t^*,t^*) - C(t^*,T)C(T,T)^{-1}C(T,t^*)) \tag{4.24}$$

Therefore, prediction y^* has the mean and variance

$$\begin{aligned} E(y^*) &= \mu(t^*) + H(y - \mu(t)) \\ \sigma^* &= C(t^*,t^*) - HC(T,T)^{-1}H', \end{aligned} \tag{4.25}$$

where $H = C(t^*, T)C(T, T)^{-1}$.

With the assumption of independence of multiple objects, the aforementioned prediction can be repeated N times, where $E(y_{(m)^*})$ and $Var(y^*_{(m)})$ can be obtained. The case of unbalanced data can be well addressed because the conditional distribution of predictions given the measurements of objects are estimated independently. Based on the above $E(y_{(m)^*})$ and $Var(y^*_{(m)})$, the prediction for the response associated with a input t^* is given as

$$\begin{aligned} \hat{y^*} &= \frac{\sum_{n=1}^{N} \hat{y}^*_{(n)}}{N} \\ y^*_{(n)} &= \mu(t^*) + H_{(n)}(y_n - \mu_n(t)) \\ \sigma^{*2} &= \frac{\sum_{n=1}^{N} \hat{\sigma}^2_{(n)}}{N} + (\sum_{n=1}^{N} \hat{y}^{*2}_{(n)}/N - \hat{y}^{*2}). \end{aligned} \tag{4.26}$$

For every cycle or time point, a Gaussian distribution of the degradation state can be retained based on the observed data Y. The probability that the capacity at cycle t is less or equal to the threshold values, i.e., $Pr(y^* \leq D_T)$, can be obtained, based on which the distribution can be estimated (see Figure 4.2).

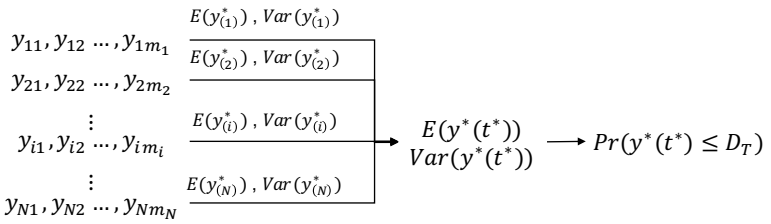

Figure 4.2: Prediction of degradation and cycle to failure estimation procedure

4.4 A joint modeling method for Lithium Ion battery capacity fade modeling and prognostics

This section presents the cycling test and the result analysis using the proposed joint modeling method. The MCD method and joint modeling method are presented. Validation of the proposed method is implemented through the battery capacity fading prognostics.

4.4.1 Experimental result

This section presents the cycling test and the result analysis using the proposed method. Four $LiCoO_2$ Lithium Ion batteries, whose rated capacities are 1.1 Ah, are tested in the room temperature under constant-current-constant-voltage (CCCV) charge/discharge profile. The profile is as follows: charge the battery at current 1C until the voltage reaches 4.2 V; continue to charge with the constant voltage of 4.2 V until the current decreases to 0.05 mA; and then discharge at current 1C until the voltage decreases to 2.5V. Table 4.1 and Figure 4.3 summarize the experimental results where the total number of cycles and cycle to failure are recorded.

4.4.2 Covariate selection in battery capacity fading

To improve the interpretation of prognostics model and its prediction performance, covariates in JMM are selected based on the physical understanding of the degradation process of the capacity fading.

Table 4.1: Summary of experiment data

Sample	Discharge rate	Number of cycles	Cycle to failure (80%)
CS2_35	1C	930	645
CS2_36	1C	970	549
CS2_37	1C	1036	625
CS2_38	1C	1050	672

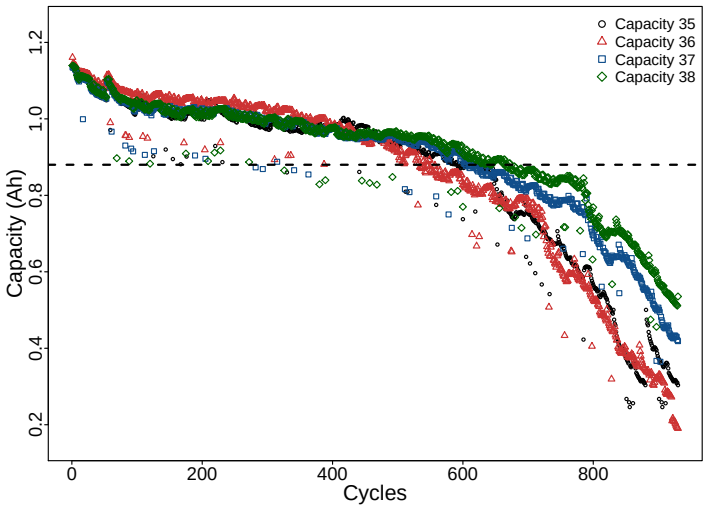

Figure 4.3: The capacity fade of tested batteries over cycles

The selection of functions and their covariates in the joint model mainly depends on the physical understanding of the degradation process. In Equation (4.6), $\boldsymbol{X}_i$ are the covariates that are used to model the mean of the measurements. The selection of $\boldsymbol{X}_i$ might mainly depend on the analytic model of the degradation process, i.e., the physics of failure model. For example, in the battery capacity degradation, it is a natural idea that the covariates can be selected from the factors affecting SEI formation during the electro-chemical reaction, such as the number of cycles, state of charge (SOC), depth of discharge (DOD), charge rate and discharge rate, temperature, etc. These factors are also good candidates of $\boldsymbol{Z}_i$, the covariates of the logarithm square variance of the jth observations. The correlation between

measurements $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ decay over the time difference between the measurements. Thus $\boldsymbol{W}_{ijk}$ should include time varying covariates that may depend on measurement time t_{ik} and t_{ij} as it is capturing the correlation between the responses at these two measurement times.

The mean functions $\boldsymbol{\mu}_i$, $i = 1, \cdots, n$, are to describe the deterministic degradation process based on the physical understandings of the degradation process. For the battery capacity degradation, it is a natural idea to select the number of cycles. These factors are also good candidates of $\boldsymbol{Z}_i$, the covariates of the log-variance of observations of the the ith subject. The correlation between measurements y_{ik} and y_{ij} decays over the time lag between the measurements. Thus W_{ijk} should include time-varying covariates that depend on time t_{ik} and t_{ij}, which can captures the correlation between the capacity at these two cycles.

4.4.3 Modeling the mean, log-variance, and correlation of capacity fading

This section aims to identify the basic forms of functions of the mean, log-variance, and angles and determine their unknown parameters.

4.4.3.1 Mean function

The baic form of the mean function is determined by the capacity degradation mechanism. The chemical degradation mechanism has been widely recognized as a major cause of loss of active lithium ions, especially the SEI formation. The capacity fade analysis models based on chemical degradation mechanism attract interest. The basic form of functions in the proposed mean function incorporates the chemical degradation model in [142], where the four scenarios of active lithium ions loss are combined: SEI formation at the first cycle, SEI thickness growth on initial surface, SEI formation due to crack propagation, and SEI thickness growth on cracked surface,

$$\hat{Q} = 1 + \frac{2-m}{C} BAL_0 \Big[(1 - CN)^{2/(2-m)} - (1 - C)^{2/(2-m)} \Big] \tag{4.27}$$

where $\hat{Q}$ represents the fraction capacity and N is the number of cycles. B, A, L_0, m, C are constant parameters determined by the material of batteries. It can be concluded that the value of m determines covariates of the mean function. It is reasonable to rewrite Equation (4.27) into the pth order polynomial function of the number of cycles, see Equation (4.28).

$$\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\alpha} = \begin{pmatrix} 1 & x_{i1} & x_{i1}^2 & \cdots & x_{i1}^p \\ 1 & x_{i2} & x_{i2}^2 & \cdots & x_{i2}^p \\ 1 & x_{i3} & x_{i3}^2 & \cdots & x_{i3}^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{in_i} & x_{in_i}^2 & \cdots & x_{in_i}^p \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix} \tag{4.28}$$

where $\boldsymbol{\mu}_i$ is the mean capacity of the ith unit, $\alpha_i, i = 0, 1, 2, \ldots, p$ are the coefficients of the polynomial terms, and $x_{ij}, j = 1, \ldots, p$ are the number of measured cycles of the ith unit.

4.4.3.2 Log-variance and angle function

Given the assumption of $n_i = n$ and $\boldsymbol{E}_i = \sigma^2\boldsymbol{I}$, the covariance matrix $\boldsymbol{\Sigma}_i$ of each unit can be written as $\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{\Sigma}_b\boldsymbol{Z}^T + \sigma^2\boldsymbol{I}$. Similarly, the first 930 measurements are used to find the basic functions of the log-variance and angles. A correlation matrix based on the historical data is sometimes non-positive definite due to missing data, noises and linearity of components. A "broken" correlation matrix with some negative eigenvalues with very small absolute values can be fixed. This research uses a simulation-based method to create a positive define matrix out of a broken correlation matrix, where the small negative eigenvalues are replaced by a normal distribution with the mean of 0 and small variance. The simulated covariance matrix is consistent of the orignal matrix [170]. Given Equation (4.5) and the simulated covariance matrix, the logarithm variance and angles ϕ_{ijk} can be obtained.

For the log-variance function, the number of cycles is also a natural choice. The angle function has a unique relationship with the entities of the correlation matrix. Thus the cycle-lag – the time difference between measurements– is selected as the covariate of the angle

function. Statistical tests are implemented to demonstrate the statistical significance of the selected covariate to the log-variance and angle function for correlations. It is difficult to detect equality of variance since only one observation for the log-variance is available. To investigate the variance of the log-variance over cycles, the bootstrapping method is employed to generate samples. The bootstrapping samples of any three batteries capacity observations are randomly sampled from four batteries, and log-variances of three capacity observations at each cycle are calculated. The ANOVA assumptions of homoscedasticity and residual normality in the log-variance are checked. It can be concluded that these two assumptions are violated with significant p-values (see Table 4.2). In addition, sample sizes

Table 4.2: Assumptions of equality of variance and residual normality for the log-variance

Test of equality of variance- Studentized Breusch-Pagan test	
BP value	325.52
Degrees of free	1
p-value	$< 2.2e-16$
Test of residual normality - Shapiro-Wilk normality test	
W value	0.85919
p-value	$< 2.2e-16$

of the angle functions over cycles are unbalanced which leads to the one-way AVONA test being too sensitive to inequality of variances and becoming non-applicable. Therefore, the nonparametric test of Friedman test is employed considering its robustness in dealing with non-normality, heterscedasticity, and outliers. The null hypothesis of the Friedman test is that there are no differences between the predictive variables. If p-value is significant, it can be concluded that at least 2 of the variables are significantly different from each other. The results of Friedman tests over log-variance and angle functions for correlations are shown in Table 4.3, where p-values for both log-variance and angles are significant. We can reject the null hypothesis of tests for both log-variances and angles. It can be concluded that the number of cycles is statistically significant to log-variance and cycle lags

are statistically significant to angles. Due to the lack of analytic models for log-variance and angles, this research uses data-driven methods to explore the basic functional forms with balanced repeated measurements. It turns out that polynomials can also be used to model log-variance and angles (see red lines in Figure 4.4 and 4.5).

Table 4.3: Friedman test for covariate significance of log-variance and angles for correlations

log-variance	
Chi-squared value	626.53
Degrees of freedom	3
p-value	$< 2.2e-16$
Angle functions for correlations	
Chi-squared value	28570
Degrees of freedom	928
p-value	$< 2.2e-16$

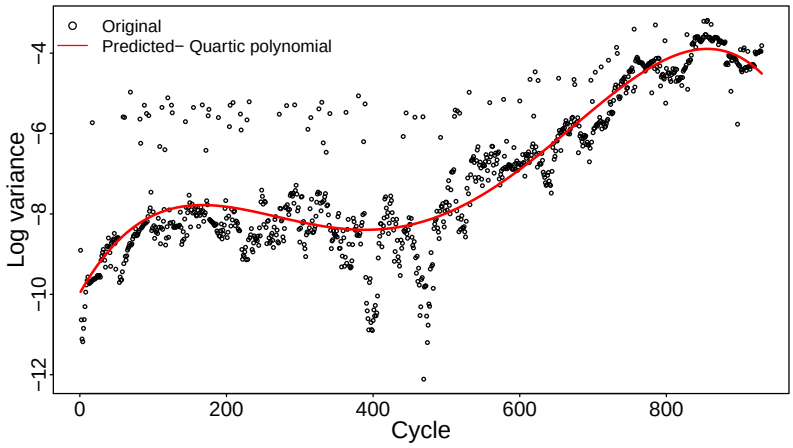

Figure 4.4: Log-variance over cycles

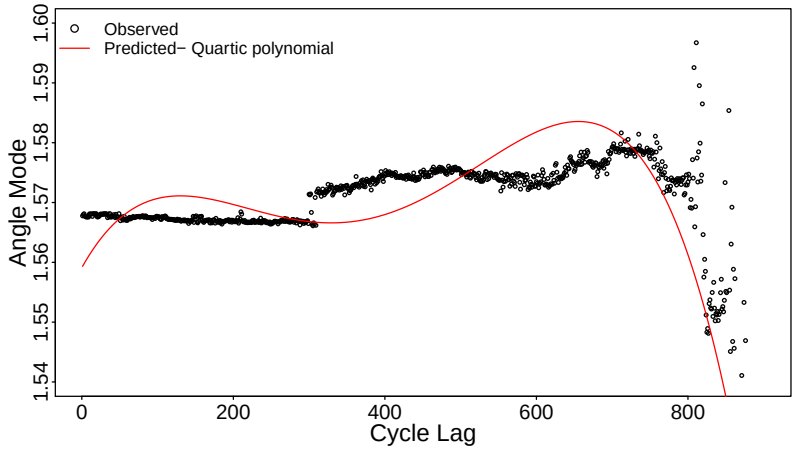

Figure 4.5: The angle mode versus the cycle lag

4.4.4 A mean-covariance decomposition modeling method for battery capacity prognostics

As mentioned in Chapter 3, the initial parameters in the mean-variance-angles are very important for the convergence of the Newton-Raphson algorithm. Inspired by the mean and covariance decomposition, the mean, log-variance, and angle function in Equation (4.6) are modeled independently in this section. Due to the high dimension covariance matrix, the concept of the moving window is also used. The optimal moving window is determined through the same method in Section 4.2. This research fits the mean, log-variance, and angle function through non-parametric methods under the optimal moving window. Using the estimated mean and covariance, the multivariate distribution of the capacity over cycles can be used to illustrate the degradation process of the capacity. Given the failure threshold of the capacity, the distribution of the cycle to time can be obtained through sample numerical simulations.

4.4.4.1 Modeling the mean function

As discussed in Section 4.4.2, the mean function of the capacity degradation process can be written as function of cycles. To enable this model to analyze various types of lithium ion batteries, Equation (4.27) is simplified into a general form. Based on the empirical values of m and C in [171], it is reasonable to rewrite Equation (4.27) into the third order polynomial function of the number of cycles, see Equation (4.29).

$$\begin{aligned} Q(N) &= a_n N^{2/(2-m)} + a_{n-1} N^{2/(2-m)-1} + \cdots + a_1 N + a_0; \\ Q(N) &= a_3 N^3 + a_2 N^2 + a_1 N + a_0; \end{aligned} \tag{4.29}$$

where $Q(N)$ is the battery capacity at cycle N, $a_i, i = 3, 2, 1$, are the coefficients of the polynomial terms and a_0 represents the constant. The parameters can be determined by balancing the residuals and test errors. The training data and test data in this research is 80% and 20%. Equation (4.30) presents the estimated mean function. Another commonly used physics failure model two-term exponential function [143], i.e. $a \cdot \exp(b \cdot N) + a \cdot \exp(d \cdot N)$, is also used. But the proposed polynomial model has smaller RSME of 0.0147 than the two-term exponential function of 0.0914. Figure 4.6 illustrate the polynomial predictive model, two-term exponential model and average capacity over the batteries.

$$\begin{aligned} \mu = 1.109 &- 7.679 \times 10^{-4} N + 1.947 \times 10^{-6} N^2 \\ &- 2.171 \times 10^{-9} N^3. \end{aligned} \tag{4.30}$$

4.4.4.2 Modeling log-variance and correlation

The concept of moving window is used for variance and correlation modeling due to the high dimension of the covariance in this case. Based on the same procedure mentioned in Chapter 3, the optimal size of moving-window is 46 with the minimum mean absolute deviation (MAD), $1.215e-06$. Compared with the optimal length of 195 in the linear mixed effects model, less historical information is used for prediction (see Figure 4.7).

Since there is no analytic model for the variance and angle function, non-parametric learning algorithms are used to model the logarithm variance and angles. The concept

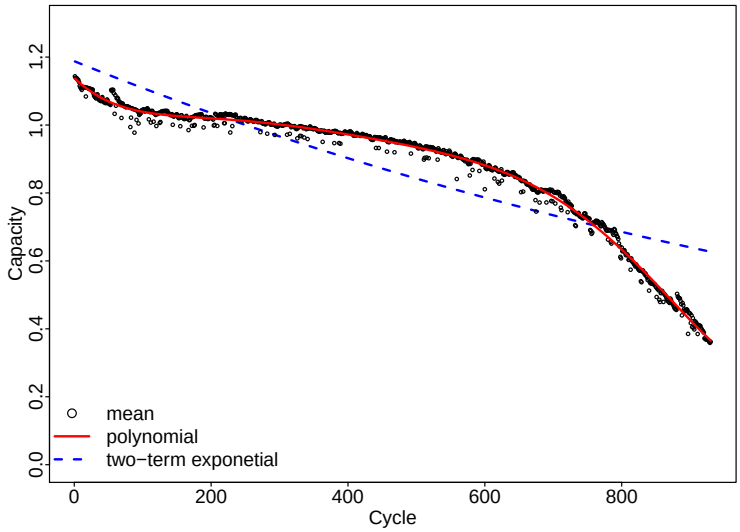

Figure 4.6: Comparison of the polynomial and two-term exponential predictive models

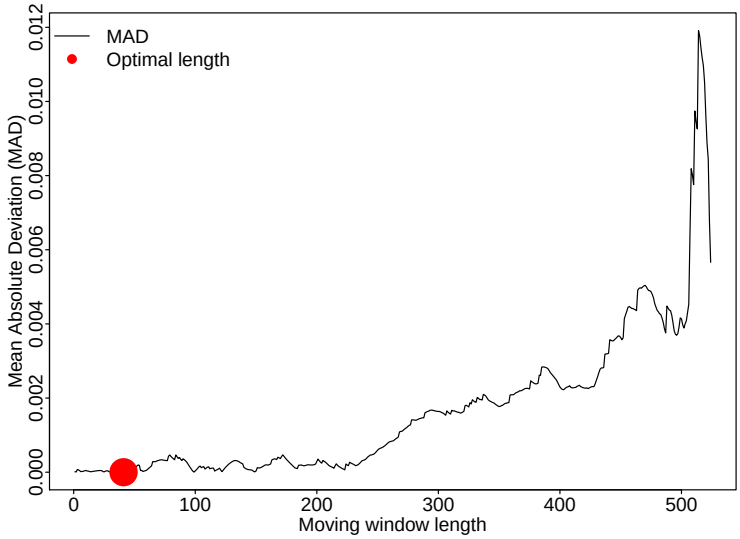

Figure 4.7: The optimal moving window

of SVM is applied to model the logarithm variance, whose core is to train support vector machines so that the data points lie in between the two borders of the margin, which is maximized under suitable conditions to avoid outliers inclusion. For the angles which

indicate the correlations of longitudinal measurements, there would be large, densely sampled data. Due to the lack of theoretical understanding, non-parametric regression methods are natural choices. Within a moving window, the fitted curve based on SVM is obtained where the predictions of future values yield. Figure 4.8 shows one example of data points of logarithm variance, and the predicted curve and Figure 4.9 shows an example of the angles versus cycle lag and the fitted curve based on SVM regression. Obviously the moving window significantly reduce the computation complexity.

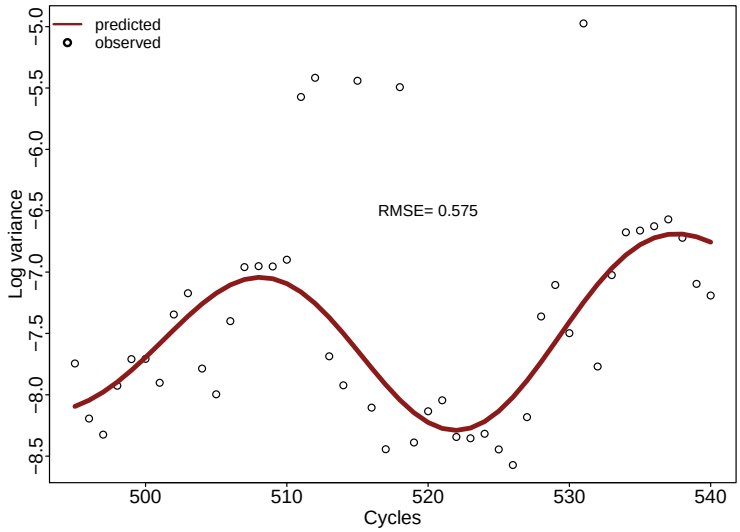

Figure 4.8: The actual and predicted log-variance

4.4.4.3 Performance prediction using mean-covariance decomposition method

To have a better understanding of the cycle to failure estimation, the mean-covariance decomposition model with the optimal moving window length of 46 is used to predict the capacity around the last 100 cycles. Given the end life cycles, the data of the first 540 cycles are used to predict the cycle to failure at which the capacity of batteries are less or equal to 80% of the initial capacity. The prediction based on the proposed method using the optimal window is done for the last 15 cycles. That is, the predictions of the capacity

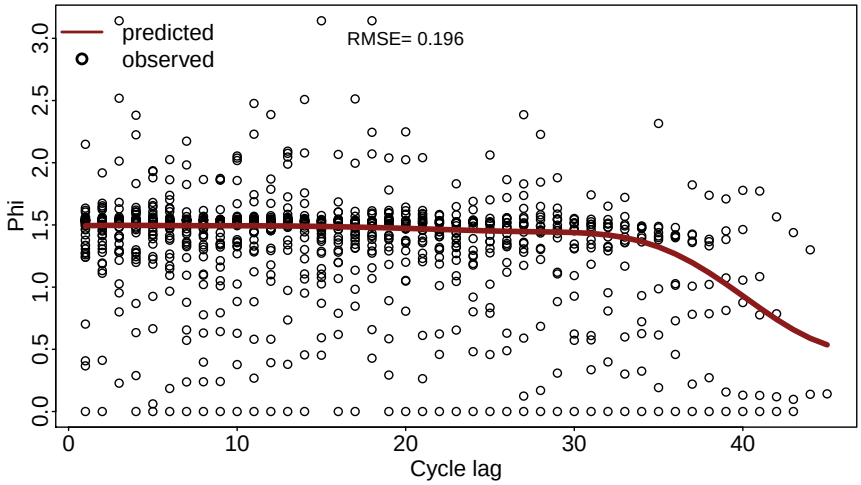

Figure 4.9: The actual and predicted angles versus the cycle lags

at cycle 541 are made based on the historical capacity data at cycle 495 to 540, and that at cycle 542 is based on 496 to 541, and so on. The prediction of future values is provided in the form of a multivariate normal distribution, where each variate– the prediction of the capacity– follows a normal distribution. The prediction and the 90% confidence interval is shown in Figure 4.10. To illustrate the performance of the proposed method, the mixed effects model suggested in [59] is also used to predict the capacity at cycle 541 to 560. Since the predicted results are both in the form of statistical distributions. Compared with the mixed effects model in [59], the width of 90% confidence interval is 0.108, while that of the mixed effects model is 0.044. The prediction based on mean-covariance decomposition method has high variation compared with that of the mixed-effect model proposed in [59]. However, the mean function modeling is more accurate in terms of lower RMSE and short moving window.

A mean-covariance modeling method is proposed to model the longitudinal and between-sample uncertainties. Through the covariance matrix of the multivariate normal distribution of the repeated measurement, mean-covariance decomposition can effectively deal with unbalance data through decomposed covariance matrix and the time-vary random effects. With the characteristics of the correlation matrix, a trigonometric function is used

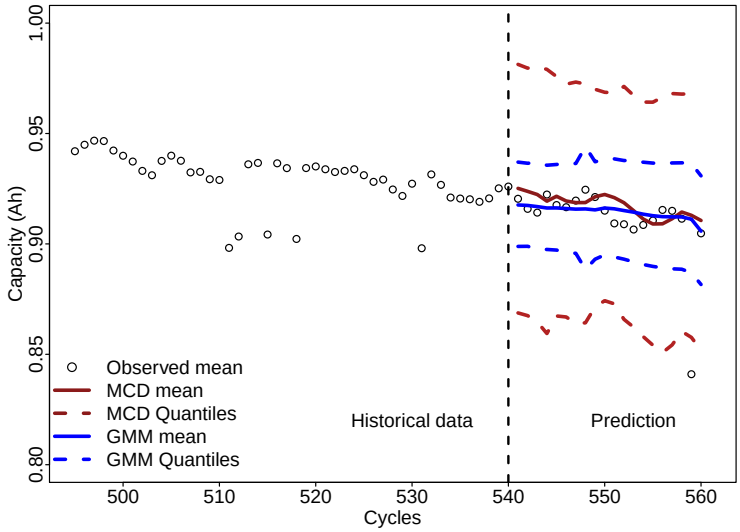

Figure 4.10: Performance comparison of the proposed method and the mixed effects model

to reparameterize the correlation matrix, which can reduce the time-varying factors of the correlation matrix. To improve the interpretation of the degradation model, the analytic model from the electrochemical viewpoint is employed as the basic form of the mean function. For the slow degradation process, the moving-window scheme is used to include the most recent information for predictions. Within the optimal moving window, the parameters in the mean-covariance models are estimated through balancing the goodness-of-fit of the capacity data and the model complexity. Compared with the mixed effects model, the proposed method needs fewer historical data with the moving window with smaller length, which improves the accuracy of the prediction of the mean function. Consideriing either the MCD method and the mixed effects model can not easily deal with unbalanced data, the joint modeling method is used for prognostics using unbalanced data.

4.4.5 Parameter estimation of joint modeling method

Based on the physics of failure and fitted polynomials of log-variance and angles, the mean, log-variance, and angle function can be written as

$$\boldsymbol{\mu}_i = \boldsymbol{X}_i\boldsymbol{\alpha}, \log \boldsymbol{\sigma}_i^2 = \boldsymbol{Z}_i\boldsymbol{\beta}, \phi_{ijk} = \boldsymbol{W}_{ijk}\boldsymbol{\gamma}. \tag{4.31}$$

The highest polynomial orders of μ_{ij}, $\log \sigma_{ij}^2$, and ϕ_{ijk} functions are specified as 4, 4, and 6, respectively. Through PMLE, significant covariates can be obtained. The tuning parameter λ assigned through a grid search in (0, 100) by the step of 0.1. The estimated parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are shown in Equation (4.32) with the tuning parameter λ of 3.9 and BIC of -112.4.

$$\boldsymbol{\alpha} = \begin{pmatrix} 1.1506 \\ 0 \\ 0.0045 \\ 0.34 \\ 0 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} -9.59 \\ 3.42 \\ 0 \\ -0.23 \\ 0 \end{pmatrix}, \boldsymbol{\gamma} = \begin{pmatrix} 1.54 \\ 0 \\ 0.47 \\ 0.12 \\ -0.21 \\ 0 \\ -0.10 \end{pmatrix} \tag{4.32}$$

4.4.6 Cycle to failure assessment

Using the estimated mean and covariance function, the distribution of $\boldsymbol{y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ can be obtained based on Equation (4.5). Cycle to failure can be assessed through numerical simulations based on Figure 4.2. Figure 4.11 illustrates the cycle to failure distribution by setting 0.88 Ah as the threshold of failure. It turns out that cycle to failure normally distributes with the mean of 634.6 and the standard deviation of 10.2.

4.4.7 Conclusion and discussion

A joint modeling method (JMM) of mean-variance-correlation is proposed to model Lithium Ion battery capacity fade in this research. The mean, variance, and correlation

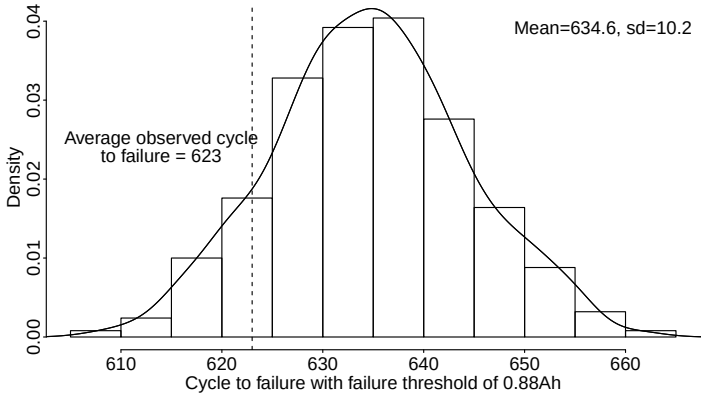

Figure 4.11: Cycle to failure of $D_F = 0.88$Ah

of repeated measurements of capacity are individually explored in JMM, through which the within and between-subject uncertainties as well as time-varying random effects in the battery capacity degradation process are effectively modeled. The unstructured correlation matrix in JMM can easily avoid the negative effects of unbalanced or missing data. Considering the characteristics of the correlation matrix, the trigonometric function can be used to reparameterize the correlation matrix, which reduces the complexity of modeling correlations and improves the interpretation of entities of the decomposed matrix. The analytic model of formation and growth of SEI over cycles are used to provide the basic polynomial function for the capacity mean in JMM. The basic form of log-variance and angle mode are estimated through leaning methods. Polynomials are selected for the practical interest. For the purpose of comparison, the mean-covariance decomposition (MCD) method is also investigated, where parameters of the mean, variance, and angles for correlations are estimated independently and whose result can be assigned as the initial values of JMM. Penalized maximum likelihood estimate (PMLE) is used to select covariates and estimate parameters in the above three functions. Based on the capacity model, cycle to failure can be obtained through simulations.

5. OTHER PROGNOSTICS METHODS AND PERFORMANCE EVALUATION

This chapter introduces two other prognostics methods– prognostics based on the functional principal component analysis (FPCA) and spatio-temporal method, which can be used to solve challenges using repeated measurement data of multiple units, such as high dimensionality and dependent units. FPCA is the extension of the conventional principal component analysis, which can provide more stable estimations and avoid the high-dimensional covariance matrix. Eigenfunctions of the estimated covariance kernel can be easily computed and provide insight into the model of variability among the groups of degradation signals. Through representing each degradation trajectory with a specific combination of eigenfunctions, challenges along with unbalanced data can be well addressed in FPCA. The assumption of independence among multiple units is impractical in many cases. More importantly, components/subsystems in complex systems are correlated with each other during operations. Such correlation is critical to the lifetime of the system. For example, the capacity degradation behavior of a battery pack with parallel connected cells depends on the correlation between these cells during cycling usage. Spatio-temporal models are utilized to model the mentioned correlation between units, which initially aims to model spatial and temporal trends and uncertainties in geostatistics. In addition, performance comparison among proposed methods is implemented in terms of prediction accuracy of degradation and time to failure.

5.1 Prognostics based on functional principal component analysis

This research proposes a method based on FPCA, where the mean function and covariance kernel are modeled separately. The mean function can be estimated easily through the existing learning algorithms, such as support vector machine (SVM). The mean function is

determined through the local smoothing methods by assuming a deterministic mean function representing the underlying physics-based degradation mechanisms. Eigenvalues and eigenfunctions are used to represent the covariance kernel based on the Karhunen-Loéve decomposition.

5.1.1 Functional data analysis

The term "functional data" refers to data where each observation is a curve, a surface, or a hypersurface, as opposed to a point or a finite-dimensional vector [172–174]. For example, the degradation mechanism analysis of a certain type of electronic part is expected. The test data collected over time of each sample produces a curve over the time interval. The smoothness property of the observation data, though observations could be viewed as discrete, facilitates the functional intepretation [175]. Functional data analysis (FDA) is to explore the smoothness of data through dealing with the analysis and theory of data that are in the form of functions, images and shapes, or more general objects. By considering observations as curves, FDA has advantages in reducing dimension and dealing with highly correlated data. Moreover, smooth functional representations can allow us to evaluate the trend and fluctuation of degradation processes at any time point, even in the case of unbalanced data.

Assume the observed data, $\{(y_{i1}, t_{il}), \ldots, (y_{im_i}, t_{im_i})\}$, where y_{ij} is the observation of the ith latent curve Y_i at time t_{ij}. $Y_i(t)$, $i = 1, ..., N$, are independent realizations of the stochastic process $Y(t)$ whose mean is $\mu(t) = E(Y(t))$ and covariance function $\Sigma(s,t) = cov(Y(s), Y(t))$. With white noises, $y_{ij} = Y_i(t_{ij}) + \epsilon_{ij}$, where ϵ_{ij} are usually i.i.d Gaussian distributions. A commonly adopted approach to interpolate the stochastic process Y_i is through basis expansion. Let $\{\psi_1(\cdot), \psi_2(\cdot), \ldots, \psi_K(\cdot)\}$ be a set of basis functions. The observations y_{ij} can be represented as $y_{ij} = \sum_{k=1}^{K} c_{ik}\psi_k(t_{ij}) + \epsilon_{ij}$, where c_{ik} are the subject-specified scores. Basis functions $\psi(t)$ can be either pre-specified or data-driven. A set of pre-specified basis functions could be B-spline, wavelets, and Fourier. The choice of basis functions is based on the data characteristics, e.g. B-splines for Gaussian process data,

wavelets for signal data, and Fourier series for periodic data. For any curve, the number of pre-specified basis functions $\psi(t)$ as well as scores c_{ik} are the parameters to be determined through minimizing the sum of squared errors or penalized sum of squared errors. For the data-driven bases, one convenient choice is the eigenbasis of $Y(t)$. Referring to Karhunen-Loéve expansion, the related methods is called functional principal component analysis (FPCA), which is considered as an extension of the principal component analysis (PCA) [176] in functional data.

5.1.2 Functional principal component analysis

The comprehensive frame of functional principal component analysis (FPCA) was first developed by Dauxios and Pousse [177]. FPCA becomes the most important tool in FDA because of its advantages in dimension reduction through facilitating the conversion of infinite-dimensional functional data into finite-dimension random vectors [174, 178–183]. FPCA provides a data-driven basis, based on which measurements can be represented as $y_{ij} = \mu(t_{ij}) + \sum_{k=1}^{K} c_{ik}\psi_k(s_{ij}) + \epsilon_i$. The unknown set of bases $\{\psi_1(t), \psi_2(t), \ldots\}$ describes the main directions of variability in the observation data. Similar with PCA, othonormal functional principal components are obtained through

$$\begin{cases} \psi_1 = \text{argmax} \int \psi_1(t)[Y_i(t) - \mu(t)]dt, & s.t. \parallel \phi_1(td) \parallel^2 = 1 \\ \psi_k = \text{argmax} \int \psi_k(t)[Y_i(t) - \mu(t)]dt, & s.t. \int \psi_{k-1}(t)\psi_k(t)dt = 0, \\ & \int \psi_k(t)\psi_{k-1}(t)dt = 0 \end{cases} \tag{5.1}$$

where $\mu(t) = E[Y_i(t)]$ and ψ_k are the bases.

Let the continuous systematic function $\Sigma(s,t)$ be the covariance function of Y, such that

$$\Sigma(t,s) = cov(Y(t), Y(s)) = E\{[Y(t) - \mu(t)][Y(s) - \mu(s)]\} \tag{5.2}$$

$\Sigma(t,s)$ includes the kernel operator Σ defined by $\Sigma\psi(t) = \int \Sigma(t,s)\psi(s)ds$. According to the Mercer's theorem, there is an orthonormal basis function ψ_k and non-increasing positive value λ_k, i.e., $\lambda_1 \geq \lambda_2 \geq \lambda_k \geq \ldots > 0$ such that

$$\begin{aligned} \Sigma\psi_k &= \int \Sigma(t,s)\psi_k(s)ds = \lambda_k\psi_k \\ \text{and,} \Sigma(t,s) &= \sum_{k=1}^{\infty} \lambda_k\psi_k(t)\psi_k(s). \end{aligned} \tag{5.3}$$

Under the assumption of Mercer's theorem, the stochastic process can be expressed as

$$Y(t) = \mu(t) + \sum_{k=1}^{\infty} \sqrt{\lambda_k}\xi_k\psi_k(t), \tag{5.4}$$

where $\xi_k = \frac{1}{\sqrt{\lambda_k}} \int Y(t)\psi_k(t)dt$ is a random variable with mean $E(\xi_k) = 0$ and variance $E(\xi_j\xi_k) = \delta_{jk}$– the Kronecker delta. According to the Mercer's theorem, Equation (5.4) converges uniformly, i.e., $E[Y(t) - \mu(t) - \sum_{k=1}^{K} \sqrt{\lambda_k}\xi_k\psi_k(t)]^2 \to 0$. Therefore,

$$Y_i(t) = \mu(t) + \sum_{k=1}^{K} \sqrt{\lambda_k}\xi_{ik}\psi_k(t) + \epsilon_i, \tag{5.5}$$

which is named as Karhunen-Loève expansion that illustrates that $Y_i(t)$ is projecting into a K-dimension space. The idea of FPCA is to use the first K terms of KL expansion as an approximation of $Y(t)$.

5.1.3 Degradation modeling based on functional data analysis

Based on Equation (4.1), the degradation trajectory of the ith unit can be rewritten as

$$Y_i(t) = \mu(t) + S_i(t) + \epsilon_i(t), \tag{5.6}$$

where $\mu(t)$ is the mean of the degradation trend, S_i is the stochastic process from the underlying degradation mechanism, and ϵ_i is the noise. For a degradation process, the mean function can be assumed to be a deterministic model. Physical understanding of the degradation process provides the basic knowledge of this deterministic model, such as the function form of the model, covariates, and so on.

5.1.3.1 Mean function $\mu(t)$

For a degradation process, the mean function $\mu(t)$ can be assumed to be a deterministic model. Physical understanding of the degradation process provides the basic knowledge of this deterministic model, such as the functional form of the model, covariates, and so on. Take lithium ion battery degradation for instance, where the formation, growth, and repair of SEI cause most loss of the active lithium ions. Factors, such as cycles, temperature, state of charge, depth of discharge, and charge/discharge rate are other common covariates. Many parametric and nonparametric methods can be used to model the mean function, such as general regression methods, functional regression and other data-driven algorithms. Based on the discussion in Chapter 4, the mean function is modeled through SVM.

5.1.3.2 Stochastic process $S_i(t)$

The stochastic process $S_i = Y_i - \mu$ has mean zero and covariance kernel. The covariance kernel $\Sigma_i(t, s)$ can be decomposed as Equation (5.7) according to the Karhunen-Lovéve decomposition.

$$cov(t, t') = \sum_{k=1}^{\infty} \lambda_k \psi_k(t) \psi_k(t'), \tag{5.7}$$

where $t, t' \in \{t_{ij}\}_{j=1,\ldots,m_i}$. $\psi_k(t), k = 1, \ldots$ are the eigenfunction and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$ are the ordered eigenvalues. Based on this decomposition, S_i can be rewritten as

$$S_i = \sum_{k=1}^{\infty} \xi_{ik} \psi_k(t), \tag{5.8}$$

where ξ_{ik} are the scores that describes the uncorrelated random effects with mean zero and variance λ_k. Since only a small number of eigenvalues are significantly nonzero, finite are nonzero. Therefore, Equation (5.8) can be approximated as

$$S_i(t) = \sum_{k=1}^{K} \xi_{ik} \psi_k(t), \tag{5.9}$$

where K is the number of significant nonzero eigenvalues. The eigenfunction $\psi_k(t)$ can be represented as a linear combination of basis functions, *i.e.*, $\psi_k(t) = \sum_{m=1}^{M} b_{km} \phi_m(t)$

where $\phi_m(t), m = 1, \ldots, M$ are known bases, such as Bsplines [184, 185]. Then for every t,

$$\begin{aligned} \psi(t) &= (\psi_1(t), \ldots, \psi_K(t))^T, \\ &= B^T(\phi_1(t), \ldots, \phi_M(t))^T, \end{aligned} \tag{5.10}$$

where B is a $M \times K$ matrix with the factor b_{mk} and $B^T B = I$.

5.1.3.3 Parameter estimation and time to failure evaluation

Similarly to the joint modeling method, maximum likelihood estimation is used to estimate the parameters, B, Λ, and σ^2. The negative log-likelihood of $S_i(t) = Y_i - \mu$ is

$$\begin{aligned} -\log(B, \Lambda, \sigma^2) = &\frac{1}{2}\sum_{i=1}^{n} tr[(\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i)^{-1} S_i^T S_i] \\ &+ \frac{1}{2}\sum_{i=1}^{n} \log|\sigma^2 I_{m_i} + \Phi_i^T B \Lambda B^T \Phi_i|. \end{aligned} \tag{5.11}$$

Peng et al. [186] proposed a geometric approach to solve this optimization problem, where B, Λ, and σ^2 are considered as the vectors in the Stiefel manifold and two-step procedure of Newton-Raphson algorithm is applied to estimate the parameter. More comprehensive details of the algorithm can be found in Peng's work. With B, Λ, σ^2, we can obtain a complete degradation model which describes the variability from within-unit and between-unit. Therefore, cycle to failure can be obtained based on the simulation.

5.1.4 Prognostics of Lithium Ion battery using functional principal component analysis

Functional principal component analysis can be used for battery capacity prognostics. The observed degradation signal is decomposed into the mean function and variance-covariance function. Functional principal components of the variance-covariance function are represented and modeled through eigenfunctions, which are further approximated and estimated using a combination of B-Splines. For the battery capacity prognostics, three

eigenfunctions explained 99.98% of the total variation. Capacity prediction and cycle to failure distribution are also analyzed and evaluated based on the proposed method.

As discussed in Chapter 4, support vector regression is used to estimate the mean function, which provides a parsimonious algorithm with the certain level of accuracy. The mean absolute deviation (MAD) using SVM of 0.008 is less than that using location quadratic smoothing as 0.012. Three eigenvalues are included $\lambda = (2.497, 0.056, 0.005)$. The error variance σ^2 is 0.0007 with 6 cubic splines are included. Over 99% of total variation is modeled using the first three principal components. Using the estimated mean function and covariance kernel, the capacity can be predicted based on the simple kriging shown in Equation (4.22) (see Figure 5.1). Figure 5.3 and 5.4 shows the cycle to failure distribution by setting 0.88 Ah as the threshold. It turns out that the mean of the cycle to failure is 635 and the standard deviation is 15.3.

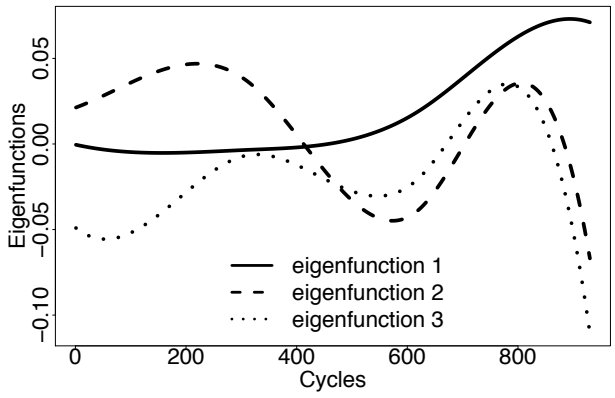

Figure 5.1: The first three eigenfunctions over cycles

5.2 Prognostics based on general spatio-temporal model

All aforementioned methods investigate the degradation model of the ith unit with the observations $\boldsymbol{y}_i$, $i = 1, \ldots, n$. Given the assumption that these units are independent from each other, maximum likelihood estimation (MLE) is used to estimated the fixed and random effects coefficients. But no such model that includes unit-specific variation, temporal

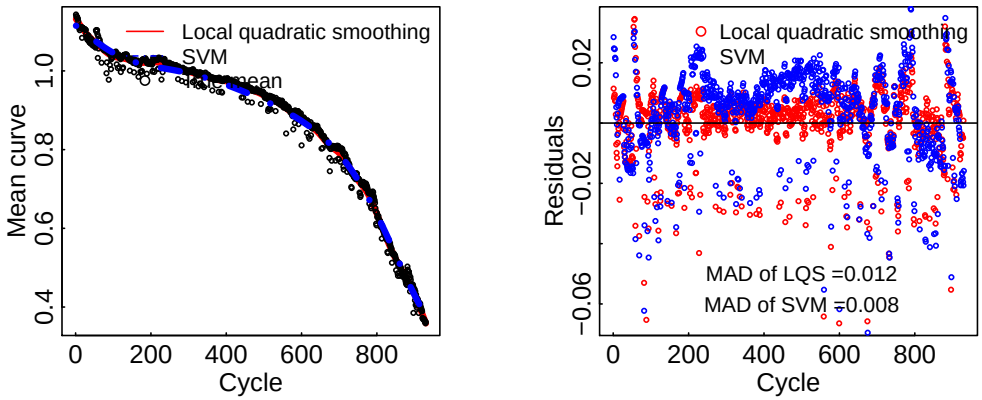

Figure 5.2: The comparison of location quadratic smoothing and SVM

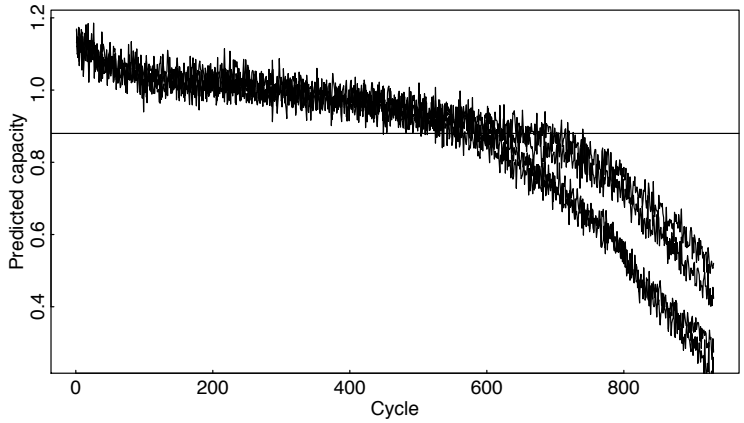

Figure 5.3: The simulated capacity degradation process based on the proposed method

variation, and unit-temporal variation has been studied for the repeated measurement data of multiple units. This research proposes a general spatio-temporal model to investigate these variations.

5.2.1 General spatio-temporal model

Initially spatio-temporal models arise when data is collected across time as well as space [187]. A typical example would be that of a monitoring network of an atmospheric

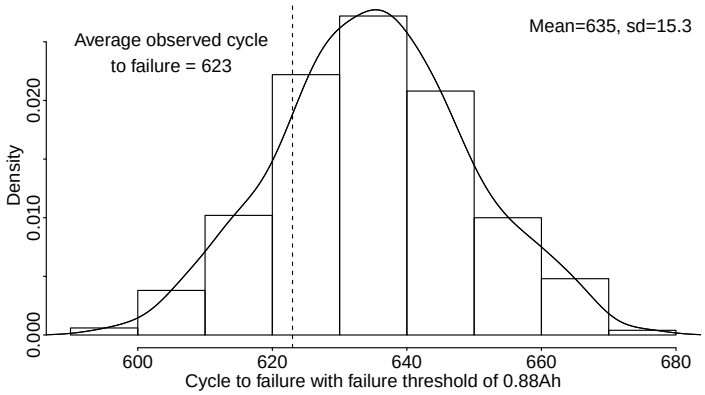

Figure 5.4: Cycle to failure estimation and the observed average cycle to failure

pollutant, or a network of meteorological stations on which data are collected at regular intervals. Thus the data analysis has to take account of spatial dependence among the monitors, but also that the observations at each monitor typically are not independent, but form a time series. Temporal correlations and spatial correlations are taken into account in this model. Wikle et al. [188] summarized the general spatio-temporal model as

$$\boldsymbol{Y}(s,t) = \mu(s;\theta_t) + \gamma(t;\theta_s) + \kappa(s;t;\theta_{s,t}) + \epsilon(s;t), \tag{5.12}$$

where $Y(s,t)$ denotes a spatio-temporal random process. s and t are the spatial and temporal factors, respectively. $\mu(s;\theta_s)$ is the spatial trend function which can be used to describe the difference between multiple units and θ_t is the spatial trend function parameter that change over the temporal factor t. $\gamma(t;\theta_s)$ is the temporal trend with the spatially varying parameter θ_s. $\gamma(t;\theta_s)$ can be used to illustrate the degradation process of each unit. $\kappa(s;t;\theta_{s,t})$ describes the spatio-temporal dependency. $\epsilon_{s,t}$ represents the error over (s,t). $\kappa(s;t;\theta_{s,t})$ is the spatio-temporal dependence. $\boldsymbol{Y}(s,t)$ performs as a spatio-temporal process which can be evaluated by its mean and the associated covariance function. Without losing generality, we can let $EY(s;t) = Z(s;t) = \mu(s;\theta_t) + \gamma(t;\theta_s) + \kappa(s;t;\theta_{s,t})$. Therefore,

$$Y(s;t) = Z(s;t) + \epsilon_{s;t}, \tag{5.13}$$

where $Z(s;t)$ can be modeled based on the physical understanding or empirical data. Usually the case with continuous spatial and discrete time is investigated. For example, Brown et al. [189] and Liu et al. [190] provided comprehensive discussions of the statistical modeling of spatio-degradation data, where the spatio-temporal degradation process $Y(s;t)$ is obtained by smoothing $Y(s;t-\Delta), \forall \Delta \geq 0$ with the spatial trend and a stochastic propagation process. To extend the application of spatio-temporal methods, the requirement of continuous spatial variables is less demanding. For example, the capacity degradation of a battery pack is the degradation propagation of cells which distribute spatially in the pack. In the case of discrete spatial variables, $Z(s;t)$ can be easily obtained. The covariance of $Y(s;t)$ is often assumed to be stationary. That is, the covariance of $\boldsymbol{Y}(s,t)$ and $\boldsymbol{Y}(s+h,t+u)$ is given as $C(h,u) = cov(\boldsymbol{Y}(s,t), \boldsymbol{Y}(s+h,t+u))$. Though investigations of non-stationary covariance have been done, methods are still mainly based on the stationary covariance. For example, Sigrist et al. [191, 192] have tried to add Markov chain and auto-regression coefficient into stationary covariance function for the non-stationary saptio-temporal process. Hefley et al. [193] introduced random effects into spatial covariance.

This research aims to extend the spatio-temporal model in prognostics modeling with repeated measurement data of multiple units to quantify the correlation between units. Multiple families of stationary covariances and variogram estimators are discussed. The capacity degradation of battery packs is used to validate the proposed method, where the correlation of cells during usage cycling is investigated. To estimate cycle to failure, the prediction based on the spatio-temporal model can be easily done through the simple kriging predictor.

5.2.2 Model selection and parameter

$Z(s;t)$ represents the mean of the degradation process over space and time. A general linear model $Z = X\alpha$ can used to model $Z(s;t)$. As mentioned in Chapters 3 and 4, physical understanding can provide covariance, basic function form, etc. For a spatio-

temporal stochastic process $Y(s;t)$, spatio-dependency is usually chararicterized by the covariance function, $cov(Y(s_1;t_2), Y(s_2;t_2)) = C(s_1, s_2; t_1, t_2)$. A stationary covariance is defined by the fact that the covariance depends on only the location distance and time difference, which is $C(h,u) = C(Y(s+h;t+u), Y(s;t))$. The separable covariance composed of the pure spatial and temporal covariances is a natural idea to construct the covariance matrix of $Y(s;t)$. However, separable covariance often fails to model the space-time interaction. Cressie et al. [194] proved that continuous spatio-temporal stationary covariance function can be written as the product of two parts based on Bochner's theorem. This is

$$\begin{aligned} C(h,u) &= \int\int \epsilon^{ih'w+iu\delta} g(w;\delta) dw d\delta \\ g(w;\delta) &= (2\pi)^{-1} \int \exp(-iu\delta) h(w;u) du \\ h(w;u) &= \int g(w;\delta) d\delta. \end{aligned} \tag{5.14}$$

By simply defining $\rho(w,u) = \dfrac{h(w;u)}{\int g(w;\delta)d\delta}$ and $k(w) = \int g(w;\delta)d\delta$, we can obtain $h(w;u) = \rho(w;u)k(w)$. The covariance function of $Y(s;t)$ can be obtained by specifying $\rho(w;u)$ and $k(w)$. Seven examples were given in [194, 195], which can be summarized into the following three major categories with difference spatial margins, such as Gaussian, Whittle-Matérn function, and exponential, whose generalized forms are shown as follows:

(1) Covariance function with Gaussian related spatial margins:

$$C(h;u|\theta) = \frac{\sigma^2}{(a|u|+1)^{d/2}} \exp\left(\frac{b^2 \parallel h \parallel^2}{a|u|+1}\right) \tag{5.15}$$

where $h \in \Re^d$, $u \in \Re$, and $\theta = (\sigma^2, a, b)$ is the set of unknown parameters.

(2) Covariance function with Whittle-Matérn function related spatial margins:

$$C(h;u|\theta) = \begin{cases} \dfrac{\sigma^2(2c^{d/2})}{(a^2u^2+1)^v(a^2u^2+c)^{d/2}\Gamma(v)} \left[\dfrac{b}{2}\left(\dfrac{a^2u^2+1}{a^2u^2+c}\right)^{1/2} \parallel h \parallel\right]^v & \parallel h \parallel \neq 0 \\ \dfrac{\sigma^2 c^{d/2}}{(a^2u^2+1)^v(a^2u^2+c)^{d/2}} & \parallel h \parallel = 0 \end{cases} \tag{5.16}$$

where $h \in \Re^d$, $u \in \Re$, and $theta = (a, b, c, d, v)$ is the set of unknown parameter.
(3) Covariance function with exponential related spatial margins:

$$C(h; u|\theta) = \frac{\sigma^2(a|u| + 1)}{[(a|u| + 1)^2 + b^2 \parallel h \parallel^2]^{d+1}/2} \tag{5.17}$$

where $h \in \Re^d$, $u \in \Re$, and $\theta = (a, b)$.

The selection of the covariance function of a spatio-temporal process is usually based on the the characteristics of variogram, such as trend and concaveness. Variogram depicts the spatial auto-correlation of the measured sample points. The variogram $2\gamma(h; u|\theta) = var(Y(s + h; t + u), Y(s, h))$. For a stationary covariance function, $var(Y(s + h; t + u), Y(s, h)) = C(0; 0|\theta) - C(h; u|\theta)$. Then the variogram can be calculated via

$$2\gamma(h; u|\theta) = C(0; 0|\theta) - C(h; u|\theta). \tag{5.18}$$

For a spatio-temporal process, the empirical variogram estimator is needed because sample information of $Y(s; t)$ is not available for every location. The commonly used empirical variogram is originally defined by Matheron [196], which is the average squared difference between measurements separated by distance h and time u. That is,

$$2\hat{\gamma} = \frac{1}{|N(h; u)|} \sum_{i,j,t,t'} (Y(s_i; t) - Y(s_j; t'))^2, \tag{5.19}$$

where $N(h; u)$ is the set of all pairwise distances $s_i - s_j = h$ and $t - t' = u$, and $|N(h; u)$ is the number of distinct pairs in $N(h; u)$. Because this classic estimator is not effective in the case of the existence of outliers, this research applies robust estimators mentioned in [197] to estimate parameters from spatio-temporal measurements.

Parameters θ in the covariance function can be estimated by fitting $2\hat{\gamma}$ to the variogram calculated in Equation (5.18). Regression methods, such as least-square methods and maximum likelihood methods, can be used. Cressie [198–200] proposed weighted-least-squares methods, where the parameter θ is estimated by minimizing weights in Equation (5.20), which is solved through the gradient decent method in this research.

$$W(\theta) = \sum_{h=1}^{H} \sum_{u=0}^{U} |N(s; t)| \Big(\frac{\gamma(\hat{h}; u)}{\gamma(h; u)} - 1 \Big)^2. \tag{5.20}$$

5.2.3 Prognostics modeling battery pack capacity

A battery pack is composed of cells that are managed through battery management system. In the battery pack, the battery management system balances all the cells in the battery pack by intelligently bleeding off excess energy from cells that are charged more than others. This provides the maximum amount of usable energy (capacity) from the battery pack since the pack is only as strong as the weakest cell. The "spatial" distance of cells can be expressed by the way in which they are connected. The spatio-temporal model is used to model the capacity degradation with the focus of cell correlation during the charge and discharge process. Because of lack of measurements of capacity of cells and unclear correlation between cells, the sample entropy method is used to approximate the capacity change over cycles. The capacity values at the end of each operation cycle are usually recorded so that the state of health can be estimated. The open circuit voltage and current change over time in one operation cycle and vary from the charge/discharge profile as well as the number of cycles. For example, in typical constant-current-constant voltage (CCCV) charge/discharge profile, the voltage of the test battery increases first, then stays steady, and eventually drops. Moreover, the curves of the voltage profile vary from the cycles. For the same battery, the voltages curves of each cycle are different from each other. Based on the sample entropy algorithm, there exists a certain sample entropy value and a test capacity value for each cycle. With predicted capacity of each cells, the spatio-temporal model can be used to estimate the mean and covariance function of cells in the tested pack. Prediction of capacity degradation of cells can be easily obtained based on the simple kriging. Figure 5.5 illustrates the scheme of the proposed prognostics model based on the spatio-temporal model.

5.2.3.1 Experiment design

Two paralleled connected LiCoO2-graphite batteries, shown in Figure 5.6, with the rated capacity 3.4 Ah and rated voltage 4.2 V are tested at the room temperature under constant-current-constant-voltage (CCCV) charge/discharge profile. The cycling test was

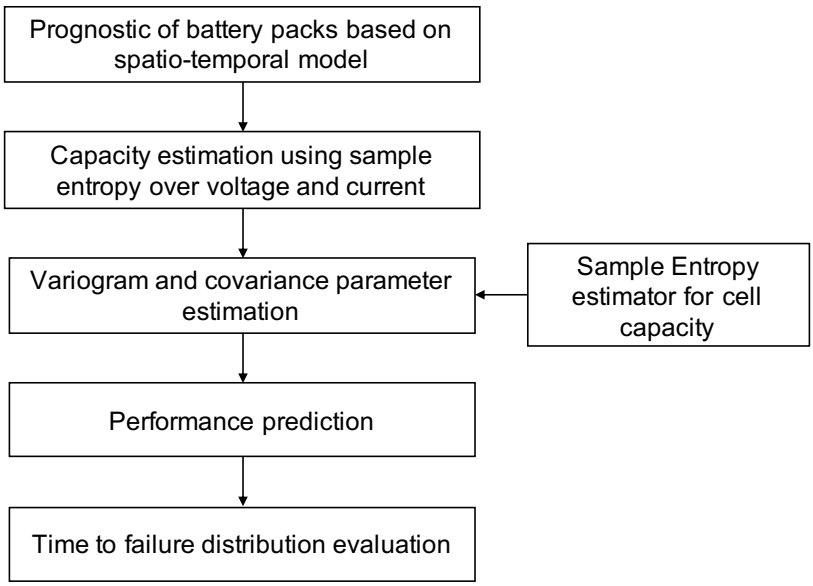

Figure 5.5: Scheme of prognostics based on the spatio-temporal model

conducted under the following parameters settings: 1) the constant current charge rate is 0.5C, 2) the cutoff voltage is 4.2 V, 3) The voltage is 4.2 for the constant voltage charge, 4) the discharge rate is 0.5C, and 5) the discharge cutoff voltage is 2.5V. Typical voltage and current during a charge/discharge cycle is shown in Figure 5.7. Figure 5.8 summarizes the experimental results where the observations of the tested battery pack over 125 cycles were recorded. Meanwhile, current and voltage changes of each cell in this pack are also monitored.

5.2.3.2 Estimation of capacity using the sample entropy-based method

Entropy is defined as the average amount of information contained in each message received. Approximate entropy (ApEn) and sample entropy (SamEn) are mathematical algorithms created to measure the repeatability or predictability within a time series. It is proven that SamEn is more accurate when the data length varies. For a N points time-series, $X_i^m = \{s(i), s(i+1), \ldots, s(i+m-1)\}$ can be obtained, where m is the run length. Then the distance d_{ij}^m between the vector X_i^m and X_j^m is defined as $d_{ij}^m = \max|s(i+k)-s(j+k)|$. The

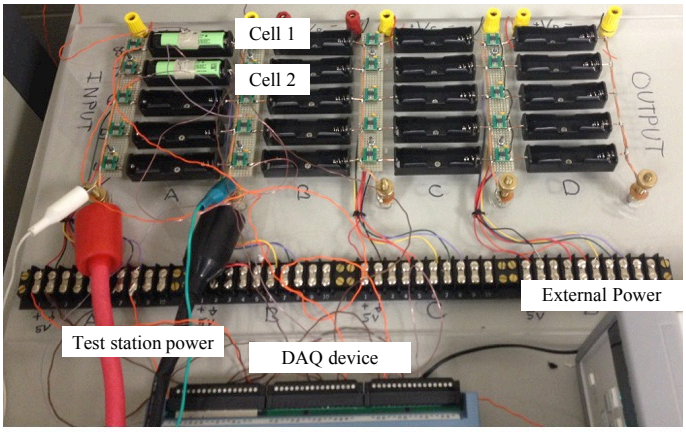

Figure 5.6: The battery pack with two parallel connected cells

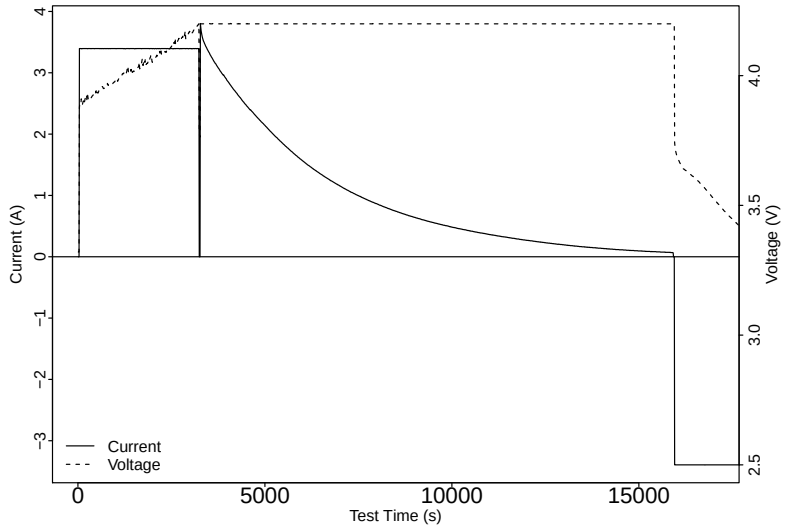

Figure 5.7: The voltage and current profile in a CCCV charge/discharge cycle

similarity degree D_{ij}^m between X_i^m and X_j^m is determined by a fuzzy membership function. That is, $D_{ij}^m = f_z(d_{ij}^m, r)$ where $f_z(\cdot)$ is the fuzzy membership function. Gaussian, Sigmoid and logistic functions are common choices of fuzzy functions. The Sigmoid function, $f_z(d_{ij}^m, r) = \dfrac{1}{1+\exp((d_{ij}^m + 0.5)/r)}$ is used in this research, d_{ij}^m is the distance between X_i^m and X_j^m, and r is a tolerance window. For each vector X_i^m, averaging all the similarity degree of its neighboring vectors X_j^m, $C_r^m(i) = 1/(N-m-1)\sum_(j=1, i \neq j)^{(N-m)} D_{ij}^m$

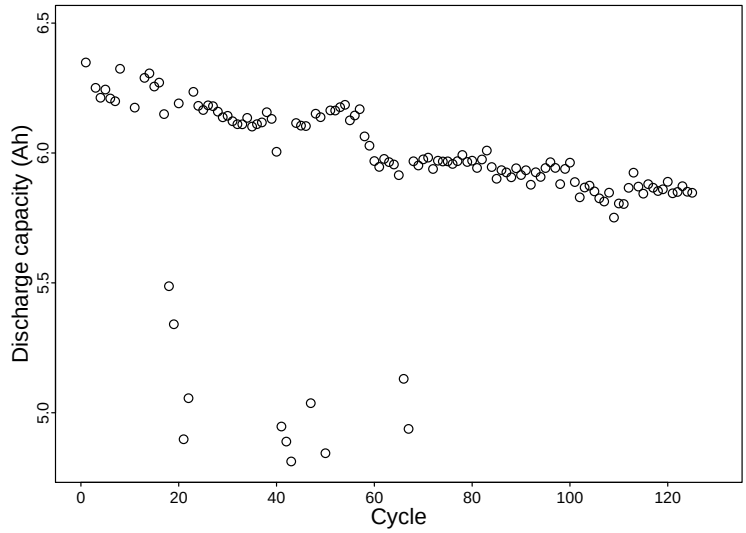

Figure 5.8: The capacity fade of tested battery pack over cycles

can be obtained and the cumulative probability $C_r^m = 1/(N-m)\sum_(i = 1)^{(N-m)}[C_r^m(i)]$. Define $C_r^{(m+1)}$ with the same logic of C_r^m. Then the sample entropy can be calculated as follows:

$$\text{SampEn}(r, m, N) = -\log[(C_r^{(m+1)})/(C_r^m)], \tag{5.21}$$

where r, m, and N are the tolerance window, the run length, and the total number of data points, respectively.

The sample entropy is not new in the lithium ion capacity assessment. There are several works applying sample entropy in assessing capacity, but the discussion over dynamic charge/discharge profile is missing. Moreover, the voltage time series is the only input of the SampEn calculation. For example, Sun et al. [127] analyzed the lead-acid battery state of health under the fully charge/discharge condition. Widodo, et al. [128] used sample entropy to analyze the capacity degradation mechanism where batteries are full charge and discharged. Hu et al. [201] fitted the regression model between the capacity and sample entropy of lithium ion batteries under various ambient temperatures. Considering the characteristics of voltage and current in a parallel connected battery pack, this research investigates the SampEn based on both voltage and current, and quantify the relationship

between the SampEn and capacity degradation. Figure 5.9 shows the SampEn over cycles. The general capacity SampEn-based estimator is proposed as follows,

$$Y = f(\text{SampEn}_v, \text{SampEn}_c) + \epsilon, \tag{5.22}$$

where SampEn_v and SampEn_c are the sample entropy based on voltage and current, respectively. $f(\cdot)$ is the function that represents the relationship of SampEn values and capacity. Without losing generality, polynomials function is selected $f(\cdot)$ in this research. Guo et al. [202] provided a $Y = 1.9899 - 0.2594 SampEn_v - 0.0012 SampEn_v^2 + 0.3112 SampEn_c$ based on the data single cells provided by NASA [22], which is used to predict the capacity in the battery pack.

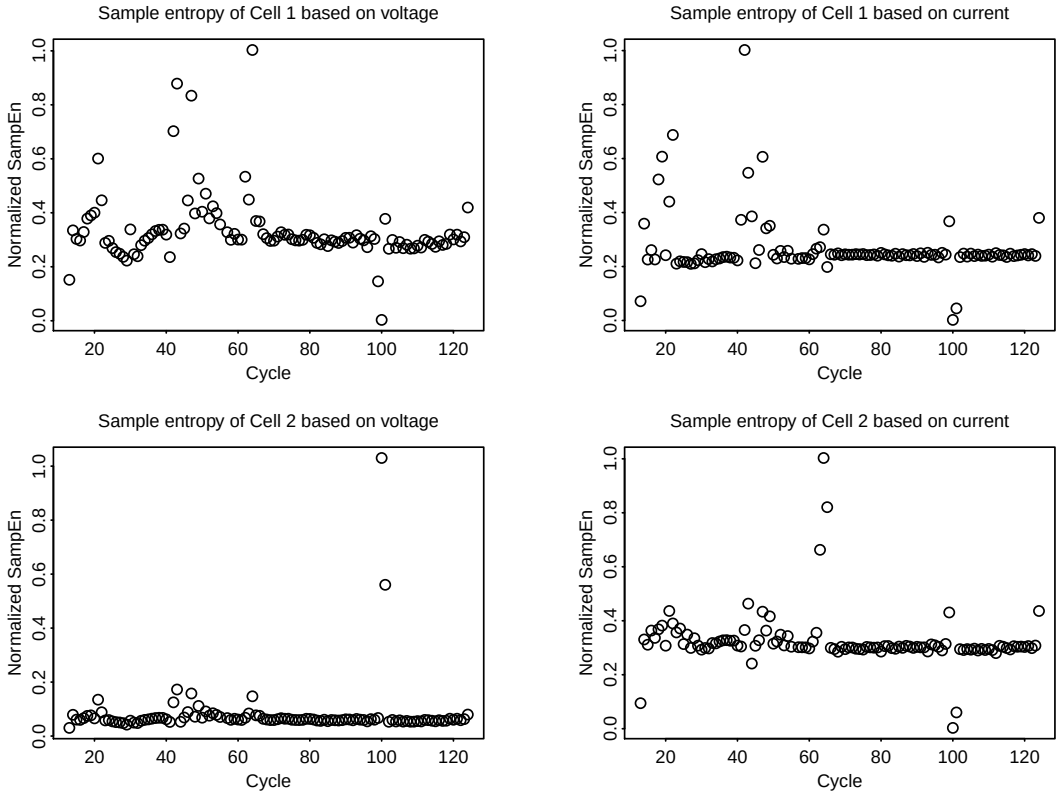

Figure 5.9: The sample entropy of cells in the tested battery pack

5.2.3.3 Prognostics of battery pack using the spatio-temporal model

The generalized robust variogram estimator, proposed by Cressie and Hawkins [203], $\hat{\gamma}(h;u) = \frac{[N(s;t)\sum_{N(s;t)}|Y(s_i;t)-Y(s_j;t')|^{1/2}]^4}{0.457+0.494/|N(s;t)|+0.045/|N(s;t)|^2}$, is used considering the existence of outliers. The covariance function with Gaussian spatial margins is selected. For $d = 2$, with the covariance function of Equation (5.15) and (5.18), the variogram can be defined as

$$\gamma(h;u|\theta) = \begin{cases} 0 & \| h \| = u = 0; \\ \sigma^2\Big[1 - \frac{1}{a|u|+1}\exp\Big(\frac{b^2 \| h \|^2}{a|u|+1}\Big)\Big] + r^2 + \alpha_1 \| h \|^{\alpha_2} & \text{otherwise} \end{cases}, \tag{5.23}$$

where $\theta = (a, b, \sigma^2, r)$ is the set of unknown parameters. Since there are two cells in the pack, $\| h \|$ is considered as 1 for the purpose of simplicity. By minimizing Equation (5.20), the covariance matrix parameter θ is $(0.2, 0.003, 0, 0, 4, 0.05)$. The covariance of the two cells decreases over various time intervals (see Figure 5.10).

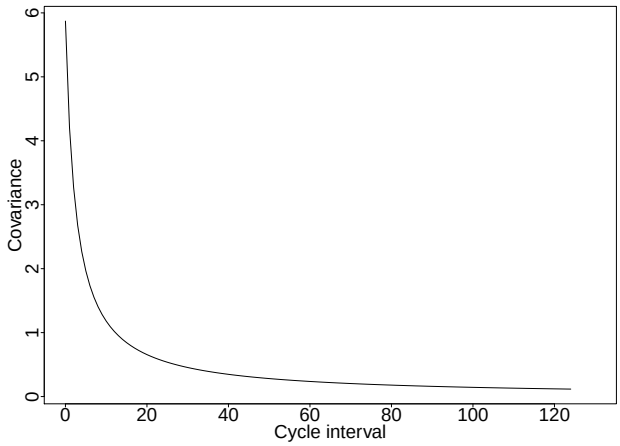

Figure 5.10: Covariance of two cells over the time with the spatial lag of 1

5.2.3.4 Conclusion and discussion

The spatio-temporal model is used to capture the spatial correlation and temporal uncertainties. The correlation of cells in a battery pack is considered as the function of "spatial" variables. A pack of two-parallel cells is used to validate the proposed method. Stationary covariance function with Gaussian related margins is employed, whose parameters are estimated based on the robust empirical variogram. Due to the complex correlation between cells and lack of direction measurements of capacity, the sample entropy-based estimator is used to approximate the capacity of each cell. With the approximated capacity of cells, the covariance function of two cells is estimated through the weighed least square method. Assuming the spatial distance of two-parallel connected cells as 1, the covariance function of cells shows that correlation of two cells decreases with large temporal intervals. In the future, prognostics based on spatio-temporal model for more complex systems will be investigated. Considering dynamic models and processes are encountered in many case, non-stationary covariance function will be investigated, which can the represent time-vary decay rate of the degradation process.

5.3 Performance evaluation

The above-proposed methods as well as those in Chapters 3 and 4 are proposed to solve various challenges in prognostics modeling using repeated measurements of multiple units. The relationship between the proposed method is shown in Figure 5.11. The general mixed-effect regression model in Chapter 3 aims to quantify between-unit variability through random effects, but there are the computational difficulties in the existence of unbalanced data. Joint modeling method and mean-covariance decomposition method in Chapter 4 attempts to solve the hamper in prognostics modeling using unbalanced data by assuming the degradation process as a multivariate normal distribution based on the general mixed-effect regression model. To process large volumes of data such as dense observations over time, functional principal component analysis (FPCA)-based regression is introduced in prognostics modeling, where covariances are modeled as a linear combina-

tion of eigenfunctions. Compare with other prognostics modeling techniques, the proposed models aim to model the degradation mechanism and estimate the remaining useful life (RUL) with the focus of within-between-unit uncertainty quantification. Without the assumption of independence between multiple test units, spatio-temporal models investigate both "spatial" and temporal trends and uncertainties. Despite the differences in modeling techniques, those methods above share the goal of modeling the degradation process and evaluating time to failure. This section serves as a performance comparison of the proposed methods regarding degradation prediction accuracy and time to failure distribution.

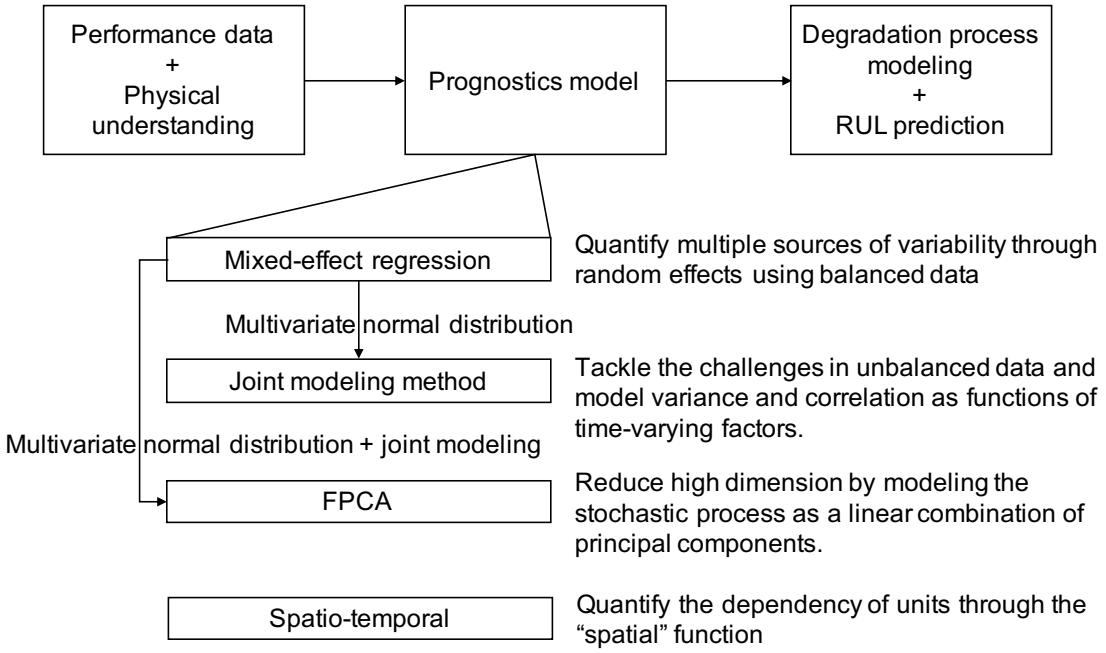

Figure 5.11: Relationship between proposed methdos

Numerical simulation is needed to predict degradation and estimate the time to failure distribution. The simulation of stochastic processes based on mixed-effect regression models can be implemented via generating random numbers from the parameter distribution shown in Equation (3.17). Both joint modeling method and functional principal component regression have the same assumption of Gaussian degradation process. The simple kriging predictor shown in Equation (4.22) can be used for performance prediction. For

lithium ion battery capacity, degradation prediction and cycle to failure evaluation can be implemented through

1. Estimate unknown parameters in the proposed methods.

2. Simulate the capacity from the determined models at Step 1

    - for mixed-effect regression model: (1) simulate parameter distribution (Equation (3.17)), (2) calculate the capacity at cycle t based on the parameters generated above, and (3) repeat the above two steps
    - for joint modeling method and FPCA-based method: (1) estimate the distribution of capacity at cycle t through the simple kriging predictor (Equation (4.22)) and (2) simulate capacity based on the distribution in (1)

3. Evaluate cycle to failure distribution using the result of Step 2

    - for mixed-effect regression model: calculate cycle to failure using the simulated parameters in Step 2 and given failure threshold
    - for joint modeling method and FPCA-based method: calculate the probability of failure at cycle t based on the estimated distribution in Step 2

Figure 5.12 presents the capacity prediction accuracy comparison of proposed methods. Predictions based on mixed-effect regression and joint modeling method are close to each other because that when the input of data is balanced, these two methods are consistent. Because the joint modeling method can well deal with unbalanced data, measurements of all batteries are used in the joint modeling method while only the first 549 measurements are used in the mixed-effect model. Therefore, in the late stage, the confidence interval of the joint modeling method is smaller than that of the mixed-effect model. Compared with the above two methods, FPCA-based and mean-covariance decomposition (MCD) methods model the mean and covariance separately, which leads to accurate mean prediction with large variations.

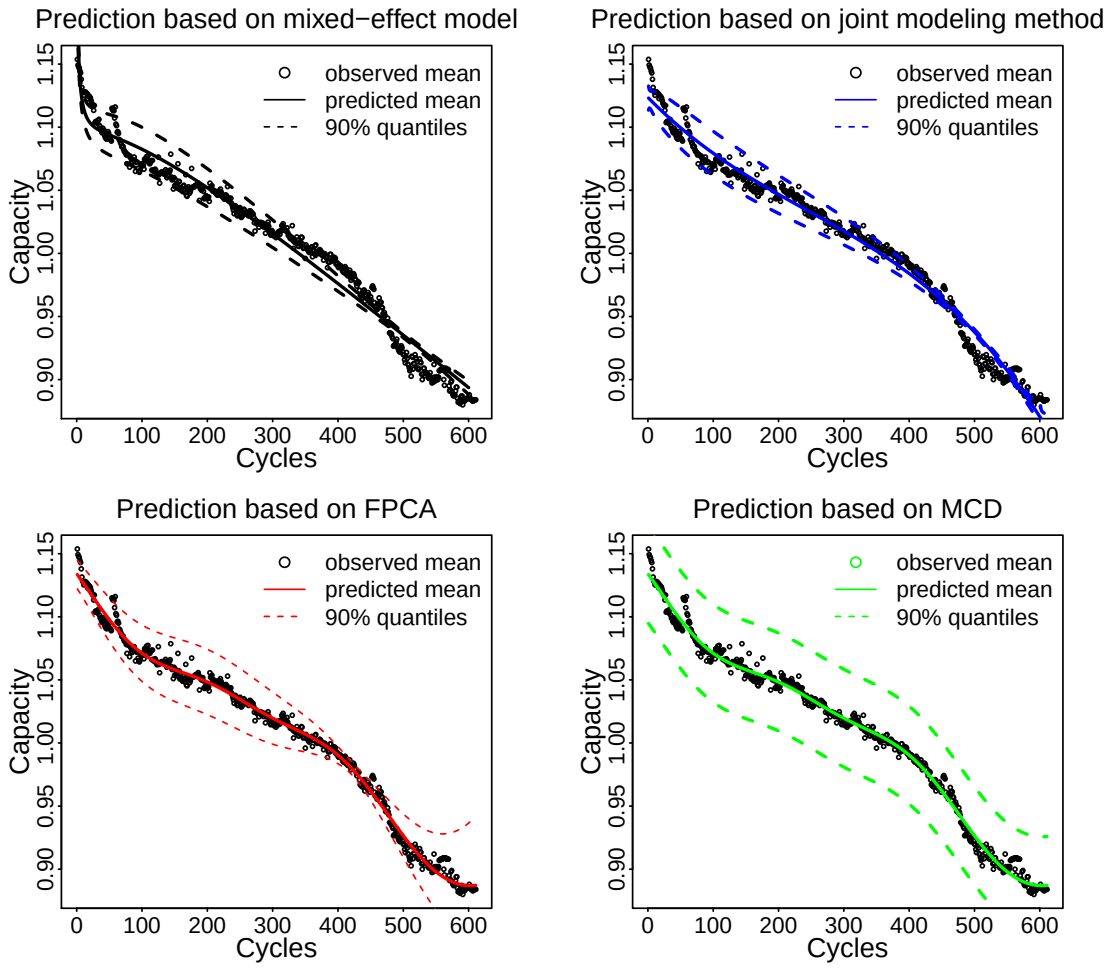

Figure 5.12: Comparison of capacity prediction accuracy

Table 5.1: Mean and standard deviation of cycle to failure based on various prognostics models

	Mixed-effect model	Joint modeling method	FPCA-based method	MCD-based method
Mean	635.2	634.6	635	637 .0
Standard deviation	11.2	10.3	15.3	29.4

Cycles to failure distribution, estimated based on the capacity prediction, have the same characteristics (see Figure 5.13). The mean and standard deviation of cycle to failure estimated based on mixed-effect are shown in Table 5.1. The observed average cycle to failure

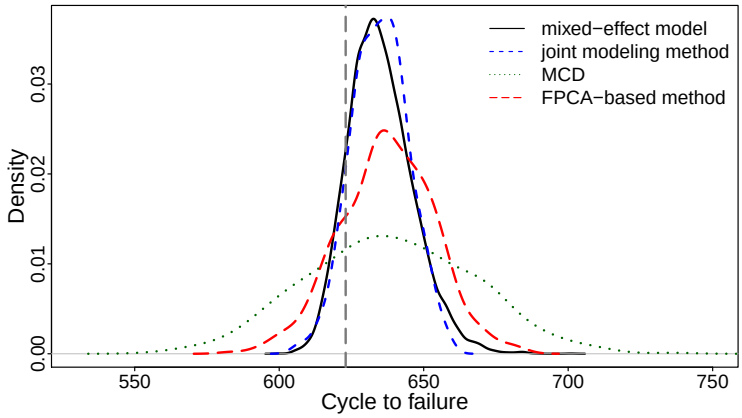

Figure 5.13: Comparison of cycle to failure distribution

over the four batteries is 623 less than the estimated cycle to failure because battery CS2 36 fails at the cycle of 549 and other batteries fails at the cycle over 620. When more samples of batteries are available, the estimated mean and observed mean are closed to each other.

6. CONCLUSION AND FUTURE WORK

The fundamental problem that this research addresses is the prognostics modeling with repeated measurement data of multiple units. In the previous chapters, statement of problem, literature review, proposed methods, and battery capacity prognostics application are discussed. This final chapter summarized the main contributions and discuss the future research work.

6.1 Summary

Chapter 1 introduces the major research problems. As the key process of prognostics and health management (PHM) and predictive maintenance (PdM), prognostics aims to predict the system's remaining useful lifetime (RUL). There are four major uncertainties in prognostics modeling: input, model, measurement, and operational uncertainties. Repeated measurement studies are designed to characterize the input uncertainties. Challenges along with the repeated measurement data require advanced modeling techniques so that model errors can be reduced. Four research objectives are proposed, including 1) to quantify multiple sources of variability, 2) to deal with complex data structure, 3) to explore general spatio-temporal model for repeated measurement data, and 4) to develop robust algorithms for prognostics models.

Chapter 2 summarizes the existing prognostics approaches and their application in prognostics with repeated measurement data. Based on the literature review, research gaps are identified. Prognostics approaches are classified as data-driven, physics-based, and hybrid approaches, according to the usage of physical knowledge. There are limited works in prognostics modeling with repeated measurement. This research focuses on developing adequate models in this area.

Chapter 3 illustrates the general mixed-effect model to tackle the challenges along with repeated measurement data. This research first uses the general linear mixed-effect model to quantify the multiple sources of variability. The incorporation of physical knowledge of the degradation mechanism is done through covariates identification and model selection.

Chapter 4 presents the joint modeling method based prognostics. Due to the drawback of general linear mixed-effect models in unbalanced data, joint modeling method (JMM)-based prognostics assumes the measurements of multiple units over time as a multivariate Gaussian distribution. Trigonometric function is used for the unconstrained parameteration of the correlation matrix.

Chapter 5 introduces two other prognostics modeling methods, functional principal component analysis (FPCA) and spatio-temporal modeling-based prognostics. FPCA is used to solve the computation issue of aforementioned methods due to the high dimensionality of repeated measurement data. Based on the above methods, a general spatio-temporal model is proposed to quantify unit-specific variation, temporal variation, and unit-temporal variation. Moreover, the performance of proposed models is evaluated in terms of prediction accuracy, adaptability, and computational complexity.

In Chapters 3 to 5, the application of proposed methods in battery capacity prognostics is used for the validation of the proposed methods. The proposed methods are used to investigate battery capacity fade over repetitive cycles by considering both within-battery and between-battery variations. Physics-based covariates are integrated with functional forms for modeling the capacity fade. A systematic approach based on covariate identification, model selection, and a strategy for prognostics data selection is presented for all these proposed methods. The results of preliminary work show that the proposed methods can improve the prediction accuracy and can also be used to understand the battery capacity degradation mechanism.

6.2 Future work

The discussions at Chapters 2, 3, 4, and 5 have addressed the future research directions. This section summarizes those and suggests other potential opportunities.

The major goal of prognostics modeling is to predict the remaining useful life. In the case of repeated measurements of multiple units, modeling multiple sources of uncertainties in prognostics becomes a challenge with the complex data structures, such as large scale data sets, unbalanced data, heterogeneous measurements, etc. As discussed in previous chapters, it would be necessary to develop robust methods that can perform well in various cases. On this subject, future work could be done in the following directions.

1. Hybrid methods of physics/model-based and data-driven methods are widely used for prognostics due to their strength inheriting from both data-driven and physics-based models, which can help to improve predict accuracy and model interpretation. However, existing hybrid methods used in repeated measurement data, mainly regarding the development of data-driven techniques and physical understanding are relatively less incorporated. In the future, fusion of physics-based models and cutting-edge data-driven methods, such as artificial intelligence (AI) method, became one of important topics in prognostics such as the integration material computational models and deep learning methods.

2. The general spatio-temporal model can effectively model and effectively quantify the multiple sources of variability, especially when the "spatial" dependency exists. Development of nonstationary convariance functions and their parameter estimations are top challenges for highly dynamic degradation process. The generalization of spatio-temporal models using repeated multiple units can be used in engineering systems prognostics when sensing signals of correlated components are available. Moreover, non-Gaussian models and processes are encountered in many natural and applied science fields. Constructing non-Gaussian process with given correlation structures will be an interesting research area.

3. Online monitoring and prognostics is an emerging trend. Prognostics modeling with online data is expected to be adaptive. When recovery or maintenance actions are taken, the system degradation state estimation scheme based on the health state of sub-

systems/components should be updated automatically. Moreover, storage and analysis of large-scale degradation data are also challenges. The concept of moving-window can be promoted to deal with online prognostic modeling. Sensing data can be available in forms of complex data structures, such as few data and unbalanced or missing data. Bayesian inference attracts growing attentions. For the case of few data or no data available, physics-based models can provide prior knowledge for the degradation process. However, how to validate the expertise judgment remains a challenge. The major challenge in prognostics modeling over unbalanced/missing data is to incorporate all information of multiple subjects.

References

[1] R. Gouriveau, M. Kamal, and Z. Noureddine,“ From prognostics and health systems management to predictive maintenance 1: monitoring and prognostics,” John Wiley & Sons, 2016.

[2] Joint Strike Fighter Program Office (2016) Joint strike fighter PHM vision, 2016.

[3] J. Lee, J. Ni, D. Djurdjanovic,H. Qiu, and H. Liao, “Intelligent prognostics tools and emaintenance,” Computers in Industry, vol. 57, pp. 476-489.

[4] S. Uckun, K. Goebel, and P. Lucas, “Standardizing research methods for prognostics, International Conference on Prognostics and Health Management,” Denver, CO, USA, 2008.

[5] K. Goebel, A. Saxena, M. Daigle, J. Celaya, and I. Roychoudhury, “Introduction to prognostics,” European PHM conference, 2012.

[6] A. Hess, J. Stecki, and S. Rudov-Clark, “The maintenance aware design environment: Development of an aerospace phm software tool,” Annual Conference on the Prognostics and Health Management, 2008.

[7] E. Zio and F. Maio, “A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system,” Reliability Engineering and System Safety, vol. 95, pp. 49-57, 2010.

[8] N. Clements, “Introduction to prognostics,” Annual Conference of the Prognostics and Health Management Society, 2011.

[9] ISO 13381-1, “Condition monitoring and diagnostics of machines Prognostics Part 1: General guidelines,” International Standards Organization, 2015.

[10] M. Pecht, “Prognostics and health management of electronics,” John Wiley & Sons, 2008.

[11] N. Kim, D. An, and J. Choi, “Prognostics and health management of engineering system,” Springer, Switzerland, 2017.

[12] Y. Peng, Ming Dong, and M. Zuo, “Current status of machine prognostics in condition-based maintenance: a review, International Journal of Advanced Manufacture Technology,” vol. 50, pp. 297-313, 2010.

[13] Y. Ying, Y. Cao, S. Li, J. Li, and J. Guo, “Study on gas turbine engine fault diagnostic approach with a hybrid of gray relation theory and gas-path analysis,” Advances in Mechanical Engineering, vol. 8, pp. 1-14, 2016.

[14] J. Celaya, A. Saxena, and K. Goebel, “Uncertainty representation and interpretation in model-based prognostics algorithms based on Kalman Filter estimation,” Annual Conference on the Prognostics and Health Management Society, 2012.

[15] S. Sankararaman and K. Goebel, "Uncertainty in Prognostics and Systems Health Management," International Journal of Prognostics and Health Management, vol. 6, pp. 1-14, 2015.

[16] A. Hedayat and K. Afsarinejad, "Repeated measurement designs, I, A Survey of Statistical Design and Linear Models," North-Holland, Amsterdam, 1975.

[17] A. Hedayat and K. Afsarinejad, "Repeated measurement design, II, The Annuals of Statistics," vol. 6, pp. 619-628, 1978.

[18] D. Hedeker, and R. Gibbons, "Longitudinal data analysis, Wiley Series in Probability and Statistics," Wiley & Sons, New York, NY, 2006.

[19] B. Bebout, L. Profert-Bebout, E. Fleming, A. Detweiller, and K. Goebel, "Algae raceway data set," NASA Ames Prognostics Data Repository.

[20] A. Agogino and K. Goebel, "Mill data set. BEST lab, UC Berkeley," NASA Ames Prognostics Data Repository.

[21] University of Cincinnati, "Bearing data set," NASA Ames Prognostics Data Repository.

[22] B. Saha and K. Goebel, "Battery data set," NASA Ames Prognostic Data Repository.

[23] A. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical System and Signal Processing," vol. 20, pp. 1483-1510, 2006

[24] M. Schwabacher and K. Goebel, "A survey of artificial intelligence for prognostics," AIAA Fall symposium, 2007.

[25] A. Heng, S. Zhang, A. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," Mechanical Systems and Signal Processing, vol. 23, pp. 724-739, 2009.

[26] K. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and Health Management: A Review on Data Driven Approaches," Mathematical Problems in Engineering, vol. 2015, 2015.

[27] X. Si, W. Wang, C. Hu, and D. Zhou, "Remaining useful life estimation- A review on the statistical data driven approaches," European Journal of Operational Research, vol. 213, pp. 1-14, 2011.

[28] C. Okoh, R. Roy, J. Mehnen, and L. Redding, "Overview of remaining useful life prediction techniques in through-Life engineering services," the 6th CIRP Conference on Industrial Product-Service Systems, 2014.

[29] H. Elatter, H. Elminir, and A. Riad, "Prognostics: a literature review," Complex & Intelligence Systems, vol. 2, pp. 125-154, 2016.

[30] O. Eker,F. Camci, and I. Jennions, "Major challenges in prognostics: Study on benchmarking prognostics datasets," PHM Conference in Europe, 2012.

[31] D. An, N. Kim, and J. Choi, "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews," Reliability Engineering and System Safety, vol. 133, pp. 223-236, 2015.

[32] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, "Intelligent fault diagnosis and prognosis for engineering systems," Wiley, 2006.

[33] K. Javed, "A robust & reliable Data-driven prognostics approach based on extreme learning machine and fuzzy clustering," PhD Thesis, The University of Franche-Comté, 2014.

[34] O. Dragomir, R. Gouriveau, F. Dragomir, E.Minca, and N. Zerhouni, "Review of prognostic problem in condition-based maintenance," European Control Conference, ECC09, 1585-1592, Budapest, Hungry, 2009.

[35] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," Journal of Intelligent Manufacturing, vol. 23, pp. 227-237, 2012.

[36] Y. Fan and C. Li, "Diagnostic rule extraction from trained feedforward neural networks," Mechanical Systems and Signal Processing vol. 16, pp. 1073-1081, 2002.

[37] N. Gebraeel and M. Lawley, "A neural network degradation model for computing and updating residual life distributions," IEEE Transaction Automatic Science and Engineering, vol. 5, pp. 387-401, 2008.

[38] J. Liu, A. Saxena, K. Goebel, B. Saha, and W. Wang, "An adaptive recurrent neural network for remaining useful life prediction of Lithium-ion batteries," Annual Conference of the Prognostics and Health Management Society, 2010.

[39] C. Zhang, J. Jiang, W. Zhang, Y. Wang, S. Sharkh, and R. Xiong, "A novel data-driven fast capacity estimation of spent electric vehicle lithium-ion batteries," Energies, vol. 7, pp. 8076-8094, 2014.

[40] R. Chinnam, and P. Mohan, "Online reliability estimation of physical systems using neural networks and wavelets," International Journal of Smart Engineering System Design, vol. 4, pp. 253-264, 2002.

[41] S. Amin, C. Byington, and M. Watson, "Fuzzy Inference and Fusion for Health State Diagnosis of Hydraulic Pumps and Motors," Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society, 2005.

[42] A. Volponi, "Data fusion for enhanced aircraft engine prognostics and health management," NASA Contractor Report CR-2005-214055.

[43] R. Silva, R. Gouriveau, S. Jemei, D. Hissel, L. Boulon, K. Agbossou, and N. Yousfi Steiner, "Proton exchange membrane fuel cell degradation prediction based on adaptive neuro-fuzzy inference systems," International Journal of Hydrogen Energy, vol. 39, no. 21, pp. 11128-11144, 2014.

[44] B. Zupan, J. Demsar, M. Kattan, J. Beck and I. Bratko, "Machine learning for survival analysis: a case study on recurrence of prostate cancer," Artificial Intelligence in Medicine, vol. 20, no. 1, pp. 59-75, 2000.

[45] C. Xie, D. Yang, Y. Huang, and D. Sun, "Feature extraction and ensemble decision tree classifier in plant failure detection," Annual Conference of the Prognostics and health management society, 2015.

[46] N. Ibrahim and A. Kudus, "Decision tree for prognostic classification of multivariate survival data and competing risks," Recent Advances in Technologies, InTech, DOI: 10.5772/7429, 2009.

[47] M. Verduijn, N. Peek, P. Rosseel, E. de Jonge, and B. de Mol, "Prognostic Bayesian networks: I: rationale, learning procedure, and clinical use," Journal of Biomedical Informatics, vol. 40, pp. 609-618, 2007.

[48] M. Verduijn, N. Peek, P. Rosseel, E. de Jonge, and B. de Mol, "Prognostic Bayesian networks: II: an application in the domain of cardiac surgery," Journal of Biomedical Informatics, vol. 40, pp. 619-630, 2007.

[49] K. McNaught and A. Zagorecki, "Using dynamic Bayesian networks for prognostic modelling to inform maintenance decision making," IEEE International Conference on Industrial Engineering and Engineering Management, December 8-11, 2009.

[50] K. Medjaher, J. Moya, and N. Zerhouni, "Failure prognostic by using dynamic Bayesian networks," Proceedings of IFAC, vol. 42, no.5, pp. 257-262, 2009.

[51] S. Ferrerio, A. Arnaiz, B. Sierra, and I. Irigoien, "Application of Bayesian networks in prognostics for a new integrated vehicle health management concept," Expert Systems with Applications, vol. 39, pp. 6402-6418, 2012.

[52] A. Nuhic, T. Terzimehic, T. Soczka-Guth, M. Buchholz, and K. Dietmayer, "Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods," Journal of Power Sources, vol.239, pp. 680-688, 2013.

[53] S. Yang, C. Liu, X. Zhou, W. Liang, and Q. Miao, "Investigation on data-driven life prediction methods," International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012.

[54] H. Huang, H. Wang, Y. Li, and Z. Liu, "Support vector machine based estimation of remaining useful life: current research status and future trends," Journal of Mechanical Science and Technology, vol. 29, pp. 151-163, 2015.

[55] J. Shi, B. Wang, R. Murray-Smith, and D. Titterington, "Gaussian process functional regression modeling for batch data," Biometrics, vol. 63, pp. 714-723, 2007.

[56] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, "Prognostics for state of health estimation of lithium ion batteries based on combination Gaussian process functional regression," Microelectronics Reliability, vol. 53, pp. 832-839, 2013

[57] Q. Zhou, J. Son, S. Zhou, X. Mao, and M. Salman, "Remaining useful life prediction of individual units subject to hard failure," IIE Transactions on Quality and Reliability Engineering, vol. 46, pp. 1017-1030, 2014.

[58] R. Kontar, J. Son, S. Zhou, C. Sankavaram, Y. Zhang, and X. Du, "Remaining useful life prediction based on the mixed effects model with mixture prior distribution," IISE Transactions, vol. 49, pp. 1-16, 2016.

[59] J. Guo, Z. Li, and M. Pecht, "A Bayesian approach for Li-Ion battery capacity fade modeling and cycles to failure prognostics," Journal of Power Sources, vol. 281, pp. 173-184, 2015.

[60] X. Zhang and P. Pisu, "An unscented Kalman filter based approach for the health-monitoring and prognostics of a polymer electrolyte membrane fuel cell," Annual Conference of the Prognostics and Health Management Society, 2012.

[61] E. Jantunen, "Prognosis of rolling bearing failure based on regression analysis and fuzzy logic," Proceedings of the VETOMAC-3 and ACSIM-2004, New Delhi, India, pp. 836-846, 2004.

[62] C. Rasmussen and K. Williams, "Gaussian process for machine learning," MIT Press, 2006.

[63] M. Davidian and D. Giltinan, "Nonlinear models for repeated measurement data," CRC, Boston, 1995.

[64] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive prognostics for rolling element bearing condition," Mechanical Systems and Signal Processing, vol. 13, pp. 103-113, 1999.

[65] Y. Li, T. Kurfess, and S. Liang, "Stochastic prognostics for rolling element bearings," Mechanical Systems and Signal Processing, vol. 14, pp. 747-762, 2000.

[66] C. Li and H. Lee, "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics," Mechanical Systems and Signal Processing, vol. 19, pp. 836-846, 2005.

[67] A. Ray and S. Tangirala, "Stochastic modeling of fatigue crack dynamics for on-line failure prognostics," IEEE Transactions on Control Systems Technology, vol. 4, pp. 443-451, 1996.

[68] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," Mechanical Systems and Signal Processing, vol. 60-61, pp. 252-272, 2015.

[69] V. Nistane and S. Harsha, "Failure evaluation of ball bearing for prognostics," Procedia Thechnology, vol. 23, pp. 179-186, 2016.

[70] G. Kacprzynski, A. Sarlashkar, M. Roemer, A. Hess, and B. Hardman, "Predicting remaining life by fusing the physics of failure modeling with diagnostics," JOM, vol. 56, pp. 29-35, 2004

[71] T. Addabbo, A. Fort, R. Garbin, M. Mugnaini, S. Rocchi, and V. Vignoli, "Theoretical characterization of a gas path debris detection monitoring system based on electrostatic sensors and charge amplifiers," Measurement, vol. 64, pp. 138-146, 2015.

[72] J. Qiu, B. Seth, S. Liang, and C. Zhang, "Damage mechanism approach for bearing lifetime prognostics," Mechanical System and Signal Processing, vol. 16, pp. 817-829, 2002.

[73] W. He, W. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," Journal of Power Sources, vol. 196, pp. 10314-10321, 2011.

[74] M. Jouin, R. Gouriveau, D. Hissel, M. Pera, and N. Zerhouni, "Particle filter-based prognostics: review, discussion and perspectives," Mechanical Systems and Signal Processing, vol. 72-73, pp. 2-31, 2016.

[75] S. Kumar, M. Torres, Y. Chan, and M. Pecht, "A hybrid prognostics methodology for electric products," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, June 1-8, 2008.

[76] S.Cheng and M. Pecht, "A fusion prognostics method for remaining useful life prediction of eletronic products," 5th Annual IEEE Conference on Automation Science and Engineering, Bangalore, India, August 22-25, 2009.

[77] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar, and M. Pecht, "Model-based and data-driven Prognosis of Automotive and Electronic Systems," IEEE International Conference on Automation Science and Engineering, Bangalore, India, September 9, 2009.

[78] L. Liao and F. Köttig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery Life prediction," IEEE Transactions on Reliability, vol. 63, pp. 191-207, 2014.

[79] M. Orchard and G. Vachtsevanos, "A particle filtering approach for on-line failure prognosis in a planetary carrier plate," International Journal of Fuzzy Logic Intelligence System, vol. 7, pp. 221-227, 2007.

[80] H. Zhang, R. Kang, and M. Pecht, "A hybrid prognostics and health management approach for condition-based maintenance," International Conference on Industrial Engineering and Engineering Management, pp. 1165-1169, 2009.

[81] C.Chen, G.Vachtsevanos, and M.Orchard, "Machine remaining useful life prediction: an integrated adaptive neuro-fuzzy and high-order particle filtering approach," Mechanical System and Signal Process, vol. 28, pp. 597-607, 2011.

[82] J. Liu, W. Wang, F. Ma, Y. Yang, and C. Yang, "A data-model-fusion prognostic framework for dynamic system state forecasting," Engineering Application and Artificial Intelligence, vol. 25, pp. 814-823, 2012.

[83] Y. Lei, N. Li, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," IEEE Transactions on Reliability, vol. 65, pp. 1314-1326, 2016.

[84] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," IEEE Transactions on Industrial Electronics, vol. 62, pp. 7762-7773, 2015.

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

[86] N. Laird and J. Ware, "Random-effects models for longitudinal data," Biometrics, vol. 38, pp. 963-974, 1982.

[87] A. Cnaan, N. Laird, and P. Slasor, "Tutorial in Biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data," Statistics in Medicine, vol. 16, pp. 2349-2380, 1997.

[88] D. Bates, M. Machler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," arXiv preprint arXiv:1406.5823.

[89] R-CARN, http://cran.r-project.org/

[90] Y. Fan and R. Li, "Variable selection in linear mixed effects models," The Annals of Statistics, vol. 40, pp. 2043-2068, 2012.

[91] J. Guo, Z. Li, T. Keyser, and Y. Deng, "Modeling Li-ion battery capacity fade using designed experiments," Proceedings of the 2014 Industrial and Systems Engineering Research Conference, Montreal, CA, 2014.

[92] J. Ryoo, "Model selection with the linear mixed effects model for longitudinal data," PhD Thesis, The University of Minnesota, 2010.

[93] M. Pinson, and M. Bazant, "Theory of SEI formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction," Journal of The Electrochemical Society, vol. 160, pp. A243-A250, 2013.

[94] G. Bartram and S. Mahadevan, "Prognostics and health monitoring in the presence of heterogeneous information," Annual Conference of the Prognostics and Health Management Society, 2012.

[95] J. Fergus, "Recent developments in cathode materials for lithium ion batteries," Journal of Power Sources, vol. 195, pp. 939-954, 2010.

[96] S. Kabitz, J. Gerschler, M. Ecker, Y. Yurdagel, B. Emmermacher, D. Andre, T. Mitsch, and D. Sauer, "Cycle and calendar life study of a graphite $LiNi_{1/3}Mn_{1/3}Co_{1/3}O_2$ Li-ion high energy system. Part A: Full Cell Characterization," Journal of Power Sources, vol. 239, pp. 572-583, 2013.

[97] M. Doyle, T. Fuller, and J. Newman, "Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion Cell," Journal of The Electrochemical Society, vol. 140, pp. 1526-1533, 1993.

[98] S. Lee, H. Cui, M. Rezvanizaniani, and J. Ni, "Battery prognostics: SOC and SOH prediction," Proceedings of the ASME 2012 International Manufacturing Science and Engineering Conference, 2012.

[99] S. Peteson, J. Apt, and J. Whitacre, "Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization," Journal of Power Sources, vol. 195, pp. 2385-2392, 2010.

[100] R. Jungst, G. Nagasubrahanian, H. Case, A. Liaw, T. Paez, and D. Doughty, "Accelerated calendar and pulse life analysis of Lithium-ion cells," Journal of Power Sources, vol. 119/121, pp. 870-873, 2003.

[101] E. Thomas, I. Bloom, J. Christophersen, and V. Battaglia, "Statistical method for predicting the life of lithium-ion cells via accelerated degradation testing," Journal of Power Sources, vol. 184, pp. 312-317, 2008.

[102] M. Broussely, P. Biensan, F. Bonhomme, P. Blanchard, S. Herreyre, K. Nechev, and R. Staniewicz, "Main aging mechanisms in Li ion batteries," Journal of Power Sources, vol. 146, pp. 90-96, 2005.

[103] T. Utsunomiya, O. Hatozaki, N. Yoshimoto, M. Egashira, and M. Morita, "Influence of particle size on the self-discharge behavior of graphite electrodes in lithium-ion batteries," Journal of Power Sources, vol. 196, pp. 8675-8682, 2011.

[104] H. Wu, Z. Guo, H. Wen, and M. Yang, "Study the fading mechanism of $LiMn_2O_4$ battery, with spherical and flake type graphite as anode material," Journal of Power Sources, vol. 146, pp. 736-740, 2005.

[105] G. Sikha, P. Ramadass, B. Haran, R. White, and B. Popov, "Comparison of the capacity fade of Sony US 18650 cells charged with different protocols," Journal of Power Sources, vol. 122, pp. 67-76, 2003.

[106] B. Scrosati and J. Garche, "Lithium Batteries: Status, Prospects and Future," Journal of Power Sources, vol. 195, pp. 2419-2430, 2010.

[107] M. Wohlfahrt-Mehrens, C. Vogler, and J. Garche, "Aging mechanisms of lithium cathode materials," Journal of Power Sources, vol. 127, pp. 58-64, 2004.

[108] L. Ji, H. Zheng, A. Ismach, Z.Tan, S. Xun, E. Lin, V. Battaglia, V. Srinivasan, and Y. Zhang, "Graphene/Si multilayer structure anodes for advanced half and full lithium-ion cells," Nano Energy, vol. 1, pp. 164-171, 2012.

[109] M. Broussely, S. Herreyre, P. Biensan, P. Kasztejna, K. Nechev, and R. Staniewicz, "Aging mechanism in Li-ion cells and calendar life predictions," Journal of Power Sources, vol. 97/98, pp. 13-21, 2001.

[110] P. Liu, P. J. Wang, J. Hicks-Garner, E. Sherman, S. Soukiazian, M. Verbrugge, H. Tataria, J. Musser, and P. Finamoure, "Aging mechanism of $LiFePO_4$ batteries deduced by electrochemical and structural analysis," Journal of Electrochemical Society, vol. 157, pp. A499-A507, 2010.

[111] K. Striebel, J. Shim, E. Cairns, R. Kostecki, Y. Lee, J. Reimer, T. Richardson, P. Ross, X. Song, and G. Zhuang, "Diagnostic analysis of electrodes from high-power lithium-ion cells cycled under different conditions," Journal of Electrochemical Society, vol. 151, pp. A857-A866, 2004.

[112] A. Smith, J. Burns, D. Xiong, and J. Dahn, "Interpreting high precision Coulometry results on Li-ion cells," Journal of Electrochemical Society, vol. 158, pp. A1136-A1142, 2011.

[113] J. Dai, D. Das, M. Ohadi, and M. Pecht, "Reliability risk mitigation of free air cooling through prognostics and health management," Applied Energy, vol. 111, pp. 104-112, 2013.

[114] T. Hansen and C. Wang, "Support vector based battery state of charge estimator," Journal of Power Sources, vol. 141, pp. 351-358, 2005.

[115] Q. Shi, C. Zhang, and N. Cui, "Estimation of battery state of charge using v-support vector regression algorithm," International Journal of Automotive Technology, vol. 9, pp. 759-764, 2008.

[116] B. Pattipati, C. Sankavaram, and K. Pattipati, "System identification and estimation framework for pivotal automotive battery management system characteristics," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, pp. 1-16, 2011.

[117] B. Saha, K. Goebel, S. Poll, and J. Christophersen, "Prognostic methods for battery health monitoring using a Bayesian framework," IEEE Transactions on Instrumentation and Measurement, vol. 58, pp. 291-296, 2009.

[118] C. Hu, G. Jain, C. Schmidt, C. Strief, and M. Sullivan, "Online estimation of lithium-ion battery capacity using sparse Bayesian learning," Journal of Power Sources, vol. 289, pp. 105-113, 2015.

[119] C. Hu, G. Jain, P. Zhang, C. Schmidt, P. Gomadam, and T. Gorka, "Data-driven approach based on particle swarm optimization and k-nearest neighbor regression for estimating capacity of lithium-ion battery," Applied Energy, vol. 129, pp. 49-55, 2014.

[120] W. Shen, C. Chan, E. Lo, and K. Chau, "A new battery available capacity indicator for electric veihles using neral network," Energy Conversion Management, vol. 43, pp. 817-826, 2002.

[121] Y. Lee, W. Wang, and T. Kuo, "Soft computing for battery state-of-charge (BSOC) estimation in battery string systems," IEEE Transaction on Industrial Electronics, vol. 55, pp. 229-39, 2005.

[122] I. Li, W. Wang, S. Su, and Y. Lee, "A merged fuzzy neural network and its applications in battery state-of-charge estimation," IEEE Transaction on Energy Conversion, vol. 22, pp. 697-708, 2007.

[123] B. Chang, Z. Bai, and B. Cao, "State of charge estimation based on evolutionary neural network," Energy Conversion Management, vol. 49, pp. 2788-2794, 2008.

[124] T. Weigert, Q. Tian, and K. Lian, "State of charge prediction of batteries and battery supercapacitor hybrids using artificial neural networks," Journal of Power Sources, vol. 196, pp. 4061-4064, 2011.

[125] A. Eddaheck, O. Briat, N. Bertrand, J. Deletage, and J. Vinassa, "Behavior and state of health monitoring of Li-ion batteries using impedance spectroscopy and recurring neural networks," International Journal of Electrical Power Energy Systems, vol. 42, pp. 487-494, 2012.

[126] J. Kim, S. Lee, and B. Cho, "Complementary cooperation algorithm based on DFKF combined with pattern cognition for SOC/capacity estimation and SOH prediction," IEEE Transaction on Power Electronics, vol. 27, pp. 436-451, 2012.

[127] Y. Sun, H. Jou, and J. Wu, "Auxiliary diagnosis method for lead-acid battery health based on sample entropy," Energy Conversion and Management, vol. 50, pp. 2250-2256, 2009.

[128] A. Widodo, M. Shim, W. Caesarendra, and B.Yang, "Intelligent prognostics for battery health monitoring based on sample entropy," Expert Systems with Applications, vol. 38, pp. 11763-11769, 2011.

[129] G. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part3. State and parameter estimation," Journal of Power Sources, vol. 134, pp. 277-292, 2004.

[130] G. Plett, "Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. simultaneous state and parameter estimation," Journal of Power Sources, vol. 161, pp. 1369-1384, 2006.

[131] J. Lee, O. Nam, and B. Cho, "Li-ion battery SOC estimation method based on the reduced order extended Kalman filtering," Journal of Power Sources, vol. 174, pp. 9-15, 2007.

[132] S. Lee, J. Kim, J. Lee, and B. Cho, "State-of-charge and capacity estimation of lithium ion battery using a new open-circuit voltage versus state-of-charge," Journal of Power Sources, vol. 185, pp. 1367-1373, 2008.

[133] B. Saha and K. Goebel, "Modeling Li-ion battery capacity depletion in a particle filtering framework," Proceedings of annual conference of the PHM Society, San Diego, CA, September 27 - October 1, 2009.

[134] S. Santhanagophlan and R. White, "State of charge estimation using an unscented filter for high power lithium ion cells," International Journal of Energy Research, vol. 34, pp. 152-163, 2010.

[135] F. Sun and X. Hu, Y. Zou, and S. Li, "Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles," Energy, vol. 36, pp. 3531-3540, 2011.

[136] C. Hu, B.Youn, and J. Chung, "A multiscale frame work with extended Kalman filter for lithium-ion battery SOC and capacity estimation," Applied Energy, vol. 92, pp. 694-704, 2012.

[137] Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," Microelectron Reliability, vol. 53, pp. 805-810, 2013.

[138] W. He, N. Williad, C. Chao, and M. Pecht, "State of charge estimation for electric vehicle batteries using unscented kalman filtering," Microeletronics Reliability, vol. 53, pp. 840-847, 2013.

[139] R. Xiong, F. Sun, Z. Chen, and H. He, "A data-driven multi-scale extended Kalman filtering based parameter and state estimation apparoache of lithium-ion battery in electric vehicles," Applied Energy, vol. 113, pp. 463-476, 2014.

[140] J.Vetter, P.Novák, M.Wagner, C.Veit, K. Möller, J. Besenhard, M. Winter, M.Wohlfahrt-Mehrens, C.Vogler, and A.Hammouche, "Ageing mechanism in lithium-ion batteries," Journal of Power Sources, vol. 147, pp. 269-281, 2005.

[141] J. Yan, B. Xia, Y. Su, X. Zhou, J. Zhang, and X. Zhang, "Phenomenologically modeling the formation and evolution of the solid electrolyte interface on the graphite electrode for lithium-ion batteries," Journal of Power Sources, vol. 53, pp. 7069-7078, 2008.

[142] R. Deshpande, M. Verbrugge, Y. Chen, J. Wang, and P. Liu, "Battery cycle life prediction with coupled chemical degradation and fatigue mechanics," Journal of Electrochemical Society, vol. 159, pp. A1730-A1738, 2012.

[143] B. Xu, "Degradation-limiting optimization of battery energy storage systems operation," Master's thesis, ETH Zurich, Zurich, Switzerland, 2013.

[144] M. Safari, M. Morcrette, A. Teyssot, and C. Delacourt, "Multimodal physics-based aging model for life prediction of Li-Ion batteries," Journal of Electrochemical Society, vol. 156, pp. A145-A153, 2009.

[145] G. Ning, B. Haran, and B. Popov, "Capacity fade study of lithium-ion batteries cycled at high discharge rates," Journal of Power Sources, vol. 117, pp. 160-169, 2003.

[146] G. Jin, D. Matthews, and Z. Zhou, "A Bayesian framework for on-line degradation assessment and residual life prediction of secondary batteries in spacecraft," Reliability Engineering and System Safe, vol. 113, pp. 7-20, 2013.

[147] W. He, W. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," Journal of Power Sources, vol. 196, pp. 10314-10321, 2013.

[148] V. Muenzel, "A multi-factor battery cycle life prediction methodology for optimal battery management," Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems, ACM, 2015.

[149] M. Ecker, J. Gerschler, J. Vogel, S. Kabitz, F. Hust, P. Dechent, and D. Sauer, "Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data," Journal of Power Sources, vol. 215, pp. 248-257, 2012.

[150] I. Bloom, B. Cole, J. Sohn, S. Jones, E. Polzin, V. Battaglia, G. Henriksen, C. Motloch, R. Richardson, T. Unkelhaeuser, D. Ingersoll, and H. Case, "An accelerated calendar and cycle life study of Li-ion cells," Journal of Power Sources, vol. 101, pp. 238-247, 2001.

[151] J. Guo, Z. Li, "Bivariate gamma processes for modeling Lithium ion battery aging mechanism," the 2017 Industrial and Systems Engineering Research Conference, Pittsburgh, PA, USA, 2017.

[152] A. Sklar, "Fonctions de répartition á n dimensions et leurs marges," Publications de L'Institut de Statistique de L'Université de Paris, vol. 8, pp. 229-231, 1959.

[153] J. Pinheiro and D. Bates, "Unconstrained parameterizations for variance-covariance matrices," Statistics and Computing, vol. 6, pp. 289-296, 1996.

[154] M.Pourahmadi, "Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation," Biometrika, vol. 86, pp. 677-690, 1999.

[155] M. Smith and R. Kohn, "Parsimonious covariance matrix estimation for longitudinal data," Journal of the American Statistical Association, vol. 97, pp. 1141-1153, 2002.

[156] Z. Chen and D. Dunson, "Random effects selection in linear mixed models," Biometrics, vol. 59, pp. 762-769, 2003.

[157] W. Zhang, and C. Leng, "A moving average cholesky factor model in covariance modeling for longitudinal data," Biometrika, vol. 99, pp. 141-150, 2012.

[158] P. Jaeckel and R. Rebonato, "The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes," Journal of risk 2.2, pp. 17-28, 1999.

[159] D. Creal, S. Koopman, and A. Lucas, "A dynamic multivariate heavy tailed model for time-varying volatilities and correlations," Journal of Business and Economic Statistics, vol. 29, pp. 552-563, 2011.

[160] W. Zhang, C. Leng, and C. Tang, "A joint modeling approach for longitudinal studies," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 77, pp. 219-238, 2015.

[161] S. Morita, H. Shinzawa, R. Tsenkova, I. Noda, and Y. Ozaki, "Computational simulations and a practical application of moving-window two-dimensional correlation spectroscopy," Journal of Molecular Structure, vol. 799, pp. 111-120, 2006.

[162] C. Cortes and V. Vapnik, "Support-vector network," Machine Learning, vol. 20, pp. 1-25, 1995.

[163] A. Smola and B. Scholkopf, "A tutorial support vector regression," Technical Report, 1998.

[164] A. Karatzoglou, D. Meyer, and K. Hornik, "Support vector machines in R," Journal of Statistical Software, vol. 15, pp. 1-28, 2006.

[165] C. Chang and C. Lin, LIBSVM: A library for support vector machines.

[166] E. Dimitriadou, K. Hornik,F. Leisch, D. Meyer, and A. Weingessel, "SVM, e1071 Package," The R project for statistical computing, 2004.

[167] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," Journal of Machine Learning Research, vol. 13, pp. 281-305, 2012.

[168] V. LaRiccia and P. Eggermont, "Maximum Penalized Likelihood Estimation: Volume I: Regression," Springer, 2009.

[169] V. LaRiccia and P. Eggermont, "Maximum Penalized Likelihood Estimation: Volume II: Regression," Springer, 2009.

[170] F. Brissette, M. Khalili, and R. Leconte, "Efficient stochastic generation of multi-site synthetic precipitation data," Journal of Hydrology, vol. 345, pp. 121-133, 2007.

[171] S. Suresh, "Fatigue of materials," Cambridge university press, 1998.

[172] J. Ramsay, "When the data are functions," Psychometrika, vol. 47, pp. 379-396, 1982.

[173] J. Ramsay and B. Silverman, "Applied functional data analysis: methods and case studies," Springer, New York, NY, 2002.

[174] J. Ramsay and B. Silverman, "Functional data analysis," Springer, New York, NY, 2005.

[175] N. Tran, "An introduction to theoretical properties of functional principal component analysis," Thesis, University of Melbourne, 2008.

[176] H. Hotelling, "Analysis of a complex of statistical variable into principal components," Journal of Educational Psychology, vol. 24, pp. 417-441, 1933.

[177] J. Dauxois, A. Pousse, and Y. Romain, "Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference," Journal of multivariate analysis, vol. 12, pp. 136-154, 1982.

[178] F. Yao, H. Muller, and J. Wang, "Functional linear regression analysis for longitudinal data," The Annals of Statistics, vol. 33, pp. 2873-2903, 2005.

[179] F. Yao, H. Muller, and J. Wang, "Functional data analysis for sparse longitudinal data," Journal of the American Statistical Association, vol. 100, pp. 577-590, 2005.

[180] J. Wang, J. Chiou, and H. Muller, "Review of functional data analysis," Annual Review of Statistics and Its Application, Vol. 3, pp. 257-295, 2016.

[181] F. Ferraty and P. Vieu, "Nonparametric functional data analysis: theory and practice," Springer, New York, 2006.

[182] F. Yao and H. Muller, "Functional quadratic regression," Biometrika, vol. 97, pp. 49-64, 2010.

[183] H. Zhu, F. Yao, and H. Zhang, "Structured functional additive regression in reproducing kernel Hilbert spaces," Journal of the Royal Statistical Society Series B, Statistical methodology, vol. 76, pp. 581-603, 2014.

[184] C. Chui, "Multivariate Splines," SIAM, 1987.

[185] J. Peng and P. Debashis, Package "fpca".

[186] J. Peng and P. Debashis, "A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data," Journal of Computational and Graphical Statistics, vol. 18, pp. 995-1015, 2009.

[187] P. Diggle, "Statistical analysis of spatial and spatio-temporal point patterns," CRC Press, Boca Raton, FL, USA, 2014.

[188] C. Wikle, "Modern perspectives on statistics for spatio-temporal data," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 7, pp. 86-98, 2015.

[189] P. Brown, K. Karesen, G. Roberts, and S. Tonellato, "Blur-generated non-separable space-time models," Journal of the Royal Statistical Society, Series B (Statistical Methodology), vol. 62, pp. 847-860, 2000.

[190] X. Liu, K. Yeo, and J. Kalagnanam, "A statistical modeling approach for spatio-temporal degradation data," arXiv: 1609.07217v2.

[191] F. Sigrist, H. Kunsch, and W. Stahel, "A dynamic nonstationary spatio-temporal model for short term prediction of precipitation," arXiv:1102.4210.

[192] F. Sigrist, H. Kunsch, and W. Stahel, "An autoregressive spatio-temporal precipitation model," Procedia Enironmental Sciences, vol. 3, pp. 2-7, 2011.

[193] T.Hefley, M. Hooten, E. Hanks, R. Russell, and D. Walsh, "Dynamic spatio-temporal models for spatial data," Spatial Statistics, vol. 20, pp. 206-220, 2017.

[194] N. Cressie and H. Huang, "Classes of nonseparable, spatio-temporal stationary covariance functions," Journal of the American Statistical Association, vol. 94, pp. 1330-1340, 1999.

[195] C. Ma, "Families of spatio-temporal stationary covariance models," Journal of Statistical Planning and Inference, vol. 116, pp. 489-501, 2003.

[196] G. Matheron, "Principles of Geostatistics," Economic Geology, vol. 58, pp. 1246-1266, 1963.

[197] R. Lark, Bellamy, and R. Rawlins, "Sptio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring," Geoderma, vol. 143, pp. 363-379, 2006.

[198] N. Cressie, "Fitting variogram models by weighted least squares," Journal of the International Association for Mathematical Geology, vol. 17, pp. 563-586, 1985.

[199] N. Cressie, "Statistics for spatial data," John Wiley & Sons, Inc, New York, NY.

[200] R. Carrol, R. Chen, E. George, T. Li, H. Newton, H. Schmiediche, and N. Wang, "Ozone exposure and population density in Harris County, Texas," Journal of the American Statistical Associate, vol. 92, pp. 392-404, 1997.

[201] X. Hu, S. Li, Z. Jia, and B. Egardt, "Enhanced sample entropy-based health management of Li-ion battery, for electrified vehicles," Energy, vol. 64, pp. 953-960, 2014.

[202] J. Guo, J. Li, and W. Si, "A Sample entropy-based method for estimating Lithium Ion battery capacity under non-standard operation condition," ISERC 2017, Pittsburgh, PA, 2017.

[203] N. Cressie and D. Hawkins, "Robust estimation of the variogram," Journal of the International Association of Mathematical Geology, vol. 12, pp. 115-125, 1980.

[204] B. Silverman, "A fast and efficient cross-validation method for smoothing parameter choice in spline regression," Journal of the American Statistical Association, vol. 79, pp. 584-589, 1984.