1
Introduction

1.1 Background and Motivation

Today, we are facing a data rich world that is changing faster than ever before. The ubiquitous availability of data provides great opportunities for industrial enterprises to improve their process quality and productivity. Indeed, the fast development of sensing, communication, and information technology has turned modern industrial systems into a data rich environment. For example, in a modern manufacturing process, it is now common to conduct a 100% inspection of product quality through automatic inspection stations. In addition, many modern manufacturing machines are numerically-controlled and equipped with many sensors and can provide various sensing data of the working conditions to the outside world.

One particularly important enabling technology in this trend is the Internet of Things (IoT) technology. IoT represents a network of physical devices, which enables ubiquitous data collection, communication, and sharing. One typical application of the IoT technology is the remote condition monitoring, diagnosis, and failure prognosis system for after-sales services. Such a system typically consists of three major components as shown in Figure 1.1: (i) the in-field units (e.g., cars on the road), (ii) the communication network, and (iii) the back-office/cloud data processing center. The sensors embedded in the in-field unit continuously generate data, which are transmitted through the communication network to the back office. The aggregated data are then processed and analyzed at the back-office to assess system status and produce prognosis. The analytics results and the service alerts are passed individually to the in-field unit. Such a remote monitoring system can effectively improve the user

Industrial Data Analytics for Diagnosis and Prognosis: A Random Effects Modelling Approach, First Edition. Shiyu Zhou and Yong Chen.

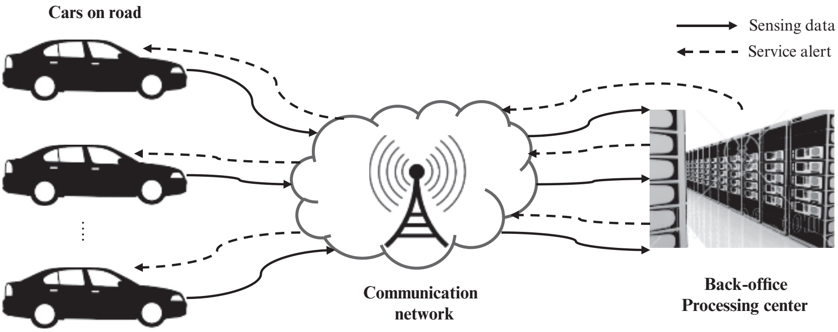

Figure 1.1 A diagram of an IoT enabled remote condition monitoring system.

experience, enhance the product safety, lower the ownership cost, and eventually gain competitive advantage for the manufacturer. Driven by the rapid development of information technology and the critical needs of providing fast and effective after-sales services to the products in a globalized market, the remote monitoring systems are becoming increasingly available.

The unprecedented data availability provides great opportunities for more precise and contextualized system condition monitoring, diagnosis, and prognosis, which are very challenging to achieve if only scarce data are available. Industrial data analytics is the process of collecting, exploring, and analyzing data generated from industrial operations throughout the product life cycle in order to gain insights and improve decision-making. Industrial data analytics encompasses a vast set of applied statistics and machine learning tools and techniques, including data visualization, data-driven process modeling, statistical process monitoring, root cause identification and diagnosis, predictive analytics, system reliability and robustness, and design of experiments, to name just a few. The focus of this book is industrial data analytics approaches that can take advantage of the unprecedented data availability. Particularly, we focus on the concept of random effects and its applications in system diagnosis and prognosis.

The terms *diagnosis* and *prognosis* were originally used in the medical field. Diagnosis is the identification of the disease that is responsible to the symptoms of the patient's illness, and prognosis is a forecast of the likely course of the disease. In the field of engineering, these terms have similar meanings: for an industrial system, diagnosis is the identification of the root cause of a system failure or abnormal working condition; and prognosis is the prediction of the system degradation status and the future failure or break down. Obviously, diagnosis and prognosis play a critical role in assuring smooth, efficient, and safe system operations. Indeed, diagnosis and prognosis have attracted ever-growing interest in recent years. This trend has been driven by the fact that capital goods manufacturers have been coming under increasing pressure to open up new sources of revenue and profit in recent years. Maintenance service costs constitute around 60–90% of the life-cycle costs of industrial machinery and installations. Systematic extension of the after-sales service business will be an increasingly important driver of profitable growth.

Due to the importance of diagnosis and prognosis in industrial system operations, a relatively large number of books/research monographs exist on this topic [Lewis et al., 2011, Niu, 2017, Wu et al., 2006, Talebi et al., 2009, Gertler, 1998, Chen and Patton, 2012, Witczak, 2007, Isermann, 2011, Ding, 2008, Si et al., 2017]. As implied by their titles, many of these books focus on model-based diagnosis and prognosis problems in dynamic systems. A model-based approach adopts a dynamic model, often in the form of a state space model, as the basis for diagnosis and prognosis. Then the difference between the observations and the model predictions, called residuals, are examined to achieve fault identification and diagnosis. For the prognosis, data-driven dynamic forecasting methods, such as time series modeling methods, are used to predict the future values of the interested system signals. The modeling and analysis of the system dynamics are the focus of the existing literature.

Different from the existing literature, this book focuses on *the concept of random effects and its applications in system diagnosis and prognosis*. Random effects, as the name implies, refer to the underlying random factors in an industrial process that impact on the outcome of the process. In diagnosis and prognosis applications, random effects can be used to model the sources of variation in a process and the variation among individual characteristics of multiple heterogeneous units. The following two examples illustrate the random effects in industrial processes.

Example 1.1 Random effects in automotive body sheet metal assembly processes

The concept of variation source is illustrated for an assembly operation in which two parts are welded together. In an automotive sheet metal assembly process, the sheet metals will be positioned and clamped on the fixture system through the matching of the locators (also called pins) on the fixture system and the holes on the sheet metals. Then the sheet metals will be welded together. Clearly, the accuracy of the positions of the locating pins and the tightness of the matching between the pins and the holes significantly influence the dimensional accuracy of the final assembly. Figure 1.2(a) shows the final product as designed. The assembly process is as follows: Part 1 is first located on the fixture and constrained by 4-way Pin L_1 and 2-way Pin L_2. A 4-way pin constrains the movement in two directions, while a 2-way pin only constrains the movement in one direction. Then, Part 2 is located by 4-way Pin L_3 and 2-way Pin L_4. The two parts are then welded together in a joining operation and released from the fixture.

If the position or diameter of Pin L_1 deviates from design nominal, then Part 1 will consequently not be in its designed nominal position, as shown in Figure 1.2(b). After joining Part 1 and Part 2, the dimensions of the final parts will deviate from the designed nominal values. One critical point that needs to be emphasized is that Figure 1.2(b) only shows one possible realization of produced assemblies. If we produce another assembly, the deviation of the position of Part 1 could be different. For instance, if the diameter of a pin is reduced due to pin wear, then the matching between the pin and the corresponding hole will be loose, which will lead to random wobble of the final position of part. This will in turn cause increased variation in the dimension of the produced final assemblies. As a result, mislocations of the pin can be manifested by either mean shift or variance change in the dimensional quality measurement such as M_1 and M_2 in the figure. In the case of mean shift error (for example due to a fixed position shift of the pin), the error can be compensated by process adjustment such as realignment of the locators. The variance change errors (for example due to a worn-out pin or the excessive looseness of a pin) cannot be easily compensated for in most cases. Also, note that each locator in the process is a potential source of the variance change errors, which is referred to as a *variation source*. The variation sources are random effects in the process that will impact on the final assembly quality. In most assembly processes, the pin wear is difficult to measure so the random effects are not directly observed. In a modern automotive body assembly process, hundreds of locators are used

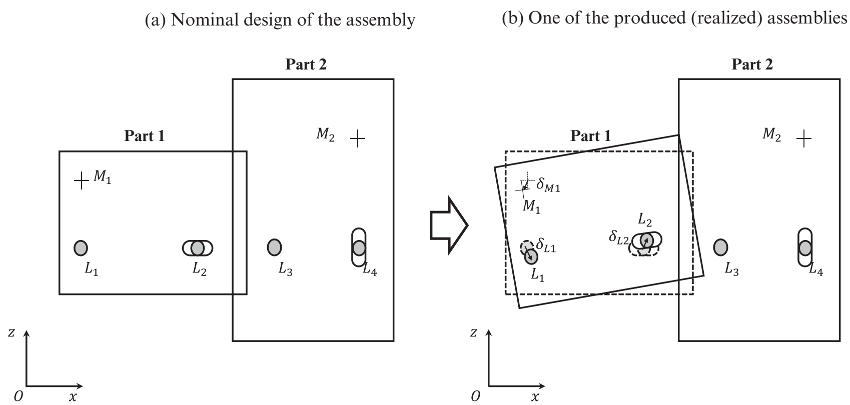

Figure 1.2 Random effects in an assembly operation.

to position a large number of parts and sub-assemblies. An important and challenging diagnosis problem is to estimate and identify the variation sources in the process based on the observed quality measurements.

Example 1.2 Random effects in battery degradation processes

In industrial applications, the reliability of a critical unit is crucial to guarantee the overall functional capabilities of the entire system. Failure of such a unit can be catastrophic. Turbine engines of airplanes, power supplies of computers, and batteries of automobiles are typical examples where failure of the unit would lead to breakdown of the entire system. For these reasons, the working condition of such critical units must be monitored and the remaining useful life (RUL) of such units should be predicted so that we can take preventive actions before catastrophic failure occurs. Many system failure mechanisms can be traced back to some underlying degradation processes. An important prognosis problem is to predict RUL based on the degradation signals collected, which are often strongly associated with the failure of the unit. For example, Figure 1.3 shows the evolution of the internal resistance signals of multiple automotive lead-acid batteries. The internal resistance measurement is known to be one of the best condition monitoring signals for the battery life prognosis [Eddahech et al., 2012]. As we can see from Figure 1.3, the internal resistance measurement generally increases with the service time of the battery, which indicates that the health status of the battery is deteriorating.

We can clearly see from Figure 1.3 that although similar, the progression paths of the internal resistance over time of different batteries are not identical. The difference is certainly expected due to many random factors in the material, manufacturing processes, and the working environment that vary

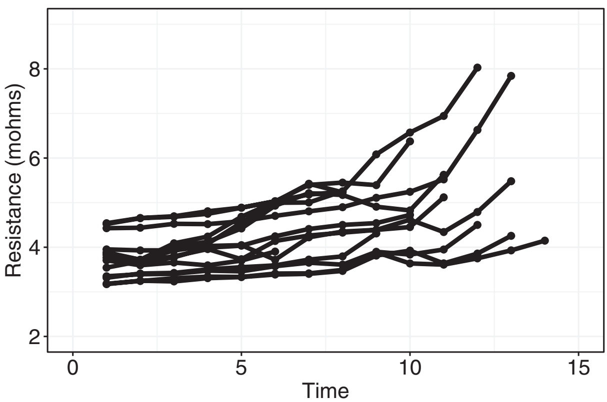

Figure 1.3 Internal resistance measures from multiple batteries over time.

from unit-to-unit. The random characteristics of degradation paths are random effects, which impact the observed degradation signals of multiple batteries.

The available data from multiple similar units/machines poses interesting intellectual opportunities and challenges for prognosis. As for opportunities, since we have observations from potentially a very large number of similar units, we can compare their operations/conditions, share the information, and extract common knowledge to enable accurate prediction and control at the individual level. As for challenges, because the data are collected in the field and not in a controlled environment, the data contain significant variation and heterogeneity due to the large variations in working conditions for different units. The data analytics approaches should not only be general (so that the common information can be learned and shared), but also flexible (so that the behavior of an individual subject can be captured and controlled).

Random effects always exist in industrial processes. The process variation caused by random effects is detrimental and thus random effects should be modeled, analyzed, and controlled, particularly in system diagnosis and prognosis. However, due to the limitation in the data availability, the data analytics approaches considering random effects have not been widely adopted in industrial practices. Indeed, before the significant advancement in communication and information technology, data collection in industries often occurs *locally* in very similar environments. With such limited data, the impact of random effects cannot be exposed and modeled easily. This situation has changed significantly in recent years due to the digital revolution as mentioned at the beginning of the section.

The statistical methods for random effects provide a powerful set of tools for us to model and analyze the random variation in an industrial process. The goal of this book is to provide a textbook for engineering students and a reference book for researchers and industrial practitioners to adapt and bring the theory and techniques of random effects to the application area of industrial system diagnosis and prognosis. The detailed scope of the book is summarized in the next section.

1.2 Scope and Organization of the Book

This book focuses on industrial data analytics methods for system diagnosis and prognosis with an emphasis on random effects in the system. Diagnosis concerns identification of the root cause of a failure or an abnormal working condition. In the context of random effects, the goal of diagnosis is to identify the variation sources in the system. Prognosis concerns using data to predict what will happen in the future. Regarding random effects, prognosis focuses on addressing unit-to-unit variation and making degradation/failure predictions for each individual unit considering the unique characteristic of the unit.

The book contains two main parts:

1. **Statistical Methods and Foundation for Industrial Data Analytics**
   This part covers general statistical concepts, methods, and theory useful for describing and modelling the variation, the fixed effects, and the random effects for both univariate and multivariate data. This part provides necessary background for later chapters in part II. In part I, Chapter 2 introduces the basic statistical methods for visualizing and describing data variation. Chapter 3 introduces the concept of random vectors and multivariate normal distribution. Basic concepts in statistical modeling and inference will also be introduced. Chapter 4 focuses on the principal component analysis (PCA) method. PCA is a powerful method to expose and describe the variations in multivariate data. PCA has broad applications in variation source identification. Chapter 5 focuses on linear regression models, which are useful in modeling the fixed effects in a dataset. Statistical inference in linear regression including parameter estimation and hypothesis testing approaches will be discussed. Chapter 6 focuses on the basic theory of the linear mixed effects model, which captures both the fixed effects and the random effects in the data.

2. **Random Effects Approaches for Diagnosis and Prognosis**
   This part covers the applications of the random effects modeling approach to diagnosis of variation sources and to failure prognosis in industrial processes/systems. Through industrial application examples, we will present variation pattern based variation source identification in Chapter 7. Variation source estimation methods based on the linear mixed effects model will be introduced in Chapter 8. A detailed performance comparison of different methods for practical applications is presented as well. In Chapter 9, the diagnosability issue for the variation source diagnosis problem will be studied. Chapter 10 introduces the mixed effects longitudinal modeling approach for forecasting system degradation and predicting remaining useful life based on the first time hitting probability. Some variations of the basic method such as the method considering mixture prior for unbalanced data in remaining useful life prediction are also presented. Chapter 11 introduces the concept of Gaussian processes as a nonparametric way for the modeling and analysis of multiple longitudinal signals. The application of the multi-output Gaussian process for failure prognosis will be presented as well. Chapter 12 introduces the method for failure prognosis combining the degradation signals and time-to-event data. The advanced joint prognosis model which integrates the survival regression model and the mixed effects regression model is presented.

1.3 How to Use This Book

This book is intended for students, engineers, and researchers who are interested in using modern statistical methods for variation modeling, diagnosis, and prediction in industrial systems.

This book can be used as a textbook for a graduate level or advanced undergraduate level courses on industrial data analytics. The book is fairly self-contained, although background in basic probability and statistics such as the concept of random variable, probability distribution, moments, and basic knowledge in linear algebra such as matrix operations and matrix decomposition would be useful. The appendix at the end of the book provides a summary of the necessary concepts and results in linear space and matrix theory. The materials in Part II of the book are relatively independent. So the instructor could combine selected chapters in Part II with Part I as the basic materials for different courses. For example, topics in Part I can be used for an advanced undergraduate level course on introduction to industrial data analytics. The materials in Part I and some selected chapters in Part II (e.g., Chapters 7, 8, and 9) can be used in a master's level statistical quality control course. Similarly, materials in Part I and selected later chapters in Part II (e.g., Chapters 10, 11, 12) can be used in a master's level course with emphasis on prognosis and reliability applications. Finally, Part II alone can be used as the textbook for an advanced graduate level course on diagnosis and prognosis.

One important feature of this book is that we provide detailed descriptions of software implementation for most of the methods and algorithms. We adopt the statistical programming language `R` in this book. `R` language is versatile and has a very large number of up-to-date packages implementing various statistical methods [R Core Team, 2020]. This feature makes this book fit well with the needs of practitioners in engineering fields to self study and implement the statistical modeling and analysis methods. All the `R` codes and data sets used in this book can be found at the book companion website.

Bibliographic Notes

Some examples of good books on system diagnosis and prognosis in engineering area are Lewis et al. [2011], Niu [2017], Wu et al. [2006], Talebi et al. [2009], Gertler [1998], Chen and Patton [2012], Witczak [2007], Isermann [2011], Ding [2008], Si et al. [2017]. Many good textbooks are available on industrial statistics. For example, Montgomery [2009], DeVor et al. [2007], Colosimo and Del Castillo [2006], Wu and Hamada [2011] are on statistical monitoring and design. On the failure event analysis and prognosis, Meeker and Escobar [2014], Rausand et al. [2004], Elsayed [2012] are commonly cited references.