Integrated Prognostics for Component Health Management

by

Fuqiong Zhao

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Engineering Management

Department of Mechanical Engineering
University of Alberta

Abstract

Prognostics in engineering field is dedicated to predicting how long further a component or a system will perform their intended functions before failure. Prognostics is an essential building block in condition based maintenance. Accurate prediction of the component remaining useful life provides valuable information to decision making on maintenance planning, mission planning and logistics. Preventive actions based on remaining useful life prediction can dramatically avoid unscheduled downtime, reduce operational risk and cost, and improve the safety of the working environment.

This thesis is devoted to developing integrated prognostics methods for the remaining useful life prediction of a specific component by integrating physics of failure and condition monitoring data. The first contribution of the thesis is that by combining physics and data effectively, the proposed method overcomes the limitations of existing prognostics approaches, which are mainly either physics based or data-driven. To account for the uncertainty in failure times of units in population, parameters are treated as random variables in physical degradation models. By noticing the uniqueness of the failure time for a specific unit, this study utilizes Bayesian inference to reduce the uncertainty in model parameters, which leads to a more accurate prediction on remaining useful life of the specific unit.

This thesis also proposes an integrated prognostics method for the component operating under time-varying operating conditions. The capability to directly relate the load to the degradation rate is a key advantage of the proposed method over the existing

data-driven methods when dealing with time-varying operating conditions. This is the second main contribution of this thesis.

To cater to real-time applications of condition based maintenance, an efficient spectral method named polynomial chaos expansion is investigated for uncertainty quantification in prognostics. The proposed method is able to accelerate the uncertainty quantification in the integrated prognostics method and the computational efficiency is significantly improved, which is the third main contribution of this thesis.

In addition, this thesis accounts for two important factors when developing integrated prognostics method: uncertainty in damage initiation time and shock in the degradation. These two factors have not been explicitly considered for prognostics purpose in the existing research. By simultaneously adjusting both the damage initiation time and the model parameters, the prediction accuracy is improved. The failure time reduction caused by the shock is accommodated by identifying a virtual damage initiation time. This work consists of the fourth main contribution.

The integrated prognostics methods developed in this thesis are applied to spur gears. Two types of failure modes are considered. One is the tooth fracture due to bending stress and the other one is the surface wear due to sliding contact. Validation is conducted using a run-to-failure experiment on a planetary gearbox.

Acknowledgments

First and foremost I would like to sincerely thank my supervisor, Dr. Zhigang Tian, who supports my PhD study and research with his patience, enthusiasm, and immense knowledge. His guidance and encouragement made the experience of working with him so enjoyable and rewarding. He offers a lot of help in the course of my study, research and this thesis writing. Without him this thesis would not have been completed. Meanwhile, I also offer the sincere gratitude to my co-supervisor, Dr. Yong Zeng. He always encourages me to follow the heart and to bravely pursue the dreams. I admire his persistence, wisdom and effort and I thank his help, guidance and suggestion. I feel so lucky to have these two supervisors. I simply could not wish for better ones.

Besides my supervisors, I would like to thank the rest of my thesis committee: Dr. Qingyu Yang, Dr. Ming J. Zuo, Dr. Zengtao Chen and Dr. John Doucette, for their time to review this thesis, insightful comments and hard questions to improve the quality of this thesis.

I would like to thank Dr. Ming J. Zuo for his help in my career and for his kindness to provide the experimental data from his research group for validation of my research work.

In addition, I would like to express my gratitude to Dr. Eric Bechhoefer. He is a so knowledgeable engineer and a so nice friend. Every discussion with him inspires me and I appreciate a lot for him to kindly share with me his valuable field experience in equipment diagnostics and prognostics.

Last but not least, words cannot express my love and gratitude to my family. I thank my parents, Hanyun Zhao and Aiyun Li, for bringing me to this wonderful world, and for teaching me how to become a happy, helpful and nice person; I thank my husband, Qinchun Qian, for his constant support, encouragement and endless love; and I thank my lovely son, Benjamin, for making my life even more bright and delightful.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CBM	condition based maintenance
PHM	proportional hazard model
FE	finite element
RUL	remaining useful life
SIF	stress intensity factor
PCE	polynomial chaos expansion
MCMC	Markov chain Monte Carlo
DOF	degrees of freedom
CIT	crack initiation time
PDF	probability density function
ANN	artificial neural network
IID	independent and identically distributed

Chapter 1.   Introduction

1.1   Overview, motivations and objectives

Condition based maintenance (CBM) which is an innovative technology in reliability and maintainability engineering has emerged over recent years. CBM changes the way complex engineering systems are designed, monitored and maintained to ensure their safety, reliability and availability. The core idea of CBM is to put the engineering system under a continuous surveillance and to give maintenance suggestions according to the health condition of the system. CBM allows for preventive actions before failure occurs, so that the unscheduled downtime can be avoided and the operational risk and cost can be reduced. Due to the great potential of CBM, intensive research efforts have been invested from both industrial and academic sides to enhance its capability in various application areas, such as aerospace, power & energy, ground vehicle, manufacturing industry, as well as electronics industry.

CBM architecture consists of three parts. 1) *Diagnostics*: to detect, identify and assess the faults. 2) *Prognostics and Uncertainty Management*: to predict remaining useful life (RUL) and its confidence level. 3) *Maintenance Optimization*: to schedule maintenance based on the failure probability predicted by prognostics and other information available on logistics. The research work in this thesis focuses on prognostics and uncertainty management.

Prognostics is dedicated to predicting how long further a component or a system will perform their intended function before failure. Prognostics is an essential building block in CBM because accurate prediction of the component RUL provides valuable

information to decision making for corrective actions. Uncertainty inherent in prognostics imposes major challenges for prognostic methods development. Hence, another important task in prognostics is to quantify the uncertainty in RUL prediction.

During operation in the life span, component health condition degrades gradually, resulting in the failure sooner or later. Condition monitoring systems are built to oversee the health condition of the component. Some health condition indices are extracted to represent the degradation process based on sensor measurements, including but not limited to vibration, acoustics and oil debris. There are two basic ways to predict the degradation. One is to analyze the sensor data to establish the relationship between the data and the degradation; the other one is to utilize the physics behind the degradation to build the predictive model. These two ways lead to two types of prognostics approaches: data-driven methods and physics-based methods. The third type of prognostics approaches tries to combine physics of failure and sensor data to benefit from both, which is termed in this paper as integrated prognostics methods.

Gearbox is a critical system for power transmission. It is widely used in various engineering systems, such as ground vehicles, helicopters and wind turbines. Unexpected failure of the gearbox will cause catastrophic failure of the whole system. Hence, RUL prediction for the gear is important. In this thesis, we investigate two main failure modes of the gear. One is the tooth fracture caused by crack propagation due to cyclic bending stress at tooth root. The other one is the surface wear due to sliding contact during the mesh process. The failure mode of surface wear is studied for validation purpose.

Existing prognostics methods for the gear are either data-driven or physics-based. The two types of methods have their respective drawbacks. Data-driven methods purely

depend on data without any consideration of physics, therefore, they sometimes end up with meaningless extrapolation. Furthermore, data-driven methods are inadequate to deal with time-varying operating conditions simply because it is impossible to build a library to cover all the possible operating conditions. While, physics-based methods usually assume model parameters to be deterministic without considering uncertainty. In addition, they require intensive efforts to build physical models with high fidelity.

By noticing these limitations of the existing prognostics methods, the first objective of this thesis is to develop an integrated prognostics method for gear RUL prediction, which is able to combine physics of failure and sensor data to exceed these limitations. Although the gear is used to demonstrate the method, by appropriate model adjustment, the methodology can be extended to other critical component in engineering systems.

The second research work is motivated by the inefficiency of the current uncertainty quantification methods in prognostics. The slow convergence rate prohibits the full exploration of the effects of uncertainty on RUL prediction. Hence, the second objective is to propose an efficient uncertainty quantification method in the integrated prognostics.

The third motivation comes from the difficulties of data-driven prognostics methods in dealing with time-varying operating conditions. The third objective is to develop an integrated prognostics method that is able to account for changes in operating conditions.

The fourth work is motived by the lack of research to consider uncertainty in damage initiation time and shock in the degradation in the integrated prognostics. The damage initiation time determines the starting point of prediction, whose uncertainty has not yet been explicitly considered in the existing prognostics methods. The shock causes a sudden amount of damage accumulation. Because of the shock, the degradation path will

deviate from the original one, resulting in a shortened lifetime of the component. Both factors have influences on the failure time. Hence the fourth objective is to accommodate their effects in the integrated prognostics.

1.2 Approaches and validation

There is a large variation in failure times of component population. Thus failure time is treated as random variable. Failure is often defined by the exceedance to a predefined threshold of an indicator. This indicator is usually the output of the prognostic model. Hence, when the prognostic model is used to predict the degradation, the uncertainty is usually assigned to model parameters. However, for an individual component that is currently under monitoring, its failure time is unique. The model parameters for this individual component should be deterministic or at least have very small uncertainty. In the proposed integrated prognostics methods, we can reduce the uncertainty in model parameters by utilizing the sensor data. After adjustments, the model parameters can better characterize the degradation path for this specific component. The predicted failure time will be more accurate. Bayesian inference is naturally amenable to uncertainty quantification. It is used to reduce the uncertainty in the model parameters.

The proposed integrated method is able to deal with time-varying operating conditions in that the damage (crack) propagation model naturally relates the load to the degradation rate. The stress intensity factor (SIF) is calculated as a function of both crack size and load in the theory of linear elastic fracture mechanics. The time-varying loading condition considered in this thesis has a profile of piece-wise constant.

Uncertainty propagation is needed in Bayesian inference and failure time distribution computation. Monte Carlo simulation is often used to realize uncertainty propagation. However, its slow convergence rate makes it computationally infeasible to deal with high dimensional uncertainty. This thesis investigates a spectral method named polynomial chaos expansion (PCE) to improve the computational efficiency for uncertainty quantification in integrated prognostics.

The prognostic model in the proposed integrated method is based on physics of failure. Hence, physical models are needed to describe the engineering system and the damage progression. Usually, the available physical models are used for system simulation under the healthy condition. However, since prognosis is to predict the damage progression, the physical models with faults are in need. Moreover, it is desirable to characterize the interactions of components in the system to achieve high fidelity. In the proposed integrated prognostics method, physical models include damage progression model, finite element model and gearbox dynamics model. These physical models can give good simulations for both the faulty gearbox dynamics and the fault progression process.

Prognosis starts when a fault is detected and assessed by diagnosis. However, the inaccuracy of diagnosis leads to the inaccuracy of start time of prognosis. Early start time will underestimate the failure time, while late start time will overestimate the failure time. Hence, the damage initiation time should be considered as a source of uncertainty. By considering this type of uncertainty, we are able to deal with the lifetime reduction due to shock occurrence. The idea is to identify a different damage initiation time from the

original one. By adjusting both the damage initiation time and the model parameters simultaneously, the prediction accuracy is improved and the shock is accommodated.

Validation of the proposed method is conducted by utilizing a set of experimental data from a run-to-failure test on a planetary gearbox. Mass loss of the sun gear is recoreded. It is assumed that the mass loss is caused by both rolling and sliding contact when gears are in mesh. Wear model is established for describing the wear progression. The recorded data on the mass loss are used to update the wear coefficient, which is the model parameter in the wear model. The proposed integrated method is able to give mass loss predictions that well agree with the actual mass loss.

1.3   Main contributions

- Propose an integrated prognostics method for gear RUL prediction with failure mode of tooth fracture. The proposed method is better than physics-based method in that the physical model parameters can be adjusted for a specific component; and is better than data-driven method in that massive data trending is not necessary.

- Propose an efficient uncertainty quantification method in the integrated prognostics method based on PCE. The proposed method is able to accelerate the update process of model parameters as well as the uncertainty propagation in RUL distribution computation.

- Propose an integrated prognostics method under time-varying operating conditions. The capability to directly relate the load to the degradation rate is a key advantage of the proposed method over existing data-driven methods.

- Propose an integrated prognostics method considering uncertainty in damage initiation time and shock in the degradation. By simultaneously adjusting both the damage initiation time and the model parameters, the prediction accuracy is improved and the shock effect is accommodated.

- Propose an integrated prognostics method for the gear with failure mode of tooth surface wear, which is caused by sliding contact during the mesh process. The proposed method is validated using experimental data from a run-to-failure test.

1.4   Organization and outline

The thesis is organized as follows.

Chapter 2 is the literature review on CBM, current prognostics approaches, physical models and uncertainty quantification.

Chapter 3 is devoted to presenting the gearbox dynamics model that is used in this thesis. This dynamics model outputs the dynamic load that the gear tooth (healthy and faulty) experiences during the mesh process. Potential energy method is used to calculate the stiffness of the gear tooth that bears a curved crack initiated near the tooth root.

Chapter 4 presents an integrated prognostic method for RUL prediction of the gear with the failure mode of tooth fracture. There are three basic parts in the proposed method: physical models, data and Bayesian inference. The logic is that, we will use data to update physical model through Bayesian inference. The physical models include a damage progression model, a finite element model and a gearbox dynamics model. The damage progression model adopted is the well-known Paris' law, which describes the crack evolution with loading cycles. The parameters in the damage progression model are

7

to be updated in Bayesian inference. Finite element (FE) model is used to calculate SIF, which is an important quantity in Paris' law. The load applied on the FE gear model is calculated by the gearbox dynamics model discussed in Chapter 3. The data is of two types: degradation histories and sensor data. The degradation histories are used to determine an initial prior distribution for the model parameters; the sensor data are collected from condition monitoring and are used to update the model parameters.

Chapter 5 presents an efficient uncertainty quantification method based on PCE in the integrated prognostics method. The sources of uncertainty considered are classified into three categories: model input uncertainty, measurement uncertainty and model uncertainty, among which, the model input uncertainty is further divided into updating-uncertainty and non-updating-uncertainty. A form of likelihood in Bayesian inference considering different roles of different types of uncertainty is defined based on this classification. PCE is proposed to accelerate the update process and the uncertainty propagation in RUL prediction.

Chapter 6 presents the integrated prognostics method under time-varying operating conditions. One of the most important operating conditions that the gearbox is subject to is loading. This chapter considers the time-varying loading profile of piece-wise constant. In the theory of linear elastic fracture mechanics, the surface of SIF is obtained as a function of both crack size and load. An efficient method based on PCE is used to accelerate Markov Chain Monte Carlo (MCMC) sampling in Bayesian inference implementation. The model parameters in damage progression model are considered to be correlated. Their joint probability density function (PDF) is updated when a new

observation on crack size is available at inspection time. Uncertainty propagation in RUL prediction is implemented using PCE considering the parameter correlation.

Chapter 7 presents the integrated prognostics method considering uncertainty in damage initiation time and shock in degradation. The damage initiation time determines the start time of prognostics, which is the intercept of degradation path with time axis at the initial damage size. The model parameters determine the slope of the degradation path. Hence, simultaneous updates of damage initiation time and model parameters manifest the updates of both the slope and the intercept of the degradation path. In this way, the degradation path can be better characterized. Shock will cause a sudden damage accumulation, deviate degradation path from the original one, and shorten the gear lifetime. The proposed method accommodates the lifetime reduction by identifying a virtual damage initiation time.

Chapter 8 presents the experimental validation for the integrated prognostics method. The failure mode considered is surface wear due to sliding contact during the mesh process. A run-to-failure test is conducted on a planetary gearbox. During the test, the weight loss of the sun gear is recorded at several inspection times. The surface wear and pitting are assumed to be responsible for the mass loss. A wear model for the sun gear tooth is established to predict the mass loss. In the integrated prognostics method, the wear coefficient in this wear model is updated by utilizing the measurements on mass loss collected during the experiment.

Chapter 9 concludes the thesis and suggests areas for further research.

Chapter 2.   Background and literature review

2.1   CBM

Reliability describes the ability of a product to perform its intended functions. High requirement of reliability is crucial to the engineering systems which are responsible for production, manufacturing, power generation and transportation. Unexpected failures of these systems will cause catastrophic loss of asset, undermine the revenue and the reputation of enterprises, and more dangerously, jeopardize human lives. The primary task in maintenance is to take actions to retain a system in a state in which it can perform the required function. Corrective maintenance is a maintenance strategy that suggests maintenance actions after failure occurs. It is totally reactive which allows for the unexpected failure to happen. Hence, this type of maintenance suffers from poor availability and high outage rate. Instead, preventive maintenance (Jardine and Tsang 2005) suggests a time interval to perform the component replacement. This time interval can be optimized in terms of total cost and/or reliability based on the failure time distribution of the component. Even though the preventive maintenance can prevent unexpected failures to some extent, it is still resource wasteful, because the component may be replaced when it still has good life left.

In the recent decade, CBM has emerged as a maintenance technology (Vachtsevanos et al. 2006). The strategy is to take maintenance actions according to the equipment health condition. CBM can achieve reliable and cost-effective operations of engineering systems.

In CBM, condition monitoring system is used to monitor the equipment health by a variety of sensors. Based on the sensor signals, faults are detected and assessed by diagnostics; future degradation and RUL are predicted by prognostics; and maintenance actions are taken accordingly. Diagnostics, prognostics and maintenance optimization are the three basic components in CBM (Jardine et al. 2006).

2.2 Prognostics methods

Classification of prognostics methods can be considered from different angles. Byington et al. (2002) classified the prognostic approaches into experience-based prognostics, evolutionary or trending models and model-based prognostics. Jaw and Wang (2004) proposed a classification of prognostics systems in terms of off-line and on-line. It depends on whether the system performs tasks by feeding it with real time data while the equipment is in operation or by retrieving the previous stored data to do the analysis for the past operations. Hines and Usynin (2008) categorized the prognostics methods by the type of information they use: failure data-based, stress-based, effects-based and combined type. The combined type was defined as methods which made use of a combination of failure data, stress data and effects data. A review paper on the prognostics methods in the field of rotating machinery (Heng et al. 2009) divided the prognostics models into traditional reliability models, prognostic models (physics-based, data-driven) and integrated approach. While in the paper (An et al. 2011), the authors categorized the methods into data-driven, physical model-based and hybrid approaches. The hybrid approaches specifically include particle filters and Bayesian techniques. Zio (2012) proposed classification among prognostic approaches as first-principle model-based, reliability model-based and process sensor data-driven approaches. The author

also discussed the ways to hybrid information and data from various sources. By noticing these existing classifications, this thesis partitions the prognostics methods in its own way: data-driven methods, physics-based methods and integrated methods. The integrated methods are defined as a type of methods in which the data will be used to affect the physical model. In the following sections, the literature review will be conducted according to this classification.

2.2.1 Data-driven prognostics methods

Generally speaking, data-driven prognostics methods depend on data. They basically have two directions to proceed. One direction is based on the concept of reliability, where the failure time distribution is estimated by parametric or non-parametric models. The other direction is based on diagnostics feature extraction and intelligent training.

Prognostics is closely related to reliability. Component reliability describes the probability of how long the component will fail. Time to failure of a component is a basic quantity of interest in reliability engineering. It is treated as random variable characterized by its probability distribution. The times when the operation is stopped but the component has not yet failed are termed as censor data. The history associated with the censor data is called suspension history. The suspension history provides the reliability information of the component surviving certain time. Both failure history and suspension history are useful in determining the failure time distribution. Prognostics based on failure time distributions, like Weibull analysis (Abernethy 1996), is one of the data-driven methods. When operating environment is considered, the reliability is also a function of the environment. The environment variables include temperature, humidity,

pressure, stress, speed, etc., which influence the failure time of the component. These variables are termed as "covariates". The well-known proportional hazard model (PHM) proposed by Cox (1972) is a data-driven prognostics method considering covariates. In addition to environment variables, covariates in PHM can also represent health indicators. The hazard rate (2.1) in PHM consists of two parts: hazard baseline and covariate effects.

$$h(t \mid \tilde{\beta}) = h_0(t) \exp\left(\sum_{i=1}^{M} \tilde{\beta}_i z_i\right) \tag{2.1}$$

The coefficients $\tilde{\beta}$ could be estimated by *partial likelihood*. Anderson and Gill (1982) further investigated the scenario in which the hazards were non-proportional and the covariates were time-dependent. Markov models were often used to account for the influence of the covariates on component failure behavior (Samanta et al. 1991; Yeh, 1997; Banjevic et al. 2001; Marseguerra et al. 2002). A disadvantage of reliability model is that it usually requires a large number of degradation states and distribution parameters to be fitted by data. As the equipment becomes more and more reliable, failure data will be insufficient for fitting the distributions.

Lu and Meeker (1993) proposed a two-stage method to estimate the failure time distribution using degradation measures. General path model (linear or nonlinear) were developed, which was able to account for both population and individual effects. The model parameters were estimated using historical data.

Different from the scarce of failure data, in CBM rich sensor data are available collected from condition monitoring of the component. Extracted features from the sensor data are expected to represent the component health condition. The features that are suitable for prognostics are expected to be damage sensitive and to have an increasing

trend with time as the damage accumulates. With these features, data-driven methods can be developed to identify the relationship between degradation status and aging of the component. Such approaches are mainly statistical methods, like regression, time series analysis (Hines and Garvey 2007), and artificial intelligence algorithms such as neural networks, fuzzy logic systems and support vector machines (Yan et al. 2004; Wang et al. 2004; Sotiris and Pecht 2007; Peng et al. 2010). Among these artificial intelligence methods, artificial neural network (ANN) is a promising and commonly used data-driven method for prognostics. Gebraeel et al. (2004, 2008) developed a feedforward neural network for ball bearing RUL prediction. The output of ANN model was the vibration magnitude. In the paper (Wu et al. 2007), the ANN output was life percentage. This type of output facilitates the threshold definition. Tian et al. (2010) developed a prognostics method based on ANN utilizing both failure and suspension histories, in which age of the component and condition monitoring data were the inputs, and life percentage was the output. A recurrent wavelet neural network was developed by Wang and Vachtsevanos (2001) to predict the rolling element bearing crack propagation. Other published work on component health prognostics based on ANN techniques are found in (Wang et al. 2004; Shao and Nezu 2000; Huang et al. 2007; Tse and Atherton 1999; Tian 2012). These data-driven methods can directly use the sensor data for equipment failure prognostics. However, because they are purely depend on sensor data without considering any physics, massive data training is needed and sometime the methods end up with meaningless extrapolation.

2.2.2 Physics-based prognostics methods

Physics-based prognostics methods predict the component degradation based on physics of failure. The physical model is typically a mathematical equation that describes the damage evolution, as known from first-principle laws of physics. Because physics-based methods are based on physics of failure, they have high prediction accuracy when applicable. In this type of methods, faulty system modeling is often required to calculate the physical quantities that determine the damage evolution.

The faulty system modeling is a challenging issue in physics-based prognostics. Since analytical approaches are impossible or computational prohibited for many real applications, numerical methods become the indispensable tools to solve problems. FE method is a powerful numerical method to solve problems in structural mechanics, fluid mechanics, heat transfer as well as electromagnetics. In physics-based prognostics methods, the physical models often involve finite element analysis for a single stand-alone component or a complex system considering the interactions of all the composing components. System dynamics modeling is also often used for the simulation of system vibration. In particular, the system dynamics at the faulty state is of our interest for the purposes of diagnostics and prognostics. Considering the system dynamics often facilitates to improve the fidelity of physical models.

After obtaining the physical quantities by FE analysis and system dynamics modeling, the damage progression model will predict the future degradation and estimate the RUL of the component. Most approaches track the degradation and define the time to failure as the time when the degradation indicator exceeds a predefined threshold. The

physical models used in the literature for component failure prediction are discussed as follows. Paris' law (Paris and Erdogen 1963) and its variants (Collipriest 1972; Inoue et al. 1999; Wheeler 1972) are often used to describe crack growth process. Kacprzynski et al. (2004) developed a 3D FE model for a cracked spiral bevel gear used in the helicopter transmission. Li and Lee (2005) and Li and Choi (2002) proposed a method of predicting RUL based on Paris' law for spur gears. The method to predict the lifetime of a shaft with a crack was found in (Oppenheimer and Loparo 2002). Researchers also employed Paris' law for spall expansion (Li et al. 1999; Kotzalas and Harris 2001; Roemer et al. 2008; Hannes and Alfredsson 2012) under rolling contact. Damage mechanics theory was investigated in (Xu and Sadeghi 1996; Marble and Morton 2006; Qiu et al. 2002) to predict the pitting process. The Archard's model was used to describe the surface wear evolution of gears in (Wu and Cheng 1993; Flodin and Andersson 1997; Bajpai and Kahraman 2004). For electrical devices, like battery, the degradation models include Arrhenius model and Eyring model, etc.

Physics based methods have high prediction accuracy, however, building an accurate physical model is not a trivial task, especially for complex engineering systems that involve many components, multiple failure modes and complicated working environment. Even if a high fidelity model is available, too much computational time may prohibit it from being used in real-time prognostics applications. In addition, physics based methods usually use deterministic model parameters without considering uncertainty. However, the times to failure exhibit a large variation even for identical components under well controlled environment. Hence, physics based methods fail to address the uncertainty in failure times.

2.2.3 Integrated prognostics methods

Data affect model in various ways. For example, prediction is usually performed after the fault is detected and assessed by analyzing the sensor data in diagnostics. In other words, the size of the initial fault actually serves as the initial condition in the prediction model. Data can also be used to calibrate model parameters and to estimate the fault state. Particle filter (Orchard and George 2007; Cadini et al. 2009; Zio and Peloni 2011) is a state-of-the-art integrated prognostics method. When the damage progression is modeled as a dynamic system, particle filter can deal with non-linear state transition and non-Gaussian noise. The fault state and the parameters can be estimated and updated using observations sequentially. Various uncertainty sources were considered by Sankararaman et al. (2011) including loading conditions, material properties, data uncertainty and model uncertainty. However, Bayesian inference was used to calibrate only three selected parameters based on a sensitivity analysis because Monte Carlo simulation prevented the full consideration of the effects of all the uncertainty sources on the failure prediction. Byington et al. (2009) used vibration diagnostics to update the model prediction for a cracked shaft. The model inputs were considered to be normal random variables. Monte Carlo simulation was used to obtain a distribution characterizing the dispersion in the fatigue life. Coppe et al. (2010) considered the RUL prediction for a structure that had a crack in it. A Bayesian framework was used to reduce the uncertainty in the material parameters in Paris' law. With the uncertainty reduction, prediction on RUL was more accurate. An et al. (2011) extended Coppe's work to identify correlated parameters simultaneously via Bayesian inference. Standard MCMC method was used to calculate Bayesian posterior joint distribution of the correlated parameters.

2.3   Physics of failure

Physics-based prognostics methods take advantage of physics of failure. This thesis focuses on the life prediction of gears. Two failure modes of gears will be considered. One is the tooth fracture due to cyclic bending stress, and the other one is the surface wear due to sliding contact during the mesh process. This section presents the physics laws behind the physical models of these two failure modes.

2.3.1 Fracture

Fracture is a serious and commonly seen failure mode in engineering structures and mechanical components, such as ship deck, aircraft fuselage, bridge, oil pipeline, bearing, gear and shaft. Fracture is caused by crack propagation. In this section, fracture mechanics will be discussed in terms of its FE solution. The methods are presented to calculate SIF and crack growth direction. In addition, fatigue life models are discussed which are used to describe the crack growth with loading cycles.

2.3.1.1   Fracture mechanics

Unexpected fractural failures jeopardize human safety and engineering system integrity, which stimulate the research efforts to predict and control their occurrence. The work in this thesis is constrained within the theory of linear elastic fracture mechanics. FE method is widely used for stress analysis in fracture problem, where the discontinuity in the geometry introduced by the crack gives rise to a singularity of the stress and strain fields near the crack tip. Special elements are needed to simulate this singularity in fracture FE models. The fracture modes in the continuum can be considered in three

18

forms according to the loading modes: Mode I – opening; Mode II – in-plane sliding; Mode III – anti-plane tearing. These three modes are shown in Figure 1.
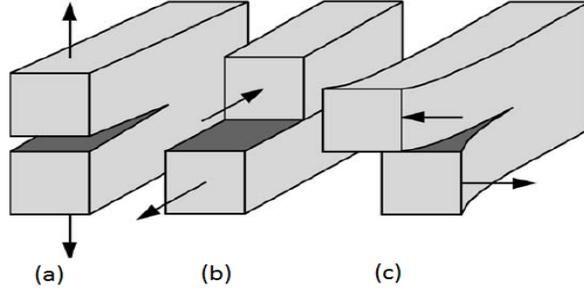
Figure 1. Three types of loading on a cracked body (Anderson 2005).

In the opening mode, the cracked body is loaded by tensile forces tending to open the crack; the sliding mode refers to in-plane shear loading, and the tearing mode corresponds to out-of-plane shear loading. SIF is an important quantity in fracture mechanics because it describes the stress field in the region near the crack tip and determines the crack propagation direction. Pertinent to the three loading conditions, there are three types of SIF: opening mode K_I, sliding mode K_{II}, and tearing mode K_{III}.

2.3.1.2   Crack tip singularity

From Williams (1957), the stress field near the crack tip in 2D fracture problem is shown in Figure 2. The normal stress $\sigma_{\theta\theta}, \sigma_{rr}$ and shear stress $\sigma_{r\theta}$ are expressed approximately in cylindrical coordinate system as follows:

$$\sigma_{\theta\theta} = \frac{1}{\sqrt{2\pi r}} \cos\frac{\theta}{2} \left[K_I \cos^2\frac{\theta}{2} - \frac{3}{2} K_{II} \sin\theta \right] \tag{2.2}$$

$$\sigma_{rr} = \frac{1}{\sqrt{2\pi r}} \cos\frac{\theta}{2} \left[K_I \left(1 + \sin^2\frac{\theta}{2}\right) + \frac{3}{2} K_{II} \sin\theta - 2K_{II} \tan\frac{\theta}{2} \right] \tag{2.3}$$

$$\sigma_{r\theta} = \frac{1}{2\sqrt{2\pi r}} \cos\frac{\theta}{2} \left[K_I \sin\theta + K_{II}(3\cos\theta - 1) \right] \tag{2.4}$$

19
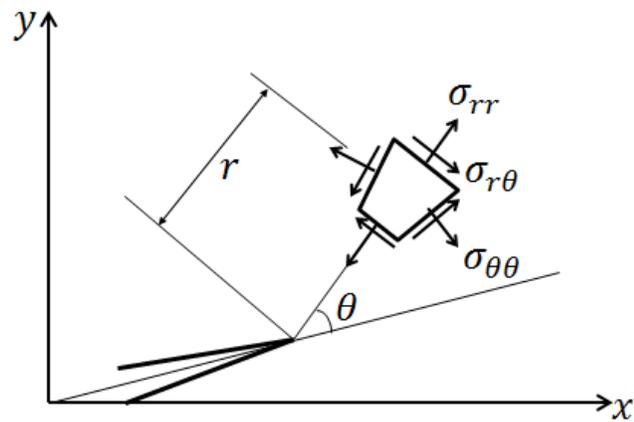
Figure 2.Two dimensional stress field near the crack tip.

A singularity of $r^{-1/2}$ in the stress along the ray originating from the crack tip is seen in the above expressions. The usual element type with polynomial shape function has difficulty in simulating such singularity, which motivates the invention of the special element called "singular" element. The essential novelty of the singular element is to create the $r^{-1/2}$ singularity along the element edge as the point approaches the crack tip by moving the element node from the middle to the quarter of the element edge, as shown in Figure 3.
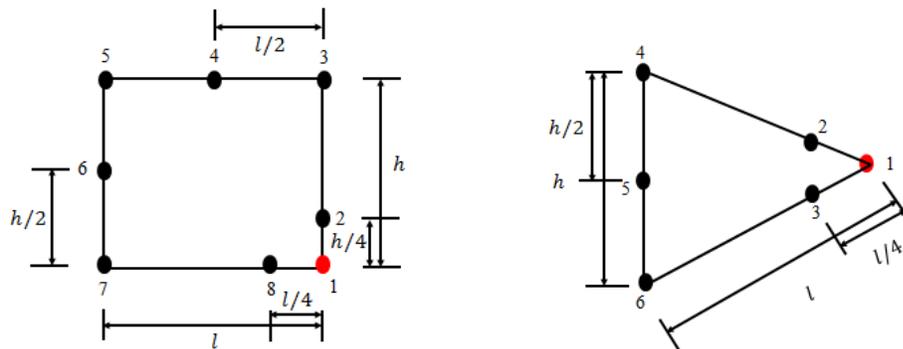
Figure 3. Singular elements.

The singular element reproduces the $r^{-1/2}$ singularity near the crack tip in FE formulation. Higher order elements have capabilities to approximate the real stress field more accurately. Within the framework of linear elastic fracture mechanics, stress and displacement fields in the area near the crack tip are determined by the SIFs. Techniques to calculate the SIFs include direct approach and energy approach. To be more specific, three methods are commonly used for computing SIFs under mixed Mode I and Mode II loading: (1) displacement correlation technique (Shih et al. 1976), (2) potential energy release rate technique (Narayana and Dattaguru 1996), (3) J-integral (Raju and Shivakumar 1990). The first technique is a direct approach and the rest two are energy based methods. In this thesis, the software of FRANC2D is adopted to compute the SIFs. FRANC2D is able to calculate SIFs using the three above methods. Direct approach based on the singular element is presented in the following section.

2.3.1.3   SIF calculation using direct approach

SIF dominates the stress states near the crack front fields under the condition of small-scale yielding. In order to predict the fracture process, including crack growth rate and trajectory, SIF should be the first to calculate.

The direct approach is a displacement correlation method. The idea is to correlate the local displacements with their theoretical asymptotic values. The SIFs appearing in these equalities are extracted as a scaling parameter. If the triangular singular elements are used around the crack tip, as shown in the Figure 4, the displacement correlation method gives the following formulas for SIFs:

21

$$K_I = \frac{G\sqrt{2\pi}}{\sqrt{l}(2 - 2v)}[4(v_b - v_d) + v_e - v_c] \qquad (2.5)$$

$$K_{II} = \frac{G\sqrt{2\pi}}{\sqrt{l}(2 - 2v)}[4(u_b - u_d) + u_e - u_c] \qquad (2.6)$$

where $u_i, v_i, i = a, b, c, d, e$ are the displacements obtained by FE method along x and y directions respectively; l is the length of element edge.
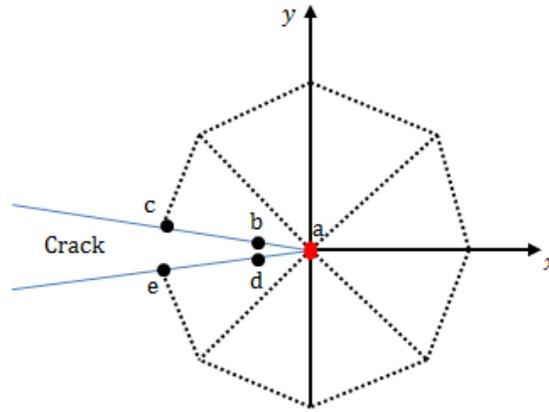
Figure 4. Triangular singular element near the crack tip.

Direct approach is simple and it expresses the SIFs as functions of the nodal displacements. The correlation form could have other variants. The accuracy depends on the local meshing. The use of singular element reduces the heavy computational work of refining the mesh size near the crack tip. The direct approach is used for the problem considered in this thesis.

2.3.1.4 Crack growth trajectory

Crack trajectory prediction is an important topic in fracture mechanics. It is also essential in health management in mechanical component suffering from crack fracture. Take the gear for example. For the sake of lightness, transmission system in the aircraft

22

and helicopter often utilizes gear set with thin rim. If the crack at the gear root grows towards to the rim, the failure will be catastrophic and the broken gear will cause damage to other critical components in the system. The consequence is more severe than the transverse crack trajectory across the tooth thickness. Lewicki and Ballarini (1997) did a research on the gear crack propagation path. Figure 5 shows the predicted crack propagation paths for gears with different backup ratio, from which it can be seen the crack towards to the rim in the thin-rimmed gear.
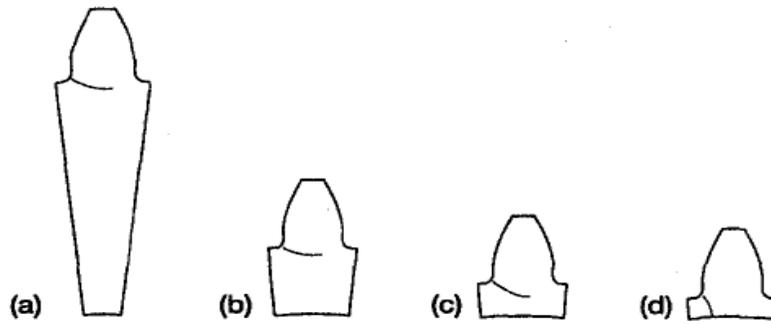
Figure 5. Crack propagation paths with different backup ratio (Lewicki and Ballarini 1997).

To determine the crack propagation direction, the stress state near the crack tip needs to be investigated. According to Williams (1957), the leading stress term is singular in r and dominates the stress field near the crack tip in the elastic solution. Erdogan and Sih (1963) proposed the first-order maximum tangential stress theory for isotropic materials. It stated that the crack extension started at the crack tip and would grow in the direction normal to the maximum tangential tensile stress. Mathematically, by taking the derivative of $\sigma_{\theta\theta}$ in (2.2) with respect to θ, the crack propagation angle is

$$\theta_c = 2 \tan^{-1}\left(\frac{1 - \sqrt{1 + 8(K_{II}/K_I)^2}}{4K_{II}/K_I}\right). \tag{2.7}$$

To obtain the 2D crack trajectory, a static stress analysis should be done for each crack growth step. Then the trace of the crack tip through all the extension steps consists of the trajectory.

2.3.1.5   Crack propagation life model

Many mechanical components in engineering systems are subject to cyclic loading and susceptible to fatigue crack. For example, the shaft is loaded by tension and compression alternatively; the gear tooth is loaded only when it meshes. The overall fatigue life of the mechanical component can be described by three stages: crack initiation, crack propagation and fast rupture. Once the initial crack is observed, fracture mechanics may be applied to predict the crack propagation. Component failure is usually defined by a threshold value of crack size. Crack propagation life model describes the crack evolution with loading cycles.

The fundamental crack propagation life model was proposed by Paris and Erdogan (1963). The model is well-known as Paris' law. It asserted that the crack growth rate was linear with SIF in a log-log scale. Paris' law is applicable in the stable crack propagation phase. The range of SIF during one loading cycle determines the crack size increment. By considering other factors which may affect the crack growth, Paris' law has other variants. The Collipriest crack growth model (Collipriest 1972) took three other factors into account: load ratio, crack instability near rupture, and SIF threshold. Inoue et al. (1991) developed a life model for the actual application of gears whose surface was treated. Hence, the fatigue life was influenced by different properties of the case and the core. Elber (1971) proposed an adaptation of Paris' law by taking into account the important

24

effect of plasticity-induced crack closure. The crack propagation life model can be further modified to consider other factors that affect the crack propagation, such as oxidation, corrosion, variable amplitude loading, random loading, etc.

2.3.2 Surface wear

Wear occurs when two surfaces are loaded together and have relative sliding and/or rolling motion. The asperity contacts will result in worn particles. Archard (1953) proposed a wear model (2.8) in which the wear rate, expressed as the worn volume removed per unit sliding distance, is proportional to the load.

$$\frac{V}{s} = K \frac{F_N}{H} \tag{2.8}$$

where V is the wear volume, s is the sliding distance, K is the dimensionless wear coefficient, H is the hardness of the softer contact surface, and F_N is the normal load. The wear coefficient K is interpreted as the probability that an asperity interaction results in the formation of a wear particle. Because wear depth is of more interest by engineers than wear volume, by dividing both sides of (2.8) by the contact area, and replacing K/H with a dimensional wear coefficient k, the wear model can have the following form in terms of the wear depth, h,

$$\frac{h}{s} = kp \tag{2.9}$$

where p is the contact pressure.

The Archard's wear model is now generally accepted as a suitable framework within which the quantitative analysis on wear progression can be discussed (Peterson and

Winer 1980). The wear process was seen as an initial value problem in (Andersson 2010) and was described by a differential equation as

$$\frac{dh}{ds} = kp \qquad (2.10)$$

2.4 Dynamic models of gearbox

Vibration signature is often used for fault diagnostics of the geared system. The faults, such as crack and pitting, affect the tooth stiffness and further affect the vibration of the system. A great deal of research work is invested to build dynamic models for the geared system to understand the torsional and/or lateral vibrations of the gearbox. These models assist in identifying the vibration sources in gear transmission.

Based on the physical laws in dynamics, many mathematical models to describe the gearbox dynamics were developed in (Bartelmus 2001). Mesh stiffness variation is one of the vibration excitation sources. Numerical and analytical methods to compute mesh stiffness were proposed in (Wang and Howard 2004; Howard et al. 2001; Chaari et al. 2009; Yang and Lin 1987). A one-stage gear set model with four degrees-of-freedom (DOF) considering torsional vibration was analyzed in by Lin et al. (1988), based on which the computer program DANST was developed in (Oswald et al. 1996). DANST can output the static transmission error, dynamic load, bending stress and other properties of spur gears. DANST was employed by Liou et al. (1992) to investigate the effect of contact ratio on spur gear dynamic load. Additionally, authors of (Lin et al. 1994; Tavakoli and Houser 1986) focused on the work of minimizing the dynamic load and transmission error by modifying the profile of spur gear. The research work in (Cornell

and Westervelt 1978; Kasuba and Evans 1981; Vedmar and Andersson 2003; Velex and Ajmi 2007) developed approaches for computation of dynamic load.

2.5 Uncertainty quantification

Prognosis as an essential part of CBM imposes major challenge for CBM system designer and practitioner because of large-grain inherent uncertainty. Prediction of the stochastic fault progression requires effective uncertainty management which is usually conducted in the framework of probability theory. In the problem of uncertainty quantification, the uncertainty sources should be firstly captured and represented so that the model could be randomizd and transformed into a stochastic form. Then, the uncertainty propagation is executed to quantify the effects of the random inputs on the output of interest, which could be the health indicators or RUL in the prognostic model. Based on the model output with certain confidence level, decisions can be made for maintenance planning, mission planning and logistics to meet the reliability and availability requirement. Due to the uncertainty inherent in prognostics, the predicted RUL takes the form of probability distribution rather than a single value. However, large uncertainty provides little information on decision making. Hence, a major objective of uncertainty management in prognostics is to reduce the uncertainty in the predicted RUL distribution, which indicates a better prediction precision.

Uncertainty propagation is part of uncertainty quantification. Its objective is to evaluate the uncertainty in model output given the uncertainty in model inputs while there is no explicit expression between the model inputs and the output. Monte Carlo simulation is often used for uncertainty propagation, in which a series of calls for the deterministic evaluations of the model output are needed to obtain sufficient samples of

model output. If the model evaluation is computationally intensive, it will be time-consuming to obtain a large number of samples on the model output. Especially when high dimensional uncertainty is considered, the convergence rate of Monte Carlo simulation is unacceptably slow. The computational burden to deal with the high dimensional uncertainty is known as "curse of dimensionality". Hence, an effective sampling strategy is of importance to improve the computational efficiency.

This section starts with the introduction of two uncertainty propagation methods: Monte Carlo simulation in section 2.5.1., and the PCE in section 2.5.2. After that, Bayesian inference is discussed as a natural way for uncertainty quantification in section 2.5.3.

2.5.1 Monte Carlo simulation

Monte Carlo simulation is the most widely used method for uncertainty propagation. It is a statistical sampling method popularized in the 1940s. Its distinguished merit is the flexibility to be implemented easily for almost any general model $Y = \mathcal{H}(X)$, which relates the input X and output Y point-wisely. To obtain the statistical properties of Y, a set of samples $\{X^{(1)}, X^{(2)}, \cdots, X^{(M)}\}$ are drawn randomly from the joint distribution $f_X(X)$. The output $Y^{(i)}$ corresponding to the input $X^{(i)}$ is $Y^{(i)} = \mathcal{H}(X^{(i)}), i = 1,2,\cdots,M.$ The mean and standard deviation of Y are estimated from these samples of output:

$$\hat{\mu}_Y = \frac{1}{M} \sum_{i=1}^{M} Y^{(i)} \tag{2.11}$$

$$\hat{\sigma}_Y^2 = \frac{1}{M-1} \sum_{i=1}^{M} (Y^{(i)} - \hat{\mu}_Y)^2. \tag{2.12}$$

From the central limit theorem, the distribution of $\hat{\mu}_Y$, as $M \to \infty$, will converge to a Gaussian distribution with mean μ_Y and standard deviation $\sigma_Y/\sqrt{M}$. Hence the convergence rate of Monte Carlo simulation is inversely proportional to the square root of the number of samples, denoted as $O(M^{-1/2})$ (Xiu 2010). This square root convergence rate is quite slow. For the large and complex system, of which a single model realization is time-consuming, Monte Carlo simulation becomes unacceptable in terms of computational time taken to achieve a good accuracy. However, the convergence rate of $O(M^{-1/2})$ is independent of the dimension of random space, which is a notable advantage. Based on the set of output samples $\{Y^{(1)}, Y^{(2)}, \cdots, Y^{(M)}\}$, the PDF of the output can be estimated using an histogram. Smoother representations may be obtained through kernel smoothing techniques, see (Wand and Jones 1995).

2.5.2 PCE

Instead of sampling, PCE aims to approximate the model output by a simple polynomial expression. It is an efficient method to analyze and quantify the effects of random inputs on the stochastic process governed by ordinary/partial differential equations. PCE is essentially a spectral representation in random space, and exhibits fast convergence rate when the expanded function depends smoothly on the random parameters (Xiu 2007). By selecting the optimal type of orthogonal polynomials, the convergence rate could reach exponential if the function is analytic, i.e., infinitely smooth (Xiu and Karniadakis 2002).

A large number of research papers have been published on the computational efficiency and accuracy achieved by PCE in many fields, most of which are nonlinear

problems with high dimensional uncertainty, e.g. fluid dynamics (Pettersson et al. 2009; Chantrasmi et al. 2006; Chantrasmi et al. 2009; Canuto et al. 1998; Najm 2009), chemistry reaction (Najm et al. 2009), solid mechanics (Eldred 2009; Ghosh and Farhat, 2007), etc. For highly-nonlinear problem, a sufficiently high order polynomial space is required, as reported in (Debusschere et al. 2004; Eldred et al. 2008).

The quantities to be solved in PCE are those coefficients in the polynomial expansion. There are two basic methods to calculate them: stochastic Galerkin method and stochastic collocation method. Stochastic Galerkin method is well developed to solve the stochastic problems. Reagan et al. (2004) and Najm et al. (2009) investigated uncertainty quantification in chemical systems using stochastic Galerkin method. In computational fluid dynamics, stochastic Galerkin method has many applications as well (Canuto et al. 1988; Najm 2009). However, this method reformulates the stochastic problem into a larger set of coupled equations of coefficients, and thus the computational work is still cumbersome. Furthermore, when the original problem takes a highly complex and nonlinear form, the derivation of the Galerkin model can be nontrivial, sometimes impossible. Due to these limitations of stochastic Galerkin method, there has been a growing interest in another class of method known as stochastic collocation method. Stochastic collocation method has the power to take advantage of both stochastic Galerkin method and Monte Carlo simulation to achieve high resolution with the ease of implementation (Xiu and Hesthaven, 2005). It is widely acknowledged that stochastic collocation method collaborated with sparse grid has appealing capability in handling high-dimensional random space. With the uncertainty dimension increases, sparse grid algorithm is able to mitigate the "curse of dimensionality", and meanwhile maintains the

numerical integration accuracy inherited from 1D integration rule as much as possible.

In the stochastic collocation method, high dimensional integration is involved when a large number of random variables are considered. These integrations are computational intractable that must resort to numerical methods, in which a set of integration points and corresponding weights is determined according to the weight function in the integration. From one-dimensional to multi-dimensional integration, the natural extension is to use tensor product. If the variable dimension is d and in each dimension n points are used, the total number of integration points is n^d. As the dimension increases, the number of points grows very fast resulting in the well-known "curse of dimensionality". To mitigate this, Smolyak (1964) proposed an algorithm to select a nodal set which consists of an algebraic sum of low-order tensor products based on one-dimensional quadrature rule in such a way that an integration property for one dimension is preserved for high dimensions as much as possible (Xiu, 2010). The nodal set obtained by Smolyak algorithm is called the sparse grid. Figure 6 shows the comparison of integration points for 2D integration. In the figure, the points are obtained based on extrema of Chebyshev polynomials, and a great reduction of number of nodes is observed in sparse grid compared to tensor product.
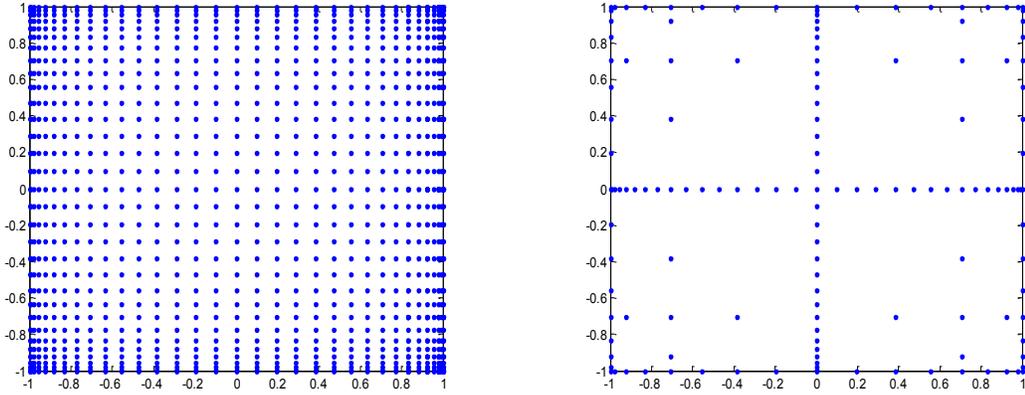
(a). 1089 points using tensor product.   (b). 145 points using sparse grid.

Figure 6. Comparison of nodal sets for tensor product and sparse grid.

2.5.3 Bayesian inference

Bayesian inference is naturally amenable to uncertainty quantification for model parameters. Consider the parameters of interest, denoted by θ, with a prior distribution $\pi(\theta)$. Meanwhile a set of observations $\boldsymbol{d} = \{d_1, d_2, \cdots, d_n\}$ is available and its marginal density function is $f(\boldsymbol{d})$. Bayesian inference generates a posterior distribution $\pi(\theta|\boldsymbol{d})$ for parameters given the prior density and the observational data:

$$\pi(\theta|\boldsymbol{d}) = \frac{L(\boldsymbol{d}|\theta)\pi(\theta)}{f(\boldsymbol{d})} \qquad (2.13)$$

where $L(\boldsymbol{d}|\theta)$ describes the likelihood to obtain these observations given parameters. Because the observations are fixed, the denominator in (2.13) is a constant, which is used to normalize the posterior density so that its integral is one. Hence, the posterior density is proportional to the product of the likelihood and the prior density.

In most cases, the denominator in Bayesian inference is a high dimensional integral, which is computational intractable. Except when the prior and the likelihood form the so

called "conjugate pair", numerical methods are needed to compute this integration. In contrast, MCMC method is to directly obtain the samples from posterior distribution.

MCMC method is based on the fact that the generated samples in the stationary Markov chain is exactly the samples from the posterior distribution. MCMC can be applied when the target density is known up to a constant: $p(x) \propto \tilde{p}(x)$. In this thesis, the implementation of MCMC uses Hasting-Metropolis sampling algorithm and its procedure is as follows:

1. Initialize $i = 0$ and choose a starting point x_0, and the list $X = \{x_0\}$.

2. Repeat the following steps until the convergence is reached

   - Sample $x^* \sim q(x^*|x_i)$, where $q(\cdot)$ is the proposal density function.

   - Calculate the acceptance ratio $\alpha(x^*, x_i) = \min[1, \frac{\tilde{p}(x^*)q(x_i|x^*)}{\tilde{p}(x_i)q(x^*|x_i)}]$.

   - Sample $u \sim$ Uniform $(0,1)$.

   - If $u < \alpha$, set $x_{i+1} = x^*$, otherwise, set $x_{i+1} = x_i$.

   - $X = \{x_i, x_{i+1}\}$.

   - $i = i + 1$

3. The samples X are used to construct the approximation to $p(x)$.

The time for the chain to reach its stationary distribution is affected by the choice of the initial staring value x_0 and the proposal density function. The most commonly used proposal density is the random walk around the origin. The step size of move is critical for the algorithm performance. If the step size is very small, most samples will be accepted and the chain will take long time to converge. If the step size is very large, the samples have more chance to fall into the tails of the target distribution, which results in

very low acceptance ratio. The series obtained by MCMC have strong correlation so that

a large number of samples are needed to make inference about the posterior distribution.

Chapter 3.   Dynamics simulation for one-stage gearbox

3.1   Overview

Vibration analysis of the mechanical system is the most commonly used technique to diagnose the fault of the mechanical components, because the fault occurrence will affect the overall vibration of the whole system. Tooth fracture is a commonly seen failure mode of gears. Crack often occurs near the tooth root where the highest bending stress appears due to repetitive cyclic loading. The crack appearance affects the tooth stiffness. Therefore, the total mesh stiffness will be affected accordingly when the cracked tooth comes into mesh. Since the force acting on the tooth is determined by the mesh stiffness and the tooth deflection, the crack will affect the dynamic load on the tooth. Dynamic load can be considered as the sum of the transmitted load and an incremental load. The transmitted load is the static load. The incremental load is due to many factors, such as spacing errors, unbalance, deformation of teeth and velocity. Many publications consider the crack propagation under a static load, which excludes the effect of crack propagation on the load acting on the tooth. Li and Lee (2005) took this effect into account to calculate the dynamic load on the cracked tooth. The mesh stiffness was calculated using Fourier series approximation. In this thesis, we also consider the crack effect to obtain an accurate load to drive the crack propagation. Tian et al. (2004) applied the potential energy method to compute the stiffness of the cracked gear tooth assuming the crack propagates along a straight line. The work in this chapter extends the method to compute the stiffness of the cracked tooth with a crack that has a curved shape. A 6-DOF dynamic

35

motion system is used to simulate one-stage gearbox dynamics. Dynamic load on the cracked tooth is calculated for a complete mesh period. The materials in this chapter have been published in (Zhao et al. 2013a).

3.2   Spur gears

Spur gear is the most commonly used gear type. It has simple parallel and symmetric geometry and can change the velocity of the transmission system. Spur gear bears an involute profile to minimize the transmission error and the vibration noise. In this subsection, a brief introduction is presented about the involute geometry and the mesh process of spur gears.

Involute curve can be considered as a path traced by the end of a string, which is originally wrapped on a circumference of a circle, when the string is unwrapped from the circle. This circle is called base circle. The driving gear (pinion) and the driven gear (gear) have the base circles with the radiuses of R_{b1}, R_{b2} respectively. The movement of contact points when spur gears are meshing is along the line of action, which is the common tangent to the base circles and meanwhile is also the common normal to the tooth profiles. The top points of the teeth consist of outside circle with radiuses of R_{o1}, R_{o2}. The basic parameters characterize a gear set include diametral pitch P, number of teeth N_1, N_2 and pressure angle α_0. The circles which have the radiuses of $R_{p1} = R_{b1}/\cos\alpha_0$, $R_{p2} = R_{b2}/\cos\alpha_0$ are called pitch circles. A pair of meshing gears is ought to have the same diametral pitch. The diametral pitch is defined as the ratio of teeth number and pitch circle diameter. The pressure angle is the complementary angle between the line of action and the line connecting two centers O_1, O_2 of base circles. $\overline{O_1 O_2}$ is called center distance.

According to Townsend (1992) and Litvin (1994), the following expressions are used to determine these parameters mentioned above:

$$R_{b1} = \frac{N_1}{2P} \cos \alpha_0 \tag{31}$$

$$R_{b2} = \frac{N_2}{2P} \cos \alpha_0 \tag{3.2}$$

$$R_{O1} = \frac{N_1 + 2}{2P} \tag{3.3}$$

$$\overline{O_1 O_2} = \frac{N_1 + N_2}{2P} \tag{3.4}$$

The main parameters of the gear and pinion used in this thesis are the same as those in Wu et al. (2008), shown in Table 1.

Table 1. Main parameters of the gears

Young's modulus	$E = 2.068 \times 10^{11}$ Pa
Poisson's ratio	$v = 0.3$
Pressure angle	$20°$
Diametral pitch	8 in^{-1}
Width of teeth	$W = 0.016$m
Number of teeth on pinion	$N_1 = 19$
Number of teeth on gear	$N_2 = 48$

3.3   Potential energy method for mesh stiffness calculation

Yang and Lin (1987) modeled the gear tooth as a cantilevered beam. The base of the beam was assumed not to experience any deflection. The authors partitioned the total potential energy stored in the meshing gear set into three categories: Hertzian energy, bending energy and axial compressive energy. Tian et al. (2004) added shear energy, which was proven to be a significant contribution to the total effective mesh stiffness. Also, the authors gave the formulas to calculate the mesh stiffness for the gear tooth with a straight crack, inclined at a constant intersection angle to upright vertical. This thesis extends the stiffness calculation to a tooth with a curved crack. This section is organized as follows. The potential energy method to calculate the stiffness of a healthy gear tooth is presented in section 3.3.1. Details are found in (Yang and Lin 1987) and (Tian et al. 2004). In section 3.3.2, the formulas to calculate the stiffness of a tooth with a straight crack is given, referenced on (Tian et al. 2004). In section 3.3.3, the method is developed to calculate the stiffness of a tooth with a curved crack.

3.3.1 Stiffness of a healthy tooth

The Hertzian, bending, shear and axial compressive energy stored in a tooth are denoted by $U_h, U_b, U_s,$ and U_a, respectively:

$$U_h = \frac{F^2}{2k_h}, U_b = \frac{F^2}{2k_b}, U_s = \frac{F^2}{2k_s}, U_a = \frac{F^2}{2k_a}. \tag{3.5}$$

where k_h, k_b, k_s, k_a stand for the Hertzian, bending, shear and axial compressive stiffness. F is the force acting on the mating tooth at the contact point.

The Hertzian contact stiffness of a pair of meshing teeth with the same material is considered to be constant, independent of the contact position (Yang and Sun 1985). The Hertzian stiffness is expressed as

$$k_h = \frac{\pi E W}{4(1 - v^2)} \, ,$$
(3.6)

where E, W, v represent Young's modulus, width of tooth, and Poisson's ratio. The other three types of stiffness are dependent on the contact point. The involved parameters are depicted in Figure **7**. Suppose the current contact point is C and the corresponding pedal on the base circle is Z_C. The angle between $\overline{CZ_C}$ and the horizontal line is α_1, which determines the contact point as the gears rotate. α_1 is also the angle for the force decomposition into horizontal and vertical: $F_b = F\cos\alpha_1$, $F_a = F\sin\alpha_1$. Denote any point on the tooth profile below C as B, with the pedal Z_B. The angle between $\overline{BZ_B}$ and the horizontal line is α which is the variable that locates all the points below C on the tooth profile and is also the integration variable in the potential energy. α_2 as shown in the Figure 7 is the half of the base tooth angle, which serves as the upper limit in integration for energy.

Figure 7. Spur gear geometry

The bending energy stored in a meshing gear tooth, based on beam theory, can be obtained by

$$U_b = \int_0^d \frac{M^2}{2EI_x} dx = \int_0^d \frac{[F_b(d - x) - F_a h]^2}{2EI_x} dx, \tag{3.7}$$

The shear energy is given by

$$U_s = \int_0^d \frac{1.2 F_b{}^2}{2GA_x} dx, \tag{3.8}$$

$$G = \frac{E}{2(1 + v)}. \tag{3.9}$$

40

And the axial compressive energy is

$$U_s = \int_0^d \frac{F_a{}^2}{2EA_x} \, dx. \tag{3.10}$$

In the above formulas, I_x and A_x represent the area moment of inertia of the section, and the area of the section, where the distance from the tooth root is x,

$$I_x = \frac{1}{12}(2h_x)^3 W = \frac{2}{3}h_x^3 W, \tag{3.11}$$

$$A_x = 2h_x L, \tag{3.12}$$

$$h_x = R_b[(\alpha + \alpha_2)\cos\alpha - \sin\alpha]. \tag{3.13}$$

The formulas for bending, shear and axial compressive stiffnesses are as follows:

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{3\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2 - \alpha)\cos\alpha}{2EW[\sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} \, d\alpha, \tag{3.14}$$

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{1.2(1 + v)(\alpha_2 - \alpha)\cos\alpha \cos^2\alpha_1}{EW[\sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} \, d\alpha, \tag{3.15}$$

$$\frac{1}{k_a} = \int_{-\alpha_1}^{\alpha_2} \frac{(\alpha_2 - \alpha)\cos\alpha \sin^2\alpha_1}{2EW[\sin\alpha + (\alpha_2 - \alpha)\cos\alpha]} \, d\alpha. \tag{3.16}$$

The total effective mesh stiffness between a pair of meshing teeth is

$$\frac{1}{k_t} = \frac{1}{k_h} + \frac{1}{k_{b1}} + \frac{1}{k_{s1}} + \frac{1}{k_{a1}} + \frac{1}{k_{b2}} + \frac{1}{k_{s2}} + \frac{1}{k_{a2}}, \tag{3.17}$$

where the subscripts 1 and 2 represent driving gear and driven gear, respectively. It can be seen from the expression that the mesh stiffness is a function of contact point. To

characterize the mesh stiffness varying with time, it requires to know which pairs of teeth are in mesh and where are the contact points. The contact ratio of the considered gear set is between 1 and 2. Hence, the single-pair-contact and double-pair-contact will alternate as gears rotate. So for the double-pair-contact, the total effective mesh stiffness is

$$k_t = \sum_{j=1}^{2} \frac{1}{\frac{1}{k_{h,j}} + \frac{1}{k_{b1,j}} + \frac{1}{k_{s1,j}} + \frac{1}{k_{a1,j}} + \frac{1}{k_{b2,j}} + \frac{1}{k_{s2,j}} + \frac{1}{k_{a2,j}}}, \qquad (3.18)$$

where $j = 1$ represents the first pair of meshing teeth, and $j = 2$ represents the second pair.

3.3.2 Stiffness of a cracked tooth - straight crack

As mentioned, the crack occurrence will affect the tooth stiffness and the total effective mesh stiffness. To incorporate the dynamic load into the crack propagation, which will be discussed in Chapter 4, the stiffness of a cracked tooth should be calculated first. Because pinion is the driving gear, experiencing more frequent fatigue loading, this thesis considers a crack at the pinion root. Tian et al. (2004) considered the shallow crack, whose length was not passing the central line. Wu et al. (2008) extended the crack length across the whole tooth thickness. Both of the work assumed the intersection angle ν between the crack path and the central line of the tooth as constant, which means the crack was assumed to propagate along a straight line. The Hertzian and axial compressive stiffness were not affected by the crack. The crack will influence the bending and the shear stiffness because the crack causes the change in the effective area moment of inertia and the area of the cross section. As shown in Figure 8, a crack with length q is inserted at the tooth root. The inclined angle with vertical is ν. The area moment of inertia and

the area of cross section at a distance of x from the tooth root (when $q \sin v - R_{b1} \sin \alpha_2 < 0$) are

$$I_{xc} = \begin{cases} \dfrac{1}{12}(h_c + h_x)^3 W, & \text{if } x \le g_c, \\ \dfrac{1}{12}(2h_x)^3 W = \dfrac{2}{3} h_x^3 W, & \text{if } x > g_c, \end{cases} \tag{3.19}$$

$$A_{xc} = \begin{cases} (h_c + h_x)W, & \text{if } x \le g_c, \\ 2h_x W, & \text{if } x > g_c, \end{cases} \tag{3.20}$$

$$h_c = R_{b1} \sin \alpha_1 - q \sin v. \tag{3.21}$$

By noticing the integration limits may cover different expressions for the area moment of inertia and the cross section area, the calculation may proceed according to different cases, depending on the contact position and the crack tip position. The complete set of formulas can be found in (Wu et al. 2008).
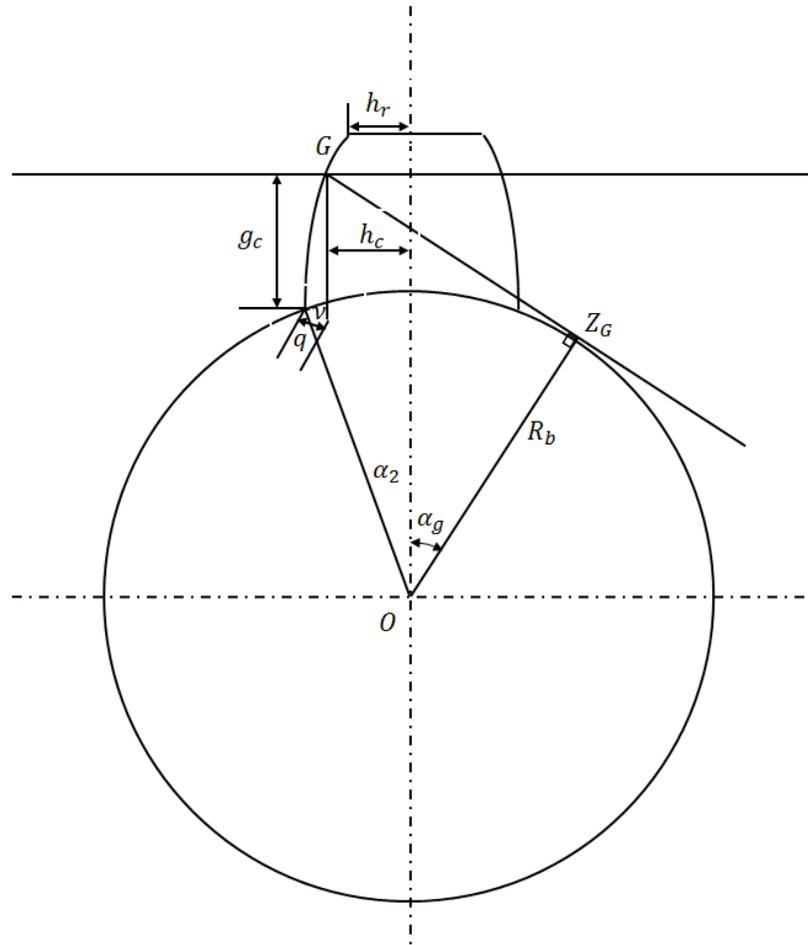
Figure 8. Cracked tooth geometry.

3.3.3 Stiffness of a cracked tooth - curved crack

The crack propagation direction is determined by the stress field near the crack tip, not necessarily along a straight line. Actually, in most cases the 2D crack has a curved shape. This thesis removes the assumption of a straight crack path made in (Yang and Lin 1987) and (Tian et al. 2004), and uses the FE stress analysis to determine the direction of crack propagation. The curved crack propagation path is formed by connecting a series of straight crack increments. Hence, compared to the straight crack path, the intersection angle β varies as crack grows, not a constant any more. Because the crack occurrence

only affects the bending stiffness and the shear stiffness, in this section only the formulas associated with these two types of tooth stiffness are given. Actually, when the gear tooth is considered as cantilever beam, the area of moment inertia and the cross section area are the only differences between a healthy tooth and a cracked tooth. Therefore, the method can be easily extended to deal with the curved crack as long as the locus of the crack tips at each crack propagation step is available.

The initial crack is inserted at the pinion tooth root with size of a_0. The procedure to calculate the tooth stiffness with a curved crack path is given as follows. As shown in Figure 9, the crack increment at each crack extension step is set to Δa. The crack tip is denoted by T_i, where the index i represents the crack propagation step. The crack grows by Δa in the direction determined by (2.7). Because the associated formulas to compute cracked tooth stiffness are related to four different cases, depending on the teeth contact point in mesh and the crack tip position as well, the index of $i = 1, 2, 3, 4$ in Figure 9 only symbolizes the four mentioned typical cases, and it does not mean that there are only these four crack tips.

The base circle of pinion centers at O with the radius of R_{b1}. The contact point C is travelling along the tooth profile $\widetilde{SM}$, and the angle of α_1 is determined by the tangential line passing C. Because the force F is applied at the contact point C, perpendicular to the tangential line, the angle α_1 also serves as the force decomposition angle to the horizontal direction $F_b = F\cos\alpha_1$ and vertical direction $F_a = F\sin\alpha_1$. Additionally, the points G_i represent the intersectional points between the vertical line passing the crack tip and the tooth profile. And Z_i are the pedals on the base circle of the tangential line passing G_i. Accordingly, g_i is the distance from G_i to the tooth root S, and α_{gi} is the angle between

G_iZ_i and OZ_i. If the crack tip passes the central line, G_i' and G_i are symmetric about the central line OP, and the associated α_{gi} is defined as the angle between $G_i'Z_i$ and OZ_i. Lastly, α_2 represents half of the base tooth angle.

Figure 9. Tooth geometry with a curved crack.

Essentially, the different ways in calculating I_x and A_x according to the crack tip location as well as the contact point position determine the four mentioned circumstances to calculate the tooth stiffness of a cracked tooth. These four cases are addressed below. Let the distance between tooth root S and G_iT_i be u_i. As said before, the purpose of the index $i = 1, 2, 3, 4$ is to indicate the four cases, not meaning there are only these four crack tip locations.

Case 1. Crack tip $= T_1$ (i.e., $h_{c1} \geq h_r$),

In this case,

$$I_x = \begin{cases} \dfrac{1}{12}(h_{c1} + h_x)^3 W, & \text{if } x \leq g_1, \\ \dfrac{1}{12}(2h_x)^3 W, & \text{if } x > g_1, \end{cases} \tag{3.22}$$

$$A_x = \begin{cases} (h_{c1} + h_x)W, & \text{if } x \leq g_1, \\ 2h_x W, & \text{if } x > g_1. \end{cases} \tag{3.23}$$

Case 1.1. Contact point is above G_1 (i.e., $\alpha_1 > \alpha_{g1}$) .

The bending stiffness of the cracked tooth is

$$\frac{1}{k_b}$$

$$= \int_{-\alpha_{g1}}^{\alpha_2} \frac{12\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2 (\alpha_2 - \alpha)\cos\alpha}{EW\left[\sin\alpha_2 - \dfrac{u_1}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha\right]^3} d\alpha$$

$$+ \int_{-\alpha_1}^{-\alpha_{g1}} \frac{3\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2 (\alpha_2 - \alpha)\cos\alpha}{2EW[\sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} d\alpha. \tag{3.24}$$

The shear stiffness is

$$\frac{1}{k_s}$$

$$= \int_{-\alpha_{g1}}^{\alpha_2} \frac{2.4(1+v)(\alpha_2-\alpha)\cos\alpha\cos^2\alpha_1}{EW\left[\sin\alpha_2 - \frac{u_1}{R_{b1}} + \sin\alpha + (\alpha_2-\alpha)\cos\alpha\right]}d\alpha$$

$$+ \int_{-\alpha_1}^{-\alpha_{g1}} \frac{1.2(1+v)(\alpha_2-\alpha)\cos\alpha\cos^2\alpha_1}{EW[\sin\alpha + (\alpha_2-\alpha)\cos\alpha]}d\alpha. \tag{3.25}$$

Case 1.2. Contact point is below G_1 (i.e., $\alpha_1 \le \alpha_{g1}$) .

The bending stiffness and shear stiffness are given by

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{12\{1+\cos\alpha_1[(\alpha_2-\alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2-\alpha)\cos\alpha}{EW\left[\sin\alpha_2 - \frac{u_1}{R_{b1}} + \sin\alpha + (\alpha_2-\alpha)\cos\alpha\right]^3}d\alpha, \tag{3.26}$$

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{2.4(1+v)(\alpha_2-\alpha)\cos\alpha\cos^2\alpha_1}{EW\left[\sin\alpha_2 - \frac{u_1}{R_{b1}} + \sin\alpha + (\alpha_2-\alpha)\cos\alpha\right]}d\alpha. \tag{3.27}$$

Case 2. Crack tip $= T_2$ (i.e., $h_{c2} < h_r$)

In this case,

$$I_x = \frac{1}{12}(h_{c1} + h_x)^3 W, \tag{3.28}$$

$$A_x = (h_{c2} + h_x)W, \tag{3.29}$$

based on which, the bending stiffness, and shear stiffness are respectively obtained by

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{12\{1+\cos\alpha_1[(\alpha_2-\alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2-\alpha)\cos\alpha}{EW\left[\sin\alpha_2 - \frac{u_2}{R_{b1}} + \sin\alpha + (\alpha_2-\alpha)\cos\alpha\right]^3}d\alpha, \tag{3.30}$$

48

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{2.4(1+v)(\alpha_2 - \alpha)\cos\alpha\cos^2\alpha_1}{EW[\sin\alpha_2 - \frac{u_2}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]} d\alpha. \tag{3.31}$$

Case 3. Crack tip $= T_3$ (i.e., $h_{c3} < h_r$)

In this case,

$$I_x = \frac{1}{12}(h_x - h_{c3})^3 W, \tag{3.32}$$

$$A_x = (h_x - h_{c3})W, \tag{3.33}$$

so the bending, and shear stiffness are

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{12\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2 - \alpha)\cos\alpha}{EW[-\frac{u_3}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} d\alpha \tag{3.34}$$

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{2.4(1+v)(\alpha_2 - \alpha)\cos\alpha\cos^2\alpha_1}{EW\left[-\frac{u_3}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha\right]} d\alpha. \tag{3.35}$$

Case 4. Crack tip $= T_4$ (i.e., $h_{c4} \geq h_r$)

In this case,

$$I_x = \frac{1}{12}(h_x - h_{c4})^3 W, \tag{3.36}$$

$$A_x = (h_x - h_{c4})W. \tag{3.37}$$

Case 4.1. Contact point is above G_4' (i.e., $\alpha_1 > \alpha_{g4}$) .

$$\frac{1}{k_b} = \int_{-\alpha_{g4}}^{\alpha_2} \frac{12\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2 - \alpha)\cos\alpha}{EW[-\frac{u_4}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} d\alpha \tag{3.38}$$

$$\frac{1}{k_s} = \int_{-\alpha_{g4}}^{\alpha_2} \frac{2.4(1+v)(\alpha_2 - \alpha)\cos\alpha\cos^2\alpha_1}{EW[-\frac{u_4}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]} d\alpha. \qquad (3.39)$$

Case 4.2. Contact point is below G_4' (i.e., $\alpha_1 \leq \alpha_{g4}$) .

$$\frac{1}{k_b} = \int_{-\alpha_1}^{\alpha_2} \frac{12\{1 + \cos\alpha_1[(\alpha_2 - \alpha)\sin\alpha - \cos\alpha]\}^2(\alpha_2 - \alpha)\cos\alpha}{EW[-\frac{u_4}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]^3} d\alpha \qquad (3.40)$$

$$\frac{1}{k_s} = \int_{-\alpha_1}^{\alpha_2} \frac{2.4(1+v)(\alpha_2 - \alpha)\cos\alpha\cos^2\alpha_1}{EW[-\frac{u_4}{R_{b1}} + \sin\alpha + (\alpha_2 - \alpha)\cos\alpha]} d\alpha. \qquad (3.41)$$

So far, we have obtained the formulas to calculate the bending stiffness, and shear stiffness of a cracked tooth at any crack tip position, and any contact point. No matter what the crack shape is, as long as the crack tip position is identified, i.e., u_i is known, these two types of stiffness could be derived by the above formulas. With the Hertzian stiffness in (3.6), and the axial compressive stiffness in (3.16), the total effective mesh stiffness is ready to use in the set of dynamic equations.

3.4   A 6-DOF dynamic model of one-stage gearbox

Once the total effective mesh stiffness is known, the dynamic load on the cracked tooth can be calculated. In this thesis, a 6-DOF dynamic model developed by Wu et al. (2008) is used to simulate the dynamic motion of a one-stage gearbox, shown in Figure 10. This dynamic model includes torsional and lateral motions. It is driven by an electric motor with the torque M_1 and loaded with the torque M_2. The motor shaft and the shaft that the pinion mounts on are coupled by a flexible input coupling. Similarly, the output coupling couples the output shaft and the shaft that driven gear mounts on. The shafts are

50

supported by bearings. The stiffness and the damping considered are for meshing gears, couplings and bearings. The friction during meshing is ignored here so the vibration in x direction is a free vibration, for which when the system is in a stable state, the vibration in x direction disappears due to the damping.

The governing motion equations (Wu et al. 2008) of the gearbox dynamics are as follows:

$$m_1 \ddot{y}_1 = k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2) + c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2) - k_1 y_1 - c_1 \dot{y}_1 \quad (3.42)$$

$$m_2 \ddot{y}_2 = k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2) + c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2) - k_2 y_2 - c_2 \dot{y}_2 \quad (3.43)$$

$$I_1 \ddot{\theta}_1 = k_p(\theta_m - \theta_1) + c_p(\dot{\theta}_m - \dot{\theta}_1)$$

$$-R_{b1}\left(k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2) + c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2)\right) \quad (3.44)$$

$$I_2 \ddot{\theta}_2 = -k_g(\theta_2 - \theta_b) - c_g(\dot{\theta}_2 - \dot{\theta}_b)$$

$$-R_{b2}\left(k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2) + c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2)\right) \quad (3.45)$$

$$I_m \ddot{\theta}_m = M_1 - k_p(\theta_m - \theta_1) - c_p(\dot{\theta}_m - \dot{\theta}_1) \quad (3.46)$$

$$I_b \ddot{\theta}_b = -M_2 + k_g(\theta_2 - \theta_b) + c_g(\dot{\theta}_2 - \dot{\theta}_b). \quad (3.47)$$
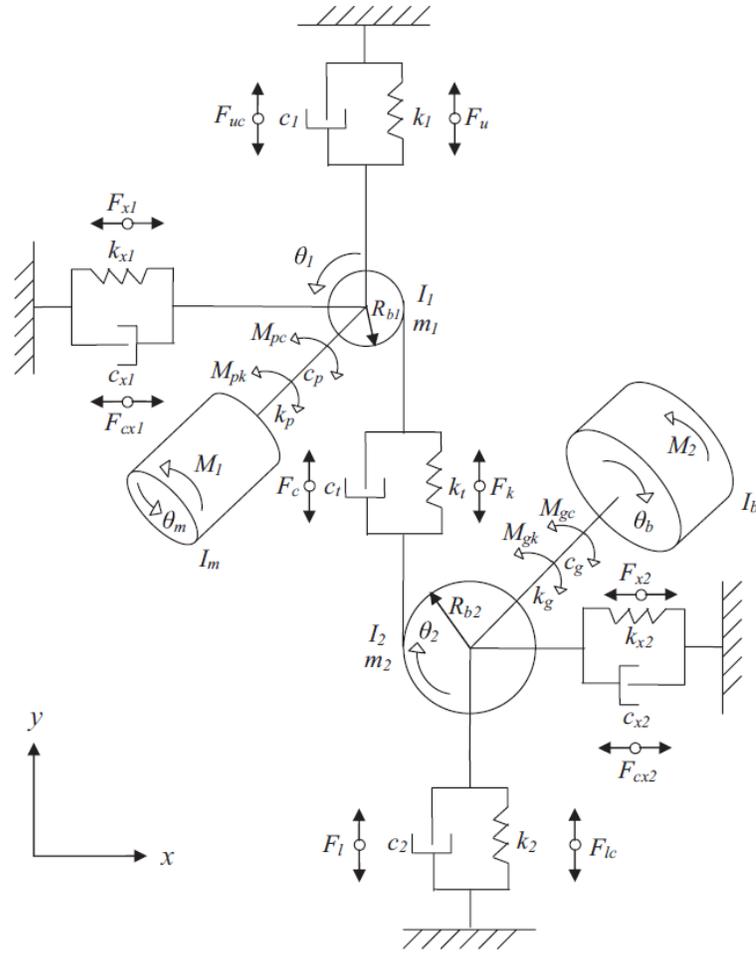
Figure 10. Dynamics model of gearbox with 6 DOFs (Wu et al. 2008).

where

I_m/I_b	mass moment of inertia of the motor/load
I_1/I_2	mass moment of inertia of the pinion/gear
M_1/M_2	input/output torque
m_1/m_2	mass of the pinion/gear
R_{b1}/R_{b2}	base circle radius of pinion/gear
k_p/k_g	torsional stiffness of the input/output flexible coupling
c_p/c_g	damping coefficient of the input/output flexible coupling

k_1/k_2 vertical radial stiffness of the input/output bearings

c_1/c_2 vertical radial damping coefficient of the input/output bearings

k_t total mesh stiffness

c_t mesh damping coefficient

y_1/y_2 linear displacement of pinion/gear in the y direction

θ_1/θ_2 angular displacement of pinion/gear

θ_m/θ_b angular displacement of motor/load

3.5   Dynamic load calculation for the cracked tooth

The system (3.42) - (3.47) is solved using MATLAB's ODE15s function.

Let δ represent the backlash. The dynamic tooth load F is calculated based on the formulas given by Lin et al. (1988). Here, because lateral motion is also considered, its effect is added to the elastic and the damping force calculation.

Case (i) $R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2 > 0$, which is the case of normal operation:

$$F = k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2) + c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2) \quad (3.48)$$

Case (ii) $R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2 \leq 0$, and $|R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2| \leq \delta$, where the teeth pair will separate:

$$F = 0 \qquad\qquad (3.49)$$

Case (iii) $R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2 < 0$, and $|R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2| > \delta$, where the gear teeth will collide backside:

$$F = k_t(R_{b2}\theta_2 - R_{b1}\theta_1 - y_2 + y_1) + c_t(R_{b2}\dot{\theta}_2 - R_{b1}\dot{\theta}_1 - \dot{y}_2 + \dot{y}_1). \quad (3.50)$$

The dynamic load on a tooth at the contact point is the sum of the elastic force $k_t(R_{b1}\theta_1 - R_{b2}\theta_2 - y_1 + y_2)$ and the damping force $c_t(R_{b1}\dot{\theta}_1 - R_{b2}\dot{\theta}_2 - \dot{y}_1 + \dot{y}_2)$. Because both the torsional and the lateral vibration are considered in this dynamic model, the effect of lateral vibration on relative gear tooth displacements as well as on velocities should be taken into account. In this study, the dynamic load F in case (iii) is considered to be zero for simplicity.

The dynamic load discussed above combines all the stiffness of teeth pairs being meshed. To focus on the cracked tooth of the pinion, the mesh stiffness of the teeth pair containing this cracked tooth will be extracted from the total effective mesh stiffness. The stiffness is a periodic function of rotation angle. During one mesh period, every tooth will experience three stages, shown in Figure 11, from the instant the tooth comes into contact with its mating tooth on the driven gear until the instant it departs from its mating tooth (a-b-c-d). Let's consider the cracked pinion tooth .The first stage (a-b) is a double-pair-contact stage, during which two pairs of teeth are in contact and share the transmission load. One pair of the two is composed by this cracked tooth and its mating tooth. The second stage (b-c) is a single-pair-contact stage, during which only one pair that contains the cracked tooth, is in mesh. The third stage (c-d) is again a double-pair-contact stage, in which the pair that contains the cracked tooth will share the load with the next pair of teeth.
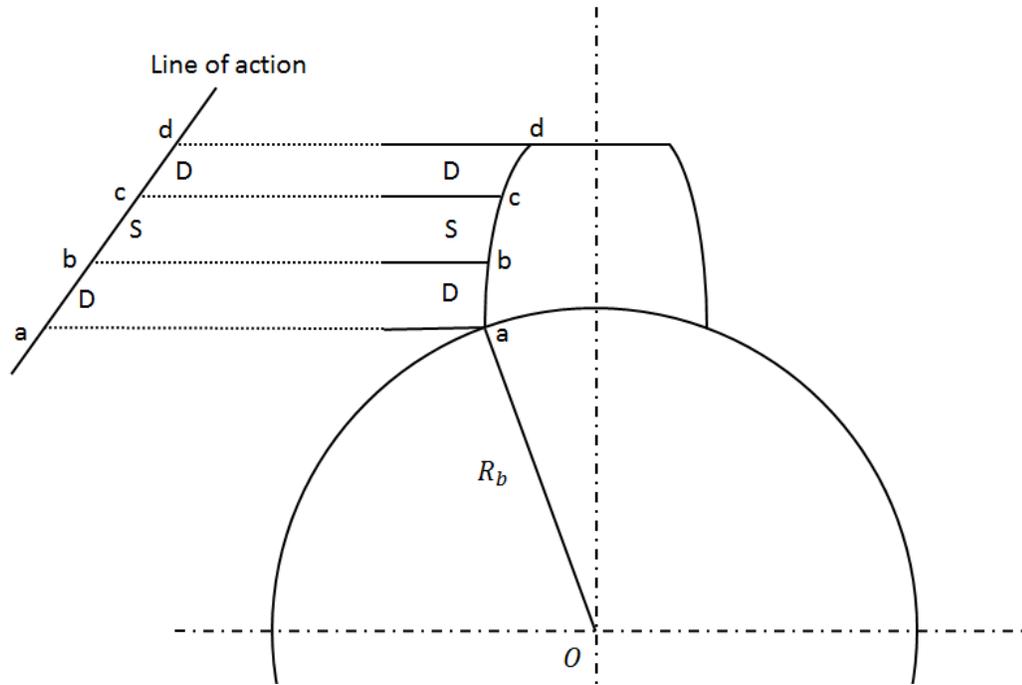
Figure 11. Two contact stages in the gear mesh process.

The mesh stiffness is a periodic function with the period as the meshing period when the teeth are healthy. One meshing period is selected to illustrate the change in the mesh stiffness after the crack appears at the tooth root. The crack is inserted at the first tooth from the reference position. The parameters of the gearbox are as the same to (Wu et al. 2008) shown in Table 2, except for the input and the output torque. With these parameters, the total mesh stiffness and the stiffness of one pair that contains the cracked tooth are plotted in Figure 12 and Figure 13, for the healthy tooth and the cracked tooth with crack size of 3.4 mm, respectively. In these two figures, the blue solid line represents the total mesh stiffness contributed by all the teeth pairs in mesh, and the mauve dashed line represents the mesh stiffness of one gear pair. These two comparative figures clearly show the decrease of the tooth stiffness due to crack appearance.

Table 2. Gearbox dynamics specifications (Wu et al. 2008)

Mass of the pinion	$m_1 = 0.96$ kg
Mass of the gear	$m_2 = 2.88$ kg
Contact ratio	$C_r = 1.6456$
Mass moment of inertia of the motor	$I_m = 0.0021$ kg m^2
Mass moment of inertia of the load	$I_b = 0.0105$ kg m^2
Mass moment of inertia of the pinion	$I_1 = 4.3659 \times 10^{-4}$ kg m^2
Mass moment of inertia of the gear	$I_1 = 8.3602 \times 10^{-3}$ kg m^2
Input shaft frequency	$f = 30$ Hz
Torsional stiffness of the coupling	$k_p = k_g = k_c = 4.4 \times 10^4$ Nm/rad
Damping coefficient of the coupling	$c_p = c_g = c_c = 5.0 \times 10^5$ Nms/rad
Radial stiffness of the bearing	$k_1 = k_2 = k_r = 6.56 \times 10^7$ N/m
Damping coefficient of the bearing	$c_1 = c_2 = c_r = 1.8 \times 10^5$ Ns/m
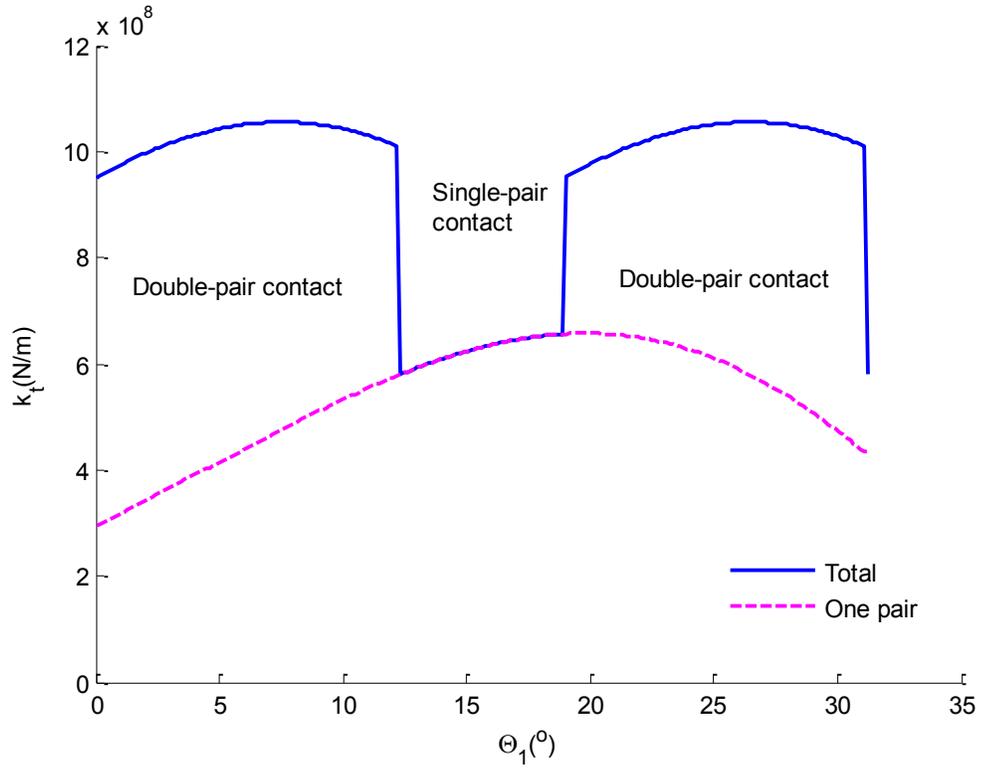
Figure 12. Mesh stiffness of healthy teeth.
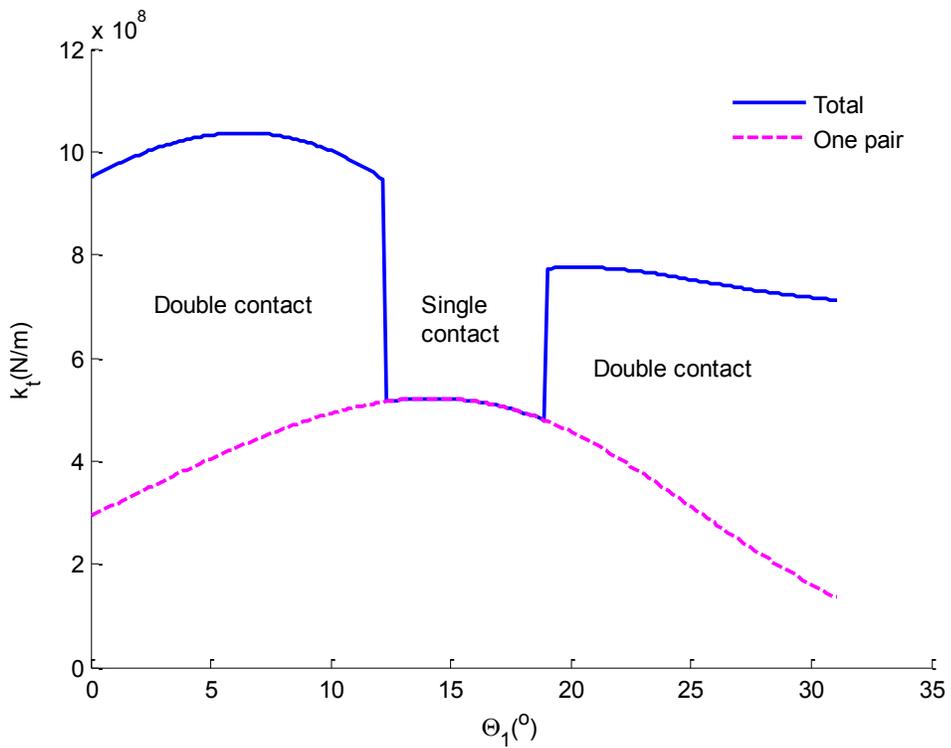
Figure 13. Mesh stiffness when a cracked tooth is involved.

57

The static load and the dynamic load on the cracked tooth are plotted in Figure 14. The static load sharing between the two teeth pairs in mesh in the double-pair-contact stage is determined by the mesh stiffness sharing. As shown in Figure 14, during this one mesh period, the range of dynamic load is larger than that of the static load. A larger range of load will result in faster crack propagation. Hence, to produce conservative prediction of RUL, the maximum dynamic load is selected to apply on the cracked tooth. Because the minimum dynamic load that the tooth experiences is zero, the maximum dynamic load is actually the range of dynamic load during one mesh period.
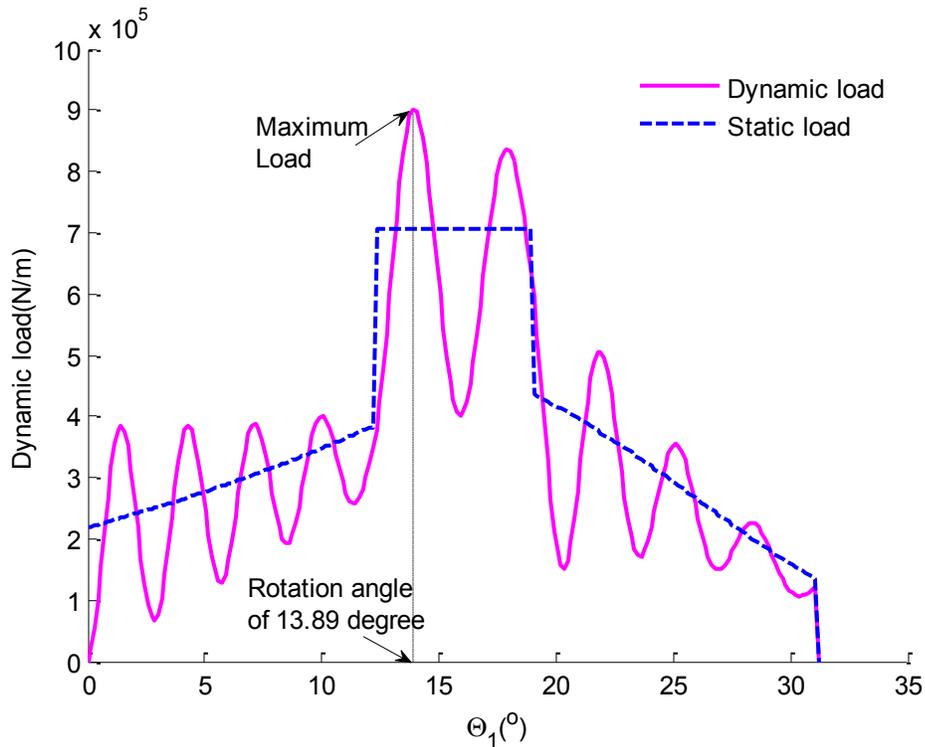
Figure 14. Static and dynamic load on the cracked tooth.

3.6   Conclusions

In this chapter, the dynamic load on cracked gear tooth is calculated using a 6-DOF one-stage gearbox dynamic model. This dynamic load will apply in FE model to drive the

crack propagation in the gear tooth. Incorporation of the crack effect to the dynamic load

and the realistic curved crack propagation path will increase the fidelity of physical

models in the integrated prognostics method proposed in the next chapter.

Chapter 4. An integrated prognostics method for gears

4.1 Overview

Gearbox is a critical system for power transmission in engineering systems. The gears are subject to various failure modes, one of which is tooth fracture. Due to cyclic bending stress, a crack is prone to appear near the tooth root area. As loading cycles continue, the crack will propagate and eventually cause the tooth breakage. The tooth breakage may lead to catastrophic failure of the whole system. The existing prognostics approaches for gears are either physics-based or data-driven. As discussed in Chapter 2, both methods have their drawbacks. For example, physics-based methods usually use deterministic model without considering uncertainty, while data-driven methods may end up with meaningless extrapolation. In this chapter, an integrated prognostics method is proposed to predict the RUL of gears considering the failure mode of tooth fracture.

In the proposed method, the physical models include a damage propagation model, Paris' law, which is a well-known equation that describes the crack evolution with loading cycles; a FE model which is used to calculate SIF in Paris' law; and a dynamics model which is used to calculate dynamic load that the gear tooth experiences. The parameters in Paris' law are treated as random variables to account for the large amount of variation in failure times of the gear population. However, a specific gear has a unique failure time, which means the parameters in Paris' law should be deterministic or at least have small uncertainty to predict the failure time of this specific gear. To achieve this goal, condition monitoring data are used to estimate the crack size at inspection times and

to assist in identifying the model parameters through Bayesian inference for a specific gear. Every time a new estimation of crack size is available, the model parameters will be updated. The update process can reduce the uncertainty in model parameters and make the parameter distribution become narrower. Moreover, the position of the distribution will approach the actual values of the parameters. Therefore, the RUL predicted based on the updated model parameters will get more accurate.

This chapter is organized as follows. Section 4.2 introduces the framework of the proposed integrated prognostics method for gear RUL prediction considering uncertainty. In Section 4.3, physical models used to describe the physics of failure are presented. In Section 4.4, uncertainty sources are defined to capture the uncertainty encountered in the integrated prognostics. The RUL calculation is presented in Section 4.5. Section 4.6 addresses the update process for model parameter using Bayesian inference as well as the way to obtain prior distribution for the model parameter using historical data. In section 4.7, examples are given to show the update process and to verify the method. Section 4.8 concludes the chapter. The materials in this chapter have been published in (Zhao et al. 2013a).

4.2   Framework of the proposed integrated prognostics method

The objective of integrated gear health prognostics is to predict gear RUL at a given time by fusing physical models with condition monitoring data. Uncertainty exists in both the model-based part and the data-driven part of the integrated prognostics method. These uncertainty sources are the key causes of the predicted RUL distribution. Hence, uncertainty quantification in the RUL is critical. Moreover, by incorporating the

observations on damage condition, the uncertainty in the predicted RUL is expected to be reduced, resulting in more accurate prognostics.

The framework of the proposed integrated prognostics method is shown in Figure 15. There are basically two parts separated by a dashed line in the figure: the part of models on the left hand side, and the part of data on the right hand side. Bayesian inference is connecting both parts. On one hand, the model part consists of three physical models: degradation model, FE model and gear dynamic model. The degradation model (Paris' law) is used to predict gear RUL. The purpose of the FE model is to calculate SIF that is the quantity used to solve Paris' law. As discussed in Chapter 3, the gear dynamic model outputs the dynamic load which is applied in FE model to drive the crack propagation. On the other hand, the data part consists of crack estimations obtained by analyzing the condition monitoring data. There is uncertainty associated with the crack estimations. The crack estimations will be incorporated into Bayesian inference to update the distribution of model parameters. The RUL distribution is then predicted by propagating the uncertainty in the model parameters through the degradation model. In this framework, the condition monitoring data are able to assist in reducing the uncertainty in model parameters and in making them more accurate for a specific gear. Hence, the RUL prediction will get more accurate accordingly. Details of each part of the framework will be elaborated in the following subsections.
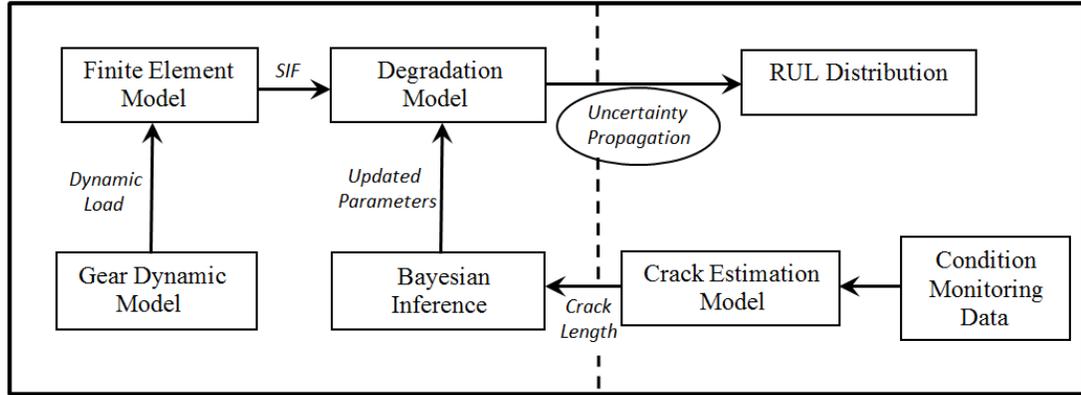
Figure 15. Integrated prognostics framework

4.3   Physical models

Without knowing the physics of failure, data-driven method must "train" or "fit" a large amount of data to identify certain pattern of trend in order to predict the future system behaviour. The effectiveness depends heavily on the availability of a set of dense and well-distributed data. In contrast, the predictive model in physics-based method is established based on the first principle laws of physics. Hence it can be directly used for failure prediction with high accuracy. In this section, three physical models needed in this chapter are introduced.

4.3.1 The degradation model

The degradation model is used for describing the damage accumulation on the component with time. A failure is usually defined by the first time at which the damage indicator crosses a threshold. Most of the existing damage propagation models are built on the empirical Paris' law. Through experimental data regression, parameters in Paris' law are identified to give a fit to the degradation process.

In this chapter, the basic Paris' law is selected as the damage propagation model,

63

which has the form of

$$\frac{\mathrm{d}a}{\mathrm{d}N} = C(\Delta K)^m, \tag{4.1}$$

where $\mathrm{d}a/\mathrm{d}N$ is crack growth rate, a is crack size, N is loading cycles, ΔK is the range of opening SIF, C and m are material dependent parameters and experimentally estimated by fitting fatigue test data. Paris' law asserts a linear relationship between crack growth rate and SIF in a log-log scale.

4.3.2 FE model

FE models are widely used for solid mechanics analysis of components with complicated geometry and loading condition. That said, analytic solutions are difficult to obtain or do not exist at all (Li and Lee 2005; Kacprzynski et al. 2004; Glodez et al. 2002; Orchard and Vachtsevanos 2007; Sankararaman et al. 2001). Especially in terms of fracture problems, discontinuity of geometry can give rise to a singularity of strain near the crack tip in linear elastic fracture theory. Computational fracture mechanics provides an effective way to obtain the approximate solution to fracture problem by using either FE method or boundary element method. Commonly used methods, e.g., direct approach, energy based approach, etc., can calculate important parameters such as SIF using FE outputs. The readers can refer to Section 2.3.1 for more background of computational fracture mechanics.

FE software packages facilitate the widespread applications of FE method in various areas, e.g., civil construction, machine design, system simulation, etc. The software of FRANC2D is designed specifically for the simulation of two-dimensional fracture process, and has been verified and used to do the analysis in many applications.

FRANC2D has an appealing feature to alter the structure body geometry and re-mesh near the crack tip automatically after the crack increments. Opening mode and sliding mode SIFs K_I and K_{II} are readily calculated using the built-in functions. In this thesis, FRANC2D is used to build a two-dimensional FE model for a spur gear tooth with a crack at root and to calculate $K_I(a)$ and $K_{II}(a)$ as functions of crack size. Since opening mode SIF K_I dominates the crack propagation in spur gear tooth, the SIF mentioned in sequel refers to the opening mode SIF.

4.3.3 Gear dynamics model

Most of the studies on the gear crack propagation problem considered constant static load on the meshing teeth. They investigated how the crack propagated under a fixed force on the tooth, which is a stand-alone component. However, the appearance of a crack could reduce the tooth stiffness and will affect the whole dynamic characteristics of the gearbox system. Hence, the dynamic load that the cracked tooth experiences is also affected. The purpose of the gear dynamics model in this thesis is similar to that in (Li and Lee 2005), which is to calculate the dynamic load on a cracked tooth at different crack sizes. With certain crack size, the maximum dynamic load is selected to apply on the cracked tooth to drive the crack extension. The details of this dynamic model can be found in Chapter 3.

4.4   Uncertainty sources

Three main uncertainty sources are considered in the integrated prognostics method: material parameter uncertainty, model uncertainty, and measurement uncertainty.

Manufacturing process variability may result in differences in material at micro-

structural level, such as different grain orientations. Thus, even physically identical components made of the same type of material could demonstrate different fatigue behaviors. Experimental data on fatigue crack propagation (Virkler et al. 1979) showed that, even under carefully controlled conditions, both the number of cycles taken for the crack to reach a given size, and the crack size researched given a number of cycles, displayed a large amount of scatter. The experiment confirms the stochastic nature of crack propagation process. Researchers assigned randomness in the parameters of Paris' law by treating them as random variables (Coppe et al. 2010; An et al. 2012; Sankararaman et al. 2011). Hence, the uncertainty in model parameters is considered to be responsible for the scatter in failure times among identical units made out of the same type of material.

The degradation model adopts the basic form of Paris' law as the crack propagation model without considering other possible parameters which may have impact on crack propagation, such as crack closure retard, fracture toughness, load ratio, etc. Therefore, an error term is introduced to represent the model uncertainty, and denoted by ε. For simplicity, assume ε follows normal distribution in this chapter. Considering this model uncertainty, the modified Paris' law is written as

$$\frac{da}{dN} = C(\Delta K)^m \, \varepsilon. \tag{4.2}$$

In addition, measurement error e is also considered. The real crack size is never known, but it can be estimated by in-situ condition monitoring and diagnostics techniques. There is uncertainty in the crack size estimation, and the error between the real crack size and the estimated one is assumed to be a zero-mean Gaussian white noise

with a standard deviation of σ. The real crack size is denoted as a^{real}, and the estimated one is a^{obs}, so the measurement error is defined as $e = a^{obs} - a^{real}$. Thus,

$$e \sim N(0, \sigma^2), \tag{4.3}$$

or equivalently,

$$a^{obs} \sim N(a^{real}, \sigma^2). \tag{4.4}$$

4.5   RUL prediction

The RUL prediction is performed at every inspection time when the current crack size is estimated. At the inspection time N_t, suppose the observed current crack size is a_t. The crack will propagate according to Paris' law until it reaches the critical size a_c, when the gear is considered failed.

By exchanging the position of differentiation, the modified Paris' law considering model error is written in (4.5).

$$\frac{dN}{da} = \frac{1}{C\big(\Delta K(a)\big)^m \varepsilon} \tag{4.5}$$

Let the crack increment be Δa. The modified Paris' law is discretized by finite difference method:

$$\Delta N_i = N_{i+1} - N_i = \Delta a \left[C\big(\Delta K(a_i)\big)^m \varepsilon \right]^{-1}, \qquad i = t, t+1, \cdots. \tag{4.6}$$

The summation $\sum_{i=t} \Delta N_i$ from the current crack size at inspection time N_t until the critical size a_c is the RUL. The entire failure time could be obtained by $N_t + \sum_{i=t} \Delta N_i$.

Because of the uncertainty in both the degradation model and the model parameters, there is uncertainty in the predicted RUL. Monte Carlo simulation is employed in this

chapter to propagate the uncertainty to the predicted RUL.

4.6   Parameter update through Bayesian inference

The appealing feature of the integrated prognostics method is the ability to incorporate condition monitoring data into physical model adjustment. The condition monitoring data are collected for a specific gear operating in a specific environment under monitoring. Hence, the information that the data bear is specific for this individual gear. In the proposed integrated prognostics method, the condition monitoring data will be used to adjust the physical model parameters to make the RUL prediction of this specific gear more precise and more distinguished from other gears.

Material parameters in Paris' law should have narrow distributions or even deterministic values to describe the crack propagation in a specific gear. For the gear population, though, the distributions of material parameters should be much wider than those for a specific gear. By taking advantage of the condition monitoring data for the specific gear under monitoring, we should be able to reduce the uncertainty in these parameters. It is achieved by updating the parameter distribution through Bayesian inference at inspection times.

Consider for example a simplified case where only the distribution of parameter m is to be updated, while assuming the other material parameter C being constant. The prior distribution for m is denoted by $f_{prior}(m)$, and the likelihood to obtain the current crack estimation is $l(a|m)$. Thus, the Bayesian formula to obtain the posterior distribution of m is given as

$$f_{post}(m|a) = \frac{l(a|m)f_{prior}(m)}{\int l(a|m)f_{prior}(m)\,dm}. \tag{4.7}$$

To calculate the likelihood function in (4.7), the crack size predicted by Paris' law at the inspection time is needed. To solve the Paris' equation, it is discretized using a first-order Euler method. Let the initial crack size be a_0, and the incremental loading cycles be ΔN; then the discretized Paris' law is

$$\begin{cases} a\big((i+1)\Delta N\big) = a(i\Delta N) + (\Delta N)C\big[\Delta K\big(a(i\Delta N)\big)\big]^m \varepsilon \\ a(0) = a_0 \end{cases}, \quad i = 0,1,2\cdots. \tag{4.8}$$

The iteration sequentially proceeds until the current inspection time is reached. The crack size obtained in (4.8) through this discretization is denoted as a^{sim}. In this process, model error is sampled from its assumed known distribution randomly at each iteration step. Because of the measurement error, the crack estimation a^{obs} has a distribution of

$$a^{obs} \sim N\big(a^{sim}, \sigma^2\big) \tag{4.9}$$

Thus, the PDF of the normal distribution in (4.9) will be the likelihood function $l(a|m)$,

$$l(a|m) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\big(a_k^{obs} - a_k^{sim}\big)^2\right]. \tag{4.10}$$

Here, the effect of the model error ε on the a^{sim} mainly relies on its mean because of central limit theorem. Hence, without much loss of accuracy, the likelihood function is considered to be determined only by measurement error.

The last problem to address is how to select the prior distribution of the parameter m. First assume that N historical paths of degradation data are available, denoted in (4.19).

Each path $\mathcal{P}_i$ composes of M inspection time instances $t_i^j, j = 1, \cdots, M$ and the associated crack size $a_i^{j\_obs}$ estimated at such inspection times.

$$\mathcal{P} = \{\mathcal{P}_i|, i = 1, \cdots, N\} \tag{4.19}$$

$$\mathcal{P}_i = (t_i^j, a_i^{j\_obs}), \ j = 1, \cdots, M \tag{4.20}$$

Next, apply Paris' law to generate degradation paths $\mathcal{P}_i'$ with parameter m, obtaining the approximate crack size $a_i^{j\_app}(m)$ at the inspection time $t_i^j, j = 1, \cdots, M$.

$$\mathcal{P}_i' = (t_i^j, a_i^{j\_app}(m)), \ j = 1, \cdots, M \tag{4.21}$$

Define $e_i^j(m) = a_i^{j\_obs} - a_i^{j\_app}(m)$, then the optimal m_{op}^i is found in the sense of least-square, such that

$$\sum_{j=1}^{M} (e_i^j(m_{op}^i))^2 \leq \sum_{j=1}^{M} (e_i^j(m))^2, \ \forall m \tag{4.22}$$

Lastly, prior distribution of m will be determined by fitting $\{m_{op}^i, \ i = 1, \cdots, N\}$ using normal distribution.

$$f_{prior}(m) \sim N\left(\mu_{prior}^m, \left(\sigma_{prior}^m\right)^2\right). \tag{4.23}$$

This prior distribution can be considered as the parameter distribution for the gear population. Once the condition monitoring data are available, the Bayesian inference will be implemented to update this prior distribution to obtain the posterior distribution which becomes specific to an individual unit. The posterior distribution will serve as the prior distribution for the next update conducted at the next inspection time.

4.7 Example

In this section, a numerical example is presented to demonstrate the proposed

integrated prognostics method. Simulated degradation paths are generated by considering various uncertainty sources. The degradation paths are divided into two sets: the training set is used to obtain the prior distribution, and the test set is used to test the prediction performance of the proposed method.

4.7.1 Introduction

The material and geometry properties of the spur gear used in this example are listed in Table 3. The FE model for the gear tooth is shown in Figure 16. Suppose the critical crack size is $a_c = 5.2$ mm, which is 80% of the total circular thickness of the tooth. Beyond this threshold, the crack will propagate very fast, and the tooth breakage is imminent.

The gear dynamic system discussed in Chapter 3 is used to calculate the dynamic load on this cracked tooth. The input torque is selected as 320Nm, and the output load torque is 640Nm. The rotation speed of gearbox is 30Hz. The other specifications of the gearbox are listed in Table 2. The maximum dynamic load appears at the rotation angle of 13.89 degrees, higher than the static load. The results show that, for the entire crack path, the position of the maximum dynamic load will move forward a little bit as the crack size increases, but the movement is less than 1 degree so that the load is considered being applied at a fixed position, which corresponds to the rotation angle of 13.89 degrees.

Table 3. Material properties and main geometry parameters of the spur gear

Young's modulus (Pa)	Poisson's ratio	Module (mm)	Diametral pitch (in^{-1})	Base circle radius (mm)	Outer circle (mm)	Pressure angle (degree)	Teeth No.
2.07e11	0.3	3.2	8	28.34	33.3	20	19
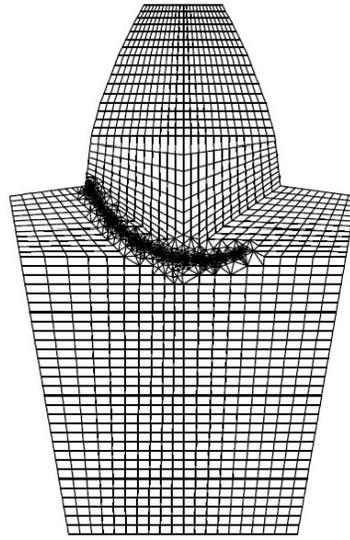
Figure 16. 2D FE model for spur gear tooth.

Referring to Figure 9 in Section 3.3.3, the procedure to obtain the SIF history as the crack grows from the initial size to the critical size under varying dynamic load is summarized next.

1. Select the initial crack tip T_j, $j = 0$, such that the angle of $\beta_0 = 45$ degrees, and the initial crack length $a_0 = 0.1$mm.

2. Calculate u_j, which is the distance between tooth root S and T_jG_j. The total mesh stiffness k_t is then obtained by the formulas proposed in Section 3.2.2 depending on where the crack tip is, and how many degrees the rotation angle is.

3. Solve gear dynamic equations in MATLAB by plugging k_t obtained in step 2, and the dynamic load is computed using (3.48) – (3.50).

4. Apply the maximum dynamic load at the contact point on the FE model of the cracked pinion tooth in FRANC2D, which corresponds to the rotation angle of 13.89 degrees. The modes I and II SIFs, as well as the crack propagation angle, are calculated.

5. Propagate the crack in the direction obtained in step 4 with an increment of Δa=0.1mm.

6. $j = j + 1$, return to step 2 until the crack size reaches the critical value.

Following the procedure above, the histories of the two modes of SIFs are obtained, and shown in Figure 17. The mode I SIF K_I is dominant just as reported in other published papers. So in Paris' law, only ΔK_I is used to calculate the crack propagation rate. The third order polynomial is used to fit the discrete values of K_I obtained by FRANC2D, thus $K_I(a)$ has its continuous form, and then the value of K_I at any given crack size is available. Additionally, because the minimum load on the cracked tooth is zero, the range of the SIF during one loading cycle is the SIF obtained under the maximum dynamic load. Figure 18 plots the maximum dynamic load at different crack sizes. Applying the maximum dynamic load to the cracked tooth produces a larger SIF compared to that under the static load; and under this circumstance, the crack bears a faster propagation rate, which will lead to a relatively shorter RUL.
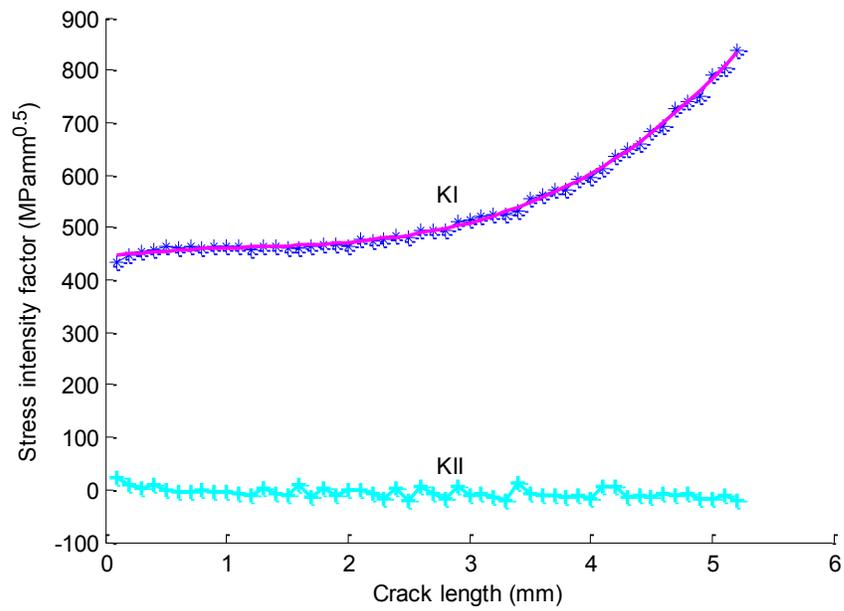
Figure 17. Stress intensity factor as a function of crack size.
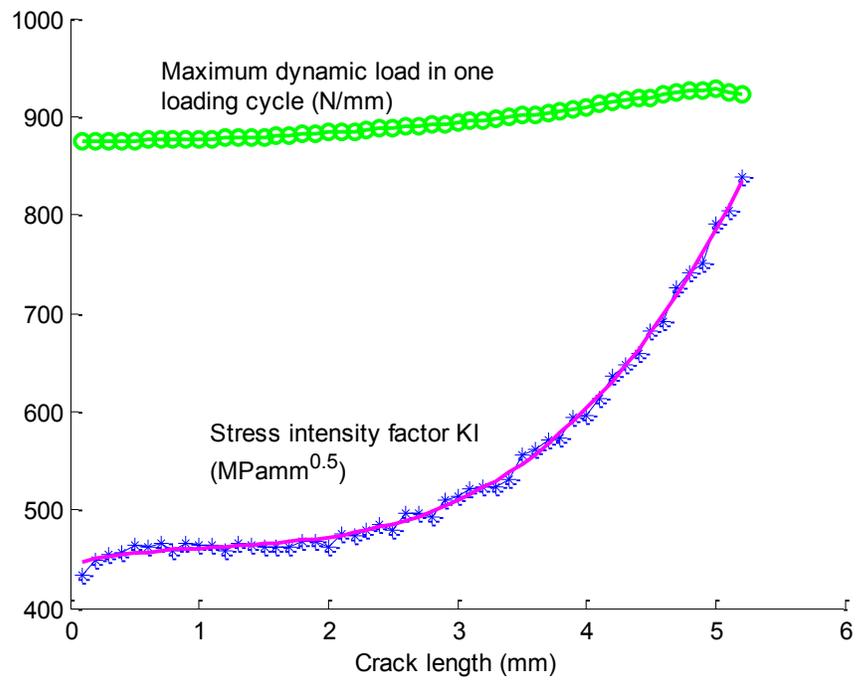
Figure 18. Mode I stress intensity factor and maximum dynamic load.

A set of crack degradation paths $\mathcal{P}$ is generated using Paris' law by adding measurement errors. All these paths, as well as the values of parameter m^i that are used to generate these paths, termed here as *real m^i*, are recorded. The paths in $\mathcal{P}$ are divided into two sets: training set H, and test set R. The training set is used to obtain the prior distribution for parameter m, and the test set is used to validate the proposed method.

To generate the degradation paths, we assume the following values and distributions for the parameters involved: $C = 9.12e - 11$, $\sigma = 0.2$, $m \sim N(1.4354, 0.2^2)$, $\varepsilon \sim N(2.5, 0.5^2)$. Note that here the uncertainty regarding m is related to the gear population, not of the specific gear being monitored.

In this example, 10 degradation paths are generated using Paris' law up, as shown in Figure 19. Select $H = \{1, 2, 3, 5, 7, 8, 10\}$ and $R = \{4, 6, 9\}$. The three test paths #4, #6, and #9 are bolded in Figure 19. Then, for each path $i \in H$, the optimal $m_{op}^i, i = 1, 2, 3, 5, 7, 8, 10$ satisfying (4.22) can be found using the least-square criterion, termed here as *trained m^i*. Then a normal distribution is used to fit them to obtain a prior distribution of m.
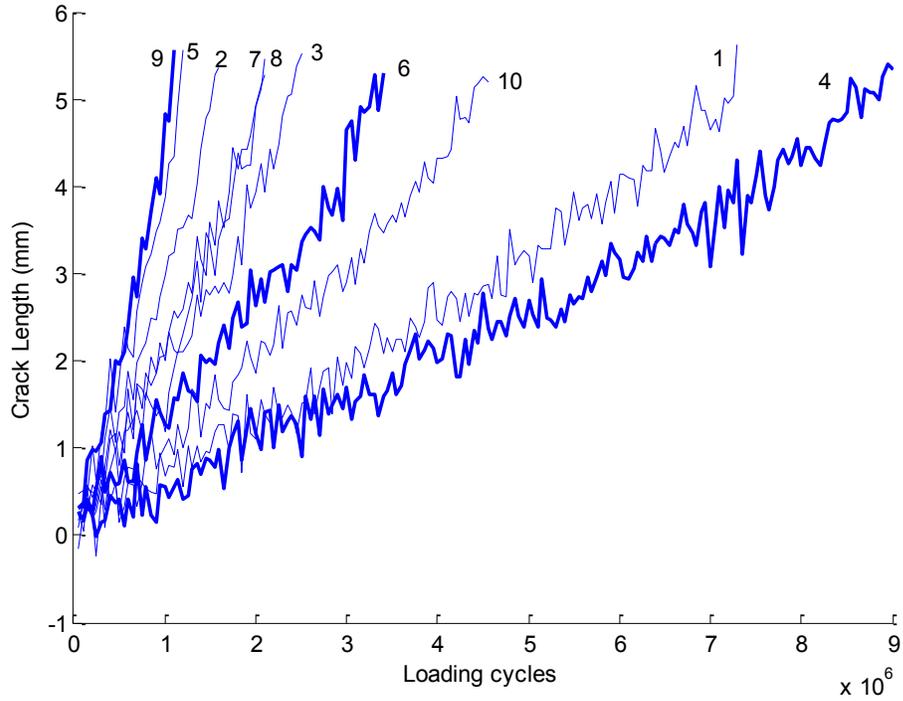
Figure 19. Ten degradation paths generated using prescribed parameters.

4.7.2 Results on update process

The ten real values of m and the corresponding seven trained values are tabulated in Table 4, based on which the prior distribution of m is calculated as

$$f_{prior}(m) \sim N(1.454, 0.1004^2).$$

To validate the proposed integrated prognostics method, we take paths #4, #6, and #9 for test. In path #4, 9×10^6 cycles are consumed for the crack to reach the critical size. The updating history for path #4 is shown in Table 5. In path #6, the actual failure time is 3.4×10^6 cycles, and in path #9, it is 1.1×10^6 cycles. The update processes for distributions of parameter m in path #6, and path #9 are shown in Table 6, and Table 7, respectively.

Table 4. The real values and the trained values of m

Path #	Real m	Trained m
1	1.2836	1.284
2	1.5302	1.5328
3	1.4569	1.4589
4	1.2495	-
5	1.5724	1.5729
6	1.407	-
7	1.4823	1.4807
8	1.4844	1.4904
9	1.5897	-
10	1.3585	1.3583

Table 5. Testing results for path #4 (real m=1.2495)

Inspection cycle	Crack length (mm)	Mean of m	Std of m
0	0.1	1.454	0.1004
2×10^6	1.1656	1.2746	0.027
4×10^6	1.9857	1.2514	0.0194
6×10^6	3.1521	1.2556	0.016
8×10^6	4.2336	1.2445	0.0121

Table 6. Testing results for path #6 (real m=1.407)

Inspection cycle	Crack length (mm)	Mean of m	Std of m
0	0.1	1.454	0.1004
0.7×10^6	0.9349	1.3956	0.037
1.4×10^6	2.0607	1.4194	0.0253
2.1×10^6	2.68	1.3931	0.0186
2.8×10^6	3.7607	1.3967	0.0156

Table 7. Testing results for path #9 (real m=1.5897)

Inspection cycle	Crack length (mm)	Mean of m	Std of m
0	0.1	1.454	0.1004
0.25×10^6	0.9629	1.5409	0.0458
0.5×10^6	1.9648	1.5675	0.0298
0.75×10^6	3.3989	1.6053	0.0201
1×10^6	4.8369	1.5849	0.0111

The results show that the Bayesian updates adjusted the mean of m from the initial value 1.454 to its real values gradually, as more condition monitoring data become available. Moreover, the standard deviation of m is reduced, which means that the uncertainty in m is reduced. To demonstrate, Figure 20 shows the updated distributions of m for path #4.

The update processes of failure time distributions for path #4, #6, and #9 are shown in Figure 21, Figure 23 and Figure 24 respectively, from which we can see, the predicted failure time distribution also becomes narrower, and its mean is approaching the real failure time as update continues. The updated RUL at each inspection time for path #4 is computed as shown in Figure 22, and the vertical lines represent the real RUL at the inspection times.
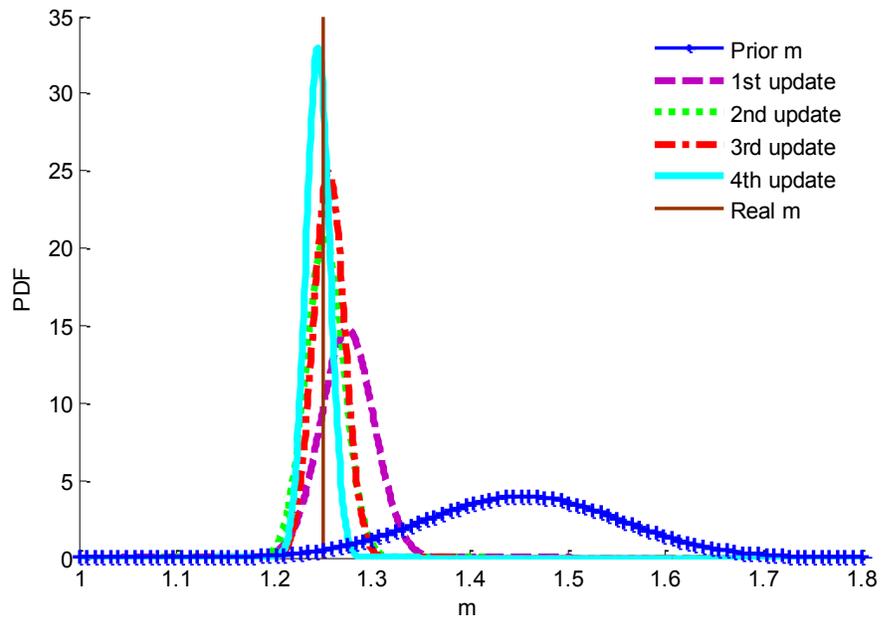
Figure 20. Updated distributions of m for path #4.
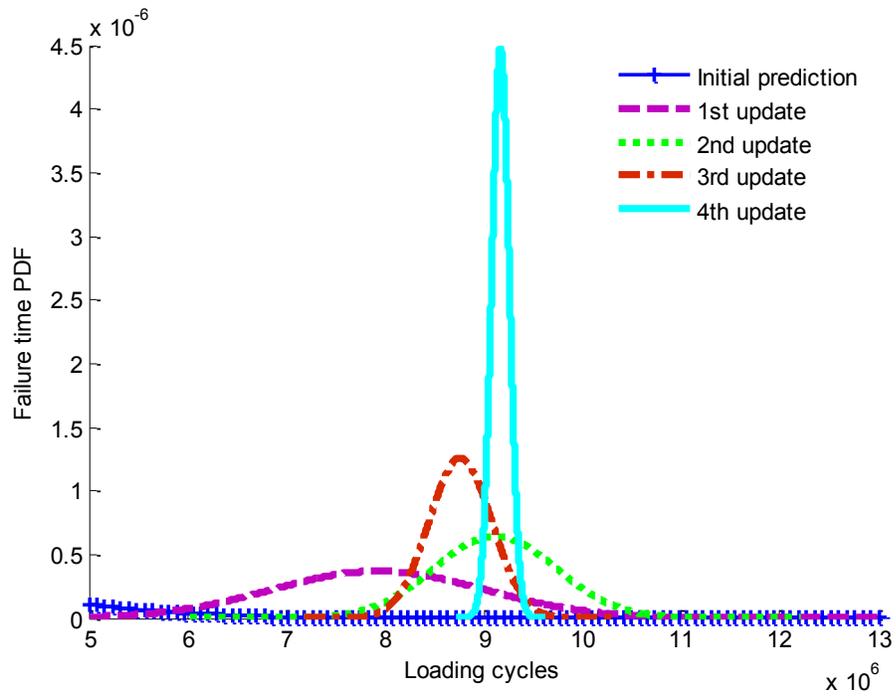
Figure 21. Updated failure time distributions for path #4.
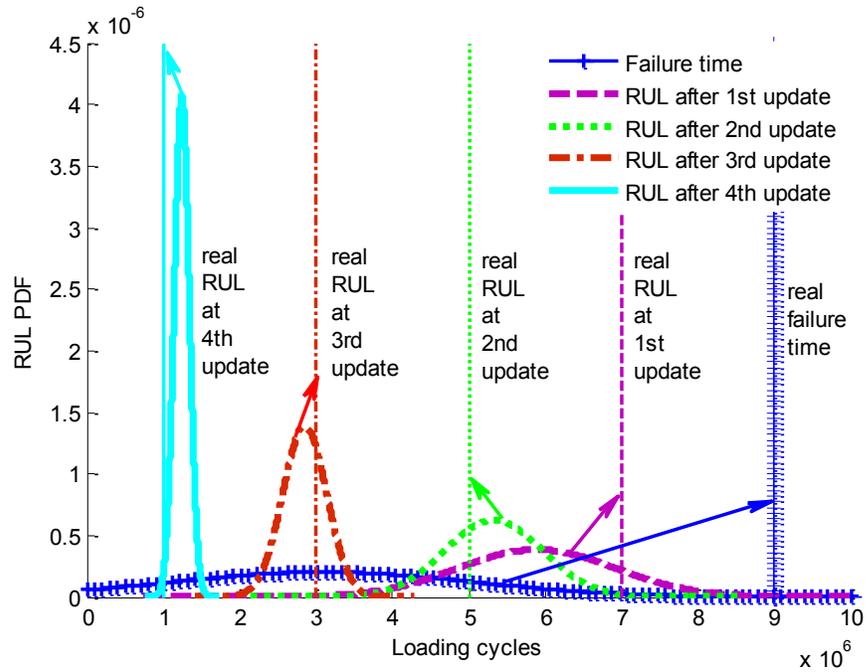
Figure 22. Updated RUL for path #4.
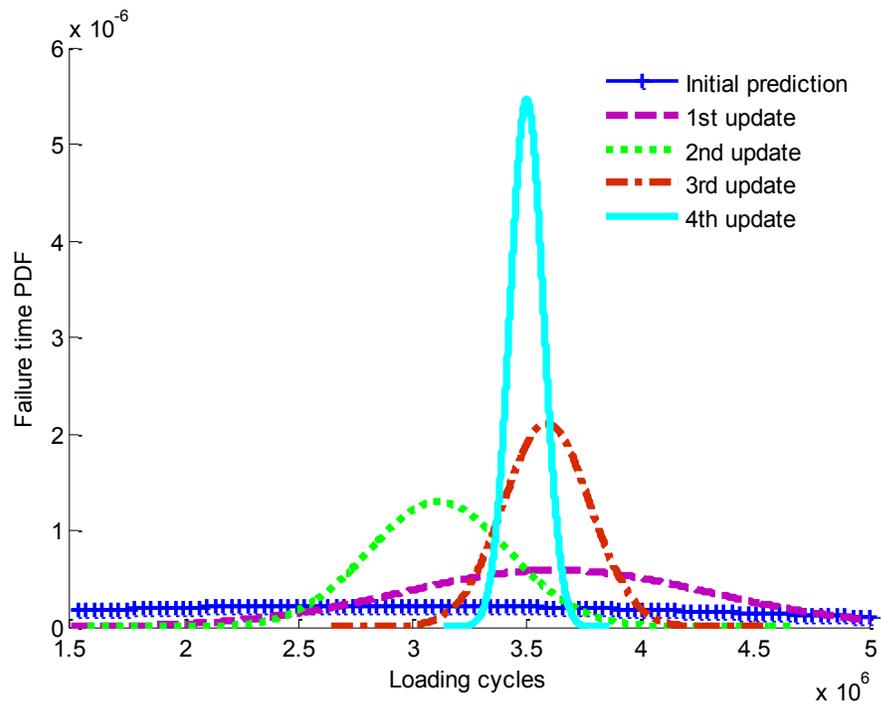
Figure 23. Updated failure time distributions for path #6.
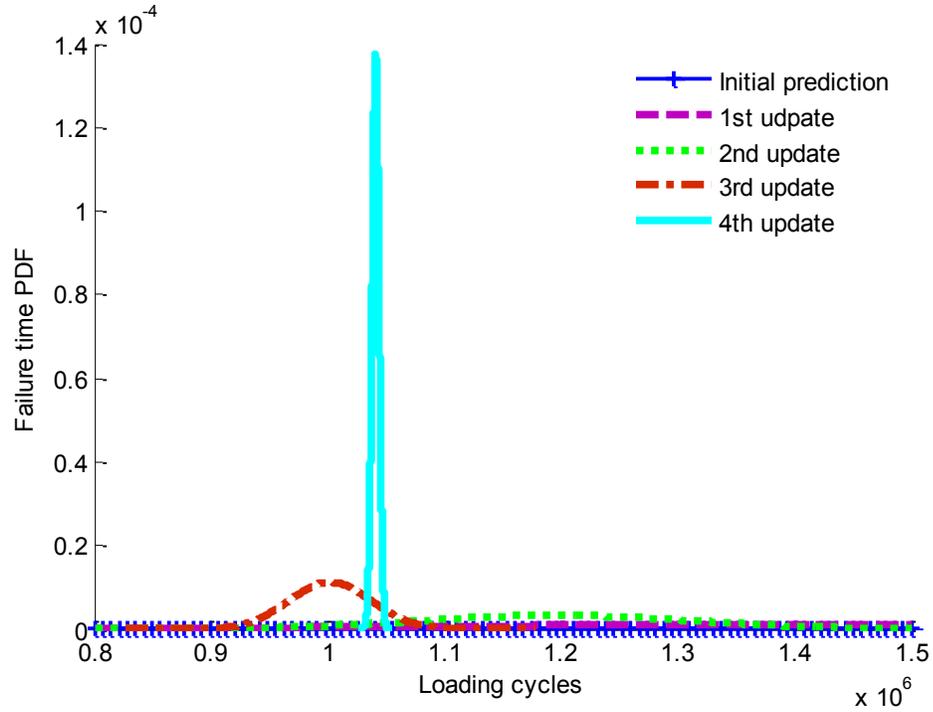
Figure 24. Updated failure time distributions for path #9.

4.8   Conclusions

Accurate health prognosis is critical for ensuring equipment reliability and reducing the overall life-cycle costs, by taking full advantage of the useful life of the equipment. In this chapter, an integrated prognostics method is developed for gear RUL prediction, which utilizes both gear physical models, and real-time condition monitoring data. The physical models specific to gear include a FE model for gear stress analysis, a gear dynamics model for dynamic load calculation, and a damage propagation model described using Paris' law. Material uncertainty and model uncertainty are considered to account for differences of failure times among different units. A Bayesian method is used to fuse the collected condition monitoring data in updating the distribution of the model parameter for the specific unit being monitored, and to achieve the updated RUL prediction.

An example based on simulated degradation paths is used to demonstrate the effectiveness of the proposed method. The results demonstrate that the proposed integrated prognostics method can effectively adjust the model parameter based on the observed degradation data, and thus lead to more accurate RUL predictions; and the prediction uncertainty can be reduced.

The proposed method can only be used after crack initiation. Diagnostics methods are needed to detect the crack initiation, and after that, the prognostics methods, such as the one we propose in this chapter, can be used to predict the failure time.

Chapter 5. A stochastic collocation approach based on PCE for efficient integrated gear health prognosis

5.1 Overview

In prognostics, uncertainty quantification in RUL prediction is a major task. Monte Carlo simulation is a commonly used technique, which generates samples on the model output through the model evaluation. Its slow convergence rate makes Monte Carlo simulation infeasible for high dimensional uncertainty quantification when the model evaluation is computationally intensive. In this chapter, based on the integrated prognostics method proposed in Chapter 4, a stochastic collocation approach based on PCE is developed for efficient uncertainty quantification in the integrated gear health prognosis. The proposed method is able to improve the computational efficiency significantly.

Figure 25 sketches one update process in the integrated prognostics method between two consecutive inspection times. This figure shows that the Bayesian inference is the pivot to combine physics models (FE model, gear dynamic model and degradation model) and condition monitoring data. The output of Bayesian inference is the updated distribution of input parameters in degradation model. And PCE is used in the calculation of the likelihood function and the failure time distribution.
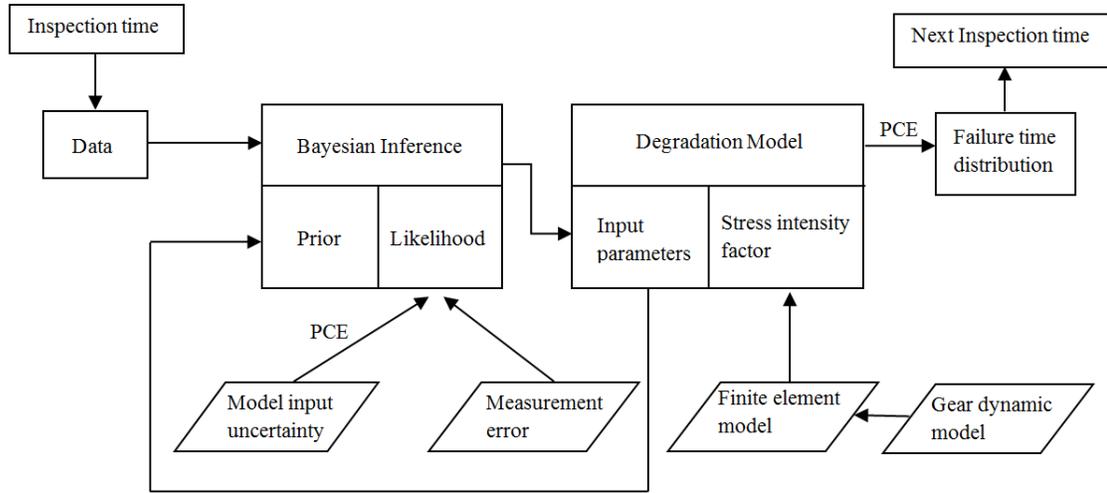
Figure 25. Update process between two consecutive inspection times.

This chapter is organized as follows. Section 5.2 presents fundamentals of PCE and stochastic collocation method. In Section 5.3, an efficient method based on PCE is proposed to improve the computational efficiency of uncertainty quantification in the integrated prognostics method. The uncertain inputs in the degradation model are further classified into updating-uncertainty and non-updating-uncertainty. Model uncertainty and measurement uncertainty are also considered as in Chapter 4 except that the model uncertainty is modeled by a lognormal distribution rather than a normal distribution. Examples are given in Section 5.4, illustrating the effectiveness and the efficiency of the proposed method. Two comparative studies are given to demonstrate the performance of PCE in terms of accuracy and efficiency. Section 5.5 concludes this chapter. The materials in this chapter have been published in (Zhao et al. 2013b).

5.2   Fundamentals of PCE and stochastic collocation method

In this section, fundamentals of PCE and stochastic collocation method on sparse grid are briefly presented. PCE can be considered as an efficient approximation to stochastic

processes and random variables. To introduce this technique and to further investigate the approximation properties, a probability space, $(\Omega, \mathrm{F}, \mathrm{P})$, needs to be properly defined, where Ω is the event space equipped with σ-field $\mathcal{F}$ and probability measure $\mathcal{P}$.

5.2.1 Background

Consider a computational model $H: \mathbf{Z} = H(\mathbf{Y})$, mapping input vector $\mathbf{Y} \in \mathbb{R}^d$ into output, $Z \in \mathbb{R}$, which is the quantity of interest. Here, for the sake of simplicity, Z is considered as a scalar. When it is a vector, the derivation holds component wise. In the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, $\mathbf{Y}$ and Z become random variables, which are functions of random event $\omega \in \Omega$. The purpose of uncertainty quantification is to study the effects of uncertainty in $\mathbf{Y}(\omega)$ on the statistical property of $Z(\omega)$.

PCE is essentially a spectral method in probabilistic context. It relies on the fact that a random response of a computational model can be approximated by the polynomials coordinated in a suitable finite orthogonal basis. Let $\boldsymbol{i} = (i_1, \cdots, i_d)$ be a multi-index with $|\boldsymbol{i}| = i_1 + \cdots + i_d$. By using PCE, the model output can be expressed as

$$Z = H(\mathbf{Y}) \approx H_P(\mathbf{Y}) = \sum_{|\boldsymbol{i}| \leq N} \hat{f}_{\boldsymbol{i}} \, \Phi_{\boldsymbol{i}}(\mathbf{Y}), \qquad (5.1)$$

where $\{\Phi_{\boldsymbol{i}}(\mathbf{Y})\}_{|\boldsymbol{i}| \leq N}$ are the basis orthogonal polynomial functions in d-variate Nth-degree polynomial space $\mathbb{P}_N^d$. H_P can be seen as a orthogonal projection $P_N: L^2 \rightarrow \mathbb{P}_N^d$. The series converges in the sense of L^2- norm given that both $\mathbf{Y}$ and Z have finite variances:

$$\|H - P_N H\|_{L^2} \rightarrow 0, \quad N \rightarrow \infty. \qquad (5.2)$$

The selection of basis polynomial function depends on the type of the distribution of input variables. There exists a correspondence between the basis function type and the

distribution type of the inputs. For example, if the input variable follows Gaussian distribution, Hermite polynomials are selected as basis. Basis functions of multivariable can be generated by the products of univariate basis functions when the components in the input random vectors are independent. For the general cases when the dependency exists among random variables, Soize and Ghanem (2004) conducted a theoretical study to clarify the mathematical structure of the associated functional space. But it may be difficult to find the orthogonal basis because of the unavailability of the joint PDF of random variables. However, for the dependence structure in multivariate Gaussian distribution, the correlated Gaussian random variables can be transformed into uncorrelated standard Gaussian random variables using Cholesky decomposition for covariance matrix. This is also the approach adopted in this chapter to tackle the uncertainty in the Gaussian distributed material parameters.

The main computational task in PCE is to calculate the coefficients $\hat{f}_i$ in $H_P(Y)$. As discussed in Section 2.5.2, stochastic Galerkin method and stochastic collocation method are the two ways to obtain these coefficients. To save the efforts for complex model reformulation required by Galerkin method, the stochastic collocation method is introduced here briefly and adopted in this thesis.

5.2.2 General PCE formulation

Let Y be a d-variate continuous random vector in the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, having independent and identically distributed (i.i.d.) components: $Y = (Y_1, \cdots, Y_d)$. The joint PDF of Y with respect to the support Γ_Y is denoted as $p_Y(\mathbf{y}) = \prod_{i=1}^{d} p_{Y_i}(y_i)$, where $p_{Y_i}(y_i)$ is the marginal PDF of Y_i.

Starting with one-dimensional case, let $\mathbb{P}_N(Y_i)$ be univariate polynomial space with the degree up to N, and $\{\phi_k(Y_i), k = 0, \cdots, N\} \subset \mathbb{P}_N(Y_i)$. We call $\{\phi_k(Y_i), k = 0, \cdots, N\}$ univariate PCE basis functions in $\mathbb{P}_N(Y_i)$ if they satisfy the following orthogonality condition

$$\mathbb{E}[\phi_m(Y_i)\phi_n(Y_i)] = \int \phi_m(y)\phi_n(y)p_{Y_i}(y)dy = \gamma_m\delta_{mn} , \quad 0 \le m, n \le N \quad (5.3)$$

where

$$\gamma_m = \mathbb{E}\big[{\phi_m}^2(Y_i)\big] = \int {\phi_m}^2(y)p_{Y_i}(y)dy \quad (5.4)$$

is a normalized factors and δ_{mn} is Kronecker delta function. The type of orthogonal polynomial is determined by the distribution type of random variable such that the orthogonality will hold with respect to the associated weight function $p_{Y_i}(y)$.

The basis functions in multiple variable polynomial space $\mathbb{P}_N^d$ are the products of those of univariate:

$$\Phi_i(Y) = \phi_{i_1}(Y_1) \cdots \phi_{i_d}(Y_d), \quad 0 \le |i| \le N \quad (5.5)$$

So that $\mathbb{P}_N^d$ can be defined as the expansion of $\{\Phi_i(Y), \ 0 \le |i| \le N\}$,

$$\mathbb{P}_N^d \equiv \left\{ p : \Gamma_Y \to \mathbb{R} \mid p(Y) = \sum_{|i| \le N} c_i\Phi_i(Y) \right\} \quad (5.6)$$

It follows directly from (5.3) and (5.5) that,

$$\mathbb{E}\big[\Phi_i(Y)\Phi_j(Y)\big] = \int_{\Gamma_Y} \Phi_i(y)\Phi_j(y)p_Y(y)dy = \gamma_i\delta_{ij}, \quad 0 \le m, n \le N \quad (5.7)$$

where γ_i is a multiplication of one-dimensional normalized factors and δ_{ij} is d-variate Kronecker delta function.

To address the approximation property, define the weighted L^2 space:

$$L^2(\Gamma_Y) \equiv \left\{ f : \Gamma_Y \to \mathbb{R} \mid \mathbb{E}[f^2(Y)] = \int_{\Gamma_Y} f^2(\mathbf{y}) p_Y(\mathbf{y}) d\mathbf{y} < \infty \right\} \quad (5.8)$$

with the inner product

$$< w, v >_{L^2} = \int_{\Gamma_Y} w(\mathbf{y}) v(\mathbf{y}) p_Y(\mathbf{y}) d\mathbf{y} \quad (5.9)$$

and the L^2- norm

$$\|w\|_{L^2} = \left(\int_{\Gamma_Y} w^2(\mathbf{y}) p_Y(\mathbf{y}) d\mathbf{y} \right)^{\frac{1}{2}} \quad (5.10)$$

Based on the PCE basis functions in polynomial space $\mathbb{P}_N^d$ and the weighted L^2 space defined above, from classical approximation theory, the following conclusion holds:

For any $f \in L^2(\Gamma_Y)$, define Nth-degree PCE orthogonal projection as

$$P_N f = \sum_{|i| \leq N} \hat{f}_i \Phi_i(Y) \quad (5.11)$$

$$\hat{f}_i = \frac{1}{\gamma_i} \int_{\Gamma_Y} f(\mathbf{y}) \Phi_i(\mathbf{y}) p_Y(\mathbf{y}) d\mathbf{y}, \quad \gamma_i = \gamma_{i_1} \cdots \gamma_{i_d} \quad (5.12)$$

Then

$$\|f - P_N f\|_{L^2} \to 0, \quad N \to \infty. \quad (5.13)$$

That is to say, the function of random variables can be approximated in a form of orthogonal polynomial expansion. The convergence rate depends on the regularity of the function f and the type of orthogonal polynomials $\Phi_i(Y)$. This kind of convergence is referred to as spectral convergence, and this expansion error due to the truncated degree of polynomial space is referred to as PCE projection error. More details of this section could be found in Xiu (2010).

5.2.3 Stochastic collocation method

Stochastic collocation method aims at estimating the coefficients in the polynomial expansion by utilizing the orthogonality of the basis functions. By multiplying the basis function to the polynomial expansion and taking the expectation in L^2-norm, one gets the theoretical expression of each coefficient:

$$\hat{f}_i = \int_{\Gamma_Y} f(\mathbf{y})\Phi_i(\mathbf{y})\, p_Y(\mathbf{y})d\mathbf{y}. \tag{5.14}$$

However, the integration in (5.14) is intractable when high dimensional random space is involved. In practice, the numerical integration rules provide approximate solutions by a weighted sum using the pre-selected points $\boldsymbol{p}^{(j)} \in \mathbb{R}^d$ and the associated weights $\alpha^{(j)} \in \mathbb{R}, j = 1, \cdots, Q$, such that

$$\tilde{f}_i = \sum_{j=1}^{Q} f(\boldsymbol{p}^{(j)})\alpha^{(j)} \to \hat{f}_i, \qquad Q \to \infty. \tag{5.15}$$

With $\tilde{f}_i$ available, define the operator $I_N: L^2 \to \mathbb{P}_N^d$, such that

$$I_N f = \sum_{|\boldsymbol{i}| \leq N} \tilde{f}_{\boldsymbol{i}} \, \Phi_{\boldsymbol{i}}(\boldsymbol{Y}). \tag{5.16}$$

Various schemes of summation in (5.16) may be used which differ in the selections of the integration rules. For one-dimensional case, this problem is readily solved using the available numerical integration rules, such as Gaussian quadrature rule and Clenshaw-Curtis rule. The natural generalization to multi-dimensional case is to use tensor product. However, the number of points in the tensor product will grow rapidly as the dimension increases, which is known as "curse of dimensionality". When high dimensional uncertainty is considered, the computational burden will be very heavy using tensor product collocation.

Smolyak sparse grid can mitigate the "curse of dimensionality" by constructing the nodal set which consists of an algebraic sum of low-order tensor products based on one-dimensional quadrature rule. Such construction will result in a great reduction of nodes for integration compared to tensor products.

Define the integration operator $\mathcal{L}_i[f]$ in the i-th dimension

$$\mathcal{L}_i[f] \equiv \sum_{j=1}^{Q} f\!\left(p_i^j\right) \alpha_i^j \tag{5.17}$$

then the Smolyak algorithm gives the following multi-dimensional operator

$$\mathcal{L}[f] = \sum_{K-d \leq |\boldsymbol{i}| \leq K-1} (-1)^{K-d-|\boldsymbol{i}|} \cdot \binom{d-1}{d-K+|\boldsymbol{i}|} \cdot \left(\mathcal{L}_{i_1} \otimes \cdots \otimes \mathcal{L}_{i_d}\right) \tag{5.18}$$

The difference between $P_N f$ and $I_N f$ is caused by the approximation of the coefficients $\tilde{f}_{\boldsymbol{i}} \to \hat{f}_{\boldsymbol{i}}$, and the consequent error $\|P_N f - I_N f\|_{L^2}$ is called aliasing error.

When the numerical integration rule converges, the aliasing error tends to zero, which means $I_N f$ becomes a good approximation to f.

5.3 The proposed stochastic collocation method for uncertainty quantification in integrated gear prognosis

This section will elaborate the proposed stochastic collocation method for uncertainty quantification in the integrated prognostics method. Firstly in Section 5.3.1, the uncertainty sources as well as their roles in the integrated prognostics are defined. By classifying the model parameters into updating-uncertainty and non-updating-uncertainty, a likelihood function is defined to combine the contribution from both measurement error and non-updating-uncertainty. Then Section 5.3.2 and 5.3.3 present the stochastic collocation method based on PCE for uncertainty quantification in RUL and Bayesian inference, respectively.

5.3.1 Modeling of uncertainty sources

In this chapter, it is assumed that all the uncertainty is categorized into three main sources: model inputs uncertainty, measurement uncertainty and model uncertainty. The model inputs uncertainty is further classified into updating-uncertainty and non-updating-uncertainty.

Paris' law and its evolutions are widely used physical models to describe the crack propagation process. When the associated inputs in the model are treated as random variables, Paris' law becomes a stochastic equation. Denote the set of all these random inputs appearing in Paris' model as $\boldsymbol{\Theta}$, and divide them into two subsets: updating-uncertainty $\mathbf{U}$, and non-updating-uncertainty $\boldsymbol{R}$. The distributions of uncertain inputs

belonging to **U** are unknown, and only *a* priori may be assumed. The updating-uncertainty will be updated through Bayesian inference using the condition monitoring data. With the uncertainty reduction, the distribution of updating-uncertainty will be narrowed and may converge to the real value for a specific unit. While, those uncertain inputs belonging to **R** will take well-known distributions, and they are treated as contributors to the likelihood function in the Bayesian inference. So this type of uncertainty is named as likelihood-uncertainty. Both the likelihood-uncertainty and the updating-uncertainty contribute to RUL distribution. In addition to the likelihood-uncertainty, another contributor to the likelihood function is the measurement error, which has been defined in the integrated prognostics method presented in Chapter 4.

The existence of a number of different physical models to describe the crack propagation process implies that there is no perfect model for all the circumstances. The models are usually selected by leveraging the accuracy and the complexity. The authors in (Ortiz and Kiremidjian 1988) assigned the reason of randomness nature of the crack growth rate to the random unpredictable resistance in material's microstructure. Yang et al. (1985) proposed a model by multiplying a stochastic term ε to the deterministic crack growth model after investigating the crack propagation in the fastener holes of aircrafts under spectrum loading, and experiments were also conducted to validate this model. The random term ε was assumed to be a lognormally distributed random variable and "accounts for the crack growth rate variability, such as the variabilities due to material cracking resistance, crack geometry, crack modeling, spectrum loading, etc." (Yang et al. 1983). This model has the simplest stochastic form and produces conservative and reasonable results. Many related work adopted this multiplicative form of model error for

crack propagation study in other applications (Sankararaman et al. 2011; Ortiz and Kiremidjian 1988; Yang et al. 1983; Yang et al. 1987; Willhauck et al. 2008). The statistical property of ε could be acquired by a least-square fitting of Paris' law in a log-log scale using the information of crack sizes and associated cycles obtained in an experiment of fatigue crack propagation, as reported in (Yang et al. 1987; Virkler et al. 1978). The experimental data showed the residual ζ in the regression model has a Gaussian distribution, and thus, the distribution of ε could be obtained by noticing the relationship of $\varepsilon = \exp(\zeta)$. Based on the discussion above, the distribution of ε is assumed to be log-normal in this chapter.

In a word, the uncertainty considered in this chapter can be categorized as follows:

$$\begin{cases} \text{measurement error } e \\ \text{model error } \varepsilon \\ \text{Random model inputs} \begin{cases} \text{updating} - \text{uncertainty } \boldsymbol{U} \\ \text{non} - \text{updating} - \text{uncertainty } \boldsymbol{R}. \end{cases} \end{cases}$$

As for the problem that among the random model inputs which belong to which subset, is a matter to be decided according to the specific application.

5.3.2 Stochastic collocation method based on PCE for uncertainty quantification in RUL distribution

The crack propagation process is stochastic in nature, which is affected by a variety of uncertainty, such as material, lubrication, speed, loading and damage initial conditions. If all the uncertain factors are considered as the random inputs in the crack propagation model, the computation of RUL distribution will become a problem of uncertainty propagation.

We parameterize the probability space by finding the finite set of uncertain factors in the gear health degradation. The random parameters in this set are denoted by $\xi = (\xi_1, \cdots \xi_d)$, and are assumed to be i.i.d. so that the density function can be written as $\rho(\xi) = \prod_{i=1}^{d} \rho_i(\xi_i)$. Given the distribution of ξ, the corresponding orthogonal basis polynomials $\{\Phi_i(\xi) \in \mathbb{P}_N^d, \ 0 \leq |i| \leq N\}$ are selected. Denote the failure time as T, which is the quantity of interest. T will be a function of these random inputs ξ through Paris' law,

$$T = T(\xi). \tag{5.19}$$

Firstly, nodal set is selected using Smolyak algorithm, which is $\{\xi^{(j)}, \ j = 1, \cdots, Q\}$. Based on the proper integration rule, the associated weights $\alpha^{(j)}, j = 1, \cdots, Q$, are also available. Secondly, the failure times at these nodes are obtained by propagating the crack through Paris' law to the critical size in a deterministic way. Denote them as $\tilde{T}^j = T(\xi^{(j)}), j = 1, \cdots, Q$. After that, we use the truncated N-th degree polynomial expansion, $T_N = P_N T$, to approximate T,

$$T_N = \sum_{l=1}^{M} \widehat{\omega}_l \, \Phi_l(\xi), \tag{5.20}$$

$$T_N \to T \quad \text{as} \quad M \to \infty. \tag{5.21}$$

From (5.12), we have

$$\widehat{\omega}_l = \int_\Gamma T(\xi) \, \Phi_l(\xi) \rho(\xi) \, d\xi. \tag{5.22}$$

Based on (5.13), numerical integration could be used to calculate $\widetilde{\omega}_l^N$ to approximate $\widehat{\omega}_l$,

$$\widetilde{\omega}_l^N = \sum_{l=1}^{Q} \tilde{T}^j \cdot \Phi_l(\xi^{(j)}) \cdot \alpha^{(j)}, \tag{5.23}$$

$$\widetilde{\omega}_l^N \to \widehat{\omega}_l \quad \text{as} \quad Q \to \infty. \tag{5.24}$$

94

Finally, we replace $\hat{\omega}_l$ in (5.20) by $\tilde{\omega}_l^N$, obtaining

$$I_N T = \bar{T}_N = \sum_{l=1}^{M} \tilde{\omega}_l^N \Phi_l(\xi), \qquad (5.25)$$

$$\bar{T}_N \to T_N \quad \text{as} \quad Q \to \infty. \qquad (5.26)$$

Through the triangular inequality involving *projection error* and *aliasing error*, $\bar{T}_N$ will converge to T, which is guaranteed by (5.21) and (5.26). $\bar{T}_N$ can approximate T to a required accuracy by increasing the order of polynomial and the number of nodes for integration. If the error of numerically solving T is also considered, the approximation is still valid as long as such numerical method gives convergent solution to the deterministic problem (Xiu 2007).

The failure time distribution obtained by PCE is an approximative solution, of which the accuracy can be refined by increasing the polynomial degree and the number of integration points. Comparing to Monte Carlo sampling, two merits of PCE are distinct. One is the reduction of computational time because the executions of the deterministic problem, say, Paris' law equation here, are needed only at the selected nodes in the sparse grid. The other merit is that the post-processing work is simple because the evaluation of polynomials is a trivial task. Since the failure time is expressed in a form of polynomial expansion, its density function and the associated moments can be obtained in a very efficient manner.

5.3.3 Stochastic collocation method based on PCE for uncertainty quantification in Bayesian inference

As discussed in Section 5.3.1, we divide the set of all random model inputs, Θ, appearing in the degradation model into two subsets: updating-uncertainty U, and non-

updating-uncertainty (likelihood-uncertainty) $\boldsymbol{R}$. The condition monitoring data provide the estimated crack size with uncertainty at inspection time. Denote the random variable of the estimated crack size as a, the updating-uncertainty vector as $\boldsymbol{u}$, and the likelihood-uncertainty vector as $\boldsymbol{r}$. The formula of Bayesian rule is:

$$f_{post}(\boldsymbol{u}|a) = \frac{L(a|\boldsymbol{u})f_{prior}(\boldsymbol{u})}{\int L(a|\boldsymbol{u})f_{prior}(\boldsymbol{u})\, d\boldsymbol{u}} , \qquad (5.27)$$

Given a fixed value of updating-uncertainty $\boldsymbol{u}$, the likelihood to observe a crack size at a given inspection time depends on two factors: measurement error e and likelihood-uncertainty $\boldsymbol{r}$.

Let a_j^{sim} denote the crack size obtained by Paris' law at inspection time j. Because of the uncertainty in the non-updating-uncertainty $\boldsymbol{r}$, a_j^{sim} becomes a random variable with density $h_j(a)$. Meanwhile, by considering the measurement error, the estimated crack size at such inspection time has the following normal distribution:

$$a_j^{obs} \sim N\!\left(a_j^{sim}(\boldsymbol{r}), \sigma^2\right) \qquad (5.28)$$

Denote the density function in (5.28) as $g_j(a)$. Then the likelihood to observe the crack size a_j^{obs} at inspection time j is defined as

$$L\!\left(a_j^{obs} \mid \boldsymbol{u}\right) = \int g_j\!\left(a_j^{obs} \mid a_j^{sim}(\boldsymbol{r}), \boldsymbol{u}\right) \cdot h_j\!\left(a_j^{sim}(\boldsymbol{r}) \mid \boldsymbol{u}\right) \mathrm{d}a_j^{sim}(\boldsymbol{r}) \quad (5.29)$$

This formulation of the likelihood function accounts for both effects of the non-updating-uncertainty and the measurement error. In the likelihood function (5.29), $g_j(a)$ is known as a Gaussian PDF. However $h_j(a)$ is unknown, and can only be obtained through uncertainty propagation. If we have high dimensional uncertainty in the set of $\boldsymbol{R}$,

getting $h_j(a)$ is a time-consuming task using Monte Carlo simulation. Thus, to improve the computation efficiency, we employ stochastic collocation method based on PCE to replace Monte Carlo simulation to perform this task.

Suppose the likelihood-uncertainty in R to be i.i.d random vector $r = (r_1, \cdots, r_D)$. The crack size $a_j^{sim}(r)$ at inspection time j is a random variable due to uncertainty in r. It can be approximated by PCE as,

$$a_j^{sim}(r) = \sum_{i \leq N} \widetilde{\mu}_{ij} \Phi_i(r), \quad j = 1,2, \cdots \tag{5.30}$$

$$\widetilde{\mu}_{ij} = \sum_{l=1}^{Q} a_j^R(r^l) \Phi_i(r) \alpha^l, \qquad l = 1,2, \cdots, Q \tag{5.31}$$

where $r^{(l)}$ are the pre-selected nodes and $\alpha^{(l)}$ are the associated weights for integration.

The prior distribution of updating parameters in U can be obtained using the similar procedure discussed in Section 4.6.

With the prior distribution and the likelihood function available, Bayesian inference as shown in (5.27) can be applied to update the distribution of uncertain parameters in the set of U. With the updated distribution of parameters, the RUL distribution can be predicted using PCE stochastic collocation method discussed in Section 5.3.2.

5.4   Example

5.4.1 Introduction

In this section, we present a numerical example on integrated gear prognosis using the proposed PCE stochastic collocation method for uncertainty quantification. The physical

models were discussed in Chapters 3 and 4. The FE model of spur gear tooth in Figure 16 with the property in Table 3 is utilized in this chapter. The SIF in Figure 17 and the dynamic load in Figure 18 are adopted as well in this chapter. The uncertainty sources considered include model error ε, measurement error e, and two material random parameters in Paris' law, C and m. Other parameters appearing in Paris' law are treated as deterministic. Divide the random inputs appearing in Paris' law into two subsets: $\boldsymbol{R} = \{C\}, \boldsymbol{U} = \{m\}$. The crack observations are simulated by Paris' law with artificially added measurement error.

To generate the simulated degradation paths, the following values and distributions for the parameters are assumed: $m \sim N(1.4354, 0.2^2), C \sim N(9.12e - 11, (1e - 12)^2)$, $\sigma = 0.15$, $\varepsilon \sim Ln(0.8924, 0.2128^2)$. It is worth noting that here the uncertainty regarding m is for the gear population, not for the specific gear being monitored.

In this example, ten degradation paths are generated, as shown in Figure 26. Seven of them consist of the training set, and the three remaining in bold are for test. The selection of the two groups is random in theory while there is a underlying preference. That is, the paths bearing the value of "m" which are far from the mean of prior distribution have the priority to be selected in order to better show the method's capability of tuning parameters. In Figure 26, the paths with the longest (path #1) and the shortest (path #4) lifetimes are selected in the test set. Furthermore, path #7 which has similar value of "m" as the mean of prior distribution is also targeted to illustrate the stability of proposed method in an opposite perspective.
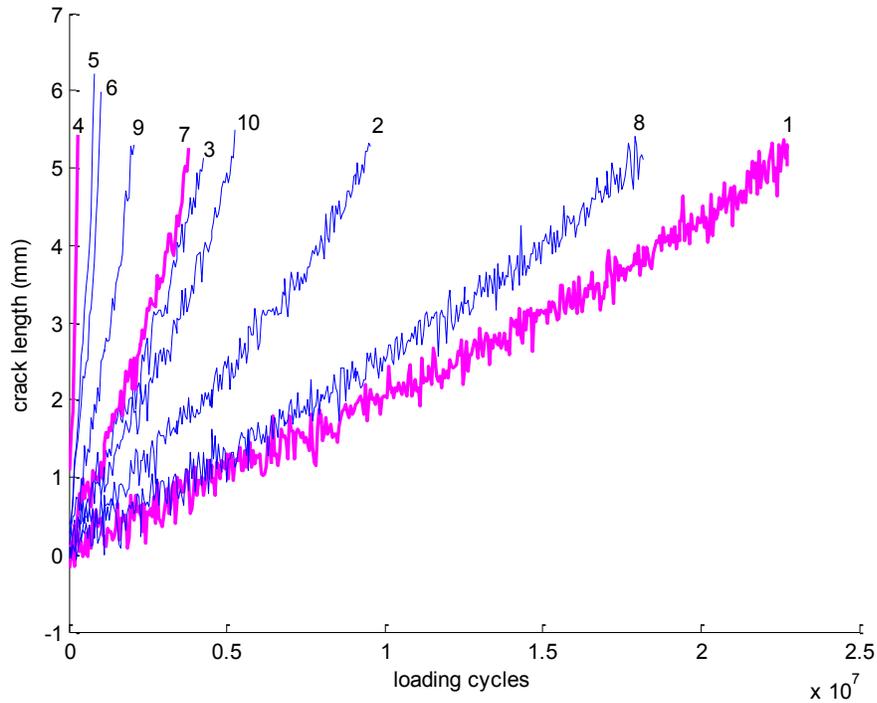
Figure 26. Ten degradation paths generated using prescribed parameters

The purposes of the example are two-fold. One is to show the effectiveness of the integrated prognostics method with PCE stochastic collocation method for uncertainty quantification. The other purpose is to demonstrate the computational efficiency of PCE by comparing with Monte Carlo simulation. The results are shown in the next two subsections.

5.4.2 Results on update process

The ten generated degradation paths are summarized in Table 8. This table tabulates the failure times, the real values of m generating the associated paths as well as the trained values of m. The test set consists of paths #1, #4 and #7. Normal distribution is used to fit the other seven paths in the training set to obtain the prior distribution of m:

$$m \sim N(1.4029, 0.1878^2) \qquad\qquad (5.32)$$

Table 8. The real values and the trained values of m

Path #	Failure cycles	Real value	Trained value
1	**22.75e6**	**1.1009**	-
2	9.55e6	1.2403	1.2365
3	4.25e6	1.3724	1.3671
4	**0.3e6**	**1.7968**	-
5	0.8e6	1.6454	1.646
6	1.0e6	1.6098	1.6064
7	**3.8e6**	**1.3884**	-
8	18.15e6	1.1388	1.1377
9	2.05e6	1.4863	1.4893
10	5.3e6	1.3364	1.3375

The only likelihood-uncertainty in this example is the material parameter C following normal distribution. Hence, up to third-order Hermite orthogonal polynomials are used in PCE and a sparse grid containing 7 points is selected. The updating-uncertainty in this example is the material parameter m, which also follows normal distribution. At each inspection time, this distribution is updated by Bayesian inference. From that inspection time, the RUL is calculated based on the distributions of both C and m. To achieve this, a two-dimensional sparse grid with 13 nodes is selected. The third-order polynomial space is selected in PCE to approximate the RUL distribution.

Table 9, Table 10 and Table 11 show the updating results for both m and failure time. From these results, it is observed the convergence of the prior distribution to the actual values is obvious. For example, starting with the prior value 1.4029, m approaches its

respective real values in the three paths gradually, as the condition monitoring data on the crack size are fused into Bayesian inference. Meanwhile, the uncertainty reduction in both m and failure time is apparent. The failure time distributions for path #1, #4 and #7 are shown in Figure 27, Figure 28 and Figure 29 respectively, from which we can observe that with the updated m, the predicted failure time distribution becomes narrower and its mean is approaching the real failure time. This example only utilizes low degree of polynomial and few collocation nodes in PCE, however, the prediction accuracy is satisfactory. In the following subsection, comparative study is conducted to illustrate the effects of the truncated polynomial degree and the number of collocation nodes on the accuracy of PCE. It shows that with the increase of both factors, the PCE accuracy can be improved.

Table 9. Testing results for path #1 (real m=1.1009, real failure time=2.275e7 cycles)

Inspection cycle	Crack length (mm)	Mean of m	Std of m	Mean of Failure time	Std of Failure time
0	0.1	1.4029	0.1878	6.9703e6	9.2389e6
0.7e7	1.6849	1.1257	0.0153	1.947e7	1.2064e6
1.4e7	2.9092	1.1024	0.0099	2.2493e7	5.5313e5
2.1e7	4.6018	1.0991	0.0073	2.2789e7	9.3263e4

Table 10. Testing results for path #4 (real m=1.7968, real failure time =0.3e6 cycles)

Inspection cycle	Crack length (mm)	Mean of m	Std of m	Mean of Failure time	Std of Failure time
0	0.1	1.4029	0.1878	6.7118e6	8.8709e6
0.1e6	1.4971	1.7924	0.0184	3.0302e5	2.3163e4
0.2e6	3.1079	1.8008	0.0114	2.8527e5	7.4258e3
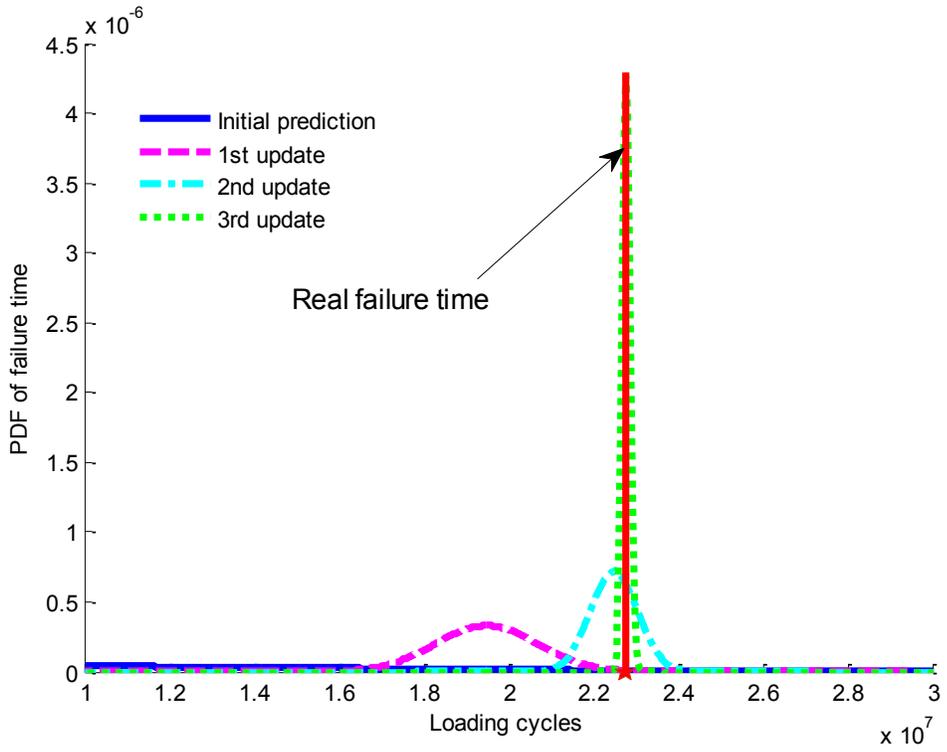
Figure 27. Updated failure time distributions for path #1
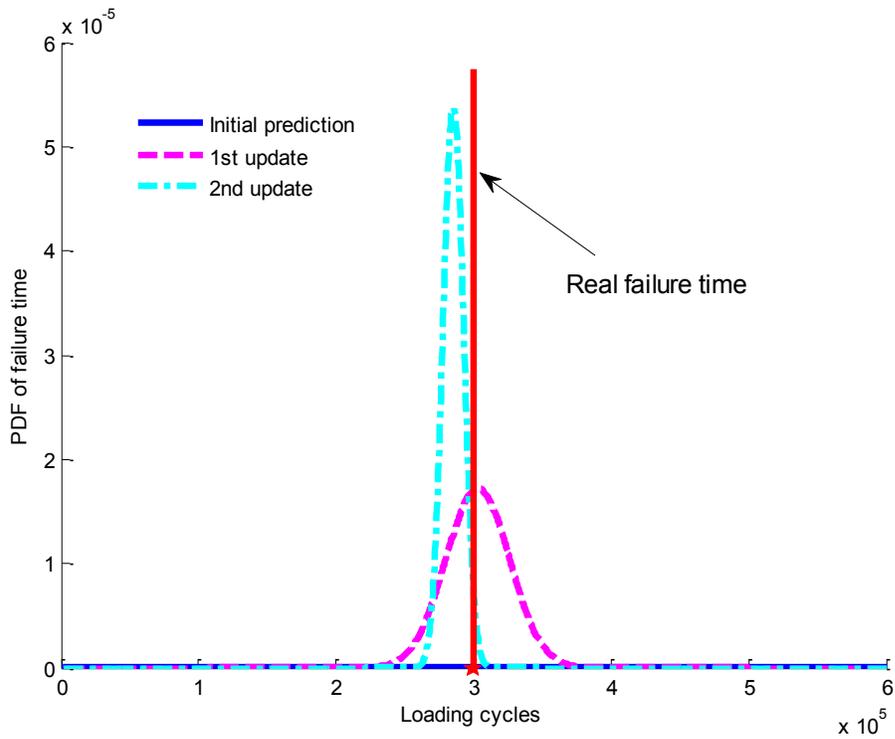
Figure 28. Updated failure time distributions for path #4

Table 11. Testing results for path #7 (real m=1.3884, real failure time=3.8e6 cycles)

Inspection cycle	Crack length (mm)	Mean of m	Std of m	Mean of Failure time	Std of Failure time
0	0.1	1.4029	0.1878	6.7092e6	8.8759e6
1.2e6	1.4387	1.3839	0.0185	3.8888e6	3.1581e5
2.4e6	3.022	1.3947	0.0119	3.6268e6	9.7808e4
3.6e6	4.7519	1.3799	0.007	3.8027e6	1.0637e4
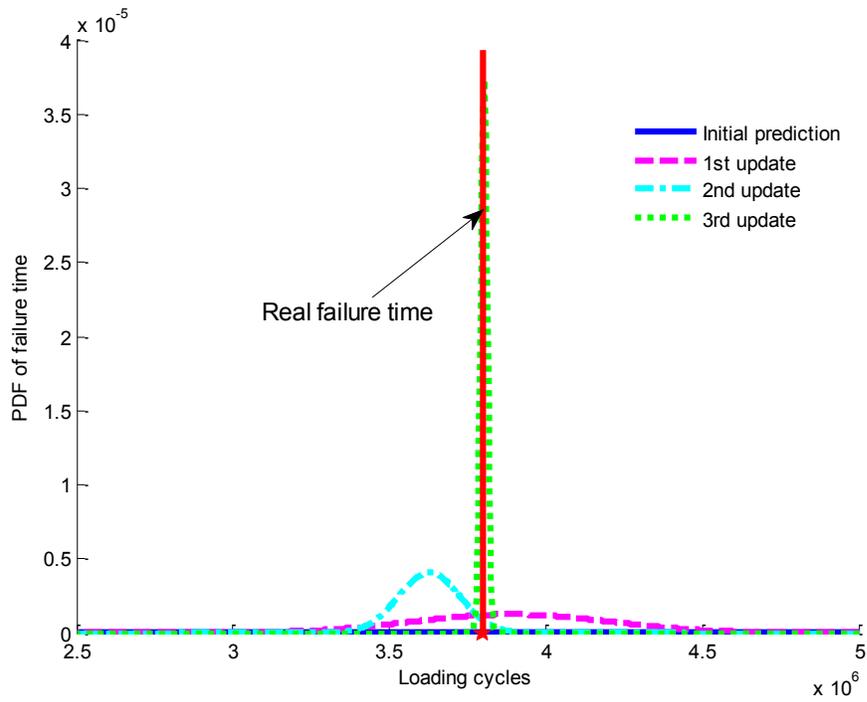
Figure 29. Updated failure time distributions for path #7

5.4.2 Comparative study on PCE accuracy

The accuracy of PCE is controlled by the error of projection in (5.13) and the error of numerical integration in (5.15). Hence, two controllable factors, which are the truncated polynomial degree and the number of collocation nodes in the numerical integration, determine the accuracy of PCE stochastic collocation method. A comparative study is

conducted in this subsection to illustrate the effects of the two factors. In this investigation, we select the RUL prediction as the representative case, in which two uncertainty factors, C and m are the random inputs. And we focus on the entire failure time during which the crack is propagated from its initial size to the critical value.

Due to the lack of explicit form of the exact solution, we consider the result obtained using Monte Carlo simulation as the real distribution of the failure time, so as to compare with the results obtained by PCE stochastic collocation method. Comparative studies are conducted to demonstrate 1) the effect caused by truncated polynomial degree by fixing the number of integration nodes, and 2) the effect caused by number of integration nodes by fixing the truncated polynomial degree. The mean and standard deviation of RUL distribution predicted by PCE are compared with those of the real distribution obtained by Monte Carlo simulation.

The comparison results are tabulated in Table 12 and Table 13, respectively. As shown in Table 12, with a fixed and sufficient degree of polynomial space, the mean of and the standard deviation of the RUL approach those of the real RUL, as the number of integration nodes increase. On the other hand, Table 13 shows the accuracy improvement by increasing the truncated polynomial degree with a fixed number of integration nodes. It is worth noting that the latter type of improvement is limited by the number of integration nodes. It is possible that the higher order polynomial may not perform better than the lower order one, unless we increase the integration accuracy at the same time, i.e., the number of integration nodes. It is concluded that PCE accuracy can be improved by increasing both the truncated polynomial degree and the number of integration nodes.

But it should be noted that the increase in polynomial degree and number of integration nodes will lead to the increase in computation burden.

Table 12. Effect of number of integration nodes with 6th order polynomial space

RUL	5 nodes	13 nodes	29 nodes	53 nodes	89 nodes	137 nodes	5000 MC
Mean	5.3733e6	5.9348e6	6.038e6	6.1427e6	6.1837e6	6.1943e6	6.1974e6
Std	9.5945e9	1.1115e7	1.214e7	1.28e7	1.3591e7	1.3867e7	1.3912e7

Table 13. Effect of truncated polynomial degree with 137 integration nodes

RUL	2nd order	3rd order	4th order	6th order	7th order	8th order	5000 MC
Mean	6.0472e6	6.087e6	6.161e6	6.1943e6	6.1914e6	6.2072e6	6.1974e6
Std	1.0315e7	1.1954e7	1.3042e7	1.3867e7	1.3922e7	1.413e7	1.3912e7

5.4.3 Comparative study on PCE efficiency

The implementation of the integrated prognostics method is conducted in MATLAB. The code mainly includes two phases: Bayesian update and RUL prediction. In this subsection, the comparison will be conducted on the computational time consumed by PCE stochastic collocation and Monte Carlo simulation. Path #1 is selected as an example to demonstrate the comparison results.

Monte Carlo simulation is a random sampling method. The model needs to be evaluated at all the samples. As a matter of fact, the nature of PCE stochastic collocation method is also a process of sampling. However, it samples following "rules" which decide the selection of integration points. The model only needs to be evaluated at these integration points to obtain the coefficients in the polynomial expansion. After obtaining the coefficients, the remaining work is polynomial evaluation which costs negligible

computational time compared to the model evaluation. Hence, the computational time saved by PCE is mainly attributed to the reduction of times to evaluate the model.

Let's assume it will take around 2 seconds to compute the failure time using discrete Paris' law. If Monte Carlo simulation is used, at least 1000 iterations is required to get a good picture of failure time statistics, which will take around 90 minutes. In contrast, PCE stochastic collocation only needs 13 nodes to give agreed results, which means the Paris' equation only needs to be solved deterministically at these 13 predetermined points. So the time consumed by PCE is only around 26 seconds. That is, the proposed integrated prognosis approach is seventy times faster than the simulation method for the case considered here.

To compare the results obtained by Monte Carlo simulation and PCE stochastic collocation method, the error between the predicted failure time and the real failure time defined in (5.33) is used to measure the prediction performance of these two approaches. Denote $\rho(t)$ as the PDF of failure time predicted at the last inspection time. The real failure time is t_r. The prediction error is defined in a weighted L^2-norm,

$$\left(\int (t - t_r)^2 \rho(t) \mathrm{dt} \right)^{\frac{1}{2}}. \tag{5.33}$$

Table 14 shows the comparison of the prediction errors as well as the computational time of Monte Carlo simulation and stochastic collocation method. From this table, it is seen that PCE not only produces satisfactory results but also saves much computational efforts.

106

Table 14. Comparison of Monte Carlo simulation and PCE collocation method

Approach	Time in two phases		Total time	RUL prediction error (cycle)
	Bayesian update	RUL prediction	Integrated method	
Monte Carlo simulation	1000 loops	1000 loops	-	9.9325e4
1st update	146830 sec	5066.5 sec	≈42 hours and 12 min	
2nd update	151540 sec	3439.5 sec	≈43 hours and 30 min	
3rd update	147470 sec	725.5 sec	≈41 hours and 10 min	
PCE collocation method	7 points	13 points	-	9.8624e4
1st update	959.7 sec	65.1 sec	≈17.1 min	
2nd update	1016.7 sec	44.2 sec	≈17.7 min	
3rd update	963.5 sec	9.6 sec	≈16.2 min	

5.5   Conclusions

In this chapter, within the integrated prognostics framework for gear RUL prediction, a stochastic collocation approach based on PCE is developed for efficient integrated gear health prognosis. Instead of using simulation, stochastic collocation method based on PCE is employed to evaluate the uncertainty in gear RUL prediction and to compute the likelihood function in Bayesian inference. Two categories of random parameters appearing in Paris' law are also defined.

The results in the numerical example demonstrate that the integrated prognostics method with stochastic collocation method for uncertainty quantification can effectively

and efficiently adjust the model parameters based on the observed degradation data, and thus lead to more accurate RUL prediction. The significant improvement of computational efficiency provided by PCE enables us to consider more uncertain factors in a practical way. This method has the potential to be applied to other rotating components, such as bearings, shafts, and structures, such as aircraft structures, bridges, pipelines, pressure vessels, etc. These potential applications require investigations in building the physical models for these components and structures, integration of the PCE methods, updating methodologies, etc.

Chapter 6.　An integrated prognostics method under time-varying operating conditions

5　Overview

The recent interest in prognostics under variable loads is fuelled by operations and maintenance personnel's need for decision support tools. Equipment subjected to time-varying operating conditions imposes a demanding requirement on a prognostic method because of its dynamic nature. Prognostics should account for changes in operating conditions and report an accurate RUL in a timely manner. The research efforts to address prognostics under time-varying operating conditions are driven by the need of on-line prognostics for manufacturing process, numerical controlled machining as well as other scenarios where the changes in operating conditions during production are unavoidable. A gearbox under time-varying operating conditions is investigated in this chapter. The time-varying environment could be due to the changes in temperature, load, lubrication, speed, etc. This chapter specifically addresses changes in load, which is the most important operating condition factor for a power transmission system.

The existing studies concerning prognostics under time-varying environment are mostly data-driven. In (Gebraeel and Pan 2008), a linear degradation model was assumed. The effects of time-varying operating conditions on the degradation signal were taken into account by the coefficients assigned to these time-varying environmental parameters. Bayesian methods were used to derive the posterior distribution of these coefficients. Because of the linear model assumption, analytical expressions for the posterior distribution as well as that for the residual life distribution were available. Liao and Tian

(2013) extended the prognostics in time-varying operating conditions to non-linear models, in which the degradation process was assumed to be governed by the Brownian motion with linear drift. Stress changes were accounted for in the instantaneous drift parameter. Bayesian inference was employed to estimate the posterior distribution of coefficients in the drift parameter. These approaches tackle the prognostics under time-varying operating conditions through the coefficients estimation in the degradation model. Unfortunately, these data-driven methods do not address the physical mechanism of the degradation, and hence the load has no direct relationship with the parameters in the degradation model. In addition, the effectiveness of data-driven methods also depends heavily on the availability of a set of dense, well-distributed data. It is thus particularly challenging for time-varying operating conditions because it is unlikely that the training set encompasses all the operating conditions.

By noticing the limitations of existing data-driven prognostics methods, the present chapter develops an integrated prognostics approach to deal with time-varying operating conditions. In the integrated prognostics approach, the degradation model is built on the physics of damage progression, which is usually a function of environmental parameters. Any changes of these environmental parameters, such as load, temperature, and speed, can be manifested immediately in the physical model. Hence, a key advantage of using the integrated prognostics method to deal with time-varying operating conditions is its capability to directly relate the environmental parameters to the degradation model. The proposed framework can apply to different mechanical components, given the corresponding physical models.

By noticing the efficiency of stochastic collocation method based on PCE, we will apply it for the uncertainty quantification in this chapter. Considering the uncertainty in the two correlated parameters of Paris' law, this study applies PCE technique to improve the efficiency of MCMC algorithm (Marzouk et al. 2007, Marzouk and Xiu 2009) when updating these parameters via Bayesian inference. A specific PCE formulation is given for the uncertainty quantification. By expressing the likelihood as an explicit function of material parameters, this formulation allows a large amount of samples to ensure MCMC convergence in an efficient manner, hence enabling a fast update of the joint PDF of the two correlated material parameters. Because the updated joint PDF of material parameters can better characterize the degradation process, the failure time distribution based on this joint PDF is expected to be more accurate.
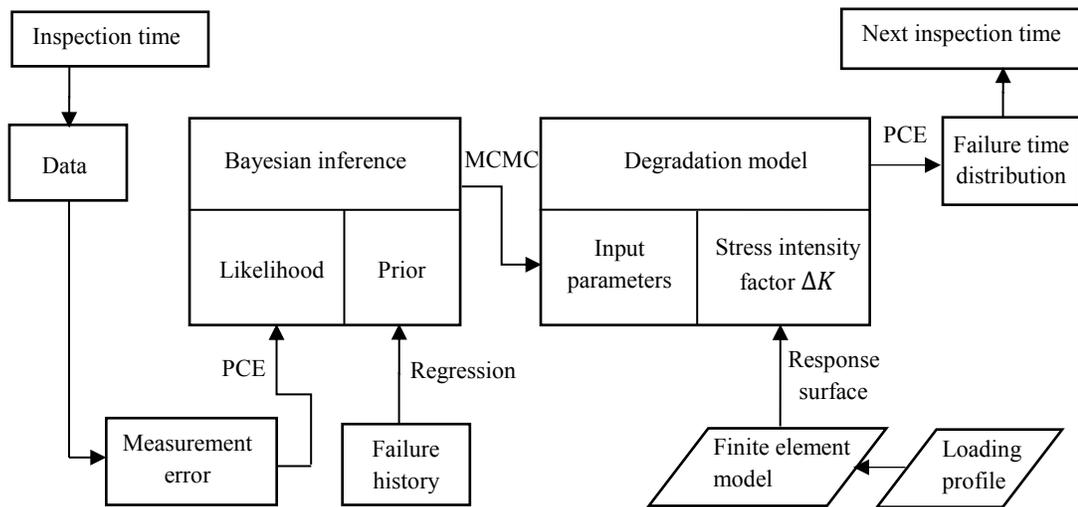
Figure 30. Update process between two consecutive inspection times under time-varying loading conditions.

Figure 30 shows the update process between two consecutive inspection times in the integrated prognostics method considering time-varying loading condition. Since this chapter considers the effect of loading condition on prognostics, loading profile is extracted to be one independent module as an input to FE model. As a consequence, SIF becomes a surface with respect to both crack size and load. Because this surface is obtained offline by running the FE model at a baseline load and a series of selected crack sizes, there is no line connecting data module to FE model.

This chapter is organized as follows. The physical models and the load profile are presented in Section 6.2. In Section 6.3, an integrated prognostics method is proposed under time-varying loading conditions, which includes the Bayesian inference framework, the specific PCE formulation to update two correlated material parameters, the way to obtain prior joint distribution, and the method for RUL prediction. In Section 6.4 examples are given to demonstrate the effectiveness of the method. Section 6.5 concludes the work. The materials in this chapter have been published in (Zhao et al. 2015a).

6.2   Physical models and load change

The physical models include the FE model and the degradation model. When a component is subject to a time-varying loading condition, the degradation process described by the degradation model (e.g., Paris' law) will depend on the load, denoted by lo. To be specific, the expression of SIF range $\Delta K = \Delta K(a, lo)$ suggests that the model degradation rate is a function of crack size and loading condition. In order to capture the degradation pattern of a cracked component using Paris' law, the response surface of ΔK with respect to crack size and load is needed.

A piece of equipment under operation may be exposed to a series of varying loads according to the user's needs. The work logs should record two facts: the time when a loading change occurs, and the amplitude of such a change. The general case may be described as follows. Assume that totally n loading changes happen at time $t_1, t_2 \cdots, t_n$, respectively, and that the load amplitude during $[t_i, t_j)$ is $F_{i,j}$, as shown in Figure 31. This is a loading profile of piece-wise constant.
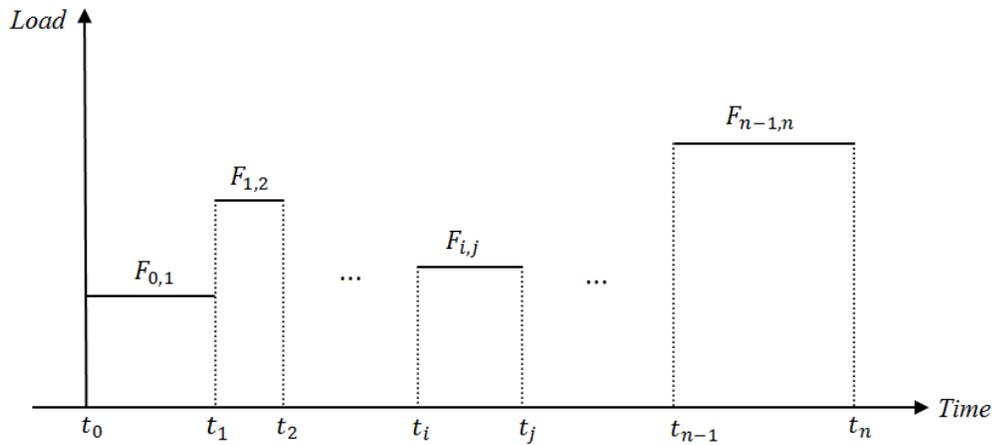
Figure 31. General load changes history.

As discussed, the key value that needs to be determined in the degradation model is the SIF (which determines ΔK in Paris' law), which is primarily a function of crack size, structure geometry and loading condition. The inputs of crack size and structure geometry have already been taken into account in the geometric modeling in the FE model, while the input of loading condition is provided by the load history recorded in the work logs. For complex situations, such as non-linearity and non-elasticity in the materials, the relationship between ΔK and load needs to be obtained by running the FE analysis each time the new load is applied. In this thesis, the stress analysis is constrained within the framework of linear elastic fracture mechanics. As a result, ΔK has a linear

113

relationship with the load. By following this convention, the baseline relationship of ΔK and load may be derived by running FE model once for a baseline load only. The remaining work is merely to multiply the derived ΔK by the ratio between the new and the baseline loads. In this way, a surface of ΔK with respect to different crack sizes and different loads can be obtained. For transmission systems, such as a gearbox, the load is determined by the input torque. Such an observation can significantly improve the efficiency of implementing the proposed prognostics approach.

6.3 Uncertainty quantification for correlated material parameters in Paris' law

Due to its stochastic nature, the crack propagation process should be investigated within a framework of probability theory. Paris' law is used to describe fatigue crack growth undergoing cyclic loading during its stable growth period, in which the material parameters m and C are obtained by fitting the experimental fatigue data. The variability of the crack propagation process should be reflected by the randomness in m and C. It was reported in (Annis 2003) that a strong correlation between m and C must be taken into consideration to achieve acceptable prediction accuracy. In practice, assume a set of experimental trajectories of stochastic crack growth is available, i.e., a set of curves representing crack growth rate da/dN vs ΔK. By linear regression using Paris' law in a log-log scale, for each trajectory, we can determine a pair of $(m, \log C)$ which minimizes the discrepancy between the measurements and the predicted values. A standard statistical analysis can be applied to the sample set of $(m, \log C)$ to infer the best joint probabilistic distribution of $(m, \log C)$. It is found that the bivariate normal distribution is

usually a valid hypothesis for the joint density of $(m, \log C)$ (Ortiz and Kiremidjian, 1988; Annis 2003; Ditlevsen and Olesen 1986; Kotulski 1998). The density obtained in this way can be considered as prior information of the parameters.

As discussed before, precise values are often unknown for these material parameters for a specific unit. Sometimes only the prior distribution is available based on the population failure histories. In this section, the joint distribution of m and $\log C$ will be updated via MCMC in Bayesian inference, by taking advantage of the condition monitoring data on crack size. As more condition data come in, the uncertainty in the material parameters will be reduced, and the mean values may approach the real values.

6.3.1 Updates of the joint distribution of material parameters in Paris' law

Considering the model error, the non-linear dependence of crack growth on the loading cycle is embedded in Paris' law:

$$\frac{da}{dN} = C(\Delta K(a, lo))^m \varepsilon. \tag{6.1}$$

A surface of $\Delta K(a, lo)$ is derived using the method in Section 6.2. It is assumed that material parameters, $(m, \log C)$, is a bivariate Gaussian random vector. The joint distribution of $(m, \log C)$ will be updated through Bayesian inference given the crack sizes estimated at given inspection times.

Equation (6.1) is solved by discretizing it using the first-order Euler method. Let the initial crack length be a_0, and the incremental loading cycles be ΔN; then the discretized Paris' law is

$$\begin{cases} a\big((i+1)\Delta N\big) = a(i\Delta N) + (\Delta N)C[\Delta K(a(i\Delta N), lo(i\Delta N))]^m \varepsilon \\ a(0) = a_0 \end{cases}, \quad i = 0,1,2 \cdots. \quad (6.2)$$

The iteration sequentially proceeds until the inspection time for update is reached. The model error is sampled from its assumed known distribution randomly at each iteration step. The crack size simulated through this discretization is denoted as a^{sim}.

As before, Bayesian inference is used to update the distribution of model parameters, given the estimated crack size as the inspection time. In particular for this chapter, it is the joint distribution of m and $\log C$ that is to be updated because they are correlated with each other.

Assume that during the whole crack propagation process, from the initial detected crack a_0 to the critical crack size a_C where the failure occurs, there are totally U updates at inspection times T_1, $T_2, \cdots, T_U$. At each update time T_j, suppose that the material parameters $\vec{\xi}_j = (m_j, \log C_j)^{\mathrm{T}}$ follow a bivariate normal distribution $N(\vec{\mu}_j, \Sigma_j)$, where $\vec{\mu}_j = (\mu_{m_j}, \mu_{C_j})^{\mathrm{T}}$ is the mean vector, and Σ_j is the covariance matrix with the covariance coefficient ρ_j, where

$$\Sigma_j = \begin{bmatrix} \sigma_{m_j}^2 & \rho_j \sigma_{m_j} \sigma_{C_j} \\ \rho_j \sigma_{m_j} \sigma_{C_j} & \sigma_{C_j}^2 \end{bmatrix}. \quad (6.3)$$

The crack sizes $\vec{a}_{1:j} = (a_1^{obs}, a_2^{obs}, \cdots, a_j^{obs})$ at the inspection times T_1, $T_2, \cdots$ up to T_j are estimated through diagnostic methods. Then, denote the PDF of $N(\vec{\mu}_j, \Sigma_j)$ as $\Pi(\vec{\xi}_j)$,

$$\Pi(\vec{\xi}_j) = \frac{1}{(\sqrt{2\pi})^2} \frac{1}{\sqrt{\det(\Sigma_j)}} \exp\left[-\frac{1}{2} (\vec{\xi}_j - \vec{\mu}_j)^T \Sigma_j^{-1} (\vec{\xi}_j - \vec{\mu}_j) \right]. \quad (6.4)$$

At the next update time T_{j+1}, the crack size a_{j+1}^{obs} is estimated from sensor data. The posterior distribution of $\vec{\xi}_{j+1}$ is obtained using the Bayesian inference formula

$$f_{post}(\vec{\xi}_{j+1}|\vec{a}_{1:j+1}) = \frac{l(\vec{a}_{1:j+1}|\vec{\xi}_j)f_{prior}(\vec{\xi}_j)}{\int l(\vec{a}_{1:j+1}|\vec{\xi}_j)f_{prior}(\vec{\xi}_j)\,d\vec{\xi}_j}. \tag{6.5}$$

Given the assumption that the measurement error, $e_k = a_k^{obs} - a_k^{sim}, k = 1, \cdots, j+1$, are statistically i.i.d. random variables, then the likelihood $l(\vec{a}_{1:j+1}|\vec{\xi}_j)$ is calculated as

$$l(\vec{a}_{1:j+1}|\vec{\xi}_j) = \prod_{k=1}^{j+1} \Phi_k(a_k^{obs}|\vec{\xi}_j), \tag{6.6}$$

where

$$\Phi_k(a_k^{obs}|\vec{\xi}_j) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(a_k^{obs} - a_k^{sim})^2\right]. \tag{6.7}$$

As time passes, new samples are collected. Accordingly, the posterior distribution of $\vec{\xi}_j = (m_j, \log C_j)^{\mathrm{T}}$ is obtained via (6.5) sequentially as j increases. In this way, the joint distribution of material parameters is updated to be more accurate for this specific unit under monitoring. Thus, the failure time predictions based on these parameters will be more accurate and reliable.

When Metropolis-Hastings algorithm in MCMC is used to implement Bayesian inference, the time-consuming part is the calculation of the likelihood function $l(\vec{a}_{1:j+1}|\vec{\xi}_j) = \prod_{k=1}^{j+1} \Phi_k(a_k^{obs}|\vec{\xi}_j)$. For any sample $\vec{\xi}_j$ generated by a random walk, the discretized Paris' law needs to be evaluated to obtain the crack sizes up to the current inspection time T_j. A large number of samples are required to ensure the convergence of the Markov chain so that the posterior distribution will be the stationary state of this

117

Markov chain. This task is computationally prohibitive for an on-line prognostics mission. To improve the computation efficiency, the stochastic collocation method based on PCE is used, as presented in the following.

6.3.2 Stochastic collocation method in Bayesian inference

In Chapter 5, a PCE stochastic collocation method was proposed for the uncertainty quantification in the integrated prognostics method. The specific PCE formulation was presented for updating one material parameter m in the example. This subsection presents the specific formulation of employing PCE to update two correlated material parameters m and C in Paris' law through Bayesian inference. By expressing the likelihood as an explicit function of m and C, this formulation allows a large number of samples to ensure convergence of MCMC in an efficient manner, hence enabling fast update of the joint distribution of the two material parameters.

Consider the material parameters $\vec{\xi} = (m, \log C)$ in Paris' law as a random vector with two components. In this chapter, $\vec{\xi}$ is assumed to follow a bivariate normal distribution with joint density function $\Pi(\vec{\xi})$, where the mean is $\vec{\mu}$, and the covariance matrix is $\Sigma = \begin{bmatrix} \sigma_m^2 & \rho\sigma_m\sigma_c \\ \rho\sigma_m\sigma_c & \sigma_c^2 \end{bmatrix}$. To take advantage of the orthogonality of the basis polynomial functions to reduce the computational work, $\vec{\xi}$ needs to be converted into a random vector $\vec{\zeta}$ whose components are standard statistically i.i.d. Gaussian variables by Cholesky decomposition,

$$\Sigma = L^T L, \tag{6.8}$$

$$\vec{\zeta} = L^{-1}(\vec{\xi} - \vec{\mu}). \tag{6.9}$$

After the change of variable, define the crack size at inspection time T obtained by Paris' law as $a^{sim}(\vec{\xi}) = a^{sim}(\vec{\mu} + L\vec{\zeta}) = a_s^{sim}(\vec{\zeta})$. The polynomial approximation to $a_s^{sim}(\vec{\zeta})$ is denoted by $a_{s,N}^{sim}(\vec{\zeta})$, which is the projection in N-th order polynomial space. Following the notation in Section 5.2, we can write $a_{s,N}^{sim}(\vec{\zeta}) = P_N a_s^{sim}(\vec{\zeta})$. Let $\boldsymbol{i} = (i_1, i_2)$ be an index with $|\boldsymbol{i}| = i_1 + i_2$; then,

$$a^{sim}(\vec{\xi}) = a_s^{sim}(\vec{\zeta}) \approx a_{s,N}^{sim}(\vec{\zeta}) = \sum_{|\boldsymbol{i}|=0}^{N} \widehat{w}_{\boldsymbol{i}} \, \Psi_{\boldsymbol{i}}(\vec{\zeta}). \tag{6.10}$$

The coefficients $\widehat{w}_{\boldsymbol{i}}$ are calculated as

$$\widehat{w}_{\boldsymbol{i}} = \mathbb{E}\left(a_s^{sim}(\vec{\zeta}) \Psi_{\boldsymbol{i}}(\vec{\zeta})\right) = \int a_s^{sim}(\vec{\zeta}) \, \Psi_{\boldsymbol{i}}(\vec{\zeta}) \, X(\vec{\zeta}) \, d\vec{\zeta}; \tag{6.11}$$

and $\Psi_{\boldsymbol{i}}(\vec{\zeta})$ are the orthogonal basis functions defined as products of one-dimensional orthogonal polynomials, satisfying, after normalization, the equality

$$\mathbb{E}\left(\Psi_{\boldsymbol{h}}(\vec{\zeta})\Psi_s(\vec{\zeta})\right) = \delta_{h,s} = \begin{cases} 1, & when \ \boldsymbol{h} = \boldsymbol{s} \\ 0, & otherwise \end{cases}, \tag{6.12}$$

and $X(\vec{\zeta}) = \varphi(\zeta_1)\varphi(\zeta_2)$, $\varphi(x) = (1/\sqrt{2\pi})\exp(-x^2/2)$. Because the random vector is assumed to follow a bivariate normal distribution, a set of Hermite polynomial is selected as basis functions in polynomial space. To reduce computational work, a sparse grid containing R pairs of integration points and associated weights, $\{(\vec{\zeta}^{(j)}, \beta^{(j)}), j = 1, \cdots, R\}$, is generated for computing the coefficients in (6.11) numerically as

$$\widetilde{w}_{\boldsymbol{i}} = \sum_{j=1}^{R} a_s^{sim}(\vec{\zeta}^{(j)}) \, \Psi_{\boldsymbol{i}}(\vec{\zeta}^{(j)}) \, \beta^{(j)}. \tag{6.13}$$

If we define

$$I_N a_s^{sim}(\vec{\zeta}) = a_{s,I}^{sim}(\vec{\zeta}) \triangleq \sum_{|\boldsymbol{i}|=0}^{N} \widetilde{w}_i \, \Psi_i(\vec{\zeta}) \,. \tag{6.14}$$

then, we have

$$a_{s,I}^{sim}(\vec{\zeta}) \to a_s^{sim}(\vec{\zeta}) = a^{sim}(\vec{\xi}) \quad as \; N \to \infty, R \to \infty. \tag{6.15}$$

Equation (6.15) is of essential importance for accelerating Bayesian inference implementation because it provides an efficient way to calculate the likelihood function $l(\vec{a}_{1:j+1}|\vec{\xi}_j) = \prod_{k=1}^{j+1} \Phi_k(a_k^{esti}|\vec{\xi}_j)$. As mentioned previously, to obtain the posterior joint distribution, each random walk in MCMC needs the evaluation of Paris' law once. A large number of MCMC samples could consume a large amount of computational time for Markov chain to converge. With the availability of $a_{s,I}^{sim}(\vec{\zeta})$ as an approximation to $a^{sim}(\vec{\xi})$ as shown in (6.15), the expression of $a_{s,I}^{sim}(\vec{\zeta})$ is simply a combination of polynomials. For each random walk, $a^{sim}(\vec{\xi})$ is easily approximated by $a_{s,I}^{sim}(\vec{\zeta})$. The performance of such an approximation depends on the order of polynomial space as well as the number of points in the collocation set.

6.3.3 Prior distribution in Bayesian inference

To initiate the Bayesian inference, the information of the prior distribution of the material parameters is needed. Because an individual component is the focus, the population can be naturally considered as a candidate for a priori. Hence, to get the prior distribution of $(m, \log C)$, one assumes historical data are available for F identical gear sets under identical constant loading condition. Each history serves as a degradation path, with loading cycles and the associated crack sizes. Following the standard crack fatigue

test procedure (ASTM E647-00, 2000), the linear regression of $(m_i, \log C_i)$, $i = 1, 2, \cdots, F$, can be obtained for each failure history. Let

$$\bar{m} = \frac{1}{F} \sum_{i=1}^{F} m_i, \quad s_{mm} = \frac{1}{n} \sum_{i=1}^{F} (m_i - \bar{m})^2 ; \tag{6.18}$$

$$\overline{\log C} = \frac{1}{F} \sum_{i=1}^{F} \log C_i, \quad s_{cc} = \frac{1}{F-1} \sum_{i=1}^{F} (\log C_i - \overline{\log C})^2 \tag{6.19}$$

$$s_{mc} = \frac{1}{F-1} \sum_{i=1}^{F} (m_i - \bar{m})(\log C_i - \overline{\log C}) ; \tag{6.20}$$

$$\boldsymbol{S}_{mc} = \begin{bmatrix} s_{mm} & s_{mc} \\ s_{mc} & s_{cc} \end{bmatrix}. \tag{6.21}$$

Then the prior distribution is selected to be $N\left(\left[\bar{m}, \overline{\log C}\right]^T, \boldsymbol{S}_{mc}\right)$. The regression process can also be implemented by simulation based optimization, see Section 4.6. The objective is to find the optimal value $\left(m_i^{op}, \log C_i^{op}\right), i = 1, 2, \cdots, F$, which generates the degradation path that has the minimum difference from the real degradation path in a least-square sense.

6.3.4 RUL prediction

By using the crack size estimation as the observation, Bayesian inference is able to update the joint distribution of the material parameters. The RUL or failure time prediction is conducted after the updated distribution is available. Paris' law can be written in its reciprocal form as follows,

$$\frac{\mathrm{d}N}{\mathrm{d}a} = \frac{1}{C(\Delta K(a, lo))^m \varepsilon}. \tag{6.16}$$

121

Let the current inspection cycle be N_t and the crack increment be Δa. The RUL is calculated by discretizing (6.16) in the following way,

$$\Delta N_i = N_{i+1} - N_i = \Delta a [C \Delta K(a_i, lo_i)^m \varepsilon]^{-1}, i = t, t+1, \cdots \quad (6.17)$$

The summation $\sum_{i=t} \Delta N_i$ from current inspection cycle to the cycle where failure occurs is the RUL. Accordingly, the failure time is expressed as $N_t + \sum_{i=t} \Delta N_i$.

PCE is used to quantify the uncertainty of material parameters in the RUL or the failure time. The procedure was detailed in Section 5.3.2 for the case when the two material parameters were considered independent to each other. In this chapter, the two parameters are considered to be correlated. Hence, the first step to is to use the change of variable in (6.9) to transform them into i.i.d. random variables so that PCE can be applied. After this step, the remaining steps are similar to those provided in Section 5.3.2.

6.4   Examples

The crack propagation at the root of a spur gear is taken as an example to demonstrate the proposed method. To deal with time-varying loading condition, the baseline torque is selected as 40 N-m. The response surface of SIF as a function of crack size and load is shown in Figure 32. According to the linear elastic fracture mechanics theory, this surface is linear with respect to load and nonlinear with respect to crack size. Here, cubic polynomial is used to fit this nonlinearity. With this surface available, the SIF at any combination of load and crack size is obtained by simply looking up the corresponding value in this surface. Hence, it is unnecessary to run the FE model for every case during online prognostics, which saves considerable computational time.

Ten degradation paths are generated using the following parameters: $\mu_m = 1.4354$, $\mu_C = -23.118$, $\sigma_m = 0.2$, $\sigma_c = 0.5$, $\rho = -0.99$. Two examples are conducted in this section. The loading change pattern in Example 1 is a two-step load change, in which the stress is held constant during two consecutive times when load changes. The purpose is to demonstrate that, as more crack estimations are incorporated into Bayesian inference, the updated distributions will gradually converge to the corresponding real values. Furthermore, in Example 2, we will show that the proposed method is effective even when the loading profile changes. A loading profile of a three-step load change is used. It demonstrates that two different loading histories will have similar narrow posterior distributions for $(m, \log C)$.
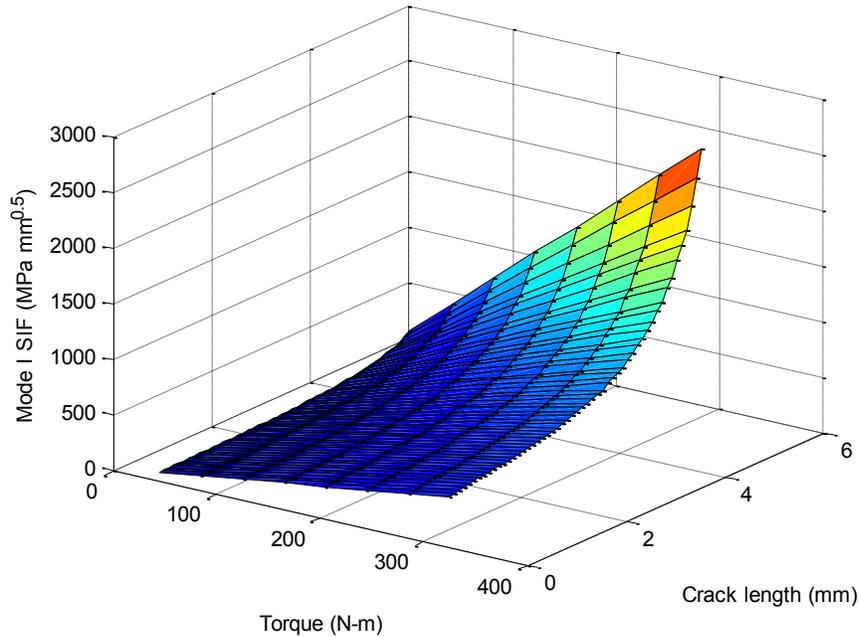
Figure 32. Surface of SIF as a function of load and crack size.

6.4.1 Example 1

The simulated degradation paths with measurement error $\sigma = 0.15$ mm are shown in Figure 33. Consider that the torque increases from 40 N-m to 120 N-m at 0.5×10^7 cycles, and returns to 40 N-m at 2×10^7 cycles. Under this two-step load change condition, two distinct changes in the slope of the degradation path are observed in Figure 33 because the crack growth rate is changed when the load changes. The prior joint distribution is obtained based on the first 8 paths among these ten paths, giving

$$(m, \log C) \sim N \left(\begin{bmatrix} 1.4472 \\ -23.12 \end{bmatrix}, \quad \begin{bmatrix} 0.0229 & -0.052 \\ -0.052 & 0.1203 \end{bmatrix} \right). \tag{6.18}$$
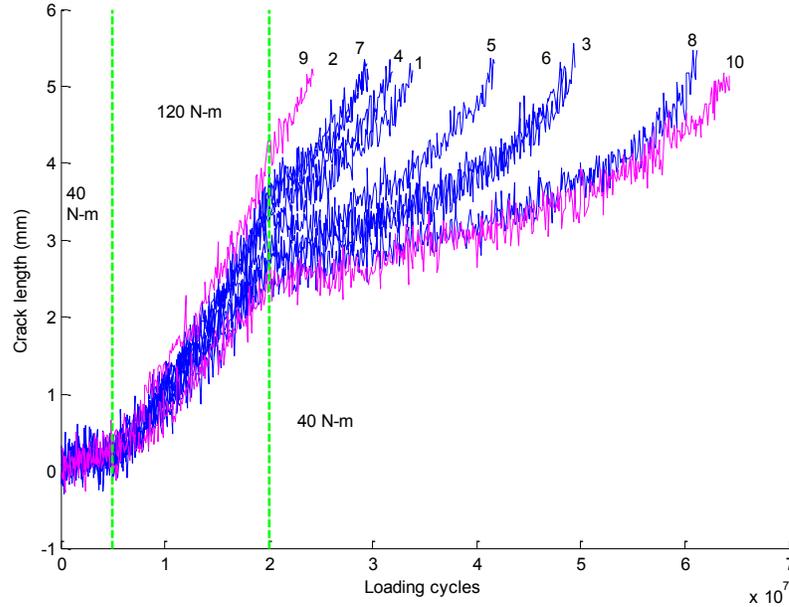
Figure 33. Ten degradation paths generated under two-step load change.

To test the proposed method, two extreme paths, path #9 and path #10, are selected, which have the shortest failure time and the longest one, respectively. The Bayesian update is performed at each inspection time, which is equally spaced. Some of the intermediate update steps will be tabulated in the table to show the trend of distribution adjustment as more data on crack size become available.

124

The real material parameters used to generate the path #9 is: $(m, \log C) = (1.1495, -22.4311)$ and the real failure time is 2.43×10^7 cycles. The inspection time interval is 4×10^6 cycles. The updating results are shown in Table 15 from which it can be observed that the mean of the material parameters are approaching their real values as more observations are available. Figure 34 displays the contours of the prior and the posterior distribution of the last update. The update process of failure time distribution obtained using the updated material parameters distribution is displayed in Figure 35. The failure time distribution gets narrower and approximates to the real failure time as expected. The uncertainty in the predicted failure time is thus reduced.

Table 15. Testing results for path #9

Inspection cycle	Crack length (mm)	μ_m	μ_C
0	0.1000	1.4472	-23.1200
8×10^6	0.6373	1.4590	-23.1479
12×10^6	2.2764	1.1871	-22.4771
16×10^6	2.8742	1.1637	-22.4067
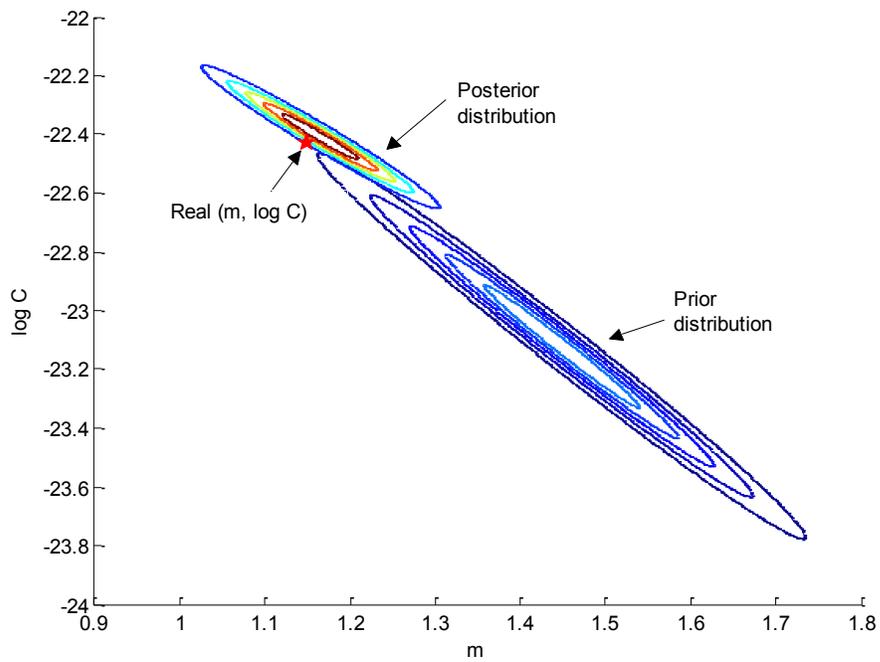
Figure 34. Contours of prior and posterior distribution of $(m, \log C)$ for path #9.
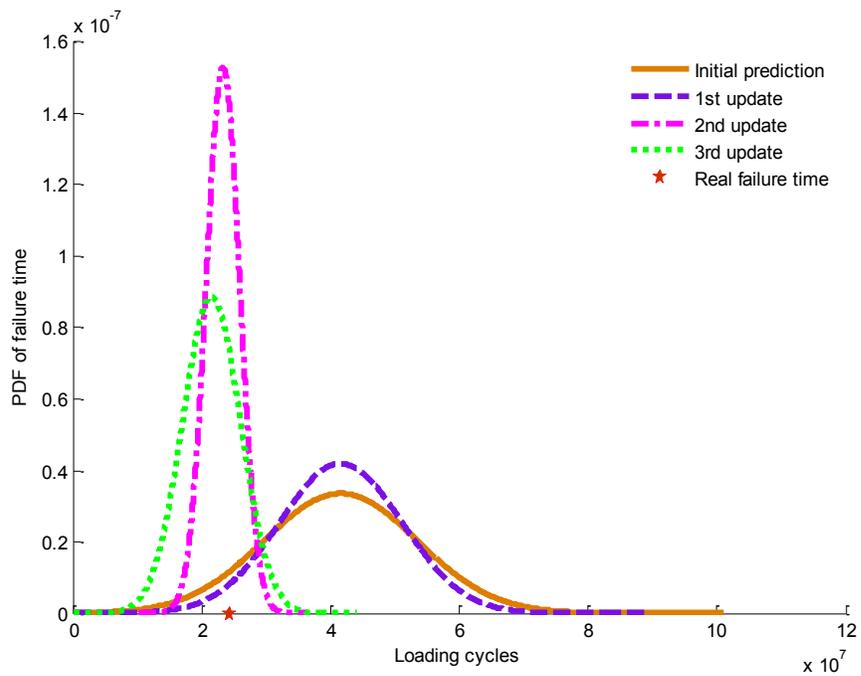
Figure 35. Updated failure time distribution for path #9.

Similarly, the updating results for path #10, which has the longest failure time, are shown in

Table 13. The real material parameters used to generate the Path #10 is $(m, \log C) = (1.6336, -23.6258)$ and the real failure time is 6.43×10^7 cycles. The inspection time interval is 10×10^6 cycles. The contours of the prior and the posterior distribution of the last update are shown in Figure 36. The predicted failure time distributions are presented in Figure 37. The parameters adjust themselves to get close to their real values. Accordingly, the uncertainty in the failure time distribution is reduced gradually, the mean of which approaches the real failure time.

Table 16. Testing results for path #10

| Inspection cycle	Crack length (mm)	μ_m	μ_C
0	0.1000	1.4472	-23.1200
10×10^6	0.7293	1.5670	-23.4063
20×10^6	2.4450	1.6515	-23.6117
60×10^6	4.3758	1.6510	-23.6457
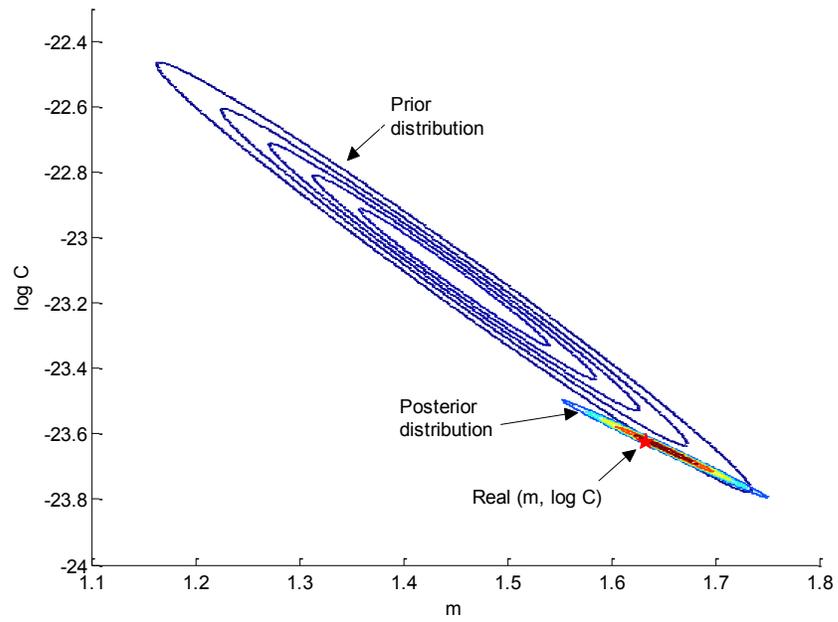
Figure 36. Contours of prior and posterior distribution of $(m, \log C)$ for path #10.
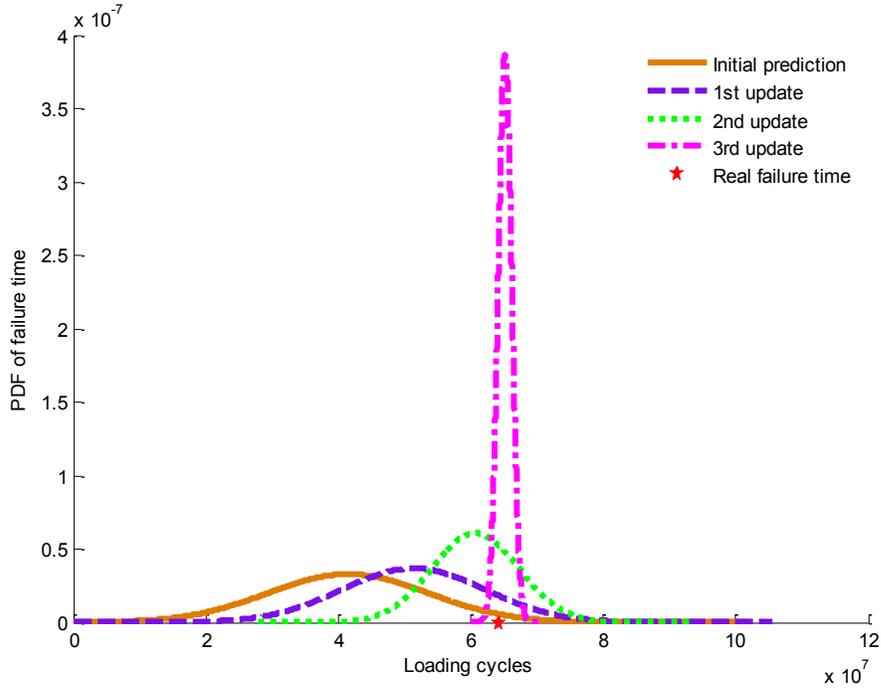
Figure 37. Updated failure time distribution for path #10.

6.4.2 Example 2

For a specific gear, the values of $(m, \log C)$ are material dependent, and are not supposed to change with different loading profiles. In this example, path #11 is subject to a loading profile different from that in Example 1, and a three-step load change is used. At 0.5×10^7 cycles, the torque increases from 40 N-m to 160 N-m; at 1.5×10^7 cycles, the torque returns to 40 N-m; and at 3×10^7 cycles, the torque goes up to 120 N-m until the component failed. The degradation path is shown in Figure 38. The true values of the material parameters for this component are $(m, \log C) = (1.6336, -23.6258)$, the same as that in path #10 in Example 1. The real failure time is 3.44×10^7 cycles. The inspection time interval is 4×10^6 cycles. The update process is listed in Table 17. Even though the last update of $(m, \log C)$ deviates a

small amount from their real values, the failure time distribution gets closer to the real failure time, shown in Figure 39. This example demonstrates that the proposed method is effective, even when the current loading profile is different from the loading profile under which historical data were collected.

It may be worth noticing that in Paris' law, different combinations of m and C could lead to the same crack size at a given loading cycles. When the measurement error is too large for the Bayesian inference to discriminate the real values of $(m, \log C)$ from the noise, there is a possibility that the updated $(m, \log C)$ deviates from the true value. However, the failure time still approaches the real failure time. A similar conclusion was made in (An et al. 2012).
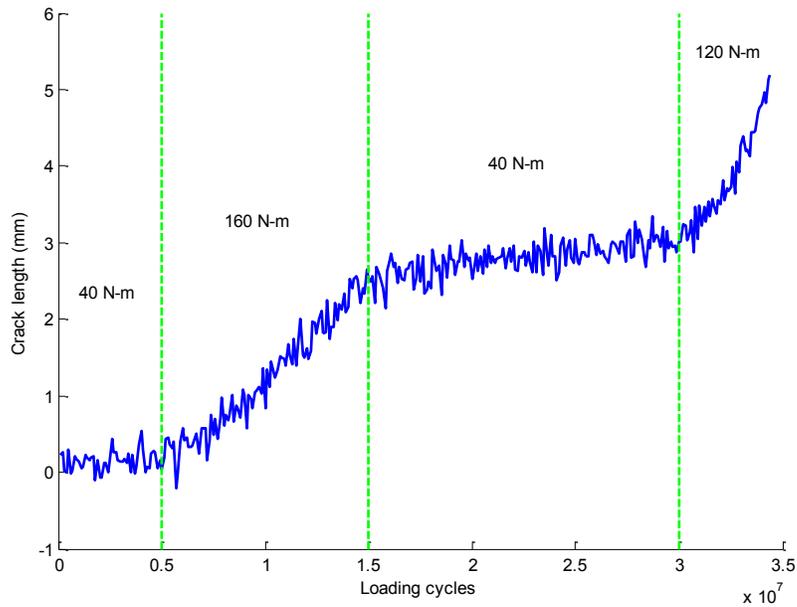
Figure 38. Degradation path of #11 generated under three-step load changes.

Table 17. Testing results for path #11

Inspection cycle	Crack length (mm)	μ_m	μ_c
0	0.1000	1.4472	-23.1200
12×10^6	1.6192	1.6036	-23.5285
20×10^6	2.5303	1.6547	-23.6579
32×10^6	3.5021	1.6972	-23.7521
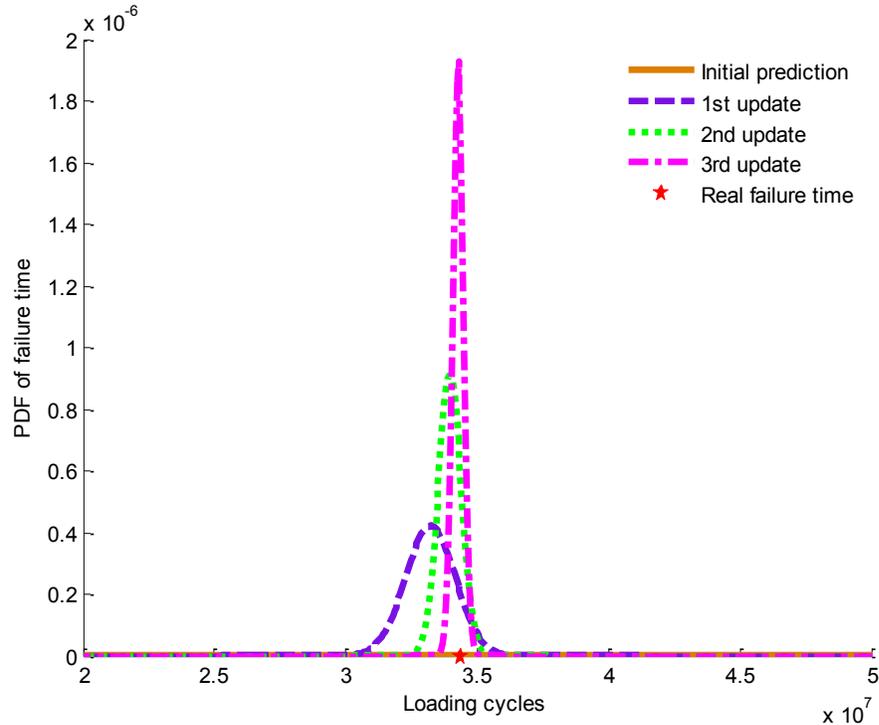
Figure 39. Updated failure time distributions for path #11.

6.5 Conclusions

An integrated prognostics method considering time-varying operating conditions is developed in this chapter, which integrates physical models and sensor data. By taking advantage of stress analysis in FE modeling, the degradation process governed by Paris' law can adjust itself immediately to respond to the changes of the operating conditions. The capability to directly relate the load to the damage propagation is a key advantage of

the proposed integrated prognostics approach over the existing data-driven methods for dealing with time-varying operating conditions.

In the proposed method, uncertainty in material parameters is considered responsible for randomness in the predicted failure time. The joint distribution of the material parameters is updated as the sensor data is available. The updated distributions better characterize the material parameters and reduce the uncertainty in them for the specific unit under monitoring. The update process is realized via Bayesian inference. To reduce the computational effort, the stochastic collocation method based on PCE is applied for uncertainty quantification. Examples are given to demonstrate the effectiveness of the proposed method, which is effective even when the current loading profile is different from the loading profile under which historical data were collected. Even though the gearbox is considered in this chapter, the proposed method is also applicable to various other components and structures subject to the similar fatigue loading profile, after the appropriate adjustment of physical models.

Chapter 7.　Gear integrated health prognosis considering crack initiation time uncertainty and random shock

7.1　Overview

A large grain of inherent uncertainty imposes major challenges in prognostic methods development. Uncertainty may arise from various sources, such as micro-structure of material, operating conditions, working environment, measurement as well as human factors. There are many research efforts being conducted on how to identify, capture and manage these multiple sources of uncertainty to make the RUL prediction more accurate, precise and reliable. The existing prognostic approaches usually start the prediction at an assumed time instant when a fault at certain severity is detected (Coppe et al. 2010; An et al. 2012; Orchard and Vachtsevanos, 2007). This is based on an assumption that the starting point of prognosis is accurate. However, due to the limitations of the fault detection and diagnostic technologies, there is a large variation in the accuracy of fault detection. This variation affects the prediction accuracy accordingly: an early starting point of prognostics will lead to underestimated RUL, and a late starting point will lead to overestimated RUL. Crack initiation time (CIT) determines the starting point of prognostics. Hence, it is needed to explicitly consider the uncertainty in CIT.

Among existing research on real time prognostic algorithms development, most of them investigated components/systems undergoing a gradual degradation with time. Such gradual degradation is governed by a dynamic model, either data driven or physics-of-failure based, which usually leads to a continuous degradation. Very few of them have

considered the degradation path containing shock, which appears as a sudden jump in the degradation. The shock will cause sudden damage increase and accelerate the degradation rate. In reliability engineering, researchers have investigated the ways of modeling shock process in system reliability analysis (Wang et al. 2011; Mallor and Santos 2003; Jiang et al. 2012). The purpose of these studies was either to investigate system reliability properties or maintenance strategies considering the shock effects. However in the present study, we look into the shock effects on real time RUL prediction. This purpose is also the main focus of prognostic algorithms development in CBM.

In the crack propagation problem, model parameters in Paris' law include material dependent coefficients, which should not be affected by external forcing factors, e.g., overload causing shock. Hence, another uncertainty source rather than material parameters is needed to account for the shock effect on the degradation prediction. Because sudden damage increment results in a discontinuity in the degradation path, the lifetime is shortened accordingly. Note that, if the slope of the degradation path is given as fixed, the degradation path with a shortened lifetime can be considered as equivalent to a gradual degradation path which has an earlier CIT. Therefore, the variation in CIT provides a degree of freedom in translational adjustment for the degradation path. With both slope and translational adjustments, the integrated prognostics method is expected to reduce the uncertainty in RUL prediction and to capture the shock effect on the degradation as well.

In this chapter, an integrated prognostics method is developed to account for both the uncertainty in CIT and the shock in gear degradation. We define CIT as the time instant when a crack is detected and the prognosis starts. In fact, when the CIT is adjusted, it is

the "intercept" with the time axis of the degradation path at the initial crack size that is adjusted, which is a different mechanism of adjustment. The combination of adjustment in both "slope" and "intercept" will better characterize the real degradation path given the crack observations. Section 7.2 presents the integrated prognostics framework which incorporates the uncertainty in CIT. In Section 7.3, the proposed method is elaborated in detail. Three cases are considered: gradual degradation considering both model parameter uncertainty and CIT uncertainty; shock degradation with known shock occurrence time; and shock degradation with unknown shock occurrence time. In Section 7.4, examples are given to show the effectiveness and the efficacy of the method. Section 7.5 concludes the work. The materials in this chapter have been documented in (Zhao et al. 2015b).

7.2   Integrated prognostics framework

An integrated prognostics framework is proposed in this chapter for gear RUL prediction considering CIT uncertainty. Two types of degradation are considered: gradual degradation and shock degradation.

The gradual degradation was described by Paris' law. The equation is shown in (7.1)

$$\begin{cases} \dfrac{da}{dN} = C(\Delta K(a))^m \\ \quad a(t_0) = a_0 \end{cases} \tag{7.1}$$

where t_0 is CIT which is treated as a random variable. The Paris' law is discretized using first-order Euler's rule as in (7.2) to obtain the crack size at inspection times.

$$\begin{cases} a(i\Delta N) = a((i-1)\Delta N) + (\Delta N)C[\Delta K((a(i-1)\Delta N))]^m \\ \quad a(t_0) = a_0 \end{cases}. \tag{7.2}$$

The iteration proceeds until the inspection time t_i is reached. The crack size obtained is denoted as $a_i(t_0, \xi)$.

The other type of degradation contains shock. The shock causes a sudden damage accumulation due to external impact, such as a transient overload. In the case of fracture, the phenomenon of shock is a sudden increase in crack size. The degradation containing shock is defined as the shock degradation.

The core idea of "integrated" prognostics method is to combine condition monitoring data (e.g., crack size observations) and physical models in a way that crack size observations can be utilized to adjust the physical model. After the adjustment, the physical model is expected to predict the RUL better. The adjustment is incurred when a new observation is available. The sequential adjustments form a series of updates, which are triggered at every inspection time. Figure 40 shows the data flow in the integrated prognostics method proposed in this chapter. The model update is achieved by updating the distributions of uncertain parameters through Bayesian inference. The posterior distribution of the uncertain parameters is applied in Paris' law to calculate RUL. Meanwhile, the posterior distribution is fed into the next iteration as the prior distribution. The first prior distribution in absence of any observations is obtained by regressing the existing historical degradation paths.

When the update process is executed, three cases are considered: Case 1) no shock occurs; Case 2) shock occurs at a known time; and Case 3) shock occurs at an unknown time. As shown in Figure 40, when the gradual degradation without shock is considered (Case 1), the posterior distribution is directly used as the prior distribution for the next iteration. However, when shock is considered (Case 2 and 3), an adjustment is added for

marginal posterior distribution of CIT before it is used as the prior distribution for the next iteration. The purpose of this adjustment is to make sure that a virtual gradual degradation can be identified equivalently by searching an appropriate CIT which is earlier than original CIT. The equivalence is defined in the sense of failure time. Accordingly, the likelihood function also needs modifications to eliminate the adverse effects brought out by the observations before shock occurs. The details of the three cases will be addressed in Section 7.3.
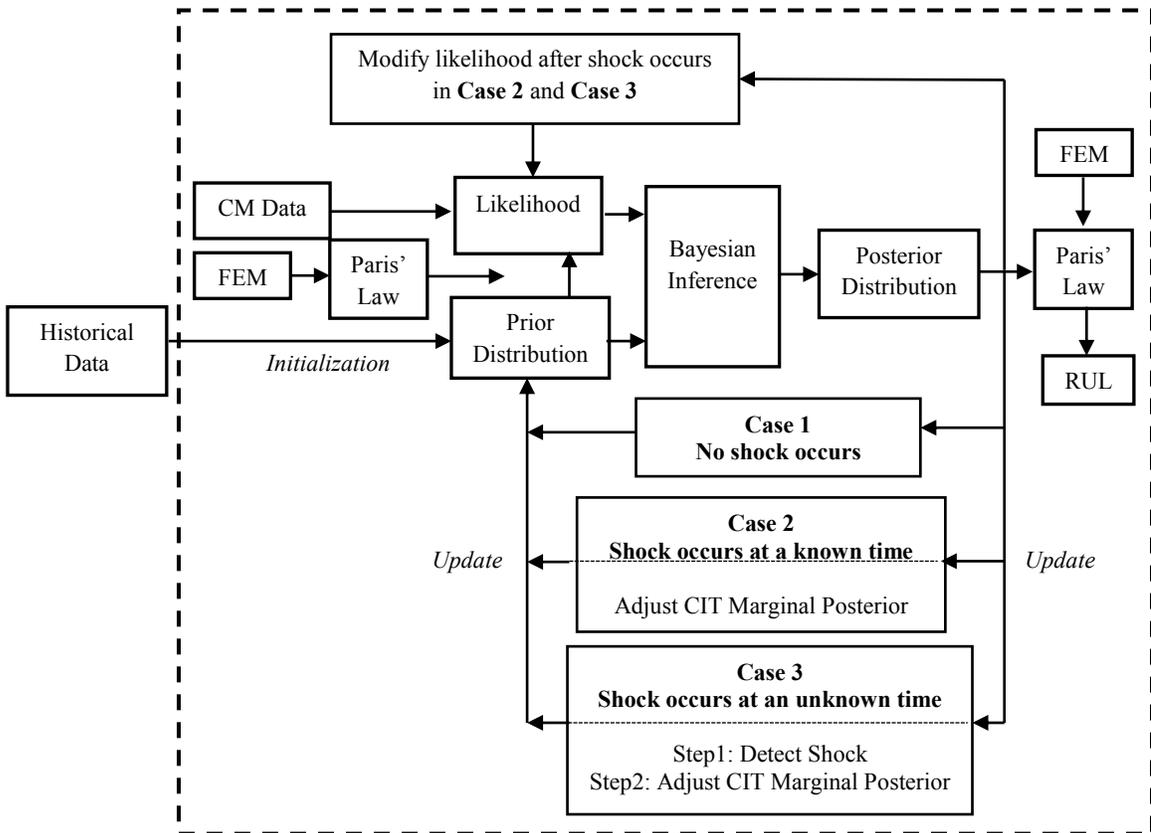
Figure 40. Data flow in the proposed integrated prognostics method

7.3 Integrated prognostics method considering CIT uncertainty and shock in the degradation

Damage degradation governed by dynamic equations is usually a continuous function with time regardless of measurement error. A discontinuity in the degradation path occurs due to sudden increase in damage size, i.e., shock, which may be caused by transient overload. Such increase in damage size will shorten the lifetime. The sudden change is due to external forces, rather than the uncertainty in material dependent model parameters. By noticing the equivalence between a degradation path with a shortened lifetime due to shock and a gradual degradation path with an earlier CIT, the uncertainty in CIT is considered to account for the shock effect. In this section, we firstly present the definition of uncertainty in CIT as well as the process to incorporate it in the Bayesian update process. After that, the integrated prognostics method considering shock degradation is proposed, which is realized by considering uncertainty in both material parameters and CIT. The distribution of RUL can be obtained using the updated uncertainty parameters.

7.3.1 Uncertainty in CIT

The well-known Paris' law describes the crack propagation rate using the principle of linear elastic fracture mechanics. It represents the crack propagation rate as a function of SIF and material dependent parameters m and C. Hence, given the applied loading, both m and C are factors that determine the "slope" of the degradation path in a scale of crack size versus time. The previous chapters focused on the adjustment of $\xi = (m, C)$, which represents the "slope" of the degradation path. The motivation to identify CIT as another source of uncertainty is that this uncertainty is difficult to account for by "slope". It is

actually a translational movement along the time axis. In another word, the CIT is the "intercept" with the time axis of the degradation path at the initial crack size. By adjusting the "slope" and the "intercept" simultaneously, it is expected to obtain an optimal approximation to the real degradation path.

Denoting the initial crack size as a_0 and the critical crack size as a_c , as shown in Figure 41, the degradation paths are generated by varying CIT and physical model parameters. As a result, the variation in the crack size at a certain inspection time t_i is contributed by uncertainty in both CIT and the physical model parameter. The path in the red dashed line represents the actual degradation path, with actual values for both the CIT and the physical model parameters. However, these actual values are not known exactly beforehand. The objective to use the Bayesian inference is to obtain their distributions by feeding the crack observations.
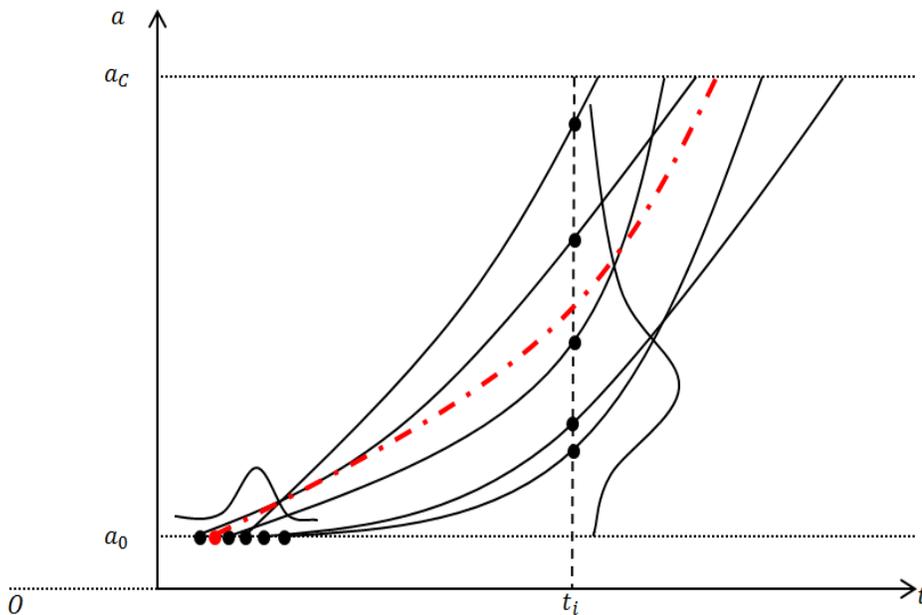
Figure 41. Degradation paths generated by varying CIT and physical model parameters.

Define a set of random variable, $\xi = (m, C)$, representing the uncertainty from model material parameters. As before, t_0 stands for the CIT. Here, t_0 is also considered as a random variable to account for the uncertainty in CIT. Suppose that several failure histories are available with the information on inspection times and associated crack sizes. Then a prior distribution $f_{prior}(t_0, \xi)$ can be obtained by regression and statistical fitting. If it is assumed that the crack measurement error follows zero-mean Gaussian distribution with σ as the standard deviation, at a certain inspection time t_u the likelihood to observe a crack size of $a_u^{obs} = a^{obs}(t_u)$ is

$$l(a_u^{obs}|t_0, \xi) = \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{\left(a_u^{obs} - a_u(t_0, \xi)\right)^2}{2\sigma^2}\right). \tag{7.3}$$

In the Bayesian inference framework, a posterior distribution $f_{post}(t_0, \xi)$ can be obtained by

$$f_{post}(t_0, \xi|a_u^{obs}) = \frac{l(a_u^{obs}|t_0, \xi)f_{prior}(t_0, \xi)}{\int l(a_u^{obs}|t_0, \xi)f_{prior}(t_0, \xi)dt_0d\xi}. \tag{7.4}$$

The update process is executed at the inspection time when new observation on crack size is available. The posterior distribution of the current update process will serve as the prior distribution for the next update process. To circumvent the intractable integration in the Bayesian formula, importance sampling technique is used to obtain samples which follow the posterior distribution. In order to simplify the problem and to emphasize the different effects of "slope" and "intercept", in the following discussions, only uncertainty in m and t_0 are considered while C is treated as a constant.

7.3.2 Physical model update considering shock in the degradation

Besides gradual degradation, shock degradation is another degradation type where the damage is accumulated suddenly leading to a discontinuity in the degradation path. In Figure 42, the blue line depicts a gradual degradation path with CIT t_{0_actual}, and the marks of circle on the path represent crack sizes at inspection times. Assume that a shock happens between the third and the fourth inspection times. The shock will result in a sudden crack increment, denoted by purple dotted line segment. After the shock, the gradual degradation continues at an accelerated rate. A shock degradation path is thus formed, which is depicted by a magenta line with square marks. Because of the shock, the lifetime is shortened with an extent denoted by a green dash line segment. The shock degradation is thus a discontinuous curve. Material dependent model parameters should be static because the material is not changing. It also explains why the two degradation paths are parallel after the shock occurs. Hence, a new uncertainty source should be uncovered to account for the change of the degradation path due to shock. By noticing that our interest is to predict the future performance instead of regressing over the past performance, we can assume a virtual crack growth history back propagating from the time when shock occurs. This virtual history is indicated in yellow dot-dash line in Figure 42. Therefore, the objective of predicting the degradation after the shock occurrence can be achieved by identifying a gradual degradation with a different CIT, to be more specific, an earlier one, denoted by $t_{0\_virtual}$. Based on the foregoing analysis, uncertainty in both the CIT and the material parameter will be considered to deal with shock degradation. We investigate two cases in the following discussions: shock occurrence time instant is known and shock occurrence time instant is unknown.
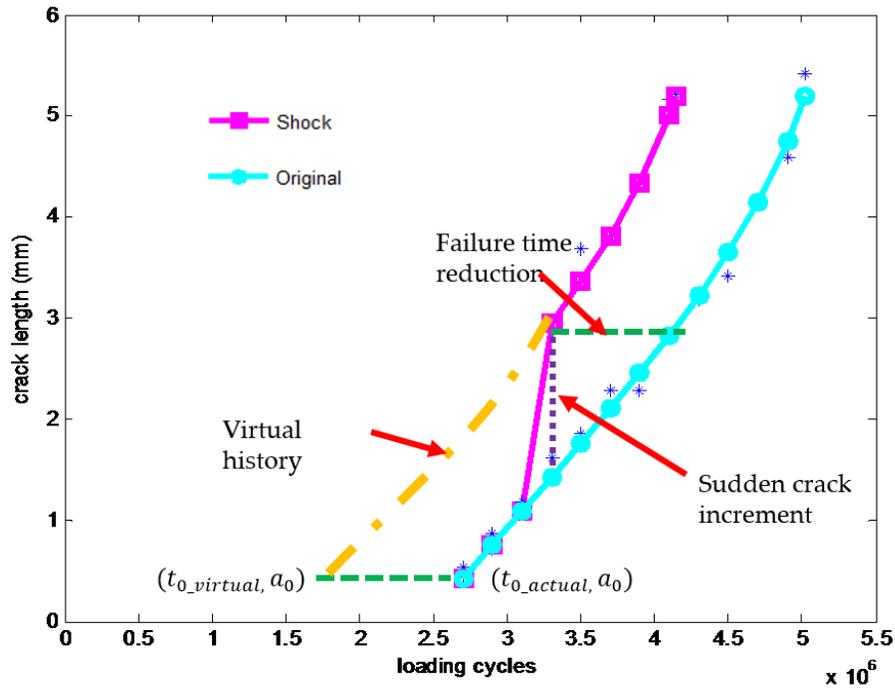
Figure 42. Shock degradation path for illustration purpose

7.3.2.1 Shock occurrence time is known

The update process is executed every time when observation on crack size is available. The posterior distribution in the last iteration will be used as the prior distribution in the next iteration. By assimilating the observations, the variance of uncertainty parameter is expected to be reduced and the mean to approach the actual value. The shrinkage of uncertainty leads to a narrower distribution, which appears to have a peak covering a very small range of parameter values. When we implement the Bayesian inference (7.4), importance sampling technique is utilized, where the density is represented by samples from the associated distribution. This discretization of the continuous function excludes the samples with low density, especially when sample size is limited. After several updates before the shock occurs, the prior distribution becomes

very narrow and its density has a peak near the actual CIT. As a consequence, if the virtual CIT is far from the actual CIT, there will be little chance for the values near $t_{0\_virtual}$ to be sampled. When the shock occurs, there is a sudden increase in crack size, which results in a noticeable difference between the predicted crack size and the actual observation. In order to attribute this difference to an earlier CIT, it needs a fair extent of adjustment of CIT in its posterior distribution. Hence, the coverage of prior distribution must be large enough to give moderate density for the virtual CIT $t_{0\_virtual}$, otherwise the samples could not carry useful information. Therefore, the posterior distribution before shock occurs is adjusted in a way that the variance of CIT marginal distribution is artificially increased to make sure the vicinity of the targeted virtual CIT $t_{0\_virtual}$ has moderate chance to be sampled.

Let the inspection times be a set $\{t_u: u = 1, 2, \cdots, U\}$. At these inspection times, the observed crack sizes form a set $\{a_u^{obs}: u = 1, 2, \cdots, U\}$ and the uncertain parameters are denoted by $\left\{\left(t_0^{(u)}, \xi^{(u)}\right): u = 1, 2, \cdots, U\right\}$. Assume that the measurement errors are independent among different inspection times; hence, the likelihood function to observe a crack growth history up to an inspection time t_s is

$$l_s\left(a_{1:s}^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right) = \prod_{u=1}^{s} l\left(a_u^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right), \tag{7.5}$$

where $l\left(a_u^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right)$ was defined in (7.3). The $a_{1:s}^{obs}$ refers to the set of observation history $\{a_u^{obs}: u = 1, 2, \cdots, s\}$ up to the inspection time t_s. The update process at t_s can be obtained accordingly by

$$f_{post}\left(t_0^{(s-1)}, \xi^{(s-1)} \big| a_{1:s}^{obs}\right)$$

$$= \frac{l_s\left(a_{1:s}^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right) f_{prior}\left(t_0^{(s-1)}, \xi^{(s-1)}\right)}{\int l_s\left(a_{1:s}^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right) f_{prior}\left(t_0^{(s-1)}, \xi^{(s-1)}\right) dt_0^{(s-1)} d\xi^{(s-1)}}. \tag{7.6}$$

The inspection times divide the whole lifetime into several disjoint intervals. If the degradation is gradually happening without shock, the update of uncertain parameters is to directly assign the posterior distribution of uncertain parameters at the current inspection time to be the prior distribution at the next inspection time. However, if the shock occurs at a known time instant during interval $[t_{v-1}, t_v)$, the posterior distribution will be adjusted before being assigned to the prior distribution for next update. As discussed earlier, the adjustment is to increase the variance of marginal distribution for CIT to cover $t_{0\_virtual}$. Denote the adjusted distribution as $\tilde{f}_{post}\left(t_0^{(v-1)}, \xi^{(v-1)}\right)$. In addition, the likelihood also needs modification. The actual observations before the shock occurs have adverse effects on the likelihood of observing the crack sizes after the shock. Hence, only the observations after the shock will be used to define the likelihood function in Bayesian formula for the updates executed after shock occurrence time. The update process considering the shock occurring at a known time instant thus is modified to be

$$l_s\left(a_{1:s}^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right) = \begin{cases} \prod_{u=1}^{s} l\left(a_u^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right), & \text{when } s < v \\ \prod_{u=v}^{s} l\left(a_u^{obs} \big| t_0^{(s-1)}, \xi^{(s-1)}\right), & \text{when } s \geq v \end{cases} \tag{7.7}$$

$$f_{prior}\left(t_0^{(s)}, \xi^{(s)}\right) = \begin{cases} f_{post}\left(t_0^{(s-1)}, \xi^{(s-1)}\right), & \text{when } s \neq v \\ \tilde{f}_{post}\left(t_0^{(s-1)}, \xi^{(s-1)}\right), & \text{when } s = v \end{cases}. \tag{7.8}$$

7.3.2.2 Shock occurrence time is unknown

If the shock occurs at an unknown time, we could adopt the same adjustment strategy presented in Section 7.3.2.1 for each update, to ensure that the virtual CIT will be covered within the distribution samples. However, this strategy will result in a large amount of uncertainty in RUL prediction, which provides little useful information in decision making. Hence, an additional step for shock detection is proposed to add into the update process to deal with the case when shock occurrence time is unknown.

Shock occurrence will cause sudden increase of the crack size. The amount of increase is assumed to be far out of the range of measurement error. A large adjustment of CIT from the last inspection time is expected. The average of the predicted crack size using the distribution of CIT at the last inspection time ought to deviate a lot from the observed crack size at the inspection time right after shock occurs. Therefore, a shock is said to be detected if the amount of such deviation exceeds a predefined threshold, δ. The criterion is that, shock occurs during $[t_{v-1}, t_v)$ if

$$\left| \mathbb{E}_{\left(t_0^{(v-1)}, \xi^{(v-1)}\right)}\left[a\left(t_v \middle| t_0^{(v-1)}, \xi^{(v-1)}\right)\right] - a_v^{obs} \right| > \delta. \tag{7.9}$$

The symbol of $\mathbb{E}$ denotes the operator for expectation. After identifying the shock occurrence time, update process (7.7) - (7.8) will apply.

7.3.3 RUL prediction

The ultimate goal of updating uncertain parameters is to predict the RUL of the cracked gear more accurately. With the updated parameter distribution obtained through Bayesian inference, the RUL prediction is given accordingly at the inspection time. Paris' law can be written in its reciprocal form as in (7.10),

$$\frac{dN}{da} = \frac{1}{C(\Delta K(a))^m}.$$ (7.10)

Let the current inspection cycle be $N_u(t_0)$ and the crack increment be Δa. The RUL is calculated by discretizing (7.10) in the following way,

$$\Delta N_i(\xi) = N_{i+1}(\xi) - N_i(\xi) = \Delta a \left[C\left(\Delta K(a_i)\right)^m \right]^{-1}.$$ (7.11)

The RUL is the summation $\sum_i \Delta N_i(\xi)$ from the current inspection cycle to the cycle when failure occurs. Therefore, the total failure time is expressed as $F(t_0, \xi) = N_u(t_0) + \sum_i \Delta N_i(\xi|t_0)$. This expression shows that the uncertainty in total failure time prediction is determined by the uncertainty in both the CIT and the physical model parameters, (t_0, ξ). Thus more accurate values of (t_0, ξ) will produce more accurate $F(t_0, \xi)$. An update of $F(t_0, \xi)$ will be triggered after the update of (t_0, ξ) so as to adjust the failure time prediction.

7.4   Examples

Three cases are discussed in Section 7.3 in which case 3 can be considered as a generalization of case 2 with an additional step for shock detection. Hence, it is sufficient to only verify case 3 for case 2 to be valid. In this section, we will show the effectiveness of the proposed method for case 1 and 3: gradual degradation considering CIT and degradation with shock occurring at an unknown time.

7.4.1 Case 1: Gradual degradation considering CIT

In this case, the prior distributions of the uncertain parameters are assumed as: $m \sim N(1.4, 0.2)$, $t_0 \sim N(2e6, 2e5)$. The measurement error is $e \sim N(0, 0.15^2)$, The values

of the constants are set to be: $C = 9.12\text{e} - 11$, $a_0 = 0.1\text{mm}$, $a_c = 5.2\text{mm}$. The history

of SIF is adopted from the example in Chapter 4, in which the input torque is 320 Nm

and the effect of dynamic load is considered. The characteristics of the two actual

degradation paths are shown in Table 18. These two test degradation paths share the same

value of real m ($m = 1.6$) but the different crack initiation times.

Table 18. Characteristics of two actual degradation paths #1 and #2

Path #	m	t_0 (cycles)	Failure time (cycles)	Inspection interval (cycles)
1	1.6	1.7e6	4,215,500	3e5
2	1.6	2.3e6	4,815,500	5e5

The crack observations and updating results for the two paths are listed in Table 19

and Table 20, respectively. The updated PDFs for distributions of t_0 and m are displayed

in Figure 43 and Figure 44 for path #1, and in Figure 45 and Figure 46 for path #2.

Table 19. Update process for path #1

Update #	Inspection time (cycles)	Crack size (mm)	Mean of m	Mean of t_0 (cycles)	Mean of $F(t_0, m)$ (cycles)
Prior			1.4	2e6	17,159,000
1	3,015,700	2.1900	1.6277	1.9572e6	
2	3,315,700	3.0425	1.6420	2.0105e6	
3	3,615,700	3.5473	1.6275	1.9325e6	
4	3,915,700	4.2072	1.6134	1.8354e6	3,989,700
Actual value			1.6	1.7e6	4,215,500

The updating results show that, as more crack observations are fed into Bayesian

inference, the means of m and t_0 approach progressively to their actual values.

147

Meanwhile, the standard deviation is decreasing, which shrinks the distribution shape. This is beneficial for improving the performance of prognostic algorithm since narrower distribution indicates reduced uncertainty, and it is useful in making more accurate and cost-effective maintenance decisions. Based on the updated uncertain parameters, the mean of the failure time approaches its actual value accordingly. The actual values are denoted using star marks in the Figures 43-46.
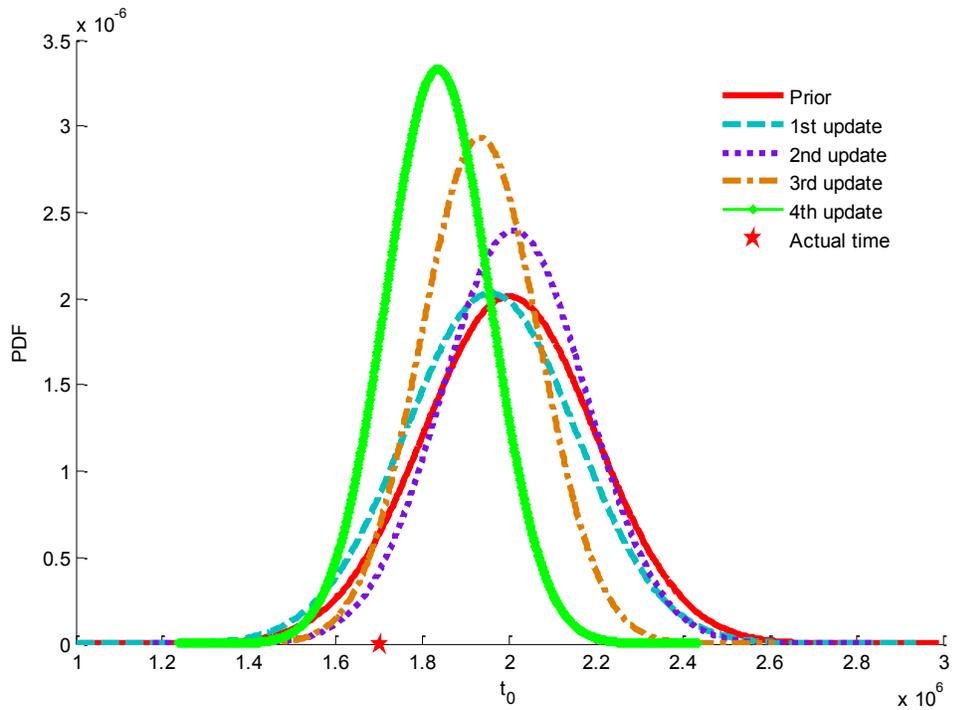
Figure 43. Updated distributions of t_0 for path #1
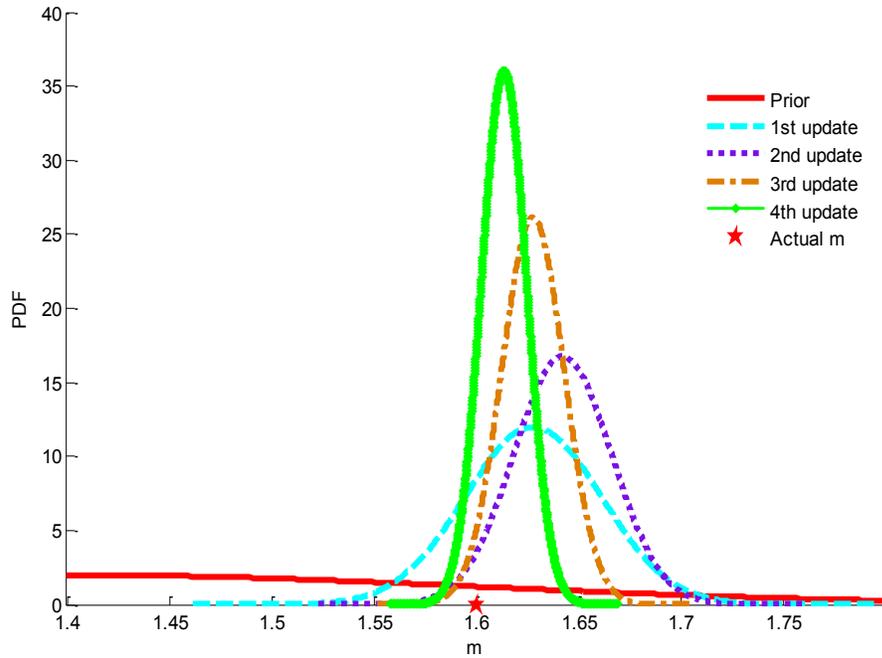
Figure 44. Updated distributions of m for path #1

Table 20. Update process for path #2

Update #	Inspection time(cycles)	Crack size (mm)	Mean of m	Mean of t_0(cycles)	Mean of $F(t_0, m)$ (cycles)
Prior			1.4	2e6	17,159,000
1	3215700	1.5910	1.5469	1.9827e6	
2	3715700	2.3063	1.5545	1.9999e6	
3	4215700	3.4307	1.5821	2.1583e6	
4	4715700	5.0218	1.6073	2.3604e6	5025200
Actual value			1.6	2.3e6	4,815,500
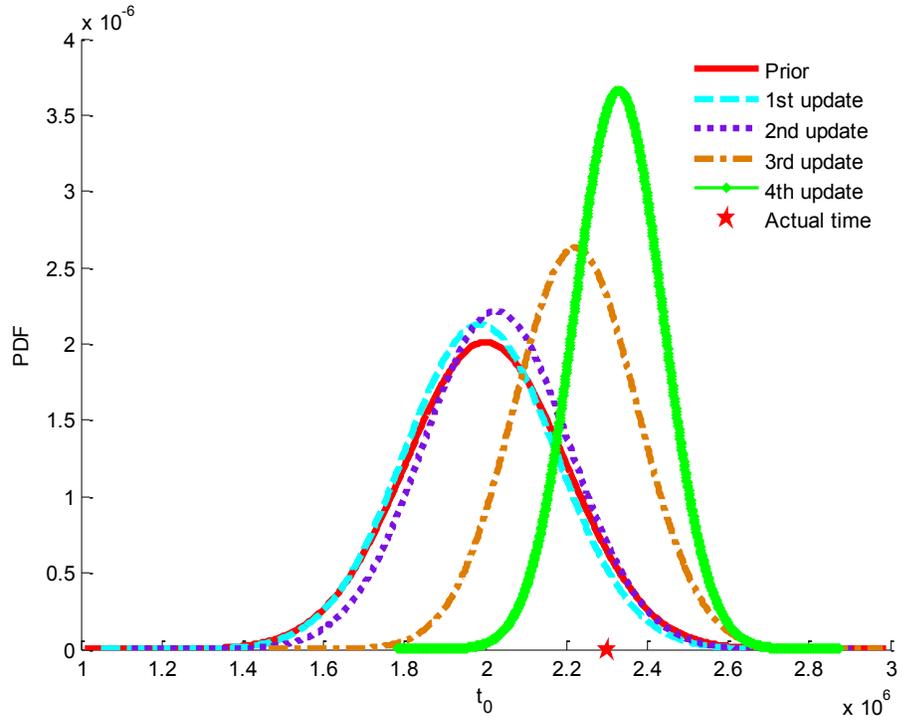
Figure 45. Updated distributions of t_0 for path #2
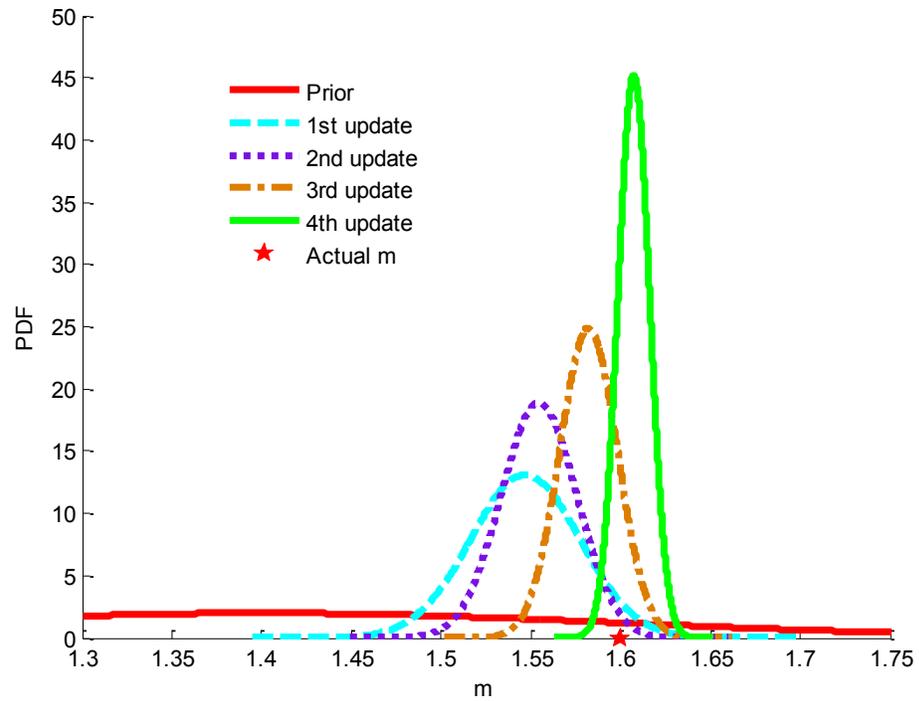
Figure 46. Updated distributions of m for path #2

7.4.2 Case 2 & 3: Degradation with shock

In this case, the prior distributions of the uncertain parameters are given as: $m \sim N(1.6, 0.2)$, $t_0 \sim N(6.5e7, 0.5e7)$. The measurement error is $e \sim N(0, 0.15^2)$ while the values of the constants are set to be $C = 9.12e - 11$, $a_0 = 0.1$mm, $a_c = 5.2$mm. The computed SIF values versus crack size are shown in Figure 47, for which the input torque is 40 Nm and the effect of dynamic load is not considered. One actual degradation path, path #3, is used to demonstrate this case, and the characteristics of the degradation path are shown in Table 21. The actual degradation path #3 is depicted in Figure 48. The marks of circle represent the actual crack sizes, which are unknown. The marks of star represent the observed crack sizes, which deviate from the corresponding actual values due to measurement error. Shock occurs between observation #3 and #4, and a sudden increase in crack size can be seen in Figure 48.
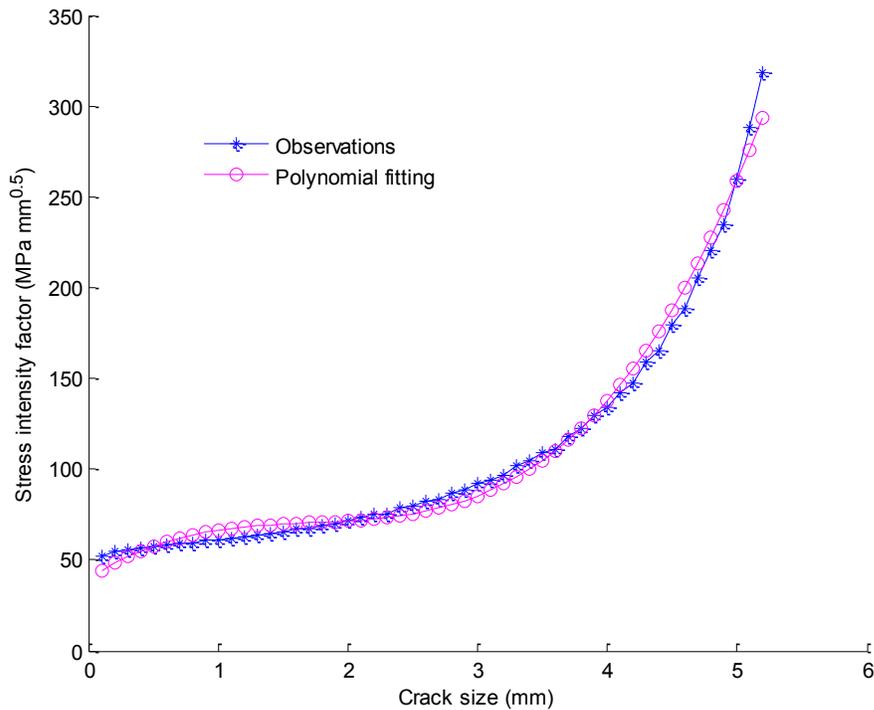
Figure 47. History of stress intensity factor for path #3

Table 21. Characteristics of actual degradation path #3

Path #	m	t_{0_actual} (cycles)	$t_{0\_virtual}$ (cycles)	Failure time (cycles)	Inspection interval(cycles)
3	1.4354	7.5e7	4.9e7	1.474e8	1e7
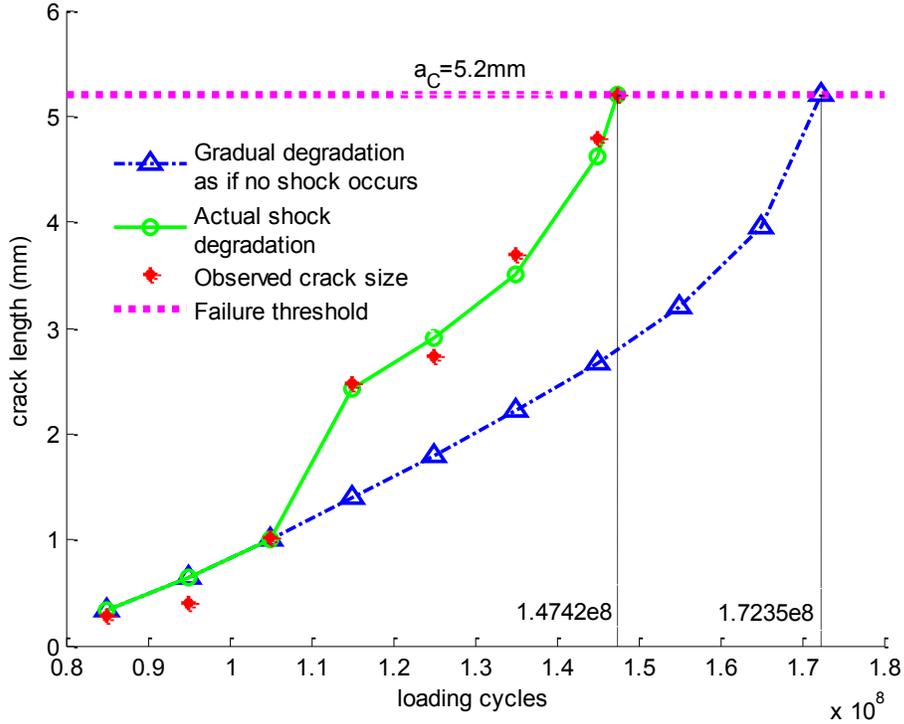
Figure 48. Actual shock degradation path #3

Table 22. Update process for path #3

Update #	Inspection time(cycles)	Crack size (mm)	Mean of m	Mean of t_0 (cycles)	Mean of $F(t_0, m)$(cycles)
Prior			1.6	6.5e7	1.331e8
1	8.5e7	0.2775	1.3006	6.7311e7	2.639e8
2	9.5e7	0.3968	1.2382	6.8447e7	3.126e8
3	10.5e7	1.0214	1.3204	6.8454e7	2.183e8
4	11.5e7	2.4759	1.4303	4.7962e7	1.480e8
5	12.5e7	2.7370	1.4233	4.7805e7	1.524e8
6	13.5e7	3.6924	1.4301	4.7735e7	1.454e8
7	14.5e7	4.7976	1.4334	4.8387e7	1.465e8
Actual value			1.4354	5.0060e7	1.474e8

The shock could be detected between update #3 and #4 using the criterion proposed in Section 7.3.2. The horizontal boarder between these two updates points in Table 22 separates the update process into two phases: history before shock occurrence and history after shock occurrence. From Table 22, it is observed a sudden transition of changing pattern for mean of t_0 after the shock is detected. Before the shock occurs, the mean of t_0 tends to increase to reach $t_{0\_actual} = 7.5e7$ cycles, which is the actual CIT of the shock degradation path. Accordingly, the mean of failure time $F(t_0, m)$ tends to approach 1.7235e8 cycles, which is the failure time of a gradual degradation if no shock occurs, as depicted by dot-dash line with marks of triangle in Figure 48. However, after the shock occurs, the mean of CIT reverses the changing direction to approach $t_{0\_virtual} = 5.006e7$ cycles. Compared to the actual shock degradation, this virtual gradual degradation path compensates the reduction of failure time due to shock occurrence by identifying an earlier CIT. As a material dependent parameter, there is no such transition in m because its value is dependent on material, which should not be changed by shock occurrence. Based on the accurate information on parameters t_0 and m, the predicted mean of failure time $F(t_0, m)$ successfully approach the actual failure time in a shock degradation, which is 1.4742e8, as depicted by solid line with marks of circle in Figure 48.

Figure 49 shows the updates of PDF for m, in which the bold line indicates the updates after the shock occurs. It can be seen that, as more observations are available, the mean of m approach the actual value. Also the variance of PDF is reduced progressively which provides more precise information. Figure 50 shows the updates of PDF for t_0, in which two neighbourhoods are apparent to observe. One is in the vicinity of $t_{0\_actual} = 7.5e7$, denoted using the mark of diamond, and the other is in the vicinity of $t_{0\_virtual} =$

5.006$e7$ denoted using the mark of star. As discussed before, this phenomenon explains the adjustment of t_0 due to the shock occurrence, and the mean of t_0 approaches $t_{0\_virtual}$ after the shock occurs.
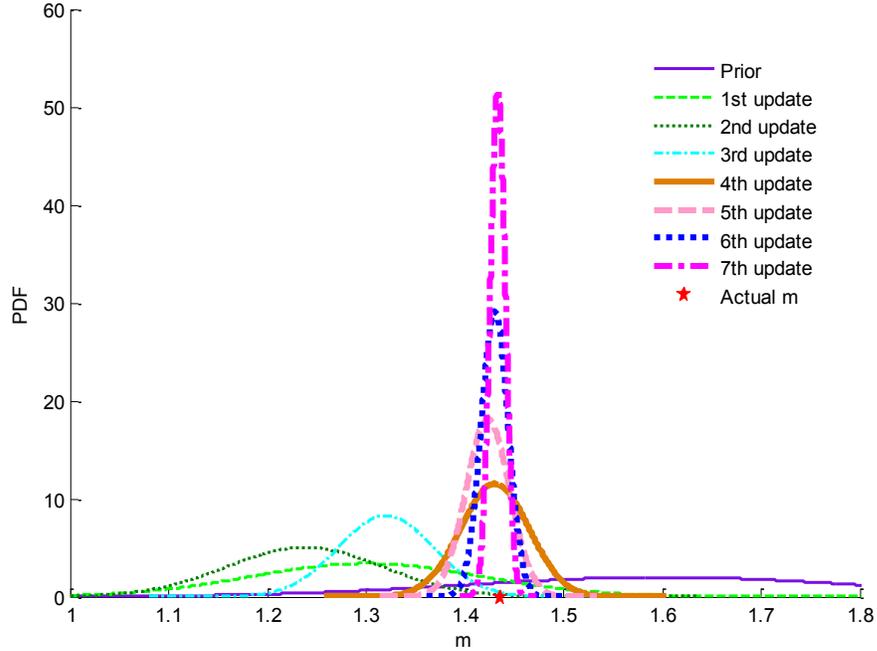
Figure 49. Updated distributions of m for path #3
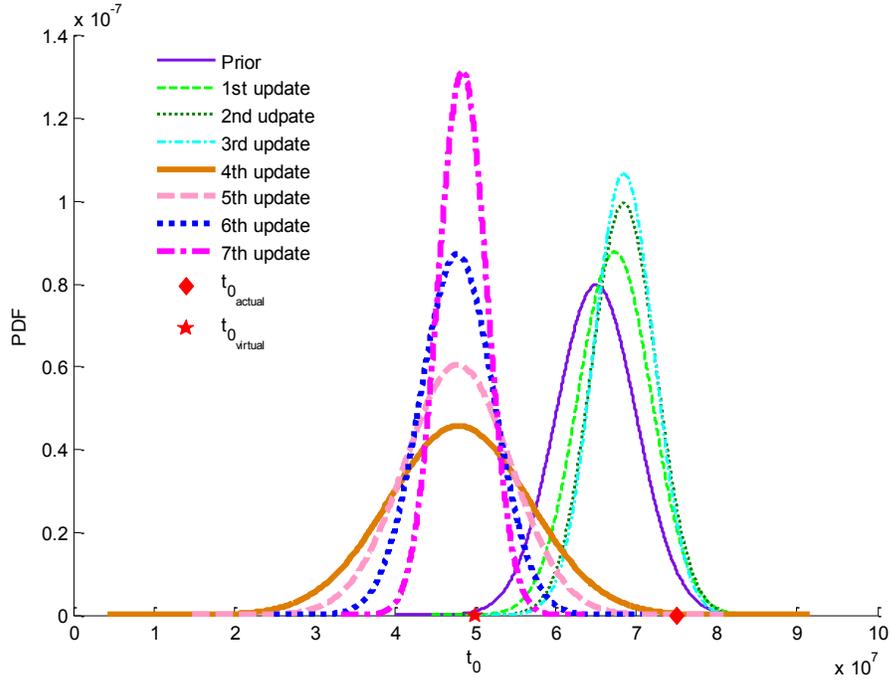
Figure 50. Updated distributions of t_0 for path #3

7.5   Conclusions

Prognostics tools are used for RUL prediction starting from the CIT. However there is uncertainty in CIT due to the limited capability of existing fault detection tools, and such uncertainty has not been explicitly considered in the literature for integrated prognosis. A shock causes a sudden damage increase and creates a jump in the degradation path, which shortens the total lifetime, and it has not been considered before either in the integrated prognostics framework. This chapter proposes an integrated prognostics method considering these two important factors, the uncertainty in CIT and the shock in gear degradation. In the proposed integrated prognostics method, CIT is considered as an uncertain parameter, which is updated using condition monitoring data. To deal with the sudden damage increase and the reduction of lifetime, a virtual gradual degradation path with an earlier CIT is utilized, and the effect of shock is captured through identifying an appropriate CIT. Examples demonstrate the effectiveness of the proposed method in predicting RUL considering shock and uncertainty in CIT.

Chapter 8.   Experimental validation

8.1   Overview

In the several preceding chapters, the aspects of an integrated prognostics method were researched for gear life prediction: the framework, the uncertainty quantification, considering time-varying operating conditions, and considering crack initiation time and shock in the degradation. The failure mode investigated in the preceding chapters was the tooth fracture which is caused by cyclic bending stress at the tooth root. In addition to the tooth fracture, gears also suffer from another failure mode of surface wear due to sliding contact in the mesh process. Researchers have proposed various wear models which are dedicated to forecasting the wear progression with time. Among them, the Archard's wear model (Archard 1953) is now generally accepted as a suitable framework within which the quantitative analysis on wear progression can be discussed (Peterson and Winer 1980). The wear coefficient k in Archard's wear model is an important parameter that determines the wear rate. There exists a large amount of variation in the value of wear coefficient among different units due to contacting material property. Flodin and Andersson (1997) used Archard's model to predict the wear in spur gears. The sliding distance was calculated based on the involute profile of spur gear tooth. The contact pressure was obtained by a Winkler surface model. When calculating the load on gear tooth, the mesh stiffness was assumed to be constant for simplicity. The wear coefficient k was selected as a parameter with constant value. There was no experimental result reported in that paper to validate the method. A wear prediction methodology was proposed in (Bajpai et al. 2004) for parallel axis gear pairs, which was also based on

Archard's model. Commercial software was employed to calculate the contact pressure during the mesh process. Thanks to the FE model and the technique of surface integral formulation, the method was able to calculate the contact pressure and the sliding distance with accounting for the geometry change due to wear and tooth manufacturing imperfections. In the experimental validation, four tests were conducted. The purpose of the first experiment was to determine the wear coefficient k. Then the predictions were conducted using the same k obtained in the first experiment and were compared to the other three tests because the authors assumed that all the test specimens were manufactured with the same material, heat treatment, and manufacturing process, and also assumed the same test conditions. These assumptions exclude the variations in wear coefficient among different components and impose risks of applying this wear coefficient to other components. This chapter researches the integrated prognostics method for the failure mode of surface wear and presents the experimental validation using a run-to-failure test on a planetary gearbox. We consider the wear coefficient as a random variable to account for its variation in different components. However, for a specific component, the uncertainty of the wear coefficient is much less than that of the population. Therefore, we propose to use inspection data during the wear process to reduce the uncertainty in wear coefficient for the specific component. Because the wear removes material of contacting surface, the mass loss would be an effective indicator for the wear status. The mass loss data will be integrated into the Bayesian inference to update the prior distribution of the wear coefficient into its posterior distribution. The posterior distribution is expected to better characterize the specific wear process as a result of the data assimilation. The organization of this chapter is as follows. In Section

8.2, the wear model for spur gear is established. Section 8.3 presents the integrated prognostics method proposed for the failure mode of surface wear. Section 8.4 presents the run-to-failure experiment on a planetary gearbox. The experimental validation is given in Section 8.5. Section 8.6 concludes the work.

8.2   Wear model for spur gear

During the mesh process of spur gears, there is a relative movement between the two meshing teeth because of the difference in tangential velocity, except for the pitch point. The sliding movement causes wear of tooth surface. The material particles will be removed due to the surface wear and the tooth profile will be altered. As discussed in Section 2.3.2, the Archard's wear model (2.10) describes the wear rate.

$$\frac{dh}{ds} = kp$$

We discretize the model into the form in (6.1)

$$h(i) = h(i-1) + kps \tag{6.1}$$

where h is the wear depth, s is the sliding distance, p is contact pressure and k is the wear coefficient.

In order to use Archard's wear model to predict the gear tooth surface wear, two important quantities are needed: the contact pressure and the sliding distance of all the points on the tooth flank during the mesh process. In order to analyze the two quantities for all the points on the tooth flank, a coordinate system is established first. In this coordinate system, the origin is located at the pitch point; y axis is along the line of action; x axis is perpendicular to the y axis. For spur gear with involute tooth profile, the

contact points are moving along the y axis. As shown in Figure 51, P is the pitch point; B_1 and B_2 are the tangential points of base cycles; the tooth comes into mesh at point Q_2 and departs the engagement at point Q_1; the two points Q_2' and Q_1' are the transition positions for the two types of contact: single-pair-contact and double-pair-contact. The load carried by this tooth is plotted above the y axis without scale.
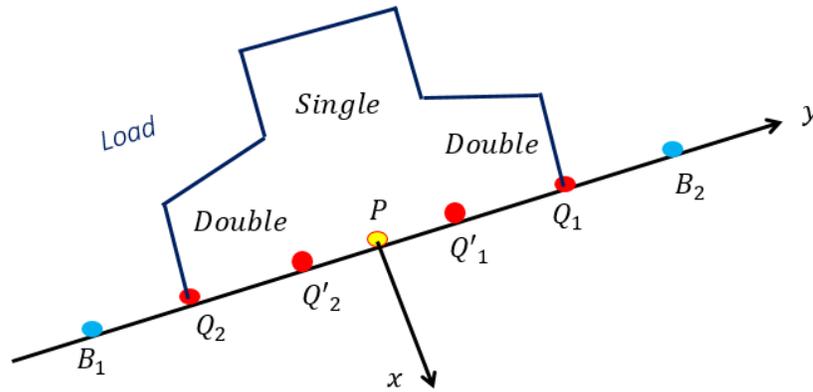
Figure 51. The coordinate system and the load distribution

When the meshing gears are loaded together by normal load F, the surface will deform to form the area of contact, as shown in Figure 52. The contact can be regarded as the cylinder on cylinder line contact. Hertz contact theory will be applied to calculate the contact pressure for each point on the tooth flank when it comes into mesh. Suppose one point on the tooth flank starts the engagement at point N which has the coordinate of $(0, y)$. As shown in Figure 52, the contact width is $2a_H$. The radiuses of curvature at the contact point for the pinion and the gear are R_1 and R_2 respectively.

$$R_1 = R_{b1} \tan \alpha_0 + y, \tag{6.2}$$

$$R_2 = R_{b1} \tan \alpha_0 - y, \tag{6.3}$$

Then, the effective radius of curvature R^* is defined as

$$\frac{1}{R^*} = \frac{1}{R_1} + \frac{1}{R_2}.$$ (6.4)

In a similar manner, the effective modulus of elasticity E^* is defined as

$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2}.$$ (6.5)
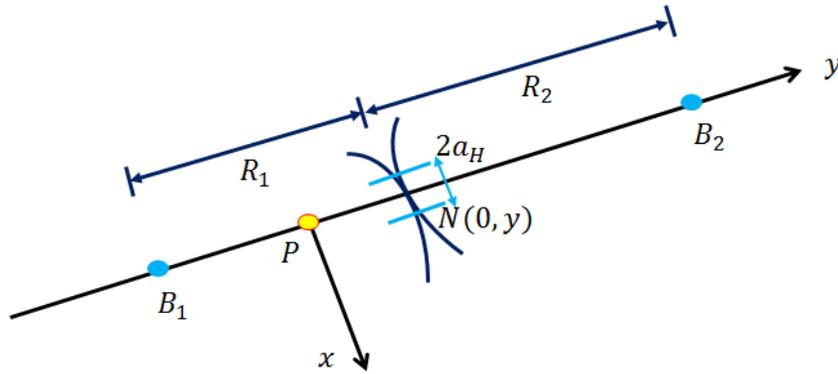
Figure 52. Contact of a pair of spur gear teeth

According to Hertz contact theory, the half contact width is

$$a_H = \sqrt{\frac{4FR^*}{\pi E^*}}$$ (6.6)

and the mean contact pressure is

$$\bar{P}_N = \frac{4F}{3\pi a_H}.$$ (6.7)

After obtaining the contact pressure, another quantity that is needed in Archard's wear model is the sliding distance. Andersson (1975) gave an analytical formula for sliding distance s_1 of the point $N\,(0, y)$ on pinion flank as follows.

$$s_1 = a_H - \sqrt{(R_2)^2 - \left(R_{p2} \sin \alpha_0 - y_{1d}\right)} + R_{p2} \cos \alpha_0 \qquad (6.8)$$

$$R_2 = \sqrt{(R_{p2} \cos \alpha_0 - a_H)^2 + \left(R_{p2} \sin \alpha_0 - y_{1e}\right)^2} \qquad (6.9)$$

$$y_{1e} = \sqrt{R_1 - (R_{p1} \cos \alpha_0 + a_H)^2} - R_{p1} \sin \alpha_0 \qquad (6.10)$$

$$y_{1d} = \sqrt{R_1 - (R_{p1} \cos \alpha_0 - a_H)^2} - R_{p1} \sin \alpha_0 \qquad (6.11)$$

$$R_1 = \sqrt{(R_{p1} \cos \alpha_0)^2 + \left(R_{p2} \sin \alpha_0 - y\right)^2} \qquad (6.12)$$

With the contact pressure and the sliding distance available, the determination of wear coefficient is required before using Archard's model to predict the wear depth evolution.

8.3 Wear model update through Bayesian inference

In this chapter, Archard's wear model is used to predict the wear depth evolution at each point on the tooth surface. More accurate wear coefficient in the model leads to more accurate wear prediction. However, different gears most likely have different wear evolution processes due to unavoidable variations in material property, manufacturing process and working conditions. We therefore consider the wear coefficient as a random variable to account for the uncertainty in wear evolution process from the population point of view. Meanwhile, the health condition of an individual gear is of our interest. The uncertainty in the failure time of a specific gear is much less than that in the population. Hence, a mechanism of uncertainty reduction is needed in the wear prediction. By noticing the material removal as a direct consequence of gear wear process, the gear

mass loss would be a good indicator of wear status. The Bayesian inference will take the data on gear mass loss as observations to update the distribution of wear coefficient. The formula is given in (6.13)

$$f_{post}(k|ml) = \frac{l(ml|k)f_{prior}(k)}{\int l(ml|k)f_{prior}(k)\,dk}.$$

(6.13)

where k is the wear coefficient, and ml stands for the mass loss. The update on wear coefficient distribution is conducted at each inspection time when a new measurement on gear weight is available. The posterior distribution will serve as the prior distribution for the next update at the next inspection time.

At each inspection time T_j, the measured mass loss is ml_j^{obs}. With the wear coefficient k^{j-1} obtained in the last inspection time T_{j-1}, the predicted mass loss at inspection times T_1 up to T_j are thus denoted as $ml_{1:j}^{mod}$, which are computed with the predicted wear depth and material density. We define the measurement error as $e = ml^{obs} - ml^{mod}$ and assume it follows zero-mean Gaussian distribution with standard deviation σ. It is further assumed that the measurement errors at different inspection times are i.i.d., therefore the likelihood to observe the mass loss at inspection times up to T_j is

$$l\left(ml_{1:j}^{obs}|k^{j-1}\right) = \prod_{i=1}^{j} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(ml_i^{obs} - ml_i^{mod})^2}{2\sigma^2}\right)$$

(6.14)

where $ml_{1:j}^{obs}$ stands for the mass loss measured at inspection times T_1 up to T_j.

8.4 A run-to-failure experiment on a planetary gearbox

Planetary gearbox is extensively used in helicopter, wind turbine and other engineering systems which require high power transmission and good torque ratio in a

small volume. There are three types of gears in a planetary gearbox: sun gear, planet gear and ring gear. By holding one type of gear stationary and letting the other two types rotate around it, the planetary gearbox can have different configurations. The configuration to be investigated in this experiment is to keep ring gear stationary and to provide power to the sun gear. All the gears are spur type.

A run-to-failure experiment was conducted on a planetary gearbox test rig. In this experiment, data collected include vibration signals, current signals of the drive/load motor, encoder signal and torque signal. In addition, the metal particle counter data and the weight loss of gears were also recorded. The physical parameters of this planetary gear set are listed in Table 23. More details on this experiment can be found in the technical report by Do et al. (2010).

Table 23. The physical parameters of the planetary gear set (Do et al. 2010)

Parameters	Sun gear	Planet gear	Ring gear
Number of teeth	19	31	81
Module (mm)	3.2	3.2	3.2
Pressure angle	20°	20°	20°
Mass (kg)	0.7	1.822	5.982
Face width (m)	0.0381	0.0381	0.0381
Young's modulus (Pa)	2.068×10^{11}	2.068×10^{11}	2.068×10^{11}
Poisson's ratio	0.3	0.3	0.3
Base circle radius (mm)	28.3	46.2	120.8

Because the surface wear removes material of the gear and causes the weight loss, the recorded history of weight loss will be a good indicator for wear propagation. Hence, the weight loss is used as observations to adjust the wear model in the integrated prognostics method. Among all the gears, sun gear experiences the most severe wear. Therefore, the wear prediction of the sun gear is of interest.

Before the gear was considered failed, the planetary gearbox experienced 19 runs. The photographs in Figure 53 show the tooth profile changes of the sun gear during this experiment. In these photographs, we can see the gear tooth becomes thinner gradually as runs continue and much material was removed. At the end of run #19, the thickness of the teeth was reduced to about 50% of the original thickness. According to the tooth profile changes and the metal particle data, this run-to-failure experiment can be divided into three stages:

- Runs 1-6: Normal operations to damage initiation.

- Runs 7-11: Initial to severe damage progression.

- Runs 12-19: Severe damage progression and profile change.

The weight loss of the sun gear is plotted in Figure 54. During these 19 runs, the input torque was kept to a constant level starting from run #5. Also considering the complicated burn-in period and the relatively large measurement error in the early stage, the data on the weight loss during run #1 to run #4 are not used in the validation.
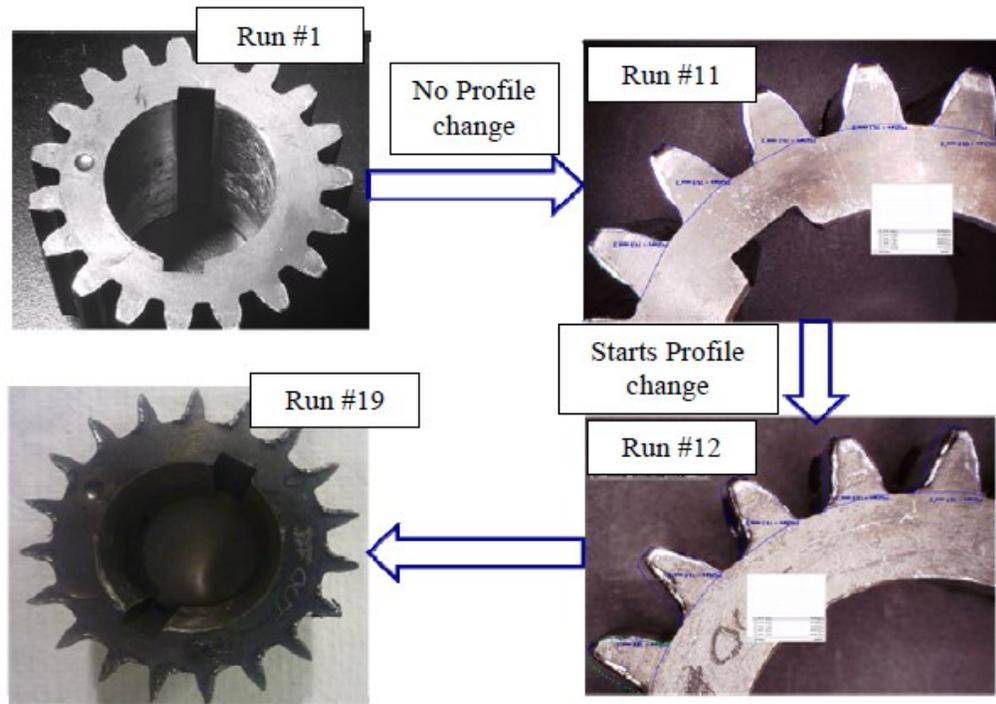
Figure 53. Tooth profile changes of the sun gear (Do et al. 2010)

Figure 54. Measured mass loss of the sun gear

8.5   Validation of the integrated prognostics method

In the wear model (6.1), the wear coefficient will determine the wear rate. The purpose of the integrated prognostics method is to use observations of mass loss of sun gear to update the wear coefficient in the wear model so that the wear process for this specific gear can be predicted more accurately. The Bayesian formula can be modified into (6.13). Similar to the failure mode of tooth fracture, the update on wear coefficient distribution is conducted at each inspection time.

The contact pressure and the sliding distance for each point on the sun gear tooth flank are calculated based on the methods given in Section 8.2.1. The contact pressure and the sliding distance for each point on tooth surface are assumed to be constant along the wear progression. By noticing the fact that the history of mass loss roughly has a linear trend from run #5 to run #14 as shown in Figure 54, the assumption of constant pressure and sliding distance being constant during the wear process is acceptable for these runs. During the late stages after run #14, we believe that this assumption is not valid anymore because the wear rate is accelerating and the predicted mass loss will be less than the actual mass loss. Therefore, for the purpose of validating the integrated prognostics method, the mass loss records from run #5 to run #14 are used to update the wear coefficient.

In addition, the gear mesh process is actually a combination of rolling and sliding contact. Besides surface wear due to the sliding contact, the material removal is also contributed by the rolling contact fatigue. During the experiment, pitting was observed on the gear tooth which was caused by rolling contact. To account for the effect of pitting on

the mass loss, we select the maximum wear depth as the wear depth for all the points. The reason is that, in the photograph of tooth profile, the wear volume change is roughly uniform for all the points on the tooth flank. Furthermore, the photograph also suggests similar wear condition of all the teeth of the sun gear. Therefore, we additionally assume that all the teeth of sun gear experience the same amount of wear. According to the configuration of the planetary gearbox, each tooth on the sun gear will mesh 3.2 times with all the four planet gears during one cycle of rotation. Hence, if we denote the wear depth increment on the sun gear tooth during one mesh period as h_s and the mass of material removed as $m(h_s)$, then the total mass loss of the sun gear within one cycle of rotation will be $m(h_s) \times 3.2 \times 19$ (# of teeth of sun gear).

For a given value of wear coefficient, we can use discretized Archard's model (6.1) to predict the wear depth. For example, when $k = 1.27 \times 10^{-15} \mathrm{Pa}^{-1}$, Figure 55 displays the predicted wear depths from run #5 to run #14 only considering sliding wear effect. From the figure, we can see that the maximum wear depth occurs near the root area.
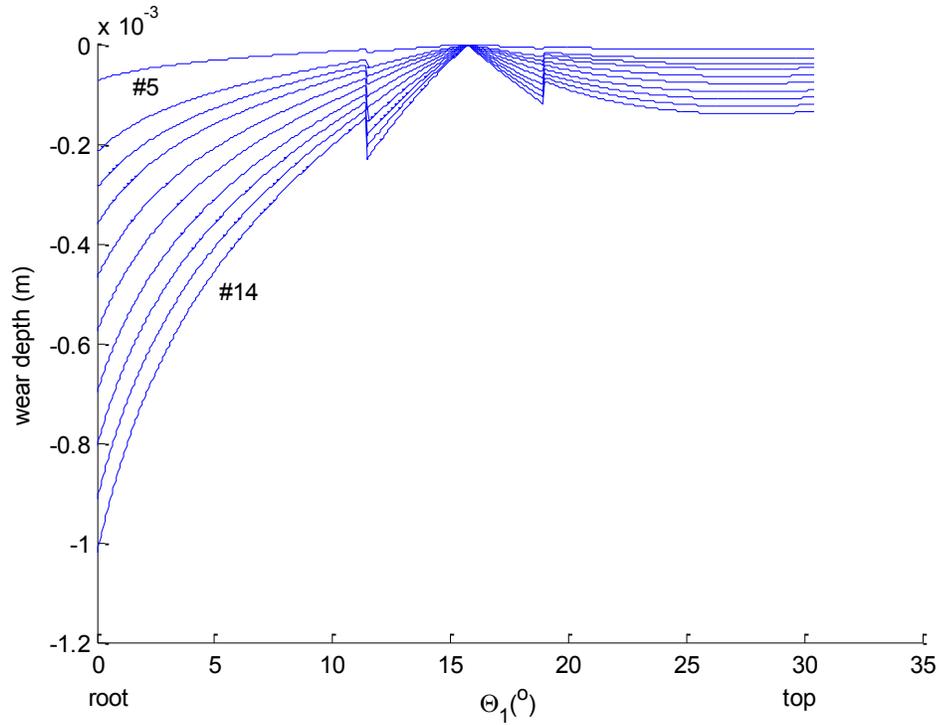
Figure 55. Predicted wear depths of the sun gear

Next, the update process for wear coefficient is presented.

The prior distribution of wear coefficient is assumed to follow normal distribution $N(3 \times 10^{-15}, (1 \times 10^{-15})^2)$, and the measurement error follows another normal distribution $N(0, 0.5^2)$. By selecting the maximum wear depth as the wear depth for all the points, the predicted mass loss is calculated at every inspection time. By applying Bayesian update formula in (6.13) and (6.14), the update process is tabulated in Table 24. Several selected distribution updates are shown in Figure 56.

Table 24. Update process for wear coefficient k

Run #	Measured mass loss	Mean of k	Std of k
	0	3e-15	1e-15
5	1.4	1.0552e-15	2.9315e-16
6	4.45	0.9288e-15	9.2107e-17
7	8.12	1.0664e-15	4.9948e-17
8	10.9	1.1636e-15	3.2156e-17
9	13.81	1.2186e-15	2.2289e-17
10	17	1.2513e-15	1.6210e-17
11	20.13	1.2672e-15	1.2228e-17
12	26.04	1.2748e-15	9.5061e-18
13	28.71	1.2776e-15	7.5751e-18
14	34.53	1.2770e-15	6.1681e-18
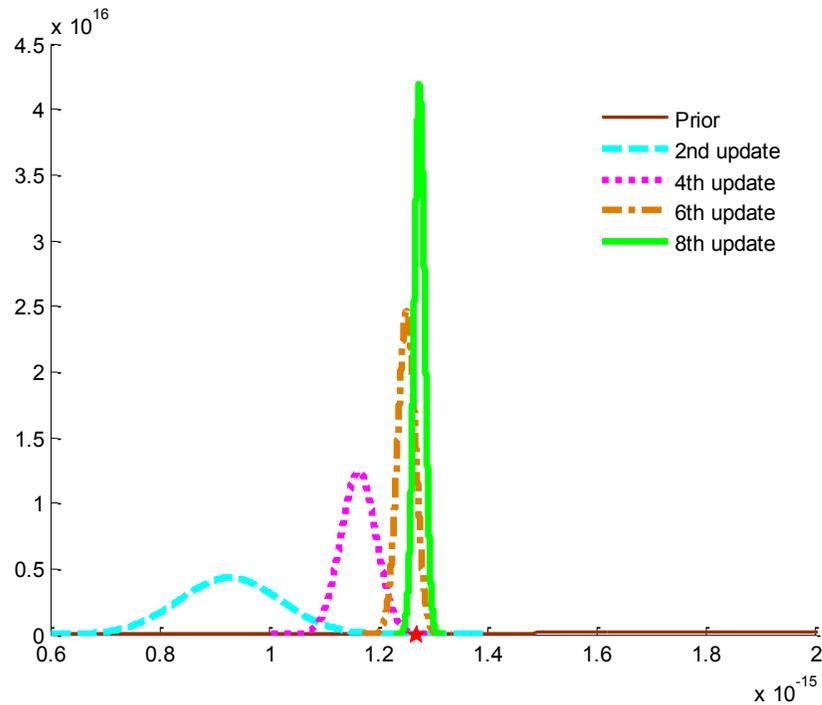
Figure 56. Updated distributions of wear coefficient k

169

From the results of update process, it is observed that the mean of wear coefficient k has been adjusted from the prior value $3 \times 10^{-15} \text{Pa}^{-1}$ to the value around $1.27 \times 10^{-15} \text{Pa}^{-1}$. Moreover, the shape of the distribution gets narrower, which indicates that the uncertainty is reduced gradually as more data of mass loss are available.
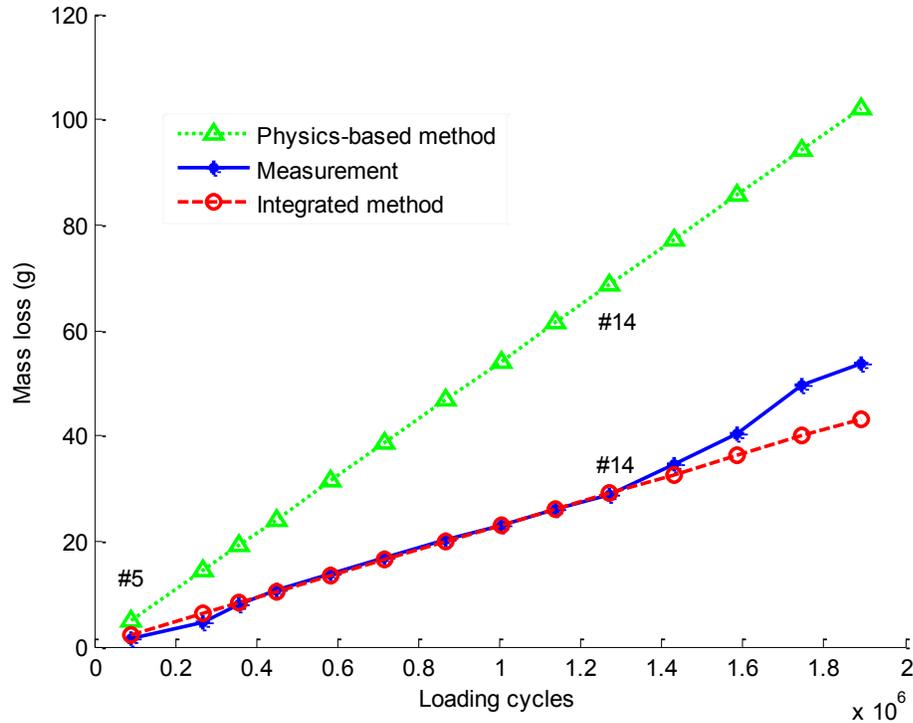
Figure 57. Measured and predicted mass loss of the sun gear

The prediction results are compared between the physics-based method and the integrated prognostics method, shown in Figure 57. The green dotted line with marks of triangle represents the mass loss predicted by physics-based method, which uses a prior value of wear coefficient $k = 3 \times 10^{-15} \text{Pa}^{-1}$. While, the red dashed line with marks of circle represents the mass loss predicted by the integrated prognostics method, which uses the mean value of posterior distribution of the wear coefficient in the last update

$k = 1.27 \times 10^{-15} \text{Pa}^{-1}$. From the figure, it is obvious that the predicted results obtained by the integrated prognostics method agree well with the measured data from run #5 and run #14. In contrast, the mass loss predicted by the physics-based method has large discrepancy with the measured data. The results show that, the integrated prognostics method is able to accurately identify the wear coefficient for a specific unit under monitoring by using measurement on gear mass loss. With the updated wear coefficient obtained through Bayesian inference, the wear process is predicted with good accuracy.

8.6　Conclusions

In this chapter, the integrated prognostics method is developed for the failure mode of surface wear of the gear. The validation is conducted using a run-to-failure experiment. The wear model is only used to predict the surface wear due to sliding contact. In order to account for the contribution of rolling contact to the mass loss, we propose to remove the material using the maximum wear depth. In addition, the assumption of contact pressure and sliding distance being constant during the wear process leads to a linear prediction, which is not suitable anymore for the late stage of the wear process.

In order to obtain a more accurate prediction on wear process, in the future work, the FE model will be used to calculate the actual contact pressure and the sliding distance, accounting for the effects of tooth profile changes due to material removal. Moreover, a more reasonable model will be built to account for the interactions between pitting and surface wear process.

Chapter 9.   Conclusions and future work

9.1   Conclusions

Integrated prognostics methods for gear life prediction have been proposed in this thesis. The methods can make use of condition monitoring data and update the physical model to achieve more accurate failure time prediction.

Several aspects in prognostic method development are researched, including uncertainty quantification, time-varying operating conditions, crack initiation time and shock in the degradation.

An efficient method is developed to improve the computational efficiency in uncertainty quantification in the integrated prognostics. The method is the stochastic collocation method based on PCE. It can accelerate the process of model update and RUL prediction significantly. By using this method, it is possible to fully explore the effect of uncertainty from multiple sources on the life prediction.

The integrated prognostics method can deal with time-varying operating conditions. In this thesis, the varying loading condition is considered since it is the most important operating factor for a power transmission system. The loading profile takes the form of piece-wise constant. With the capability to directly relate the load to the degradation rate, the integrated prognostics method can account for the loading change in RUL prediction.

The CIT affects the failure time prediction. By considering it as a source of uncertainty, the prediction accuracy is improved by adjusting both the slope and the intercept of the degradation path. It also facilitates the prediction of degradation with

shock in it. By searching a different CIT, the reduction in lifetime that is caused by the shock can be accommodated.

We also develop the integrated prognostics method for another failure mode of the gear: surface wear caused by sliding contact. The proposed method is validated using a run-to-failure experiment on a planetary gearbox. The recorded data on mass loss in the experiment are used to update the wear coefficient in the integrated prognostics method. By observing the tooth profile change in the photographs, the maximum wear depth is used to remove material, which accounts for the rolling contact to some extent. The predicted mass loss matches well with the measured mass loss during the middle stage of wear process. The wear model is no longer valid at the late stage because of the assumption of contact pressure and sliding distance being constant. More sophisticated methods are needed to calculate the actual contact pressure and the sliding distance in the future work.

9.2   Future work

The future work may include the following aspects.

- It is expected that the wear process should be affected by the wear itself. That is, the contact pressure and the sliding distance should be a function of wear depth. In order to improve the wear model, FE modeling will be used to calculate the contact states and the sliding distance. The assumption of constant values for them can be removed to produce more reasonable results.

- The wear process is actually involving more than one wear mechanisms. It is a complicated process that all the mechanisms can affect each other. For example, the

mild wear due to sliding contact can expose some areas of the tooth body under stress concentration, hence to accelerate the fatigue process due to rolling contact. More effort can be invested to build an effective physical model to predict the pitting process and to investigate the interactions between different wear mechanisms.

- In this thesis, the crack size is taken as the observations to update model parameters. However, it may be difficult to obtain an accurate mapping between crack size and condition feature. Under such circumstance, the integrated prognostics method should have the capability to use condition feature directly to update model parameters. Hence, the research effort will be further devoted to extracting effective and representative condition features for prognostics and to incorporate their uncertainty.

- The time-varying loading condition considered in the current work takes the form of piece-wise constant. We assume the theory of linear elastic fracture mechanics can always apply. However, the variable amplitude loading condition may result in crack closure and plasticity, which makes the linear elastic fracture mechanics theory not suitable. The future work will consider the physical model for life prediction that can account for more factors in the crack propagation.

9.3 Paper published from the thesis work

A. Peer-Reviewed Journal Papers

1. Fuqiong Zhao, Zhigang Tian and Yong Zeng, "Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method," *IEEE Transactions on Reliability*, vol. 62, no. 1, pp. 146-159. 2013.

2. Fuqiong Zhao, Zhigang Tian and Yong Zeng, "A stochastic collocation approach for efficient integrated gear health prognosis," *Mechanical Systems and Signal Processing*, vol. 39, pp. 372-387, 2013.

3. Fuqiong Zhao, Zhigang Tian, Eric Bechhoefer and Yong Zeng, "An integrated prognostics method under time-varying operating conditions," *IEEE Transactions on Reliability*, vol. PP, no. 99, pp. 1-14, 2015.

4. Fuqiong Zhao, Zhigang Tian and Yong Zeng, "Gear integrated prognostics considering shock in degradation and uncertainty in crack initiation time," submitted.

5. Fuqiong Zhao and Zhigang Tian "Integrated prognostics method for gear wear prediction", in preparation.

B. Peer-Reviewed Conference Papers

1. Fuqiong Zhao and Zhigang Tian, "Gear integrated prognosis considering crack initiation time uncertainty". In *Proceedings of the 20th ISSAT International Conference on Reliability and Quality in Design*, Seattle, USA, August, 2014.

2. Fuqiong Zhao and Zhigang Tian, "Gear remaining useful life prediction using generalized polynomial chaos collocation method". In *Proceedings of the 18th ISSAT International Conference on Reliability and Quality in Design*, Boston, USA, July, 2012.

3. Fuqiong Zhao and Zhigang Tian. ''Crack propagation simulation in spur gear tooth root using ANSYS''. In *Proceedings of the 17th ISSAT International Conference on Reliability and Quality in Design*, Vancouver, Canada, August, 2011.

References

[1] Abernethy, R. B. *The New Weibull Handbook*. 2nd ed. Abernethy, North Palm Beach, 1996.

[2] An, D., Choi, J. and Kim, N. H. Identification of correlated damage parameters under noise and bias using Bayesian inference, *Structural health monitoring*, 11(3), 293-303, 2012.

[3] Andersen, P. K. and Gill, R. D. Cox's regression model for counting processes: a large sample study, *The annals of statistics*, 10(4), 1100-1120, 1982.

[4] Anderson, T. L. *Fracture mechanics: fundamentals and applications*, 3rd ed., Taylor & Francis, 2005.

[5] Andersson, S. Partial EHD theory and initial wear of gears, Doctoral thesis, Royal institute of technology, Stockholm, 1975.

[6] Andersson, S. Wear simulation, Advance knowledge application in practice, Lgor Fuerstner (Ed.), InTech, 2010.

[7] Annis, C. Probabilistic life prediction isn't as easy as it looks, Probabilistic aspects of life prediction, ASTM STP-1450, W. S. Johnson and B. M. Hillberry, Eds., ASTM International, West Conshohocken, PA, 2003.

[8] Archard, J. F. Contact and rubbing of flat surface, *Journal of applied physics*, 24(8), 981-988, 1953.

[9] ASTM E647-00 Standard test method for measurement of fatigue crack growth rates, ASTM International, 2000.

[10] Bajpai, P., Kahraman, A. and Anderson, A. surface wear prediction methodology for parallel-axis gear pairs, *Journal of tribology*, 126(3), 2004.

[11] Banjevic, D., Jardine, A. K. S., Makis, V. and Ennis, M. A control-limit policy and software for condition-based maintenance optimization, *INFOR*, 39(1), 32-50, 2001.

[12] Bartelmus, W. Mathematical modeling and computer simulations as an aid to gearbox diagnostics, *Mechanical systems and signal processing*, 15, 855-871, 2001.

[13] Byington, C. S., Watson, M. J., Sheldon, J. S. and Swerdon, G. M. Shaft coupling model-based prognostics enhanced by vibration diagnostics, *Helicopter prognostics*, 51, 420-425, 2009.

[14] Cadini, F., Zio, E. and Avram, D. Monte Carlo-based filtering for fatigue crack growth estimation, *Probabilistic engineering mechanics*, 24, 367-373, 2009b.

[15] Canuto, C., Hussaini, M. Y., Quarteroni, A., Zang, T. A. *Spectral methods in fluid dynamics*, Springer-Verlag, New York, 1988.

[16] Chaari, F., Fakhfakh, T., Haddar, M. Analytical modeling of spur gear tooth crack and influence on gear mesh stiffness, *European journal of mechanics a/solids*, 28, 461-468, 2009.

[17] Chantrasmi, T., Constantine, P., Etemadiz, N., Iaccarino, G. and Wang, Q. Uncertainty quantification in simple linear and non-linear problems, *Annual research briefs,* 2006.

[18] Chantrasmi, T., Doostan, A., Iaccarino, G. and Padé–Legendre, Approximants for uncertainty analysis with discontinuous response surfaces, *Journal of computational physics*, 228, 7159–7180, 2009.

[19] Choi, S. and Li, C. J. Practical gear crack prognosis via gear condition index fusion, gear dynamic simulator, and fast crack growth model, *Proceedings of the institution of mechanical engineers, part I: journal of systems and control engineering,* 221(3), 465-473, 2007.

[20] Collipriest, J. E. An experimentalist's view of the surface flaw problem, *The surface crack: physical problems and computational solutions, American society for mechanical Engineers*, 43-61, 1972.

[21] Coppe, A., Haftka, R. T. and Kim, N. H. Uncertainty reduction of damage growth properties using structural health monitoring, *Journal of aircraft*, 47(6), 2030-2038, 2010.

[22] Cornell, R. W. and Westervelt, W. W. Dynamic tooth loads and stressing for high contact ratio spur gears, *Journal of mechanical design*, 100, 69-76, 1978.

[23] Cox, D. R. Regression models and life-tables, *Journal of the Royal Statistical*

Society, Series B, 34(2), 187-220, 1972.

[24] Debusschere, B. J., Najm, H. N., Lébay, P. P., Knio, O. M., Ghanem, R. G. and Le Maître, O. P. Numerical challenges in the use of polynomial chaos representations for stochastic process, *Siam Journal on scientific Compu*ting, 26, 698-719, 2004.

[25] Ditlevsen, O. and Olesen, R. Statistical analysis of the Virkler data on fatigue crack, *Engineering fracture mechanics*, 25(2), 177-195, 1986.

[26] Do, V. T., Hoseini, M. R., Sahoo, A., Pandey, M. and Zuo, M. J. Metal Scan data, gear weight loss data, and gear damage images recorded during the run-to-failure experiments, Technical report, Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, 2010.

[27] Elber, W. The significance of crack closure. *Damage tolerance in aircraft structure*, Philadelphia: American Society for Testing and Materials. 230-242, 1971.

[28] Eldred, M. S. Recent Advances in Non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design, in *Proceedings of the 11th AIAA Nondeterministic Approaches Conference, No. AIAA-2009-2274*, Palm Springs, CA, 2009.

[29] Eldred, M. S., Webster, C. G. and Constantine, P. Evaluation of non-intrusive approaches for Wiener-Askey generalized polynomial chaos, in *Proceedings of the 10th AIAA Nondeterministic Approaches Conference, No. AIAA-2008-1892*, Schaumburg, IL, 2008.

[30] Erdogan, F. and Sih, G. C. On the crack extension in plates under plane loading and transverse shear, *Journal of Basic Engineering,* 85(4), 519-525, 1963.

[31] Flodin, A. and Andersson, S. Simulation of mild wear in spur gears, *wear*, 207, 16-23, 1997.

[32] Gebraeel, N. and Lawley, M. A. A neural network degradation model for computing and updating residual life distributions, *IEEE transactions on automation science and engineering*, 5(1), 154–163, 2008.

[33] Gebraeel, N., Lawley, M. A. and Liu, R. Residual life, predictions from vibration-based degradation signals: a neural network approach, *IEEE transactions on industrial electronics*, 51(3), 694–700, 2004.

[34] Gebraeel, N. and Pan, J. Prognostic degradation models for computing and updating residual life distributions in a time-varying environment, *IEEE transactions on reliability*, 57(4), 539-550, 2008.

[35] Ghosh, D. and Farhat, C. Strain and stress computations in stochastic finite element methods, *International journal for numerical methods in engineering*, 74(8), 1219-1239, 2008.

[36] Glodez, S., Sraml, M. and Kramberger, J. A computational model for determination of service life of gears, *International journal of fatigue*, 24(10), 1013-1020, 2002.

[37] Hannes D. and Alfredsson B. A fracture mechanical life prediction method for rolling contact fatigue based on the asperity point load mechanism, *Engineering fracture mechanics*, 83, 62-74, 2012.

[38] Heng, A., Zhang, S., Tan, A. C. C. and Mathew, J. Rotating machinery prognostics: State of the art, challenges and opportunities, *Mechanical systems and signal processing*, 23, 724-739, 2009.

[39] Hines, J. W. and Garvey, D. R. Data based fault detection, diagnosis and prognosis of oil drill steering systems, Maintenance and reliability conference, Knoxville, Tennessee, USA, 2007.

[40] Hines, J. W. and Usynin, A. Current computational trends in equipment prognostics, *International journal of computational intelligence systems*, 1(1), 94-102, 2008.

[41] Howard, I., Jia, S., and Wang, J. The dynamic modeling of a spur gear in mesh including friction and a crack, *Mechanical systems and signal processing*, 15, 831-853, 2001.

[42] Huang, R. Q., Xi, L. F. and Li, X. L. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, *Mechanical systems and signal processing*, 21(1), 193–207, 2007.

[43] Inoue, K., Kato, M., Deng, G. and Takatsu, N. Fracture mechanics based of strength of carburized gear teeth, in *Proceedings of JSME international conference on motion and power transmissions*, Hiroshima, Japan, 1999.

[44] Jardine, A. K. S., Lin, D. M. and Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, 20(7), 1483-1510, 2006.

[45] Jardine, A. K. S. and Tsang, A. H. C. *Maintenance, replacement, and reliability: theory and applications*, CRC Press, 2005.

[46] Jaw, L. and Wang, W. A run-time test system for maturing intelligent system/vehicle capabilities – SIDAL. in *Proceedings of 2004 IEEE Aerospace Conference*, Big Sky, MT, USA, 2004.

[47] Jiang, L., Feng, Q. and Coit, D. W. Reliability and maintenance modeling for dependent competing failure processes with shifting failure thresholds, *IEEE transactions on reliability*, 61, 932-948, 2012.

[48] Kacprzynski, G. J., Sarlashkar, A., Roemer, M. J., Hess, A., and Hardman, G. Predicting remaining life by fusing the physics of failure modeling with diagnostics, *Journal of the minerals, metals and materials society*, 56(3), 29-35, 2004.

[49] Kasuba, R. and Evans, J. W. An extended model for determining dynamic loads in spur gearing, *Transactions of the ASME*, 103, 398-409, 1981.

[50] Kotzalas, M. N. and Harris, T. A. Fatigue failure progression in ball bearings, *Transactions of the ASME*, 123, 238-242, 2001.

[51] Kotulski, Z. A., On efficiency of identification of a stochastic crack propagation model based on Virkler experimental data, *Archives of mechanics*, 50(5), 829-847, 1998.

[52] Lewicki, D. G. and Ballarini, R. Effect of rim thickness on gear crack-propagation path, *Journal of mechanical design*, 119(1), 88-95, 1997.

[53] Lewicki, D. G., Handschuh, R. F., Spievak, L. E., Wawrzynek, P. A. and Ingraffea, A. R. Consideration of moving tooth load in gear crack propagation predictions, *Journal of mechanical design*, 123(1), 118-124, 2001.

[54] Li, C. J. and Choi, S. Spur gear root fatigue crack prognosis via crack diagnosis and fracture mechanics, in *Proceedings of the 56th Meeting of the Society for Machinery Failure Prevention Technology*, 2002.

[55] Li, C. J. and Lee, H. Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics, *Mechanical systems and signal processing,* 19(4), 836–846, 2005.

[56] Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S. and Liang, S. Adaptive prognostics for rolling element bearing condition, *Mechanical systems and signal processing*, 13(1), 103-113, 1999.

[57] Li, Y., Billington, S. Zhang, C., Kurfess, T., Danyluk, S. and Liang, S. Dynamic prognostic prediction of defect propagation on rolling element bearings, *Tribology transactions*, 42 (2), 385-392, 1999.

[58] Lin, H. H., Huston, R. L. and Coy, J. J. On dynamic loads in parallel shaft transmissions: part I – modeling and analysis, *Journal of mechanisms, transmissions, and automation in design*, 110, 221-225, 1988.

[59] Liao, H and Tian, Z. A framework for predicting the remaining useful life of a single unit under time-varying operating conditions*, IIE Transactions*, 45(9), 964-980, 2013.

[60] Lin, H. H., Oswald, F. B. and Townsend, D. P. Dynamic loading of spur gears with linear or parabolic tooth profile modifications, *Mechanism and machine theory*, 29(8). 1115-1129, 1994.

[61] Lin, H. H., Townsend, D. P. and Oswald, F. B. Profile Modification to minimize spur gear dynamic loading, *NASA technical memorandum 89901*, 1988.

[62] Liou, C. H., and Lin, H. H., Oswald, F. B. and Townsend, D. P. Effect of contact ratio on spur gear dynamic load, *NASA technical memorandum 105606*, 1992,

[63] Litvin, F. L., *Gear Geometry and Applied Theory*, Prentice-Hall, Inc., 1994,

[64] Lu, C. J., and Meeker W. Q. Using degradation measures to estimate a time-to-failure distribution, *Technometrics*, 35(2), 161-174, 1993.

[65] Mallor, F. and Santos, J. Classification of shock models in system reliability, *Monografias del semin matematicas garcia de galdeano*, 27, 405-412, 2003.

[66] Marble, S. and Morton, B. P. Predicting the remaining life of propulsion system bearings, In *Proceedings of the 2006 IEEE aerospace conference*, Big Sky, MT, USA, 2006.

[67] Marseguerra, M., Zio, E. and Podofillini, L. Condition-based maintenance optimization by means of genetic algorithms and Monte Carlo simulation, *Reliability engineering & system safety,* 77, 151-165, 2002.

[68] Marzouk, Y. M., Najm, H. N. and Rahn, L. A. Stochastic spectral methods for efficient Bayesian solution of inverse problems, *Journal of computational physics*, 224, 560-586, 2007.

[69] Marzouk, Y. and Xiu, D. A stochastic collocation approach to Bayesian inference in inverse problem, *Communications in computational physics*, 6(4), 826-847, 2009.

[70] McFadden, P. D. Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration, *Journal of vibration, acoustic, stress, and reliability in design*, 108, 165–70, 1986.

[71] Najm, H. N. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual review of fluid mechanics,* 41, 35-52, 2009.

[72] Najm, H. N., Debusschere, B. J., Marzouk, Y. M., Widmer, S. and Le Maître, O. P. Uncertainty quantification in chemical systems, International journal for numerical methods in engineering, 80, 789-814, 2009.

[73] Narayana, K. B. and Dattaguru, B. Certain aspects related to computation by modified crack closure integral (MCCI), *Engineering fracture mechanics,* 55(2), 335–339, 1996.

[74] Oppenheimer, C. H. and Loparo, K. A. Physically based diagnosis and prognosis of cracked rotor shafts, in *Proceedings of SPIE 4733*, 122-132, 2002.

[75] Orchard, M. E. and Vachtsevanos, G. J. A particle filtering approach for on-line failure prognosis in a planetary carrier plate, *International journal of fuzzy logic and*

intelligent systems, 7(4), 221-227, 2007.

[76] Ortiz, K. and Kiremidjian, A. S. Stochastic modeling of fatigue crack growth, *Engineering fracture mechanics*, 29, 317-334, 1988.

[77] Oswald, F. B., Lin, H. H. and Delgado, I. R. Dynamic analysis of spur gear transmissions (DANST), PC version 3.00 user manual, 1996.

[78] Paris, P. C. and Erdogen, F. A critical analysis of crack propagation laws, *Journal of basic engineering*, 85(4), 528-534, 1963.

[79] Peng, Y., Zhang, S., and Pan, R. Bayesian network reasoning with uncertain evidences, *International journal of uncertainty, fuzziness and knowledge-based systems,* 18(5), 539-564, 2010.

[80] Peterson, M. B. and Winer, W. O. Wear control handbook, 1980.

[81] Pettersson, P., Iaccarino, G. and Nordström, J. Numerical analysis of the Burgers' equation in the presence of uncertainty, *Journal of computational physics*, 228(22), 8394-8412, 2009.

[82] Qiu, J., Zhang, C. Seth, B. B. and Liang, S. Y. Damage mechanics approach for bearing lifetime prognostics, *Mechanical systems and signal processing*, 16(5), 817–829, 2002.

[83] Raju, I. S., Shivakumar, K. N. An equivalent domain integral method in the two dimensional analysis of mixed mode crack problems. *Engineering fracture mechanics,* 37 (4), 707–725, 1990.

[84] Reagan, M. T., Najm, H. N., Debusschere, B. J., Le Maître, O. P., Knio, O.M. and Ghanem, R. G. Spectral stochastic uncertainty quantification in chemical systems, *Combustion theory and model*ling, 8, 607-632, 2004.

[85] Roemer, M. J., Byington, C. S. and Sheldon, J. Advanced vibration analysis to support prognosis of rotating machinery components, *International journal of COMADEM*, 11(2), 2-11, 2008.

[86] Samanta, P. K., Vesely, W. E., Hsu, F. and Subudly, M. Degradation modeling with application to ageing and maintenance effectiveness evaluations, *NUREG/CR-5612. US Nuclear Regulatory Commission*, 1991.

[87] Sankararaman, S., Ling, Y. and Mahadevan, S. Uncertainty quantification and model validation for fatigue crack growth prediction, *Engineering fracture Mechanics*, 78, 1487-1504, 2001.

[88] Shao, Y. and Nezu, K. Prognosis of remaining bearing life using neural networks, *Proceedings of the institution of mechanical engineers, part I: journal of systems and control engineering*, 214(3), 217-230, 2000.

[89] Shih, C. F., De Lorenzi, H. G. and German, M. D. Crack extension modeling with singular quadratic isoparametric elements, *International journal of fracture,* 12, 647-651, 1976.

[90] Smolyak, S. Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet mathematics doklady*, 4, 240-243, 1963.

[91] Soize, C. and Ghanem, R. Physical systems with random uncertainties: chaos representations with arbitrary probability measure, *SIAM Journal on scientific computing,* 26, 395-410, 2004.

[92] Sotiris, V., and Pecht, M. Support vector prognostics analysis of electronic products and systems. AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2007.

[93] Tavakoli, M. S. and Houser, D. R. Optimum profile modifications for the minimization of static transmission errors of spur gears, *Transactions of ASME*, 108, 86-94, 1986.

[94] Tian, X. H., Zuo, M. J. and Fyfe, K. R. Analysis of the vibration response of a gearbox with gear tooth faults, in *Proceedings of IMECE04, 2004 ASME international mechanical engineering congress and exposition*, Anaheim, CA, 2004.

[95] Tian, Z., Wong, L. and Safaei, N. A neural network approach for remaining useful life prediction utilizing both failure and suspension histories, *Mechanical systems and signal processing*, 24, 1542-1555, 2010.

[96] Tian, Z. An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring, *Journal of intelligent Manufacturing*, 23(2), 227-237, 2012.

[97] Townsend, D. P., *Gear Handbook*, McGRAW-HILL, INC., 1992.

[98] Tse, P. and Atherton, D. Prediction of machine deterioration using vibration based fault trends and recurrent neural networks, *Journal of vibration and acoustics,* 121, 355-362, 1999.

[99] Vedmar, L. and Andersson, A. A method to determine dynamic loads on spur gear teeth and on bearings, *Journal of sound and vibration*, 267, 1065-1084, 2003.

[100] Velex, P. and Ajmi, M. Dynamic tooth loads and quasi-static transmission errors in helical gears – approximate dynamic factor formulae*, Mechanism and machine theory*, 42, 1512-1526, 2007.

[101] Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A. and Wu, B. *Intelligent fault diagnosis and prognosis for engineering system*, Wiley, 2006.

[102] Virkler, D. A., Hillberry, B. M. and Goel P. K., The statistical nature of fatigue crack propagation, Journal of engineering materials and technology, 101(2), 148-153, 1979.

[103] Wand, M. P. and Jones, M. C. *Kernel smoothing*. Chapman and Hall, London, 1995.

[104] Wang, J. and Howard, I. The torsional stiffness of involute spur gears, *Journal of mechanical engineering science*, 218, 131-142, 2004.

[105] Wang, P. and Vachtsevanos, G. Fault prognostics using dynamic wavelet neural networks, *Artificial intelligence for engineering design, analysis and manufacturing,* 15, 349–365, 2001.

[106] Wang, W. Q., Golnaraghi, M. F., and Ismail, F. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical systems and signal processing, 18*, 813-831, 2004.

[107] Wang, Z., Huang, H-Z, Li, Y. and Xiao, N-C An approach to reliability assessment under degradation and shock process, *IEEE transactions on reliability,* 60, 852-863, 2011.

[108] Wheeler, E. Spectrum loading and crack growth, *Journal of basic engineering, Transactions of the ASME*, 94, 181-186, 1972.

[109] Williams, M. L. On the stress distribution at the base of a stationary crack*, Journal of applied mechanics, ASME*, 24(1), 109–114, 1957.

[110] Wu, S. J., Gebraeel, N., Lawley, M. A. and Yih, Y. A neural network integrated decision support system for condition-based optimal predictive maintenance policy, *IEEE transactions on systems man and cybernetics part A – systems and humans*, 37(2), 226–236, 2007.

[111] Wu, S. and Cheng, H. S. Sliding wear calculation in spur gears, *Journal of tribology*, 115, 493-500, 1993.

[112] Wu, S., Zuo, M. J. and Parey, A. Simulation of spur gear dynamics and estimation of fault growth, *Journal of sound and vibration*, 317, 608-624, 2008.

[113] Xiu, D. Efficient collocational approach for parametric uncertainty analysis, *Communications in computational physics*, 2, 293-309, 2007.

[114] Xiu, D. *Numerical methods for stochastic computations*, Princeton, New Jersey, 2010.

[115] Xiu, D. and Hesthaven, J. S. High-order collocation methods for differential equations with random inputs, *SIAM Journal on scientific computing,* 27, 1118-1139, 2005.

[116] Xiu, D. and Karniadakis, G. E. The Wiener-Askey polynomial chaos for stochastic differential equations, *SIAM Journal on scientific computing*, 24(2), 619-644, 2002.

[117] Xu, G. and Sadeghi, F. Spall initiation and propagation due to debris denting, *wear*, 201, 106-116, 1996.

[118] Yan, J., Koç, M. and Lee, J. A Prognostic algorithm for machine performance assessment and its application, *Production planning and control,* 15(8), 796-801, 2004.

[119] Yang, D. C. H. and Lin, J. Y. Hertzian damping, tooth friction and bending elasticity in gear impact dynamics, *Journal of mechanisms, transmissions, and automation in design*, 109, 189-196, 1987.

[120] Yang, D. C. H. and Sun, Z. S. A rotary model for spur gear dynamics, *ASME Journal of mechanisms, transmissions, and automation in design*, 85-DET-2, 1985.

[121] Yeh, R. H. State-age-dependent maintenance policies for deteriorating systems with Erlang sojourn time distributions, *Reliability engineering system safety,* 58, 55-60, 1997.

[122] Zhao, F., Tian, Z. and Zeng, Y. Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method, *IEEE transactions on reliability*, 62(1), 146-159, 2013a.

[123] Zhao, F., Tian, Z. and Zeng, Y. A stochastic collocation approach for efficient integrated gear health prognosis, *Mechanical systems and signal processing*, 39, 372-387, 2013b.

[124] Zhao, F., Tian, Z., Bechhoefer E. and Zeng, Y. An integrated prognostics method under time-varying operating conditions, *IEEE transactions on reliability*, DOI: 10.1109/TR.2015.2407671, 2015a.

[125] Zhao, F., Tian, Z. and Zeng, Y. Gear integrated health prognosis considering crack initiation time uncertainty and random shock, under review, 2015b.

[126] Zio, E. and Peloni, G. Particle filtering prognostic estimation of the remaining useful life of nonlinear components, *Reliability engineering and system safety*, 96, 403-409, 2011.

[127] Zio, E. Prognostics and health management of industrial equipment, *Seifedine Kadry. Diagnostics and prognostics of engineering systems: methods and techniques, IGI Global*, 333-356, 2012.

[128] Zouari, S., Maatar, M., Fakhfakh, T. and Haddar, M. Following spur gear crack propagation in the tooth foot by finite element method, *Journal of failure analysis and prevention*, 10(6), 531-539, 2010.