An Intelligent PHM System for Bearing Fault Phase Diagnosis and Multi-Phase RUL Prediction using Oil Debris Monitors

by

Lourd Arun Raj

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Aerospace Engineering

Carleton University
Ottawa, Ontario

Abstract

Several experimental studies have found that a rolling element bearing service life can be typically divided into three phases: normal operation, gradual degradation, and accelerated degradation. Although extensive studies have been carried out on bearing life prediction, most of these studies focus only on life prediction in the accelerated degradation phase. Many studies in the literature show that Oil Debris Monitors (ODMs) have excellent potential in bearing condition monitoring. Yet, only a handful of studies explore its potential for Remaining Useful Life (RUL) prediction of bearings. Therefore, in this thesis, an intelligent bearing health management system is presented that primarily uses the information from the ODMs. The system contains two major modules namely, fault phase diagnosis and RUL prediction. Firstly, the fault phase diagnosis module is a data-driven tool based on an ensemble method of K-Nearest Neighbors and Random Forest classifiers. This diagnosis tool uses multi-sensor data along with three novel degradation indicators based on wear debris characteristic information. Secondly, the RUL prediction module is a model-based tool where a physics-based multi-phase degradation model along with a novel Particle Filter (PF) technique is utilized. Almost all model-based RUL prediction studies use a single monotonic degradation model to represent the bearing degradation process however, these models are not suitable for long-term predictions. Therefore, a multi-phase bearing degradation model is developed using the Gaussian process and Weibull distribution functions. Furthermore, since multi-phase bearing RUL prediction involves sudden trend changes in degradation data, standard PF algorithms are not capable of adapting to the trend changes. Therefore, an Enhanced Adaptive PF algorithm is proposed to tackle the primary issues of PF algorithms. The validity of the proposed methodology is demonstrated using five bearing spall propagation experiments.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Dr. Jie Liu for giving me the opportunity to work on this interesting and challenging research topic with Gastops Ltd. His mentoring, patient guidance and immense knowledge have helped me complete this thesis. Also, my deepest gratitude to Mr. Stephen Swart, who helped me to advance my research with valuable advice and in-depth discussions.

I would also like to express my special appreciation to the thesis defense committee members Prof. Patrick Dumond, Prof. Fidel Khouli, and Prof. Rong Liu for giving such thoughtful suggestions and recommendations to improve the dissertation.

This project owes its inception and progression to the strong collaboration between Carleton University and Gastops Ltd. I am grateful for the encouragement, guidance, and support by Mr. Oleg Sosnovski and Ms. Pooja Suresh during this project.

I have been fortunate to work with my friends Shengcai Deng, Hassan Mahmoud, and Alexander Gu on this project. They provided a positive and interactive environment and a constant source of reassurance.

Finally, I would like to use this opportunity to express my sincere gratitude to my beloved wife Navdeep Kaur. Her encouragement (sometimes criticism) is the greatest motivation in my life. I thank my parents and my brother for trusting and supporting me in all my decisions.

Table of Contents

Abstract ..... **ii**

Acknowledgements ..... **iii**

Table of Contents ..... **iv**

List of Tables ..... **viii**

List of Figures ..... **ix**

List of Appendices ..... **xii**

Abbreviations ..... **xiii**

Chapter 1: Introduction **1**

1.1 Overview 1

1.2 Literature Survey and Motivation 3

1.2.1 Bearing Degradation Indicators from Wear Debris Analysis 3

1.2.2 Fault Phase Diagnosis using Ensemble Method 5

1.2.3 Adaptive Particle Filtering for Model-Based RUL Prediction 10

1.2.4 Multi-Phase Degradation Model for Long-Term RUL Prediction 13

1.3 Objectives 14

1.4 Experimental setup 17

1.5 Thesis Outline 19

Chapter 2: Bearing Degradation Indicators from Wear Debris Analysis **20**

2.1 Overview 20

2.2 Wear Debris Analysis 21

2.2.1 Operating Principle of ODMs 21

2.2.2 Bearing Condition Monitoring using ODM 24

2.3 Wear Debris Characteristics 25

2.3.1 Wear Debris Characteristics of Experimental Data 27

2.3.2 Degradation Indicators Based on Wear Debris Characteristics 29
2.4 Online Change Point Detection for Degradation Indicator 31
2.4.1 ROCD based Degradation Indicator 32
2.5 Summary 34

Chapter 3: Fault Phase Diagnosis using Ensemble of Random Forest and K-Nearest Neighbors 35

3.1 Overview 35
3.2 Classifiers 36
3.2.1 K-Nearest Neighbors (KNN) 36
3.2.2 Random Forest (RF) 38
3.3 Feature Selection Techniques 39
3.3.1 Information Gain 39
3.3.2 Chi-Squared Test 40
3.3.3 Evaluation Metrics 41
3.4 Proposed Framework 42
3.4.1 Training the Classification Models 42
3.4.2 Fault Phase Diagnosis 45
3.5 Demonstration on Experimental Data 46
3.5.1 Data Acquisition and Feature Extraction 47
3.5.2 Feature Selection 49
3.5.3 Training Classifiers and Optimization 50
3.5.4 Testing the Ensemble Technique 52
3.6 Summary 55

Chapter 4: Adaptive Particle Filter for Real-Time Degradation Prediction 56

4.1 Overview 56
4.2 Particle Filtering for Prognostics 56

4.2.1 Principle of Particle Filtering 56
4.2.2 Standard Particle Filters 58
4.3 Enhanced Adaptive PF 60
4.4 Model-based Prognostics with Enhanced Adaptive Particle Filter 66
4.4.1 Model-based Prognostics 66
4.4.2 Developing a Simulated Bearing Health State 68
4.4.3 Initializing particle filter 69
4.4.4 Model Parameter Estimation 70
4.4.5 Effect of Noise on Model Parameter Estimation 73
4.5 Summary 76

Chapter 5: A Multi-Phase Degradation Model based on Mixture of Gaussian Process and Weibull Functions 77

5.1 Overview 77
5.2 Phase Transition Mass of the Bearings 78
5.3 Multi-Phase Degradation Model 80
5.3.1 Sum of Gaussians 80
5.3.2 Modified Weibull Failure Rate 82
5.3.3 Phase Transition Mass 85
5.3.4 Geometric Relations 88
5.3.5 Empirical Relations 89
5.4 Compiling the Degradation Model 90
5.5 Summary 91

Chapter 6: A PHM Framework for Long-Term and Multi-Phase RUL Prediction 92

6.1 Proposed PHM Framework 92
6.1.1 Fault Phase Diagnosis 94
6.1.2 RUL Prediction 94

6.2 Results and Discussion 97
6.2.1 RUL Prediction on Experiment 4 98
6.2.2 RUL Prediction on Experiment 5 100
6.3 Summary 103
Chapter 7: Conclusions and Future Works 104
7.1 Conclusion 104
7.2 Limitations and Future Works 105
Appendices 107
References 116

List of Tables

Table 3.1 List of features along with the processing steps. 47

Table 6.1 Database of experiments used for training 98

List of Figures

Fig. 1.1 Phases of degradation illustrated using the debris counts from ODM [3]..... 2

Fig. 1.2 (a) Examples of monotonic and trendable indicators, (b) examples of non-monotonic and non-trendable indicators [18]. 4

Fig. 1.3 Estimation of FPT in a simple two-phase degradation process [31]. 7

Fig. 1.4 The schematic of the intelligent PHM system. 16

Fig. 1.5 Schematic of the experimental setup. 17

Fig. 2.1 Illustration of ODMs from Gastops (MetalSCAN™ 1000) [94]. 21

Fig. 2.2 Cross-section of typical ODM when no debris particle is detected [12]. 22

Fig. 2.3 Cross-section of typical ODM when in an unbalanced state [12]. 23

Fig. 2.4 ODM output signal examples [22]. 24

Fig. 2.5 ODM debris counts data from a spall propagation test binned as per the average bin size (in micro meter) and mass (in micro grams). 25

Fig. 2.6 Relationship between debris generation and fault severity [21]..... 26

Fig. 2.7 Evolution of the debris distribution as the fault severity increases. 27

Fig. 2.8 Variation in the debris distribution as the fault severity increases. 28

Fig. 2.9 Standard deviation and skewness indicators from the debris counts distribution. 31

Fig. 2.10 ROCD based degradation indicator with 20 minutes sliding window identifies the areas where bearing degradation occurs. 33

Fig. 3.1 Example of KNN classification [104]. 37

Fig. 3.2 The step-by-step process for training the classifiers. 43

Fig. 3.3 The step-by-step process to implement fault phase diagnosis on the real-time data. 45
Fig. 3.4 Feature ranking from both the methods arranged in ascending order and the bottom 1/3rd features are highlighted. 50
Fig. 3.5 Comparison of the evaluation metrics on the validation experiments..... 51
Fig. 3.6 Fault phase diagnosis of the spall propagation experiment 4. 53
Fig. 3.7 Fault phase diagnosis of the spall propagation experiment 5. 54
Fig. 4.1 Flow chart of the proposed enhanced adaptive particle filter. 64
Fig. 4.2 Illustration of the extended adaptive sampling process. 66
Fig. 4.3 Simulated bearing health state. 69
Fig. 4.4 Particle filter initialization and estimation of initial α using Least Square Error (LSE) curve fitting. 70
Fig. 4.5 Standard PF: (a) State estimation and inadaptability to trend change (b) Particle range reduction and particle impoverishment. 70
Fig. 4.6 Enhanced adaptive PF: (a) Accurate state estimation (b) Particle range adapting to the sudden changes. 72
Fig. 4.7 Comparing the model parameter estimations using enhanced adaptive PF and Bayesian method with MCMC sampling. 74
Fig. 4.8. Comparing the RMSE between enhanced adaptive PF and Bayesian method with MCMC sampling. 75
Fig. 5.1 Bearing mass loss trend curves for bearings from life endurance tests [118]. 79
Fig. 5.2 Bearing mass loss trend curves for bearings from life endurance tests. All at 2.41 GPa [117]. 80

Fig. 5.3 Approximation of the debris generation rate using the sum of Gaussians (magnitude of gaussian exaggerated for clarity). 81
Fig. 5.4 Influence of β parameter on the modified Weibull failure rate function. 83
Fig. 5.5 Modified Weibull failure rate approximates the overall growth of the debris generation rate. 84
Fig. 5.6 Area under Phase 1 Gaussian approximately equal to the phase transition mass. 86
Fig. 5.7 Approximation of Gaussian integral using elemental divisions. 87
Fig. 5.8 Approximation of Gaussian Integral. 88
Fig. 5.9 Geometric relation between μ_1 and μ_2. 89
Fig. 6.1 Schematic diagram for the proposed PHM system. 93
Fig. 6.2 Degradation predictions of experiment 4 at various time instants. 99
Fig. 6.3 RUL prediction plot for experiment 4 100
Fig. 6.4 Degradation predictions of experiment 5 at various time instants. 101
Fig. 6.5 RUL prediction plot for experiment 5 102
Fig. A.1 (a) Linear regression line fitted to the data and MSE of the residuals, (b) Residuals and MSE increases as when a change point occurs. 110
Fig. A.2 (a) Quadratic regression curve with positive curvature when change point occurs, (b) Quadratic regression curve with very low curvature, (c) Quadratic regression curve with negative curvature when change point occurs. 111
Fig. A.3 Effect of variation in change point residual in the observed variable (Y). 112
Fig. A.4 Simulated signal to validate the ROCD method. 113
Fig. A.5 BOCD method identifies only one change point. 114

Fig. A.6 ROCD method with 5 minutes sliding window identifies both the change points. 115

List of Appendices

Appendix A Regression-Based Online Change Point Detection 107

A.1 Online Change Point Detection Techniques 107

A.2 Bayesian Online Change Point Detection (BOCD) 108

A.3 Regression-based Online Change Point Detection (ROCD) 109

A.4 ROCD Methodology 112

A.5 Validation on Simulated Signal 113

A.6 Conclusion 115

Abbreviations

BOCD	Bayesian Online Changepoint Detection
FN	False Negative
FP	False Positive
FPT	First Prediction Time
IG	Information Gain
ISO	International Standards Organization
KNN	K-Nearest Neighbor
LSE	Least Square Error
MCMC	Monte Carlo Markov Chain
ODM	Oil Debris Monitor
PDF	Probability Density Function
PF	Particle Filter
PHM	Prognostics & Health Management
RF	Random Forest
RMS	Root Mean Square
RMSE	Root Mean Square Error
ROCD	Regression-based Online Change Point Detection
RUL	Remaining Useful Life
SIS	Sequential Importance Sampling
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
TP	True Positive

Chapter 1: Introduction

1.1 Overview

Prognostics and Health Management (PHM) systems primarily deal with real-time fault diagnosis and prediction of the future state of machinery based on its historical data. It incorporates several engineering disciplines such as sensing technologies, signal processing, failure physics, machine learning, modern statistics, and reliability engineering [1]. Fault diagnosis is a process of determining the presence of defects, identifying their locations, and estimating the failure phase. Whereas, future state prediction is a process of forecasting the remaining useful life (RUL) of machinery [2].

Rolling element bearings are very critical parts of rotating machines and they are used in a variety of applications. They are complex systems that consists of two rings and a set of rolling elements rotating in the races between the two rings. The rolling elements can be spherical balls, cylindrical rollers, needles, or barrel rollers. These elements are enclosed in a cage that provides equal spacing and prevents inter-element contact [4]. Once a bearing fails, failure of other adjacent components and machines are accelerated, and additionally may lead to increased downtime and high maintenance costs of the machinery. As shown in Fig. 1.1, the entire bearing service life can be typically divided into three phases: normal operation, gradual degradation, and accelerated degradation [5, 6, 7]. Although extensive studies have been carried out on bearing RUL prediction, most of these studies focus only on the final phase, i.e., the accelerated degradation phase [8, 7, 2]. Whereas, not much work has been carried out associated with long-term and multi-phase RUL prediction.

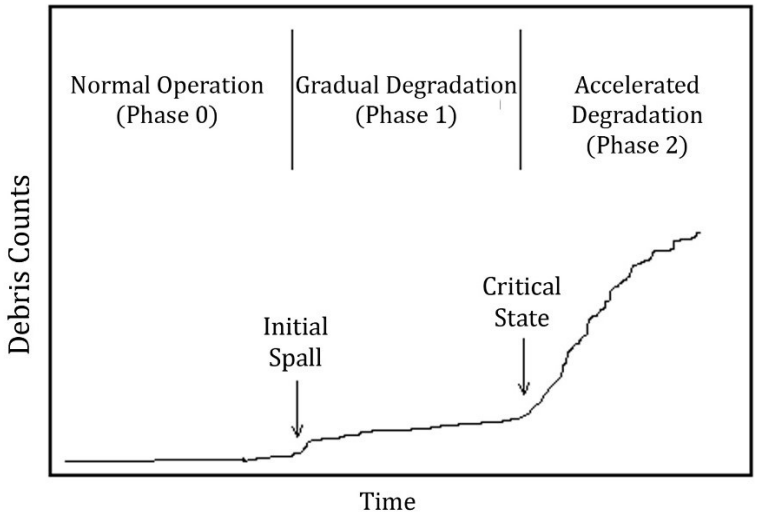

Fig. 1.1 Phases of degradation illustrated using the debris counts from oil debris monitors [3].

Some studies have shown that the characteristics (like size, and concentration) of the wear debris released into the lubricant are directly related to the fault severity in the machines [9, 10]. Hence, wear debris analysis has an immense potential to develop effective indicators for the health state of rotating machinery [11]. Inductive wear debris sensors or Oil Debris Monitors (ODMs) are widely used in the industry to detect the characteristics of these wear debris particles. ODMs are inductive devices that can detect the passage of debris particles through a magnetic field. The disruption to the magnetic field caused by the passage of the debris particle is processed to provide information about the approximate size and mass of the debris particles [12]. Over time, the ODM counts the number of debris particles and archives the information related to each debris particle. Fig. 1.1, shows that the debris counts released from bearing can be effectively used to differentiate the fault phases. Despite the excellent potential of ODMs in PHM systems of rotating machinery, very few studies explore its potential for RUL prediction [13].

Therefore, this study focuses on developing an intelligent PHM system for rolling element bearings that mainly uses the information from the ODMs for the long-term and multi-phase RUL prediction. The primary aim of this study is to perform the RUL prediction beginning from the gradual degradation phase (Phase 1) till the bearing failure.

1.2 Literature Survey and Motivation

This section introduces the theory and motivation behind the techniques proposed in this study while discussing the state-of-the-art solutions available in the literature.

1.2.1 Bearing Degradation Indicators from Wear Debris Analysis

For fault diagnosis of bearings, data-driven methods have been widely investigated [14, 15]. Data-driven methods attempt to perform the fault diagnosis of a machine using its historical data in machine learning techniques. These data-driven methods generally contain the following steps: (1) data acquisition, (2) construction of degradation indicators, (3) model training, and (4) estimation. Among these steps, constructing degradation indicators is the most important step because the accuracy of the fault diagnosis is directly related to the performance of these degradation indicators. In literature, synthesized degradation indicators are getting much attention not only in bearing PHM but also in battery PHM [16]. However, the main difference between battery and bearing PHM is that the health status of rechargeable lithium-ion batteries can be quantified and described by the battery capacity. However, for bearings, it is rare to discover a simple and direct health indicator to track the current health condition [17].

In general, for bearings, an ideal degradation indicator must be both monotonic and trendable. Monotonicity refers to a continuously increasing or decreasing trend of an indicator while the fault severity increases. Whereas, trendability represents the degree to which the indicator could be described by some functional form [19]. In reality, every bearing follows a unique degradation trajectory because bearing failure is a stochastic process. Fig. 1.2 depicts the evolution of a degradation indicator for various bearings. Fig. 1.2 (a) shows the indicators that have high monotonicity and trendability. While Fig. 1.2 (b) shows the indicators that have low monotonicity and trendability. Typically, non-monotonous degradation indicators fail to represent a machine's health state and they impair the performance of both fault diagnosis and RUL prediction process. In the data-driven fault diagnosis methods, these non-monotonous indicators are not very useful for training the machine learning models. Therefore, it is extremely important to construct degradation indicators that are both monotonous and trendable.

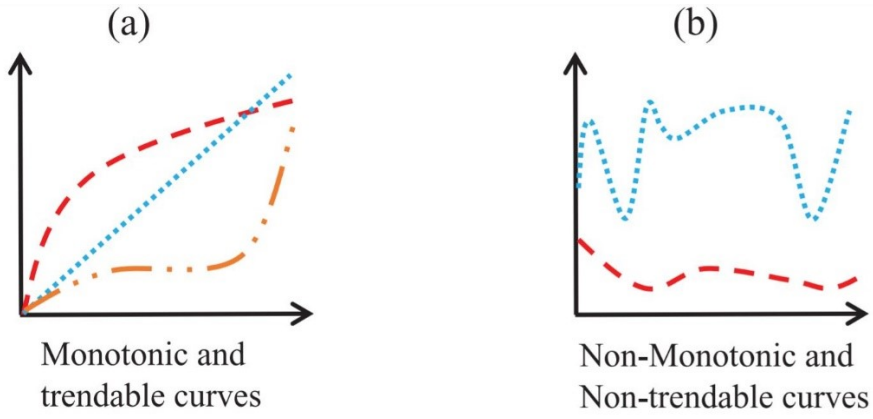

Fig. 1.2 (a) Examples of monotonic and trendable indicators, (b) examples of non-monotonic and non-trendable indicators [18].

In the literature, a lot of interest has been devoted to vibration-based degradation indicators. Specifically, frequency domain and time-frequency domain vibration-based indicators are preferred because, in theory, the bearing characteristic frequencies allow the fault signals to be separated from heavy background noises and other unwanted components in the frequency domain. More importantly, because vibration accelerometers are affordable and easily available. However, it is a difficult task to construct an ideal degradation indicator because of the stochastic nature of vibration signals. It requires complex signal processing methods to construct meaningful indicators [19]. Furthermore, vibration-based indicators have been used in several fault diagnoses and RUL prediction, but, many studies show that they are not very effective for long-term RUL prediction. Therefore, some studies suggest that it is preferable to use other sensors, such as thermal and ODMs [17, 20].

ODMs have some distinguished advantages compared to vibration-based sensors because they have a strong relationship to wear surface profile, long persistence of information, and strong anti-interference capacity [21]. As discussed earlier, ODMs measure and quantify the debris particles lost from the bearings. Thus, the wear debris information archived by the ODM has a direct relationship with the defect severity. Therefore, ODMs have a very good potential to represent the true degradation of the bearings and can it is useful in developing a physically meaningful indicator for the life of mechanical equipment. Further, ODM's strong anti-interference capacity greatly reduces

the probability of generating false alarms and thereby does not cause any complications for the maintenance crew. Although ODMs are comparatively more expensive than vibration accelerometers, their return-on-investment is better for heavy industrial machinery. Due to these reasons, using wear debris to investigate the health conditions of machines has attracted attention since the 1950s. And online debris monitoring has been applied in commercial engines, fighter engines, helicopter gearboxes, and wind turbines to increase system reliability and reduce maintenance costs [21]. However, very few studies, in the literature, have explored its potential for developing degradation indicators.

Paula Dempsey et al., conducted a series of experiments in [20, 22, 23, 24] using ODMs to monitor the wear of bearings and gears in real-time. They used the accumulated number (total count) of debris particles and/or the total mass of debris particles released as the degradation indicators. In [24], they proved that both of these indicators are directly proportional to the damaged area on the friction pairs. Additionally, in [20], they showed that the mass of debris particles, used as the indicator is comparable to using vibration-based indicators for condition monitoring of critical machines.

Further, in [25], a positive feedback-based wear process mechanism is used to develop a degradation model which captures the nonlinear relationship between debris generation rate and surface roughness. The rate of change of the total debris counts along with the total count is proposed as an indicator of mechanical failure while the peak of generation rate is used as the threshold for RUL prediction. In this study, the statistical characteristics of the debris particle distribution are explored to construct bearing degradation indicators.

1.2.2 Fault Phase Diagnosis using Ensemble Method

Bearing fault diagnosis is an essential part of a PHM system and has been widely studied for several years using signal processing approaches, and more recently using machine learning methods [26, 27]. Bearing fault diagnosis can be primarily used for three purposes: fault detection, phase estimation, and defect location estimation. Fault detection assesses whether the device is in an abnormal condition or a nominal condition, i.e., the bearing does or does not have a defect. After fault detection, assessing the appropriate fault phase

is necessary to deploy the relevant degradation models. Further, defect location estimation is performed to determine where exactly the defect has emerged; in the case of rolling element bearings, it is performed to know whether the defect is in the inner race, outer race, or on the rolling element [28]. No matter the purpose of the fault diagnosis, the steps for data-driven fault diagnosis methods remain the same (data acquisition, construction of degradation indicators, model training, and estimation). In this study, a data-driven fault diagnosis tool is developed that focuses exclusively on fault phase estimation.

Fault phase diagnosis is used to determine the initiation of various fault phases in the bearing degradation. For example, fault phase diagnosis can be used to identify whether the bearing is in the gradual degradation or accelerated degradation phase at any given time instant during its operation. Accurate fault phase diagnosis can be used to notify the operator about scheduling various maintenance activities; whereas in some cases, the operators may choose to prolong the machine operation by adjusting the operating conditions based on the fault phase. In this study, the results from the fault phase diagnosis are used for deploying and adjusting the degradation models in different fault phases.

In [29], a fault diagnosis method is proposed which performs both fault phase diagnosis and defect location estimation. The proposed method used empirical mode decomposition, linear discriminate analysis, and principal component analysis to extract relevant features from vibration data and used them to train a simplified fuzzy adaptive resonance theory map neural network. In [30], a fault-phase diagnosis method is proposed using piezoelectric sensors and convolutional neural networks to detect the healthy and various stages of damages in aluminum structures. The method achieves high accuracy even with very little data for training.

Estimating the correct time to begin the RUL prediction of the machine is an important part of bearing prognostics studies. The appropriate time to begin the RUL prediction is defined as the first prediction time (FPT) [31]. In the simple bearing degradation process in Fig. 1.3, two fault phases are shown, i.e., I: the normal operation (no fault initiation) and II: the accelerated degradation phase. The degradation index remains stable in Phase I, which means that the bearing is healthy. Once a fault is initiated,

the bearing enters the accelerated degradation phase (Phase II), where the magnitude of the degradation index increases rapidly. Typically, in Phase I, PHM systems are designed only to monitor the health state and once a fault is detected, the RUL predicting process is triggered [32]. The time when the PHM system is triggered to start RUL prediction, is the FPT, as shown in Fig. 1.3.

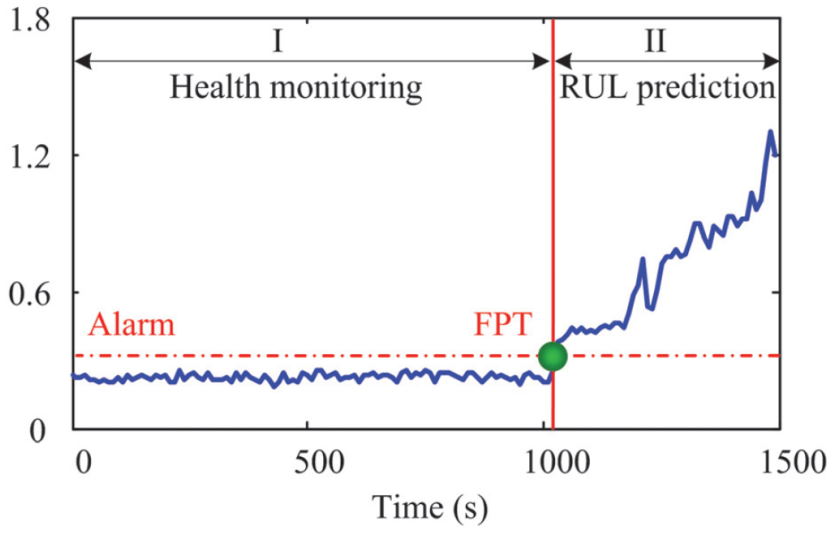

Fig. 1.3 Estimation of FPT in a simple two-phase degradation process [31].

It is important to select an appropriate FPT for all the model-based prediction techniques, so that the prediction algorithms like Kalman filters, Particle Filters (PFs), and Bayesian filters can perform accurately. If an inappropriate FPT value is chosen, interference noises (occurring during the phase transition) could be included in the prediction process or critical data may be excluded from the prediction process. In both cases, RUL prediction accuracy is affected negatively. The most appropriate FPT should be the time of fault initiation. However, detecting the fault occurrence is a very difficult task, because incipient faults of a bearing often occur randomly and the fault characteristics which show the fault initiation are very weak [26, 31].

Several studies in literature, subjectively select the FPT, whereas some studies use alarms for estimating the FPT. Engineering norm ISO 10816 and the approach based on the longest time constant of a machine and statistical properties of a candidate baseline are some of the reported techniques for setting the alarm. In these approaches, alarm limits are

set based on the statistical properties of a large number of mechanics operating in similar conditions [33, 34]. However, these alarm-based methods are not very accurate in identifying the FPT when the operating conditions of the system are changed.

In [31], a simple statistical FPT selection approach is proposed based on the standard deviation of the kurtosis of the vibration signature, where once the kurtosis crosses the three-standard deviation interval, it is considered as the FPT. The same approach has been adopted in [35]. A similar approach is proposed in [36], where a two-standard deviation interval of the kurtosis is used to select the FPT.

So far, many studies suggest using information from multiple types of sensors for the detection and diagnosis of bearing defects [29]. These sensors can be broadly classified as acoustic, temperature, wear debris detection, and vibration analysis. As discussed earlier, vibration analysis is the most used method in the field as the accelerometers are affordable and easily available. In acoustic emission, the measurement of transient sound waves emanating from bearing due to rapid local stress redistributions is obtained and analyzed for acoustic signal properties. Further, temperature, pressure, and current are monitored to detect the changes in bearing conditions. In the training step of a data-driven fault diagnosis, the information from these sensors is used as features for training the machine learning models.

The most effective set of features should only be used for training to get a higher diagnostic accuracy [37]. Therefore, a feature selection step is introduced to eliminate noisy, redundant, or irrelevant features before training the machine learning models. It also helps in considerably reducing the training time by reducing the number of features [38]. In many studies, feature selection has been considered as a dimensionality reduction problem where techniques such as Principal Component Analysis, Multidimensional Scaling, Factor Analysis, Projection Pursuit, and other techniques are used. These types of methods commonly generate dummy features with lower dimensions than the original set, however, these reduced features set lack physical significance and meaning [39].

Thus, to overcome the issues with dimensionality reduction, feature ranking methods such as Fisher score, ReliefF algorithm, Wilcoxon rank, Gain ratio, Memetic

feature selection, Chi-square, and Information gain have been used in the literature to select appropriate features and improve precision in bearing fault diagnosis [40, 38].

In the last decade, various machine learning techniques have been employed in fault diagnosis studies. Machine learning techniques like support vector machine (SVM), Extreme Learning Machine, Random Forest (RF), and K-Nearest Neighbor (KNN) have obtained good performance as classification techniques. Compared with extreme learning machines, SVM has better classification ability for small datasets. However, SVM also has certain limitations which have severely restricted its development in the fault diagnosis field [41]. Whereas, KNN and RFs have performed well and have been used in several studies for fault diagnosis of bearings [42, 43, 44, 45].

RF classification method was proposed by Breiman in 2001 [46], based on the concept of the bagging and regression tree techniques. Compared with the traditional decision tree algorithm, RF has manifested robust classification performance in solving high-dimensional and small-sample problems [39]. Additionally, the RF algorithm has also inherited the high interpretability of the tree-based models [41]. The KNN is a simple and efficient classification algorithm and it has been studied extensively in many practical applications. The KNN method has very few hyper-parameters that need to be selected; it is highly effective for statistical pattern recognition [45].

Although these techniques have good classification capability, the performance of these techniques can be improved significantly using ensemble learning strategies. This is highly sought after in the case of critical applications with limited failure dataset. Ensemble learning is also known as a multi-classifier system or committee-based learning system. It uses multiple base classifiers to obtain better classification performance [47]. There are several widely-used ensemble strategies, such as majority voting, Bayes optimal classifier, bootstrap aggregating (bagging), boosting, stacking, winner-take-all, and other user-defined strategies [48]. In this study, more ensemble-based data-driven techniques are explored to perform fault phase diagnosis and FPT selection.

1.2.3 Adaptive Particle Filtering for Model-Based RUL Prediction

One primary aim of a PHM system is to predict the degradation evolution of mechanical equipment based on current and historical operational data. Developing an industry-viable PHM system is a challenging task because the mechanical systems are complex in design and they typically have non-linear degradation patterns. Compared to the offline degradation prediction methods, online (or inline) methods are more useful in the industry since they can be implemented directly on the real-time health conditions of the system in service [49, 50].

In general, the modern bearing prognostic approaches are categorized into data-driven approaches and model-based approaches [51, 52, 53, 2]. Data-driven methods attempt to predict the degradation process of a machine from historical data using machine learning techniques. These methods rely on the assumption that the statistical characteristics of data are relatively consistent. Therefore, the prediction accuracy of data-driven methods depends not only on the quantity but also on the quality of the historical data [52, 2]. One of the major drawbacks of the data-driven approach is that it is typically difficult to collect the required amount of good quality failure data. Because firstly, industrial machines will not be allowed to run into failure conditions due to the lethal consequences. Secondly, almost all mechanical failures occur slowly, sometimes taking months or even years, which makes it uneconomic to collect failure data [54]. On the other hand, the model-based approach typically uses a physics-based model to describe the degradation of the bearings. The Paris crack growth model [55], Forman's crack growth model [56], and spall progression model [57] are some of the well-accepted physics-based models for bearing degradation. In the model-based approach, the primary aim is to estimate the model parameters from the real-time measured degradation data of the system. Although in theory, model-based approaches have several advantages, data-driven approaches are preferred in recent times; this is because the available algorithms used in the existing model-based methods are not very efficient in accurately estimating the model parameters, as they are too sensitive to noise levels and temporary fluctuations in the measured degradation data [54]. Therefore, there is a need to improve the efficiency of the existing algorithms used in model-based approaches.

Some of the most used algorithms in model-based methods are Bayesian framework methods [58, 59], Kalman filters [60], and PFs [52, 5]. In particular, the PF method is considered the most useful for nonlinear degradation systems [51, 2]. PF can be explained as a numerical approximation method to achieve Bayesian inference using the sequential Monte Carlo method. It uses a set of numbers and their associated weights (known as particles) representing probability densities to tackle nonlinear and non-Gaussian behaviour in the dynamics of a physical system [61]. PFs, in general, has four steps: state estimation, weight estimation, particle resampling, and model parameter estimation.

The existing studies in the PHM field for machine fault prognosis focus on directly applying the standard PF for RUL prediction in the final phase, i.e., accelerated degradation phase of the bearings. These PFs are generally not viable for real-time prediction because, they tend to fail in scenarios where there is high noise, varying operating conditions, sudden trend changes, anomalous regions, and other multi-source uncertainties in the measured degradation data [62]. To avoid encountering these issues, some studies developed efficient and monotonic health indicators which are smooth and do not have anomalies or fluctuations [63, 2], whereas other studies tend to develop techniques that help to avoid the fluctuations and phase change regions [7, 64, 34]. Therefore, almost all research works start the remaining useful life prediction only after the accelerated degradation phase begins [7]. Hence, there is a need to improve the existing PF algorithms so that they can directly be applied for real-time degradation prediction.

The applications of the PFs on machine fault diagnosis and prognosis have been extensively seen in the literature [2, 5-7, 13, 31, 51-53]. However, the main drawback of the standard PF algorithms is that they cannot adapt to the trend changes and fluctuations originating in the degradation data [65]. This limitation can be attributed to three problems in the PF algorithms: 1) particle degeneracy, 2) particle impoverishment and 3) initial range dependency. Particle degeneracy is a phenomenon in PFs where the weights of the particles converge to a few particles while other particles will have insignificant weights. Thereby, a large amount of the computational cost is allocated to updating the particles whose contribution to the model parameter estimation is almost negligible [66]. Particle impoverishment is the loss of particle diversity due to the recursive application of the

resampling step in PF [67, 68]. Initial range dependence is seen in almost all the PF algorithms in the literature because PFs are completely dependent on the initial range of model parameter estimates which are defined during the initialization of the PF. This is an important issue to be tackled because the range of values for the model parameters throughout the test is not accurately known in many practical applications. There are several methods proposed in the literature that deal with improving the particle degeneracy and particle impoverishment problem of the PF algorithm, however very few studies in the literature focus on solving the problem of initial range dependence.

In [69], a methodology is proposed that consists of drawing samples uniformly from a random measure. This is the basis of the sequential importance sampling method which is currently used in standard PFs. More efficient resampling methods named stratified resampling [70], systematic resampling [71], and residual resampling [72] have been proposed to solve the particle degeneracy problem; however, they are shown to increase the severity of particle impoverishment problem [73]. In [74], enhanced PF is proposed to tackle the particle impoverishment problem by introducing a new importance density function. Where a normal distribution based on the state estimation is used in the importance density function to estimate weights of the particles. A similar normality approximation of the distribution has been explored in the class of the Gaussian PFs in [75, 76, 77] which proved to work well for nonlinear applications. However, the particle diversity maintained in these methods is adequate only to adapt to small fluctuations in the degradation data.

There are few adaptive PF techniques also proposed in the literature that try to address the problems in PFs. The first adaptive PF was proposed in [78] and it uses the Kullback–Leibler distance to approximate the error in the state prediction. This method allows choosing the number of particles based on the uncertainty of state estimate. The adaptive PF in [79] utilizes an adaptive sampling scheme to estimate the required number of particles at the subsequent time step. The adaptive PF proposed in [65] uses a statistical measure to improve the propagation function for error estimation, whereas a variable bin function is explored in [66]. However, these adaptive PF techniques have focused largely on adaptively estimating the number of particles used in each time step rather than on the

range of the particles. In addition, they have high computational costs because they determine the number of particles using multiple resampling processes at each time step. In [80] adaptive PF is explored in the PHM field. It uses a self-evaluation method to detect low-weight particles. Then an adaptive weight adjustment technique is used to find the high-likelihood region on the posterior PDF using a dynamic feedback approach. However, this adaptive PF addresses only the issue of particle degeneracy and has not been tested in scenarios with adverse trend changes in the degradation data. Thus, more work needs to be performed to improve the PFs to adaptively find the range of model parameters in the case of adverse trend changes and noisy degradation data.

1.2.4 Multi-Phase Degradation Model for Long-Term RUL Prediction

Model-based prognostic approaches typically use a single physics-based degradation model to characterize the entire service life of the bearings. As described earlier, the Paris crack growth model, Forman crack growth model, and spall progression model are some of the accepted physics-based models for bearing degradation. These models work well in scenarios where the degradation trend is monotonic. However, in most cases, the degradation trend of bearings has a lot of magnitude variations due to the stochastic nature of bearing spall progression. Therefore, some studies use stochastic models such as the Wiener model, Inverse Gaussian model, and Gamma process model to characterize the entire bearing service life [81]. The main advantage of these stochastic models is that they can be modified to characterize the non-monotonic degradation processes of bearings. Therefore, they have been successfully used for RUL prediction studies in the literature [82, 83, 84, 85, 86].

However, the approximation that the entire service life of bearings can be characterized by a single degradation model has led to restricted accuracy and robustness for long-term RUL prediction [5]. Therefore, in [5], more than one degradation model is used where one model corresponds to each degradation phase of bearing. For example, a linear model is used for the gradual degradation phase and an exponential model is used for the accelerated degradation phase. Even though the results have shown good long-term

multi-phase prediction, the accuracy of the method completely depends on the algorithm rather than on the degradation model.

Therefore, to deal with this issue, attempts have been made to develop a single degradation model which can accommodate the multi-phase degradation process. In [87], a two-phase degradation process model was developed and applied to the degradation index of plasma display panels. They proposed a hierarchical Bayesian change-point regression model to fit the two-phase degradation patterns and estimate the RUL of the population. Further in [88], a multi-phase degradation model with jumps based on the Wiener process is formulated and RUL estimation is demonstrated on gyroscopes. A model identification method with off-line and on-line parts based on an expectation-maximization algorithm and Bayesian rule is proposed. Subsequently in [89], a similar, multi-phase degradation model with jumps based on the Wiener process is developed, where the drift and the diffusion parameter of each phase are treated as hyperparameters, and they are estimated based on the expectation-maximization algorithm. Apart from the Wiener process models, no other degradation model has been considered for multi-phase degradation processes so far.

The Gaussian process model is an example of a probabilistic non-parametric model that provides information about prediction uncertainties that are difficult to evaluate in nonlinear parametric models [90]. Initially, it was used in the prediction of immunological diseases [91]. Later, it was adapted for use in predictive control process studies in [90]. However, no study has considered a Gaussian process model for a multi-phase degradation application. Most importantly, so far, no multi-phase degradation models have been proposed specifically for bearing RUL estimation. Therefore, more work needs to be performed in this aspect.

1.3 Objectives

To tackle the aforementioned issues, this thesis presents an intelligent PHM system that performs both fault-phase diagnosis and long-term multi-phase RUL prediction of rolling element bearings. The study primarily uses the information obtained from ODMs. The PHM system, shown in Fig. 1.4, broadly contains two sub-systems, the fault phase

diagnosis and the RUL prediction, and these sub-systems further have internal modules. The output of the fault-phase diagnosis is used to activate and update the RUL prediction process. Since the bearing phase change is a stochastic event, in the gradual degradation phase (Phase 1), the RUL prediction gives a probabilistic range, using an upper and lower limit, for the bearing remaining life. Whereas in Phase 2, the degradation model is updated and the mean RUL is obtained. The contributions of this work are summarized below:

1. Three novel bearing degradation indicators are proposed using the wear debris characteristics obtained from ODMs.
2. A data-driven fault phase diagnostics method has been developed using an ensemble of KNN and RF classifiers.
3. An enhanced adaptive PF algorithm is proposed to address the initial range dependency problem of the standard PF. This adaptive PF is used for real-time model parameter estimation even if there are high fluctuations and trend changes in the measured degradation data.
4. A novel physics-based degradation model is proposed based on the mixture of the Gaussian process and Weibull failure rate models. The proposed model accurately characterizes the multi-phase degradation of the bearings.

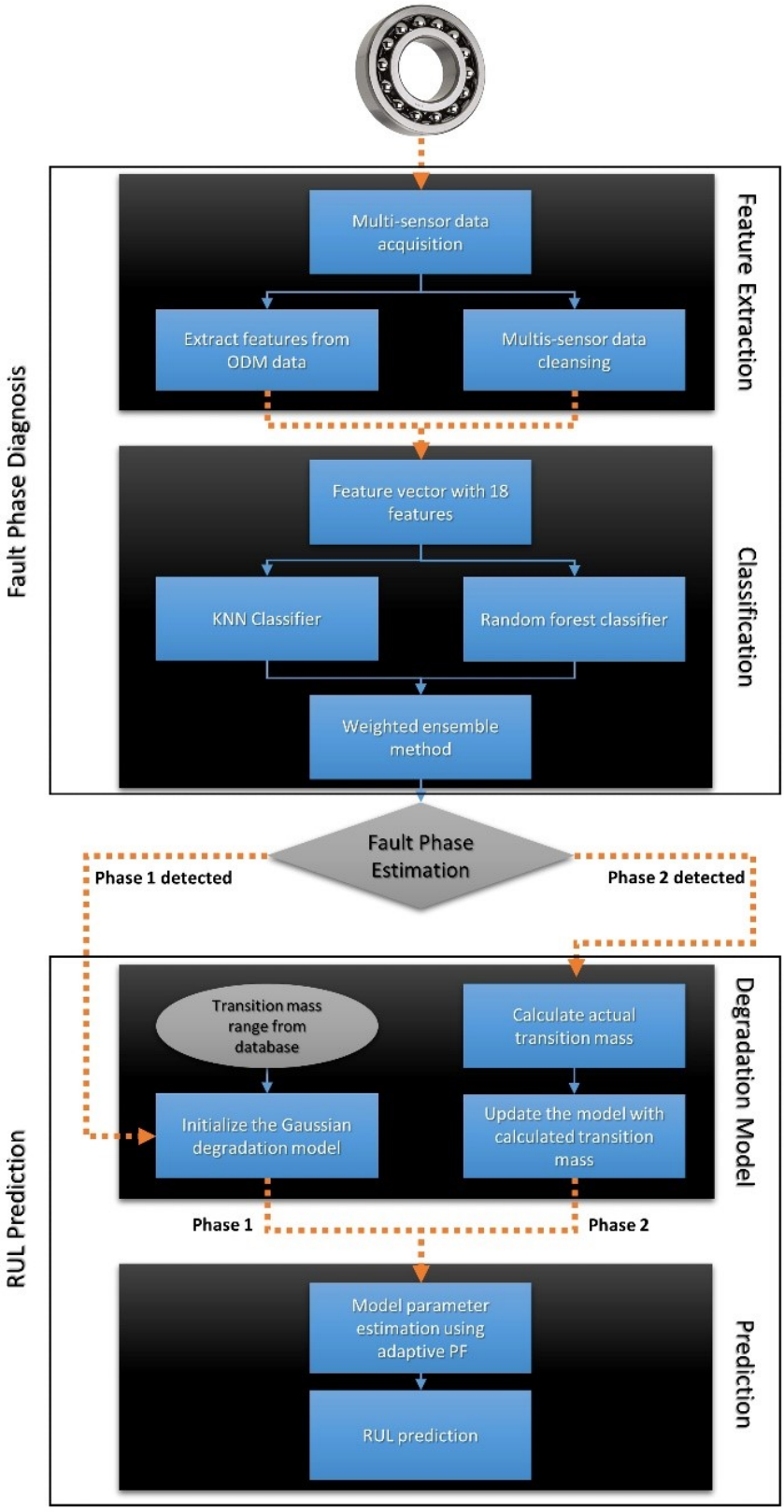

Fig. 1.4 Schematic of the intelligent PHM system.

1.4 Experimental setup

The schematic of the experimental setup employed in this research is shown in Fig.1.5. The experimental bearing is driven by a 10-kW motor and a lubrication oil is sent into the bearing housing to lubricate, clean, and cool the bearing. A supply pump on the outlet of the main tank helps to pressurize and supply the lubricant oil to the bearing housing. The lubricant is filtered by a 1 μm cartridge filter downstream of the supply pump. The temperature, pressure, and flow rate of the lubricant are monitored before the filtered oil enters the bearing housing.

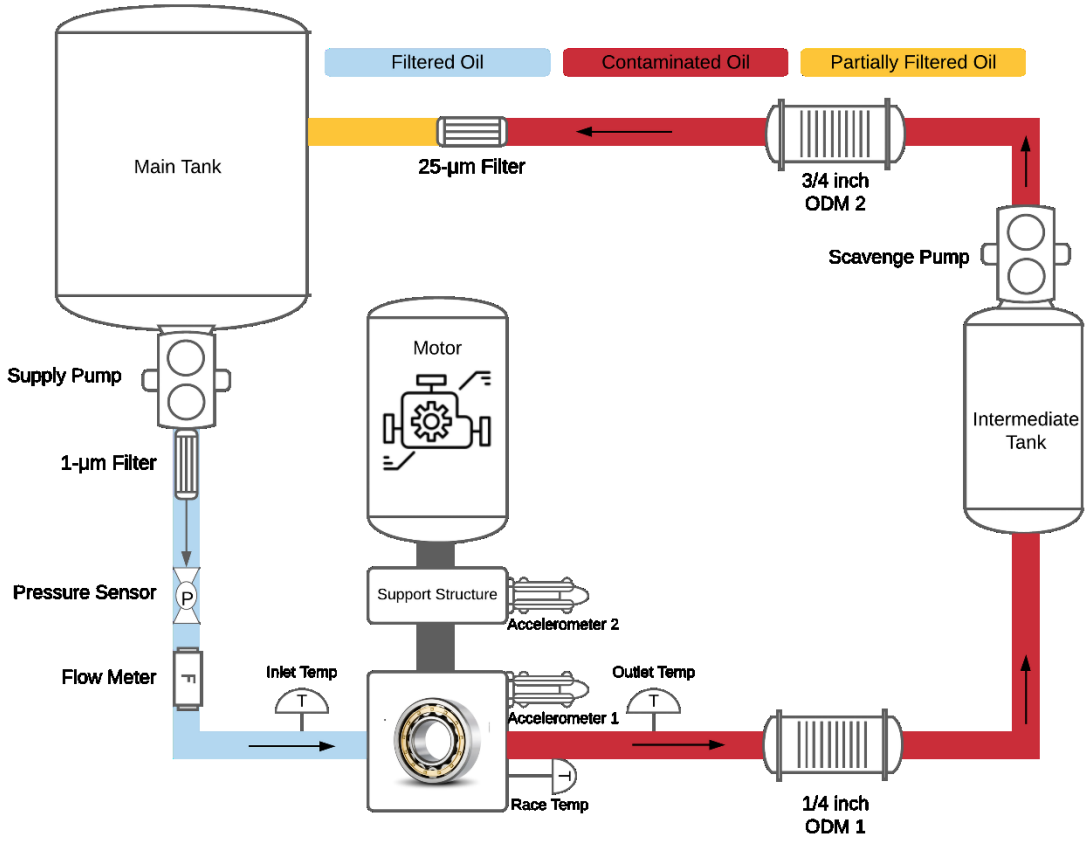

Fig. 1.5 Schematic of the experimental setup.

The motor used in this experiment allows us to monitor its torque and rotational speed. The bearing housing contains the experimental bearing; the outer race temperature and the vibration signature are captured using a temperature sensor and accelerometer

placed on the bearing housing. Additionally, a shaft support structure is available to reduce the lateral motion of the motor shaft. The vibration signature of the support structure is also monitored using Accelerometer 2 to detect if the failing experimental bearing induces any external vibration. Both the vibration accelerometers have a very high sampling frequency of 48KHz. The axial and radial load can be varied on the experimental bearing by means of pneumatic load cylinders. Two strain gauge load sensors are used to measure both axial and radial loads on the bearing.

In the downstream of the bearing housing, firstly, a temperature sensor is used to monitor the temperature change of the lubricant. Next, a 1/4-inch diameter ODM is placed to detect and quantify the debris particles released from the experimental bearing. Next, an intermediate tank is placed to increase the oil flow rate and to satisfy the flow rate requirements of the ODM. The scavenge pump in the outlet of the intermediate tank re-pressurizes the lubricant and sends it further downstream. Another ODM with a 3/4-inch diameter is available downstream of the intermediate tank to detect the debris particles. The ODMs used in this setup have different sensing ranges; and they are MetalSCANs manufactured by Gastops Ltd. ODM 1 has a sensing range from 65 to 350 μms, while the sensed mass ranges from 1 to 190 μgms for ferrous particles. Meanwhile, ODM 2 can sense debris particles from 250 to 1000 μms, while the sensed mass ranges from 90 to 4400 μgms for ferrous particles. Both of these ODMs categorize every debris particle into one of the 16 bins. A detailed explanation about the working of the ODMs is presented in Section 2.1. Finally, the contaminated oil is passed through a 25 μms cartridge filter before sending it to the main tank. Since the primary focus of this study is to validate the effectivity of the ODMs for RUL prediction, both the vibration accelerometers are used only to calculate the Root Mean Square (RMS) values of the vibration signatures. Overall, all the sensors were used to capture the data at one-second intervals.

A total of five run-to-failure bearing spall propagation tests were conducted on SKF 7208 BEP roller bearings. Before beginning the experiments, localized defects are initiated on the inner race of the bearings as a number of minor dents using wire electrical discharge machining. The spalls are allowed to propagate inside the bearing housing and the wear debris along with the vibration are monitored till the bearing fails. The motor speed, axial

and radial loads are varied intermittently during these five tests. The tests were concluded if the vibration RMS amplitude (ODM 1) increases more than 14; or if the spall length exceeded 2 times the rolling element length.

1.5 Thesis Outline

Chapter 2 presents a detailed explanation of the working and application of the inductive ODMs and also proposes three novel degradation indicators extracted from the information archived by the ODMs. Two of the indicators are derived from the statistical features of the debris particle distribution. Whereas, the third indicator is based on change-point analysis technology. The validity of these degradation indicators is demonstrated using the experimental data.

In Chapter 3 a data-driven fault diagnosis technique is presented which helps in instantaneously detecting the fault phase of the bearing spall propagation. This technique uses multi-sensor information to train two machine learning classifiers and further uses an ensemble method to integrate the classification of both the classifiers to obtain better accuracy. The validity of the technique is demonstrated using the data from three spall propagation experiments to train the classifiers and the rest of the two experiments for testing.

An enhanced adaptive PF is proposed in Chapter 4 which attempts to solve the initial range dependency problem of the standard PF. The superiority of the proposed enhanced adaptive PF over the standard PF and the Bayesian method is demonstrated using a series of simulated experiments.

Subsequently, in Chapter 5, a multi-phase degradation model using a mixture of Gaussian process and Weibull model is proposed to characterize the bearing degradation. The assumptions and the validations of the proposed model are explained in detail.

In Chapter 6, the subsystems and the modules in the proposed intelligent PHM system are explained in detail. Further, long-term multi-phase RUL prediction is performed on the experimental data to validate the proposed PHM system. The results of the accurate RUL prediction on two experimental tests are shown.

Finally, Chapter 7 presents the concluding remarks and highlights the main contributions of the work.

Chapter 2: Bearing Degradation Indicators from Wear Debris Analysis

2.1 Overview

Constructing degradation indicators is the most important step in data-driven fault diagnosis techniques. Due to the stochastic nature of bearing degradation, it is very difficult to obtain ideal degradation indicators which are both monotonic and trendable [19]. Some studies suggest that it is preferable to investigate wear debris analysis as the ODMs have some distinguished advantages compared to thermal and acoustic sensors because they have a strong relationship to wear surface profile, long persistence of information, and strong anti-interference capacity [21]. Studies reported in [22, 23, 24, 20] used the accumulated number (total count) of debris particles and/or the total mass of debris particles as the degradation indicators. Further, in [25], an indicator is developed using both the total mass of debris particles and the rate of debris mass generation for RUL prediction. Although these indicators accurately represent the overall degradation of the bearing, these indicators do not consider the detailed characteristic information of the wear debris particles.

Change points are defined as the time instants where a significant increase in the amplitude or rate of the degradation index is witnessed. Identifying the change points in the bearing degradation index will help to disclose important information about the actual degradation of the bearing. Previous works in [18], [92], and [93] demonstrated the effectivity of change point detection methods for efficient bearing condition monitoring.

In this Chapter, we firstly discuss the working and application of ODMs in bearing condition monitoring. Later, we explore the wear debris characteristics obtained from the information archived by the ODMs. Finally, we use a novel Regression-based Online Change Point Detection (ROCD) technique to construct a wear debris-based bearing degradation indicator.

2.2 Wear Debris Analysis

2.2.1 Operating Principle of ODMs

ODMs or inductive wear debris sensors, shown in Fig. 2.1, are flow-through devices that are installed anywhere between the bearing oil return lines and the oil filter, allowing the contaminated oil flow to pass through it. It detects the passage of metallic debris particles through a magnetic field that surrounds the bore of the sensor. The sensor operates by monitoring the magnetic field; once there is a signal disruption caused by the passage of the debris particle, it is detected and then processed to provide information about the type, size, and mass of the particle [12]. A typical ODM consists of two balanced magnetic field coils and a central sensing coil through which the lubricant along with the debris particles flow, as shown in Fig. 2.2.

Fig. 2.1 Illustration of ODMs from Gastops (MetalSCAN™ 1000) [94].

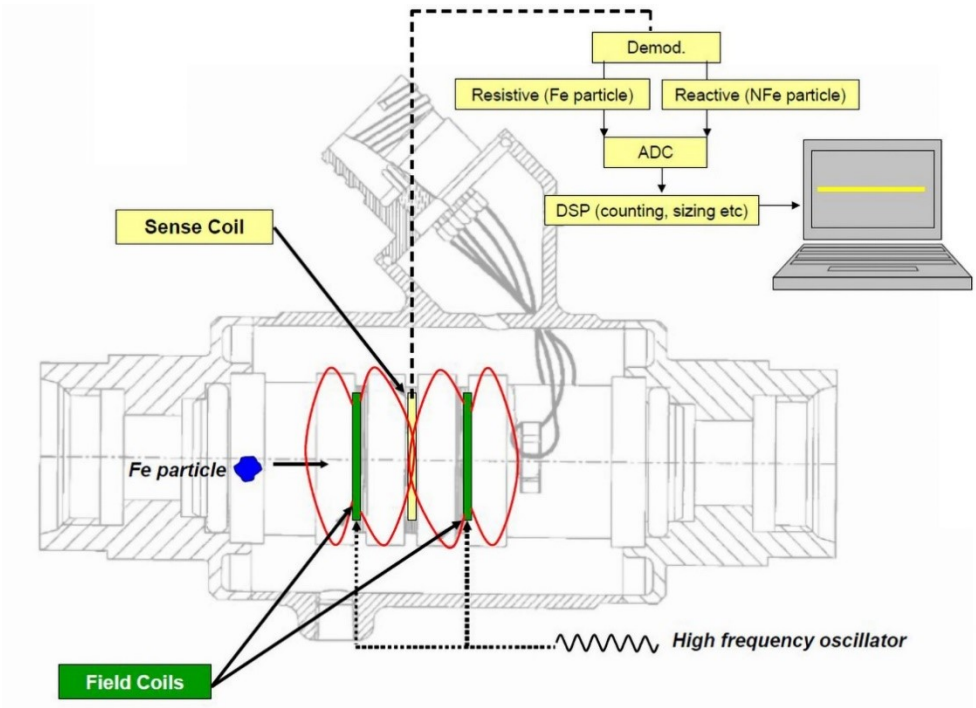

Fig. 2.2 Cross-section of typical ODM when no debris particle is detected [12].

The two field coils are oppositely wound and they are supplied with a high-frequency AC voltage that induces opposed magnetic fields. Therefore, there exists a balanced magnetic field at any point inside the sensor because the alternating magnetic field from both field coils cancels each other. In the stable state, i.e., without any debris particle present, no signal is induced into the sensing coil, as shown in Fig 2.2. However, when a debris particle passes through the sensor it disturbs the magnetic field and induces an output signal in the sensing coil, as shown in Fig. 2.3 [12, 95]. The amplitude and phase of the output signature are used to identify the size and nature of the debris particle. The amplitude of the signal is proportional to the mass of the particle for ferromagnetic metals and to the surface area of the particle for conductive non-ferromagnetic metals.

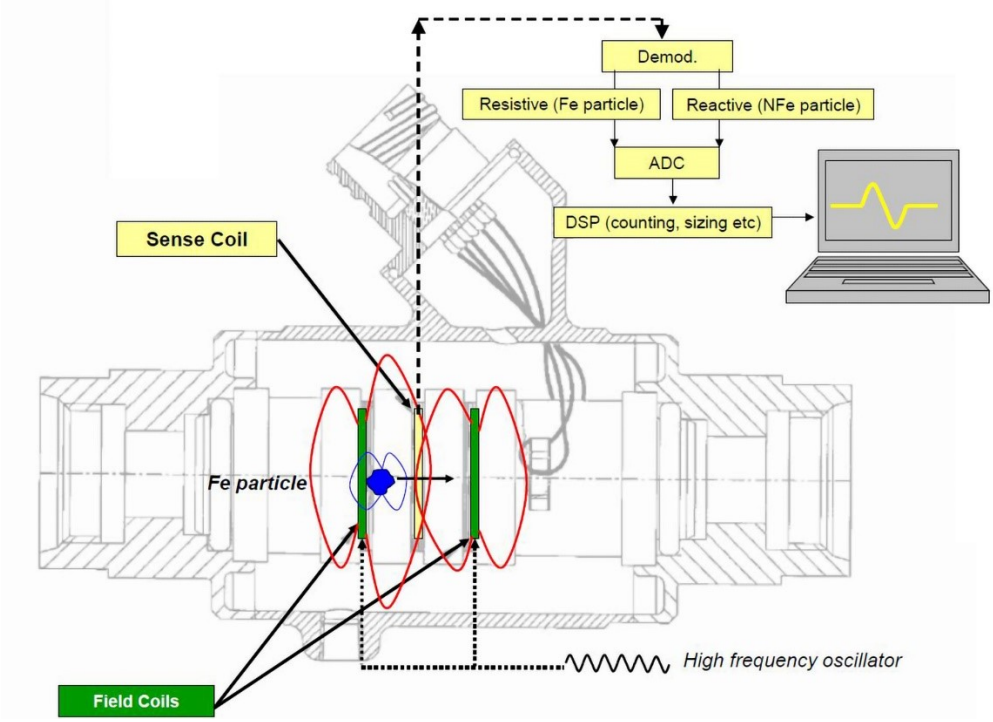

Fig. 2.3 Cross-section of typical ODM when in an unbalanced state [12].

As shown in Fig. 2.4, the phase of the signal for nonferromagnetic metals is opposite to that of ferromagnetic metals; this allows us to make a distinction between the types of debris particles. Additionally, signal conditioning using a threshold algorithm is used to categorize the debris particles based on the particle's size. Multiple size categories can be configured which allows the tracking of the distribution of debris size. The ODM used in this study categorizes each debris particle into one of sixteen bins. The minimum size particle detectable by the ODM depends on the sensor bore size [96].

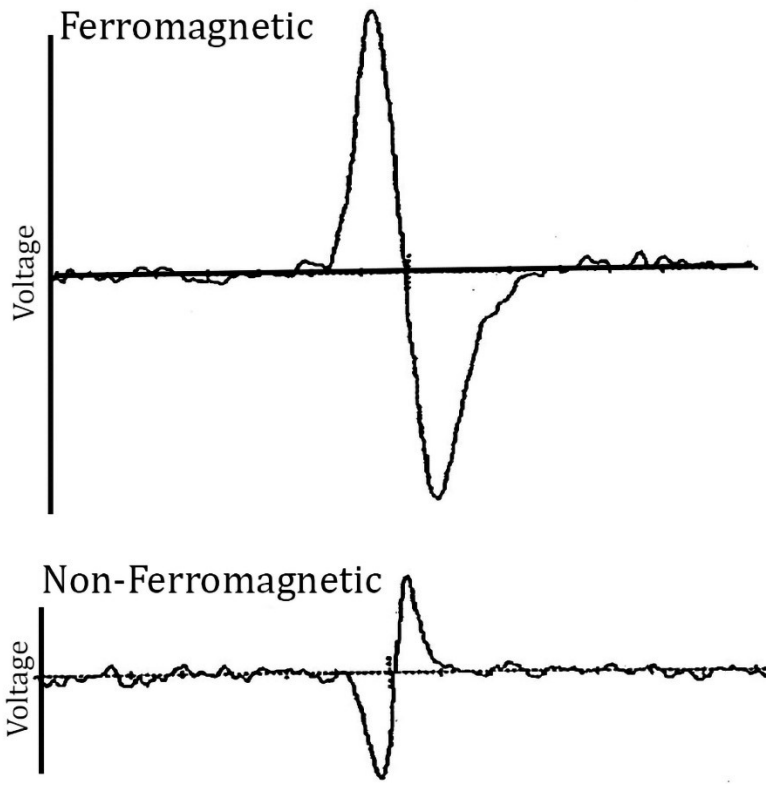

Fig. 2.4 ODM output signal examples [22].

2.2.2 Bearing Condition Monitoring using ODM

ODMs are used for bearing condition monitoring based upon the contention that a failing bearing generates sufficient and detectable quantities of debris that are washed downstream of the bearing into the lubricating oil system. In some circumstances, it is possible that some debris may not make it all the way to the ODM due to the internal architecture of the lubrication system. In practice, however, it has been found that the quantity of "trapped' debris is a very small fraction of the total debris that is found in the ODM [96].

A study in [96] demonstrated, with a series of experiments, that ODM data, such as size and quantity, have a direct correlation to the severity of damage of the bearings. Fig. 2.5 contains data from one of the bearing spall propagation tests conducted in this work. It shows how debris particles are distributed in the various bins from 1 to 16. These bins are classified based on the average size and mass of the debris particles. Apart from bearings, ODMs have also been applied to areas such as aviation [96], wind turbines [95, 97], and general industrial engines.

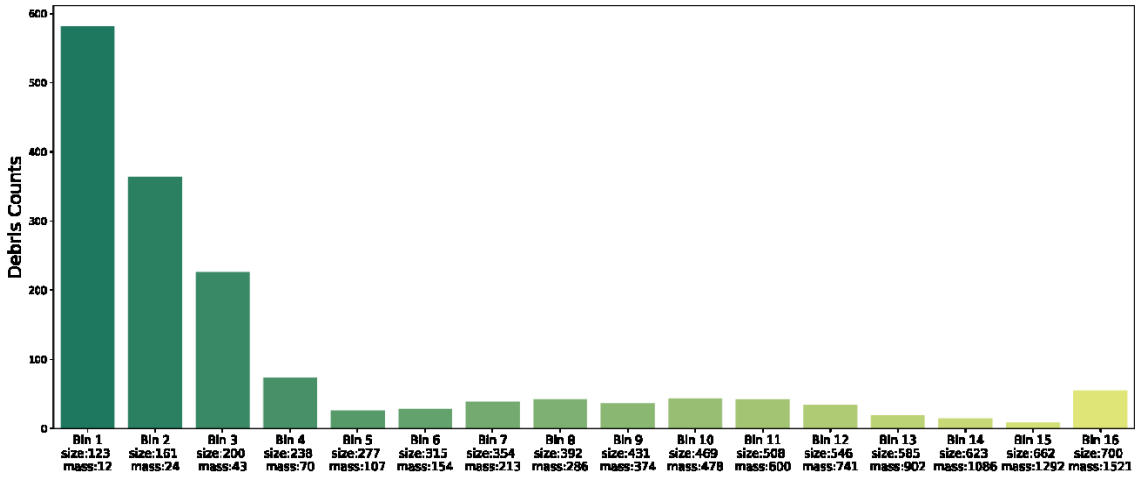

Fig. 2.5 ODM debris counts data from a spall propagation test binned as per the average bin size (in micro meters/μms) and mass (in micro grams/μgms).

2.3 Wear Debris Characteristics

Many studies in literature, found that various wear debris characteristic features such as concentration (number of debris particles), size, morphology, and composition, relate directly to the severity of the faults in mechanical equipment [21]. For instance, [22, 24] show that both debris concentration and size increase while defect size increases; and hence concentration and size directly correlate to fault severity and wear generation rate [98]. Further in [99], debris size relates to different types of wear mechanisms. Meanwhile, in [100], it is shown that debris morphology also depends on the wear mechanism of the mechanical equipment. Finally, in [101, 102], the relationship between debris generation rate and wear severity is summarized as shown in Fig. 2.6.

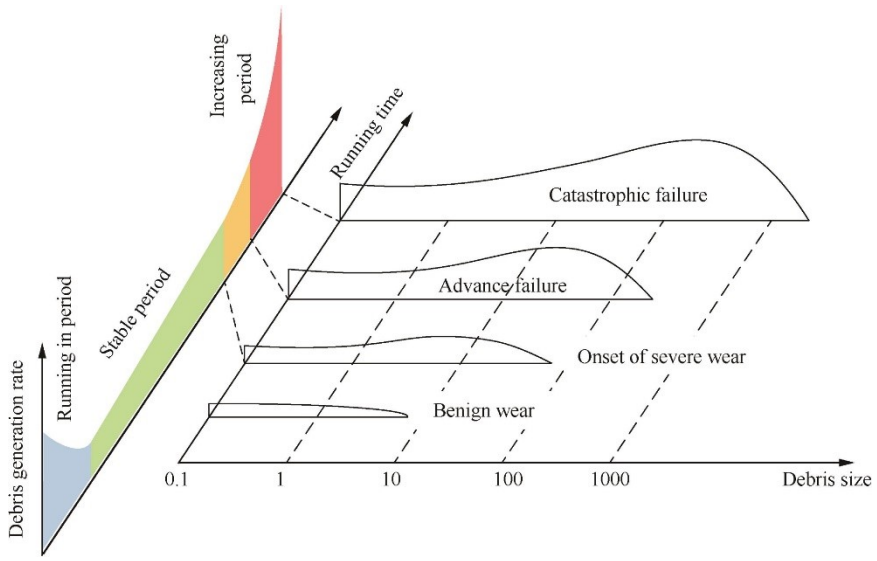

Fig. 2.6 Relationship between debris generation and fault severity [21].

As discussed earlier debris characteristics such as size and concentration can be efficiently monitored using the commercially available ODMs. However, monitoring the characteristics such as morphology and composition of the debris released from the bearings is still a challenge as the methods to monitor these characteristics require optical microscopy, and scanning electron microscopy. These monitoring methods are not affordable in the real-time working environment of the machines.

Fig. 2.6 summarizes that the size and concentration of the debris particles vary as the machine degrades from benign wear to catastrophic failure. Moreover, as the fault severity increases the mean size of the wear debris increases. However, Fig. 2.6 is only an idealized distribution (i.e. not based on actual data) and is not common to all types of mechanical equipment. Despite this being identified as an idealized distribution, this plot has been reproduced many times without much scrutiny [12]. In general, the main insight from this distribution makes us understand that there is a complicated relationship between debris characteristic features (size and concentration) and fault severity. In the case of the bearings, the real distribution depends a lot on the types of bearing, material, loads, and location of defect (outer/inner race).

In the literature, so far, no study has tried to use a relationship between the debris characteristic other than concentration for condition monitoring of equipment. Therefore, in this study, the statistical relationship between wear debris characteristics is used for condition monitoring and found that it helps distinguish bearing fault severity.

2.3.1 Wear Debris Characteristics of Experimental Data

As discussed earlier, ODMs quantify the debris particles by categorizing them into specific bins based on the size of the particles. Fig. 2.7 shows the debris counts in their respective bins at different time instances of an actual bearing spall propagation test. The curve in these plots shows a fitted distribution of the particle counts for each of the time instances. In general, we see that, as the fault severity increases (i.e., from plots with time 100 to 600 minutes), the shape of the fitted distribution changes a lot. For example, the fitted distribution at 600 minutes is starkly different from 100 minutes.

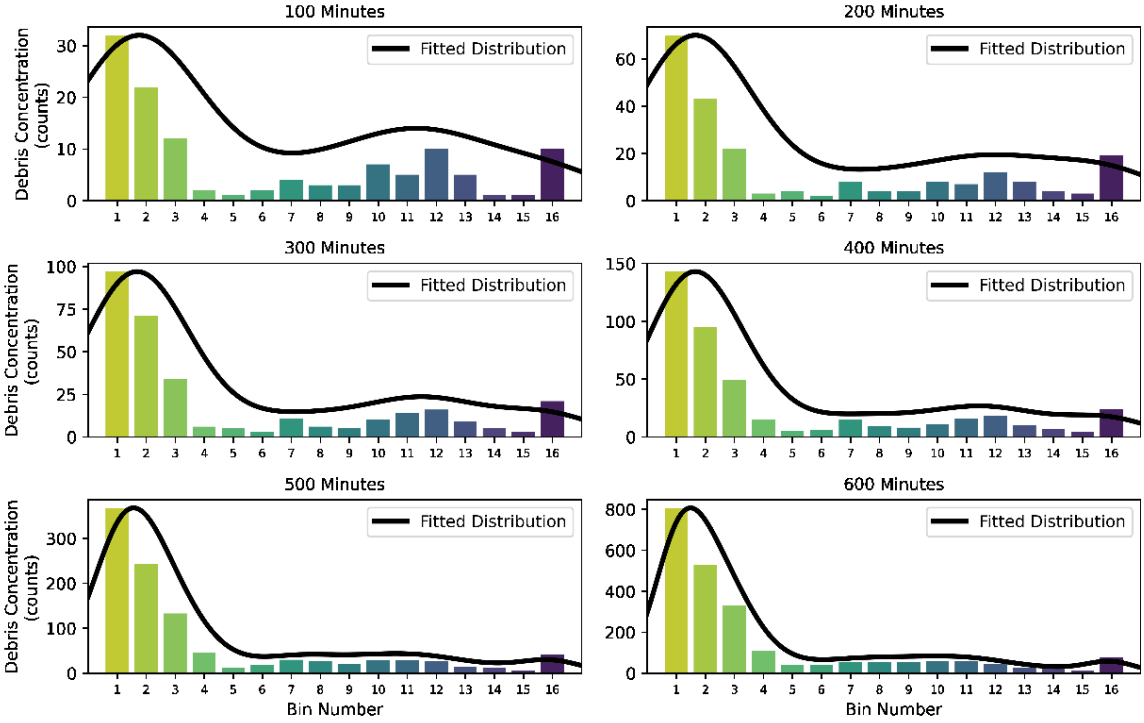

Fig. 2.7 Evolution of the debris distribution as the fault severity increases.

Meanwhile, Fig. 2.8 shows a 3D plot where the fitted distribution varies during the entire spall propagation test. The fitted distribution in this 3D plot is the particle density distribution instead of the count distribution. The spall propagation test shown in this 3D plot has a degradation phase change from gradual degradation to accelerated degradation around 350 minutes, after which the debris concentration of the smaller particles (bins 1 to 6) increases rapidly. Meanwhile, soon after the accelerated degradation phase begins, the debris concentration of the larger particles (bin 15 and 16) also increases after 400 minutes. This increase in the larger particles manifests as the double bell curve (see the plot from 400 minutes to 650 minutes).

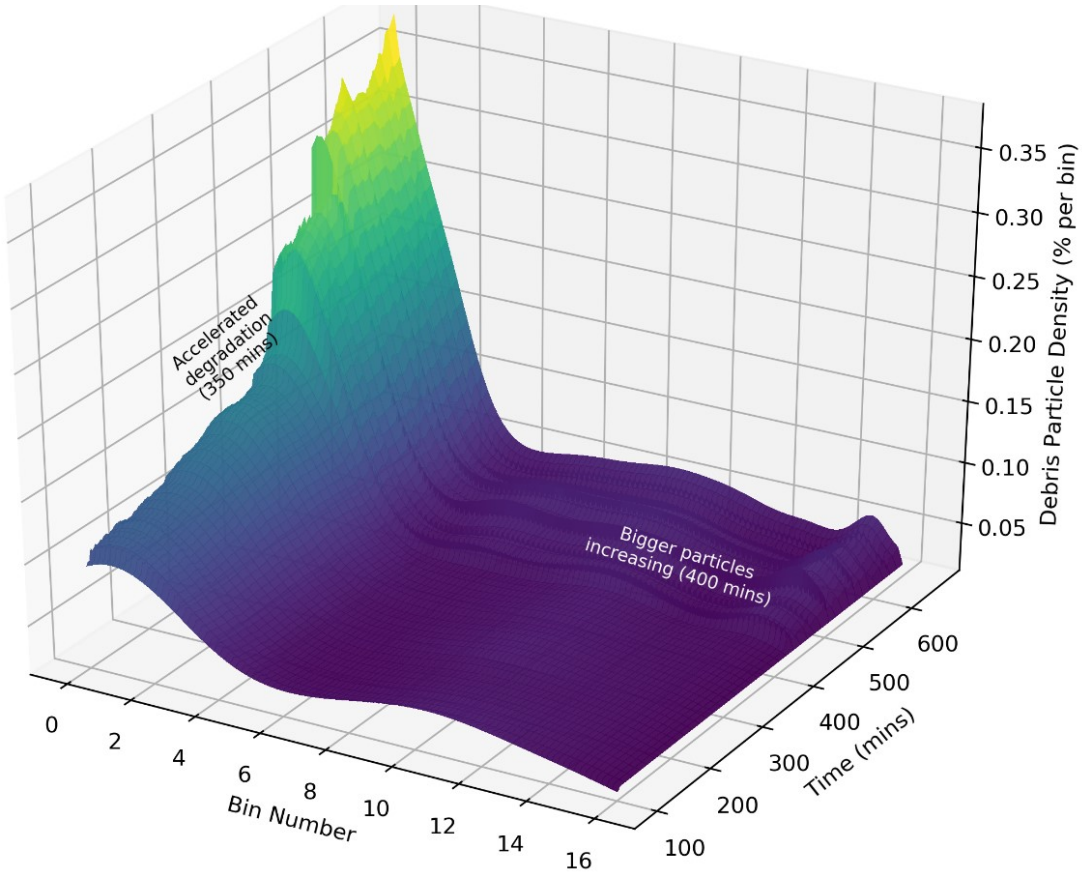

Fig. 2.8 Variation in the debris distribution as the fault severity increases.

2.3.2 Degradation Indicators Based on Wear Debris Characteristics

In this study, the aforementioned variations in the debris count distribution are quantified to generate monotonic and trendable degradation indicators. Calculating the statistical moments of these distributions has been identified, in this study, as an effective and simplest method to monitor the variations of the debris count distribution. Therefore, the skewness and standard deviation of the debris counts distribution are utilized as the two degradation indicators based on the wear debris characteristics. The other two standard statistical moments, mean and kurtosis, of the distribution, give redundant information. The process to calculate the standard deviation and skewness from the information available from the ODMs is described below:

1) *Generate bin sizes list:*
    - Generate a *bin sizes list* that represents the bin sizes of all the debris particles using the particle counts information from the ODM.
    - For example, consider the debris counts information of time 100 minutes in Fig. 2.7, we see that there are a total of 120 debris particles of different sizes with 32 particles of Bin 1, 23 particles of Bin 2, 13 particles of Bin 3 and so on. Therefore, a *bin sizes list* is generated at time 100 minutes which will contain a total of 120 individual bin sizes instances, such that, there are 32 instances of Bin 1, 23 instances of Bin 2, and 13 instances of Bin 3 and so on. Similarly, at time 200 minutes the *bin sizes list* will contain 221 individual bin sizes such that there are 67 instances of Bin 1, 43 instances of Bin 2 and so on. This *bin sizes list* will be generated at every time instant and used for calculating the standard deviation and skewness of the debris size distribution.
2) *Calculate standard deviation and skewness:*
    - The standard deviation of the debris distribution (σ_{Dist}) is calculated as:

$$\sigma_{Dist} = \sqrt{\frac{\sum(x_i - \mu)^2}{T_N}} \tag{2.1}$$

Where x_i are each individual element of the *bin sizes list* and μ is the mean and T_N is the total number of elements of the *bin sizes list.*

- Meanwhile, the skewness of the debris distribution (S_{Dist}) is calculated as:

$$S_{Dist} = \frac{\sum(x_i - \mu)^3}{(N-1){\sigma_{Dist}}^3} \tag{2.2}$$

The standard deviation and skewness indicators of one of the spall propagation tests are displayed in Fig. 2.9 corresponding to the total cumulative debris counts. As can be seen, both of these indicators are monotonic and trendable. We see that the standard deviation indicator linearly reduces in the gradual degradation phase. This is because smaller debris particles are released from the bearing during this phase which leads to the increase of debris concentration of the smaller particles bins (Bins 1 to 4). When the bearing approaches failure bigger particles tend to release more from the bearing which accelerated the failure. This can be recognized by the flattening standard deviation indicator in the accelerated degradation phase. These same changes to the releasing debris particles manifest as the increasing trend of the skewness indicator during the gradual degradation phase. Whereas the skewness indicator also flattens out when the bearing approaches failure during the accelerated degradation phase.

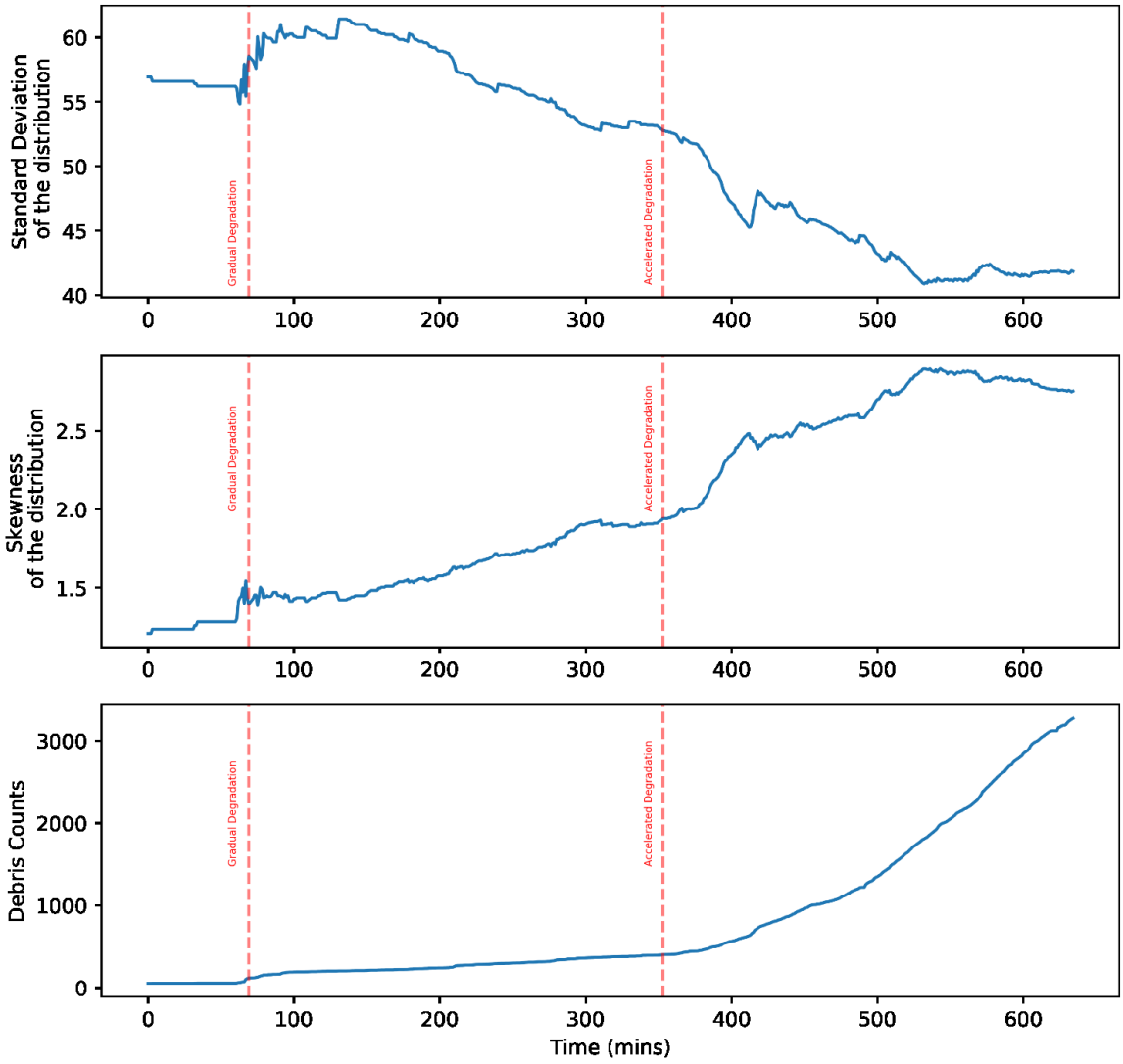

Fig. 2.9 Standard deviation and skewness indicators from the debris counts distribution.

2.4 Online Change Point Detection for Degradation Indicator

Bearing degradation is a stochastic process and it involves an elevated level of uncertainty. Typically, these uncertainties manifest in sudden trend changes and/or amplitude variations in the degradation index. The time instants when these sudden trend changes occur are defined as change points [103]. “Online” change point detection, refers to the processing and identification of the change points from the degradation index in real-time data. Identifying the exact time when these change points occur in the bearing degradation index

will help to reveal important information about the actual failure of the bearing and can be used to schedule maintenance activities [92]. Therefore, online changepoint detection techniques are used to monitor the real-time data and detect the change points in the data.

Previous works in [18], [92], and [93] demonstrated the effectivity of change point detection methods for efficient bearing condition monitoring. In [92], a methodology is proposed for the online detection of the health status of bearings into various damage stages for naturally progressing defects. It uses the Z-statistic score, based on Mahalanobis distance, to detect the shift in the status of bearing health via Chebyshev's inequality. Similarly, in [93], an algorithm based on Chebyshev's inequality is applied to the Jensen Rényi divergence-based degradation index to identify the change points. These change points are further used to remove outliers and also for fault classification as possible locations where specific defect initiation could have taken place. Further, in [18], a combination of change point detection and K-means clustering is used for feature selection in the data-driven RUL estimation algorithms.

In this study, a novel change point detection technique based on linear and polynomial regression is developed and presented in Appendix A. The performance of the proposed ROCD technique is validated by comparing it with one of the most accepted techniques named the Bayesian online changepoint detection technique. Unlike the other change point detection algorithms, the proposed ROCD needs only one hyperparameter to be preassigned which is the length of the sliding window. The ROCD method outputs an *error product* that has a high amplitude at the change points whereas, in the rest of the locations, the *error product* remains very low.

2.4.1 ROCD based Degradation Indicator

The primary aim of this section is to generate a monotonic and trendable degradation feature that can be used for data-driven fault diagnosis. However, the *error product* obtained from the ROCD method cannot be used as a degradation indicator because of two issues: the rapid decrease in the values and also the non-monotonous change. Therefore, a

cumulative summation of the *error product* is used to obtain a monotonous degradation indicator.

To validate the proposed method, the ROCD based degradation indicator is extracted from one of the spall propagation experiments. The ODM cumulative counts are used as the degradation index and a 20 minutes sliding window is selected. The ROCD indicator presented in Fig. 2.10 shows that the *error product* accurately highlights the locations where abnormalities occur in the ODM total cumulative debris counts. Further, the cumulative summation of the *error product* is the monotonous ROCD based degradation indicator. The indicator effectively highlights the time instants where the phase changes occur and this contributes a lot in training the machine learning classifiers for fault phase diagnosis.

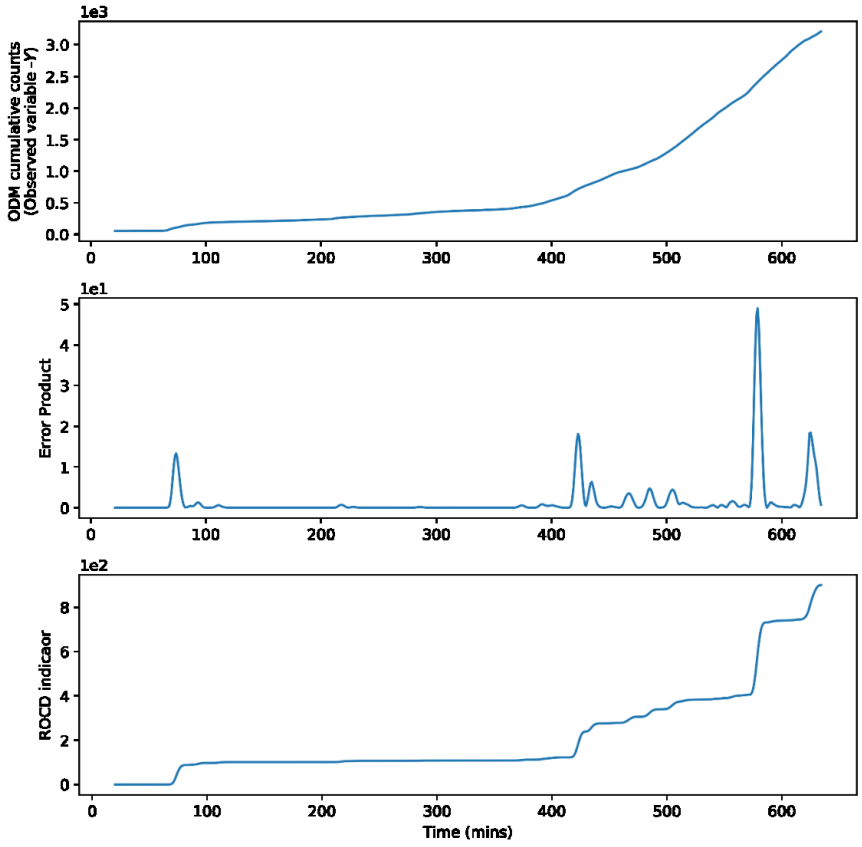

Fig. 2.10 ROCD based degradation indicator with 20 minutes sliding window identifies the areas where bearing degradation occurs.

2.5 Summary

Three novel degradation indicators based on ODM data are proposed in this chapter. The first two indicators leverage the relationship between the wear debris characteristics and the defect severity. Whereas the third degradation indicator is based on the ROCD method proposed in Appendix A. The proposed indicators are validated on an experimental spall propagation test and they satisfy the requirement of monotonicity and trendability. Moreover, they also reveal important information about the bearing degradation and its phase changes. The three proposed degradation indicators along with other multi-sensor indexes will be further used in Chapter 3 to develop a data-driven fault phase diagnosis tool.

Chapter 3: Fault Phase Diagnosis using Ensemble of Random Forest and K- Nearest Neighbors

3.1 Overview

As described earlier, the bearing degradation process can be divided into three phases namely: no-fault, gradual degradation, and accelerated degradation, and in this dissertation, these three phases are denoted as Phase 0, Phase 1, and Phase 2 respectively. Although the degradation indicators proposed in Chapter 3 perform well in identifying the phase changes, it is still difficult to rely only on these indicators for accurate estimation. Therefore, in this section, a data-driven fault phase diagnosis technique is developed using the multi-sensor data available from the experimental setup.

Fault phase diagnosis is used to determine the initiation of various fault phases in the bearing degradation. For example, fault phase diagnosis can be used to identify whether the bearing is in the gradual degradation or accelerated degradation phase at any given time instant during its operation. Accurate fault phase diagnosis can be used, in some cases, to prolong the machine operation by adjusting the operating conditions based on the fault phase. In this study, we will use the results from the fault phase diagnosis to select an appropriate FPT for the model-based prediction technique and also for refining and updating the multi-phase degradation models.

Out of all the degradation indicators and features available only the most effective set of features should be used for training the data-driven models to obtain a higher diagnostic accuracy [37]. Therefore, multiple feature selection methods are employed in this study, to eliminate noisy, redundant, or irrelevant features before using them for training. Although several machine learning techniques have been employed in fault diagnosis studies, KNN and RFs have performed very well and have been used in several studies for fault diagnosis of bearings [42, 43, 44, 45].

RF classification method was proposed by Breiman in 2001 [46], based on the concept of the bagging technique, classification, and regression tree techniques. Compared with the traditional decision tree algorithm, RF has manifested robust classification

performance in solving high-dimensional and small-sample problems [39]. Additionally, the RF algorithm has also inherited the high interpretability of the tree-based models [41]. The KNN is a simple and efficient classification algorithm and it has been studied extensively in many practical applications. The KNN method has very few hyper-parameters that need to be selected; it is highly effective for statistical pattern recognition [45].

Although these techniques have good classification capability, the performance of these techniques can be improved using ensemble learning strategies. This is highly sought after in scenarios that have fewer training datasets. Therefore, a simple but effective ensemble strategy named the weighted average method is used to obtain better performance in fault phase diagnosis. The proposed technique uses both RF and KNN algorithms simultaneously to estimate the fault phase of the bearing degradation and the probabilities are combined to get the final phase estimation.

3.2 Classifiers

3.2.1 K-Nearest Neighbors (KNN)

KNN is a conventional non-parametric classification algorithm that uses a training dataset as the basis for the classification of test samples. In general, to classify an unknown test sample represented as a point in the feature space, the KNN algorithm evaluates the distances between the test sample point and other points in the training data set. Typically, the Euclidean distance is used as the distance metric. Then, it assigns the point to the class among its K nearest neighbors (where K is an integer) [104].

Given a training set $D(x, y)$ (where x represents, the samples and y its respective classes) and a test sample $z = (x', y')$, the algorithm calculates the distance between z and all training samples, $(x, y) \in D$, to get a list of nearest neighbors. The class assignment for y' corresponding to test sample x' will be based on a majority vote of the classes of its nearest neighbors. For example, consider * is a test sample point in the feature space and the rest of the points are the training points with two different classes, as shown in Fig. 3.1.

If $K = 5$, we can delineate the nearest 5 training samples within the shown radius. Thus, we can classify the test sample point as the small circle class because the number of circles is more than the squares. However, if we reduce K such that $K = 1$, then the training samples will reduce and we will classify the test sample point as the square class.

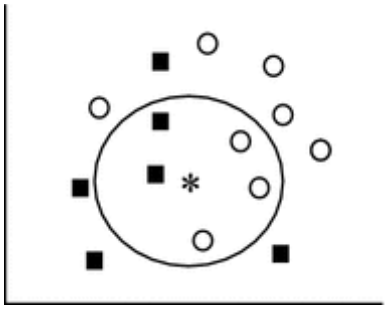

Fig. 3.1 Example of KNN classification [104].

The key point when using KNN is the choice of the distance metric and the value of K. Since we generally have a high number of features in the feature vector, the classification problem is addressed in a high-dimensional space. Thus, a simple Euclidean distance metric will not be appropriate [104]. Therefore, in the KNN algorithms, with a high number of feature vectors, a Minkowski distance metric is used. It is a distance measurement between two points in n dimensional real space. The generalized Minkowski distance is given as:

$$D = \left[\sum_{i=1}^{n} |S_i - x_i'|^p\right]^{1/p} \tag{3.1}$$

Where n is the number of feature vectors, S is the training sample point, p is the order of the norm. When $p = 1$, the Minkowski distance is equal to Manhattan distance, and when p=2, it is equal to Euclidean distance. The appropriate value for p can be optimized during the KNN algorithm optimization step.

Meanwhile, the choice of K is also very critical, if it is very small the classifier can be very sensitive to noise, on the other hand, if it is very large it might group many

neighboring samples belonging to several different classes. Therefore, a more sophisticated approach is used for class labeling; it is based on weighting the contribution of each neighboring sample as a function of the distance from the test sample point which needs to be classified. The weighting can be done such that each neighboring sample is assigned a weight factor which is a function of the inverse distance, such that the weights will be higher for the nearest neighbor.

$$\omega_i = \frac{1}{|S_i - x_i'|^p} \tag{3.2}$$

Where, ω_i is the weight of an individual neighbor.

3.2.2 Random Forest (RF)

RF is one of the most used classifiers based on multiple sets of decision trees generated from the training dataset. In this algorithm, the training dataset is divided into a number of random subsets. These subsets are used to make individual decision trees. This combines the ideas of bagging and the random selection of features; further, it helps to reduce variance and overfitting problems seen in other algorithms [39]. The train data subsets, which were selected to create the trees, are known as in-the-bag observations and the remaining data are out-of-bag observations [46]. Once a test sample vector (out-of-bag observation) is provided to the trained random forest algorithm, each decision tree gives a particular class output. The final class is chosen based on counting the majority of votes from the response of each decision tree.

To train the random forest algorithm, we need to initially determine the number of decision trees, and the attributes to form each decision tree, like the maximum depth and the number of samples in the subset to form the decision trees. Both parameters are optimized during the training step of the proposed methodology.

3.3 Feature Selection Techniques

Feature selection can be defined as a process that chooses a minimum subset of available features from the original set of N features. It optimizes the feature space and eliminates redundant and uninformative features. Feature selection is a fundamental problem in many areas like forecasting, classification, bioinformatics, and object recognition or modeling of complex technological processes. In many practical applications, only a subset of features is usually relevant to the machine learning task at hand [105, 106]. Several methods are proposed for optimal feature selection in the literature. They are generally classified into statistical and entropy-based methods. The purpose of these methods is to discard irrelevant or redundant features from a given feature vector that do not effectively contribute to the learning. Several studies recommend utilizing more than one feature selection method to find the optimal subset of features because a single feature selection method sometimes might not be the efficient feature ranking method for the current application [106]. In this study, we have used three methods for feature selection namely: Variance thresholding, Information gain, and Chi-square test. These three methods are briefly described below:

Variance thresholding is a statistical feature selection method and it is used to identify and eliminate the features which have no/very less variability. To perform variance thresholding, first, find the statistical variance of all the feature vectors in the training dataset and eliminate the features with the very low variance.

3.3.1 Information Gain

Entropy, which is generally used in information theory, is the foundation of the Information Gain (IG) ranking method. It characterizes the purity of an arbitrary collection of samples. Entropy is considered a measure of the system's unpredictability. The entropy (H) of a dataset Y is given as:

$$H\ (Y) = -\sum_{y \in Y} p(y)\log(p(y)) \tag{3.3}$$

where $p(y)$ is the marginal probability of the random variable y in the dataset Y.

First, the observed values of Y in the dataset are partitioned according to the values of a second feature X. If the entropy of Y with respect to the partitions is less than the entropy of Y prior to partitioning, then there exists a relationship between features Y and X. The entropy of Y after observing X is then:

$$H\ (Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log(p(y)) \tag{3.4}$$

where $p(y|x)$ is the conditional probability of y given x.

Since entropy is a metric that allows quantifying the amount of impurity in a dataset, we can define an information gain (IG) measure reflecting additional information about Y provided by X that represents the amount by which the entropy of Y decreases [107].

$$IG = H\ (Y) - H\ (Y|X) = H\ (X) - H\ (X|Y) \tag{3.5}$$

When the information gain method is applied to the feature set, the IG measure is obtained for every single feature (X) set with respect to the training labels (Y). The IG values are considered as the contribution of each feature to the training label.

3.3.2 Chi-Squared Test

The Chi-square test is another commonly used method in statistics to test the independence of two events. It evaluates the contribution of a feature set by computing the value of the chi-squared (X^2) statistic with respect to the given observation class. Initially, an assumption that the two features are completely independent is considered, and it is tested by the chi-squared formula:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{3.6}$$

where, O_i is the observed frequency of the class i, and E_i is the expected (theoretical) frequency of the class i. Greater the value of X^2, the greater the dependence of the observation features to the class.

When the Chi-square test is applied to the feature set, X^2 is obtained for every single feature (X) set with respect to the training labels (Y). The X^2 vales are considered as the contribution of each feature to the training label.

3.3.3 Evaluation Metrics

In machine learning, the performance of classification models is generally quantified using evaluation metrics namely, accuracy, precision, sensitivity, and F-score [32]. Accuracy is a very simple metric which computes the fraction of the number of correct predictions to the total number of predictions. Intuitions formed based on the simple accuracy metric can give misleading information especially in datasets with imbalanced class distributions. Therefore, precision, sensitivity, and F-score will be used as the evaluation metrics in this study. A brief description of these performance metrics is given in this section.

3.3.3.1 Precision

Precision can be described as the proportion of the specific class estimations that are actually true. It is given by:

$$Precision = \frac{TP}{TP + FP} \tag{3.7}$$

where TP (true positive) indicates the number of correctly classified samples that are actually correct, and FP (false positive) is the number of classified samples that are actually incorrect.

3.3.3.2 Sensitivity:

Sensitivity, also known as recall, can be described as the proportion of specific class estimations that are estimated correctly. It is mathematically defined as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.8}$$

where FN (false negative) indicates the number of incorrectly classified samples that were actually correct.

3.3.3.3 *F*-score:

F score, also known as $F1$ measure, is a way to combine the precision and sensitivity of the classification. Mathematically, it is the harmonic mean of both precision and sensitivity.

$$F\ score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{3.9}$$

In multi-class classification problems, a weighted F score is used where the weights will be the number of instances the individual class is observed in the training set.

3.4 Proposed Framework

In this section, the detailed steps used for training the classification models and using the ensemble learning strategy to accurately estimate the bearing fault phase are postulated. The proposed framework is divided into two sub-processes in the first process, the feature extraction, selection, and training of the classification models are performed. In the second process, the fault phase diagnosis using the weighted average ensemble technique on the real-time data is performed.

3.4.1 Training the Classification Models

The first sub-process, training the classification models, is summarized in Fig. 3.2 and the steps are explained below in detail.

Feature Extraction:

In this step, a training dataset is formed from the bearing spall propagation experiments performed on the experimental setup, described in Section 1.4. The training set should contain the entire data of a full spall propagation test. For example, if 5 run-to-failure experiments are performed, the entire data of three experiments from beginning to failure should be used to form the training dataset. The division of the training and testing dataset as mentioned above is very important to avoid information leakage. The steps are explained below in detail:

- The available raw data from all the available sensors like, ODMs, accelerometers, temperature, pressure, and load, etc., sensors are acquired as the training dataset.
- The degradation indicators described in Chapter 2 are extracted from the acquired ODM data and are used as features for training the classification models.
- A feature vector is formed with all the available multi-sensor information along with the extracted degradation features.

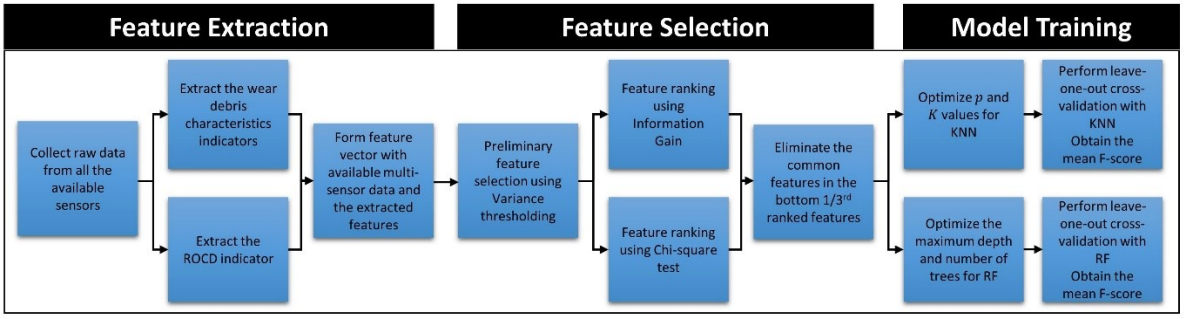

Fig. 3.2 The step-by-step process for training the classifiers.

Feature Selection:

In this step, we use three feature selection methods on the training dataset to select the most appropriate feature vectors for the classification models.

- Perform preliminary feature selection using variance thresholding by eliminating the features with zero or nearly zero variance.

- Perform feature ranking using Information Gain and Chi-square test and obtain the ranks of individual features.
- Sort the features as per the ranks and select the bottom 1/3rd features from both feature ranking methods. Find the feature vectors which are common in the bottom 1/3rd ranked features and eliminate them from further processing.
- Count the number of available features as F_n. Further, form the feature vector with the remaining features for training the classification models.

Model Training:

In this step, the optimized hyper-parameters of both the classification models are calculated and the leave-one-out cross-validation technique is used on the training dataset.

- The feature vectors in the training dataset are used to train and optimize the KNN and RF models individually.
- KNN optimization hyper-parameters and range for optimization:
    - p value of Minkowski distance: range - 1 to 5
    - K value (No. of nearest neighbors): range - 1 to 50
- Random forest optimization hyper-parameters and range for optimization
    - Number of decision trees (n_{trees}): range - 100 to 3000 with a step of 100
    - Maximum depth of each tree: range - 5 to F_n (number of features used after feature selection)
- Perform a grid search optimization with 5-fold cross-validation and the optimization metric as F -score. Where every combination of hyper-parameters is trained and the hyper-parameter combination which gives the highest F-score is selected as the optimized hyperparameters.
- Train both KNN and RF classification models with the optimized hyper-parameter combination.
- Perform leave-one-out cross-validation on the training dataset and obtain the F scores of each validation. The leave-one-out cross-validation method uses all the samples except one to train the classification model and uses the left-out samples for testing the trained model. For example, if data from three spall propagation

experiments are used, then the data from two full experiments are used to train the classification model and the one left out will be tested by the trained classification model. Similarly, this is performed for all the possible combinations and the F-score for each combination is recorded.

- The mean F-score from the cross-validation using the KNN classification model will be used as the weight for the KNN classifier in the weighted ensemble technique. Similarly, the mean F-score from the cross-validation using the RF classification model will be used as the weight for the RF classifier.

3.4.2 Fault Phase Diagnosis

In this step, the feature vectors are extracted from the real-time testing data acquired from the experimental setup are used for fault phase diagnosis using the ensemble technique. The process is summarized in Fig. 3.3 and is explained below in detail.

Feature Extraction:

- The relevant raw data from all the available sensors like, ODMs, accelerometers, temperature, pressure, and load, etc., sensors are acquired.
- The required features chosen in the feature selection process are extracted from the acquired data.
- A feature vector is formed with all the relevant features.

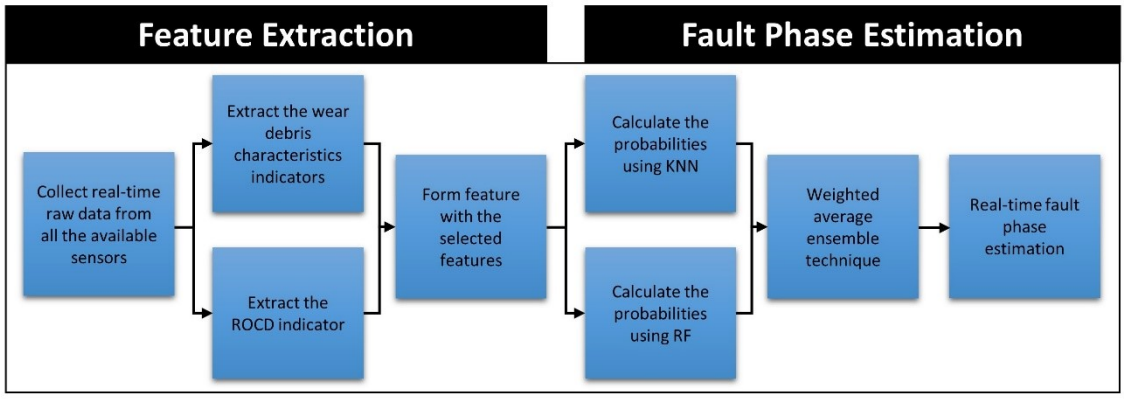

Fig. 3.3 The step-by-step process to implement fault phase diagnosis on the real-time data.

Fault Phase Estimation:

In this step, the probabilities are estimated for the testing dataset from both the classifiers and the weighted average ensemble technique is used to get the fault phase estimate.

- The F-scores obtained from both models in the leave-one-out cross-validation step are used as the weights for the respective model.
- Simultaneously, perform fault phase classification on the testing dataset and obtain the probabilities for each phase from both classifiers.
- Apply the weighted average ensemble technique to the probabilities as follows:
    - Multiply the estimated probabilities with the respective weight of the classifier. Find the weighted average of the probabilities for each fault class and assign the phase with the highest probability as the estimated fault phase.
    - For example, consider that the weight of the KNN classifier is 0.75 and the probability estimates are 0.1, 0.7, 0.45 for Phase 0, Phase 1, and Phase 2 respectively. Meanwhile, the weight of the RF classifier is 0.65 and the probability estimates are 0.3, 0.6, 0.35 for Phase 0, Phase 1, and Phase 2 respectively. First, multiply the weight of the KNN classifier by its probabilities, similarly multiply the weight of the RF classifier by its probabilities. Then find the weighted average of the probabilities which gives the ensembled probabilities as 0.135, 0.457, 0.282 for Phase 0, Phase 1, and Phase 2 respectively. Therefore, in this example, the fault phase estimated by the weighted average ensemble technique would be phase 1.

3.5 Demonstration on Experimental Data

In this section, datasets from five bearing accelerated spall propagation experiments are used to validate the proposed fault phase diagnosis technique. As discussed, a total of 5 bearing spall propagation experiments have been performed on the experimental setup described in Section 1.4. The full data from three experiments are used for training the classifiers. Subsequently, data from two experiments are used for testing the performance of the proposed ensemble technique.

3.5.1 Data Acquisition and Feature Extraction

Information from multiple sensors is used as the feature vector in this proposed methodology. As described in Section 1.4, two ODMs are used with varying sensitivity ranging from 65 to 1000 μms to obtain the relevant wear debris particle information. Further, two vibration accelerometers, three temperature sensors, a torque meter, a flowrate meter, two oil pressure sensors, two load sensors, along with the motor speed are used in this study. Since the primary agenda of this thesis is to examine the performance of the ODMs in bearing PHM, only the RMS values of the vibration signature are obtained from the two vibration accelerometers. Therefore, no frequency domain or time-frequency domain features are extracted in this study. All the features along with their data cleaning and processing steps are mentioned in Table 3.1.

Table 3.1 List of features along with the processing steps.

Sensor	Feature Label	Feature	Processing
ODM 1 (65 to 350 μms)	F_ODM1_01	Cumulative debris particle counts	Raw data from the sensor
	F_ODM1_02	Approximate mass lost	The weighted sum of the debris counts and the average bin masses
	F_ODM1_03	Debris count generation rate	The difference of cumulative debris counts between two consecutive time instances.
	F_ODM1_04	The skewness of debris distribution indicator	Process described in Chapter 2

	F_ODM1_05	The standard deviation of the debris distribution indicator	Process described in Chapter 2
	F_ODM1_06	ROCD indicator	Process described in Chapter 2
ODM 2 (250 to 1000 μms)	F_ODM2_01	Cumulative debris particle counts	Raw data from the sensor
	F_ODM2_02	Approximate mass lost	The weighted sum of the debris counts and the average bin masses
	F_ODM2_03	Debris count generation rate	The difference of cumulative debris counts between two consecutive time instances.
	F_ODM2_04	The skewness of debris distribution indicator	Process described in Chapter 2
	F_ODM2_05	The standard deviation of the debris distribution indicator	Process described in Chapter 2
	F_ODM2_06	ROCD indicator	Process described in Chapter 2
Accelerometer	F_Vib_1	Outer race vibration	RMS value of 1-second vibration signature
	F_Vib_2	Support vibration (placed on the support structure of the shaft)	RMS value of 1-second vibration signature
Temperature	F_Temp_1	Bearing housing temperature difference	Difference between outlet temperature and inlet temperature of the bearing housing

	F_Temp_2	Outlet temperature	Raw data
	F_Temp_3	Race temperature	Raw data
Speed	F_Speed	Motor speed in RPM	Raw data
Torque meter	F_Torque	Motor torque	Raw data
Load	F_Load_1	Axial load on the bearing	Raw data
	F_Load_2	Radial load on the bearing	Raw data
Pressure	F_Pressure	Inlet oil pressure	Raw data
Flowrate	F_Flow	Oil flowrate	Exponential moving average for a span of 15 minutes

3.5.2 Feature Selection

Applying the variance threshold method to the training feature dataset shows that the feature 'F_Vib_2', i.e., support structure vibration, has very low variability. Therefore, 'F_Vib_2' feature is eliminated from the feature set. Further, the information gain method and chi-square test are applied on the feature vector, the feature ranking from both the methods are shown in Fig. 3.4.

The features are arranged in ascending order as per their contribution and the 1/3rd of least informative features from the available feature vectors are highlighted. The "*" symbol marks the common features in the highlighted features. Since these features have very little contribution to the fault phase labels, these features will be eliminated from

further processing. Therefore, features namely, 'F_Flow', 'F_Pressure', 'F_Vib_1', 'F_Torque' are eliminated from the feature vector. We can also see that all the indicators from ODM contribute a lot to the fault phase diagnosis.

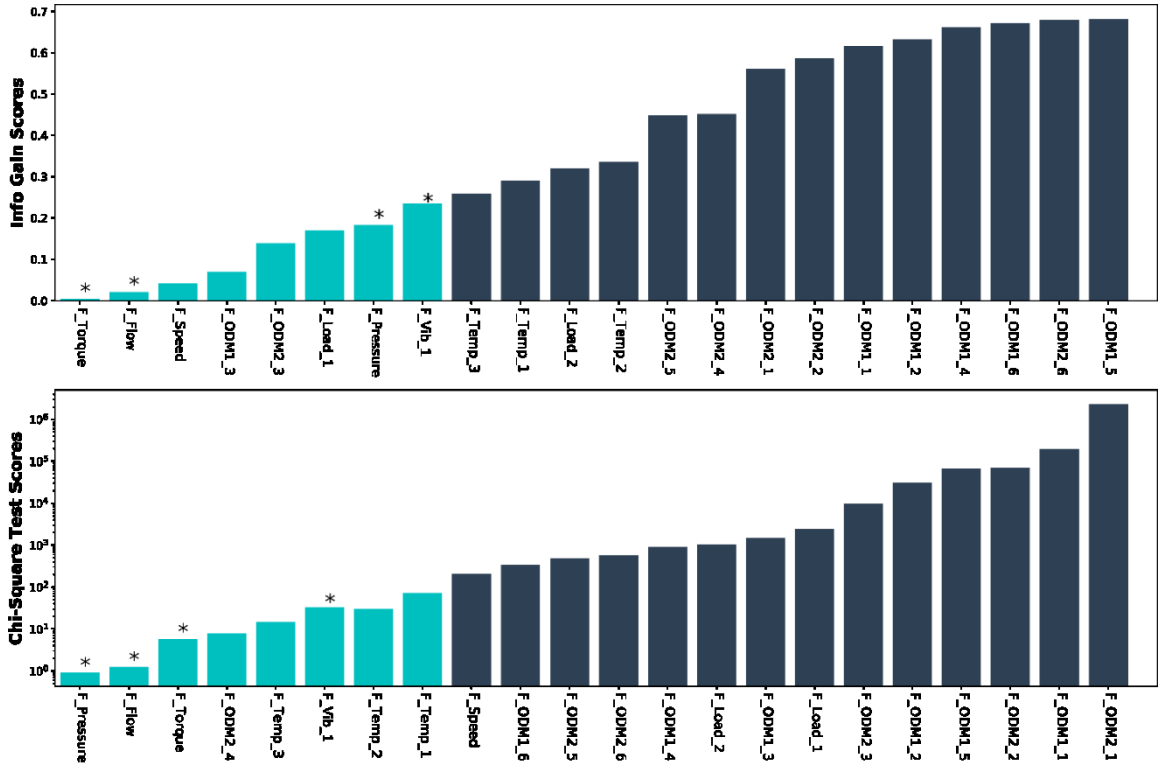

Fig. 3.4 Feature ranking from both the methods arranged in ascending order and the bottom 1/3rd features are highlighted.

3.5.3 Training Classifiers and Optimization

The training dataset is formed with the remaining 17 features to train and optimize the classification models. A grid search technique is utilized for finding the most optimum hyper-parameters for the models with F-score as the evaluation metric. The grid search is an exhaustive search technique where all the possible combinations of hyper-parameters

are evaluated. The best combination with the highest evaluation metric is chosen as the optimized hyper-parameter.

The KNN classifier is optimized for p value and K value while the RF classifier is optimized for the number of decision trees (n_{trees}) and the maximum depth of each tree. The optimized hyperparameters obtained for KNN are: p value =1.5 and the $K = 3$ whereas, the optimized hyperparameters for RF are: $n_{trees} = 1250$ and maximum depth =15. The pre-optimized and post-optimized classifiers are used for leave-one-out cross-validation and the results are shown in Fig. 3.5. As can be seen, there is a significant improvement in the performance of both the classifiers after the hyperparameter optimization. The optimized KNN classifier obtains a mean F-score of 0.696 and the optimized RF classifier obtains an F-score of 0.791. Therefore, the ensemble technique weights obtained from the training and optimization process are 0.696 and 0.791 for the KNN classifier and RF classifier respectively. Further, the entire training dataset is used to train both the classifiers for the testing process.

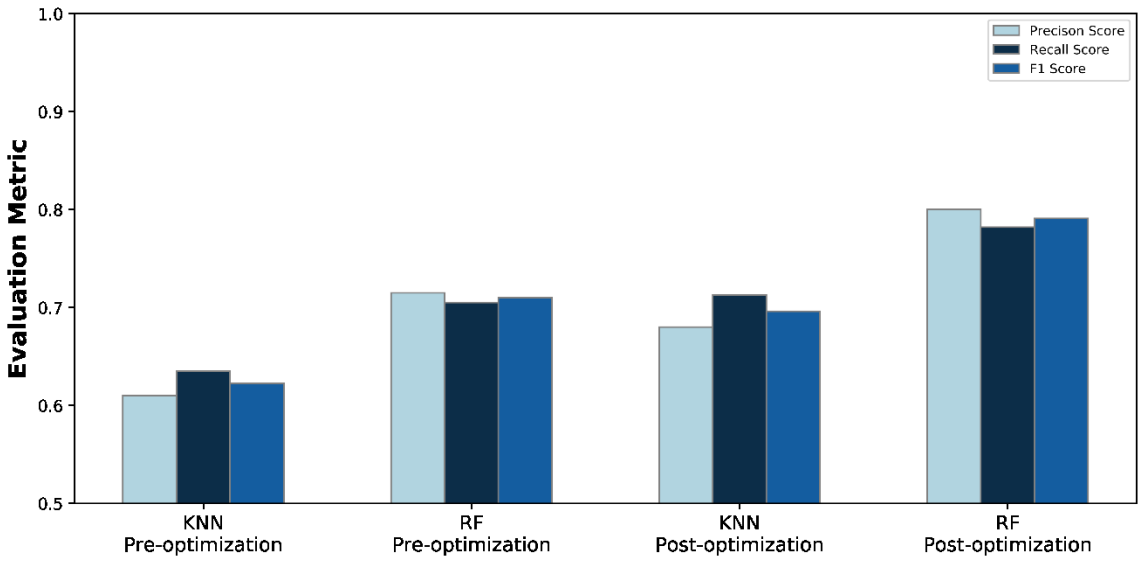

Fig. 3.5 Comparison of the evaluation metrics on the validation experiments

3.5.4 Testing the Ensemble Technique

The performance of the proposed methodology is validated on the testing dataset with two full experiments performed on the experimental setup. The trained and optimized KNN and RF classifiers are used individually in this process. As described in the earlier section, the chosen features are obtained and a feature vector is formed for the fault phase diagnosis. The probabilities estimated by the classifiers are multiplied by their respective weights. Finally, the weighted average ensemble technique is used to estimate the fault phases of the testing dataset.

Fig. 3.6 and Fig. 3.7 show the fault phase diagnosis results of the two experiments used for testing the proposed methodology. Both figures show the phase estimation either 0, 1, or 2 indicating the no-fault initiation phase, gradual degradation phase, and accelerated degradation phase of the bearing respectively. The corresponding debris counts released from the bearing are also shown for the perspective. As we can see, the phase estimation from both the classifiers has a lot of false alarms; however, the ensemble technique helps to combine the estimations of both the classifiers to provide a better diagnosis and reduces the false alarms.

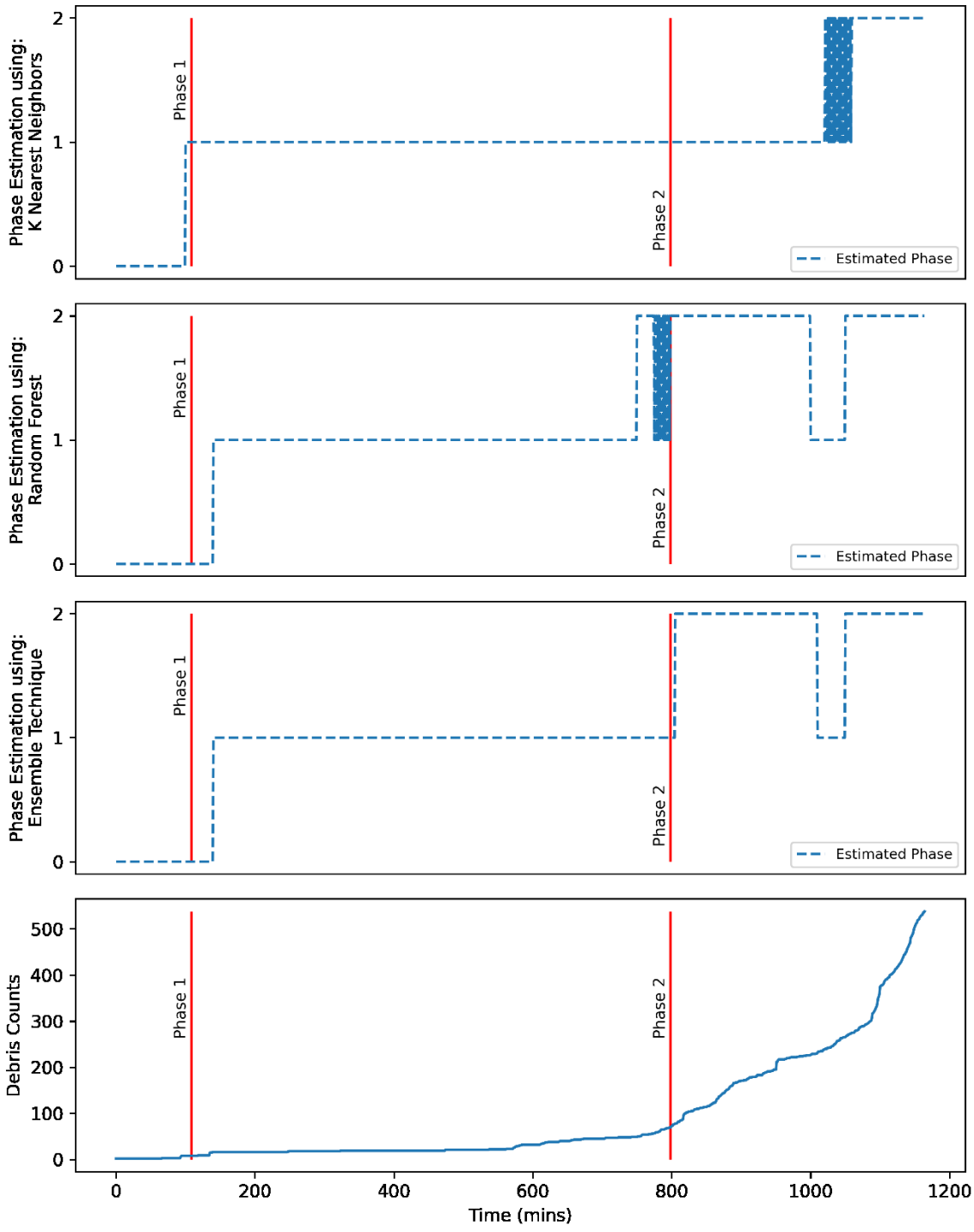

Fig. 3.6 Fault phase diagnosis of the spall propagation experiment 4.

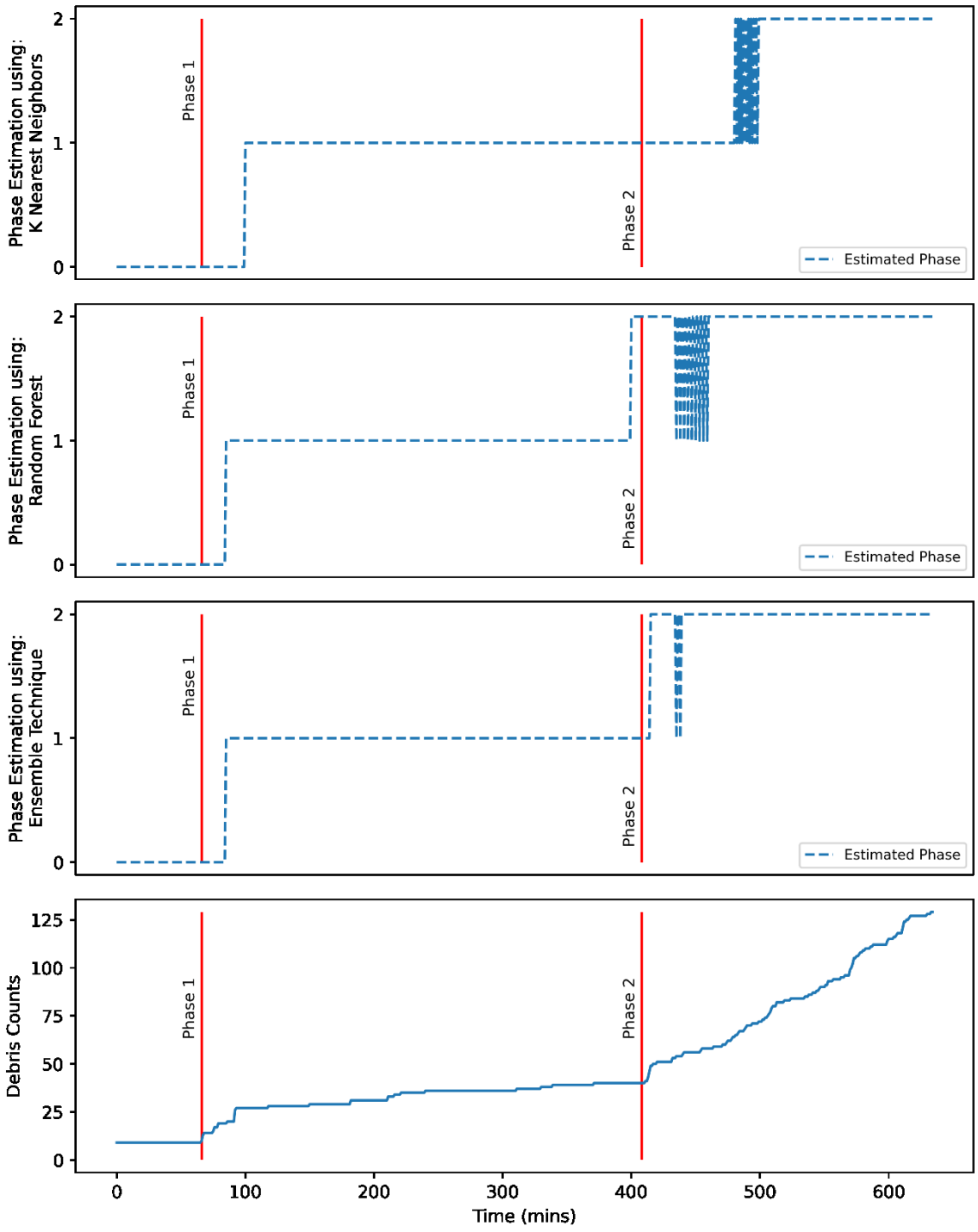

Fig. 3.7 Fault phase diagnosis of the spall propagation experiment 5.

3.6 Summary

A data-driven fault phase diagnosis technique based on an ensemble of KNN and RF classifiers is developed in this chapter. The technique is used to estimate the appropriate bearing degradation phase (no-fault, gradual degradation, and accelerated degradation phase). Multi-sensor information including the degradation indicators developed in Chapter 2 is used in the feature vector. However, only the relevant features are chosen for training the classifiers using feature selection methods namely, variance thresholding, information gain, and chi-square test. Further, the classifiers are optimized and the leave-one-out cross-validation technique is used to calculate the F-scores individually for each classifier. These F-scores are used as the weights for the respective classifier in the weighted average ensemble method. Finally, the proposed technique is validated using experimental data where two full spall propagation experiments are used for training the classifiers and 2 are used for testing.

Chapter 4: Adaptive Particle Filter for Real-Time Degradation Prediction

4.1 Overview

PFs are extensively utilized for model parameter estimation and degradation prediction in PHM systems. However, three potential issues limit the use of PFs in online systems for real-time degradation prediction. They are particle degeneracy, particle impoverishment, and initial range dependency. Although many studies substantially discuss the first two issues, very few consider the issue of initial range dependency. Several existing PF algorithms are completely dependent on the range of the initial estimates of the model parameters and they perform inaccurately when the real model parameters are outside the expected initial range. Moreover, in several practical applications, the range of the model parameters is not always known. Hence, this is a serious issue that hinders the efficiency of the PFs in estimating the model parameters when there are unforeseen trend changes in the real-time data. Therefore, in this chapter, we propose an enhanced adaptive particle filtering technique that is independent of the initial particle range and can accurately adapt to adverse trend changes in real-time data. The viability of this proposed technique for real-time degradation prediction is validated using a series of simulation tests.

4.2 Particle Filtering for Prognostics

4.2.1 Principle of Particle Filtering

In PHM we deal with the prognosis of the health state of dynamic systems. Typically, the health states cannot be measured directly using the available sensors in the machine. Therefore, we develop health indicators that represent the health state of these dynamic systems with respect to time. In the Bayesian statistical learning framework, the evolution of the system health state is described using a state-space model (f), which is known as the degradation model or system model. These degradation models are based on the underlying physics of the mechanical system. Paris law, Wiener-process model, and several others are

used as degradation models in the literature [8, 62, 63, 108]. These degradation models include few unknown model parameters (Φ_t) which are to be estimated using the real-time acquired data from the sensors. The estimation of these model parameters is the primary purpose of the PF algorithm.

Given a discrete-time, model parameter estimation (Φ_t) problem, the non-measurable health state is represented using the degradation model (f) expressed in (4.1). Since the system health state cannot be measured directly, we introduce a measurement model (h) which relates the measured sensor values to the health state as expressed in (4.2).

$$Y_t = f(\lambda_{t-1}, \Phi_t, \nu_t) \tag{4.1}$$

$$Z_t = h(\lambda_t, \omega_t) \tag{4.2}$$

Where t is the time index, Y_t is the system health state, Φ_t is a vector of unknown model parameters, Z_t is the measured data. ν_t and ω_t are the noise in the health state and measurements respectively. The probability density function (pdf) of the health state (Y_t) is recursively estimated based on the measurement data (Z_t). If Y_t has an initial density $p(Y_0)$, the transition pdf can be represented by $p(Y_t|Y_{t-1})$. The health state pdf at time $t-1$ is obtained from the measurement model as $p(Y_{t-1}|Z_{1:t-1})$. Assuming that the transition pdf is available, the prior pdf $p(Y_t|Z_{1:t-1})$ of the state at time step t can be estimated using the Chapman-Kolmogorov equation:

$$p(Y_t|Z_{1:t-1}) = \int p(Y_t|Y_{t-1})p(Y_{t-1}|Z_{1:t-1})dY_{t-1} \tag{4.3}$$

Consequently, when new measurement data is available, it is used to update the prior pdf to generate a posterior pdf using the Bayes theorem:

$$p(Y_t|Z_{1:t}) = \frac{p(Z_t|Y_t)p(Y_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})} \tag{4.4}$$

Similar methodology involving Bayes theorem is utilized in other statistical learning frameworks also, such as the Kalman Filter and Bayesian Method [59]. The recursive propagation of the posterior pdf is only a conceptual solution because it requires

the evaluation of complex high-dimensional integrals. To solve this numerically, the PF algorithm approximates the posterior pdf using a set of N number of particles and a sampling algorithm. A few of the most used sampling algorithms are discussed below.

4.2.2 Standard Particle Filters

The Sequential Importance Sampling (SIS) is the most used sampling algorithm in PFs and has been used several times in PHM [51, 61, 74]. This SIS algorithm is used to solve the recursive Bayesian filtering problem via Monte Carlo simulations. This approach is also known as bootstrap filtering, the condensation algorithm, interacting particle approximations, or survival of the fittest algorithm [68]. In this method, we represent the posterior pdf $p(Y_t|Z_{1:t})$ as a set of random sampled numbers (particles) Y_t^i ($i = 1,2,3 \ldots\ldots, N$) and their associated weights w_t^i ($i = 1,2,3 \ldots\ldots, N$). The posterior density can be approximated as:

$$p(Y_t|Z_{1:t}) \approx \sum_{i=1}^{N} w_t^i p\left(Y_t|Z_t^i\right), \qquad \sum_{i=1}^{N} w_t^i = 1 \tag{4.5}$$

Where the weights (w_t^i) are the approximation of the probability density of the corresponding particles. In a nonlinear system with non-gaussian noise, the weights (w_t^i) of a set of N particles can be recursively updated through the following:

$$w_t^i = w_{t-1}^i \frac{p\left(Z_t|Y_t^i\right)p\left(Y_t^i|Y_{t-1}^i\right)}{q\left(Y_t^i|Y_{t-1}^i, Z_t\right)} \tag{4.6}$$

Where $q\left(Y_t^i|Y_{t-1}^i, Z_t\right)$, is the importance density function and there are several ways to determine this function. The most common way is to select $q\left(Y_t^i|Y_{t-1}^i, Z_t\right) = p\left(Y_t^i|Y_{t-1}^i\right)$ such that:

$$w_t^i \approx w_{t-1}^i p\left(Z_t|Y_t^i\right) \tag{4.7}$$

In the initialization step of this PF, we need to determine the initial range of random particles for the individual model parameter $\Phi_{t=0}^i$. In the SIS step, the resampling of the relevant particles is carried out from these initial random particles $\Phi_{t=0}^i$ based on the

importance weight density. Therefore, the resampling operation makes the particles with higher weights to be resampled again, whereas the smaller weights will be replaced with the particles of higher weights. As time progresses, the new particle set at time $t = k$ will have fewer unique particles when compared to the particle set at time $t = 0$. This is known as the particle degeneracy problem where most of the particles are just repeating copies of the existing higher weight particles. Due to particle degeneracy, most of the computational effort in the PF is spent on these repeating particles. If serious degeneracy takes place, it leads to loss of particle diversity called sample/particle impoverishment. As a result, the model parameter estimation accuracy reduces inevitably.

There are other types of PFs reported in the literature, where the importance density function is varied to achieve better weight estimation. Auxiliary PF, first reported in [109], was introduced as a variant of the standard SIS PF. This filter can be derived by introducing an importance density function that uses the mean of the state estimations obtained from the degradation model (f). The advantage of the Auxiliary PF is that it generates particles from the sample depending on the current measurement and are most likely to be close to the true state. If the process noise is small, then Auxiliary PF is often not so sensitive to outliers as in SIS PF, and the weights are more uniform. However, if the process noise is large, the Auxiliary PF degrades because it has to resample based on a poor approximation of the health state.

Similarly, Regularized PF was first proposed in [110], in which the particles are resampled from a continuous approximation of the posterior density $p\left(Z_t \middle| Y_{1:t}^i\right)$ and the posterior is approximated using a kernel function. For more details on the Regularized PF, readers are recommended to read [68] and [110]. The Regularized PF has a theoretic disadvantage in that the samples are no longer guaranteed to asymptotically approximate those from the posterior. In practice, when the process noise is small, the performance of Regularized PF is better than the SIS PF. However, in processes with adverse noise and trend changes, even the Regularized PFs fail to perform well.

All these PF algorithms have a commonality that the efficiency and accuracy of the model parameter estimation depend largely on three key factors: the number of particles (N), the initial range for the model parameters, low noise in the observed health state.

4.3 Enhanced Adaptive PF

A primary disadvantage in the standard PF approach is that the initial particle range mentioned in the initialization step remains constant throughout the model parameter estimation process. In case that there is a phase change in the health state (as shown in Fig. 1.1), then the PF cannot adapt to it because there will be no relevant particles in the initial range which accommodate the phase change. Moreover, the initial range of values for the unknown model parameters (Φ_t) are not accurately known in many practical applications [65]. Therefore, after the phase change occurs, the parameter estimation will diverge from the true parameter values and continue diverging with further time steps. To overcome this issue, in the standard PFs, a large range of values are given as the initial estimates of the unknown model parameters; this compromises the fidelity of the filter. Otherwise, to maintain the fidelity, the number of particles is increased during the initialization, this adversely affects the computation requirements. To overcome these limitations, we propose an adaptive sampling process. In the adaptive sampling process, a new and relevant set of random particles are generated every new iteration to avoid particle degeneracy and impoverishment.

In practice, there is no existing closed-form solution for calculating the model parameters from the observed data. But it is possible to utilize an approximation of the model parameters, hence PFs approximate the posterior distribution of the model parameters using various approximations. In [75], a Gaussian PF is proposed which approximates the posterior distributions by a single Gaussian, similar to Gaussian filters like the extended Kalman filter and its variants. In [76], a sum of Gaussian PF is demonstrated where it approximates the filtering and predictive distributions by weighted Gaussian mixtures. Further in [73], Ensemble Gaussian PF is proposed to improve the spread of particles. These studies demonstrate the effectiveness of assuming gaussianity/normality of the particle distribution [77]. Therefore, in our proposed enhanced

adaptive PF method, the posterior distribution is approximated as a single normal distribution, shown below:

$$p(Y_t|Z_{1:t}) \approx n(\Theta_t^w, \sigma) \tag{4.8}$$

Where, Θ_t^w is the mean of the distribution and σ is the standard deviation of the distribution. The procedure to calculate the mean is explained below in the detailed steps. Whereas the standard deviation is pre-assigned during the initialization for all the model parameters. Since we approximate the posterior distributions as a normal distribution, the particle resampling step, which exists in almost all PFs, becomes unnecessary. Hence, normal distribution not only ensures the diversity of the particle distribution but also results in a computational advantage. The detailed procedure of the proposed enhanced adaptive PF is explained below, and the flow chart is shown in Fig. 4.1.:

1) *Initialization*
    a. N initial random particles are generated within expected ranges for all the model parameters.
    b. Expected standard deviations (σ_i) for all model parameters (Φ_i) are initialized.
    c. Set time step $t = t + 1$
2) *State Estimation:*
    a. Obtain the N health state estimates $Y_t^{j=1,2,\dots,N}$ using the degradation (f) and measurement models (h).
    b. Median of the N states (Y_t^{Med}) is considered as the state estimate for t.
    c. Once the measured health state (S_t) is available, a residual ratio (ϵ) is calculated as:

$$\epsilon = abs(\frac{S_t - Y_t^{Med}}{S_t})$$

3) *Likelihood Estimation:*
    a. Estimate the maximum likelihood and assign them as weights. Assuming the state-space model follows a gaussian distribution, the weights can be calculated as:

$$w_t^i = \frac{1}{\sqrt{2\pi\nu^2}} e^{\left[-\frac{(S_t - Y_t^{Med})^2}{2\nu^2}\right]}$$

where ν represents the health state or process noise:

b. Calculate the weighted average (Θ_i^w) of the distribution of the particles using the estimated weights w_t^i.

$$\Theta_i^w = \frac{\sum_{j=1}^{N} w_t^i p\left(Y_{t-1} \middle| Z_{t-1}^i\right)}{\sum_{j=1}^{N} w_t^i}$$

If: ϵ < 0.5 (No trend change detected)

Residual Ratio (ϵ) is a user-defined variable based on the application.

4) <u>Adaptive Sampling:</u>

a. The N particles for all the model parameters are replaced with N new particles within a normal distribution $n(\Theta_i^w, \sigma_i)$ with mean as the respective model parameter's weighted average (Θ_i^w) and the preassigned standard deviation (σ_i).

Else If: ϵ > 0.5 (Trend change detected)

5) <u>Extended Adaptive Sampling:</u>

While: *ϵ > 0.5 (PF Estimation of the same observation is continued till the residual ratio reduces below 0.5)*

a. The N particles for all the model parameters are replaced with N new particles within a normal distribution $n(\Theta_i^w, r\sigma_i)$ with mean as the respective model parameter weighted average (Θ_i^w) and the standard deviation ($r\sigma_i$). Where r is the spread factor defined by the user; such that $r > 1$. Larger r generates a bigger range of particles for the model parameters so that, the PF can adapt quickly to the trend change. In this paper spread factor $r = 2$.

b. The process goes back to step 2) <u>State Estimation</u>

End while

6) <u>Model Parameter Smoothing and Estimation:</u>

a. The exponential moving average of the particles from time $t-1,\ t-2,\ t-3$ ... is the resultant model parameter estimate (Φ_t) at time instant t

7) *Next Observation:*

a. Once new measurement data is available, the progress goes back to step *2) State Estimation* for the next observation step $t = t + 1$.

The likelihood-based importance density functions are discussed in the literature where the weights of the particles are approximated directly as the likelihood estimate [68]. Also in practice, the posterior is closer in similarity to the likelihood than to the prior. Using a better approximation based on the likelihood, rather than the prior, is shown to improve state estimation. Therefore, we use the maximum likelihood directly as the weights in our enhanced adaptive PF. Next, in the adaptive sampling process, new particles are generated with every new time step as a normal distribution. The normal distribution ensures that a large number of particles exist around the region of high likelihood while a low number of particles exist around the region of low likelihoods. This sampling process is different from the standard PFs because in the standard PFs the low weight particles are replaced with repeating higher weight particles but it does not generate newer particles outside the initial range.

Whenever a trend change is detected, i.e., the residual ratio is more than the prescribed value ($\epsilon > 0.5$), the extended adaptive sampling process is activated. In general, when trend change occurs the model parameters (Φ_t) of the degradation model (f) change drastically. For instance, the relatively lower values of the model parameters in the gradual degradation phase increase to higher values in the accelerated degradation phase. In such scenarios, the range of the particle distribution needs to be increased and out-of-range particles should be generated so that the model parameters can adapt to the trend change. The theoretical working of the extended adaptive sampling process is illustrated in Fig. 4.3.

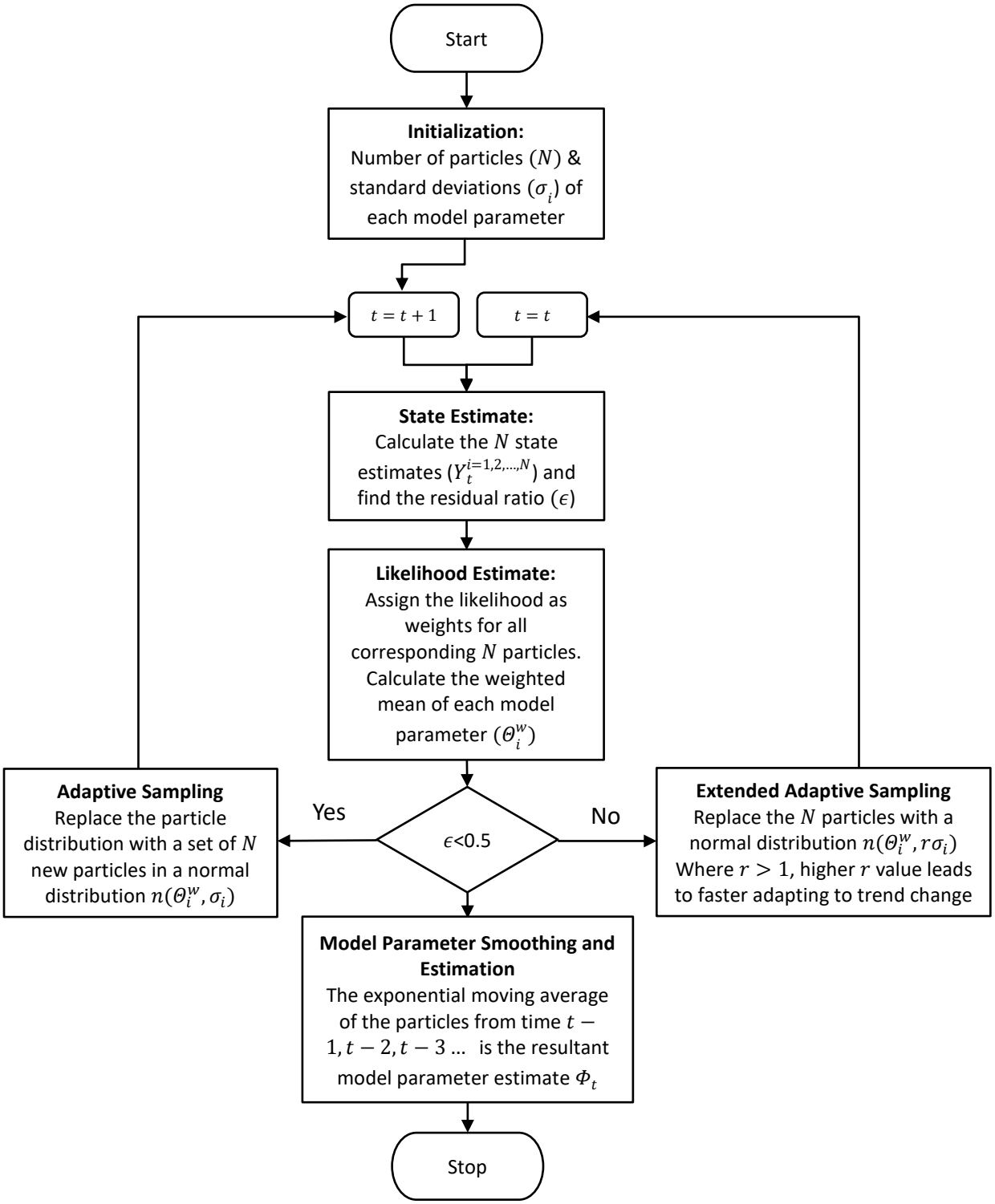

Fig. 4.1 Flow chart of the proposed enhanced adaptive particle filter.

Consider that a trend change occurred from gradual degradation phase to accelerated degradation phase and model parameter value changed drastically from $\Phi_{t-1} = 2$ to $\Phi_t = 8$. As can be seen in Fig. 4.2, the required new value of Φ_t is not contained in the particle distribution $p(Y_{t-1}|Z_{1:t-1})$ at time step $t-1$. This results in state estimate diverging from the actual observation causing the residual ratio (ϵ) to increase beyond the prescribed value ($\epsilon > 0.5$). Therefore, the extended adaptive sampling process gets activated. In general, this process recursively loops through the same time step until the particle distribution converges to the new model parameter value (Φ_t). Where, the corresponding weights obtained from the likelihood estimation step, are used to move the particle distribution towards the direction of the required value of Φ_t. As seen in Loop 1, the mean of the particle distribution moves towards the region of higher weights meanwhile the spread of the particle distribution is increased using the spread factor (r). Wider spread can be used to achieve quicker convergence (i.e., in fewer loops). To widen the spread of the particle distribution, the preassigned standard deviation is increased by r times. Hence, as described in step 5 (extended adaptive sampling), new particle distribution is $n(\Theta_i^w, r\sigma_i)$, where, r is the spread factor such that r>1.

Further, in Loop 2 the weights of the particles obtained in Loop 1 are used to move the particle distribution again towards the higher weight region. Such recursive loops are performed in the extended adaptive sampling process until the required value of Φ_t is within the particle distribution. Once the required value is finally within the particle distribution range, the residual ratio reduces below the prescribed value ($\epsilon < 0.5$), and the extended adaptive sampling process stops. Finally, once the particle distribution converges to the required value, the spread of the particle distribution is reduced and the new particle distribution is $p(Y_t|Z_{1:t}) = n(\Theta_t^w, \sigma_i)$, as shown in Fig. 4.2. This reduced spread after convergence creates enough particles around the new model parameter ($\Phi_t = 8$) while maintaining a diversity of the particles for future time steps.

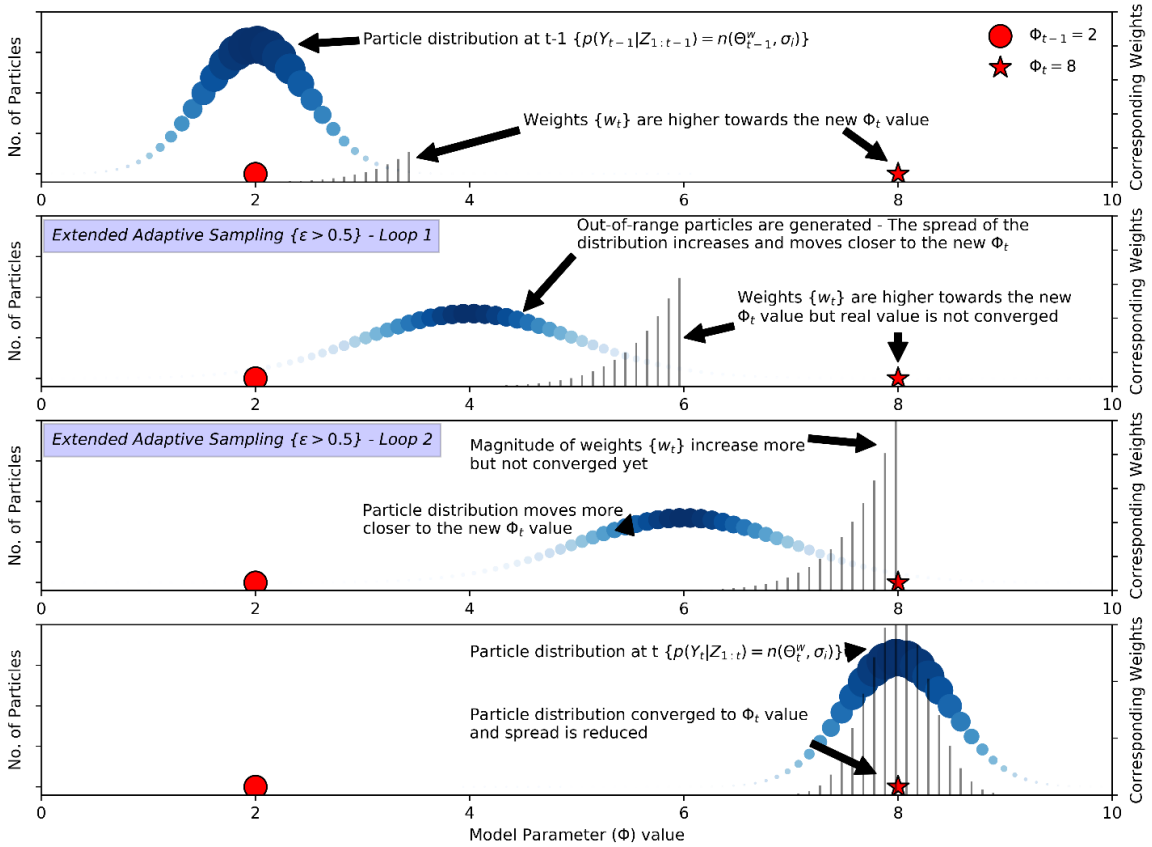

Fig. 4.2 Illustration of the extended adaptive sampling process.

Finally, the smoothing step reduces the fluctuations in the model parameter estimates and helps in obtaining consistent model parameter values. The enhanced adaptive PF algorithm has high fidelity in model parameter estimation and has lower computational requirements compared to the standard PF. The standard PF would take a significantly larger number of particles with a larger range to attain comparative results.

4.4 Model-based Prognostics with Enhanced Adaptive Particle Filter

4.4.1 Model-based Prognostics

In model-based prognostics, the degradation of a bearing or a mechanical system is expressed using a degradation model. The degradation model is also known as the system

model, is a function of the elapsed time and model parameters. In this section, we develop the degradation model using the Paris-Erdogan model. Paris-Erdogan model, also known as Paris law, is a widely used degradation model for bearings and gear system prognostics [55]. Paris law is originally expressed as:

$$\frac{dx}{dn} = c[\Delta k]^{\gamma} \tag{4.9}$$

and:

$$\Delta k = \omega\sqrt{x} \tag{4.10}$$

Where x represents the semi-crack length, n is the number of stress cycles, and c, γ and ω are material constants, these values are determined by the experimental tests. Δk is the amplitude of the stress intensity factor and it is proportional to the square root of x.

For applications in PHM, the equations (4.9) and (4.10) are transformed into a convenient form where the three model parameters, i.e., $c, \gamma,$ and ω are condensed to two model parameters α and β, such that:

$$\frac{dx}{dt} = \alpha x^{\beta} \tag{4.11}$$

Where, $\alpha = c\omega^{\gamma}$ and $\beta = {}^{\gamma}/_{2}$ and the number of stress cycles, n, is replaced by the elapsed time, t, of the experiment.

Therefore, the equation (4.11) is rewritten to form the state-space model as:

$$x_{k+1} = x_k + \alpha_k x_k^{\beta_k} \Delta t \tag{4.12}$$

Where, x_k represents the system health state at the time t_k and $\Delta t = t_{k+1} - t_k$. The model parameters α_k and β_k are the variables which are to be estimated using the progressing system state. However, during the bearing life cycle, the bearing goes through various phases of degradation like gradual degradation and accelerated degradation. Throughout

these phases, the values of the model parameters (α and β) vary drastically especially during the phase change periods. PFs are one of the most used techniques in literature to estimate these model parameters based on the available data [8]. In several papers, the standard PFs produce good estimates, but these estimates are typically only for the final phase of bearing degradation, i.e., the accelerated degradation phase, whereas the proposed enhanced adaptive PF performs well during the bearing degradation phases as well as during the phase change periods.

4.4.2 Developing a Simulated Bearing Health State

In this section, the working of the enhanced adaptive PF is illustrated by comparing it with a standard PF (PF with Sequential Importance Sampling). For this purpose, the bearing health state is simulated using the degradation model presented in (4.12). The health state, shown in Fig. 4.3, is simulated such that it contains two degradation phases and an anomalous region where the signal amplitude is irregular. The degradation model in (4.12) has two model variables alpha (α) and beta (β). In this simulation, model parameter β is kept constant such that $\beta = 0.5$. Whereas the model parameter α is varied for both the phases of degradation such that $\alpha_1 = 0.15$ in the first phase and $\alpha_2 = 0.5$ for the second phase.

The anomalous region of the bearing health state spans from $80\ sec$ to $120\ sec$ where the signal amplitude is reduced by 30%. A trend change is seeded representing the initiation of accelerated degradation phase after $t = 150\ sec$. Finally, a nominal white gaussian noise is added to the simulated bearing health state such that the signal to noise ratio (SNR_{dB}) $= 45$ dB. Signal to noise ratio (SNR) is defined as the ratio of signal power to the noise power (lower the SNR ratio, higher the noise power).

$$SNR_{dB} = 10log_{10}\left(\frac{P_{Signal}}{P_{Noise}}\right), P = E\left[S_{t=1,2\ldots}{}^2\right] \tag{4.13}$$

Where, P_{Signal} is the signal power, P_{Noise} is the noise power and E is the expected value in our case, power is given by the mean of the squared value of the signal.

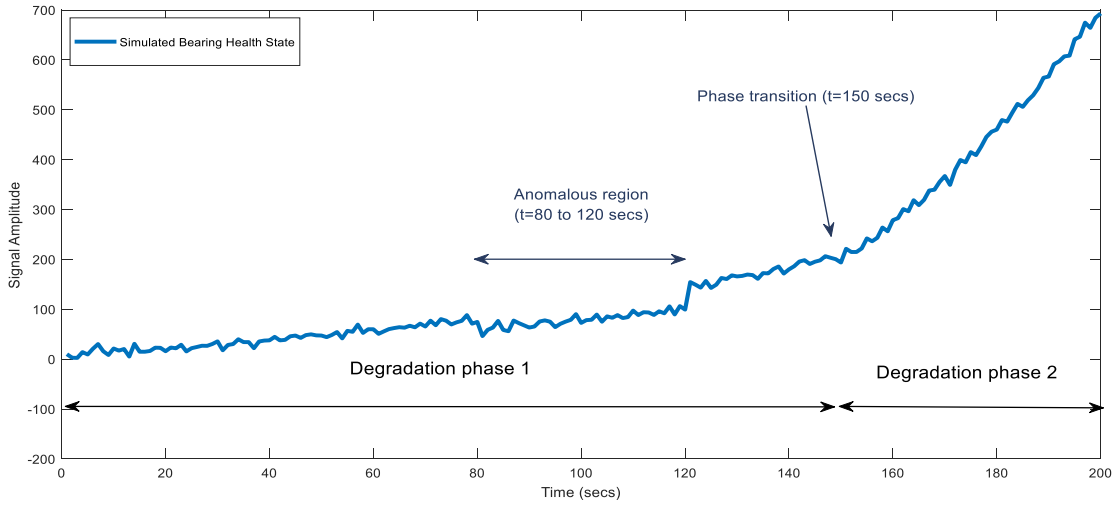

Fig. 4.3 Simulated bearing health state.

4.4.3 Initializing particle filter

The PF is initialized in between the anomalous region, i.e., at $t = 110\ sec$. Therefore, the health state from 0 to 110 sec is used to get an initial estimate of α using the LSE curve fitting method. As shown in Fig. 4.4, the curve fitting method estimates α = 0.116. Both the standard PF and enhanced adaptive PF are initialized such that α will follow a uniform distribution where the initial estimate (α = 0.116) is the mean of the distribution. The standard PF is initialized with 10000 particles whereas, the enhanced is initialized with only 1000 particles. The results of the state estimation and the model parameter estimation are shown in Section 4.4.4.

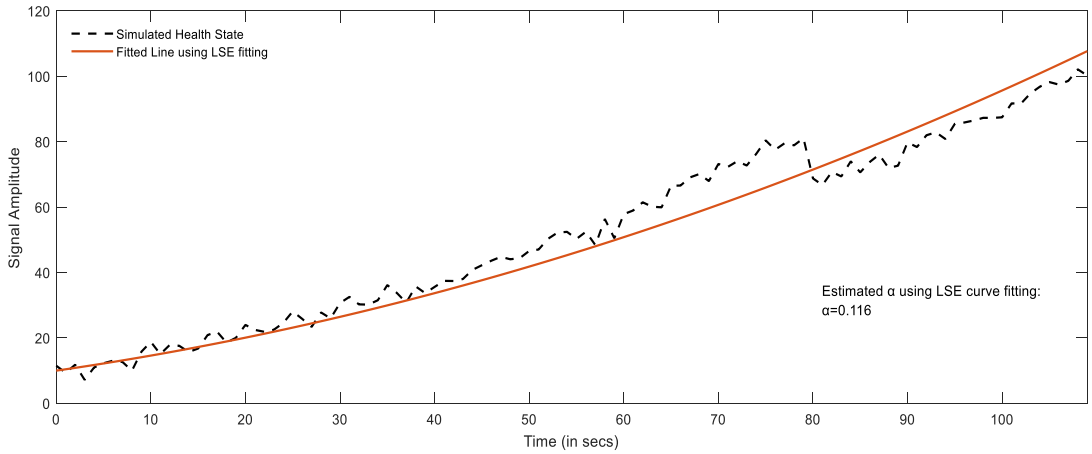

Fig. 4.4 Particle filter initialization and estimation of initial α using Least Square Error (LSE) curve fitting.

4.4.4 Model Parameter Estimation

4.4.4.1 Particle Filter with SIS

Fig. 4.5 (a) shows the result of the state estimation and Fig 4.5 (b) shows the corresponding range and the mean of the model parameter (α).

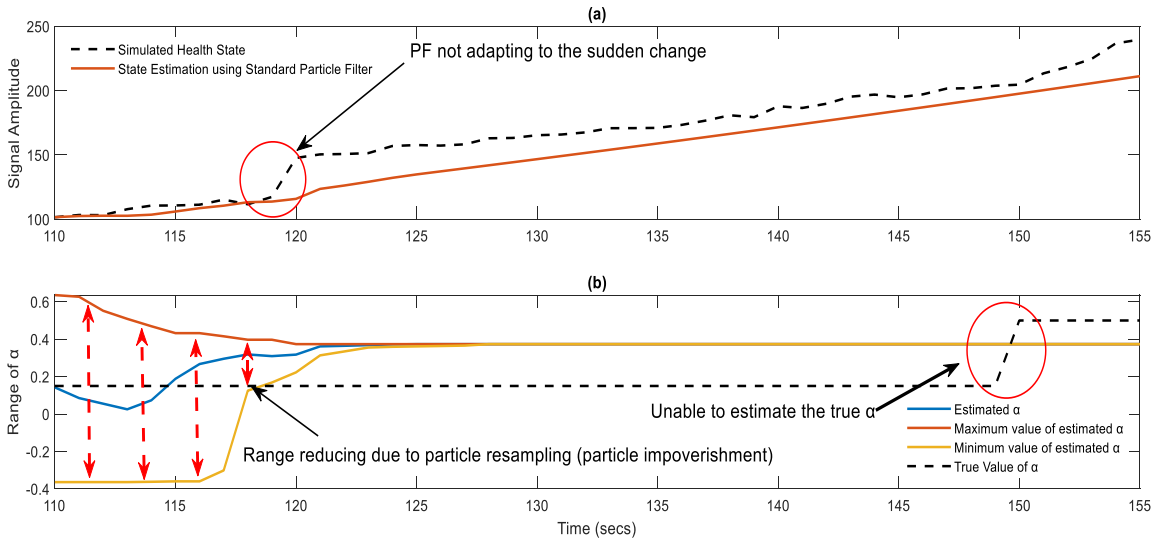

Fig. 4.5 Standard PF: (a) State estimation and inadaptability to trend change (b) Particle range reduction and particle impoverishment.

The initial range for the model parameter α (at $t = 110\ sec$) is between -0.4 and 0.6. With every time step, the SIS replaces the low-weight particles with the high-weight particles. Therefore, in Fig. 4.5 (b). we can clearly see the range of the particles reducing with time because the SIS replaces the values in the extremes of the range (low weight) with values of higher weight. This replacement of the particles gradually reduces the number of unique particles causing severe particle degeneracy. Therefore, by $t = 120\ sec$, the range of the particles reduces dramatically as many particles are already replaced with high-weight particles. After $t = 125\ sec$ there are less than 10 unique particles left from the initial 10000 unique particles. This is known as a severe case of particle impoverishment. Meanwhile, the state estimation could not adapt to the sudden change at time $t = 120\ sec$. The state estimation is well off the real value till $t = 150\ sec$. Further, when the phase change takes place, the PF cannot adapt itself to the new trend. The model parameter estimation of the standard PF is also incorrect as the mean of the range of the model parameter α, in Fig. 4.5 (b). is around 0.3, while the true α is 0.15.

4.4.4.2 Enhanced Adaptive PF

Fig. 4.6 (a) shows the result of the state estimation and 4.6 (b) shows the corresponding range and the mean model parameter (α) estimated using the enhanced adaptive PF.

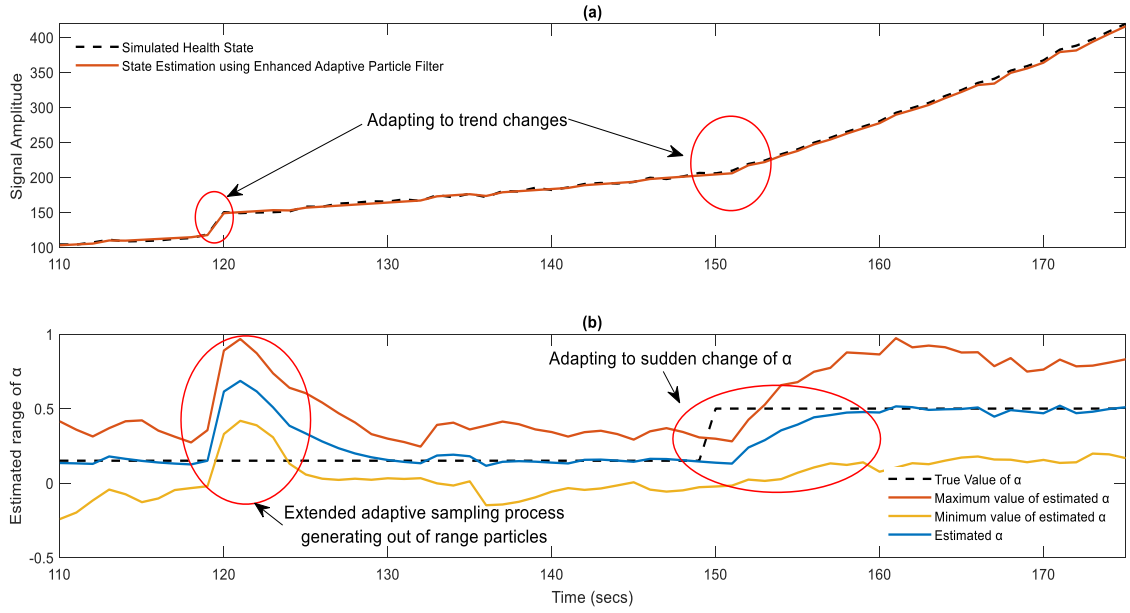

Fig. 4.6 Enhanced adaptive PF: (a) Accurate state estimation (b) Particle range adapting to the sudden changes.

In the enhanced adaptive PF, the initial range of the model parameter α is between -0.2 and 0.48. With every time step, the adaptive sampling algorithm generates newer particles such that the particles follow a normal distribution with the mean obtained using the likelihood estimation step (in Section 4.3). This makes sure that there will be more relevant particles closer to the actual value of the model parameter and fewer particles farther from it. Since there are more unique particles near the actual parameter value, the fidelity of the enhanced adaptive PF is very good even while using fewer particles compared to the standard PF. We can see that the state estimation and the model parameter estimation are both very accurate using this method. At $t = 120\ sec$, the extended adaptive sampling will be activated and it generates particles outside the normal range of particles, as shown in Fig. 4.6 (b), so that the extreme changes in the health state can also be adapted. Therefore, we see the range of the model parameter α varying dramatically, but soon coming back to the actual value in the next 10 seconds. From $t = 150\ sec$, the state estimation adapts to the trend change accurately meanwhile, the model parameter can be observed adapting to the correct value quickly. Hence, with these results, we can show that

the enhanced adaptive PF resolves all three main problems of the standard PFs (i.e., particle degeneracy, particle impoverishment, and initial range dependency).

4.4.5 Effect of Noise on Model Parameter Estimation

In this section, we evaluate the performance of the enhanced adaptive PF in model parameter estimation when the noise in the health state is increased gradually. Additionally, the performance of enhanced adaptive PF is compared with an efficient Bayesian-based prognostics algorithm using MCMC (Monte Carlo Markov Chain) sampling which is known for good noise tolerance [59]. For this evaluation, a simulated health state is generated using the state-space model in (4.12), in which there are two degradation phases. The first phase has $\alpha = 0.15$ representing gradual degradation and the second phase has $\alpha = 0.5$ representing accelerated degradation of the bearing. Finally, noise is added to the simulated health state. Noise power is gradually increased in steps (see Fig. 4.7), where the SNR_{dB} ranges from 45 dB to 30 dB (reducing SNR_{dB} means increasing noise). Root Mean Square Error (RMSE) between the true value of model parameter α and the estimated α is used as the metric to compare the performance of both techniques.

As shown in Fig. 4.7, the α estimation of enhanced adaptive PF is very accurate compared to the efficient Bayesian-based prognostics algorithm for the health states with medium to high noise, i.e., $SNR_{dB} = 45, 43, 41, 39$ & 37. However, the ability to estimate the model parameter drastically decreases for very high noise, i.e., for SNR_{dB} $35, 33$ & 30. This is due to the fact that the amplitudes of the system health states fluctuate a lot for very high noise states. Therefore, the likelihood estimation performed on these fluctuating health states causes inaccuracies in model parameter estimation. However, the estimations using enhanced adaptive PF are reasonably smoother without high oscillations compared to the Bayesian method. In both the methods, the majority of this error in parameter estimation is caused in the period immediately after the phase change ($150\ sec$).

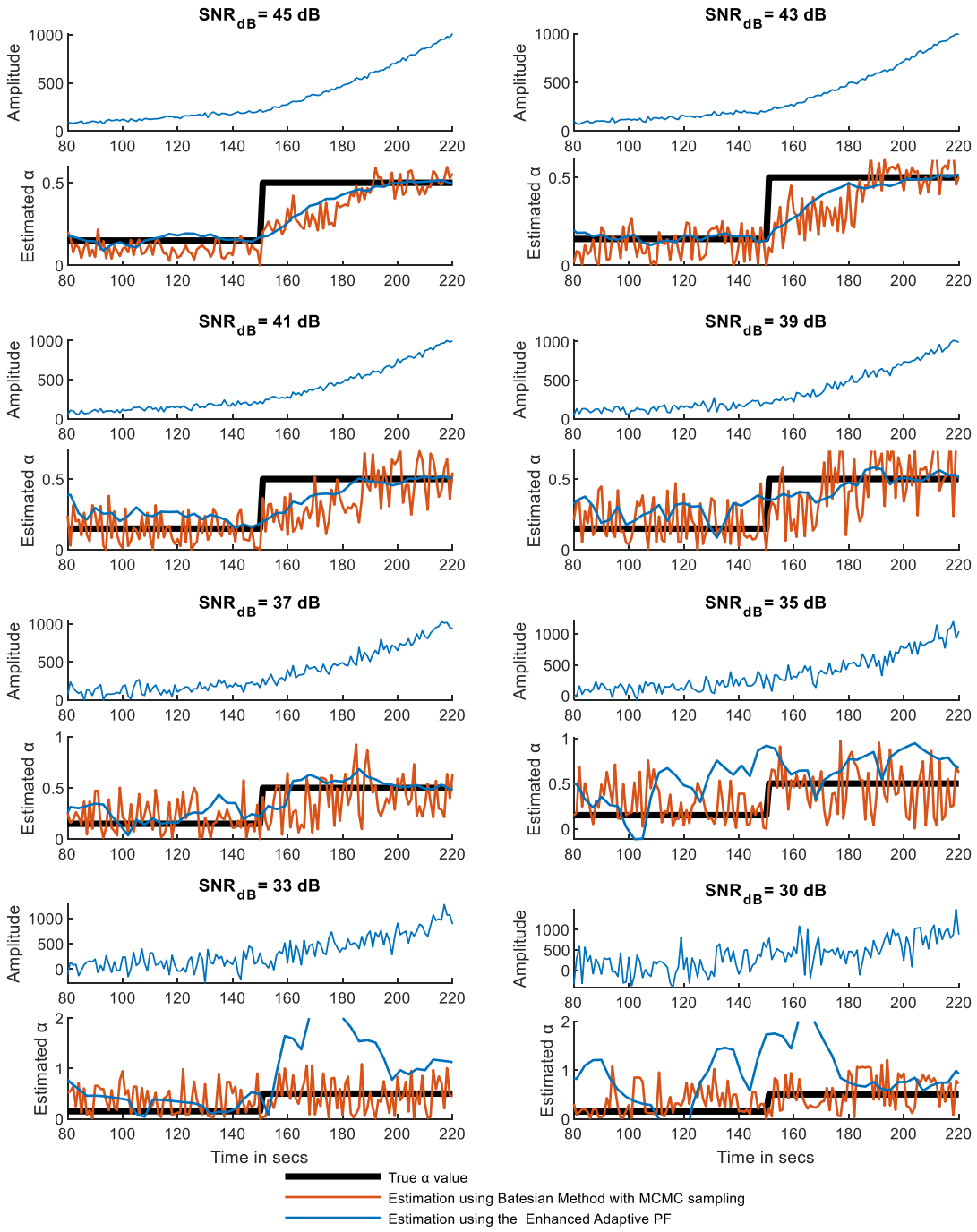

Fig. 4.7 Comparing the model parameter estimations using enhanced adaptive PF and Bayesian method with MCMC sampling.

The bar graph in Fig. 4.8 compares the RMSE values between the true α and estimated α for both methods. We can see that model parameter estimation using the enhanced adaptive PF is comparatively more accurate than the Bayesian-based method for the medium and high noise states ($SNR_{dB} = 45, 43, 41, 39$ & 37). Whereas, in the states with very high noise (SNR_{dB} $35, 33$ & 30), the errors in the model parameter estimation of both the methods are high. The RMSE of the Bayesian-based prognostics algorithm increases gradually with increasing SNR_{dB}. Whereas, the RMSE of the enhanced adaptive PF is fairly stable in the medium to high noise states but drastically increases in the very high noise states.

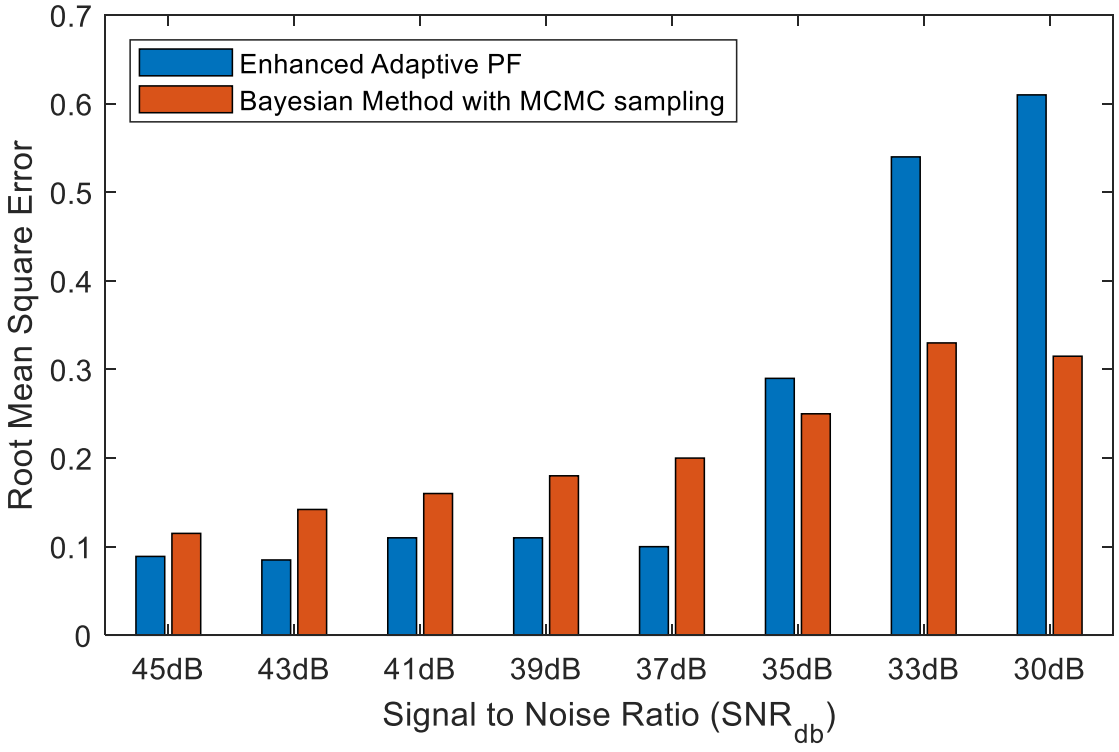

Fig. 4.8. Comparing the RMSE between enhanced adaptive PF and Bayesian method with MCMC sampling.

4.5 Summary

The proposed enhanced adaptive PF technique addresses the issues of particle degeneracy, particle impoverishment, and initial range dependence. In this technique, an adaptive sampling process is introduced which approximates the particle distribution as a normal distribution. Whereas the particle weights are approximated directly as the maximum likelihood estimate of the distribution. This helps in generating more relevant particles around the required model parameter value while maintaining diversity in the particle distribution. The normal distribution approximation eliminates the particle resampling step, which exists in almost all PF techniques, and therefore reduces the computational requirement. A residual ratio is used to activate the extended adaptive sampling process when adverse trend changes are detected. This extended adaptive sampling process generates out-of-range particles and recursively converges to the required model parameter values in the same time step to adapt to the adverse trend change. The working of the proposed technique is illustrated using simulated data with phase changes. Further, a noise sensitivity study is performed where the accuracy in model parameter estimation is tested with increasing levels of noise in the simulated data. The technique shows good accuracy in model parameter estimation and degradation prediction even in case of adverse trend changes in the degradation data.

Therefore, the proposed enhanced adaptive PF technique is used for the real-time multi-phase degradation prediction and RUL estimation in this intelligent PHM system.

Chapter 5: A Multi-Phase Degradation Model based on Mixture of Gaussian Process and Weibull Functions

5.1 Overview

Model-based prognostic approaches typically use a single physics-based degradation model to characterize the entire service life of the bearings even though the service life has more than one degradation phase. The approximation that the entire service life of bearings can be characterized by a single degradation model will lead to restricted robustness of the long-term RUL prediction process [5]. In other words, these types of approximations will work well only for short-term RUL prediction. Therefore, in [5], more than one degradation model is used where one model corresponds to each degradation phase of bearing. For example, a linear model is used for the gradual degradation phase and an exponential model is used for the accelerated degradation phase. Even though the results have shown good long-term prediction, the accuracy of the method completely depends on the algorithm rather than on the degradation model.

Therefore, to deal with this issue, attempts have been made to develop degradation models that accommodate the multi-phase degradation process of the mechanical equipment. In [87], [88], and [89], multi-phase degradation models based on the Wiener process are developed and applied to plasma panels and gyroscopes. Apart from the Wiener process models, no other degradation model has been considered for multi-phase degradation processes so far. More importantly, till now, no multi-phase degradation models have been proposed for bearing RUL estimation. Therefore, in this chapter, a multi-phase degradation model is developed which characterizes the overall degradation of the bearings.

Gaussian process models are well suited to various nonlinear dynamic systems with high fluctuations and sudden trend changes. The Gaussian process model approach to curve fitting was first introduced in [111] and was later compared to other non-linear models in [112], which led to the rapid expansion of research into Gaussian process models [113]. In the statistic and machine learning community, Gaussian models are widely utilized in

classification, and regression problems. The Gaussian process model is an example of a probabilistic model that provides information about prediction uncertainties that are difficult to evaluate in nonlinear parametric models [90]. The Gaussian process is typically defined by the mean and the variance. This is best suited in machine learning because it has fewer hyper-parameters to estimate [114]. It was used in the prediction of immunological diseases in [91], and is later adapted for use in predictive control process studies in [90].

In the literature, Weibull distribution has been effectively shown to approximate the bearing degradation process. In [63], the bearing vibration was approximated using the Weibull failure rate function and accurate RUL prediction was obtained. Further in [115], a Weibull accelerated failure time regression model was presented for bearings working under multiple operating conditions for the RUL prediction. In [116], a log-linear Weibull distribution-based method was presented to estimate the RUL of bearings.

Therefore, in this study, a mixture of the Gaussian process and Weibull failure rate model is considered to characterize the multi-phase degradation of rolling element bearings.

5.2 Phase Transition Mass of the Bearings

The experimental study in [117], identifies that the spall propagation always happens in two phases after fault initiation, i.e., gradual degradation and accelerated degradation. In 2005, US Air Force Research Laboratories conducted bearing tests with seeded defects where ODMs were used to monitor the debris quantity and generation rates [118]. It identified a strong relationship between mass lost from the bearings and the degradation phase change. Fig. 5.1 summarizes the study in [118], which shows the mass lost from several 52100 steel bearings which were conducted at different operating conditions. Even though so many operating condition variations exist, all the bearings undergo the phase change from gradual degradation to accelerated degradation phase around 60 mg

(irrespective of the operating condition). This "critical mass loss" varies for different bearings based on the geometry and materials.

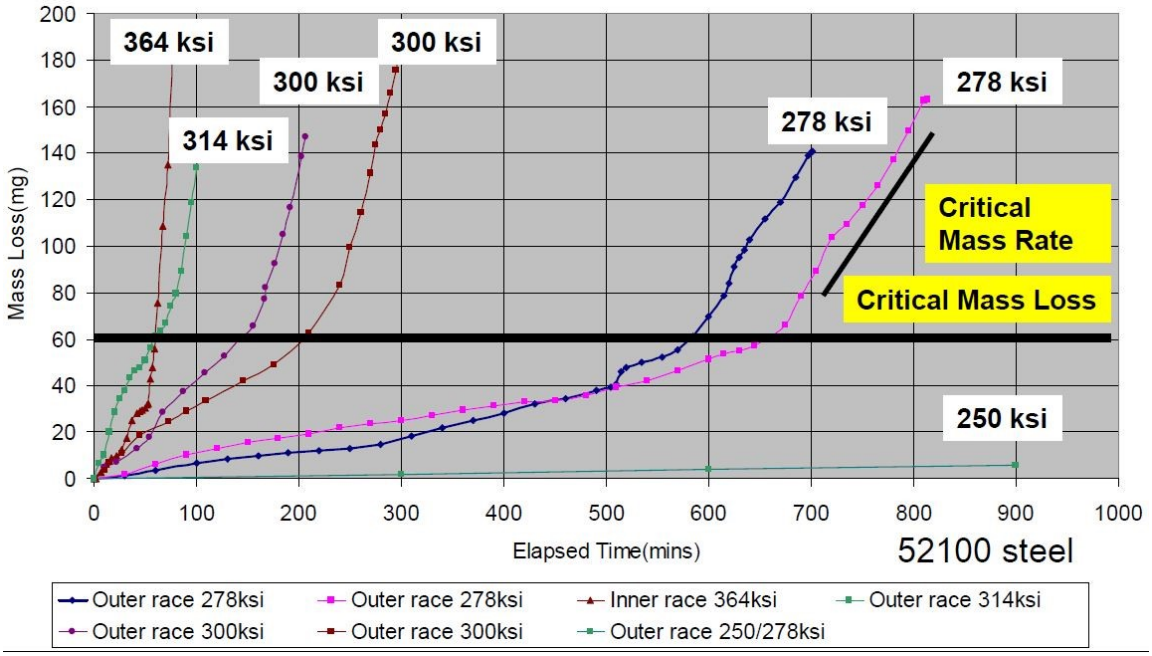

Fig. 5.1 Bearing mass loss trend curves for bearings from life endurance tests [118].

The article in [117] is one of the three-part series from Lewis Rosado; it investigates the rolling contact fatigue initiation and spall propagation characteristics of bearing with three different materials, namely, AISI 52100, VIM-VAR AISI M50, and VIM-VAR M50NiL steels. Fig. 5.2, shows the bearing mass lost from bearing spall progression tests conducted at a constant operating condition. Where we can see that the phase change occurs when the mass lost is between 15 to 40 mg. More specifically, the critical mass lost from bearings of similar material is closer to each other. For example, the M50NIL bearings have a phase change near 40 mg, whereas AISI 52100 bearings around 15 mg. However, it can be seen that the rate of spall propagation is different for every individual bearing even of the same material.

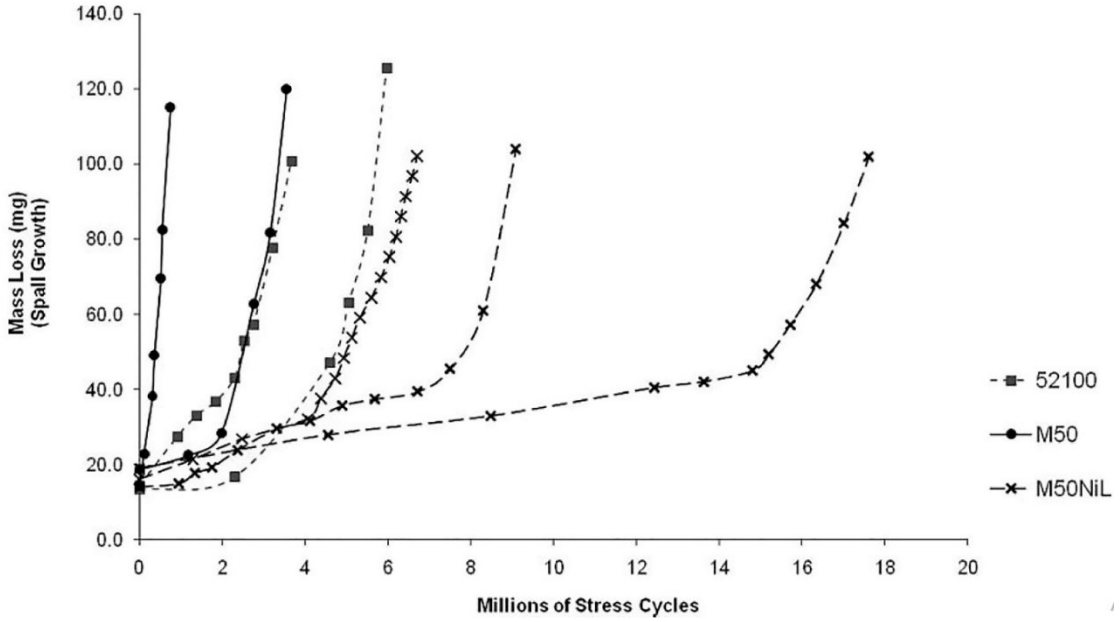

Fig. 5.2 Bearing mass loss trend curves for bearings from life endurance tests. All at 2.41 GPa [117].

Therefore, in this study, the "critical mass loss" value for the bearing is denoted as phase transition mass. The phase transition mass of individual bearing can be identified from the spall propagation tests. In reality, the phase transition mass is around a range of values because of the stochastic nature of bearing degradation and the limitation of ODMs in accurately calculating the mass lost from the bearings. The range of phase transition masses for the bearings from the spall propagation experiments is calculated and is used in this study to construct the multi-phase degradation model.

5.3 Multi-Phase Degradation Model

5.3.1 Sum of Gaussians

In this study, the bearing debris mass generation rate is approximated using the sum of two Gaussian functions. Each Gaussian represents the debris mass generation rate of one degradation phase. In other words, one Gaussian function approximates the debris generation rate of the entire Phase 1 and another Gaussian approximates the debris

generation rate of the entire Phase 2. As shown in Fig. 5.3, the debris mass generation rate during the gradual degradation phase is approximated using the Phase 1 gaussian with a smaller height. Similarly, the debris mass generation rate of the accelerated degradation phase is approximated using the Phase 2 (taller) gaussian. We also see in Fig 5.3, that this contention fits one of the spall propagation experiments very well such that both degradation phases are distinguishable.

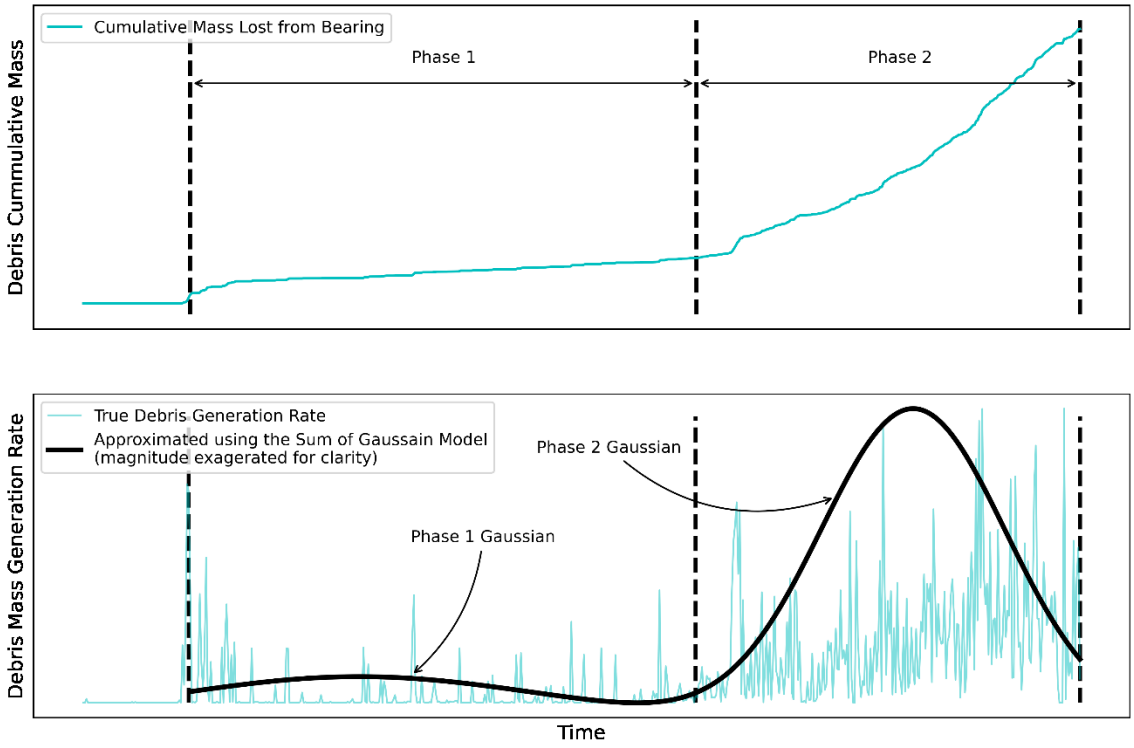

Fig. 5.3 Approximation of the debris generation rate using the sum of Gaussians (magnitude of gaussian exaggerated for clarity).

A three-parameter Gaussian is used in this model instead of the two parameters to effectively approximate the degradation index. Therefore, as per the contention the debris mass generation rate during the bearing spalling can be given as:

$$G(t) = \sum_{n=1}^{2} h_n exp\left[-\frac{(t-\mu_n)}{\sqrt{2}\sigma_n}\right]^2 \tag{5.1}$$

$$G(t) = h_1 exp\left[-\frac{(t-\mu_1)}{\sqrt{2}\sigma_1}\right]^2 + \ h_2 exp\left[-\frac{(t-\mu_2)}{\sqrt{2}\sigma_2}\right]^2 \quad (5.2)$$

Where t is the time, the subscript n denotes the degradation phase, i. e., $n = 1$ for the gradual degradation phase and $n = 2$ for the accelerated degradation phase. h is the height, μ is the mean, σ is the standard deviation of the Gaussian.

As shown in Fig. 5.3, the physical meaning of the height h represents the maximum debris mass generation rate of each phase and the mean μ corresponding to the time at which the maximum debris generation rate occurs. Whereas the standard deviation σ represents a fraction of the duration of each phase. As can be seen, σ is directly proportional to the total duration of the phase.

This sum of Gaussian model gives us 6 unknown model parameters which need to be deduced from the degradation data. However, we can use geometrical, physical and empirical relations from previous bearing spall propagation tests to reduce the number of unknown model parameters.

5.3.2 Modified Weibull Failure Rate

As described earlier, the Weibull Distribution has been found to reliably reflect the fatigue strength and fatigue life of mechanical components and it has been used as a failure model to analyze the reliability and maintainability of different mechanical systems. In several studies, Weibull distribution's failure rate function is fruitfully used to represent the evolution of the bearing failure. In [115], a bearing degradation model is developed using the Weibull failure rate function; it provides good RUL estimation even under multiple operating conditions. In [63], the Weibull distribution's failure rate function is used to approximate the vibration response of the bearing and it helps in representing the bearing degradation even during the high fluctuation zones. The probability density function and failure rate function of the Weibull distribution are shown in (5.3) and (5.4) respectively.

$$f(t,\rho,\eta,\gamma) = \frac{\rho}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\rho-1} \exp\left(-\left(\frac{t-\gamma}{\eta}\right)^{\rho}\right) \quad (5.3)$$

$$\lambda(t,\rho,\eta,\gamma) = \frac{\rho}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\rho-1} \tag{5.4}$$

Where f is the probability density function, λ is the failure rate function, t is the time instant, ρ is the shape parameter, η is the scale parameter and γ is the time position parameter. The time position parameter γ determines the position at which the failure rate begins; commonly the time position parameter is set to be zero [63]. Therefore, it can be positively assumed that the Weibull failure rate function approximates the overall growth of the debris generation rate of the bearings.

In this study, the three-parameter Weibull failure rate function (λ), is modified to a two-parameter form such that, $\gamma = 0$ and $\alpha = \eta^{\rho}/\rho$ and $\beta = \rho - 1$, as shown in (5.5).

$$\lambda(t,\alpha,\beta) = \frac{t^{\beta}}{\alpha} \tag{5.5}$$

This influence of the new shape parameter β on the modified Weibull distribution's failure rate function is shown in Fig. 5.4. The shape parameter $\beta > 0$ represents an increasing failure rate, $\beta < 0$ represents a decreasing failure rate and $\beta = 0$ represents a constant failure rate. Whereas the scale parameter (α) influences the rate of evolution, i.e., higher α leads to faster evolution of λ.

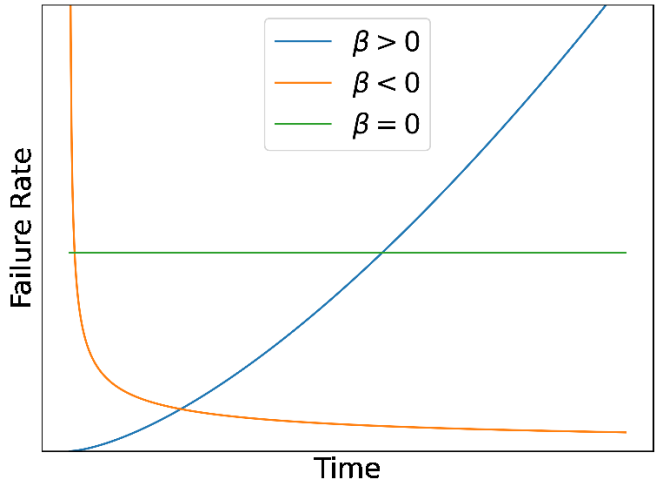

Fig. 5.4 Influence of β parameter on the modified Weibull failure rate function.

As explained earlier, that the modified two-parameter Weibull failure rate function is used to represent the overall debris generation rate of bearings. Therefore, this helps us to deduce that the peaks of the two Gaussians can be quantified using the modified Weibull failure rate function, in (5.5).

Fig. 5.5 shows the modified Weibull failure rate and the sum of Gaussians model (equation 5.2) fitted to the true debris generation rate of a bearing spall propagation experiment. We can see that the peaks of the two Gaussians, i.e., h_1 and h_2 touch the modified Weibull failure rate function. Therefore, the locations μ_1 and μ_2 can be substituted in (5.5), to get the heights of the Gaussians.

$$h_1 = \frac{{\mu_1}^{\beta}}{\alpha} \;\&\; h_2 = \frac{{\mu_2}^{\beta}}{\alpha} \tag{5.6}$$

$$h_2 = h_1 \left(\frac{\mu_2}{\mu_1}\right)^{\beta} \tag{5.7}$$

Further, a relation between h_1 and h_2 is generated by simplifying (5.6). This relation between the heights can help us eliminate one model parameter from the model in (5.2).

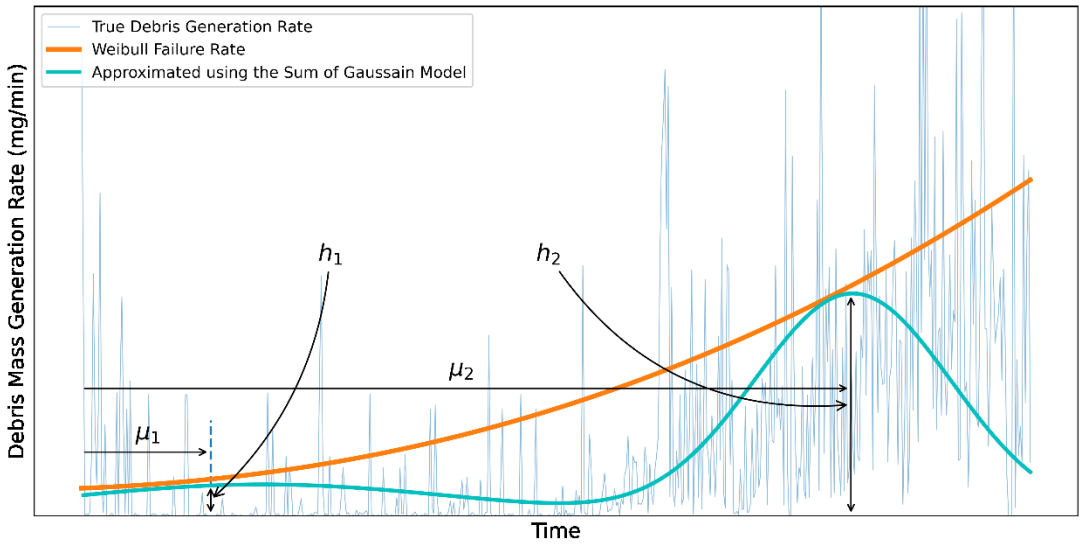

Fig. 5.5 Modified Weibull failure rate approximates the overall growth of the debris generation rate.

5.3.3 Phase Transition Mass

In this section, the physical significance of the phase transition mass is used to form a mathematical relationship between the model parameters in the sum of gaussian model. As discussed earlier, the phase transition mass is the cumulative mass lost from bearing after which the gradual degradation phase changes to the accelerated degradation phase. As per the initial assumption of the model, each Gaussian represents the debris generation rate of one entire phase. Therefore, the area under Phase 1 gaussian (gradual degradation) should be equivalent to the phase transition mass (M_T) of the bearing (see Fig. 5.6). Since we know that 95.5% of the area under the standard Gaussian distribution is between the 2 standard deviations. Thus, the phase transition mass can be assumed to be approximately equal to the area under the Phase 1 gaussian between the 2 standard deviations. In other words, the area of Phase 1 gaussian between $\mu_1 - 2\sigma_1$ and $\mu_1 + 2\sigma_1$ will be approximately equal to the phase transition mass. See Fig. 5.6, the bottom graph shows the debris mass generation rate approximated by the two Gaussians whereas, the upper graph shows the mass lost from the bearing as a cumulative sum of the bottom graph. Therefore, as described, the highlighted area under Phase 1 Gaussian is approximately equal to the phase transition mass.

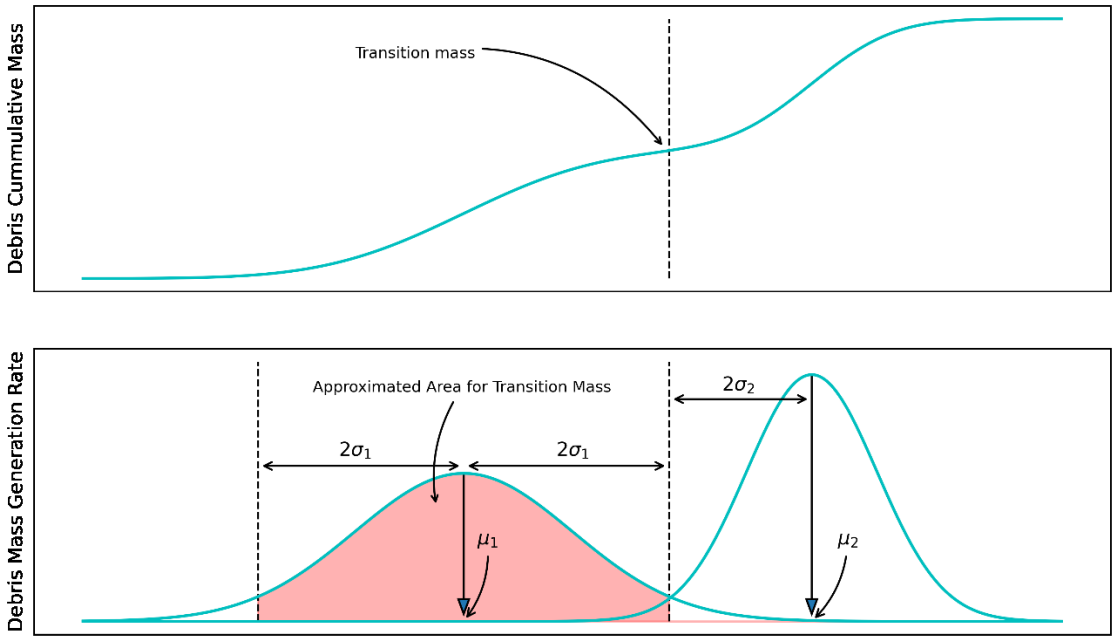

Fig. 5.6 Area under Phase 1 Gaussian approximately equal to the phase transition mass.

$$M_T \cong \int_{\mu_1-2\sigma_1}^{\mu_1+2\sigma_1} h_1 exp\left[-\frac{(t-\mu_1)}{\sqrt{2}\sigma_1}\right]^2 dt \tag{5.8}$$

Equation (5.8) is known as the gaussian integral and it is not possible to solve the integral using fundamentals of calculus. Although some straightforward solutions are available for the indefinite integral for the Gaussians, i.e., $\int_{-\infty}^{+\infty} e^{-x^2} dx$, the integration between two finite limits is very difficult to evaluate. Therefore, the Gaussian Integral is approximated using an elemental division method. In this elemental division method, a finite number of elemental rectangles are used to approximate the Gaussian, as shown in Fig. 5.7. The precision of this approximation is directly proportional to the number of elemental divisions of the Gaussian [119, 120].

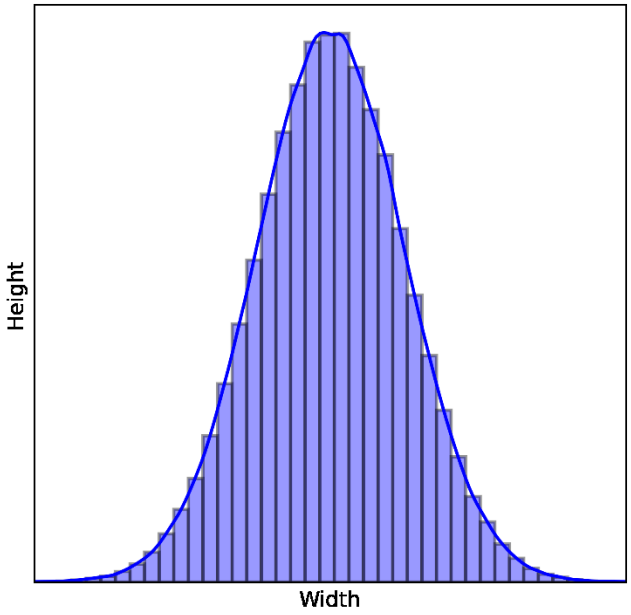

Fig. 5.7 Approximation of Gaussian integral using elemental divisions.

Consider the ratio of the area Gaussian (in equation 5.8) to a rectangle with the height of the gaussian (h) and the width from $\mu_1 - 2\sigma_1$ to $\mu_1 + 2\sigma_1$. Fig. 5.8 (a) shows the Gaussian embedded inside the rectangle. This ratio is evaluated using the elemental division method with the increasing number of elemental divisions of the gaussian. As we can see in Fig. 5.8 (b), the ratio asymptotes to a constant value as the number of divisions are increased. Therefore, using this ratio, the value for transition mass (M_T) can be approximated. This value is a good approximation for any three-parameter Gaussians. Finally, from Fig. 5.8, we get the ratio of the area of Gaussian to the rectangle as:

$$\frac{\int_{\mu_1-2\sigma_1}^{\mu_1+2\sigma_1} h_1 exp\left[-\frac{(t-\mu_1)}{\sqrt{2}\sigma_1}\right]^2 dt}{h_1 * 4\sigma_1} = 0.59813 \tag{5.9}$$

$$M_T \cong 2.392 h_1 \sigma_1 \tag{5.10}$$

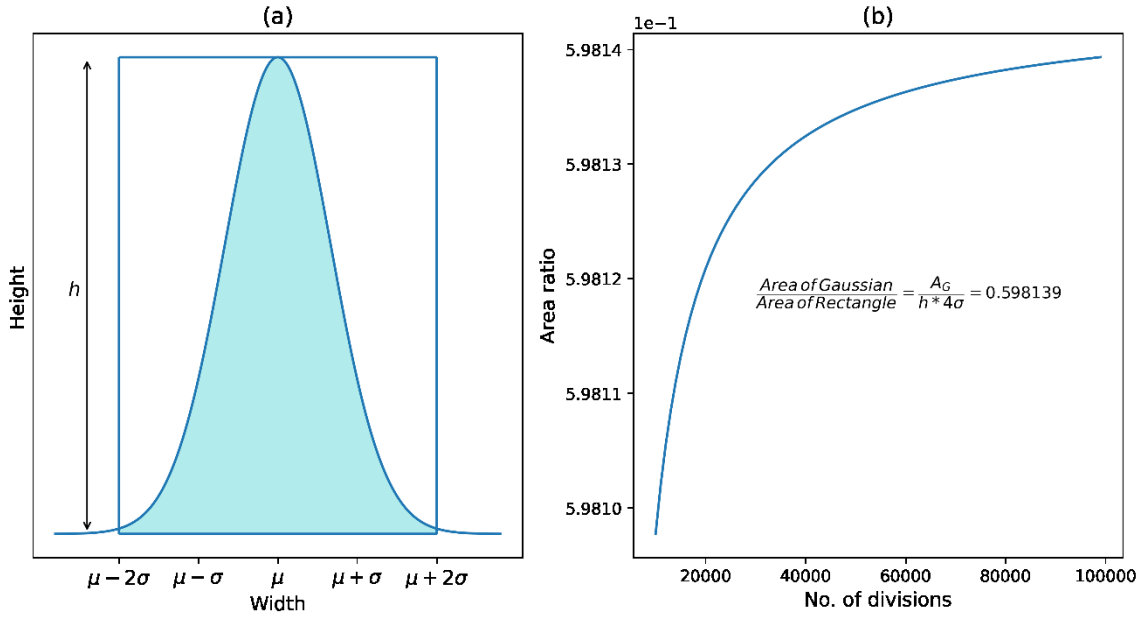

Fig. 5.8 Approximation of Gaussian Integral. (b) Area of gaussian compared with the area of the rectangle. (c) Area ratio asymptotes as the number of elemental divisions increase.

5.3.4 Geometric Relations

As seen in the previous section, the approximation that the Gaussians between the two standard deviations distance represents one degradation phase helped us extract the relationship between phase transition mass and the debris generation rate. This can also be extended to extract further geometric relationships between the unknown model parameters. As shown in Fig. 5.9, the approximation leads to a transition area where both phases contribute to the debris generation rate when the phase transition occurs from gradual degradation to accelerated degradation phase. Furthermore, since both the phases occur subsequently after each other, we can use the geometry of the Gaussians, in Fig. 5.9, to form a mathematical relation between μ_1 and μ_2.

$$\mu_2 = \mu_1 + 2(\sigma_1 + \sigma_2) \tag{5.11}$$

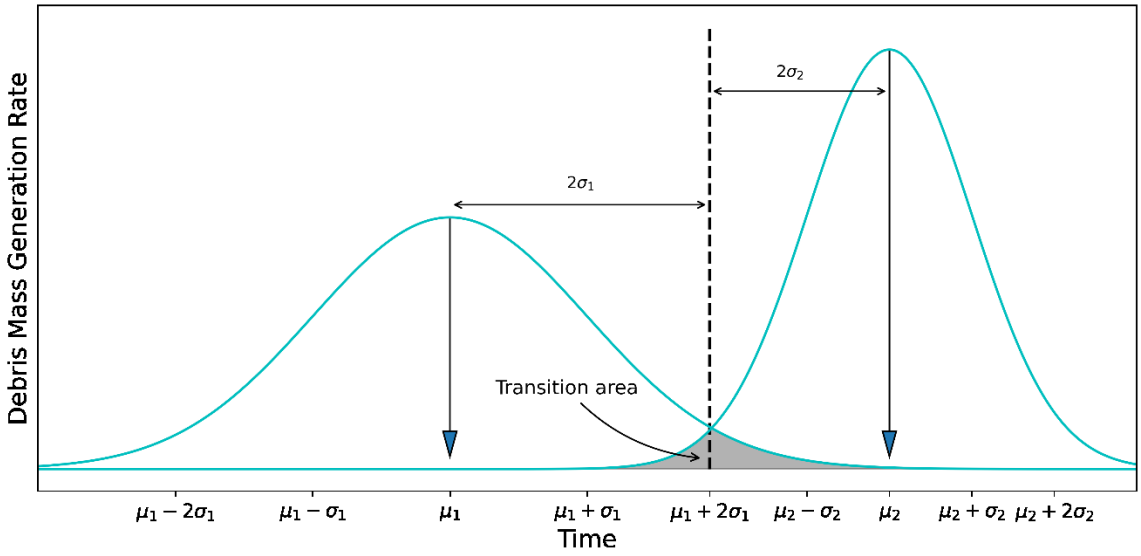

Fig. 5.9 Geometric relation between μ_1 and μ_2.

5.3.5 Empirical Relations

Several bearing spall propagation tests in this study, which were performed under relatively similar operating conditions show an empirical relation between the total duration of the phases. We observe that the ratio of Phase 2 duration and the Phase 1 duration tends to be a constant value. In other words, the duration *ratio* of the accelerated degradation phase *to* the gradual degradation phase remains constant for similar bearings. From Fig. 5.9, the total duration of each phase is equal to 4σ, we can form an empirical relation between σ_1 and σ_2 as shown:

$$\frac{T_2}{T_1} \cong \frac{\sigma_2}{\sigma_1} = \sigma_r \tag{5.12}$$

$$\sigma_2 = \sigma_r \sigma_1 \tag{5.13}$$

Where, T_1 and T_2 are the total duration of Phase 1 and Phase 2 respectively. σ_r is the ratio between both standard deviations. The value of σ_r can be obtained empirically from

previous spall propagation experiments. Chapter 6 discusses the exact empirical value for σ_r for the bearing spall propagation tests performed in this study.

5.4 Compiling the Degradation Model

To form the final degradation model, we can use the above relations to eliminate some of the unknown model parameters. Therefore, by substituting (5.7), (5.10), (5.11), (5.13) in (5.2), we obtain the final relation for the debris generation mass as follows:

$$G(t) = h_1 exp\left[-\frac{(t-\mu_1)}{\sqrt{2}\sigma_1}\right]^2 + h_1\left[\frac{\mu_1 + 2\sigma_1(1+\sigma_r)}{\mu_1}\right]^{\beta} exp\left[-\frac{(t-\mu_1-2\sigma_1(1+\sigma_r))}{\sqrt{2}\sigma_r\sigma_1}\right]^2 \quad (5.14)$$

$$h_1 = \frac{M_T}{2.392\sigma_1} \quad (5.15)$$

This is the consolidated degradation model that has three unknown model parameters β, σ_1 and μ_1. Whereas, M_T and σ_r are the phase transition mass and phase duration ratio respectively. Both M_T and σ_r are constants and are unique to every bearing; they are obtained using the empirical relationship from the previous spall propagation experiments.

Finally, the model is changed to a Markov chain form so that it can be in a convenient format for the enhanced adaptive PF.

$$C_m(t+1) = C_m(t) + \int_t^{t+1} G(t)dt \quad (5.16)$$

Where $C_m(t)$ is the cumulative mass lost from the bearing at time t, and $G(t)$ is the debris generation rate obtained from (5.15).

5.5 Summary

Although few multi-phase degradation models have been presented for non-linear applications, so far, no study has considered any multi-phase degradation model for rolling element bearings. In this chapter, we developed a multi-phase degradation model using the mixture of the Gaussian process model and Weibull failure rate function. The primary contention of this degradation model is that the debris generation rate for each bearing degradation phase can be approximated using a three-parameter Gaussian model. This contention is validated using the examples shown where the proposed model fits very well to the bearing spall propagation experiment. Thus, the sum of two Gaussians forms a two-phase degradation model however, this model has 6 unknown model parameters. Therefore, we use a modified Weibull failure rate function to form a relationship between two model parameters. Further, geometric and empirical relationships between the two Gaussians are used to reduce the number of unknown model parameters.

Finally, the degradation model is reduced to only three unknown model parameters to be estimated. Additionally, two empirical relationships namely the phase transition mass and phase duration ratio are extracted from previous spall propagation experiments of the same bearing. This multi-phase degradation model will be used for long-term RUL prediction in Chapter 6.

Chapter 6: A PHM Framework for Long-Term and Multi-Phase RUL Prediction

In this chapter, the techniques discussed so far will be compiled to create a PHM framework for fault phase diagnosis and long-term RUL prediction. In Chapter 3, the proposed data-driven ensemble-based fault phase diagnosis was developed and the validity of the tool was proven on two bearing spall propagation experiments. The results from those fault phase diagnoses are used in this chapter to perform multi-phase RUL prediction on the two spall propagation experiments. The Multi-phase degradation model proposed in Chapter 5 will be used as the state-space model in the enhanced adaptive PF (proposed in Chapter 4).

6.1 Proposed PHM Framework

The schematic diagram shown in Fig 6.1 summarizes the proposed intelligent PHM system. The proposed system has two major sub-systems namely, fault phase diagnosis and RUL prediction, and both of these sub-systems again have internal modules. The fault phase diagnosis tool runs at every time instant and estimates the fault phase, i.e., the tool estimates whether the bearing is in Phase 0, Phase 1, or Phase 2 which represents, normal operation, gradual degradation, and accelerated degradation phase respectively. Whereas RUL prediction works only when the bearing fault phase is in Phase 1 and Phase 2. The working of each sub-system is detailed below:

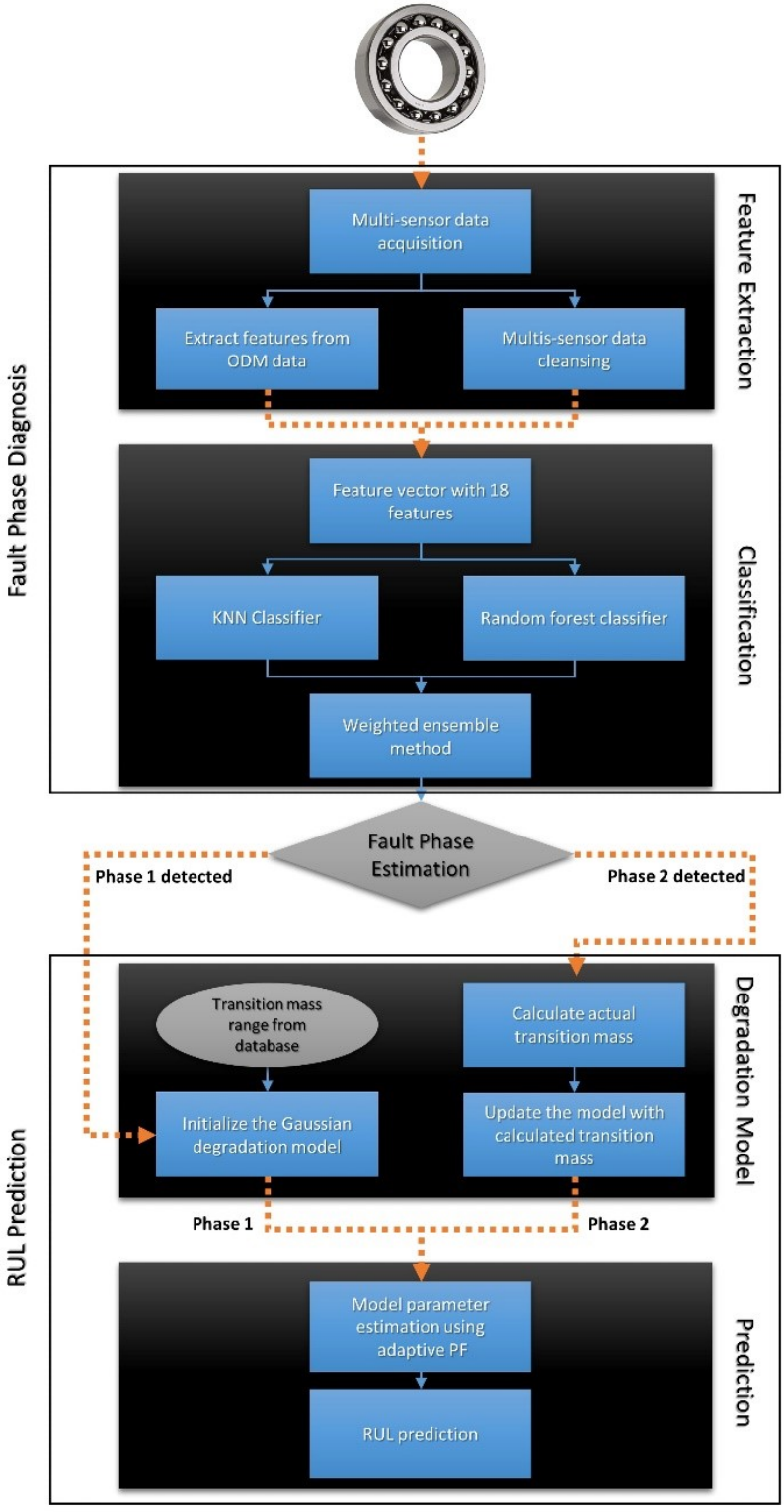

Fig. 6.1 Schematic diagram for the proposed PHM system.

6.1.1 Fault Phase Diagnosis

The fault phase diagnosis tool has two internal modules namely, feature extraction and fault phase classification, and the detailed working of this tool is explained in detail in Chapter 3. This section focuses exclusively on the aspect of its application in the proposed PHM system.

1. **Feature Extraction:**

All the multi-sensor data from the experimental setup is collected for data cleansing and feature extraction at every time instant from the experimental setup. All the features extracted along with the cleaning processes are listed in detail in Table 3.1. However, only the features selected after the feature selection process will be used to form the feature vector. Therefore, the features with labels, 'F_Flow', 'F_Pressure', 'F_Vib_1', 'F_Vib_2', and 'F_Torque' are not extracted. A total of 12 features are extracted from the data of both the ODMs. The approximate mass lost (i.e., F_ODM2_02) is later used as the degradation index for the RUL prediction.

2. **Fault Phase Classification:**

Once the required features are extracted, a feature vector is formed and is used for classification with KNN and RF classifiers. The probabilities obtained from these classifiers are used by the weighted ensemble method to estimate the fault phase. If the estimated phase is normal operation (Phase 0), no action will be taken. When the gradual degradation phase (Phase 1) is estimated for the first time, then the RUL prediction process is activated and the degradation model is initialized.

6.1.2 RUL Prediction

RUL is predicted for every time instant after the bearing enters the gradual degradation phase. This sub-system involves two modules that relate to the degradation model and the degradation prediction using the enhanced adaptive PF.

1. **Degradation Model:**

As described in Chapter 5, the Gaussian degradation model used in this study requires two empirical values, M_T and σ_r (the phase transition mass and phase duration ratio respectively). The phase duration ratio (σ_r) is a constant obtained from the database of previous spall propagation tests. Whereas, the phase transition mass (M_T) will change with the fault phase.

a) Phase 1:

When the current fault phase is Phase 1 (gradual degradation), we do not know when the exact phase transition will occur in the future, hence an anticipated range for M_T, obtained from the previous tests, is used. For example, if we have 5 spall propagation tests with the range for phase transition mass between $17mg$ to $25mg$, then three degradation models will be generated. The first model will have $M_T = 17mg$ and this will give the lower bound of RUL prediction. The second model will have $M_T = 25mg$ and it will give the upper bound of the RUL. Whereas the third model will have a mean value, $M_T = 21mg$, and it will give the mean RUL prediction. Three of these models will be used individually and simultaneously for the RUL prediction process using the enhanced adaptive PF.

b) Phase 2:

When the current fault phase is Phase 2 (accelerated degradation), we now know the exact instant when the phase transition has occurred using the fault phase diagnosis sub-system. Therefore, at the time instant where Phase 2 is estimated for two consecutive times, the current phase transition mass M_T of the bearing is calculated. Where the actual value of the degradation index (F_ODM1_02) is used as the phase transition mass.

Once the value for phase transition mass M_T, is calculated, it is used to update the degradation model. For instance, if the phase transition mass obtained is $M_T = 20mg$, then the degradation models used earlier during Phase 1, will be discarded and the new model, with $M_T = 20mg$, will be used for the RUL prediction process using the enhanced adaptive PF.

2. Degradation Prediction:

The enhanced adaptive PF proposed in Chapter 4 is used in this process for model parameter estimation at every time instant. The enhanced adaptive PF will estimate the values of the three unknown model parameters β, σ_1 and μ_1. The method to assign the initial ranges of the model parameters for the enhanced adaptive PF is discussed in Section 4.4.3. At every time instant, the estimated values of the three model parameters are used for predicting the degradation evolution of the bearing. RUL is calculated as the total time difference from the current time instant to the time instant where the predicted degradation evolution reaches the pre-determined threshold.

The threshold, for the RUL prediction, physically means the maximum allowable cumulative mass that can be released from the bearing during its operation. This maximum allowable cumulative mass of the bearing, i.e., the threshold, should accurately represent the end of the bearing's life. In practice, the bearing life threshold is unique to each different bearing not only based on its geometry and material but also on its application. For example, the threshold for a bearing in a very critical location of a rotating machine will be lower compared to the threshold of a similar bearing at a non-critical location. In [119], Garvey suggests a few methodologies for setting the thresholds in PHM systems and they are: expert judgement based, industry-standard generic limits based, statistical alarm based, and trend-based limits. But a number of these methods require human intervention and a lot of failure data of similar bearings [10]. Therefore, determining an appropriate threshold is a subject of further research in the field and is out of the scope of this research. In this study, the total accumulated mass at the end of the respective bearing accelerated failure experiment is used as the pre-determined threshold.

a) Phase 1:

During the gradual degradation phase (Phase 1), the enhanced adaptive PF will work with three degradation models with will translate to the upper limit, lower limit, and mean of the predicted RUL. For example, if one degradation model has $M_T = 17mg$ and the other model has $M_T = 25mg$, both the models will be individually used for model parameter estimation and degradation prediction using the enhanced adaptive PF. Where one model

with $M_T = 17mg$ will give a lower bound for the predicted RUL and the other model with $M_T = 25mg$ will give the upper bound for the predicted RUL. Whereas the third model with mean, $M_T = 21mg$, will give the mean RUL prediction.

b) Phase 2:

During the accelerated degradation process (Phase 2), only one updated degradation model will be used therefore, we will have only one mean RUL prediction without the upper bound and the lower bound. Because of the model parameter smoothing step described in Section 4.3, the RUL prediction using the enhanced adaptive PF will not have high fluctuations.

6.2 Results and Discussion

As described in Section 1.4 (experimental setup), five bearing spall propagation experiments were conducted with SKF 7208 BEP bearing. The data from three experiments are used for training the fault phase diagnosis models and to obtain the values for phase transition mass (M_T) and the phase duration ratio (σ_r). The data from the rest of the two experiments are used for testing and validating the proposed framework. The results of the fault phase diagnosis sub-system are presented in Section 3.5, and in this Section, the performance of the multi-phase degradation model and the enhanced adaptive PF is tested for the RUL prediction.

Table 6.1 shows the values of phase transition mass (M_T) and the phase duration ratio (σ_r) for the experiments used for training. Therefore, the M_T for the lower limit model will be $M_T = 10.47\ mg$ and for the upper limit model will be $M_T = 76.60\ mg$ while the mean value $M_T = 43.53\ mg$ is used for the mean RUL prediction. Whereas the duration ratio varies between 0.3 to 0.51, thus the mean value of all the tests ($\sigma_r = 0.391$) is used in all the degradation models.

Table 6.1 Database of experiments used for training

Experiment Number	Transition Mass (M_T)	Duration Ratio (σ_r)
1	$10.47\ mg$	0.5102
2	$76.60\ mg$	0.3649
3	$68.29\ mg$	0.3007

6.2.1 RUL Prediction on Experiment 4

The fault phase diagnosis ensemble method estimates that Phase 1 (gradual degradation) begins at 132 minutes (see Fig 3.6), thus the RUL prediction process is activated at this time instant. The Gaussian degradation models are initialized with the upper, lower limit, and mean phase transition masses. The estimated unknown model parameters by the enhanced adaptive PF are used for predicting the degradation evolution. The degradation prediction at various time instants of the experiment with the upper and lower bounds are shown in Fig. 6.2.

The fault phase diagnosis tool estimates that beginning of Phase 2 around 805 minutes and thus the actual transition mass is calculated from the degradation index as $M_T = 32.6867\text{mg}$. Further, the degradation model is updated with the calculated transition mass and the previous upper bound and lower bound degradation models are discarded. Therefore, the predictions from 815 minutes to 1115 minutes in Fig 6.2 only have the mean prediction without the upper and lower limits.

As shown in Fig. 6.3, the RUL estimation plot of experiment 4 is very accurate. The predictions during Phase 1 are close to the 20% margin with the lower bound of the prediction inside the margin. While the RUL predictions in Phase 2, after updating the degradation model with the calculated phase transition mass, are very accurate and inside the 20% margin.

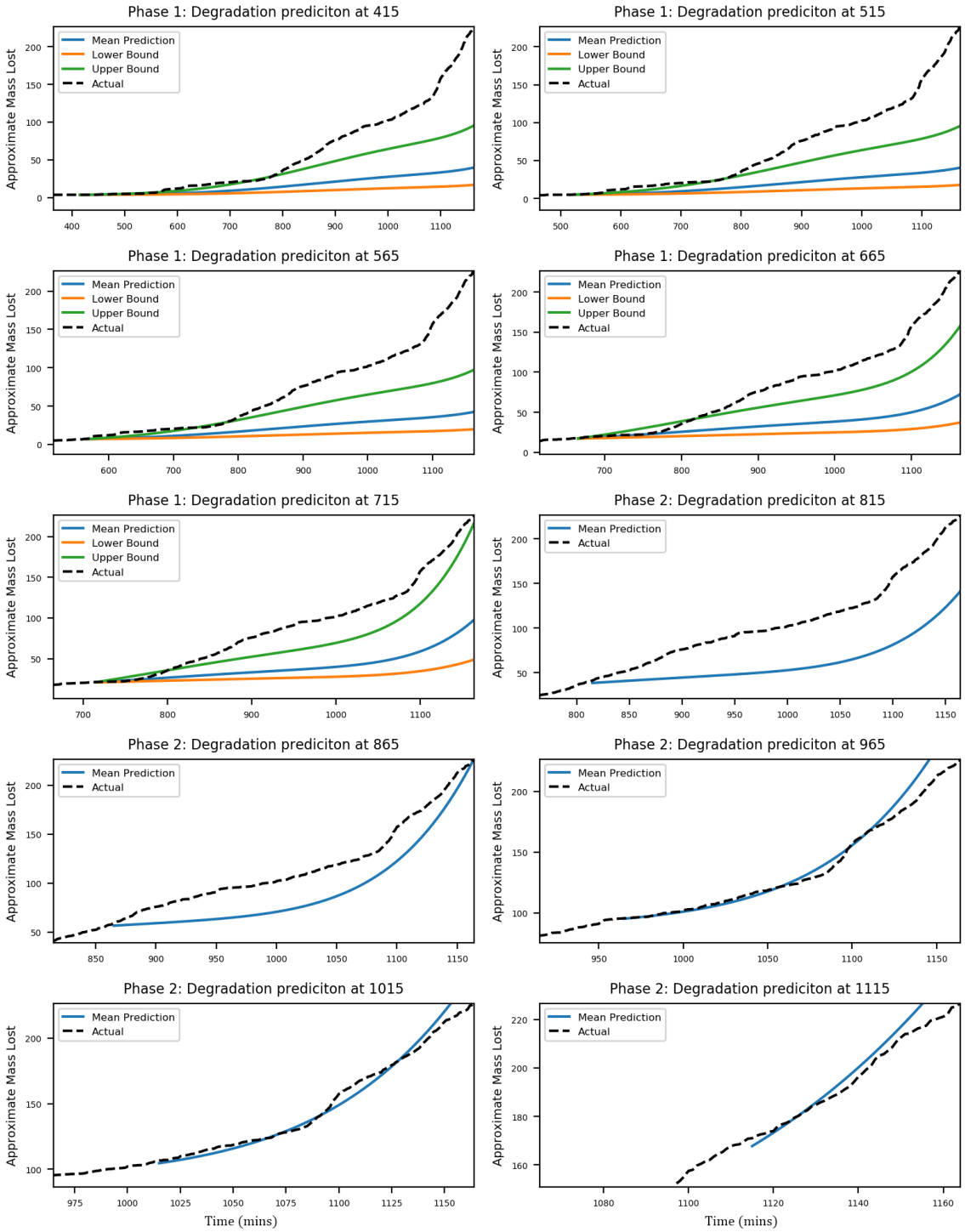

Fig. 6.2 Degradation predictions of experiment 4 at various time instants.

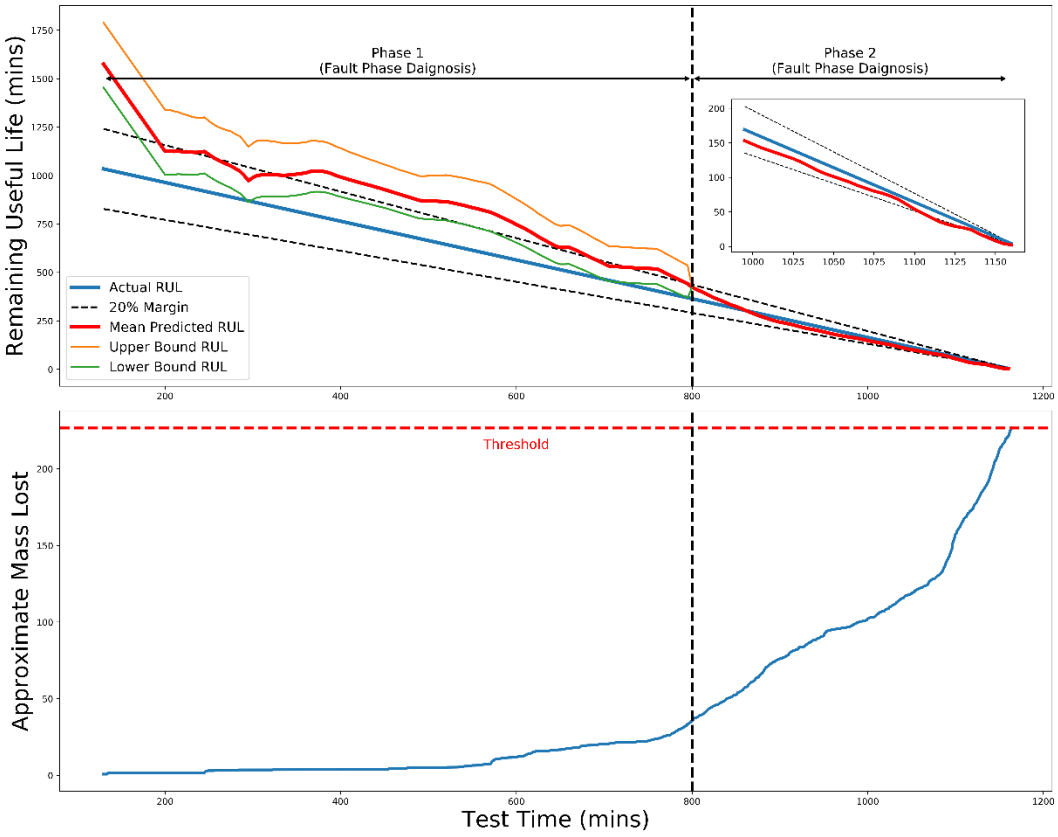

Fig. 6.3 RUL prediction plot for experiment 4.

6.2.2 RUL Prediction on Experiment 5

The fault phase diagnosis ensemble method estimates that Phase 1 begins at 85 minutes (see Fig 3.7), and therefore RUL prediction process is activated at that time instant. The Gaussian degradation models are initialized with the upper, lower limit, and mean phase transition masses as per Table 6.1. The estimated unknown model parameters by the enhanced adaptive PF are used for predicting the degradation evolution. The degradation prediction at various time instants of experiment 5 with the upper and lower bounds are shown in Fig. 6.4. The fault phase diagnosis estimates that Phase 2 begins around 415 minutes and thus the actual transition mass is calculated from the degradation index as $M_T = 18.5878\,\text{mg}$. As discussed earlier, the degradation model is updated with the calculated transition mass and the previous upper bound and lower bound degradation models are discarded.

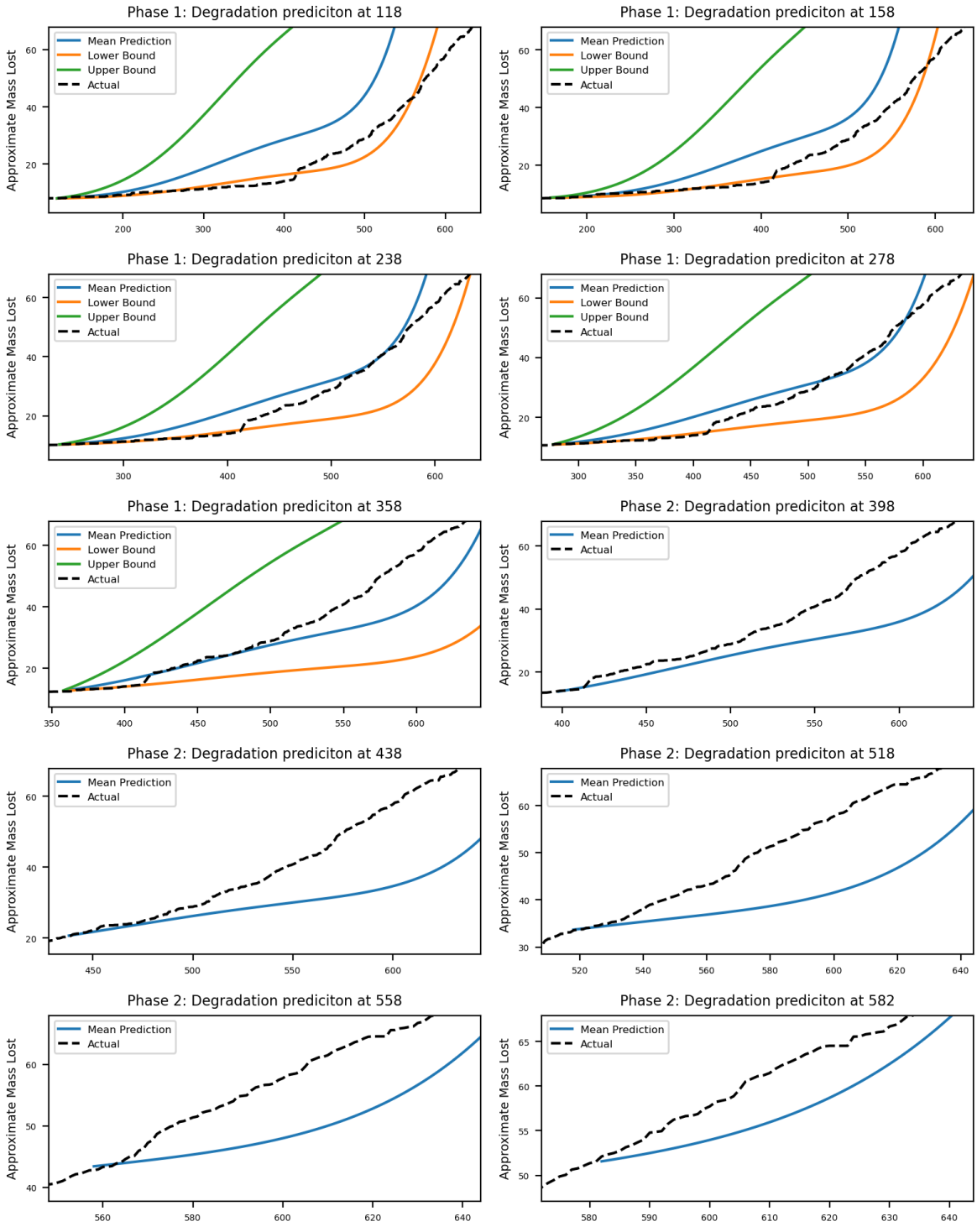

Fig. 6.4 Degradation predictions of experiment 5 at various time instants.

As shown in Fig. 6.5, the RUL estimation plot of experiment 5 is accurate during Phase 1 and the predictions are inside the 20% margin. During Phase 2 the predicted RUL is just over the 20% margin between 350 to 570 minutes of the experiment. But after 570 minutes the predictions become accurate. These inaccurate predictions between 350 to 570 minutes are due to the rapid burst of debris particles accumulated from the bearing around 570 minutes. Therefore, Phase 2 degradation predictions (shown in Fig. 6.4) follow the degradation trends which prevailed before the 570 minutes and they do not accurately forecast the future trend properly. However, soon after the sudden increase in the accumulated mass value at 570 mins the enhanced adaptive PF adapts well and gives accurate degradation predictions. The degradation prediction of 582 minutes (in Fig. 6.4) shows the gradually adapting degradation predictions.

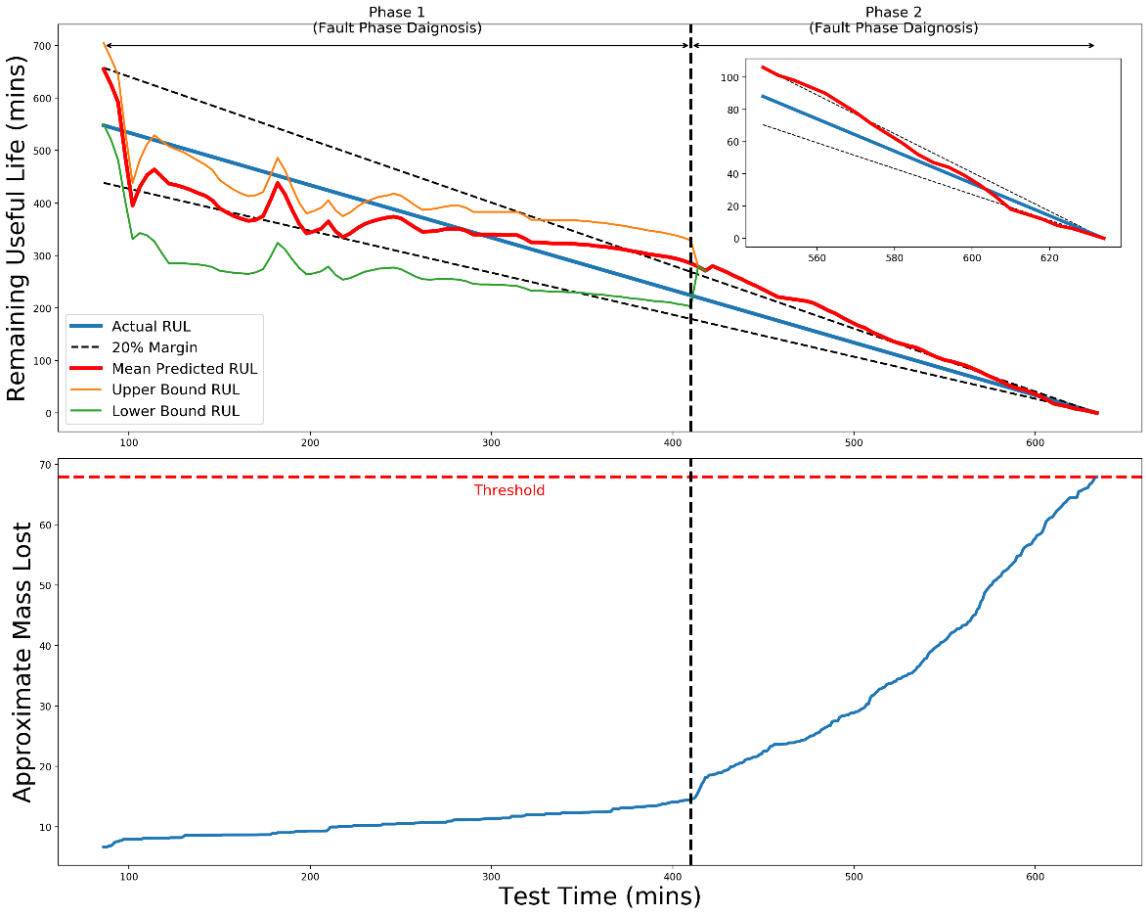

Fig. 6.5 RUL prediction plot for experiment 5.

6.3 Summary

The proposed bearing long-term RUL prediction PHM system is explained in detail and validated in this chapter. The proposed framework has two major sub-systems namely, fault phase diagnosis and RUL prediction, and both of these sub-systems have internal modules. The fault phase diagnosis runs at every time instant of the bearing operation and estimates the fault phase, i.e., Phase 0, Phase 1, or Phase 2. Whereas the RUL estimation process is activated as soon as Phase 1 is detected the first time. The degradation model initialized with the values from the database of previous experiments is used for multi-phase RUL prediction with the enhanced adaptive PF in Phases 1 and 2. The RUL prediction in Phase 1 gives an upper and lower bound while in Phase 2 only the mean RUL prediction is available. Further, the proposed methodology is validated on two spall propagation experiments conducted using the experimental setup presented in Section 1.4. The results show that the multi-phase and long-term RUL prediction results of the proposed PHM system are very accurate.

Chapter 7: Conclusions and Future Works

7.1 Conclusion

The bearing service life is typically divided into three phases: normal operation, gradual degradation, and accelerated degradation. Although extensive studies have been carried out on bearing RUL prediction, most of these studies focus only on the final phase i.e., the accelerated degradation phase. Whereas, not much work has been carried out associated with long-term and multi-phase RUL prediction. Even though many studies show that the ODMs have an excellent potential for bearing PHM systems, very few explore its potential for RUL prediction. Therefore, in this thesis, an intelligent PHM system is presented for rolling element bearings that mainly uses the information from the ODMs. The theme of this dissertation is to discuss the major issues in the existing PHM systems and subsequently present robust solutions for these issues.

Firstly, to leverage the relationship between the wear debris characteristics and the defect severity, three novel degradation indicators are developed based on ODM data. They are further used in the fault phase diagnosis tool as features. Two indicators are developed by monitoring the statistical relation of the debris particle size distribution. Whereas the third degradation indicator is based on an online change point detection technique. Next, a data-driven fault phase diagnosis technique based on an ensemble of KNN and RF classifiers is developed. Multi-sensor information including the degradation indicators previously developed is used in the feature vector. Three feature selection methods are used for selecting the most relevant features. Further, the classifiers are optimized and the leave-one-out cross-validation technique is used to calculate the F-scores which are used as the weights for each classifier. A weighted average ensemble method is finally used for the final phase estimation.

Secondly, to perform multi-phase RUL predictions, the model-based techniques need to be capable to adapt to sudden trend changes. However, the PFs used in the literature have three issues namely, particle degeneracy, particle impoverishment, and initial range dependence. Therefore, an enhanced adaptive PF algorithm is proposed to tackle these three problems. In this algorithm, a novel adaptive sampling process is introduced which

approximates the particle distribution as a normal distribution. In case of an adverse trend change in the degradation index, an extended adaptive sampling process is used that generates out-of-range particles and recursively converges to the required model parameter values. The superiority of the proposed technique is illustrated using simulated data and a noise sensitivity study. Although the enhanced adaptive PF is capable of adapting to trend changes, to get accurate long-term RUL prediction, the degradation model should also be enhanced. Many studies use a single degradation model to characterize the entire bearing service life, however, this is not suitable if the degradation process involves phase changes. Therefore, a multi-phase degradation model is developed for rolling element bearings using a mixture of the Gaussian process and Weibull function. The main contention of this model is that the debris generation rate of each degradation phase can be approximated by a three-parameter Gaussian. Then the physical, geometric, empirical relations are used to reduce the number of model parameters in the degradation model.

Finally, the proposed intelligent PHM system compiling the developed techniques is presented in detail. The proposed framework has two major sub-systems namely, fault phase diagnosis and RUL prediction. The phase estimation of the fault phase diagnosis is used to activate the RUL estimation process. In the accelerated degradation phase, the degradation model is updated with the calculated transition. Further, the performance of the proposed methodology is validated on two spall propagation experiments conducted using the experimental setup. The results verify the multi-phase and long-term RUL prediction PHM system.

7.2 Limitations and Future Works

As demonstrated with the results, the proposed PHM system works well for bearing systems with minor dynamic load changes, however, industrial applications of rotating machines operate under systematic and large load variations. Therefore, the PHM system needs to be validated for bearing systems with such major changes in the loading. Further, the developed multi-phase degradation model does not have the capability to incorporate

the effects of loading on the bearing. Therefore, this limitation shall be addressed in future works such that the degradation model will integrate the ability to utilize the current and future expected loads for RUL prediction.

Furthermore, the developed ensemble learning model for fault phase diagnosis inaccurately estimates the phases in certain scenarios. For example, in the fault phase estimation results of experiment 4 (see Fig. 3.6), the tool incorrectly estimates the phase around 1020 to 1050 minutes as Phase 1. In industrial applications, such inaccuracies will result in false alarms and may cause severe complications to the maintenance personnel. Such problems can be overcome by training the machine learning models with more and better-quality failure data. Additionally, future work will be conducted in developing a rule-based fault phase diagnosis tool that can avoid such inaccuracies.

In addition, further research work shall be conducted in the following aspects:

1. The proposed PHM system will be applied to other mechanical equipment with multi-phase degrading such as wind-turbine gearboxes.
2. More research work will be conducted related to the noise sensitivity of the enhanced adaptive PF technique.
3. The viability of the proposed PHM system will be studied in scenarios where the machines might have both gears and bearings in the same lubrication system.
4. Deep learning classification models will be employed for future fault diagnosis and RUL estimation studies.

Appendices

Appendix A Regression-Based Online Change Point Detection

A.1 Online Change Point Detection Techniques

Bearing degradation is a stochastic process and it involves an elevated level of uncertainty. Typically, these uncertainties manifest in sudden trend changes in the degradation index. The time instant when these sudden trend changes occur is known as change points in the data. These change points correspond to a significant increase in the amplitude of the degradation index in comparison to previous reference values. Identifying the exact time when sudden trend changes appear in the bearing degradation index will help to disclose important information about the actual degradation of the bearing. The "online" change detection, refers to the processing and identifying the trend changes at each measurement point (i.e., the data is processed and the change point is detected immediately instead of observing trends for a longer period).

Several supervised and unsupervised data-driven techniques have been designed, enhanced, and adapted for online change point detection techniques in the literature. The supervised techniques using Naïve Bayes, Support Vector Machines, Hidden Markov model, etc., have been proposed for online change point detection [121, 122]. Alternatively, unsupervised methods such as likelihood ratio, kernel-based methods, clustering, and probabilistic methods have been proposed [123, 124, 125, 103, 126]. One of the most used online change point detection methods is Bayesian Online Change Point Detection (BOCD) proposed in [103]. Compared to other methods which only consider pairs of consecutive samples, BOCD uses a sliding window method where the change point is estimated based on all previous intervals within the sliding window [103, 127].

This section proposes a novel methodology for online change point detection in degradation index or any time-series data. To verify this methodology, the working of this proposed technique is compared with the most famously available BOCD technique.

A.2 Bayesian Online Change Point Detection (BOCD)

BOCD is a probabilistic method with the assumption that a sequence of observations may be divided into non-overlapping states and the data within each state ρ in time series are independent and follow some probability distribution $P(x_t|\eta_\rho)$. It estimates the posterior distribution by defining an auxiliary variable run-length (r_t) which represents the time elapsed since the last change point. Given the run length at a time instant *t*, the run length at the next time point can either reset back to 0 (if a change point occurs at this time) or increase by 1 (if the current state continues for one more time unit). The run-length distribution based on Bayes' theorem can be denoted as:

$$P(r_t|x_{1:t}) = \frac{\sum_{r_{t-1}} P(r_t|r_{t-1})P\left(x_t|r_{t-1}, x_t^{(r)}\right)P(r_{t-1}|x_{1:t-1})}{\sum_{r_t} P(r_t|x_{1:t})} \qquad \text{(A.1)}$$

Where $x_t^{(r)}$ indicates the set of observations in the sliding window, and $P(r_t|r_{t-1})$, $P\left(x_t|r_{t-1}, x_t^{(r)}\right)$, $P(r_{t-1}|x_{1:t-1})$ are prior, likelihood, and recursive components of the equation. The conditional prior is non-zero at only two outcomes (r_t = 0 or $r_t = r_{t-1} + 1$) and simplifies the equation.

$$P(r_t|r_{t-1}) = \begin{cases} H(r_{t-1}+1) & if\ r_t = r_{t-1}+1 \\ 1 - H(r_{t-1}+1) & if\ r_t = r_{t-1}+1 \\ 0 & otherwise \end{cases} \qquad \text{(A.2)}$$

In this equation, $H(x) = \frac{P(x)}{\sum_{t=r}^{\infty} P(x)}$ is a hazard function which is defined as the ratio of probability density over the run to the total value of probability densities. The likelihood term represents the probability that the most recent observation point belongs to the current run. After calculating the run length distribution and updating the corresponding statistics, change point prediction is performed by comparing probability values. If r_t has the highest probability in the distribution, then a change point has occurred and the run length is reset to r_t = 0. If not, the run length is incremented by one ($r_t = r_{t-1}$ + 1) [103, 127]. Implementing BOCD on online data requires us to define the hazard and the function for

data distribution. BOCD method is very efficient for finding the online change points if the magnitude variation between the points is high; however, the efficiency of BOCD is not good when there are small magnitude changes.

A.3 Regression-based Online Change Point Detection (ROCD)

As discussed earlier, likelihood ratio, kernel, clustering, and probabilistic-based methods have been proposed for online change point detection so far in the literature. However, regression-based methods have not been explored yet.

Linear regression is a process in which a first-order equation ($Y = \alpha_1 X + \alpha_2$) is fitted to a set of data points to describe and generalize the relationship between the explanatory variable (X) and the observed variable (Y). We can use the least square approach or maximum likelihood estimation approach to find the most optimum fit called the regression line. It gives us the values for both coefficients α_1 and α_2. However, since the inevitability of the noise in physically observed variables, all the observed data points will have some distance from the linear regression line. This distance from the regression line is called residuals and it can be quantified using the Mean Square Error (MSE) metric. Fig. A.1 (a) shows a linear regression line fitted to a set of evolving observed data points and the MSE of the highlighted residuals. Consider that a trend change appears in the observed variable, the magnitude of the most recent observed data points will increase or decrease. This instant of the explanatory variable (X) is known as the change point. When this change point occurs, as shown in Fig. A.1 (b), the residuals for the last few data points will increase, and as a result, the MSE will increase. Let us define the mean residual of the last few data points during the change point as "*change point residual*" for further explanation.

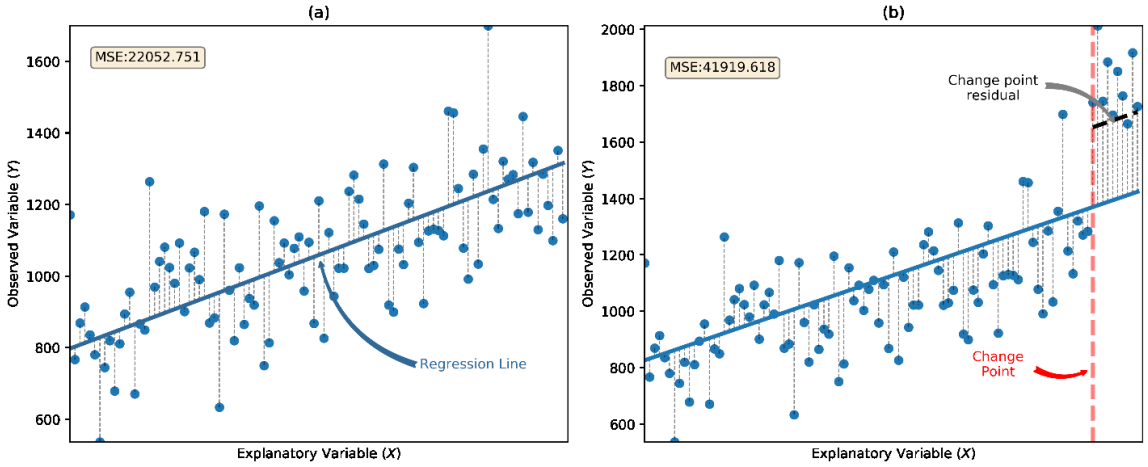

Fig. A.1 (a) Linear regression line fitted to the data and MSE of the residuals, (b) Residuals and MSE increases as when a change point occurs.

Quadratic regression is a process of fitting a second-order equation ($Y = \beta_1 X^2 + \beta_2 X + \beta_3$) to the set of observed data points. This generalizes the relation between the explanatory variable (X) and the observed variable (Y) by fitting a most optimum parabolic curve known as the quadratic regression curve. Similar to linear regression, the least-square approach or maximum likelihood estimation approach can be used to fit the regression curve and obtain the values for the coefficients β_1, β_2 and β_3. Fig. A.2 (a-c) shows the quadratic regression curve to a set of data points with and without change points. As long as the data points do not have a change point, the regression curve will not have any curvature, i.e., the regression curve will be similar to a straight line. Further, the second-order coefficient (β_1) will be close to zero ($\beta_1 \rightarrow 0$) for quadratic regression curves that do not have any curvature. When there are change points in the observed variable, the value of the is coefficient will increase for a positive curvature (Fig. A.2 (a)), and similarly, the coefficient will decrease for a negative curvature (Fig. A.2 (c)).

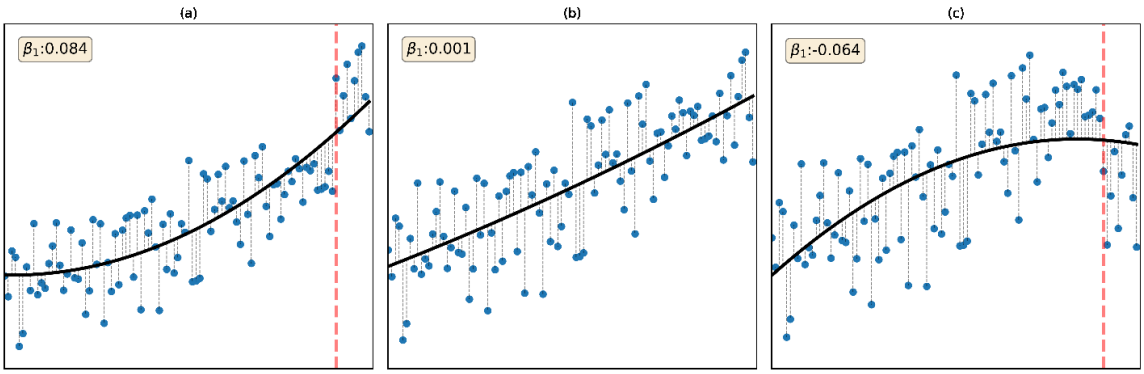

Fig. A.2 (a) Quadratic regression curve with positive curvature when change point occurs, (b) Quadratic regression curve with very low curvature, (c) Quadratic regression curve with negative curvature when change point occurs.

Although the MSE is a good metric to identify the change points in data, in practical applications it is susceptible to cause false triggers because of the noise in the observed data. Fig A.3 (a) shows the effect of varying the *change point residual* as a ratio of the noise level in the observed variable on the MSE metric. As shown noise in the data adversely affects the MSE metric. We get high MSE values even when the change point residual is not high (i. e., between -1 to 1 region). Fig. A.3 (b) shows the corresponding variation of β_1 value as the *change point* residual varies linearly. As we can see the β_1 (curvature coefficient) has less variation in the region within -1 to 1 region because the curvature in of the quadratic regression curve is less when there is low noise. Therefore, to reduce the influence of the noise, we get the product of the absolute value of the second-order coefficient (β_1) and the MSE metric and it will be called the "*error product*" for further discussions. The corresponding *error product* value for varying change point residual is shown in Fig. A.3 (c). As shown, the value of the *error product* is close to zero between the -1 to 1 region. This makes sure that we identify only the relevant change points and reduce the risk of false alarms.

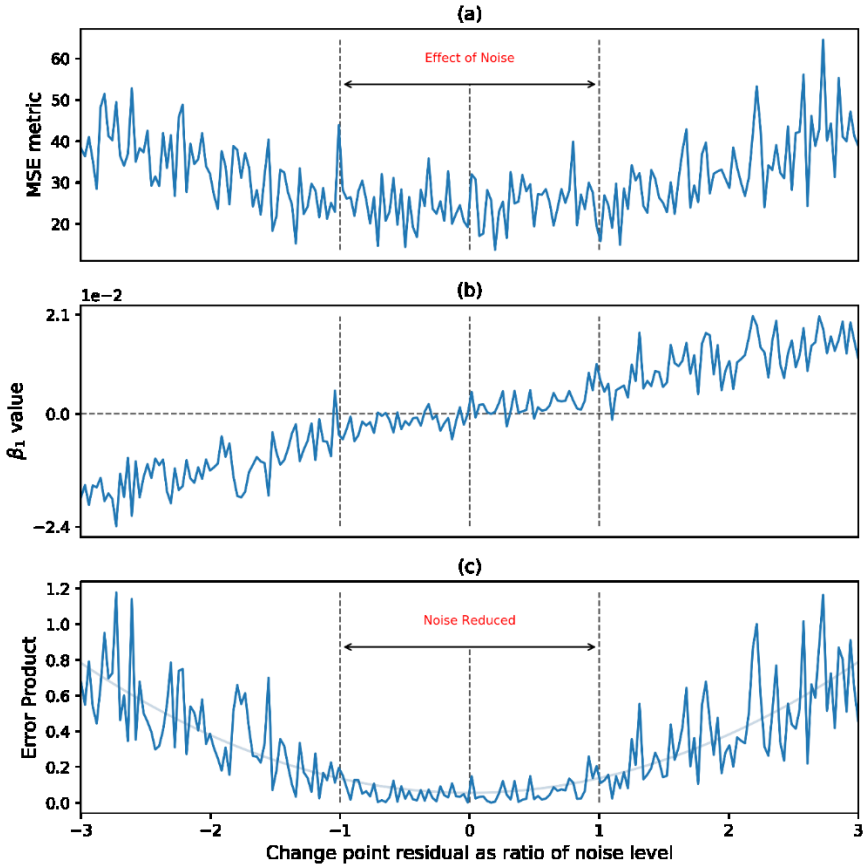

Fig. A.3 Effect of variation in *change point residual* in the observed variable (Y) (a) Variation of the MSE (b) Variation of β_1(curvature coeffiecient) (c) Variation of *error product.*

A.4 ROCD Methodology

The proposed ROCD method uses this *error product* to identify the change points in the observed degradation index. In such a case, the degradation index will be the observed response variable (Y) and the corresponding time instants will be the explanatory variable (X). Further, we will use a sliding window to consider only the recent observed data points for the online change detection. The proposed ROCD method is explained in the below steps:

1) Select the number of recent data points to be considered in the sliding window.

2) Using the LSE approach, find the linear regression line and the quadratic regression curve for the data points in the sliding window.
3) Find the residuals of the observed data points from the linear regression line and calculate the MSE of the residuals.
4) Calculate the second-order coefficient (β_1) from the quadratic regression.
5) Calculate the *error product* by multiplying the MSE and the absolute value of β_1.

The proposed ROCD method is practical for many types of applications as long as the sliding window can meaningfully fit the linear regression line. In many applications, especially in bearing fault diagnosis, even though the overall trend of degradation is exponential, the data points in a reasonable sliding window are appropriate to be used for linear regression.

A.5 Validation on Simulated Signal

In this section, the proposed ROCD method is compared with the widely accepted BOCD method for online change point detection. Consider the simulated signal, shown in Fig. A.4, with two change points with magnitude disparity at 30 minutes and 90 minutes respectively.

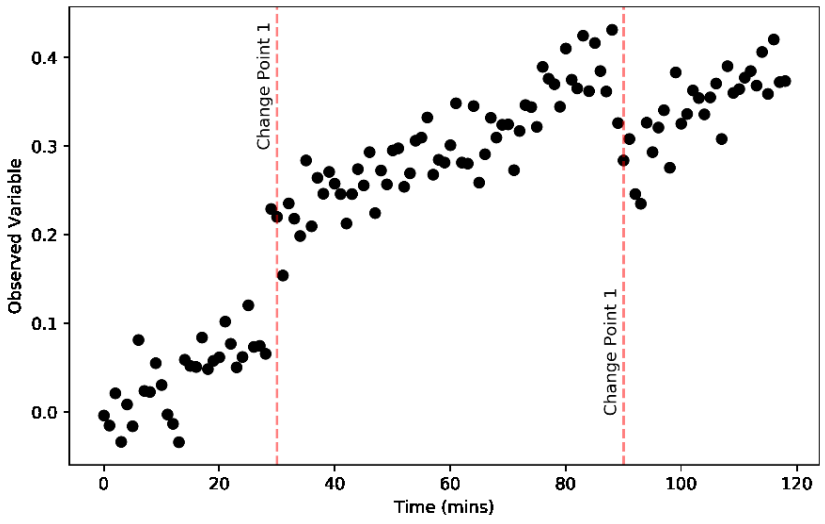

Fig. A.4 Simulated signal to validate the ROCD method.

Fig. A.5. shows results of the BOCD method applied on the simulated signal with a low hazard value and a standard Student T distribution for the observed variable. As we can see, that the change point probability is high at 30 minutes and hence, the run length becomes zero at 30 minutes. However, change point 2 at 90 minutes is not identified at all. Whereas the ROCD method applied with a 5 minutes sliding window comfortably identifies both the change points. As shown in Fig. A.6, the *error product* value increases during both the change points at 30 minutes and 90 minutes, while the *error product* for the rest of the locations is low.

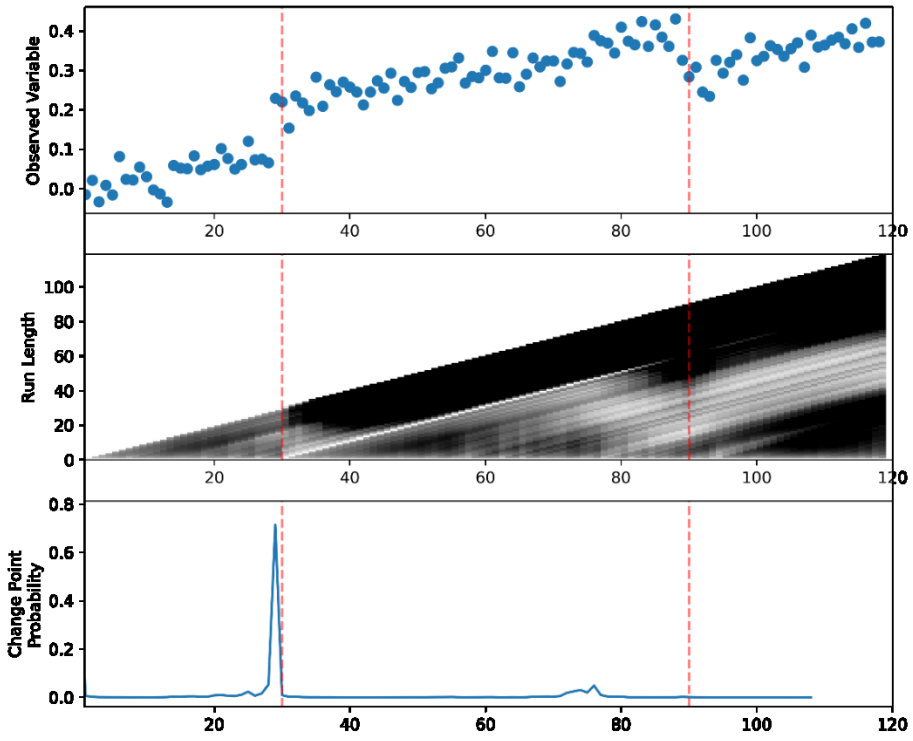

Fig. A.5 BOCD method identifies only one change point.

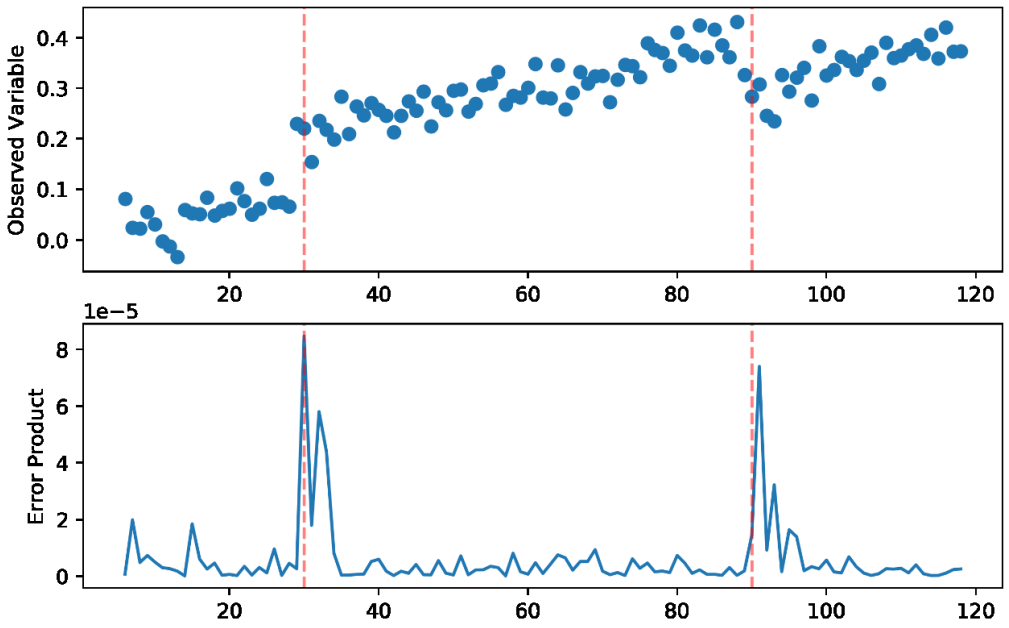

Fig. A.6 ROCD method with 5 minutes sliding window identifies both the change points.

A.6 Conclusion

Several online change point detection methods have been proposed so far in the literature. However, these methods need a few hyper-parameters to be predetermined and these parameters mostly depend on the characteristics of the observed data. Moreover, in many practical cases, it is impossible to know the exact values of hyper-parameters that give the best results for the observed data. Most of the online change point detection techniques are based on likelihood, kernels, clustering, and probabilistic methods, but regression-based methods have not been explored yet. Therefore, a regression-based change point detection technique is developed which only requires one hyper-parameter. This technique uses both linear and quadratic regression to calculate an *error product.* The *error product* is a product of the MSE of the residuals from the linear regression line and the second-order coefficient of the quadratic regression curve. The value of the *error product* increases a representing the identification of a change point. The superiority of the ROCD method over the BOCD method is verified using a simulated signal with two change points.

References

[1] Nam-Ho Kim, Dawn An, Joo-Ho Choi. Prognostics and Health Management. Prognostics and Health Management of Engineering Systems: An Introduction. Springer, 2016. ISBN 978-3-319-44740-7.

[2] Yaguo Lei, Naipeng Li, Szymon Gontarz, Jing Lin, Stanislaw Radkowski, and Jacek Dybala. A Model-Based Method for Remaining Useful Life Prediction of Machinery. IEEE Transactions On Reliability 65.3 (2016).

[3] Karen Cassidy. Qualifying an On-Line Diagnostic and Prognostic Sensor for Fixed and Rotary Wing Bearings and Gears. IEEE Aerospace Conference. Big Sky, Montana, 2008.

[4] Liu, Jie. An intelligent system for bearing condition monitoring. Waterloo: University of Waterloo, 2008.

[5] Peng Wang, Ruqiang Yan, and Robert X. Gao. Multi-Mode Particle Filter for Bearing Remaining Life Prediction. Proceedings of the ASME 2018 13th International Manufacturing Science and Engineering Conference. College Station, TX, USA, 2018.

[6] Yizhen Peng, Yu Wang , Member, IEEE, and Yanyang Zi. Switching State-Space Degradation Model With Recursive Filter/Smoother for Prognostics of Remaining Useful Life. IEEE Transactions On Industrial Informatics 15.2 (2019).

[7] Li, Naipeng and Lei, Yaguo and Lin, Jing and Ding, Steven X. An Improved Exponential Model for Predicting Remaining Useful Life of Rolling Element Bearings. IEEE Transactions on Industrial Electronics 62.12 (2015): 7762-7773.

[8] Adrian Cubillo, Suresh Perinpanayagam, Manuel Esperon-Miguez. A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. Advances in Mechanical Engineering 8.8 (2016).

[9] Roylance BJ, Williams JA, Dwyer-Joyce R. Wear debris and associated wear phenomena—fundamental research and practice. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 2000; 214(1):79-105.

[10] Wei Cao, Guangneng Dong, You-Bai Xie, Zhongxiao Peng. Prediction of wear trend of engines via on-line wear debris monitoring. Tribology International (2018): doi: 10.1016/j.triboint.2018.01.015.

[11] Jiayi Sun, Liming Wang, Jianfeng Li, Fangyi Li, Jianyong Li, Haiyang Lu. Online oil debris monitoring of rotating machinery: A detailed review of more than three decades. Mechanical Systems and Signal Processing 149 (2021): 107341, https://doi.org/10.1016/j.ymssp.2020.107341.

[12] Becker, Andrew James. A Methodology for Determining Rolling Element Bearing Limits Applied to Inductive Wear Debris Sensors. 2014

[13] Hong, Wei & Wang, Shao-Ping & Tomovic, Mileta & Liu, Haokuo & Shi, Jian & Wang, Xingjian. A Novel Indicator for Mechanical Failure and Life Prediction Based on Debris Monitoring. IEEE Transactions on Reliability (2016; DOI:10.1109/TR.2016.2628412): 1-9.

[14] Akhand Rai, S.H. Upadhyay,. A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International 96 (2016): 289-306; https://doi.org/10.1016/j.triboint.2015.12.037.

[15] Duy-Tang Hoang, Hee-Jun Kang. A survey on Deep Learning based bearing fault diagnosis. Neurocomputing 335 (2019): 327-335; https://doi.org/10.1016/j.neucom.2018.06.078.

[16] Lin, Liang Guo and Naipeng Li and Feng Jia and Yaguo Lei and Jing. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 240 (2017): 98-107.

[17] D. Wang, K. Tsui and Q. Miao. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. IEEE Access 6 (20118): 665-676; doi: 10.1109/ACCESS.2017.2774261.

[18] Singh, Jaskaran Singh and A K Darpe and S P. Bearing remaining useful life estimation using an adaptive data-driven model based on health state change point identification and K-means clustering. Measurement Science and Technology 31.8 (2020): 085601, doi: 10.1088/1361-6501/ab6671.

[19] Zhang, Dingcheng, et al. Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network . Measurement 156 (2020): 10.1016/j.measurement.2020.107585.

[20] Dempsey, Paula. A Comparison of Vibration and Oil Debris Gear Damage Detection Methods Applied to Pitting Damage. . 2000

[21] Wei Hong, Wenjian Cai, Shaoping Wang, Mileta M. Tomovic. Mechanical wear debris feature, detection, and diagnosis: A review. Chinese Journal of Aeronautics 31.5 (2018): 867-882.

[22] Dempsey, Paula J. Lewicki, David G. Decker, Harry J. Investigation of Gear and Bearing Fatigue Damage Using Debris Particle Distributions. Cleveland, OH: Glenn Research Center, National Aeronautics and Space Administration, 2004.

[23] PJ Dempsey, G Kreider, T Fichter. Investigation of Tapered Roller Bearing Damage Detection using Oil Debris Analysis. IEEE Aerospace Conference. 2006; doi: 10.1109/AERO.2006.1656082.

[24] Dempsey, P., Bolander, N., Haynes, C., & Toms, A. Investigation of Bearing Fatigue Damage Life Prediction Using Oil Debris Monitoring. 2011

[25] Wei Hong, Shaoping Wang, Mileta M. Tomovic, Haokuo Liu, Jian Shi and Xingjian Wang. A Novel Indicator for Mechanical Failure and Life Prediction Based on Debris Monitoring. IEEE Transactions on Reliability 99 (2016): 1-9.

[26] Aditya Sharma, M Amarnath, PK Kankar. Feature extraction and fault severity classification in ball bearings. Journal of Vibration and Control 22 (1).doi:10.1177/1077546314528021 (2016): 176-192.

[27] Cerrada, M., Sánchez, R.-V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., & Vásquez, R. E. A review on data-driven fault severity assessment in rolling bearings. Mechanical Systems and Signal Processing 99 (2018): 169-196 DOI:10.1016/j.ymssp.2017.06.012.

[28] Idriss El-Thalji, Erkki Jantunen. A summary of fault modelling and predictive health monitoring of rolling element bearings. Mechanical Systems and Signal Processing 60 (2015): 252-272, https://doi.org/10.1016/j.ymssp.2015.02.008.

[29] Safizadeh, M. S., Latifi, S. K. Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. Information Fusion 18 (2013): 1-8, doi:10.1016/j.inffus.2013.10.002 .

[30] de Oliveira, Mario, Andre Monteiro and Jozue Vieira Filho. A New Structural Health Monitoring Strategy Based on PZT Sensors and Convolutional Neural Network. Sensors 18.9 (2018).

[31] Naipeng Li, Yaguo Lei, Jing Lin and Steven Ding. An Improved Exponential Model for Predicting Remaining Useful Life of Rolling Element Bearings. IEEE Transactions on Industrial Electronics 62.12 (2015).

[32] Aditya Sharma, M. Amarnath, Pavan Kumar Kankar. Novel ensemble techniques for classification of rolling element bearing faults. The Journal of Brazilian Society of Mechanical Sciences and Engineering 39.DOI 10.1007/s40430-016-0540-8 (2016): 709-724.

[33] Ginart, Antonio and Barlas, Irtaza and Goldin, Jonathan and Dorrity, Jordan Lewis. Automated Feature Selection for Embeddable Prognostic and Health Monitoring (PHM) Architectures. IEEE Autotestcon. pp. 195-201, doi: 10.1109/AUTEST.2006.283625, 2006.

[34] Tran, Van Tung and Bo-Suk Yang. An intelligent condition-based maintenance platform for rotating machinery. Expert Systems with Applications 39.3, https://doi.org/10.1016/j.eswa.2011.08.159 (2012): 2977-2988.

[35] Huang, Cheng-Geng, et al. A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing. Journal of Manufacturing Systems url: https://doi.org/10.1016/j.jmsy.2021.03.012 (2021).

[36] Xiang Li, Wei Zhang, Qian Ding. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering & System Safety 182.doi: 10.1016/j.ress.2018.11.011 (2019): 208--218.

[37] Ben Ali, Jaouher, et al. Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally

progressing degradations. Engineering Applications of Artificial Intelligence 42 (2015): 67-81.

[38] Sharma, A., M. Amarnath and P. Kankar. Feature extraction and fault severity classification in ball bearings. Journal of Vibration and Control 22.1 (2014): 176-192, DOI: 10.1177/1077546314528021.

[39] Ren´e-Vinicio S´anchez, Pablo Lucero, Rafael E. V´asquez, Mariela Cerrada. Feature ranking for multi-fault diagnosis of rotating machinery by using random forest and KNN. Journal of Intelligent & Fuzzy Systems 34.DOI:10.3233/JIFS-169526 (2018): 3463-3473.

[40] Vakharia, V., V. K. Gupta and P. K. Kankar. A comparison of feature ranking techniques for fault diagnosis of ball bearing. Soft Computing - A Fusion of Foundations, Methodologies and Applications 20.4 (2016): DOI: 10.1007/s00500-015-1608-6.

[41] Xiaoming Xue, Chaoshun Li, Suqun Cao, Jinchao Sun, Liyan Liu. Fault Diagnosis of Rolling Element Bearings with a Two-Step Scheme Based on Permutation Entropy and Random Forests. Entropy (Basel) 21.1 (2019).

[42] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun and J. He. Fault Diagnosis of a Rolling Bearing Using Wavelet Packet Denoising and Random Forests. IEEE Sensors Journal 17.17 (2017): 5581-5588.

[43] V. Vinay, G.V. Kumar and K.P. Kumar. Application of chi square feature ranking technique and random forest classifier for fault classification of bearing faults. Proceedings of the 22th International Congress on Sound and Vibration. Florence, Italy, 2015.

[44] Jiantao Lu, Weiwei Qian , Shunming Li 1, Rongqing Cui. Enhanced K-nearest neighbor for intelligent fault diagnosis of rotating machinery. Applied Sciences 11.919 (2021).

[45] Shaojiang Dong, Xiangyang Xu, Renxiang Chen. Application of fuzzy C-means method and classification model of optimized K-nearest neighbor for fault diagnosis of bearing. Journal of Brazilian Society of Mechanical Sciences and Engineering 38.8 (2015).

[46] Villa-Vialaneix, Robin Genuer and Jean-Michel Poggi and Christine Tuleau-Malot and Nathalie. Random Forests for Big Data. Big Data Research 9.doi = https://doi.org/10.1016/j.bdr.2017.07.003 (2017): 28-46.

[47] Gaowei Xu, Min Liu, Zhuofu Jiang, Dirk Söffker, Weiming Shen. Bearing Fault Diagnosis Method Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning. Sensors (Basel) 19.5 (2019): 1088.

[48] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, Jibao Gu. Sentiment classification: The contribution of ensemble learning. Decision Support Systems 57 (2014): 77-93.

[49] Zhao-Qiang Wang, Chang-Hua Hu, and Hong-Dong Fan. Real-Time Remaining Useful Life Prediction for a Nonlinear Degrading System in Service: Application to Bearing Data. IEEE/ASME Transactions on Mechatronics (2017).

[50] Cheng Cheng, Guijun Ma, Yong Zhang, Mingyang Sun, Fei Teng, Han Ding, and Ye Yuan. Online Bearing Remaining Useful Life Prediction Based on a Novel Degradation Indicator and Convolutional Neural Networks. Available: https://arxiv.org/abs/1812.03315 (2018).

[51] An, Dawn, Joo-Ho Choi, and Nam Ho Kim. Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. Reliability Engineering & System Safety 115 (2013): 161-169.

[52] Liu, Jie., Wilson Wang, F. Ma, Y. B. Yang, and C. S. Yang. A data-model-fusion prognostic framework for dynamic system state forecasting. Engineering Applications of Artificial Intelligence 25.4 (2012): 814-823.

[53] Zhao, Fuqiong, Zhigang Tian and Eric Bechhoefer and Yong Zeng. An Integrated Prognostics Method Under Time-Varying Operating Conditions. IEEE Transactions on Reliability 64.2 (2015): 673-686.

[54] Ansi Zhang, Shaobo Li1, Yuxin Cui, Rongzhi Dong, and Jianjun Hu. Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning. IEEE Access (2019).

[55] Erdogan, P. C. Paris and F. A critical analysis of crack propagation laws. Fluids Eng 84.4 (1963): 528-533.

[56] Forman, R. G., V. E. Kearney and R. M. Engle. Numerical Analysis of Crack Propagation in Cyclic-Loaded Structures. Journal of Basic Engineering 89.3 (1967): 459.

[57] Choi, Youngsik and C. Richard Liu. Spall progression life model for rolling contact verified by finish hard machined surfaces. Wear 262.1-2 (2007).

[58] Kaufman, J. W. Sheppard and M. A. A Bayesian approach to diagnosis and prognosis using built-in test. IEEE Transactions on Instruments Measurements 54.3 (2005).

[59] Ting Dong, Dawn An and Nam H. Kim. Prognostics 102: Efficient Bayesian-Based Prognostics Algorithm in MATLAB. Open access peer-reviewed chapter (2019).

[60] Yuning Qian, Ruqiang Yan and Shijie Hu. Bearing Degradation Evaluation Using Recurrence Quantification Analysis and Kalman Filter. IEEE Transactions on Instrumentation and Measurement 63 (2014).

[61] Jinjiang Wang, Peng Wang and Robert X. Gao. Enhanced particle filter for tool wear prediction. Journal of Manufacturing Systems 36 (2015): 35-45.

[62] Han Wang, Xiaobing Ma and Yu Zhao. An improved Wiener process model with adaptive drift and diffusion for online remaining useful life prediction. Mechanical Systems and Signal Processing 127 (2019): 370-387.

[63] Jaouher Ben Ali, Brigitte Chebel-Morello, Lotfi Saidi, , Simon Malinowski and Farhat Fnaiech. Accurate Bearing Remaining Useful Life Prediction Based On Weibull Distribution And Artificial Neural Network. Mechanical Systems and Signal Processing 56 (2015): 150-172.

[64] Dorrity, Antonio Ginart and Irtaza Barlas and Jonathan Goldin and Jordan. Automated Feature Selection for Embeddable Prognostic and Health Monitoring ({PHM}) Architectures. IEEE Autotestcon, (2006).

[65] Soto, Alvaro. Self Adaptive Particle Filter. IJCAI (2005): 1398-1406.

[66] Zhou, Tian, et al. Adaptive particle filter based on Kullback–Leibler distance for underwater terrain aided navigation with multi-beam sonar. The Institution of Engineering and Technology 12.4 (2018).

[67] Rigatos, Gerasimos G. Particle Filtering for State Estimation in Nonlinear Industrial Systems. IEEE Transactions on Instrumentation and Measurement 58.11 (2009).

[68] Arulampalam, M.S., et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50.2 (2002).

[69] Gordon, N.J., D.J. Salmond and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation}. The Institution of Electrical Engineers 140.2 (1993): 107.

[70] Carpenter, J., P. Clifford, and P. Fearnhead. An Improved Particle Filter for Non-linear Problems. IEEE Proceedings for Radar and Sonar Navigtion (1999).

[71] Kitagawa, Genshiro. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistic 5.1 (1996): 1-25.

[72] Higuchi, Tomoyuki. Monte carlo filter using the genetic algorithm operators. Journal of Statistical Computation and Simulation 59.1 (1997): 1-23.

[73] Plaza Guingla, Douglas A., et al. Improving particle filters in rainfall-runoff models: Application of the resample-move step and the ensemble Gaussian particle filter. Water Resources Research 49.7 (2013): 4005--4021.

[74] Qian, Yuning, and Ruqiang Yan. Remaining useful life prediction of rolling bearings using an enhanced particle filter. IEEE Transactions on Instrumentation and Measurement 64.10 (2015): 2696-2707.

[75] Djuric, J. H. Kotecha and P. M. Gaussian particle filtering. IEEE Transactions on Signal Processing 51.10 (2003).

[76] —. Gaussian sum particle filtering. IEEE Transactions on Signal Processing 51.10 (2003).

[77] Frank Hutter, Richard Dearden. The Gaussian Particle Filter for Diagnosis of Non-Linear Systems. IFAC Proceedings Volumes 36.5 (2003): 909-914.

[78] Fox, Dieter. KLD-Sampling: Adaptive Particle Filters. Advances in Neural Information Processing Systems . Vancouver, British Columbia, 2001.

[79] David Lesage, Elsa D. Angelini, Gareth Funka-Lea, Isabelle Bloch. Adaptive particle filtering for coronary artery segmentation from 3D CT angiograms. Computer Vision and Image Understanding 151 (2016): 29-46.

[80] Ahwiadi, Mohamed and Wilson Wang. An Adaptive Particle Filter Technique for System State Estimation and Prognosis . IEEE Transactions on Instrumentation and Measurement 69.9 (2020).

[81] Hu, Yaogang, et al. A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process. Renewable Energy 127 (2018): 452-460.

[82] Ye, Zhi-Sheng & Chen, Nan. The Inverse Gaussian Process as a Degradation Model. Technometrics 56 (2014): Doi: 10.1080/00401706.2013.830074. .

[83] P Boškoski, Đ Juričić. Inverse Gaussian mixtures models of bearing vibrations under local faults. Mechanical Systems and Signal Processing, 66-67 (2016; https://doi.org/10.1016/j.ymssp.2015.05.010): 546-556.

[84] Elwany, Alaa and Nagi Gebraeel. Real-Time Estimation of Mean Remaining Life Using Sensor-Based Degradation Models. Journal of Manufacturing Science and Engineering 131.5 (2009): DOI: 10.1115/1.3159045.

[85] H Wang, H Liao, X Ma, R Bao. Remaining Useful Life Prediction and Optimal Maintenance Time Determination for a Single Unit Using Isotonic Regression and Gamma Process Model. Reliability Engineering & System Safety 210 (2021): 107504.

[86] Z Huang, Z Xu, X Ke, W Wang, Y Sun. Remaining useful life prediction for an adaptive skew-Wiener process model. Mechanical Systems and Signal Processing 87.Part A (2017): 294-306.

[87] Bae, Suk Joo, et al. A Bayesian approach to modeling two-phase degradation using change-point regression. Reliability Engineering & System Safety 134 (2015): 66-74.

[88] Zhang, Jianxun, et al. Lifetime Estimation for Multi-Phase Deteriorating Process with Random Abrupt Jumps. Sensors 19.6 (2019): 1472; https://doi.org/10.3390/s19061472.

[89] Guobo Liao, Hongpeng Yin, Min Chen, Zheng Lin. Remaining useful life prediction for multi-phase deteriorating process based on Wiener process. Reliability Engineering & System Safety 207 (2021): 107361; https://doi.org/10.1016/j.ress.2020.107361.

[90] Roderick Murray-smith, Carl Edward Rasmussen , Agathe Girard. Gaussian Process Model Based Predictive Control. Proceedings of 4th American Control Conference. 2004.

[91] Doksum, Kjell A. and Sharon-Lise T. Normand. Gaussian models for degradation processes-Part I: Methods for the analysis of biomarker data. Lifetime Data Analysis 1 (1995): 131-144.

[92] Piyush Shakya, Makarand S. Kulkarni, Ashish K. Darpe. A novel methodology for online detection of bearing health status for naturally progressing defect. Journal of Sound and Vibration 333.21 (2014): 5614-5629.

[93] Singh, Jaskaran, A. K. Darpe and S. P. Singh. Bearing damage assessment using Jensen-Rényi Divergence based on EEMD. Mechanical Systems and Signal Processing 87 (2017): 307-339, DOI: 10.1016/j.ymssp.2016.10.028.

[94] Metal Scan MS1000. n.d. Gastops. 20 06 2021. <https://www.gastops.com/products/metalscan/ms-1000/>.

[95] Dupuis, Richard. Application of Oil Debris Monitoring For Wind Turbine Gearbox Prognostics and Health Management. Annual Conference of the Prognostics and Health Management Society. 2010.

[96] Kitaljevich, J. L. Miller and D. In-line oil debris monitor for aircraft engine condition assessment. Proceedings of the IEEE Aerospace Conference. doi: 10.1109/AERO.2000.877882., 2000.

[97] Sheng, Shuangwen. Monitoring of Wind Turbine Gearbox Condition through Oil and Wear Debris Analysis: A Full-Scale Testing Perspective. Tribology Transactions 59.1 (2016): 149-162; DOI: 10.1080/10402004.2015.1055621.

[98] W.W. Seifert, V.C. Westcott,. A method for the study of wear particles in lubricating oil. Wear 21.1 (1972): 27-42, https://doi.org/10.1016/0043-1648(72)90247-5.

[99] DP, Anderson. Wear particle atlas Report No.: NAEC-92-165. Lakehurst (NJ): Naval Air Engineering Center, Department SEE, 1982.

[100] B.J.Roylance, S.Raadnui. The morphological attributes of wear particles — their role in identifying wear mechanisms. Wear 175.1-2 (1994): 115-121.

[101] Lockwood FF, Dally. Lubricant Analysis. Metals Handbook, Tribology, 18th edition. ASM, 1992. 300-312.

[102] B, Bushan. Wear Debris Classification. Modern Tribology Handbook. CRC Press, 2000. 309-311.

[103] Adams RP, MacKay DJC. Bayesian Online Changepoint Detection. Machine Learning (2007).

[104] Li-Yu Hu, Min-Wei Huang, Shih-Wen Ke & Chih-Fong Tsai. The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus 5 (1034).https://doi.org/10.1186/s40064-016-2941-7 (2016).

[105] Liu, H., and Motoda, H. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers, 1998.

[106] J. Novakovic, P. Strbac and D. Bulatovic. Toward optimal feature selection using ranking methods and classification algorithms,. Yugoslav Journal of Operations Research 21.1 (2011): 119-135.

[107] Hall, M.A., and Smith, L.A. Practical feature subset selection for machine learning. Proceedings of the 21st Australian Computer Science Conference. 1998.

[108] Tianyu Liu, Quan Sun, Jing Feng, Zhengqiang Pan and Qizi Huangpeng. Residual life estimation under time-varying conditions based on a Wiener process. Journal of Statistical Computation and Simulation 87.2 (2017): 211-226.

[109] Pitt, Michael K. and Neil Shephard. Filtering via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association 94.446 (1999).

[110] Musso C., Oudjane N., Le Gland F. Improving Regularised Particle Filters. Sequential Monte Carlo Methods in Practice. Statistics for Engineering and Information Science. New York: Springer, 2001. 247-271.

[111] O'Hagan, A. Curve fitting and optimal design for prediction. Journal of the Royal Statistical Society. Series B (Methodological) 40.1 (1978): 1-42.

[112] Rasmussen, C. E. Evaluation of Gaussian processes and other methods for non-linear regression. 1996

[113] Chan, Lester Lik Teck, Yi Liu and Junghui Chen. Nonlinear System Identification with Selective Recursive Gaussian Process Models. Industrial & Engineering Chemistry Research 52.51 (2013).

[114] Gregorčič, Gregor and Gordon Lightbody. Gaussian process approach for modelling of nonlinear systems. Engineering Applications of Artificial Intelligence 22.4-5 (2009).

[115] Kundu, Pradeep, Ashish K. Darpe, and Makarand S. Kulkarni. Weibull Accelerated Failure Time Regression Model For Remaining Useful Life Prediction Of Bearing Working Under Multiple Operating Conditions. Mechanical Systems and Signal Processing 134 (2019).

[116] Lv, Mingzhu, et al. General Log-Linear Weibull Model Combining Vibration and Temperature Characteristics for Remaining Useful Life Prediction of Rolling Element Bearings. Shock and Vibration, Hindawi (2020): 1070-9622.

[117] Rosado, Lewis, Nelson H. Forster and and Jason W. Cooke Kevin L. Thompson. Rolling Contact Fatigue Life and Spall Propagation of AISI M50, M50NiL, and AISI 52100, Part I: Experimental Results. Tribology Transactions 53 (2010): 29-41.

[118] N.H, Forster, et al. Assessing the Potential of a Commercial Oil Debris Sensor as a Prognostic Device for Gas Turbine Engine Bearings. AFRL presentation at ISHM conference. Cincinnati, OH, 2005.

[119] Stahl, Saul. The Evolution of the Normal Distribution. Mathematics Magazine Apr 2006: 96-104.

[120] Conard, Keith. The Gaussian Integral. University of Connecticut, 2016: 1-2.

[121] Reddy S, Mun M, Burke J, Estrin D, Hansen M, Srivastava M. Using mobile phones to determine transportation modes. ACM Transactions on Sensor Networks 6.2 (2010).

[122] Manhyung Han, La The Vinh, Young-Koo Lee and Sungyoung Lee. Comprehensive Context Recognizer Based on Multimodal Sensors in a Smartphone . sensors 12 (2012): 12588-12605; doi:10.3390/s120912588.

[123] Yoshinobu Kawahara, Masashi Sugiyama. Sequential change-point detection based on direct density-ratio estimation. Statistical Analysis and Data Mining 5.12 (2012): 144-127; https://doi.org/10.1002/sam.10124.

[124] Keogh E, Chu S, Hart D, Pazzani M. An online algorithm for segmenting time series. IEEE International Conference on Data Mining. 2001.

[125] Harchaoui Z, Vallet F, Lung-Yut-Fong A, Cappe O. A regularized kernel-based approach to unsupervised audio segmentation. IEEE International Conference on Acoustics, Speech and Signal Processing. 2009.

[126] Cho H, Fryzlewicz P. Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. Journal of the Royal Statistical Society 77.2 (2015): 475-507.

[127] Cook, Samaneh Aminikhanghahi. Diane J. A Survey of Methods for Time Series Change Point Detection. Knowledge and Information Systems 51 (2017; https://doi.org/10.1007/s10115-016-0987-z): 339-367.