Acoustic emission-based diagnostics and prognostics of slow rotating bearings using Bayesian techniques

PhD thesis

by

Sylvester Aondolumun Aye

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

Department of Mechanical and Aeronautical Engineering
Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

February, 2014

Acoustic emission-based diagnostics and prognostics of slow rotating bearings using Bayesian techniques

by

Sylvester Aondolumun Aye

Supervisor: Prof. P.S. Heyns

Department of Mechanical and Aeronautical Engineering

Degree: Doctor of Philosophy

Abstract

Diagnostics and prognostics in rotating machinery is a subject of much on-going research. There are three approaches to diagnostics and prognostics. These include experience-based approaches, data-driven techniques and model-based techniques. Bayesian data-driven techniques are gaining widespread application in diagnostics and prognostics of mechanical and allied systems including slow rotating bearings, as a result of their ability to handle the stochastic nature of the measured data well. The aim of the study is to detect incipient damage of slow rotating bearings and develop diagnostics which will be robust under changing operating conditions. Further it is required to explore and develop an optimal prognostic model for the prediction of remaining useful life (RUL) of slow rotating bearings.

This research develops a novel integrated nonlinear method for the effective feature extraction from acoustic emission (AE) signals and the construction of a degradation assessment index (DAI), which is subsequently used for the fault diagnostics of slow rotating bearings. A slow rotating bearing test rig was developed to measure AE data under variable operational conditions. The proposed novel DAI obtained by the integration of the PKPCA (polynomial kernel principal component analysis), a Gaussian mixture model (GMM) and an exponentially weighted moving average (EWMA) is shown to be effective and suitable for monitoring the degradation of slow rotating bearings and is robust under variable operating conditions. Furthermore, this study integrates the novel DAI into alternative Bayesian methods for the prediction of RUL. The DAI is used as input in several Bayesian regression models such as the

multi-layer perceptron (MLP), radial basis function (RBF), Bayesian linear regression (BLR), Gaussian mixture regression (GMR) and the Gaussian process regression (GPR) for RUL prediction. The combination of the DAI with the GPR model, otherwise, known as the DAI-GPR gives the best prediction. The findings show that the GPR model is suitable and effective in the prediction of RUL of slow rotating bearings and robust to varying operating conditions. Further, the models are also robust when the training and tests sets are obtained from dependent and independent samples.

Finally, an optimal GPR for the prediction of RUL of slow rotating bearings based on a DAI is developed. The model performance is evaluated for cases where the training and test samples from cross validation approach are dependent as well as when they are independent. The optimal GPR is obtained from the integration or combination of existing simple mean and covariance functions in order to capture the observed trend of the bearing degradation as well as the irregularities in the data. The resulting integrated GPR model provides an excellent fit to the data and improvements over the simple GPR models that are based on simple mean and covariance functions. In addition, it achieves a near zero percentage error prediction of the RUL of slow rotating bearings when the training and test sets are from dependent samples but slightly different values when the estimation is based on independent samples. These findings are robust under varying operating conditions such as loading and speed. The proposed methodology can be applied to nonlinear and non-stationary machine response signals and is useful for preventive machine maintenance purposes.

Keywords: acoustic emission, Bayesian linear regression, Bayesian techniques, covariance function, data-driven, degradation assessment index, diagnostics, experience-based, exponentially weighted moving average, Gaussian mixture model, Gaussian mixture regression, Gaussian process regression, integration, mean function, model-based, multi-layer perceptron, polynomial kernel principal component analysis, prognostics, radial basis function, remaining useful life

List of Publications

1) Aye, S.A. and Heyns, P.S. (2011). Effect of speed and torque on statistical parameters in tapered bearing fault detection, World Academy of Science, Engineering and Technology, 78, 2011, 759–761.

2) Aye, S.A. and Heyns, P.S. (2012). Condition monitoring of tapered roller bearings: A photogrammetric approach, In Proceedings of SPIE, Third Asia Pacific Optical Sensors Conference, 31 January - 3 February, 2012, Sydney, Australia.

3) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2012). Seeded damage detection on self-aligning ball bearings with acoustic emission, Condition Monitoring of Machinery in Non-Stationary Operations, Editors: Fakhfakh, T.; Bartelmus, W.; Chaari, F.; Zimroz, R.; Haddar, M., 521–526. ISBN: 978–3–642–28767–1.

4) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2014). Health diagnostics of slow rotating bearings based on Gaussian process regression, In South African Conference on Computational and Applied Mechanics Proceedings (SACAM 2014), Somerset West, Cape Town, South Africa, 14–16 January, 2014.

5) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2014). Slow rotating bearing condition assessment based on Bayesian Gaussian mixture regression, in 2014 International Measurement Confederation (IMEKO) TC3–TC5–TC22 Conference Proceedings, Cape Town, South Africa, 3–5 February, 2014.

6) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2014). Condition monitoring of slow rotating bearings based on Bayesian linear regression. In Proceedings of the International Conference on Advances in Civil, Structural and Mechanical Engineering (CSM), London, UK, 2–3 June, 2014.

7) Aye, S.A. and Heyns, P.S. (2014). An integrated Gaussian process regression for prediction of remaining useful life of slow rotating bearings based on acoustic emission, Submitted to Mechanical Systems and Signals Processing (submitted November, 2013).

8) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2014). Diagnostics of slow rotating bearings using a novel DAI based on acoustic emission, Submitted to International Journal of Applied Mechanics (submitted January, 2014).

9) Aye, S.A. and Heyns, P.S. (2014). Acoustic emission-based prognostics of slow rotating bearing using Bayesian techniques under dependent and independent samples, Submitted to Applied Artificial Intelligence (submitted January, 2014).

10) Aye, S.A., Heyns, P.S. and Thiart, C.J.H. (2014). Review of slow rotating bearing diagnostics and prognostics: advances and future, Under preparation.

Acknowledgement

I wish to thank the Almighty God – my maker, Jesus Christ – my Lord and Saviour, and the Holy Spirit – my comforter and source of inspiration for enabling me complete this thesis.

I wish to express my gratitude to my supervisor, Prof. P.S. Heyns, for his technical guidance and financial support throughout the research work, which led to the successful completion of this study. He was indeed a mentor. My thanks also go to Dr C.J.H. Thiart for his constructive criticism, Prof. N.J. Theron for his fatherly advice and counsels, Prof. J.P. Meyer for hosting me for the PhD. Programme and all staff of the Department of Mechanical and Aeronautical Engineering, University of Pretoria.

My sincere appreciation goes to my wife – Dr G.C. Aye and children – Joy and Jewel for their unflinching love and support throughout my PhD programme. To my late father – Major B.C. Aye (Rtd), my mother – Mrs C.T. Aye, my siblings – Mercy, Bridget, Alice, Amos, Blessing, Daniel, Ruth, Peace and Grace, my father in–law – Mr Oparah Benson, my mother in–law – Mrs Eileen Oparah, brothers and sisters in–laws, Ehiedu, Tunde, Akin, Taiwo, Lilian, Favour, Lawman, Ngozi, Kennedy, Glory and Fortunatus. I thank you very much for your love and moral support. You have been a source of joy and strength to me.

I sincerely appreciate the efforts made by the Centre for Asset Integrity Management (formerly the Dynamic Systems Group) to make my research in the group very fulfilling and enjoyable. My friends and colleagues – Ulrich Spangenberg, Jan-Sjoerd Van den Bergh, Donna–Lee Faure, Rudi Kroch, Johan Conradie, and all staff and colleagues in Sasol Laboratory for Structural Mechanics and the Department of Mechanical and Aeronautical Engineering, University of Pretoria – I thank you all for your regular support, as well as for your liberal social and emotional support. Special thanks to the laboratory technical instructors, George Breitenbach and Herman Booysen, for their assistance in the test setup. Also, I wish to thank the Centre secretary, Mrs M.P. Calder, for her kind assistance in administrative tasks. To all my relatives, friends and well–wishers too numerous to mention – thank you for your constant support.

Finally, I also want to thank the Tertiary Educational Trust Fund and my employer – the Department of Mechanical Engineering and Federal University of Agriculture Makurdi, Nigeria – for granting me study leave in order to further my education.

Table of Contents

List of Figures

List of Tables

Glossary

Abbreviations

Abbreviations	Description
AE	Acoustic Emission
ARD	Automatic Relevance Determination
AMGPR	Affine Mean GPR
BLR	Bayesian Linear Regression
CI	Confidence Interval
CM	Condition Monitoring
CMGPR	Constant Mean GPR
CuMGPR	Cubic Mean GPR
CV	Cross Validation
DAI	Degradation Assessment Index
DT	Degradation Threshold
EWMA	Exponentially Weighted Moving Average
GKPCA	Gaussian Kernel Principal Component Analysis
GMM	Gaussian Mixture Model
GMR	Gaussian Mixture Regression
GP	Gaussian Processes
GPR	Gaussian Process Regression
KDE	Kernel Density Estimation
LMGPR	Linear Mean GPR
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
NLL	Negative Log Likelihood
PC	Principal Component
PCA	Principal Component Analysis
PE	Prediction Error
PKPC	Polynomial Kernel Principal Component

PKPCA	Polynomial Kernel Principal Component Analysis
QMGPR	Quadratic Mean GPR
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RUL	Remaining Useful Life
SE	Squared Exponential
SKPCA	Simple Kernel Principal Component Analysis
SPE	Square Prediction Error
ZMGPR	Zero Mean GPR

Symbols

Symbols	**Description**
μ	Mean parameter
μ_o	Mean vector
σ^2	Variance parameter
τ	Precision
π	Pi
T	Exponent T which means transpose
w^j	Parameter vector
$p(D\|w)$	Likelihood
$p(y_i^j \| x_i^j, w^j)$	Likelihood of the data
j	Loading condition
δ	Kronecker delta function
F	Feature space
$\mathbf{v}$, α_1	Eigenvector, Eigenvectors
λ	Eigenvalue

$\mathbf{C}^F$	Covariance matrix
Σ_e	Covariance matrix
Σ_o	Covariance matrix
$\mathbf{K}$	$N \times N$ kernel matrix
$p(x)$	Probability density
W_t	EWMA
α	Smoothening constant
φ	Parameter vector
$N(x\|\theta_j)$	Normal probability distribution
$p(x\|\theta_j)$	Component density
π_j	Mixing coefficients
$p(x)$	Finite mixture model
M	Number of components
φ^*	Optimum vector
w_1	Numerical weights
w_{kj}	Output layer weights
$p(w\|D)$	Posterior distribution of weights or Probability density function
D	Data
$f(x^n, w)$	MLP output
ϕ_j	Basis functions
$p(w)$	Prior distribution
$p(y \mid X)$	Marginal likelihood
$\mathbf{x}$	Input
$\mathbf{X}$	Training inputs
$\mathbf{X}_N$	Input vectors
$\mathbf{X}_*$	Test inputs

$\mathbf{y}_*$	Prediction
$\mathbf{y}_N$	Target vector
$\bar{y}_*$,	MAP estimate
y_i^j	Observation
D	Length
l	A vector of length D
M	Diagonal matrix with positive ARD parameters
$f(\mathbf{x})$	Gaussian process
φ	Hyperparameters
φ_{MP}	Maximum a posteriori hyperparameter

Chapter 1 Introduction and literature review

1.1 Introduction

Bearings, including slow rotating bearings operating at varying speeds, are one of the major machinery component types used in industries like mining, automotive, power generation, railway, shipping, manufacturing, and chemical plant (Dube *et al.*, 2013). Bearing failure has been identified as the leading source of failure in rotating machinery (Li *et al.*, 1999). Research shows that bearing defects amount to about 40% of motor faults (Enzo and Ngan, 2010). Bearing failure may occur due to mechanisms such as fatigue, wear, corrosion, debris contamination and misalignment, faulty installation or improper mounting, blockade, passage of foreign particles, inadequate or improper lubrication, excessive speed and inappropriate temperature, vibration, inefficient seals, and overloading (Camci *et al*., 2012; Dube *et al.,* 2013). There are three major components of bearings that typically experience damage. These are the ball or rolling element, inner race and outer race.

There are many condition monitoring (CM) techniques for acquiring data for bearing diagnostics and prognostics. The data can be acoustic emission (AE), vibration, oil analysis, temperature, wear debris analysis, etc. collected using different sensors, such as AE sensors, accelerometers, micro-sensors, ultrasonic sensors, etc. (Tandon and Choudhury, 1999; Jardine *et al*., 2006; Randall, 2011). Vibration analysis is the most widely used. However, while vibration based methods have been shown to be effective when the defect in the bearings has already become severe, the vibration signal is not sensitive to an incipient fault. Furthermore, the vibration signal caused by bearing defects is often contaminated and distorted by other faults and mechanical noise (He *et al.,* 2009). Although, it is possible to detect typical faults using vibration analysis, this is only effective primarily in high speed machinery (Kim *et al.*, 2007).

Slow rotating bearings present special challenges such as significant speed variation, huge load variations, exceedingly high down time costs, corrosion, contamination,

wide range of failure mechanisms, etc. Traditional bearing monitoring approaches are not necessarily appropriate for CM of slow rotating bearings. Problems include inappropriateness of harmonic analysis, low frequencies which lead to low bearing accelerations, the sheer size of real life bearings which may lead to significant attenuation of measured signals from one side of the bearing to another, etc. Based on experiences from the literature, especially work done on slow rotating bearings; there are however a number of approaches which hold promise. It is clear that AE promises distinct advantages. These include viability and high success in detection of damage in slow rotating bearings. AE offers earlier fault detection due to its higher sensitivity (Mba, 2003), and could be useful for an extensive range of rotational speeds with noteworthy benefits at slow rotational speeds. Other advantages of AE include: it offers real time process information; it is non-invasive; and it can be used for dynamic CM. Despite the distinct advantages of AE for monitoring the condition of slow rotating bearings, it has some limitations which comprise: vulnerability to attenuation; vulnerability to noise in operation; and failure to match faulty AE signal to the mechanism of failure. One problem is the high frequencies that lead to large data files and the requirement for large memory space during the data acquisition phase. Another shortcoming when processing the signal is the lack of periodicity that makes it inappropriate to process the signal in the frequency domain. Hence, the time domain approach could be employed.

Several techniques have been applied to data processing of signals into interpretable features that can be used for mechanical systems degradation assessment. These are broadly categorised into three domains: the time domain, the frequency domain and the time-frequency domain. Obtaining the most beneficial statistics from these features and inputting to the diagnostics and prognostics system is a challenge. In machine learning, the accuracy of fault detection and prediction models depends on the quality and sensitivity of the features utilised to evaluate the condition and spread of the faults (Camci *et al*., 2012). The dimension of the original feature set from the time domain analysis is normally large. Thus, it is problematic selecting a priori which feature(s) is more sensitive to fault growth as various factors (e.g., signal-to-noise ratios of the data acquisition system, location of sensors, etc.) have an effect on

the efficacy of the features (Malhi and Goa, 2004). Given the fact that each feature has both advantages and disadvantages, it is crucial to develop a systematic algorithm that is capable of selecting the most characteristic features for monitoring the present machinery condition (Malhi and Gao, 2004).

An additional challenge of bearing diagnostics and prognostics is carrying out the effective evaluation of degradation utilizing the extracted features. Despite the fact that a large number of features can be extracted to characterise the AE data, earlier investigations have indicated that every feature is only effective for particular faults in particular phases (Xi *et al.*, 2000; Yu, 2011a). Moreover, it is hard to quantitatively diagnose the fault severity, especially at the early stage of fault. It is still very challenging to predict the future trend due to strong stochastic characteristics of the failure propagation process (Li *et al*., 2000). Hence, there is a need for data reduction and a methodology that captures the stochastic nature of the data.

Two important aspects in CM of bearings are diagnostics and prognostics of machinery faults. The distinction between these two aspects is clearly identified by Jardine *et al*. (2006). The act of detecting, isolating and identification of a fault when it occurs (post event analysis) is referred to as diagnostics. Fault detection is a task of detecting a deviation of the monitored system from healthy or normal operation; the isolation of faults is a task of locating the faulty asset; and identification of fault is a task of determining the type of fault after detection. “The overall theory of fault diagnostics consists of the three important tasks: fault detection– detection of the time of occurrence of faults in the functional units of the process, which lead to undesired or unacceptable behaviour of the whole system; fault isolation– localization (classification) of different faults; and fault analysis or identification– determination of the type, magnitude and cause of the fault. A fault diagnostic system, depending on its performance, is called an FD (for fault detection) or FDI (for fault detection and isolation) or FDIA (for fault detection, isolation and analysis) system, whose outputs are correspondingly alarm signals to indicate the occurrence of the faults, or classified alarm signals to show which fault has occurred, or data of defined types providing information about the type or magnitude of the fault” (Frank *et al*., 2000). However,

the most important aspect of diagnostics is fault detection and quantification (Sikorska *et al*., 2011; Zaidan *et al*., 2011) which is subsequently used for prognostics.

The question of where on the overall health curve does the component or system reside is addressed by prognostics. Prognostics also imply the prediction of the time progression of incipient damage to the time of final failure (Brotherton, 2000; Sikorska *et al*., 2011). The act of predicting or forecasting final failure before it occurs (prior event analysis) is called prognostics. Prediction of a fault refers to the task to determine the imminence of a defect and the estimation of how soon a defect will likely happen. Once the current health condition is defined, the next task is to predict the change in component health as a function of remaining useful life (RUL) based on anticipated future health states (Camci *et al.,* 2012). Jardine *et al.* (2006) defined the RUL as a conditional random variable of the time left prior to failure, given the present asset life and condition as well as history of operations. The rate of damage growth due to a given set of operating conditions (for example speed or load) is dependent upon the properties of the particular material under consideration (Bolander *et al.*, 2009).

There is no doubt that prognostics is surrounded by uncertainties arising from a variety of sources such as the current age of the asset or mechanical system, the health information or observed condition, measurement noise, process noise, modelling uncertainty and the environment in which the system is operated (Si *et al*., 2011) which makes the process inherently stochastic. Therefore, the behaviour observed from a particular run may not reflect the true nature of prediction trajectories. It is therefore expected that a prognostics algorithm should provide information about the confidence around the prediction (Saxena *et al.*, 2009). Bayesian techniques which are mainly statistical are gaining widespread application in damage detection and RUL of bearings due to their ability to handle uncertainties as opposed to traditional statistical methods (Nabney, 2002).

Hence, in this research an improved novel degradation assessment index (DAI) combining the advantages of its various models is developed for diagnostics of slow rotating bearings. This novel DAI is subsequently used in several Bayesian regression models for prognostics. The best performing prognostics model is selected and further improved upon to obtain the best prediction of RUL of the slow rotating bearings.

1.2 Literature review

In this review we investigate data acquisition methods; data processing methods; data reduction methods; diagnostics; prognostics and new evolving Bayesian prognostic techniques. The schematic of the approach is illustrated in Figure 1.1.

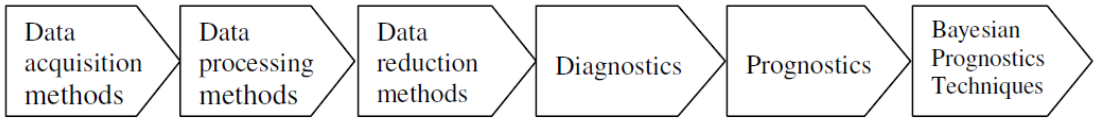

Figure 1.1: Schematics of approach

1.3 Data acquisition methods

The practice of gathering and storage of valuable data from mechanical systems is referred to as data acquisition. A vital step in diagnostics and prognostics of machinery faults is data acquisition. There are two types of data that can be acquired: the incident data and the condition data. The data on what took place (e.g., setting up, failure, repair, etc., and for what reasons) and/or what was effected (e.g., slight overhaul, precautionary repairs, oil change, etc.) to the mechanical system is referred to as incident data. On the other hand data on the health condition of the physical asset or mechanical system is referred to as condition data. Different types of condition data include AE, vibration, temperature, oil analysis data, moisture, pressure, humidity, environment, motor current, ultrasonic signals, partial discharge, etc. The most commonly used condition data are AE and vibration signals. Equally a

variety of sensors, such as AE sensors, micro-sensors, ultrasonic sensors, etc., have been designed to gather diverse forms of data (Jardine *et al.*, 2006).

1.3.1 Oil analysis

Oil analysis has been used for condition monitoring for several decades (Elforjani, 2010). Oil degradation monitoring is an important source of information for early machine failure detection (Zhu *et al.*, 2013). The essence of monitoring the oil degradation process is to provide early warning of machine failure thereby extending the operational duration of lubrication oil reducing the frequency of oil changes and consequently reduce cost of maintenance. The main test used in oil analysis include: viscosity analysis; particle counts analysis; machines wear analysis; oxidation analysis and water content analysis. Oil and wear debris analysis have proven in many instances to be a leading indicator compared to vibration analysis in machinery diagnostics (Peng *et al., 2005*). The technique of oil analysis is generally applied off-line by taking samples. However, the use of online sensors is gaining widespread usage in order to safeguard the oil quality (Elforjani, 2010; Verbruggen, 2003).

1.3.2 Vibration analysis

Vibration analysis is the most widely used method for CM. However, while vibration based methods have been shown to be effective when the defect in the bearings has already become severe, the vibration signal is not sensitive to an incipient fault. Furthermore, the vibration signal caused by bearing defects is often contaminated and distorted by other faults and mechanical noise (He *et al.,* 2009). Although, it is possible to detect typical faults using vibration analysis, this is only effective primarily in high speed machinery (Kim *et al.*, 2007).

Normal bearing monitoring techniques such as vibration and oil analysis are not necessarily appropriate for CM of slow rotating bearing rotating at slow speeds less than 100 rpm (Elforjani, 2010). AE is utilized in the CM of bearings and rotating machinery although it was initially developed for non-destructive testing of static

structures. It has the benefit of detecting damage earlier than oil and vibration analysis (Mba, 2003; Morhain and Mba, 2003).

1.3.3 AE technology

AE refers to the spontaneous release of elastic energy as transient elastic waves as a result of rapid release of strain energy initiated by a structural change in an asset under thermal or mechanical stresses (Mazal, 2009; Tandon and Choudhury, 1999; Elforjani and Mba, 2011). Balerston was one of the first to introduce AE techniques for the defect diagnostics of rolling element bearings and proposed the AE source mechanism (Nienhaus *et al.*, 2012). Since then, and especially in recent years, more and more researchers (Yoshioka, 1992; Li and Li, 1995; Shiroishi *et al.*, 1997; Mba *et al.*, 1999; Jamaludin and Mba, 2002; Mba, 2005; Elforjani and Mba, 2008; Elforjani and Mba, 2011) have investigated the application of AE techniques in the CM of bearings and allied applications. Several studies (Kakishima *et al.*, 2000; Hort *et al.,* 2010; Tavakoli, 1991; Lees *et al.,* 2011; Cai *et al.,* 2011; Miettinen, 2000; Tandon and Choudhury, 1999; Tandon and Nakra, 1992a,b; Al-Ghamd and Mba, 2006; Qi *et al.,* 2008; Miettinen and Pataniitty, 1999; Kim *et al.,* 2007; Widodo, *et al.,* 2009) also used AE in bearing damage detection and allied applications.

Materials that can be monitored using AE without destroying them include: metals, composite materials and polymers, rocks, concrete, and woods (Elforjani, 2010). Applications of AE technology can also be used for monitoring aircraft and aerospace, marine, wind turbine, pressure and civil engineering structures and equipment (Elforjani, 2010; Capgo Pty Ltd, 2009; Envirocoustics, 2007).

AE measurements are generally divided into two types namely: active type where the excitation is implemented externally and passive type where the excitation is performed by the component (Elforjani, 2010). AE signals are normally within a frequency range of 20 kHz to 1 MHz (He *et al.,* 2009; Randall and Antoni, 2011; Elforjani and Mba, 2011). Sensors that convert the surface displacement of the asset

into electrical signals are utilised for detection. The electrical signals are subsequently processed using suitable equipment and/or data processing methods, to describe the asset condition.

AE has distinct advantages which include viability and high success in detection of damage in slow rotating bearings; earlier fault detection due to its higher sensitivity (Mba, 2003), and useful for an extensive range of rotational speeds with noteworthy benefits at slow rotational speeds; it offers real time process information; it is non-invasive; and it can be used for dynamic CM. Despite the distinct advantages of AE for monitoring the condition of slow rotating bearings, it has some limitations which comprise: vulnerability to attenuation; vulnerability to noise in operation; and failure to match faulty AE signal to the mechanism of failure. Disadvantages include: high frequencies that lead to large data files and the requirement for large memory space during the data acquisition phase; lack of periodicity when processing the signal that makes it inappropriate to process the signal in the frequency domain. Therefore, the time domain approach could be employed.

1.4 Data processing methods

Usually the data is cleaned since it could contain some errors (Jardine *et al.,* 2006). The data is then subsequently used for analysis and modelling. Various signal processing methods have been established for the evaluation and interpretation of the acquired data to aid in diagnostics and prognostics. Feature extraction refers to the process of obtaining vital information from the data. There are several signal processing methods and procedures for diagnostics and prognostics of engineering systems. As earlier mentioned in paragraph 1.1, these are the time domain, frequency domain and time-frequency signal processing methods (Jardine *et al.,* 2006) discussed in the next paragraphs.

1.4.1 Time-domain analysis

Time-domain analysis refers to analysis done in the time domain. This involves the computation of descriptive statistical features such as crest factor, peak, mean, peak-to-peak interval, root mean square, standard deviation, kurtosis, skewness, etc. from the data. Several studies (Heng and Nor, 1998; Dyer and Stewart, 1978; Ferreira *et al.,* 2003; Vlok *et al.,* 2004) utilized statistical features enumerated above in fault detection of bearings. As a result of substantial speed variation in slow rotating bearings the time domain is more appropriate for CM of slow rotating bearings (Mba, 2003).

Extracting the most useful information from these features as inputs into diagnostics poses a challenge. The accuracy of fault detection models in machine learning depends on the sensitivity and quality of the features used to assess the condition and growth of the defects (Camci *et al*., 2012).

1.4.2 Frequency-domain analysis

Analysis in the frequency-domain involves the transformation of the time domain signal to the frequency domain by the use of fast Fourier transform. This enables the CM of machinery by extracting features from the frequency spectrum. This is achieved by looking at the frequencies of interest from the obtained frequency spectrum (Jardine *et al.,* 2006).

1.4.3 Time-frequency analysis

One of the shortcomings of the frequency domain is not being able to handle non-stationary data. However, time-frequency analysis, which uses time-frequency distributions, is suitable for analysing non-stationary data when mechanical system failure takes place, in the time and frequency domains. Time-frequency investigation utilizes time–frequency distributions, which characterize the energy or power of data in two-dimensional functions of frequency and time to signify failure configurations

for better diagnostics and prognostics. The most frequently used time-frequency distributions are the Wigner-Ville distribution, Short-time Fourier transform (STFT) and Hilbert-Huang Transform (Jardine *et al.,* 2006; Randall, 2011).

1.5 Data reduction methods

High dimensional data is difficult to deal with. This problem of high dimensionality can be solved by reducing the data into fewer principal components by the use of data reduction techniques which captures all the vital features. Different data reduction methods include the principal component analysis (PCA), linear graph embedding (LGE), linear discriminant analysis (LDA), locality preserving projections (LPP), neighbourhood preserving embedding (NPE) and kernel principal component analysis (KPCA).

The traditional principal component analysis (PCA) is a widely used technique for dimensionality reduction and feature extraction in machine learning. However, it is linear in nature just like the other data reduction techniques: LGE, LDA, LPP and NPE. In general, the standard PCA can only be effectively performed on a set of observations that vary linearly. When the variations are nonlinear, PCA becomes invalid.

Another technique is kernel principal component analysis. The key idea of kernel principal component analysis (KPCA) is both intuitive and generic (Schölkopf *et al.*, 1998; Schölkopf *et al.*, 1999). Unlike other nonlinear methods, such as neural networks, KPCA does not involve a nonlinear optimization procedure. Instead, KPCA maps the problem from the input space to a new higher-dimensional space (called feature space) by doing a nonlinear transformation using suitably chosen basis (kernel) functions and then use a linear model in the feature space (Lee *et al.*, 2004). As a result, KPCA performs a nonlinear PCA in the input space (Romdhani *et al.*, 1999).

1.6 Diagnostics

In industry the CM and fault diagnostics of equipment and processes is of enormous importance. Prompt discovery of defects in mechanical systems can lead to prevention of huge expenses in emergency maintenance costs. Machine fault diagnostics involves the mapping of data acquired during measurement and/or features extracted to machine failures in the failure space. The mapping procedure is also referred to as pattern recognition. It is attained by classification of data based on extracted features (Jardine *et al.,* 2006). Diagnostics is a more mature field than prognostics. Once degradation is detected, unscheduled maintenance is normally performed to prevent the failure consequences (Eker *et al.,* 2012). Machinery diagnostics is a vital research area that requires inter-disciplinary expertise which requires intelligent use of techniques from diverse domains such as machinery dynamics, signal processing, statistical analysis and machine intelligence. Due to the very complex nature of the problem, the current solutions leave much to be desired. Hence there is a continual need for new paradigms for diagnostics (Nataraj and Kappaganthu, 2011).

One of the challenges of bearing diagnostics is the effective evaluation of the degradation process, based on the features extracted. In spite of the fact that a large selection of extracted features can be used to depict the characteristics of AE signals, earlier studies have demonstrated that each feature is only effective for depicting specific defects at specific stages (Yu, 2011a). Moreover, it is difficult to quantitatively diagnose the fault severity, especially at an early stage of a fault. A number of diagnostic algorithms have been proposed in the literature. The estimation methods fall into two main approaches, namely the data-driven and model-based approaches (Randall, 2011; Sikorska *et al*., 2011). The data-driven approach is divided into statistical approaches (Phelps *et al*., 2001; Wang, 2002; Vlok *et al*., 2004; Banjevic and Jardine, 2006) and artificial intelligence (AI) approaches (Zhang and Ganesan, 1997; Yam *et al*., 2001; Wang *et al*., 2004). Several studies (Bolander *et al*., 2009; Camci *et al*., 2012) are applications of model-based approaches.

1.6.1 Data-driven diagnostic methods

The data-driven diagnostic methods comprise the AI, statistical and other approaches. These are discussed in the following paragraphs.

1.6.1.1 Artificial intelligence approaches

The application of AI methods in diagnostics of machinery has increased due to its improved performance over traditional statistical approaches. Nonetheless, it is difficult implementing AI methods as a result of the lack of proficient algorithms for acquiring data for training and particular knowledge, which are necessary for training the models. Until now, the majority of the applications simply used measured data for training of the model in the literature. Expert systems and artificial neural networks (ANNs) are the most commonly used AI methods in literature. Other AI methods in use include Fuzzy–neural networks, fuzzy logic systems, evolutionary algorithms and neural–fuzzy systems (Jardine *et al.*, 2006). AI methods have been integrated with ESs and Neural Networks in improving diagnostics of machinery (Garga *et al.,* 2001).

The ANN is a model which imitates the human intelligence in its computations. It comprises uncomplicated processing elements linked in a complex layered configuration enabling the model to estimate an intricate nonlinear function having multiple inputs and multiple outputs. A processing element is made up of a weight and a node. The ANN training process refers to adjustment of the unknown function using its weights in conjunction with input and output observations. The most commonly utilized neural network structure for defect diagnostics is the Feed forward neural network (Roemer *et al.,* 1996; Larson *et al.,* 1997; Li *et al.,* 2000).

1.6.1.2 Statistical methods

This involves the use of statistical approaches in the CM of data. The statistical methods comprise the frequentist (traditional statistics) and Bayesian approaches.

The traditional statistics or features used include crest factor, peak, mean, peak-to-peak interval, root mean square, standard deviation, kurtosis, skewness, etc. (Skormin *et al*., 1999; Artes *et al*., 2003; Schurmann, 1996). The dimension of the initial feature set from the time domain analysis is usually large. It is a huge challenge to obtain useful information from these features so that it can be used in the diagnostics system. The accuracy of fault detection depends on the quality and sensitivity of the extracted features used in the estimation of the bearing condition. However, since each feature has advantages and disadvantages, it is crucial to develop a systematic method that incorporates all the advantages of the various extracted features, capitalising on the strengths of each, thus making the method more sensitive and robust in defect detection, while at the same time reducing the number of dimensions for CM by using data reduction methods (Malhi and Gao, 2004).

The Bayesian approach develops posterior probability as a result of two antecedents, a prior probability and a likelihood function resulting from a probability model of the observed data. The posterior event (prediction) is directly proportional to the product of the maximum likelihood and the prior event.

1.6.1.3 Damage quantification indexes

Providing a quantifying degradation indication for the assessment of machine performance is vital for diagnostics and prognostics. Some existing studies developed one or more condition monitoring indexes. For instance, Lee *et al*. (2004) used the Hotelling's T^2 and squared prediction error (SPE), also known as the Q-statistic, obtained from PCA and KPCA as a monitoring index in both the simple multivariate process and the simulation benchmark of the biological waste-water treatment process. In general, the SPE performed better than the T^2 in both PCA and KPCA, but the KPCA was more effective in detecting the timing of the simulated fault and in distinguishing between faulty and healthy condition data. Yoo and Lee (2006) also developed an index using the T^2 and SPE from the KPCA-EWMA model, which

captures both the nonlinearity and dynamics (time-varying characteristics) of the system, and they compared the performance with an index from the PCA, KPCA and PCA-EWMA and found that the KPCA-EWMA model was more effective than the others. Similarly, Shu-kai *et al.* (2008) developed an index using the KPCA-EWMA model for monitoring three water tanks. Comparing their results with the traditional PCA and KPCA, they found that the KPCA-EWMA diagnosed better. Sun *et al.* (2013a) built a damage index for monitoring an aero-engine rotor and viscoelastic sandwich structure, using kernel locality preserving projection. Using simulated increasing levels of damage, their results show that the proposed method is effective in identifying the increasing trend in the damage index. Qiu *et al.* (2003) used a method based on a self-organising map and calculated a minimum quantification error, then assessed three rolling bearings based on this. When defining three stages in a run-to-failure test using the U matrix map trained by self-organising map, their results show that the proposed index is better than the RMS at reflecting the damage severity levels. Furthermore, Ma (2012) developed a health index for monitoring the condition of bearings and gearboxes using calculations based on three methods: the Gaussian order statistics, the sum of N condition indicators and normalised energy. Ma (2012) compared the index of these authors with RMS, kurtosis and the crest factor and the result showed clearly the superiority of the new health index over the traditional condition indicators.

Yu (2011a) used linear preserving projections to develop a performance degradation quantification index for bearings. The index was developed by combining the T^2 and SPE statistics obtained from locality preserving projection with EWMA. In addition, Yu (2011b) developed a performance degradation quantification index for bearings by integrating locality preserving projection with Bayesian GMM and EWMA. In both cases the indexes based on the locality preserving projection performed better than the RMS, kurtosis and the likes. Shen *et al.* (2012) also developed a damage severity index for bearings, using fuzzy support vector data description. Based on a run-to-failure test and the definition of thresholds for beginning and final failure, the results show that the proposed damage severity index better reflects the different damage growth than RMS and kurtosis. Dong and Luo (2013) used the first principal

component from a linear PCA as an index for bearing degradation quantification. Though some studies implemented the dimensionality reduction of the original features, and hence reduced the large amount of correlation in the original features, others did not. Some of these studies considered the multimodal nature of the data but not the nonlinearities. Others considered the nonlinear nature of the data but did not consider the multimodal nature of the data. Some studies captured the dynamics of the system, filtered out noise and captured non-stationarities by using EWMA, whereas a number of studies did not. The studies that used a Bayesian approach incorporated prior information and captured uncertainties in the data and parameters, others did not. Moreover, most of these studies were conducted under constant operating conditions. Overall, no single study simultaneously captured the nonlinearities and multimodal distribution of the extracted features, the dynamics of the system, noise filtering, non-stationarity and uncertainties in the data and the parameters of the model. However, it is important that the CM methodology captures the characteristics spelt above so as to give an accurate representation of true condition of the mechanical system being monitored.

1.6.2 Model-based diagnostic methods

Model-based diagnostics is the activity of locating malfunctioning components of a system solely on the basis of its structure and behaviour (Mozetič, 1992). The model based approaches could either be analytical approaches that make use of the quantitative models or the knowledge-based approaches that use qualitative models (Frank *et al*., 2000). The model-based approach uses physics definite, precise mathematical model of the mechanical system to be monitored. Residual data, depicting the presence of a defect is obtained utilizing residual generation techniques such as parameter estimation, Kalman filter, etc. which are based on the accurate model.

Loparo *et al*. (2000) studied a model-based approach in the detection and diagnostics of mechanical faults in rotating machinery. For certain types of faults, for example, raceway faults in rolling element bearings, an increase in mass unbalance, and

changes in stiffness and damping, algorithms suitable for real-time implementation were developed and evaluated using computer simulation. Jalan and Mohanty (2009) used model based technique for fault diagnostics of rotor–bearing system in their study. This was done using the residual generation technique to generate residual vibrations from experimental results for the rotor bearing system subject to misalignment and unbalance, and then the residual forces due to presence of faults were calculated. The residual forces were compared with the equivalent theoretical forces due to faults. The fault condition and location of faults were successfully detected by this model based technique.

The model-based methods are more effective than model-free methods if an accurate model is built. However, this is generally not practicable for complex systems since building their mathematical models are problematic or outrightly impossible (Jardine *et al.*, 2006; Randall, 2011).

1.7 Prognostics

Compared to diagnostics, prognostics literature is sparse. This is because prognostics is a newly evolving field (Randall, 2011). Prognostics render two types of prediction in literature. The commonly utilized prognostics method is the prediction of the time left before the occurrence of failure taken into cognisance the present state of the machine and previous operational history. RUL refers to the period left before the observance of failure. In the literature, RUL is sometimes referred to as residual life, remaining service life, or remnant life.

According to Marble and Morton (2006), bearing prognostics is the key to making the most of safety and asset availability whereas reducing to the minimum logistical costs, by ensuring maintenance to be proactive rather than reactive. However, Jardine *et al*. (2006) note that although prognostics is considerably more capable than diagnostics to attain minimum-downtime performance, diagnostics is necessary when prognostics fails and a fault occurs. Therefore, both diagnostics and prognostics are very important aspects that need to be pursued concurrently. Prognostics ensures that

maintenance is carried out at the most appropriate time once a defect is detected devoid of impairing the requirements of safety, as it is vital for effective operation and management (Camci *et al.,* 2012). The traditional approach of detecting bearing damage and failure (for example manual inspection of defect size after every machine operation) is labour intensive and leads to the shutting down of machinery, thus causing enormous time, production and financial losses (Bolander *et al*., 2009; Camci *et al*., 2012). For instance, the RUL of a bearing with a recently identified fault may be significantly higher than its L_{10} life, that is, the lifespan of 90% bearing population survival (Li *et al*., 1999; 2000). It would therefore be highly beneficial to be able to predict expected remaining bearing life with a large degree of certainty.

Prognostics is an emerging field in mechanical engineering that is gaining interest from both academia and industry resulting in the development of many algorithms for this particular application (Lee *et al*., 2014; Walker *et al*., 2013; Mosallam *et al*., 2013, Camci *et al*., 2013). There is increasing attention on the prognostics or prediction of RUL of both mechanical and non-mechanical structures such as bearings, shafts, gears, and batteries among others. This is not surprising given the huge down times and associated economic losses that often arise due to catastrophic failures. Effective prediction of the expected life of bearings after the onset of an incipient damage is therefore critical for preventive maintenance in all industries that use rotating machineries. Bayesian approaches to prognostics are therefore gaining widespread usage in the CM of bearings as a result of their ability to handle uncertainties well (see paragraph 1.1).

A number of prognostics algorithms have been proposed in the literature. The estimation approaches fall into three major groups: experience-based approaches, data-driven approaches and model-based approaches (Jardine *et al*., 2006; Randall, 2011; Sikorska *et al*., 2011).

1.7.1 Experience-based prognostic methods

In experience-based prognostics, the signal of an experienced feedback is obtained over a very long duration and utilized in the prediction of failure time or RUL. The experienced feedback data collected includes failure times, maintenance and operating data, etc. (Tobon-Mejia *et al*., 2012). The experience-based approach also known as the knowledge-based approach gives results for the average component operating under average conditions (Randall, 2011; Sikorska *et al*., 2011). The main advantage of experience-based methods is the straightforwardness of their computations which are centred on modest functions of reliability such as the Weibull and Exponential, instead of more sophisticated models of mathematics. However, their reliance on historical configurations of degradation to predict future degradation could lead to incorrect estimations during periods of changing operating conditions. Hence, the prognostic results from these methods are not as accurate as obtained from data-driven and model-based methods. This is even more pronounced once the operational circumstances are varying or in the event of different structures as a consequence of the lack of experience data.

Data-driven methods have an edge over both experience-based prognostic approaches and model-based prognostic approaches. Whereas in industry applications obtaining dependable data is easier than building physical models; on the other hand, the generated behavioural models from real condition data leads to more accurate predictive results than those gotten from experience data (Jammu and Kankar, 2011). Table 1.1 gives a summary of some of the experience-based approaches with their merits and limitations.

Table 1.1: Rotating machinery list of experience-based prediction techniques and their advantages and limits

Models	Merits	Limits
Expert systems (Sikorska *et al.,* 2011)	Easy to develop Simple to comprehend	Depends completely on knowledge of subject matter experts Substantial number of rules necessary Substantial knowledge overhead to keep knowledge base up to date Exact inputs required No confidence intervals supplied Not practicable to provide precise RUL
Fuzzy systems (Sikorska *et al*., 2011)	Less rules necessary than for expert systems Inputs can be inaccurate, noisy or incomplete Confidence intervals can be provided on the output with some type of models	Domain experts essential to develop rules

1.7.2 Data-driven prognostic methods

Data-driven methods model system behaviour using regularly collected condition data instead of using comprehensive system physics or human expertise (Heng *et al.,* 2009). Data-driven approaches are classified into two categories in general. These are statistical and machine learning approaches. Statistical approaches construct models by fitting a probabilistic model to the available data. Machine learning approaches attempt to recognize complex patterns and make intelligent decisions based on empirical data. Both statistical and machine learning methods use the degradation patterns of sufficient samples representing equipment failure progression. This requirement is the major challenge in data-driven prognostics since it is often not possible to obtain samples of failure progressions. Industrial systems are not allowed to run until failure due to its consequences especially for critical systems and failure modes. However quality and quantity (sample size) of system monitoring data has a high influence on data-driven methods. Sample sizes of prognostic datasets in the literature range from 10 to 40 (Camci and Chinnam, 2010; Baruah and Chinnam, 2005; Huang *et al.*, 2007; Gebraeel *et al.,* 2005; Eker *et al*., 2011). Finally, it is difficult to define a reasonable failure threshold, especially when limited historical failure data is available (Bolander *et al*., 2009). Table 1.2 gives a summary of some of the data-driven approaches with their merits and limitations.

Table 1.2: Rotating machinery list of data-driven prediction techniques and their advantages and limits

Models	Merits	Limits
Data-driven prognostic models (Heng *et al*., 2009; Randall, 2011)	Assumption or empirical estimation of physics parameters not required	Normally requires a huge amount of data to be precise
Simple trend projection models (Batko, 1984; Kazmierczak, 1983; Cempel, 1987)	Simplicity of calculation	Dependent on past degradation pattern and can lead to inaccurate forecasts in times of change
Time series estimation using ANNs (Tse and Atherton, 1999; Yam *et al.,* 2001; Wang and Vachtsevanos, 2001; Wang *et al.,* 2004; Wang, 2007; Shao and Nezu, 2000)	Fast in handling multivariate analysis Provide nonlinear projection Do not require a priori knowledge	Assume that condition indexes deterministically represent actual asset health Assume that failure occurs once the condition index surpasses a presumed threshold Short prediction horizon
Exponential projection using ANN (Gebraeel *et al.,* 2004)	Predicts actual failure time instead of condition index at future time steps Longer prediction horizon	Assumes that all bearing degradation follow an exponential pattern Needs training one ANN for each historical dataset
Data interpolation using ANN (Huang *et al.*, 2007)	Longer estimation horizon	Needs training one ANN for each historical dataset
Particle filtering (Orchard *et al.*, 2005)	Provision of nonlinear projection	Reduced performance with high dimensional data
Regression analysis and fuzzy logic (Jantunen, 2004)	Highlights the most recent condition information Fuzzy logic enables condition classification based on histories	Does not provide indication of time to failure or probability of failure
Recursive Bayesian technique (Zhang *et al.,* 2007)	Estimates reliability using condition data of individual assets, rather than event data	Accuracy relies strongly on the correct determination of thresholds for several trending features
Hidden Markov Model and Hidden Semi- Markov Model (Zhang *et al.,* 2005; Dong and He, 2007)	Can be trained to recognise different bearing fault types and states	Lack of relation of the defined health-state change point to the actual defect progression since it is often impractical to physically observe a defect in an operating unit. Prognostics projection relies on a failure threshold
Bearing dynamics model using system identification (Li and Shin, 2004)	Tracks defect severity based on features that are not affected by operating condition and nearby equipment	Reasonably accurate only when the signal-to-noise ratio is high, e.g. damage is severe and running speed is relatively high

1.7.3 Model-based prognostic methods

Model-based methods to prognostics entail precise expertise knowhow and theory relevant to the monitored machine. Model or physics-based approaches employ a physical understanding of the system in order to estimate the RUL of an asset. Even though samples of failure degradation are not essential in physics-based prognostics, the physical rules within the system should be known in detail. The first phase in physics-based prognostics is to employ residuals that represent the dispersion of sensed measurements from their expected values of healthy systems (Luo *et al.*, 2003). The second phase in physics-based prognostics requires mathematical modelling of failure degradation.

Physics-based approaches are usually not the most useful solutions in industrial applications as the defect type is usually distinctive from one asset to another and is difficult to be identified without interfering with the operations. Hence, the applicability of this approach is limited in practice. Table 1.3 gives a summary of some of model-based approaches with their merits and limitations.

As earlier stated in paragraph 1.7.1, data-driven methods have an advantage over both experience-based prognostic approaches and model-based prognostic approaches. This is because in industry applications obtaining dependable data (data-driven) is easier than building physical models (model-based); on the other hand, the generated behavioural models (data-driven) from real condition data leads to more accurate predictive results than those gotten from experience data (experience-based) (Jammu and Kankar, 2011).

Table 1.3: Rotating machinery list of model-based prediction techniques and their advantages and limits

Models	Merits	Limits
Model-based prognostics (Heng *et al*., 2009; Randall, 2011)	Is very accurate if physics of the models remains consistent across systems Require less data than data-driven techniques	Real-life system physics is frequently too stochastic and complex to model Defect-specific
Paris law crack growth modelling (Li *et al.*, 1999; Warrier *et al.*, 2000; Li *et al.*, 2000; Li and Lee, 2005; Wemhoff *et al*., 2007)	FEA enables material stress calculation based on bearing geometry, defect size, load and speed.	Performance relies on the correctness of crack size estimate based on vibration data; Computationally expensive.
Paris law modeling with FEA (Li and Choi, 2002; Li and Lee, 2005)	FEA enables material stress calculation based on bearing geometry, defect size, load and speed.	Performance relies on the correctness of crack size estimate based on vibration data; Computationally expensive.
Forman law crack growth modelling (Oppenheimer and Loparo, 2002)	Relays condition data and crack growth physics to life models.	Abridging assumptions need to be examined; Parameters of model yet to be determined for complex conditions e.g. in shaft loading zone and plastic zones.
Initiation of fatigue spall and model progression (Orsagh *et al*., 2003; Orsagh *et al*., 2004; Kacprzynski *et al*., 2004)	Computes the time to spall initiation and the time since spall initiation to failure; Cumulative damage since fitting is estimated with consideration of operating conditions.	Several physics parameters need to be determined.
Contact analysis for bearing prognostics (Marble and Morton, 2006)	FEA enables material stress calculation based on defect size, bearing geometry, speed and load.	Several physics parameters need to be determined; Computationally expensive.
Stiffness-based damage rule model (Qiu *et al.*, 2002, 2003)	Relays bearing component natural frequency and acceleration amplitude to the running time and failure time.	Least-square scheme similar to single-step adaptation in time series estimation; Several material constants need to be determined.

1.8 Prognostics using Bayesian techniques

Bayesian techniques are gaining widespread application in the CM of mechanical systems as result of its several advantages. These include its ability to handle uncertainties which makes them useful for risk analysis and maintenance decision making (Si *et al*., 2011), its ability to train with limited data, avoid overfitting, the use of prior probability which is a potent tool for incorporating information from previous history etc. (Neal, 1996; Titterington, 2004).

In the Bayesian method, the variables of interest (number of neurons, regularisation coefficients, weights, relevance of inputs, neural network outputs, etc.) are modelled as random variables and their prior distributions are always assumed (Hippert and Taylor, 2010). The Bayes theorem is subsequently used to derive the posterior distributions, once data is available. Bayesian methods have the ability to not only obtain point estimates for the variables of interest but also their probability distributions, which permits the researcher to compute the uncertainty using confidence intervals (CIs). The relevance of the inputs can be evaluated after the training using the automatic relevance determination (ARD) and the optimum number of neurons can be established by simply comparing the Bayesian evidence of the models. The use of Bayesian methods for neutral networks is well documented in the literature (see for e.g. Bishop, 1995; Thodberg, 1996; Penny and Roberts, 1999; Lampinen and Vehtari, 2001; Titterington, 2004; Hippert and Taylor, 2010).

The use of Bayesian theory for neural networks was introduced by Buntine and Weigend (1991). A number of Bayesian techniques have been used in implementing neural networks. These include the evidence approximation method, whereby the posterior distribution of the network weights are approximated as a Gaussian distibution. This technique simplifies the mathematical treatment, hence allows the derivation of expressions for estimating the most probable values of the hyperparameters, as well as the most probable model. Latest advances in Markov chain Monte Carlo method enable the application of Bayesian approaches to condition data. The evidence technique was introduced by Mackay (1992a, b) and has

also been implemented by a number of authors including Nabney (2002) and Hippert and Taylor (2010) amongst others. Several authors have however attempted to avoid the approximation by integrating the posterior with other techniques such as Markov chain Monte Carlo (Neal, 1996; Barber and Bishop, 1998; Lampinen and Vehtari, 2001; Nabney, 2002), or with techniques that use neither Gaussian approximations or Markov chain Monte Carlo: the variational method (Titterington, 2004), or with hybrids between Markov chain Monte Carlo and genetic algorithms (Chua and Goh, 2003; Liang, 2005), and the Bayesian conjugate prior method (Vila *et al*., 2000; Rossi and Vila, 2006) as pointed out by Hippert and Taylor (2010).

A few studies used the Bayesian approach in the CM of bearings. Chebil *et al.* (2009) identified bearing fault location and size from results obtained with the root mean square extracted from the terminal nodes of a wavelet tree of Symlet basis functions fed to a Bayesian classifier. Gebraeel *et al*. (2005) developed a Bayesian updating method that uses real-time CM information to update the stochastic parameters of exponential degradation models. Chen *et al.,* (2012) proposed a novel approach for machine health condition prognostics based on neuro-fuzzy systems and Bayesian algorithms. The neuro-fuzzy systems, after training with machine condition data, is employed as a prognostic model to forecast the evolution of the machine fault state with time. Di Maio *et al.,* (2011) estimated the RUL of degraded thrust ball bearings using a data-driven stochastic approach that relies on an iterative Naïve Bayesian Classifier for regression task. Other studies (Zhang *et al.,* 2001; Gebraeel and Lawley, 2008; Zhang *et al.,* 2007; Chen *et al.,* 2008; Hazan *et al*., 2011) also used Bayesian based techniques for bearing fault detection.

Some of the widely used Bayesian techniques includes the extension of the Bayesian method to the class of widely used ANN such as the Multi-layer perceptron (MLP) and Radial basis function (RBF) (Nabney, 2002). Also, Bayesian linear regression (BLR), Gaussian mixture regression (GMR) and Gaussian process regression (GPR) are other classes of Bayesian models that have gained wide acceptance. The use of Bayesian methods for neural networks requires more explanation given that this class of models have been implemented using the traditional approaches such as maximum

likelihood methods, cross validation (CV) for selection of variables of interest etc. Artificial neural networks are flexible models widely used for classification and regression problems in machine learning regression. However challenges remain especially when training data is limited. Bayesian methods avoid overfitting (too much variance relative to bias) commonly observed with traditional learning methods for complex neural network models (Neal, 1996; Titterington, 2004).

Predicting the RUL of bearings is surrounded by uncertainties which include randomness of bearing failure time, differences or changes in operating conditions, process noise, measurement noise, inaccurate process modes amongst others (Goebel *et al*., 2008; Hong and Zhou, 2012). These make it difficult for some models to track bearing features. However, given the ability of Bayesian models to reduce uncertainties both in the data and parameter estimates, this study is based on the use of Bayesian approach in the diagnostics and prognostics of slow rotating bearings. Some of the commonly used Bayesian methods are discussed in the following paragraphs.

1.8.1 Multi-layer perceptron regression

The multilayer perceptron is one of the most preferred feed forward artificial neural networks which make use of a supervised learning algorithm. Essentially, it has three layers which include the input layer, hidden layer and output layer (Şengüler *et al.,* 2010). It can be used in a Bayesian framework for RUL prediction.

1.8.2 Radial basis function regression

The RBF is used for nonlinear modelling. The RBF has many advantages, one of which is that it has a two stage training procedure which is considerably faster than MLP. The basis function parameters are fixed to model the unconditional data density in the first step. The second step of training entails the determination of the weights in the output layer. Secondly, it is possible to assign an interpretation to the hidden units

and also to determine the intrinsic degrees of freedom of the network (Nabney, 2002). It can also be used in a Bayesian framework for RUL prediction.

1.8.3 Bayesian linear regression

The parametric approach focuses on the use of probability distributions having specific functional forms governed by a small number of adaptive parameters, such as the mean and variance whose values are to be determined from the data set. The probability distributions include beta (binomial) and Dirichlet (multinomial) distributions for discrete random variables and the Gaussian distribution and Gaussian mixture distribution for continuous variables (Bishop, 2006). The Gaussian, also known as the normal distribution, is a widely used model for the distribution of continuous variables (Bishop, 2006).

1.8.4 Gaussian mixture regression

The standard single Gaussian distribution possesses vital analytical characteristics. Nonetheless, it has some noteworthy limitations in the modelling of real data. If a dataset forms more than one dominant clump, the simple Gaussian distribution is incapable of capturing the structure. However a linear superposition of two or more Gaussians can lead to a better characterization of the dataset. Such linear characterisation formed by taking linear combination of more basic distributions such as Gaussians, can be formulated as a probabilistic model known as mixture distribution (Bishop, 2006).

1.8.5 Gaussian process regression

Gaussian processes (GPes) are a recent development in nonlinear nonparametric modelling. In GP, the parametric model is dispensed and instead a prior probability distribution is defined over functions directly (Bishop, 2006). GP is utilized for the modelling of nonlinear functional mappings to a target space from an input space. The GP is an infinite assemblage of random variables of which the finite subsets have

joint Gaussian distributions. The Gaussian process helps smooth functions elucidate the training data clearly. The features of the smooth functions enable them to make good generalisations (Heyns *et al.,* 2012a).

The standard or basic GPR widely used by most studies assumes a zero mean function and squared exponential (SE) covariance function (Rasmussen and Williams, 2006; Lázaro-Gredilla *et al*., 2010). As acceptable as the results from the standard GPR as highlighted above might be, imposing a priori a specific mean function and or covariance function may be quite restrictive given the differing properties of datasets which includes long term trend, pronounced seasonal variations and smaller or medium irregularities (Rasmussen and Williams, 2006). When two or more of such properties exist as is the case with the data used in this study, a simple mean or covariance function may not capture the different trends well. This may then require a more complex function that takes care of these individual properties. Moreover, when the test data is "distant" from the training data, the extrapolation performance of the basic GPR deteriorates rapidly (Shi and Wang, 2007).

However, it is possible to obtain more desirable properties by allowing for more flexibility in the Gaussian process modelling. This can be achieved by constructing composite functions from the existing simple functions (Bishop, 2006). According to Rasmussen and Nickisch (2013), composite functions can be composed of either a combination of two or more simple functions or a combination of two or more composite functions, thus permitting for exceptionally exciting and flexible structures. The idea of combination of functions is relatively similar to that in the time series forecasting literature. For instance, Palm and Zellner (1992) noted that "in many situations a simple average of forecasts will achieve a substantial reduction in variance and bias through averaging out individual bias".

The ability of the Gaussian process to avoid simple parametric assumptions and still build in a lot of structure makes it a very attractive model in many application domains (Rasmussen and Williams, 2006). Rasmussen (1996) examined several cases in which GPR generalizes much better and sometimes does considerably better than

other methods of regression, particularly when the availability of sufficient training data is an issue. A practical investigation by Wang *et al.* (2008) likewise demonstrated that one of the most important advantages of the Gaussian process models is their ability to generalize small sets of data well. These results were further confirmed by Heyns *et al.* (2012a) who found that the GPR performed better than the polynomial and neural network models.

Although, there is good justification for the use of SE covariance functions in many settings (Rasmussen and Williams, 2006), it is still important to investigate very well which covariance functions are best suited for specific data. Basically, the selection among alternative covariance functions is a means of reflecting various forms of prior knowledge against the process under investigation (Ebden, 2008). This also applies to selection from alternative mean functions.

Some studies (Hong and Zhou, 2012a; Hong and Zhou, 2012b; Chen and Ren, 2009) have empirically investigated the advantage of composite covariance functions over the simple covariance functions for predicting RUL. However, these studies assumed zero mean function. The assumption of a zero mean implies that the mean of the function vanishes away and therefore does not make an impact in the prediction output. In this case, the prediction is only based on the covariance function. However, prediction maybe improved by considering and selecting the optimal mean and covariance functions simultaneously.

1.8.6 Model evaluation and comparison

A number of studies have employed at least one of these models (MLP, RBF, BLR, GMR and GPR) for prognostics. Examples include Nabney, (2002); Gebraeel *et al.*, (2004); Rasmussen and Williams, (2006); Skabar, (2007); Chen and Ren, (2009); Saxena *et al.*, (2009); Hippert, and Taylor, (2010); Wang and Wang, (2012); Hong and Zhou, (2012a), Hong and Zhou, (2012b); Liu *et al.*, (2012); Calinon, (2009); Chatzis *et al.* (2012) amongst others. However, none of these studies compared MLP, RBF, BLR, GMR and GPR to obtain the best performing model for RUL predictions.

Other studies, for instance, Hong and Zhou (2012b) evaluated the performance of GPR and wavelet neural network for prediction of bearing RUL and found GPR to show more excellent features than wavelet neural network with GPR predicting faster and having more stable prediction as well as lower prediction error (PE) in general than wavelet neural network. Goebel *et al*. (2008) compared the performance of GPR, relevance vector machine and NN-based approach for prognostics of aerospace rotating equipment. GPR seemed to have performed better than NN and relevance vector machine especially for the algorithm with specific damage estimates given the small GPR error though with late predictions than relevance vector machine. Saha *et al*. (2009) evaluated the performance of particle filter-based, autoregressive integrated moving average and extended kalman filter models and found that the particle filter framework has significant advantages over autoregressive integrated moving average and extended kalman filter for predicting the RUL of batteries. An *et al*. (2012) compared the performance of the particle filter, the overall Bayesian method, and the recursive Bayesian method and found that the performance of particle filter and overall Bayesian method differ depending on the stage of the damage state and hence should be used as complementary models. Chatzis *et al*. (2012) compared the performance of the standard GMR, Dirichlet process GMR and GPR and found the first two to be computationally less expensive for robot prognostics than GPR. An *et al*. (2013) compared NN and GPR under different levels of noise and found that GPR under a no noise case (perfect data) or small noise outperform NN while under large noise NN outperform GPR.

1.8.7 Prognostics based on dependent and independent samples

Prognostics could be based on dependent or independent samples or observations. Prognostics based on dependent observations refers to predictions where the training and the test sets are obtained using a leave-one-out CV technique that involves the division of the data set into equal samples of training and test sets. When the training and test samples are dependent, the two errors may be positively correlated, resulting in a breakdown of the CV selection approach and can equally lead to model

overfitting (Opsomer *et al*., 2001; Arlot, 2010). However, the prognostics based on independent observations refers to predictions whereby one or more different data set(s) are trained together and are hence used as the training set while the remaining data set is used as the test set.

1.9 Effects of operating conditions

Machineries are subjected to varying operational conditions in everyday conditions. The changing conditions are the main influence of variation in the energy of measured condition data. The effect of varying operating conditions in prognostics is relatively new. An example is when the operational speed of a machine increases leading to a rise in the amplitude. A prognostics model might recognise this increase in magnitude as an increase in the degradation of the asset being monitored. On the contrary, if the operational speed is reduced while a monitored asset is failing, a prognostics model might depict this as a reduction in the danger of machinery catastrophe despite the fact there is impeding failure (Heng *et al.,* 2009).

1.10 Scope of the work

The literature survey covered a broad range of topics relating to diagnostics and prognostics of rotating machinery with particular attention to slow rotating bearings. From the literature survey the scope of the work is outlined in the subsequent paragraphs.

1.10.1 Diagnostics based on a developed novel DAI

This study uses statistical data-driven Bayesian methods to develop a novel integrated methodology for slow rotating bearing fault diagnostics based on AE data obtained from run-to-failure experiments under varying operating conditions. The proposed model is capable of accounting for data dimensionality reduction and hence for the reduction of high feature correlation, nonlinearities, noise filtering, non-stationarities, uncertainties, time variation (dynamics) and multimodal distribution in the data,

under varying operating conditions. A DAI which accounts for these characteristics for the assessment of machine performance is vital for effective diagnostics.

The early identification of the presence and severity of bearing damage provides an important input, leading to an adjustment of the maintenance schedule and in this way minimising machinery downtime. In this case, bearing degradation cannot be assessed through supervised-learning models, because of the non-availability of prior information about the severity of the defect at various stages of propagation. Instead, a novel DAI is proposed for the classification of slow rotating bearing defect and severity, based on the assumption that only healthy bearing data are available. In this investigation, healthy bearings were run constantly to failure. In several applications, the operational condition of bearings (e.g. speed and load) could change as a result of excessive running operations. As a consequence, the healthy bearing data sample could result in multimodal and/or nonlinear distributions.

The broad objective of the study is to develop novel effective CM techniques for slow rotating bearings based on AE data. The following specific objectives also form part of the contributions of the study:

Firstly, a novel approach is proposed. This approach entails features selection based on the polynomial KPCA (PKPCA) which is a nonlinear data reduction approach. The fusion of information from different AE extracted features, capitalising on the strengths of each, is expected to result in the more sensitive and robust detection of defects. K-means are used to classify the defect into different health states so as to assess the usefulness of the extracted features from PKPCA. The multimodal features in the low-dimensional data space can still be conserved by the extracted features that PKPCA obtains from the high-dimensional data.

Gaussian mixture model (GMM) is an outstanding technique for handling the description of multimodal data, making it robust with high computational efficiency. Therefore, secondly, using the features extracted by PKPCA, the negative log likelihood (NLL) is obtained by using the Bayesian GMM. The GMM uses multiple

Gaussian components for describing the multimodal data distribution. This characteristic makes it very easy for GMM to render a smooth estimation of the AE signal variability of healthy bearings. This algorithm is based on unsupervised learning without specific failure-class labels, improving its usefulness in various applications in the real world. In other words, humans do not need to intervene and knowledge is not needed about the input data characteristics in the course of modelling. The only requirement is the setting of values for a few processing parameters.

Thirdly, the exponentially weighted moving average (EWMA) model is used for improving the sensitivity and dependability of the NLL with regard to the degradation of the slow rotating bearing. Finally, a novel DAI is obtained from the integration of PKPCA, the GMM and the EWMA model for the diagnostics of the slow rotating bearings. Hence, the DAI is proposed for the detection of bearing faults and the classification of fault severity, based on the assumption that measurements are necessary during the healthy operating conditions of the bearing. Therefore, in this study DAI was used for the damage assessment of slow rotating bearings.

A number of studies, as listed previously (see paragraph 1.6.1.3), have developed CM indexes under different names, but to the best of our knowledge, no previous study used the methodology proposed in this study for building a CM index. Also, no previous index has been able to capture all the features of the developed DAI. Hence, the proposed approach is considered a novel contribution to the literature. Finally, the DAI is used in the prognostics of slow rotating bearings using various approaches.

1.10.2 Prognostics using various approaches

A novel approach by integrating a newly developed DAI is used as an input in several regression models such as the MLP, RBF, BLR, GMR and GPR for RUL prediction. Secondly, the mean absolute percentage error (MAPE) and root mean square error (RMSE) were used in model evaluation to select the best performing model. There is no known study that has evaluated the performance of these set of models using the

same data set. Thirdly, the best performing model from the resulting methodology is used for the prediction of slow rotating bearing RUL.

Further, this study also makes a contribution to the prognostics literature by evaluating the performance of these models under a leave-out-one CV approach that is based on two types of samples or data set namely dependent and independent samples (see paragraph 1.8.7).

1.10.3 Prognostics based on an integrated GPR model

An integrated GPR model for prediction of RUL of slow rotating bearings is proposed. The proposed novel model combines the advantages from the individual mean and covariance functions. Moreover, the proposed technique consists of a construction of composite mean and composite covariance functions. These are respectively formed from two simple mean functions and three covariance functions selected based on acceptable scientific criteria rather than ad hoc choice as the latter may lead to model misspecification. This procedure ensures that only the mean and covariance functions which are well-suited for the problem at hand are combined in order to obtain an optimal GPR model for prognostics of slow rotating bearings. This study deviates from previous studies in that most studies implement a GPR using only the covariance function while assuming the mean function is equal zero.

Hence, another contribution of this research entails the implementation of a novel integrated GPR-based model and the scientific selection of a composite mean function and composite covariance function for a more flexible and accurate prediction of RUL estimates of slow rotating bearings. Even though there is good rationalization for the use of SE covariance function in many settings (Rasmussen and Williams, 2006), it is still vital to examine carefully which of the covariance functions is the most suitable for specific data.

1.11 Document outline

Chapter 1 gives the problem statement and literature review and scope of the study. The next chapter uses Bayesian techniques in the development of a novel DAI for diagnostics of slow rotating bearings. Subsequently, in chapter 3, the DAI is integrated into some Bayesian prognostics models for the prediction of RUL of slow rotating bearings. In chapter 4, the best performing prognostics model is further improved upon to obtain more accurate RUL predictions. Finally, chapter 5 presents the conclusions and recommendations of the study.

Chapter 2 Diagnostics of slow rotating bearings using AE

2.1 Introduction

This chapter uses statistical data-driven Bayesian methods in the development of a novel degradation assessment index (DAI) for slow rotating bearing fault diagnostics, based on acoustic emission (AE) data obtained from run-to-failure experiments. The proposed model has the capability for data dimensionality reduction and therefore for the reduction of high feature correlation, nonlinearities, noise filtering, non-stationarities, uncertainties, time variation (dynamics) and multimodal distribution in the data, under varying operating conditions. The DAI is proposed for the classification of slow rotating bearing defects and severity, based on the assumption that only healthy bearing data are available.

Firstly, the proposed approach features selection based on the nonlinear polynomial kernel principal component analysis (PKPCA) data reduction approach. The defect is classified into different health states by the use of K-means so as to assess the effectiveness of the extracted features from PKPCA. Secondly, using the extracted PKPCA features, the negative log likelihood (NLL) is obtained by using the Bayesian Gaussian mixture model (GMM). Thirdly, the exponentially weighted mean average (EWMA) model is used for improvement of the sensitivity and dependability of the NLL with regard to the degradation of the slow rotating bearing. Finally, a novel DAI is obtained from an integration of PKPCA, the GMM and the EWMA model for the diagnostics of the slow rotating bearings.

2.2 Methodology

The scheme used for developing the proposed DAI for the diagnostics of slow rotating bearings is shown in Figure 2.1. Generally the scheme involves five main steps: (1) feature extraction from AE data; (2) reducing the data dimensionality by

using PKPCA; (3) obtaining the NLL by using GMMs; (4) smoothening the NLL by means of the EWMA to obtain the DAI; (5) evaluating the proposed DAI by using benchmark studies.

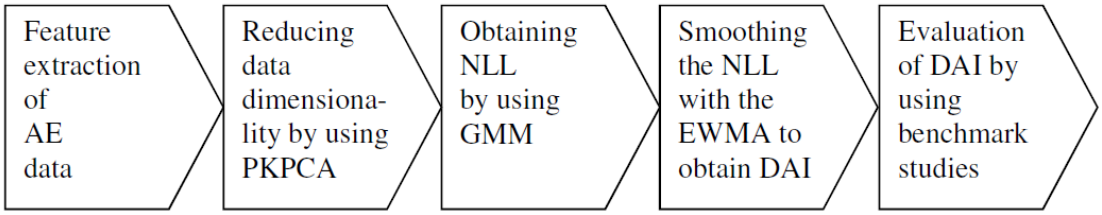

Figure 2.1: Schematics for developing DAI for diagnostics of slow rotating bearings

2.3 Feature extraction

Several features such as kurtosis, root mean square (RMS), peak-to-peak, crest factor and skewness were extracted from the time domain analysis of AE data. These high dimensional features were subsequently reduced to low dimension with the principal components still capturing all the features of the previous features (see chapter 1, paragraphs 1.4.1 and 1.6.1.1).

2.4 Kernel principal component analysis

High dimensional data is difficult to deal with. This problem can be solved by reducing the data into fewer principal components by using data reduction techniques. The traditional principal component analysis (PCA) is a widely used procedure for dimensionality reduction and feature extraction in machine learning. However, this study proposes a model based on kernel principal component analysis (KPCA). The main idea of KPCA (Schölkopf *et al*., 1998; Schölkopf *et al*., 1999) is both intuitive and generic. Generally, the standard PCA can only be effectively performed on a set of observations that vary linearly. PCA becomes invalid for nonlinear variations, as can be seen in the extracted features. The reasons for this nonlinearity are the highly stochastic nature of the data, the variable operational conditions and the changing condition of the slow rotating bearings. Unlike other nonlinear methods, such as neural networks, KPCA does not involve a nonlinear optimisation procedure. Instead,

KPCA maps the problem from the input space to a new higher-dimensional space (called feature space) by doing a nonlinear transformation using suitably chosen basis (kernel) functions, and then using a linear model in the feature space (Lee, 2004). As a result, KPCA performs a nonlinear PCA in the input space (Romdhani *et al.*, 1999). This is popularly known as the "kernel trick". Kernel methods permit the dot product of two vectors and in the feature space to be calculated as a function of corresponding vectors. Given a set of nonlinear data with a mean of zero, the covariance matrix in the feature space can be expressed as (Schölkopf *et al.*, 1998; Schölkopf *et al.*, 1999):

$$\mathbf{C}^F = \frac{1}{N}\sum_{j=1}^{N}\varphi(\mathbf{x}_j)\varphi(\mathbf{x}_j)^T \tag{2.1}$$

where it is assumed that the mapped data in the feature space are centred, i.e., $\sum_{k=1}^{N}\varphi(\mathbf{x}_k)=0$, and $\varphi(\cdot)$ is a nonlinear mapping function that projects the input vectors from the input space to F. The diagonalization of the covariance matrix requires solving the eigenvalue problem in the feature space

$$\lambda v = \mathbf{C}^F\mathbf{v} \tag{2.2}$$

where $\lambda \geq 0$ and $\mathbf{v}$ represent the eigenvalue and eigenvector, respectively. The $\mathbf{v}$ along with the largest λ obtained by Equation (2.2) becomes the first principal component (PC) in F, and the $\mathbf{v}$ along with the smallest λ becomes the last PC. Here, $\mathbf{C}^F\mathbf{v}$ can be expressed as follows:

$$\mathbf{C}^F\mathbf{v} = \left(\frac{1}{N}\sum_{j=1}^{N}\varphi(\mathbf{x}_j)\varphi(\mathbf{x}_j)^T\right)\mathbf{v} \tag{2.3}$$

Because all solutions $\mathbf{v}$ with $\lambda \neq 0$ must lie in the span of $\varphi(\mathbf{x}_1),...,\varphi(\mathbf{x}_N)$, Equation (2.2) is equivalent to

$$\lambda\langle\varphi(\mathbf{x}_k),\mathbf{v}\rangle = \langle\varphi(\mathbf{x}_k),\mathbf{C}^F\mathbf{v}\rangle \quad k=1,...,N \tag{2.4}$$

and there exist coefficients $\alpha_i (i = 1,...,N)$ such that

$$\mathbf{v} = \sum_{i=1}^{N} \alpha_i \varphi(\mathbf{x}_i) \tag{2.5}$$

Combining Equations (2.4) and (2.5), yields

$$\lambda \sum_{i=1}^{N} \alpha_i \langle \varphi(\mathbf{x}_k), \varphi(\mathbf{x}_i) \rangle = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \left\langle \varphi(\mathbf{x}_k), \sum_{j=1}^{N} \varphi(\mathbf{x}_j) \right\rangle \left\langle \varphi(\mathbf{x}_j), \varphi(\mathbf{x}_i) \right\rangle \tag{2.6}$$

for all $k = 1,...,N$. From Equation (2.6), it is clear that only the computations of the dot products of the mapped vectors in the feature space are required, which can easily be done through the kernel function (Sun *et al.*, 2013b). To obtain coefficients, α_i, define an $N \times N$ kernel matrix $\mathbf{K}$ by its ij^{th} element K_{ij},

$$K_{ij} = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle \tag{2.7}$$

Then, Equation (2.6) can be rewritten as

$$\lambda \sum_{i=1}^{N} \alpha_i K_{ki} = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \sum_{j=1}^{N} \mathbf{K}_{kj} \mathbf{K}_{ji} \tag{2.8}$$

for all $k = 1,...,N$. This yields

$$\lambda \mathbf{K} \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K}^2 \boldsymbol{\alpha} \quad j = 1,...,M \tag{2.9}$$

where $\boldsymbol{\alpha} = [\alpha_1,...,\alpha_N]^T$. The solutions of Equation (2.9), can be found through solving the eigenvalue problem

$$\lambda \boldsymbol{\alpha} = \frac{1}{N} \mathbf{K} \boldsymbol{\alpha} \tag{2.10}$$

for nonzero eigenvalues. Performing PCA in F is equivalent to resolving the eigenvalue problem of Equation (2.10). This yields eigenvectors $\alpha_1, \alpha_2,..., \alpha_N$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$. The dimensionality of the problem can be reduced by

retaining only the first p eigenvectors. Normalizing $\alpha_1, \alpha_2, ..., \alpha_p$ through scaling the corresponding eigenvector by factor $1/\sqrt{\lambda_k}$ i.e.

$$\langle \mathbf{v}_k, \mathbf{v}_k \rangle = 1 \text{ for all } k = 1, ..., p \tag{2.11}$$

where p is the number of principal components retained. The kernel PCs are then extracted by projecting $\varphi(\mathbf{x})$ onto eigenvectors v_k in F,

$$y_i(\mathbf{x}) = \langle \mathbf{v}_k, \varphi(\mathbf{x}) \rangle = \sum_{i=1}^{N} \alpha_i^k \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}) \rangle \tag{2.12}$$

where α_i^k is the i^{th} element of the eigenvector in Equation (2.10); $k = 1, ..., P$

It is assumed that the mapped or projected data are centred (i.e. have zero mean), $\sum_{k=1}^{N} \varphi(\mathbf{x}_k) = 0$, otherwise this can be realized by substituting the kernel matrix $\mathbf{K}$ with the Gram matrix

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N \tag{2.13}$$

where

$$\mathbf{1}_N = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \in R^{N \times N}$$

That is, $\mathbf{1}_N$ is the $N \times N$ matrix with all elements equal to $1/N$ (Bishop, 2006).

The power of kernel methods is that there is no need to explicitly compute or carry out the nonlinear mapping $\varphi(\cdot)$. The kernel matrix can be directly constructed from the training data set (Weinberger *et al.*, 2004; Wang, 2012). To solve the eigenvalue problem of Equation (2.10) and to project from the input space into the PKPCA space using Equation (2.12), one can avoid performing the nonlinear mappings and

computing both the dot products in the feature space by introducing a kernel function of form (Schölkopf *et al*., 1998; Schölkopf *et al*., 1999; Romdhani *et al*., 1999):

$$k(x, y) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle \tag{2.14}$$

There are a number of representative kernel functions, such as the linear, polynomial, sigmoid, Gaussian or radial basis kernels (Lee *et al*., 2004; Weinberger *et al*., 2004; Wang 2012). According to Mercer's theorem of functional analysis, there is mapping into a space where a kernel function acts as a dot product if the kernel function is a continuous kernel of a positive integral operator. The polynomial and the Gaussian kernels always satisfy Mercer's theorem (hence they are the most commonly used), whereas the sigmoid kernel satisfies it only for certain parameter values (Haykin, 1999; Lee, 2004; Wang, 2012). In this chapter, the Gaussian, simple and polynomial kernel are used and given as Equations (2.15), (2.16) and (2.17)

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2} \right) \tag{2.15}$$

$$k(x, y) = (x, y) \tag{2.16}$$

$$k(x, y) = (x, y)^d \tag{2.17}$$

with parameters σ and d

The standard steps of polynomial kernel PCA dimensionality reduction can be summarized as:

(1) Extract the relevant features (kurtosis, RMS, peak-to-peak, crest factor and skewness) from the AE signal and normalize the features using the mean and standard deviation of each feature.
(2) Compute the kernel matrix $\mathbf{K}$ using Equation (2.7).

(3) Carry out centering in the feature space for $\sum_{k=1}^{N}\varphi(\mathbf{x}_k)=0$, by computing the Gram matrix, $\tilde{\mathbf{K}}$ using Equation (2.13).

(4) Use Equation (2.10) to solve for the eigenvectors (substitute $\mathbf{K}$ with $\tilde{\mathbf{K}}$).

(5) Compute the polynomial kernel principal components $y_i(\mathbf{x})$ using Equation (2.12).

The extracted polynomial kernel PCs, henceforth, PKPC, is used as inputs in the GMM for further analysis.

However, if one intends to use the PKPCs for monitoring the condition of bearings without taking the GMM route, one would need some monitoring statistics. The two widely used monitoring statistics from KPCA-based models are the T^2 and Q-statistic or squared prediction error (SPE). The T^2 statistic accounts for the variation within the PKPCA model, whereas the SPE accounts for the variation not captured by the PKPCA model or, simply put, the SPE is a measure of the goodness of fit of a sample to the PKPCA model (Lee *et al.*, 2004; Yoo *et al.*, 2006; Sun *et al.*, 2013b). The T^2 is the sum of the normalised squared scores and is given as (Lee *et al.*, 2004):

$$T^2=[y_1,...,y_p]\Lambda^{-1}[y_1,...,y_p]^{\mathrm{T}} \tag{2.18}$$

where y_k is obtained from Equation (2.12) and is the diagonal matric matrix of the inverse of the eigenvalues associated with the retained principal components. The SPE statistic is given as (Lee *et al.*, 2004):

$$SPE=\left\|\varphi(\mathbf{x})-\hat{\varphi}_p(\mathbf{x})\right\|^2 \tag{2.19}$$

where $\hat{\varphi}_p(\mathbf{x})=\sum_{k=1}^{p}y_k\mathbf{v}_k$ is the reconstructed feature vector with p principal components in the feature space.

The confidence limits for the T^2 and SPE statistics for nonlinear methods are determined by using the kernel density estimation (KDE) (Martin and Morris, 1996) with a CI of 99.7%. This in turn was used for defining the thresholds.

2.5 Gaussian mixture model

GMM is an outstanding technique for the description of multimodal data as this technique is robust with high computational efficiency. Using several Gaussian components enables the GMM to describe multimodal data distribution. This characteristic makes it feasible for GMMs to have a smooth estimate of the AE signal variability of healthy bearings. This algorithm is based on unsupervised learning without various failure class labels, which improves its usefulness in various applications in the real world.

The distribution of feature vectors extracted by PKPCA is modelled in a GMM-based approach by a weighted sum of M mixture components. A finite mixture model can be defined as (Bishop, 2006):

$$p(x) = \sum_{j=1}^{M} \pi_j \, p(x|\theta_j) \tag{2.20}$$

where $x = [x_1, ..., x_l]^T$ is the l-dimensional data vector. Each component density $p(x|\theta_j)$ is a normal probability distribution, $N(x|\theta_j) \geq 0$, with parameters θ_j composing of the mean vector, μ_j, and covariance matrix, Σ_j, i.e. $\theta_j = [\mu_j, \Sigma_j]$. In order to provide valid probabilities, the mixing coefficients $\{\pi_j\}$ must satisfy $0 \leq \pi_j \leq 1$ (2.21)

together with

$$\sum_{j=1}^{M} \pi_j = 1 \tag{2.22}$$

If these constraints are satisfied, the resulting mixture model will also be a valid density function.

Condensing these parameters into a parameter vector $\varphi = [\pi_1,...,\pi_M;\theta_1,...\theta_M]$ and given training data set $X = \{x^{(1)},...,x^{(n)}\}$ with n independent and identically distributed samples of the random variable x, learning aims at finding the number of components M and the optimum vector $\varphi^* = [\pi_1^*,...,\pi_M{}^*;\theta_1^*,...\theta_M{}^*]$ that maximizes the likelihood function

$$\log(p(X|\varphi)) = \log\prod_{i=1}^{n} p(x^{(i)}|\varphi) \tag{2.23}$$

The method for determining the parameters of a GMM is based on maximising the data likelihood. It is convenient to recast the problem in the equivalent form of minimising the NLL of the data set which is treated as an error function (Nabney, 2002):

$$NLL = E = -\sum_{n=1}^{N} \log p\left(x^n|\varphi\right) \tag{2.24}$$

The parameters in φ are learned by maximising the likelihood function using the expectation maximisation algorithm. Some advantages of using the expectation maximisation algorithm include: it is simple to implement and understand, avoids the calculation and storage of derivatives, it is usually faster to converge than general purpose algorithms and can also be extended to deal with data sets where some points have missing values (Ghahramani and Jordan, 1994). The expectation maximisation algorithm includes two steps: the expectation step and maximisation step which are alternatively iterated until the NLL converges to a local optimum. The results obtained from GMM may be highly sensitive to the number of mixing components used. The greater the number of components in a mixture model, the greater its expressiveness and flexibility. A sufficiently expressive model may be optimised so as to accurately represent the reference signal. However, models which are too expressive may over-fit the training data. This may result in poor generalisation and subsequently impair the ability of the model to distinguish between healthy signal components and fault-related outliers (Bishop, 2006; Heyns *et al.*, 2012b). Different

numbers of mixing components are fitted and the best selected. There are several model selection criteria such as the RMSE, the leave-one-out cross validation (CV) technique, the Bayesian information criterion, Bayesian model selection and the Akaike information criterion. In this study, the leave-one-out CV technique was used to select the best number of mixing components.

The slow rotating bearing health states are monitored online by using the constructed NLL. The quantification criterion is needed to evaluate whether or not a new input is healthy. For each new input, GMM provides the unconditional probability density $p(x)$, which indicates how the input follows the probability distribution of the GMM trained by a healthy dataset. The GMM outputs corresponding to the novel data should be sufficiently smaller than the outputs of the GMM for healthy data, and below a certain threshold. By contrast, an input from the same region in input space as the training data should result in a probability density value that will be equal or greater than the threshold (Yu, 2011b). Accordingly, new inputs with smaller NLL values correspond to the expected behaviour, whereas larger values indicate the possible presence of damage.

2.6 Exponentially weighted moving average

The EWMA model is used to improve the sensitivity and dependability of the NLL to a slight degradation in the slow rotating bearing. The EWMA statistic is therefore proposed as an improved quantification index for the diagnostics of slow rotating bearings. EWMA is a type of infinite impulse response filter that applies weighting factors which decrease exponentially. The weighting for each older data point decreases exponentially, never reaching zero. The EWMA can be obtained as follows

$$W_t = (1-\alpha)W_{t-1} + \alpha NLL_t \tag{2.25}$$

where α is a smoothening constant between 0 and 1 and W_t is the average of the preliminary data. A large value of α attaches more weight to the current observation

than to historical observations. In this study, a value of 0.2 was used for α. This is in line with the general recommendation by Lucas (1990) and is also capable of providing a good balance between information from historical and current observations. Finally, the DAI, was used as an indication of the degradation of the slow rotating bearing. Hence, the DAIs on the time series may be plotted for the whole life of each bearing, to form a monitoring chart for each of the slow rotating bearings. A threshold is needed for the DAI in order to trigger the alarm for the onset of slight bearing degradation. The KDE is used to define the healthy operating region of the DAI, based on the 99.7% level of confidence (Martin and Morris, 1996).

The maximum a posteriori estimates $\bar{y}_*$ will be used for slow rotating bearing damage detection and the prognostics for the RUL of the bearing. The entire condition monitoring (CM) process is implemented in a unified framework (Figure 2.2).

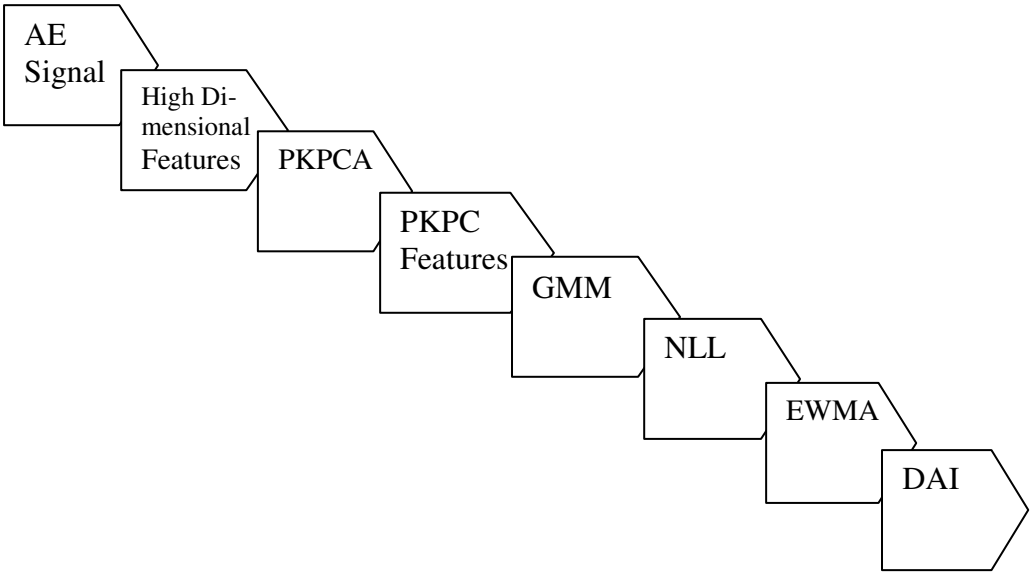

Figure 2.2: Framework for PKPCA–GMM–EWMA integrated approach to bearing diagnostics

2.7 Evaluation of the proposed DAI using benchmark studies

The effectiveness of the proposed DAI is evaluated by comparing its performance with that of other monitoring indexes. To this end, the proposed model is reordered to form three other different submodels which essentially correspond to some of the

models that may have been used in other studies for developing condition monitoring indexes. In other words, each submodel misses at least one component that is utilised in the proposed model. The models consist of the PKPCA-GMM-EWMA model, the PKPCA-EWMA model, the PKPCA-GMM model and the GMM-EWMA model.

2.8 Experimental setup

An experimental setup was used in this research to collect AE signals from slow rotating bearings. The test setup was designed so that it would be able to test slow rotating bearings. The experimental test setup is shown in Figure 2.3 (a). A servomotor controller controlled the rotational speed of the bearing. The system was driven by an AC servo motor with the speed set at 70 rpm, 80 rpm and 100 rpm for Timken Bearings 1, 2 and 3 with bearing number HR 30307 J respectively. An AE sensor was mounted on the housing of the test Timken Bearing.

The rotating bearings were loaded until failure, which occurred on the outer race for all the bearings considered here. At the start of the experiment half a teaspoon of ground metal debris was introduced gradually into Bearing 1 at the openings between the outer race, rollers and inner race to accelerate the initiation of damage. Figure 2.3 (d) and (e) show sample of Bearing 1 before and after damage. Bearings 2 and 3 were run under grease starvation conditions through-out the test period in order to speed up the bearing degradation. The bearings were loaded at various dynamic loads by using a Zonic servo-hydraulic shaker shown in Figure 2.3 (b). Bearing 1 was sinusoidally loaded over a range from 1.6 kN to -1.0 kN. Bearing 2 was loaded between 1.8 kN and -1.4 kN, whereas Bearing 3 was loaded between 2.0 kN and -1.7 kN. The excitation frequency was approximately 2 Hz.

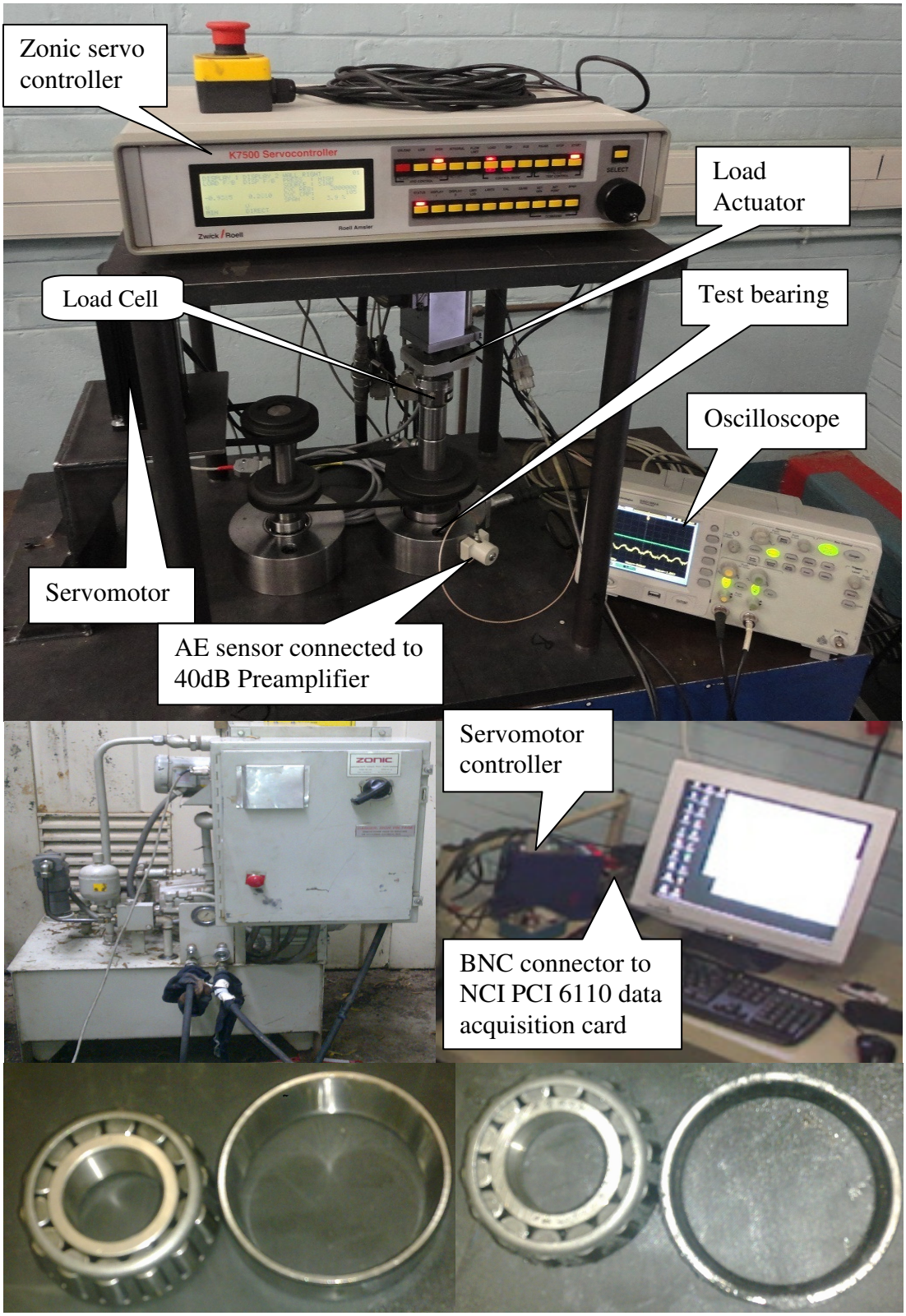

Figure 2.3: (a) Test setup (top) (b) Zonic servo hydraulic shaker (middle left) (c) Instrumentation (middle right) (d) Bearing 1 in good condition before testing (bottom left) (e) Bearing 1 with outer race damage after testing (bottom right)

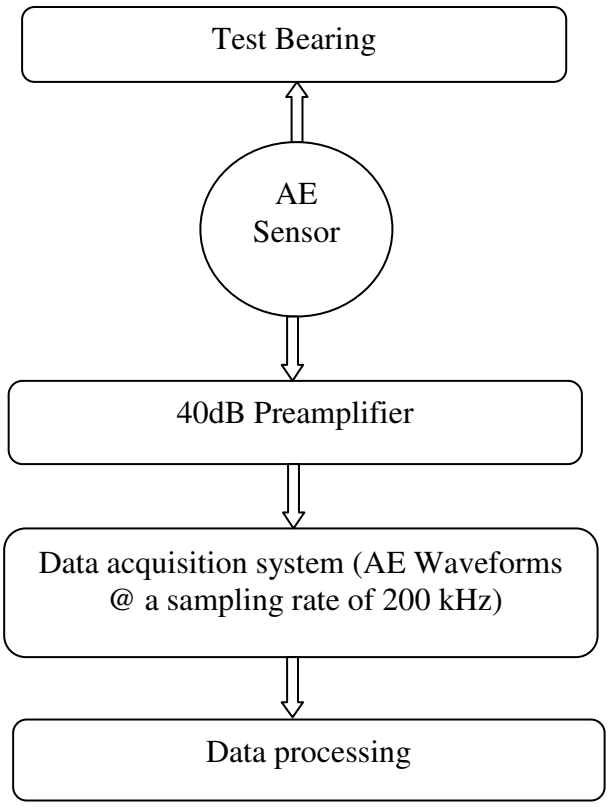

Figure 2.3: (f) Data acquisition schematics

The major components of the slow rotating bearing test setup are the Zonic Xcite 1100-4-FT System hydraulic shaker (load actuator), the load cell, the Timken tapered roller test bearings with bearing number HR 30307 J, the AC servo motor with model number 80MT-M04025, the speed controller and a National Instruments data acquisition card with a shielded BNC Connector Block.

The Soundwel AE sensor with model number SR 150 M was used for collecting the data in an analogue form. This broadband piezoelectric AE transducer was connected to a 40-dB gain pre-amplifier which amplified the AE signal and filtered out unnecessary noise. The AE transducer was mounted on the outside surface of the outer race and on the bearing housing.

The frequency of interest was 100 kHz. Hence, the AE signal was recorded at a sampling frequency of 200 kHz over a sampling period of 1 s, using the NI PCI 6110

data acquisition card with the model occupying one of the ISA slots in a host computer. The schematics of the data acquisition system is shown in Figure 2 (f). Data records were taken every 20 minutes until all three bearings failed, using the National Instruments LabVIEW software. The function for capturing the time domain and the pre-selected sampling time and interval was used. The recorded data was subsequently processed by means of dedicated Matlab programs.

2.9 Results and discussion

2.9.1 Results from alternative KPCAs and development of a DAI

Healthy slow rotating bearings were run until failure in this investigation. Figure 2.4 presents the AE signals for the healthy, slightly degraded and severely degraded states of Bearing 1. It can be seen from the topmost plot that the magnitude of the AE signal has values of about 0.2 volts for the healthy bearing state. As most of the bearing life is in the healthy state, much of the condition data is obtained during this period. Similarly, the middle plot in Figure 2.4 is a sample of the bearing when it is in a slightly degraded state. It can be seen that there is a significant increase in the magnitude of the AE signal for the slightly degraded state of the slow rotating bearing. The AE signal has relatively higher values of about 0.5 volts. This is an intermediate state and significantly shorter than the healthy state. The condition data here shows that the bearing condition has changed and that failure may not be too far away. Finally, the bottom plot in Figure 2.4 is a sample of Bearing 1 when in a severely degraded state. It can be seen that there is a significant increase in the magnitude of the AE signal for the severely degraded state of the slow rotating bearing to maximum values of about 2.0 volts, which points to imminent failure. Condition data is very scanty for this period as the bearing soon fails.

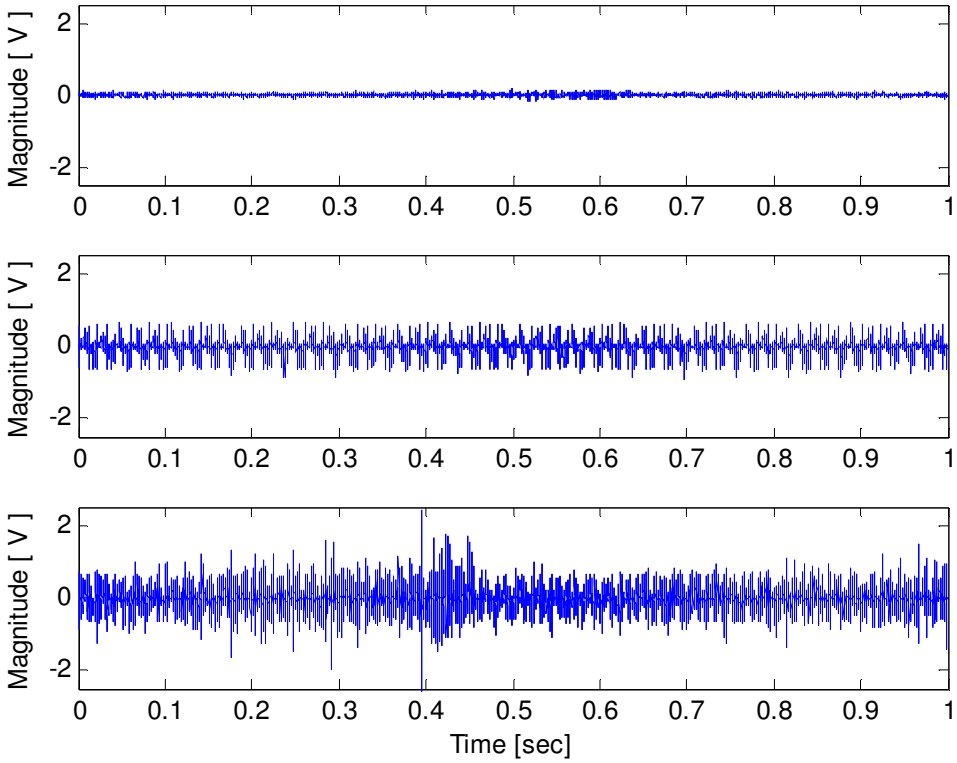

Figure 2.4: AE signal for Bearing 1: topmost (healthy state); middle (slightly degraded state) and bottom (severely degraded state)

The extracted features: RMS, kurtosis, crest factor, peak-to-peak and skewness, for Bearings 1, 2 and 3 are shown in Figure 2.5 over their entire lifespan. From the extracted bearing features such as kurtosis, RMS, peak-to-peak, crest factor and skewness, it can be shown that most of the slow rotating bearing fatigue time is consumed during the period of material damage accumulation, whereas the period of crack propagation and development is relatively shorter. Hence, for the effective maintenance of the slow rotating bearing, an early warning approach is necessary to detect the initial degradation. It can also be noted that the degradation pattern of the extracted features is inconsistent. The RMS, kurtosis, crest factor, peak value and skewness all exhibit strong inconsistencies, although all the test bearings are of the same type. This inconsistency makes it extremely difficult to assess the performance of the degradation states based on a one-feature deterministic model.

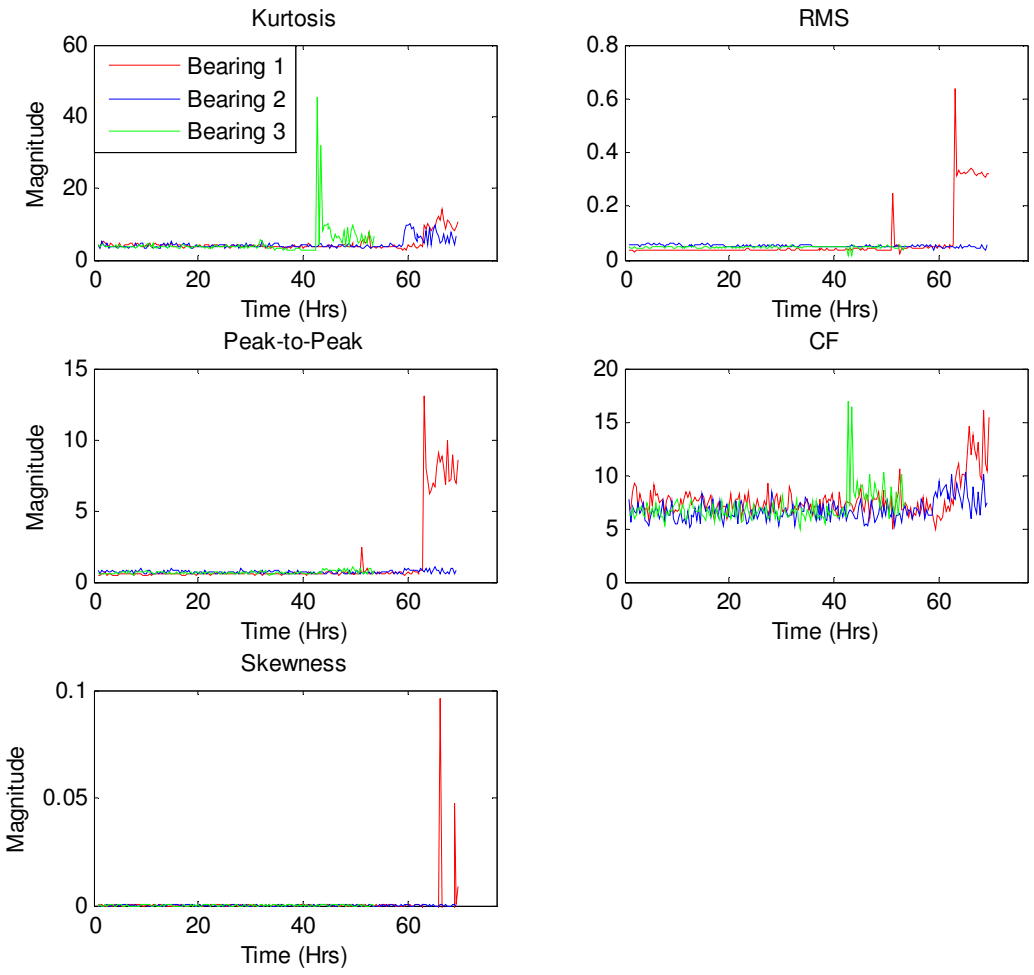

Figure 2.5: Kurtosis, RMS, peak-to-peak, CF and skewness for the complete lifespans of Bearings 1, 2 and 3.

Based on the inconsistencies observed with the original extracted features, it is not the best practice to select a specific feature for the analysis of bearing degradation. Moreover, since each feature has its strengths and weaknesses, the next task is to seek a means of combining these features into a single index for more sensitive and robust defect detection. To ensure that only the correlated and most effective features are combined, the KPCA is used for dimensionality reduction. This implies reducing a large number of original features to a manageable and more meaningful lower dimension. Different forms of KPCA were fitted. These included Gaussian KPCA (GKPCA), simple KPCA (SKPCA), and PKPCA of orders 2, 3 and 4 (i.e. PKPCA-1, PKPCA-2, and PKPCA-3 respectively). The KPCA is a nonlinear data dimensionality

reduction technique. However, the traditional linear PCA was also fitted for comparison.

One of the important criteria for determining how many principal components should be retained is to assess what proportion of variability in the data is accounted for by the selected principal components. Usually 70% – 90% is the benchmark (Martinez and Martinez, 2004). As the first two principal components of each of the three bearings accounted for over 78% of the variability, only these two components were selected for subsequent analysis. To assess the usefulness of the extracted features from KPCA, the K-means classification was used for the classification of the entire bearing data for Bearings 1, 2 and 3. The classification errors of the PCA, GKPCA, SKPCA, PKPCA-1, PKPCA-2, PKPCA-3 are shown in Figures 2.6, 2.7 and 2.8. The results show that the PKPCA of order 2 (i.e. PKPCA-1) gave the best classification result with the least classification error. The corresponding figures that generated these classification errors are plotted in Figures 2.9, 2.10 and 2.11 for Bearings 1, 2 and 3 respectively for only the best KPCA (i.e. PKPCA-1) and the traditional linear PCA. Figures 2.6, 2.7 and 2.8 and Figures 2.9, 2.10 and 2.11 show the superiority of the nonlinear PKPCA-1 over the traditional linear PCA. Hence, the PKPCs obtained by using PKPCA-1 were subsequently used in the GMM to obtain the NLLs for the bearings, over the entire lifespan of each bearing.

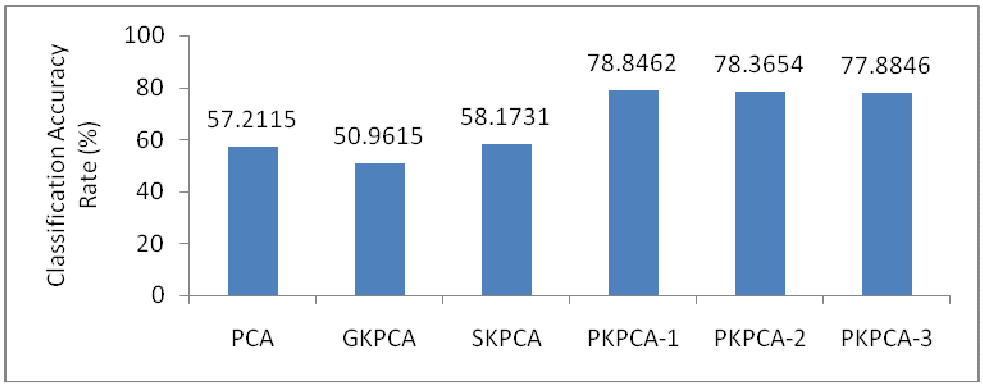

Figure 2.6: Classification accuracy rate of Bearing 1

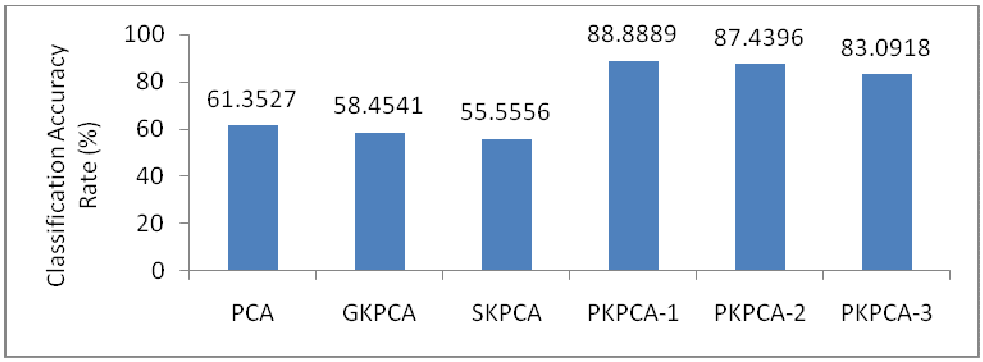

Figure 2.7: Classification accuracy rate of Bearing 2

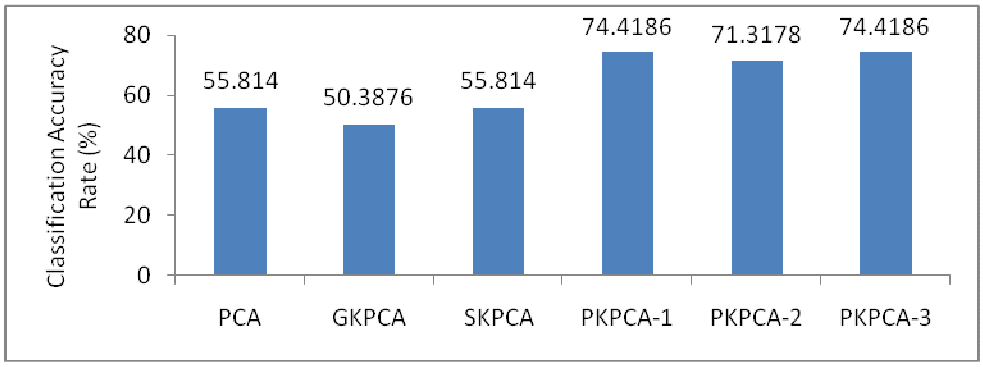

Figure 2.8: Classification accuracy rate of Bearing 3

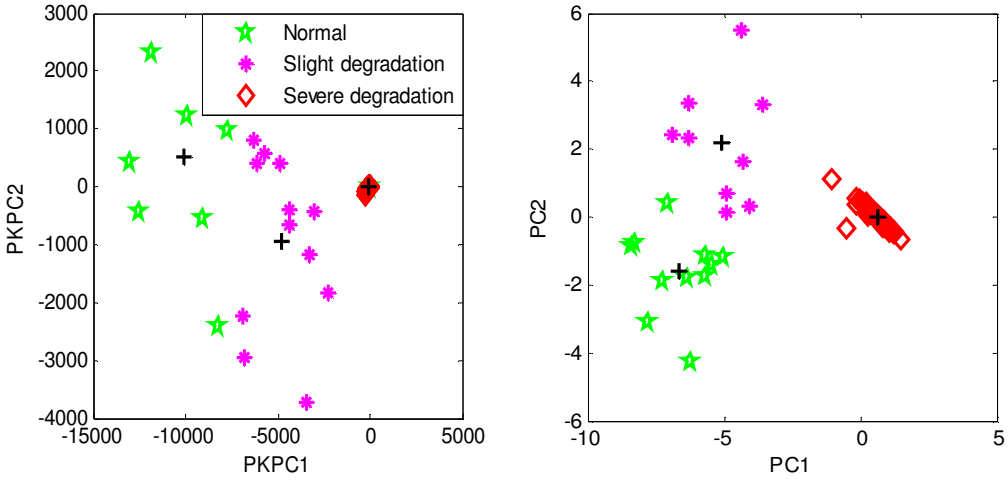

Figure 2.9: Bearing condition classification based on the first two principal components extracted by PKPCA and PCA for Bearing 1

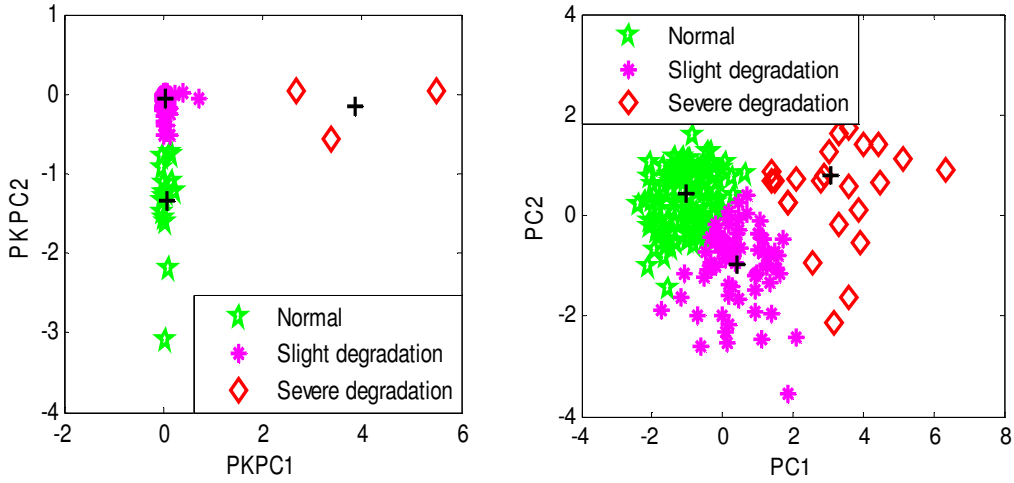

Figure 2.10: Bearing condition classification based on the first two principal components extracted by PKPCA and PCA for Bearing 2

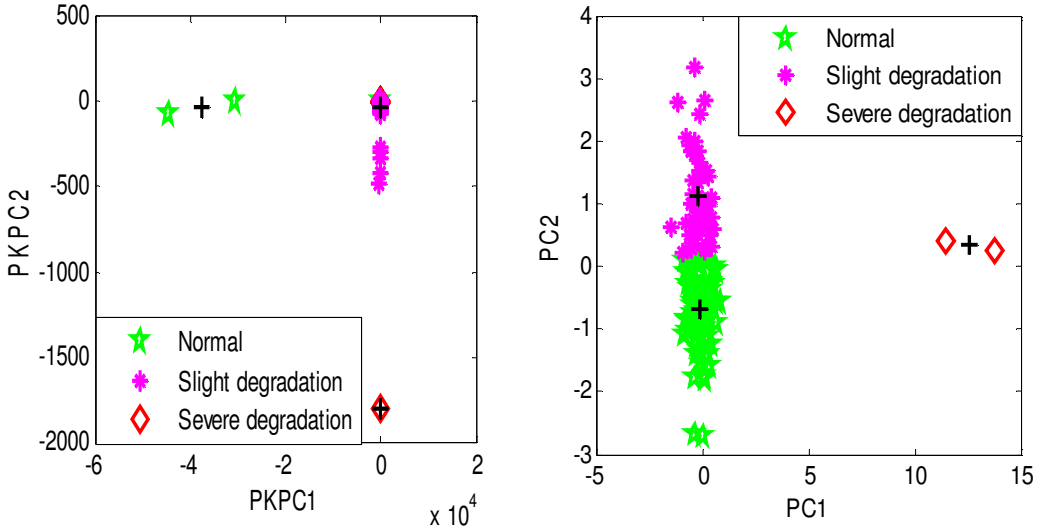

Figure 2.11: Bearing condition classification based on the first two principal components extracted by PKPCA and PCA for Bearing 3

The performance of the GMM model is sensitive to the number of mixing components (see Equation 2.20). The optimum number of components was obtained in this study by using the leave-one-out CV technique. CV is performed by partitioning the extracted features into two segments of equal length, then doing the analysis on one segment (called the training set) and validating the analysis on the other segment (called the testing or validation set). Each model is retrained three times using the expectation maximisation algorithm. The best-performing model is the one with the minimum NLL. Three mixture components were found to be

optimal for all the bearings, as shown in Figure 2.12. When more than three components are used, the performance on the validation set decreases, indicating that the models are beginning to over-fit the training set. The NLL results for Bearings 1, 2 and 3 are presented as shown in Figures 2.13a, 2.14a and 2.15a respectively. Subsequently, the DAI is obtained by smoothening the data, using the exponentially weighted mean average (EWMA) as shown in Figures 2.13b, 2.14b and 2.15b respectively.

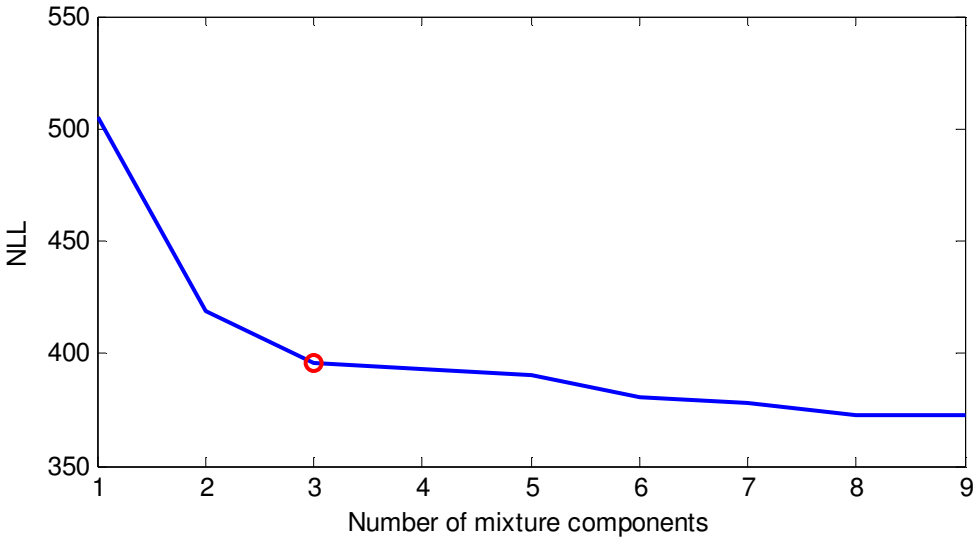

Figure 2.12: Number of mixture components

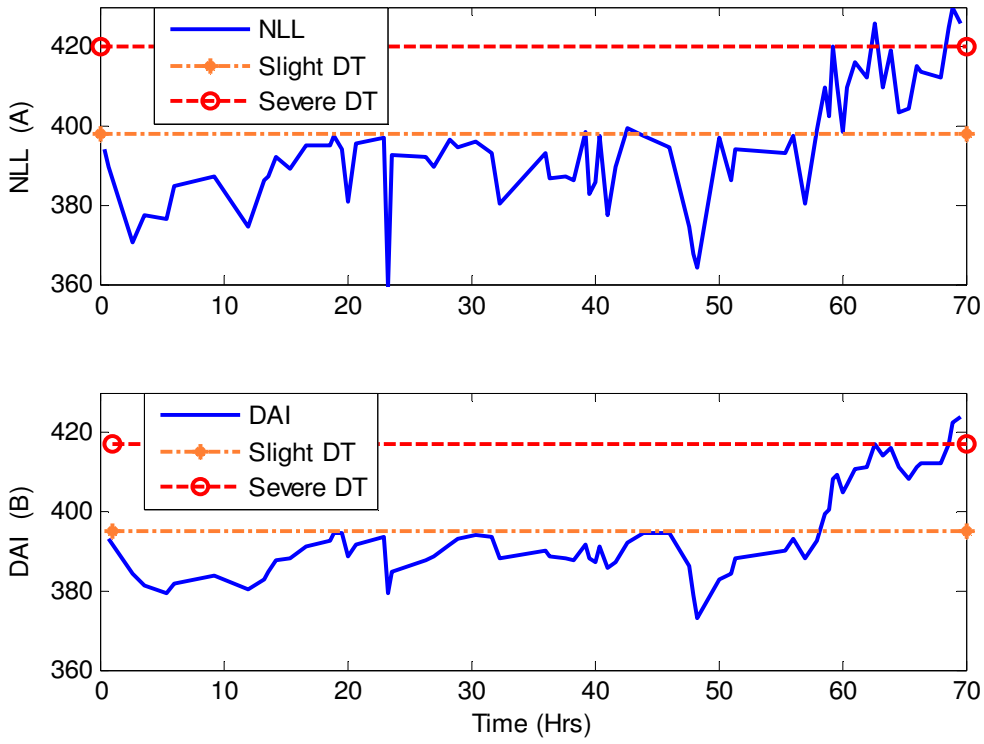

Figure 2.13: NLL and DAI for the whole lifespan of Bearing 1.

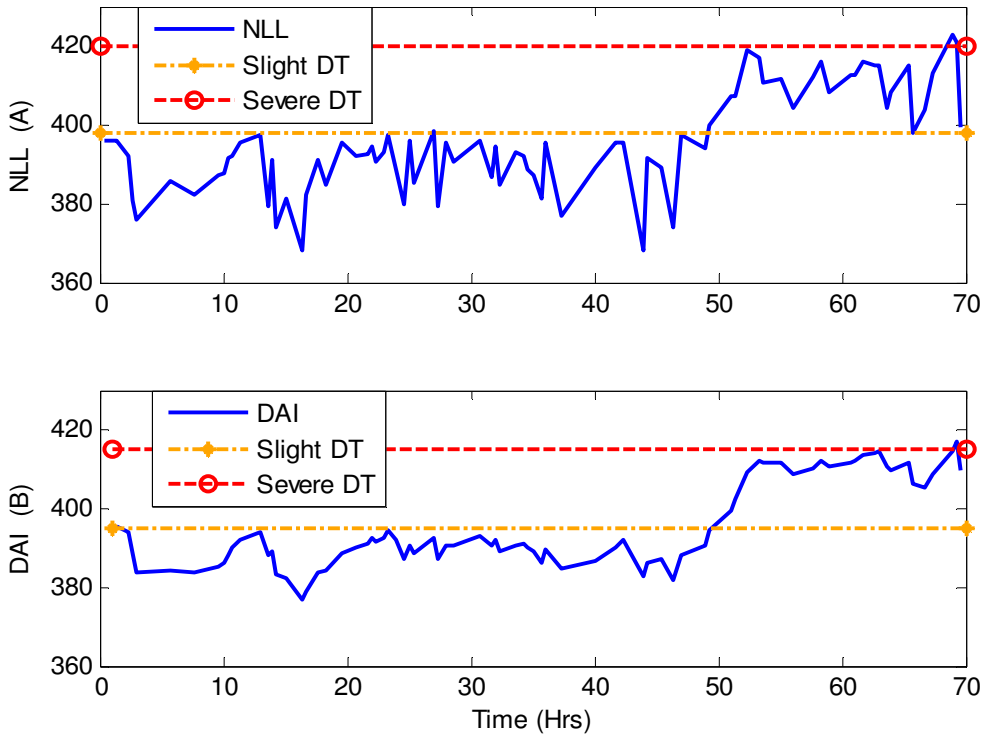

Figure 2.14: NLL and DAI for the whole lifespan of Bearing 2.

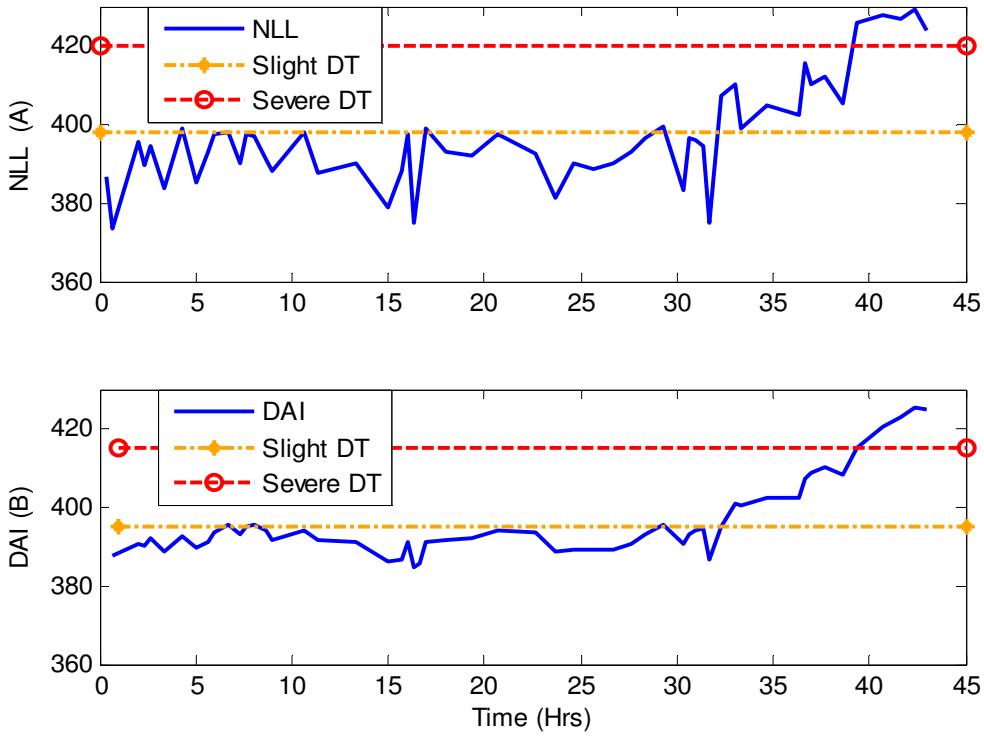

Figure 2.15: NLL and DAI for the whole lifespan of Bearing 3.

The DAI which was obtained by smoothening the NLL using EWMA (see Figure 2.1) is then used for assessing the bearing damage and failure at future time intervals. There is no need for damage detection when a bearing is in its healthy state. When the computed features are elevated above their threshold values, then it is considered that slight degradation has set in. The DAI gives an indication of whether a bearing is in a healthy, slightly degraded or severely degraded state. A DAI was developed for the assessment of the degradation of the slow rotating bearing. No prior information was available and GMM based DAI was utilised for the healthy bearing data in the first instance to obtain the degradation threshold(s) (DT). The slight DT was obtained by using the KDE method on the healthy bearing data for Bearing 1. KDE was first implemented on the healthy Bearing 1 data (i.e. the first 50 hours). The severely degraded threshold was obtained by using the KDE method on the slightly degraded bearing data (i.e. the first 50 to 65 hours).

The degradation of Bearings 1, 2 and 3 was assessed by using the proposed DAI shown in Figures 2.13(b), 2.14(b) and 2.15(b). In comparison with the RMS, kurtosis, crest factor, peak-to-peak and skewness, DAI is more effective due to its stable and clear variance trends for the degradation period for the whole life of the bearing. It can be seen from Figures 2.13(b), 2.14(b) and 2.15(b) that the bearing degradation process is presented from healthy, slight degradation and severe degradation to failure. The time from slight degradation to severe degradation is very important in the bearing life. Effective maintenance has to be planned and carried out within this period to extend the life of the bearing or avoid machinery breakdown. For this reason, some maintenance measurements should be taken more often once slight degradation has commenced to avoid catastrophic failure with its potentially disastrous consequences. Once the slight DT has been exceeded, this indicates the beginning of incipient damage in the proposed model, and in this way gives an indication that maintenance procedures should commence. Furthermore, once the severe DT has been exceeded, this indicates that the final failure is very close. DAI improves the comprehension of the damage assessment and degradation process of the slow rotating bearing in real-life applications. For Bearings 1, 2 and 3, the incipient damage commenced at time intervals of 50 hours, 60 hours and 30 hours respectively.

2.9.2 Evaluation of the proposed DAI using benchmark studies

This part of the study evaluated the effectiveness of the proposed DAI by comparing its performance with that of other monitoring indexes. To this end, the proposed model was reordered to form three additional different submodels which basically represent some of the models that may have been used in other studies for developing condition monitoring indexes. In other words, each submodel misses at least one component that is captured in the proposed model. The models consist of the PKPCA-GMM-EWMA model, the PKPCA-EWMA model, the PKPCA-GMM model and the GMM-EWMA model. The different models and their capabilities are shown in Table 2.1.

Table 2.1: Model reordering

PKPCA-GMM-EWMA		Handles non-linear data	Handles multi-modal distri-bution	Incor-porates prior know-ledge	Handles uncer-tainties	Dimen-siona-lity reduc-tion	Dynamic or time variation & filters noise	
PKPCA-GMM	PKPCA	✓	X	X	X	✓	X	PKPCA-EWMA
PKPCA-GMM	GMM	✓	✓	✓	✓	X	X	PKPCA-EWMA
GMM-EWMA	EWMA	✓	X	X	X	X	✓	PKPCA-EWMA

The PKPCA-GMM-EWMA (also known as the DAI) was the model proposed for this study. Since it is a combination of the PKPCA, the GMM and the EWMA models, it combines all the features of the three individual models. When bearings are run under varying operating conditions, there is a high likelihood that the extracted features will have nonlinear and or multimodal (multiple modes) distribution (Yu, 2011b). The PKPCA submodel is capable of reducing the dimension of the extracted AE features (kurtosis, RMS, crest factor, skewness and peak-to-peak) from a nonlinear high-dimensional data space to a linear low-dimensional data space, in this way enhancing the high correlation existing in the original features. The PKPCA model neither accounts for noise in the various AE features nor does it permit an incorporation of external knowledge (priors) about the model. Moreover, the PKPCA model is a static model as it does not capture the dynamics of the system (i.e. the historical time evolution of the monitoring statistics). Hence, the KPCA model does not account for time variation in the data.

After extracting the features from the high-dimensional data via PKPCA, the multimodal features in the low-dimension can still be preserved. Therefore, for the effective diagnostics of bearing faults and assessment of performance degradation, the GMM submodel is used to describe the multimodal distribution of the data. The GMM is accordingly capable of handling the multimodal and nonlinear features of the data by using a mixture of Gaussian components. Moreover, as the GMM is a Bayesian technique, it allows one to combine external information with the information in the data through a prior density function. Hence, all variables and parameters are considered as random or stochastic and their behaviour is described by

a probability density function. This enables the GMM to handle uncertainties in the data and parameters. Like the PKPCA, the GMM model is a static model. However, the EWMA model developed by Wold (1994) is a dynamic nonlinear model that includes a memory function by using historical data for monitoring a bearing or any other systems (Yoo *et al*., 2006). This makes it capable of improving the sensitivity and reliability of monitoring techniques to detect a slight degradation or shifts in the performance of bearings (Yu, 2011b). It also smoothens and filters out noise in the data and is therefore capable of capturing non-stationarity.

Following the above features of the three submodels, the PKPCA-GMM-EWMA model can handle the nonlinear and multimodal distribution of the data, incorporates prior knowledge, accounts for uncertainty in both the data and the parameters of the model and is dynamic. The PKPCA-EWMA is a combination of the PKPCA and EWMA models so it is capable of handling nonlinearities and the dynamic characteristics in the extracted features. However, it does not handle multimodal distribution nor account for uncertainties in the parameters of the model and it does not incorporate prior information. The PKPCA-GMM model has all the features or capabilities of the PKPCA-GMM-EWMA model except that it is not dynamic, but static. The GMM-EWMA model also has all the capabilities of the PKPCA-GMM-EWMA model, except that the concept of dimensionality reduction is not involved, meaning that the model uses all the extracted features.

There are several methods for selecting the optimal number of principal components from the PKPCA-EWMA model. These include the use of scree plots, the cumulative percentage variance explained, cross validation and the PCs with eigenvalues equal to or greater than the average eigenvalue. This study employs the cumulative percentage variance explained (Martinez and Martinez, 2004) where the number of principal components that explain a cumulative percentage variance of between 70% – 90% is selected. Two principal components were selected in the present study as they accounted for over 78% of the cumulative variance. These two retained principal components were subsequently used either to construct the traditional monitoring statistics (T^2 and SPE) or as inputs into the GMM model.

The effectiveness of the proposed model was tested by comparing its degree of accuracy in discriminating between healthy and faulty bearing conditions with that of the submodels. The evaluation of performance can be reported by using the false alarm rate and the detection rate criteria. The false alarm rate gives information about the robustness of each model to healthy system changes. The detection rate gives information about the sensitivity and efficiency of detecting faults. The rates are obtained by counting the percentage of samples that fall outside the 99.7% confidence level used in setting the thresholds. Since two thresholds have been defined, the detection rate was also classified into a slightly degraded detection rate and a severely degraded detection rate.

The plots showing the monitoring indexes obtained from the different models are discussed before discussing the qualitative assessment. Figure 2.16 shows the results for Bearing 1. A number of observations emerge from these figures. The monitoring indexes from the PKPCA-GMM-EWMA, PKPCA-EWMA-T^2 and PKPCA-EWMA-SPE models appear to perform better than the index from the GMM-EWMA model, in terms of less volatility (stability or stationarity) and also the ability to discriminate between faulty and healthy operating working conditions, with the exception of the PKPCA-GMM that appears to be as volatile as the GMM-EWMA. This is not surprising, given that the inclusion of all the original extracted features in the GMM-EWMA model can obscure its ability to monitor the degradation trends effectively. The PKPCA-GMM's non-stationarity can be attributed to the absence of the EWMA smoothening and noise filtering in this model. It can also be observed that the PKPCA-EWMA-SPE statistic shows a clearer distinction between the healthy operating condition and the faulty state than its counterpart T^2 statistic.

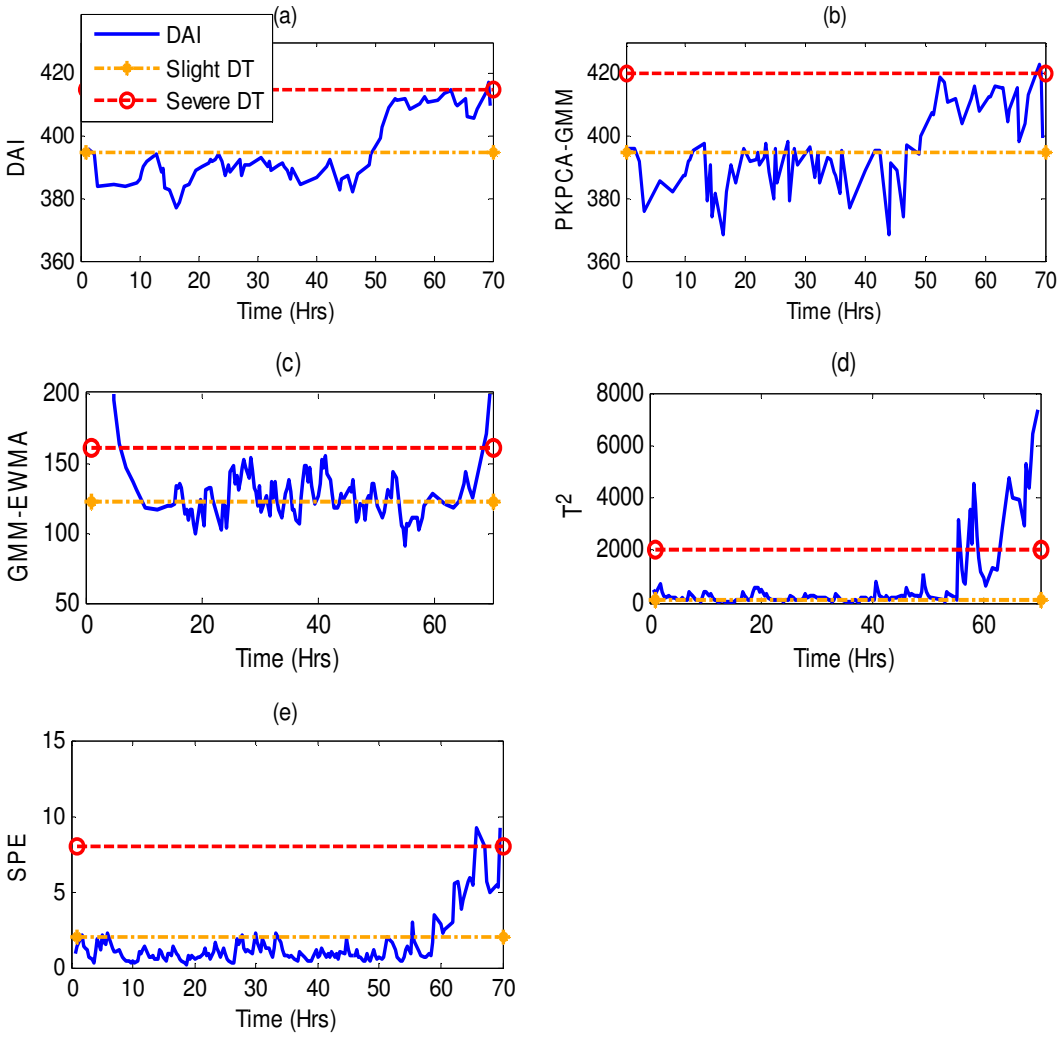

Figure 2.16: Condition monitoring indexes for Bearing 1 (a) DAI (b) PKPCA-GMM (c) GMM-EWMA (d) PKPCA-EWMA-T^2 and (e) PKPCA-EWMA-SPE.

Trends similar to those in Figure 2.16 are also observed in Figures 2.17 and 2.18 for Bearings 2 and 3, respectively. Overall, however, the proposed DAI from the PKPCA-GMM-EWMA model appears to be smoother, more stable and to exhibit clearer trends for the degradation period than the rest. The highly non-stationary trends in both the GMM-EWMA and PKPC-GMM and the inability of the PKPCA-EWMA-T^2 and PKPCA-EWMA-SPE to discriminate clearly between the health states, demonstrate that neither reducing the dimensionality of the original features alone nor capturing the uncertainties and dynamics of the system alone, can guarantee the effective diagnostics of the slow rotating bearings. The characteristics or capabilities of each of these models are complementary and therefore have to be

considered when developing an index which can effectively monitor slow rotating bearings. These complementarities are depicted in the DAI and largely explain why it appears to be performing better than the condition monitoring indexes from the different submodels.

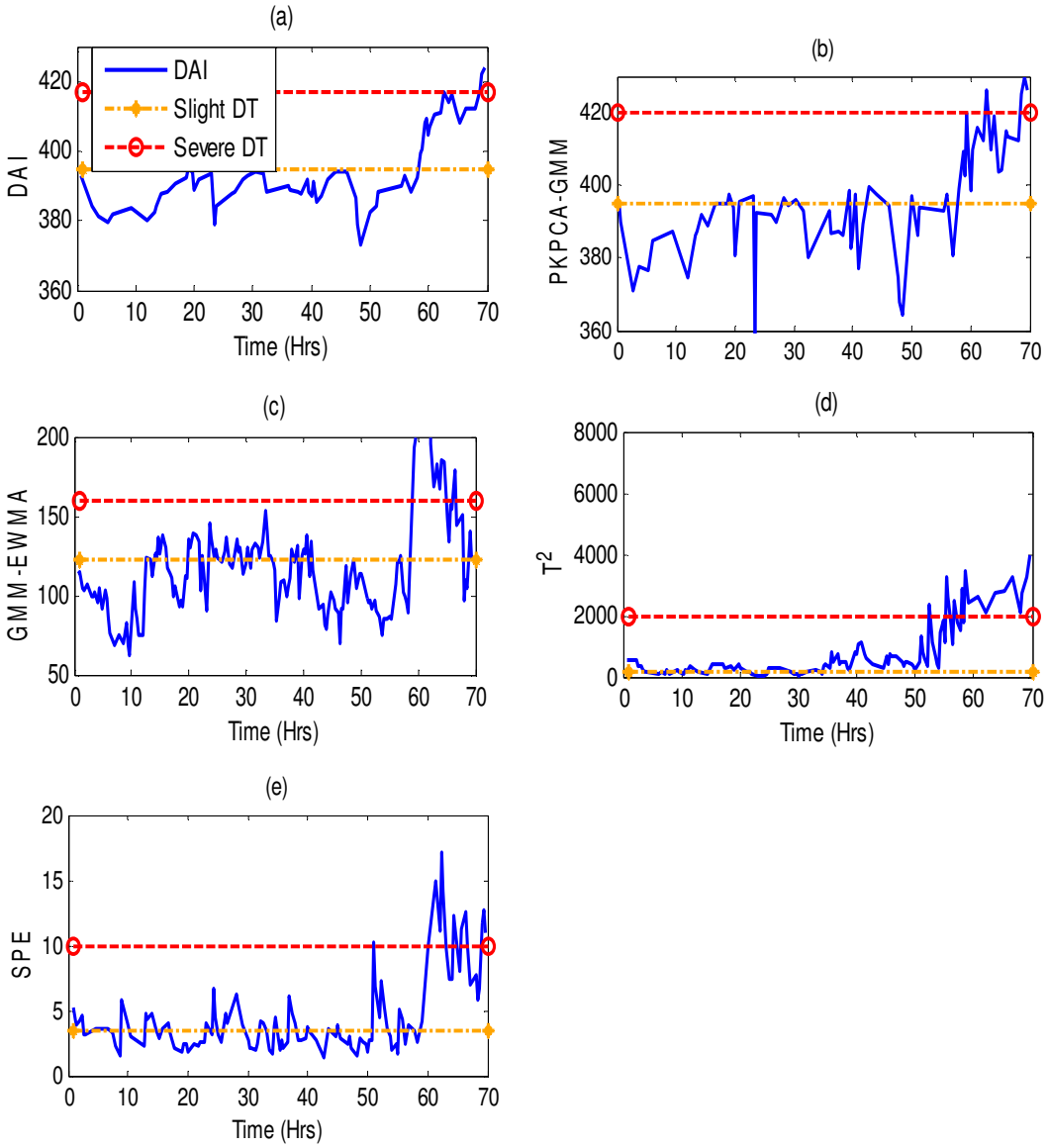

Figure 2.17: Condition monitoring indexes for Bearing 2 (a) DAI (b) PKPCA-GMM (c) GMM-EWMA (d) PKPCA-EWMA-T^2 and (e) PKPCA-EWMA-SPE.

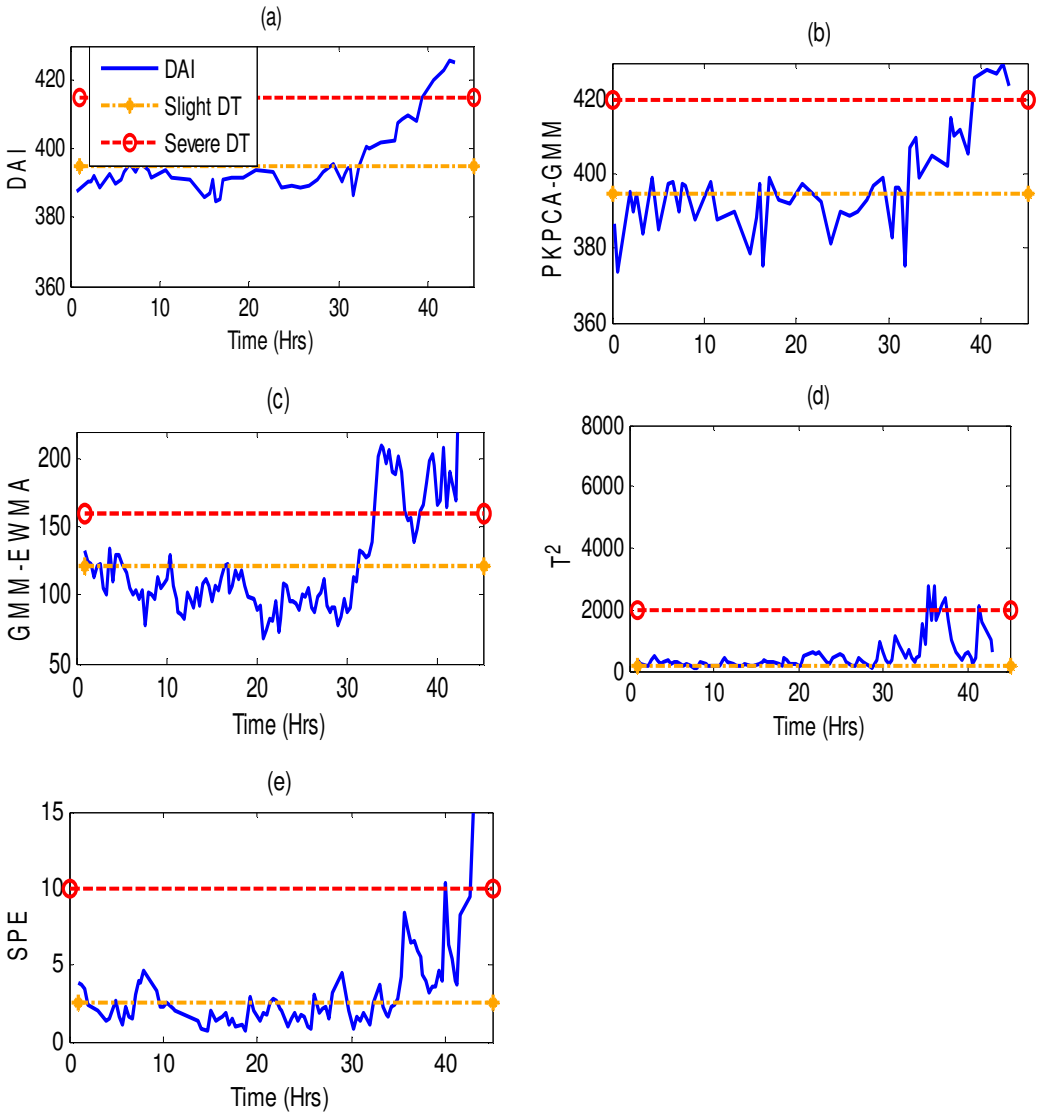

Figure 2.18: Condition monitoring indexes for Bearing 3 (a) DAI (b) PKPCA-GMM (c) GMM-EWMA (d) PKPCA-EWMA-T^2 and (e) PKPCA-EWMA-SPE.

The quantitative assessment of the effectiveness of each model in assessing the health states of each of the slow rotating bearings requires a comparison of its false alarm rates, the slight degradation detection rates and the severe degradation detection rates with those of the submodels. The results of the comparison are presented in Table 2.2 for Bearings 1, 2 and 3.

Table 2.2: False alarm and degradation detection rates of the three bearings (%)

Bearings	False alarm and fault detection rate	PKPCA-GMM-EWMA	PKPCA-GMM	GMM-EWMA	PKPCA-EWMA	
					T^2	SPE
Bearing 1	False alarm rate	1.97	25.45	55.45	42.31	4.23
	Slight degradation detection rate	100.00	75.00	42.31	61.54	60.87
	Severe degradation detection rate	50.00	50.00	50.00	60.00	42.86
Bearing 2	False alarm rate	0.00	25.00	20.00	65.48	20.53
	Slight degradation detection rate	100.00	77.78	29.73	84.21	94.12
	Severe degradation detection rate	54.55	53.85	50.00	62.00	65.00
Bearing 3	False alarm rate	7.50	32.43	12.36	72.22	7.62
	Slight degradation detection rate	88.89	76.92	56.25	75.00	83.33
	Severe degradation detection rate	100.00	100.00	80.00	92.31	60.00

First, by focusing on Bearing 1, it can be seen that the proposed index from the PKPCA-GMMA-EWMA model has the lowest false alarm rate (1.97%), meaning that in the healthy bearing operating region, the index crosses the 99.7% confidence level 1.97 times out of 100, whereas the worst index (GMM-EWMA) crosses the threshold about 55 times out of 100 when the bearing is actually healthy. The proposed model could perfectly detect a slight fault and remained within the threshold of the slight degradation state 100% of the time, followed by PKPC-GMM with a slight degradation detection rate of 75%. The proposed model's performance was poor in terms of its ability to detect severe degradation as it missed 50% of the time. In this case, the PKPCA-EWMA-T^2 performed better. For Bearing 2, the proposed model had a 0% false alarm rate whereas no other model did as well. In terms of the ability to detect a slight degradation, the proposed model performed better and had a detection rate of 100% whereas the second-best model, PKPC-EWMA-SPE, had a detection rate of 94%. For Bearing 3, the proposed model still had the lowest false alarm rate (7.50%) followed by the PKPC-EWMA-SPE model with a false alarm rate of 7.62%. In terms of detecting a slight degradation, the detection rate was

approximately 89% for the proposed model followed by the PKPCA-EWMA-SPE model; whereas it performed as well as the PKPCA-GMM model in terms of detecting severe degradation. The quantitative analysis confirms the analysis of the visual plots. Overall, the proposed model performed better than the other models in terms of the false alarm rate and the fault detection rates.

The overall superiority of the monitoring performance of the PKPCA-GMM-EWMA DAI can be attributed to the fact that it uses fewer kernel principal components from the original features, thus reducing redundancies in the data. In addition, it incorporates the dynamics of the slow rotating bearings, making it more sensitive to and reliable in detecting slight faults via EWMA; accounts for nonlinearities and multimodal distribution in the data, thus eliminating the bias arising from wrong model specifications; handles adequate filtering of noise; allows for smooth characteristics of the features through both GMM and EWMA subcomponents; reduces uncertainties in the measurement of the data as well as in the estimations of the parameters of the model; and permits the incorporation of external knowledge via the prior density function characterisation in the GMM component.

2.10 Summary

This chapter proposes a novel approach to the detection of damage in slow rotating bearings. In this study, three healthy slow rotating bearings were run until failure. A DAI was developed. The DAI was obtained from the integration of PKPCA, the GMM and the EWMA and used for assessing the damage degradation of the slow rotating bearing from incipient damage to failure.

The K-means classification was used in the state classification of the bearing data. The best-performing PKPCs were then used in the GMM to obtain the NLL. Subsequently, the DAI was gotten by the smoothening of the NLL with the EWMA and used for the assessment of bearing damage. The slight DT was obtained by using the KDE method on the healthy bearing data. Subsequently, the severe DT was obtained by using the KDE method on the slightly degraded bearing data.

The effectiveness of the DAI was investigated by comparing its performance with that of other monitoring indexes and it was found to outperform them. The DAI was the only model with all these properties: it uses fewer kernel principal components from the original features thereby reducing redundancies in the data, incorporates the dynamics of the slow rotating bearings thereby making it more sensitive to and reliable in detecting slight faults via EWMA, accounts for the nonlinearities and multimodal distribution in the data thus eliminating the bias arising from wrong model specifications, filters noise, allows for the smooth characteristics of the features through both the GMM and EWMA subcomponents, reduces the uncertainties in the measurement of the data as well as in estimations of the parameters of the model, and permits the incorporation of external knowledge via the prior density function characterisation in the GMM component.

The proposed DAI has been proven to be effective in the CM of slow rotating bearings under varying operating conditions. With further modifications it could be used for the CM of other rotating machinery such as gears.

Chapter 3 AE-based prognostics of slow rotating bearing using Bayesian techniques

3.1 Introduction

This chapter integrates both diagnostics and prognostics (RUL prediction) aspects of CM of slow rotating bearing using whole life bearing data from a laboratory experiment.

RUL refers to the time left before the observance of failure given the current machine condition and past operation profile. In the literature, RUL is sometimes referred to as residual life, remaining service life, or remnant life (Jardine *et al*., 2006). Prognosis requires knowledge (or data) on the failure mechanism. The failure mechanism is described in two ways. The first one assumes that failure depends only on the condition variables, which reflect the actual fault level, and the predetermined boundary. The most commonly used failure definition in this case is simple: failure occurs when the fault reaches a predetermined level. The second one builds a model for the failure mechanism using historical data available. In this case, different definitions of failure can be used (Jardine et al., 2006). In this study two thresholds were defined to enable the estimation of RUL. The first is the slightly degraded threshold line which indicates the onset of incipient damage. The second is the severely degraded threshold line which indicates onset of final failure. The time difference between the two thresholds gives the RUL.

Prediction of RUL is often difficult as the results depend on the models used in obtaining them. Therefore, it is important to evaluate the predictions from alternative models and choose the best based on an objective criterion. A model is deemed superior if it effectively minimizes the one-step-ahead (or multi-step-ahead) PEs by producing a lower PE than its competitors. Against this background, this study evaluates the performance of alternative Bayesian methods for slow rotating bearing fault prognostics based on AE data obtained from a run-to-failure experiment.

In this chapter, the novel DAI developed in chapter 2 was used as an input in several Bayesian regression models such as the MLP, RBF, BLR, GMR and the GPR mentioned in chapter 1, paragraph 1.7 for RUL prediction. The DAI incorporated all the advantages of the various extracted features (kurtosis, peak-to-peak, RMS, skewness, crest factor) capitalizing on the strengths of each, and thereby becoming more sensitive and robust in prognostics, while at the same time reducing the number of dimensions for CM (Malhi and Gao, 2004). The MAPE and RMSE were used in evaluating the performance of the models and hence selecting the best performing model for the prediction of slow rotating bearing RUL.

This study contributes to literature on prognostics by evaluating the performance of the MLP, RBF, BLR, GMR and GPR models using the same data set. Finally, the models are compared using both dependent and independent samples (see chapter 1, paragraph 1.8.7).

3.2 Methodology

This section describes the different models used in developing the proposed approach to prognostics of slow rotating bearings.

The entire CM process is implemented in a unified framework (Figure 3.1). Generally the process involves four main steps: (1) obtaining the DAI; (2) using the obtained DAI as input into MLP, RBF, BLR, GMR and GPR respectively; (3) the various models are then used for the RUL prediction of slow rotating bearings; (4) the five obtained models are evaluated to find out which of them gave the best prediction.

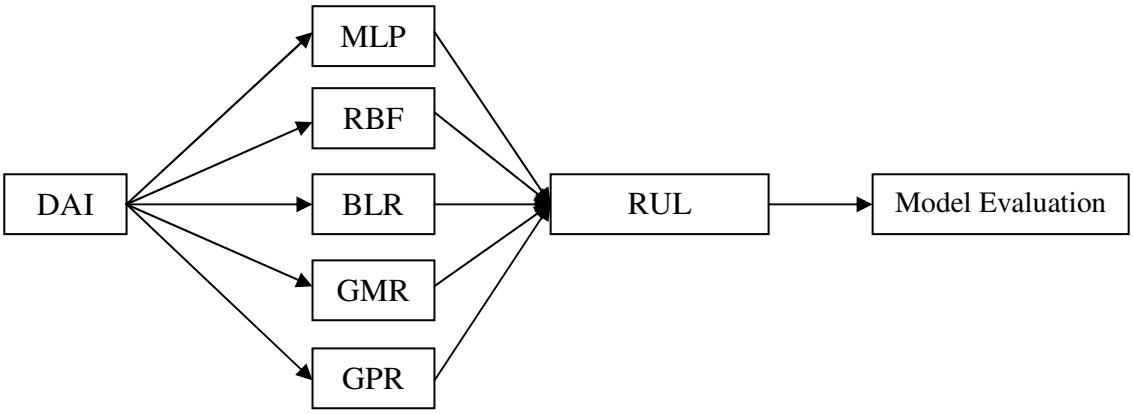

Figure 3.1: Framework for DAI integrated approach to bearing prognostics

3.2.1 The degradation assessment index

The DAI obtained in chapter 2 by the combination of the PKPCA, the GMM, and EWMA is used in this chapter as an input to the various models in the prognostics of slow rotating bearings. The PKPCA technique is used for dimensionality reduction and feature extraction. Subsequent to feature extraction by PKPCA from the high-dimensional statistics, the nonlinear (multimodal) features in low-dimensional data space can still be preserved. The PKPCs which were extracted are then used as inputs in the GMM which is an outstanding technique of complex data description, with benefits of high-performance computation and robustness. The GMM describes complex data distribution that often occurs in AE data by outputting the NLL utilizing numerous Gaussian components. The reliability and sensitivity of the NLL to the bearings slight degradation was improved by employing the EWMA statistic as an improved quantification index for prognostics of slow rotating bearing. The resulting quantification index is named DAI (see chapter 2, paragraph 2.6).

3.2.2 Multi-layer perceptron regression

The DAI is used as input into multi-layer perceptron network (MLP) for the prognostics of slow rotating bearing in order to determine its RUL. MLP is one of the generally utilized architecture for empirical usage of neural networks. It more often than not comprise of basically two layers of adaptive weights. There is a complete

linkage connecting the inputs to the hidden units, as well as another connecting the hidden units to the output units (Nabney, 2002).

The MLP is a mathematical function which has been parameterized by a set of numerical weights w_1, w_2,..., w_n, which we shall represent jointly by a vector of weight w. The Bayesian technique entails inference of the posterior distribution of weights, $p(w|D)$, given data D. It considers a functional probability distribution over the weighting space. The outputting prediction resulting from the input vector x is then determined by implementing a weighted sum of the predictions over all possible weight vectors, where the weighting coefficient for a particular weight vector is dependent on the posterior weight distribution. The predicted value is given as (Skabar, 2007):

$$\hat{y}^n = \int f(x^n, w)p(w|D)dw \tag{3.1}$$

where $f(x^n, w)$ is the MLP output, and $\hat{y}^n$ is the predicted value.

The probability density function, $p(w|D)$, can be approximated using the fact that $p(w|D) \propto p(D|w)p(w)$, where $p(D|w)$ and $p(w)$ are known respectively as the likelihood and the prior.

The prior weight distribution, $p(w)$, is the weight distribution before the observation of any data reflecting the prior knowledge of the MLP complexity. To obtain a smooth function for the reduction of the risk of overfitting, $p(w)$ is assumed to be Gaussian with zero mean and inverse square variance α which gives:

$$p(w) = \left(\frac{\alpha}{2\pi}\right)^{m/2} \exp\left(-\frac{\alpha}{2}\sum_{i=1}^{m} w_i^2\right), \tag{3.2}$$

where m is the number of weights in the MLP. Because α controls the value of other parameters (i.e. the weights) it is referred to as a hyperparameter.

Since the prior depends on α, the modification of Equation (3.1) with the inclusion of the posterior distribution over parameters of the hyperparameters gives the predicted value as:

$$\hat{y}^n = \int f(x^n, w) p(w, \alpha | D) dw d\alpha \tag{3.3}$$

where

$$p(w, \alpha | D) \propto p(D | w) p(w | \alpha) p(\alpha) \tag{3.4}$$

3.2.3 Radial basis function regression

Similarly, the DAI is used as input into RBF regression for the prognostics of slow rotating bearing in order to determine its RUL. The RBF is used for nonlinear modelling. The RBF has many advantages. One of these is that it has a two phase training process which is significantly quicker than MLP. In the first phase, the parameters of the basis functions are set to model the unconditional data density. In the second stage of training, the weights in the output layer are determined. Secondly, it is possible to assign an interpretation to the hidden units and also to determine the intrinsic degrees of freedom of the network (Nabney, 2002). The RBF network mapping could be written in the following form as shown in Equation (3.5).

$$y_k(x) = \sum_{j=1}^{M} w_{kj} \phi_j(x) + w_{k0} \tag{3.5}$$

where ϕ_j are the basis functions, w_{kj} are the output layer weights.

The bias weights can be absorbed into the summation by including an extra basis function ϕ_0 whose activation is constant value 1. This leads to Equation (3.6).

$$y_k(x) = \sum_{j=0}^{M} w_{kj} \phi_j(x) \tag{3.6}$$

Two Bayesian approaches have been found to be effective in practice to neural networks namely: Gaussian approximation to the posterior weight distribution in the weight space (known as Laplace approximation) often coupled with use of the evidence procedure for optimal hyperparameter estimation; secondly the Monte Carlo techniques, particularly the hybrid Monte Carlo (Nabney, 2002).

3.2.4 Bayesian linear regression

The parametric approach focuses on the use of probability distributions having specific functional forms governed by a small number of adaptive parameters, such as the mean and variance whose values are to be determined from the data set. The probability distributions include beta (binomial) and Dirichlet (multinomial) distributions for discrete random variables and the Gaussian distribution and Gaussian mixture distribution for continuous variables. In this study the data is continuous hence the Gaussian distribution and Gaussian mixture distributions are considered (Bishop, 2006). The Gaussian, also known as the normal distribution, is a widely used model for the distribution of continuous variables (Bishop, 2006). For the case of a single real-valued variable x, the Gaussian distribution is given as:

$$N(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{3.7}$$

where μ is the mean, and σ^2 is known as the variance.

The reciprocal variance is referred to as precision and is defined by:

$$\tau = 1/\sigma^2 \tag{3.8}$$

In this study the AE signal is extracted at different operational conditions (speeds, and dynamic loading conditions). A regression function, which measures the bearing vibration as a function of the different operating conditions is fitted. The regression function is approximated based on the parameter prior and the data-driven likelihood. The prior indicates the characteristic nature of the functions of interpolation. As such

the prior allows for more vigorous interpolation functions, particularly if only noisy and limited data are obtainable.

An observation y_i^j is given as the summation of the specific loading condition function as computed for the equivalent operational condition vector $f(x_i^j)$ and the noise term e_i

$$y_i^j = f(x_i^j) + e_i \tag{3.9}$$

The function of interpolation could be taken to an approximate linear dependency on x if the operating conditional vector is adequately expressive it may be necessary to make the assumption. The linearly dependent function is given by the parameter vector w^j:

$$f(x^j) = \{x^j\}^T w^j \tag{3.10}$$

For the least square error solutions for the reference loading condition, a multivariate Gaussian distribution is approximated. This distribution is consequently utilized as the prior distribution $p(w)$. Let the prior mean be taken as vector μ_o, and let the covariance matrix be taken as Σ_o. Hence, the prior is given as:

$$p(w) = N(w|\mu_o, \Sigma_o) \tag{3.11}$$

Based on Bayes's theorem, the prior and the data determined likelihood are utilized in obtaining a posterior distribution for the values of parameters:

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}$$

$$p(w^j \left| y^j, X^j \right.) = \frac{p(y^j \left| X^j, w^j \right.)(p(w^j))}{p(y^j \left| X^j \right.)} \tag{3.12}$$

where the posterior is normalized by the marginal likelihood $p(y \left| X \right.)$. Prior probability is the probability available before the observation. However, posterior probability is the probability obtained after the observation. The likelihood function shows the possibility of the data set observed for the settings of the vector of parameters. The posterior distribution could equally be demonstrated to be a Gaussian distribution (Bishop, 2006):

$$p(w^j \left| y^j, X^j \right.) = N(w^j \left| \mu_e^j, \Sigma_e^j \right.) \tag{3.13}$$

where the posterior mean μ_e and covariance Σ_e for loading condition j is given by:

$$\mu_e^j = \Sigma_e(\Sigma_o^{-1}\mu_o + \tau_e \left\{X^j\right\}^T X^j) \tag{3.14}$$

$$\Sigma_e^j = (\Sigma_o^{-1} + \tau_e \left\{X^j\right\}^T X^j)^{-1} \tag{3.15}$$

The likelihood of the observation of a DAI value y_* at an operational condition x_*^j while traversing bearing time interval j may be obtained from the recomputed likelihood function and is a type of a Gaussian (Bishop, 2006):

$$p(y_* \left| y^j, X^j, \mu_o, \Sigma_o, \sigma_e \right.) = N(y_* \left| w_e^T x_*^j, \left\{\sigma_*^j\right\}^2 \right.) \tag{3.16}$$

The variance $\left\{\sigma_*^j\right\}^2$ of the predictive distribution is indicative of the uncertainty in the prediction at an operational condition x_*^j defined as:

$$\left\{\sigma_*^j\right\}^2 = \sigma_e^{-2} + \left\{x_*^j\right\}^T \Sigma_n x_*^j \tag{3.17}$$

3.2.5 Gaussian mixture regression

In spite of the valuable analytical properties of the standard single Gaussian distribution it has some considerable limits in real data modelling. If a dataset forms more than one dominant clump, the basic Gaussian distribution is incapable of capturing the structure whilst the linear superposition of two or more Gaussians can give an improved description of the dataset. Such linear characterisation formed by taking a linear combination of more basic distributions such as Gaussians, can be formulated as probabilistic models known as mixture distribution (Bishop, 2006).

In this study a GMR is used to predict slow rotating bearing RUL from the DAI. Assume X represent the vector of the explanatory variables (e.g. operation conditions such as speed, time, load etc.). The explanatory variables are those variables which may have impact on the signal characteristics, but which is generally independent of the bearing condition. Y is the vector of the response or dependent variables (e.g. the DAI developed from extracted bearing features obtained from AE signal). x is the input training data $(x \in X)$ and y is the output data $(y \in Y)$. For the given x and y, the joint probability density is given as (Wang *et al.*, 2013).

$$f_{X,Y}(x,y) = \sum_{j=1}^{K} \pi_j \varphi(x,y;\mu_j,\Sigma_j) \tag{3.18}$$

where $\sum_{j=1}^{K} \pi_j = 1,\ \mu_j = \begin{bmatrix} \mu_{jx} \\ \mu_{jy} \end{bmatrix},\ \Sigma_j = \begin{bmatrix} \Sigma_{jXX} & \Sigma_{jXY} \\ \Sigma_{jYX} & \Sigma_{jYY} \end{bmatrix}$

The probability density function of the multivariate GMM is denoted by $\varphi(x,y;\mu_j,\Sigma)$. Equation (3.18) shows that the relationship between the explanatory variables and the response variable can be can be described by several GMM models. The parameters of Equation (3.18) include the number of the mixture components, K, the priors π_j, the mean value μ_j, and the variance of each Gaussian component Σ_j,

which are represented as $\theta = (\theta_1, \theta_2, ..., \theta_K)$ with $\theta_j = (\pi_j, \mu_j, \Sigma_j)$ and the constraint $\sum_{j=1}^{K} \pi_j = 1$.

As noted by each Gaussian component can be partitioned and the joint density can be rewritten as

$$f_{X,Y}(x, y) = \sum_{j=1}^{K} \pi_j \varphi(y|x; m_j(x), \sigma_j^2) \varphi(x; \mu_{jX}, \Sigma_{jX}) \tag{3.19}$$

Then marginal probability density of X is

$$f_X(x) = \int f_{X,Y}(x, y) dy = \sum_{j=1}^{K} \pi_j \varphi(x; \mu_{jX}, \Sigma_{jX}) \tag{3.20}$$

The conditional probability density function of $(Y|X)$ can deduced by combining Equation (3.19) and (3.20)

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)} = \sum_{j=1}^{K} w_j(x) \varphi(y; m_j(x), \sigma_j^2) \tag{3.21}$$

with the mixing weight

$$w_j(x) = \frac{\pi_j \varphi(x; \mu_{jX}, \Sigma_{jX})}{\sum_{j=1}^{K} \pi_j \varphi(x; \mu_{jX}, \Sigma_{jX})} \tag{3.22}$$

From Equation (3.22), the regression function for the prediction given a new input is

$$m(x) = E[Y|X = x] = \sum_{j=1}^{K} w_j(x) m_j(x) \tag{3.23}$$

and the conditional variance function is

$$v(x) = Var[Y|X = x] = \sum_{j=1}^{K} w_j(x)(m_j(x^2) + \sigma_j^2) - (\sum_{j=1}^{K} w_j(x)(m_j(x))^2 \tag{3.24}$$

where

$$m_j(x) = \mu_{jY} + \Sigma_{jYX} \sum_{jX}^{-1} (x - \mu_{jX}) \tag{3.25}$$

and

$$\sigma_j^2 = \Sigma_{jYY} - \Sigma_{jYX} \Sigma_{jX}^{-1} \Sigma_{jXY} \tag{3.26}$$

$m(x)$ in Equation (3.23) is the GMR model of index K, simply abbreviated as GMR(K) or $m(x;K)$. Although the regression function $m(x)$ from the joint mixture Gaussian density is of the form of a kernel estimator commonly used in nonparametric models, the weight function $w_j(x)$ is not determined by the local structure of the data, but by the components of a global GMM. Thus the GMR is a global parametric model with nonparametric flexibility (Wang *et al*., 2013).

A major task in fitting the GMR is the estimation of the parameters θ of GMM for the joint density $f_{X,Y}$. This can be achieved by maximizing the log likelihood function $L(\theta_k)$ denoted as

$$L(\theta_k) = \ln \prod_{i=1}^{N} p(x_i, y_i) = \sum_{i=1}^{N} \ln \sum_{j=1}^{K} \pi_j \varphi(x, y; \mu_j, \Sigma_j) \tag{3.27}$$

For the given training data, the parameters θ (comprising the means, covariances and missing coefficients) of a GMM is learnt by maximizing Equation (3.27) using the expectation maximization algorithm in the iterative means (Nabney, 2002).

The expectation maximization algorithm includes two steps:

1. Expectation step:

Calculate the posterior probability according to

$$p(j|X) = \frac{\pi_j \varphi(X, \mu_j \Sigma_j)}{p(X, \theta)} \qquad \text{with } j = 1,2,...k \tag{3.28}$$

2. Maximum step:

$$\pi_j^{New} = \frac{1}{N}\sum_{j=1}^{N}\frac{\pi_j^{old}\varphi(X,\mu_j,\Sigma_j)}{} = \frac{1}{N}\sum_{i=1}^{N} p(j|x_j) \tag{3.29}$$

$$\mu_j = \frac{1}{\pi_j N}\sum_{i=1}^{N} p(j|x_i)x_i \tag{3.30}$$

$$\Sigma_j = \frac{1}{\pi_j N}\sum_{i=1}^{N} p(j|x_i)[(x_i - m_j)(x_i - m_j)^T] \tag{3.31}$$

It is convenient to recast the maximising problem in the equivalent form of minimising the NLL of the data set (see (Equation 2.24) (Nabney, 2002).

The two steps are iterated until the model converges to a local minimum (Calinon, 2009). The entire data set is divided into training and test sets. The training set is used in estimating the parameters of the GMM while the test set is kept for prediction of bearing damage and RUL. The results obtained may be highly sensitive to the number of mixing components used. The more components a mixture model has the more expressive and flexible it becomes. A sufficiently expressive model may be optimized so as to accurately represent the reference signal (see chapter 2, paragraph 2.5). Given the test set, the GMR models can be obtained using the parameters of the GMM which has an output of a smoothened general description of the GMM encoded data and linked constraints given by matrices of the covariance (Calinon, 2009). This general smoothened description of the data is the prediction of the failure of the bearing.

3.2.6 Gaussian process regression

The use of GPR for prognostics (prediction of the RUL of slow rotating bearings based on the DAI is possible. Gaussian processes (GP) are a recent development in nonlinear nonparametric modelling. In GP, the parametric model is dispensed and instead a prior probability distribution is defined over functions directly (Bishop,

2006). A nonlinear functional mapping from an inputting space to a target space is achieved by the use of GP modelling. The GP is defined as an infinite collection of arbitrary variables of which any of the fixed subsets has joint Gaussian distributions. The Gaussian method is favourable to smooth functions and those that properly explain the training data. The smooth attribute of the function leads to its plausible generalisations (Heyns *et al*., 2012a).

To motivate the GP viewpoint, let the vector $\mathbf{x}_n$ represent the DAI in the input space. The training set of inputting vectors $\mathbf{X}_N \equiv \{\mathbf{x}^n\}_{n=1}^N$ corresponds to the targeted vector $\mathbf{y}_N \equiv \{y^n\}_{n=1}^N$. For prognostics as in this study, x is the time period while y is a novel DAI for monitoring the health state of slow rotating bearings. A Gaussian process $f(\mathbf{x})$ can be fully described by its mean and covariance (or kernel) function (Rasmussen and Williams, 2006). These functions are specified separately, and consist of a specification of a functional form as well as a set of parameters called hyperparameters.

The mean function describes the value of the function expected at any point of the input space, before the consideration of any trained data. The mean function can be defined as:

$$m(\mathbf{x}) = E(f(\mathbf{x})) \tag{3.32}$$

In supervised learning, the idea of likeness linking the various data points is vital. It is an essential similarity assumption that the points of inputs $\mathbf{x}$ which are close and likely to have similar target values y are expected to be similar. Hence, training points which are close to a test point prediction should be insightful about the prediction at that point. In the Gaussian process viewpoint, the covariance function defines the nearness or similarity (Rasmussen and William, 2006). The covariance function between two functional values evaluated at fixed points $\mathbf{x}$ and $\mathbf{x}'$ is given as

$$k(\mathbf{x}, \mathbf{x}') = E\left[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\right] \tag{3.33}$$

The covariance function enables the inference value of a function given the knowledge of the other. Thus, the covariance function $k(\mathbf{x},\mathbf{x}')$ can be interpreted as the measure of the distance between the input points $\mathbf{x}$ and $\mathbf{x}'$. The Gaussian process can then be written as:

$$f(\mathbf{x}) \sim GP[m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')] \tag{3.34}$$

The basic GPR consists of a simple zero mean and SE covariance functions. The zero mean function is given as:

$$m(\mathbf{x}) = 0 \tag{3.35}$$

for every value of $\mathbf{x}$. One of the generally used kernel functions is the SE. It assumes that the function values in close proximity in the feature space are probably going to be similar, with close to unity covariance for variables that have feature inputs that are close. The SE covariance function with ARD is given as (Rasmussen and Williams, 2006):

$$k_{SEard}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x}')^{\mathrm{T}} M (\mathbf{x}-\mathbf{x}')) \tag{3.36}$$

where M matrix is diagonal with positive ARD parameters, and $M = diag(l)$, where l is a vector of length D corresponding to the input space dimension. The characteristic length-scale parameters, also known as the ARD parameters, determine the rate of variation of the function in the direction of the corresponding inputting space. A function tends to vary faster for any variation of its component feature for its shorter length scale parameter for a specific feature component. A short length scale thus corresponds to high relevance. σ_f^2 is the signal variance linked to the general function variance.

The free parameters (i.e. hyperparameters) in the covariance function can be compressed into a matrix denoted by φ. The values of these hyperparameters are all unknown and inference is made from the trained data. It can be shown by utilization of the Bayes' rule that the maximum a posteriori hyperparameter values φ can be

obtained by maximising the marginal likelihood $p(\mathbf{y}|\mathbf{X},\varphi)$ which is the same as minimising the negative log marginal likelihood (Rasmussen and Williams, 2006):

$$\log P(\mathbf{y}|\mathbf{X},\varphi) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}}(\mathbf{K}(\mathbf{X},\mathbf{X})^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}(\mathbf{X},\mathbf{X})| - \frac{n}{2}\log 2\pi \tag{3.37}$$

where $\mathbf{K}(\mathbf{X},\mathbf{X})$ is the $N \times N$ covariance matrix between all pairs of training inputs and is computed with Equation (3.5) or (3.6). It is important to note however, that the application in this study used an $N \times 1$ matrix of time points.

Given a set of training points, one can derive the posterior distribution over functions by imposing a restriction on prior joint distribution. Once a posterior distribution is derived, it can be used to estimate predictive values for the test data points (Saxena *et al.,* 2009). Denoting $\mathbf{X}$ as the training inputs and $\mathbf{X}_*$ as the test inputs, prediction of $\mathbf{y}_*$ at the new locations $\mathbf{X}_*$ may be inferred by conditioning the joint distribution on the observed target values. For the basic GPR with zero mean, the following equations describe the predictive distribution (Rasmussen and Williams, 2006):

Prior:

$$\begin{bmatrix}\mathbf{y}\\ \mathbf{y}_*\end{bmatrix} \sim N\left(0, \begin{bmatrix}K(\mathbf{X},\mathbf{X}) & K(\mathbf{X},\mathbf{X}_*)\\ K(\mathbf{X}_*,\mathbf{X}) & K(\mathbf{X}_*,\mathbf{X}_*)\end{bmatrix}\right) \tag{3.38}$$

Posterior:

$$P(\mathbf{y}_*|\mathbf{X},\mathbf{y},\mathbf{X}_*) \sim N(\bar{\mathbf{y}}_*, \mathrm{cov}(\mathbf{y}_*)) \tag{3.39}$$

where

$$\bar{\mathbf{y}}_* = E[\mathbf{y}|\mathbf{X},\mathbf{y},\mathbf{X}_*) = \mathbf{K}(\mathbf{X}_*,\mathbf{X})\mathbf{K}(\mathbf{X},\mathbf{X})^{-1}\mathbf{y} \tag{3.40}$$

$$\mathrm{cov}(\mathbf{y}_*) = \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})\mathbf{K}(\mathbf{X},\mathbf{X})^{-1}\mathbf{K}(\mathbf{X},\mathbf{X}_*) \tag{3.41}$$

The maximum a posteriori estimates $\bar{y}_*$ can then be used as slow rotating bearing RUL metrics.

3.3 Model evaluation

The models would be evaluated using MAPE and RMSE. The MAPE and RMSE between the predicted and the original DAI were calculated using Equations (3.42) and (3.43) respectively.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{DAI_i - \hat{DAI}_i}{DAI_i}\right| \times 100 \tag{3.42}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(DAI - \hat{DAI}\right)^2}{n}} \tag{3.43}$$

where DAI_i is the actual value of the DAI for the i^{th} observation which is in this case the time point, $\hat{DAI}_i$ is the predicted value of DAI, and n is the number of observations.

The leave-one-out CV technique was used in selecting the test set for validating the predictions from each model. Two approaches were considered. The first is based on dependent samples. In this approach the bearing data set was divided into equal samples of training and test sets. The training set is the "seen" because it was used in training the parameters of the model while test set is the "unseen" as it was never fed into the model during training. However, it is been argued that when the training and test samples are dependent, the two errors may be positively correlated, resulting in a breakdown of the CV selection approach and can equally lead to model overfitting (Opsomer *et al.*, 2001; Arlot, 2010). Therefore, the second approach is based on independent samples whereby, two different sets of bearings are trained together and hence used as the training set while a third bearing data is used as the test set.

3.4 Results and discussions

3.4.1 Prediction based on dependent samples

The predictions in this section are based on dependent observations whereby the training and the test sets are obtained using a leave-one-out CV technique that involves the division of the bearing data set into equal samples of training and test sets.

The health state of a bearing is divided into three during its whole life namely, healthy or normal state, slightly degraded state and failure state. There is no need for RUL when a bearing is in its healthy state. When the computed features are above their incipient damage threshold values then it is considered that slight degradation has set in. The prediction model is then used in the prediction of the future value of the DAI.

In this investigation healthy bearings are run until they are failed. A DAI is developed to assess the degradation of the slow rotating bearing. The DAI is then used in the several regression models, namely, the MLP, RBF, BLR, GMR and the GPR models for prediction of bearing damage, RUL and failure at a future instant of time.

Secondly, the best performing model was evaluated and selected using the MAPE and RMSE. Thirdly, the best performing model from the resulting novel methodologies is recommended and used for the RUL prediction of slow rotating bearing.

3.4.1.1 RUL using MLP regression

Multi-layer perceptron is one of the most frequently used feedforward artificial neural networks which make use of a supervised learning algorithm. Essentially, it has three layers which include the input layer, hidden layer and output layer (Şengüler *et al.,* 2010).

Dimensionality of the feature vectors were reduced to 2 PKPCs from 5 bearing extracted features, using polynomial kernel principal components analysis (PKPCA) which subsequently fed into the GMM to obtain the DAI. The MLP neural network was trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE between the predicted and the actual DAI are shown in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 respectively. The MLP neural network approach was used to monitor the trend of the incipient bearing damage and RUL of Bearings 1, 2 and 3 as shown at the top right hand corner of Figures 3.5, 3.6 and 3.7 respectively.

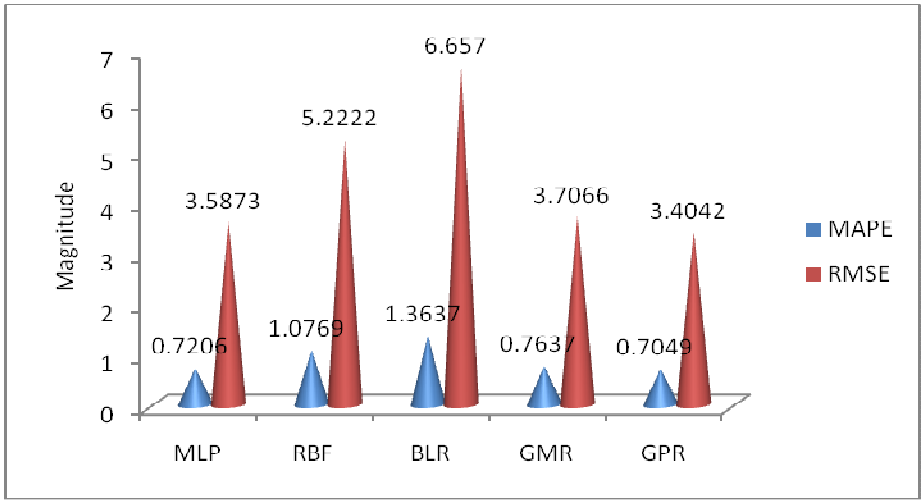

Figure 3.2: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 1 based on the dependent samples

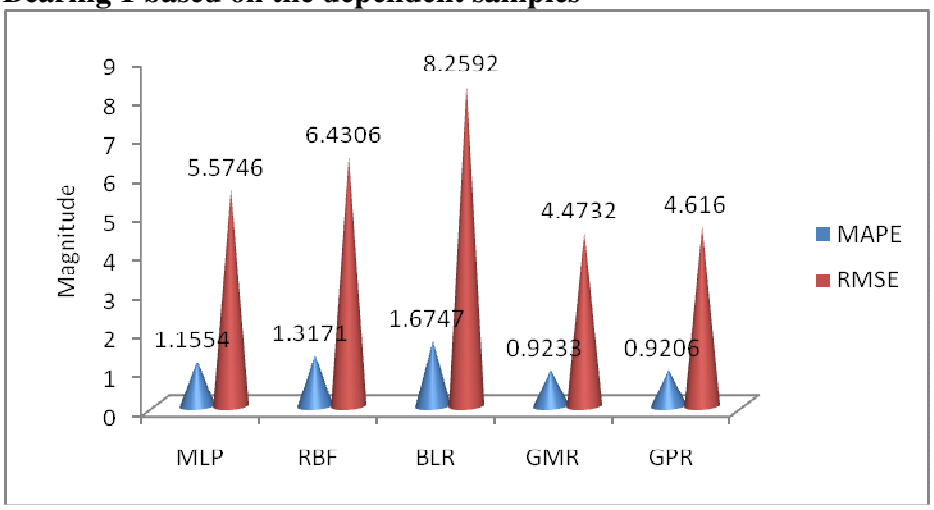

Figure 3.3: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 2 based on the dependent samples

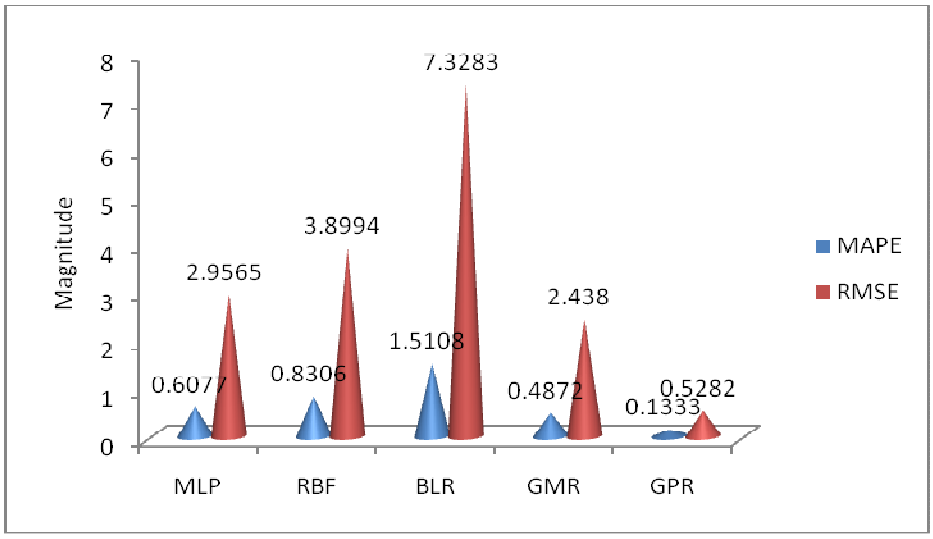

Figure 3.4: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 3 based on the dependent samples

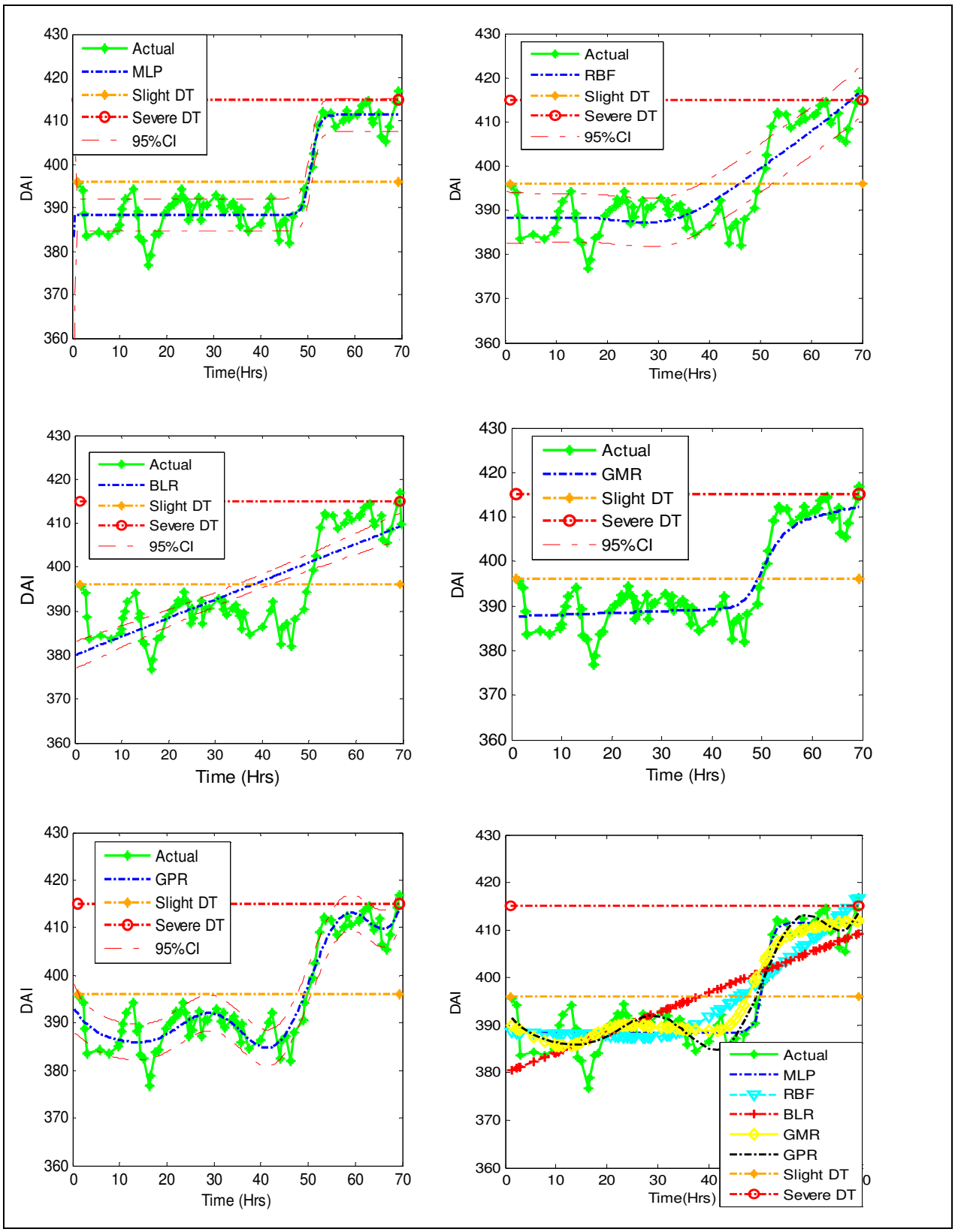

Figure 3.5: Prediction for the whole lifespan of Bearing 1 using different methodologies based on dependent samples

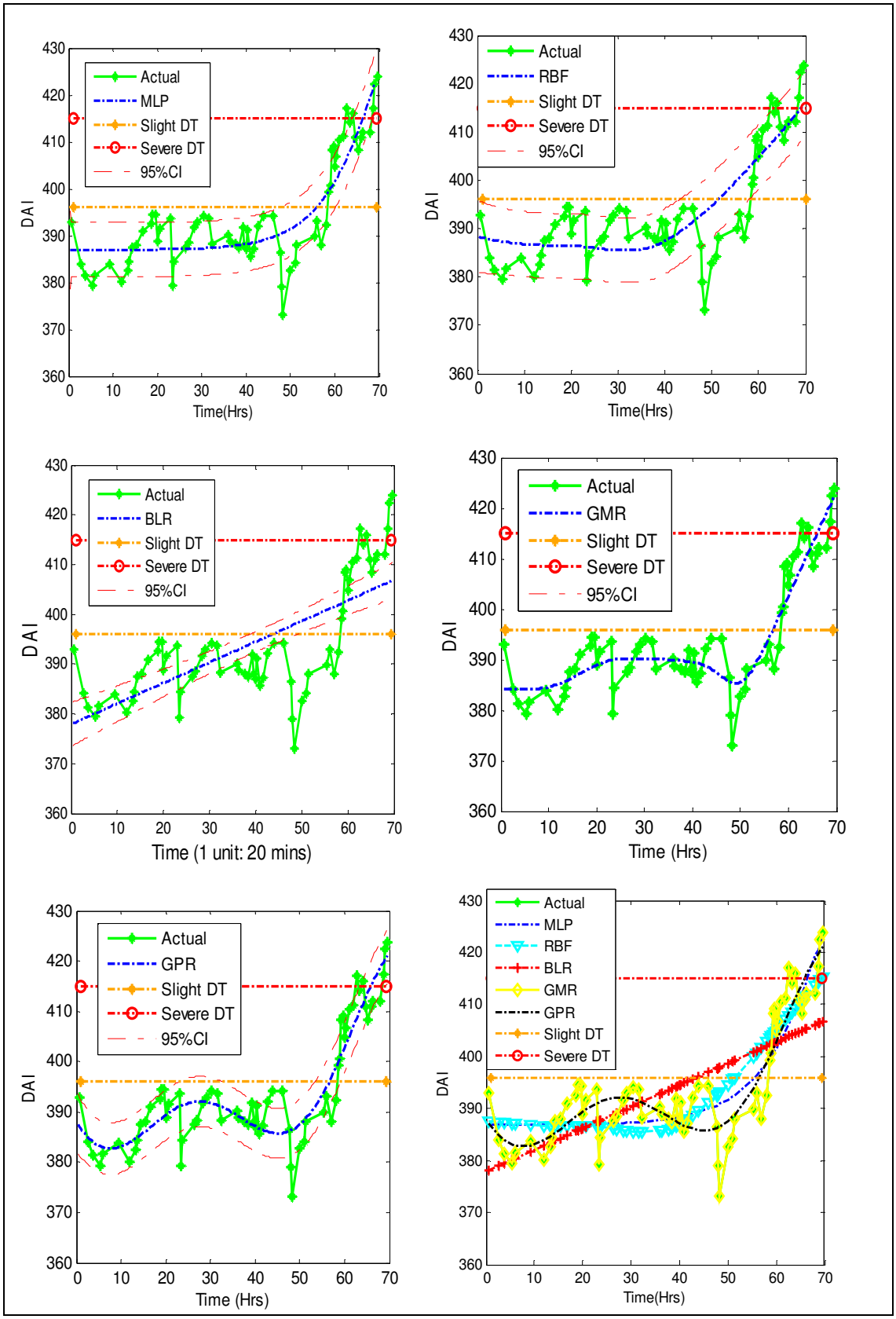

Figure 3.6: Prediction for the whole lifespan of Bearing 2 using different methodologies based on dependent samples

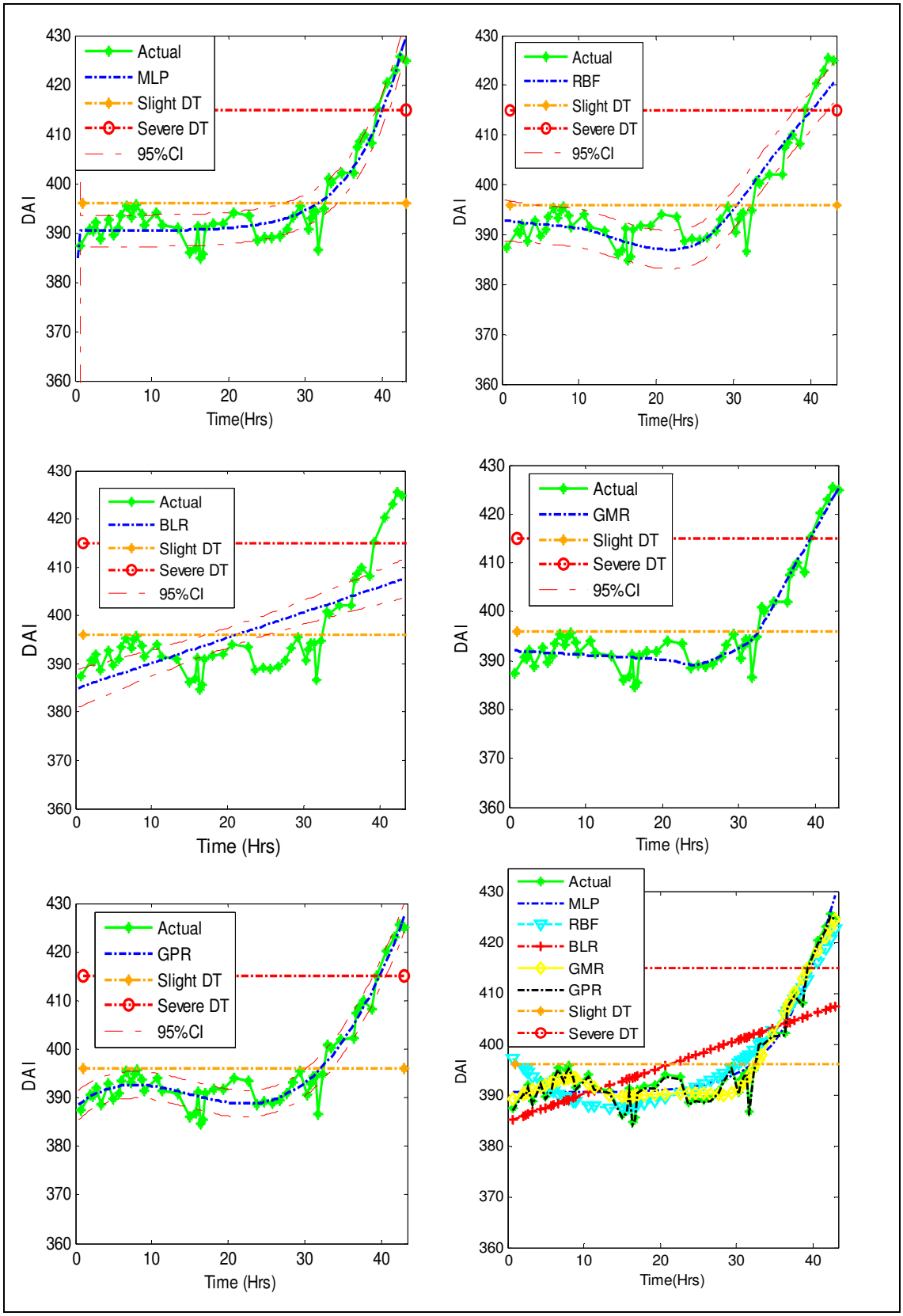

Figure 3.7: Prediction for the whole lifespan of Bearing 3 using different methodologies based on dependent samples

3.4.1.2 RUL using RBF regression

The RBF uses local hyper-sphere surfaces (nonlinear mapping) to separate the classes in the input space as a response to cluster, rather than the global hyper-planes (lines) used in MLP networks (Al-Raheem and Abdul-Karem, 2010).

The DAI was used as input into the RBF. The RBF was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the output of the trained network and the original or actual DAI was calculated using Equations (3.43) and (3.44) respectively and shown in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 respectively. The RBF predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as shown at the top right hand corner of Figures 3.5, 3.6 and 3.7 respectively.

3.4.1.3 RUL using BLR

Similarly, the DAI was used as input into the BLR. The BLR was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the output of the trained network and the original DAI was calculated using Equations (3.43) and (3.44) respectively and shown in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 respectively. The BLR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were afterwards plotted as presented in the middle left hand of Figures 3.5, 3.6 and 3.7 in turn.

3.4.1.4 RUL using GMR

Furthermore, the DAI was used as input into the GMR. The GMR was then trained with the DAI which had been obtained from the bearing data at dynamic loadingconditions. The MAPE and RMSE were again computed between the output of the trained network and the original DAI was calculated using Equations (3.43)

and (3.44) respectively and shown in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 respectively. The GMR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were then plotted as shown in the middle right hand of Figures 3.5, 3.6 and 3.7 correspondingly.

3.4.1.5 RUL using GPR

The DAI was used as input into the GPR. The GPR was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the output of the trained network and the original DAI was calculated using Equations (3.43) and (3.44) respectively and shown in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 respectively. The GPR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as presented in the bottom left hand of Figures 3.5, 3.6 and 3.7 respectively.

3.4.1.6 Model evaluation of the dependent samples

After the training process, the prediction was done with data points equal to the training data points. The predicted and actual RUL plots of all the models were plotted in the bottom right hand of Figures 3.5, 3.6 and 3.7 respectively.

The MAPE and RMSE between the output and real values are observed as plotted in Figures 3.2, 3.3 and 3.4 for Bearings 1, 2 and 3 for all the five models namely: MLP, RBF, BLR, GMR and GPR. All the models attempted to predict damage and RUL to a great degree.

For Bearing 1, the MAPE from the models was 0.7049, 0.7206, 0.7637, 1.0769 and 1.3637 for GPR, MLP, GMR, RBF, and BLR respectively for Bearing 1 from the least to the highest PEs. Similarly, the RMSE PE from the models was 3.4042, 3.5873, 3.7066, 5.2222 and 6.6570 for GPR, MLP, GMR, RBF, and BLR respectively from the least to the highest PEs. The worst prediction was that of the

BLR which was a linear line across the nonlinear model. The GMR and MLP also modelled damage and RUL with little error. However, the best predictive model for Bearing 1 was the GPR. For Bearings 2 and 3 similar results were obtained with the best predictive model being the GPR.

It could be seen that the GPR model consistently had the least error for all the three bearings. It was therefore concluded that the GPR model predicts damage and RUL better than the other models.

To simplify the discussion of RUL predictions, Table 3.1 reports the values for the onset of incipient damage and those of the final failure along with the RUL computed by taking the difference between the two values. These are reported for both actual and predicted RUL. The prediction accuracy computed by comparing the actual RUL and predicted RUL is reported in the last column of Table 3.1 for each test bearing. For Bearing 1, the GPR model produced the highest RUL prediction accuracy (100%). For Bearing 2, MLP, GMR and GPR all predicted the RUL with 100% accuracy while RBF predicted with only 40% accuracy. For Bearing 3, GMR and GPR predicted the RUL with about 83% accuracy while RBF predicted with about 33%. Overall, the GPR model appears to be the best predictor of RUL for the slow rotating bearings tested in this study under the dependent samples.

Table 3.1: RUL prediction using dependent samples

Test set		Actual			Predicted			Accuracy
Bearing 1	Model	Incipient	Failure	**RUL**	Incipient	Failure	**RUL**	
	MLP	50	69	19	51	-	-	-
	RBF	50	69	19	45	68	23	78.95
	BLR	50	69	19	38	-	-	-
	GMR	50	69	19	47	-	-	-
	GPR	50	69	19	50	69	19	100.00
Bearing 2	MLP	58	68	10	56	66	10	100.00
	RBF	58	68	10	52	68	16	40.00
	BLR	58	68	10	43	-	-	-
	GMR	58	68	10	56	66	10	100.00
	GPR	58	68	10	56	66	10	100.00
Bearing 3	MLP	33	39	6	32	40	8	-
	RBF	33	39	6	32	42	10	33.33
	BLR	33	39	6	22	-	-	-
	GMR	33	39	6	33	40	7	83.33
	GPR	33	39	6	32	39	7	83.33

Note: - indicates that the model could not predict final failure and hence RUL cannot be determined

3.4.2 Predictions based on independent samples

The predictions in this section are based on independent observations whereby two different sets of bearings are trained together and hence used as the training set while a third bearing data is used as the test set.

3.4.2.1 RUL using MLP regression

The MLP neural network was trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE between the predicted and the actual DAI are shown in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 respectively. The MLP neural network approach was used to monitor the trend of the incipient bearing damage and RUL of Bearings 1, 2 and 3 as shown at the top right hand corner of Figures 3.11, 3.12 and 3.13 respectively.

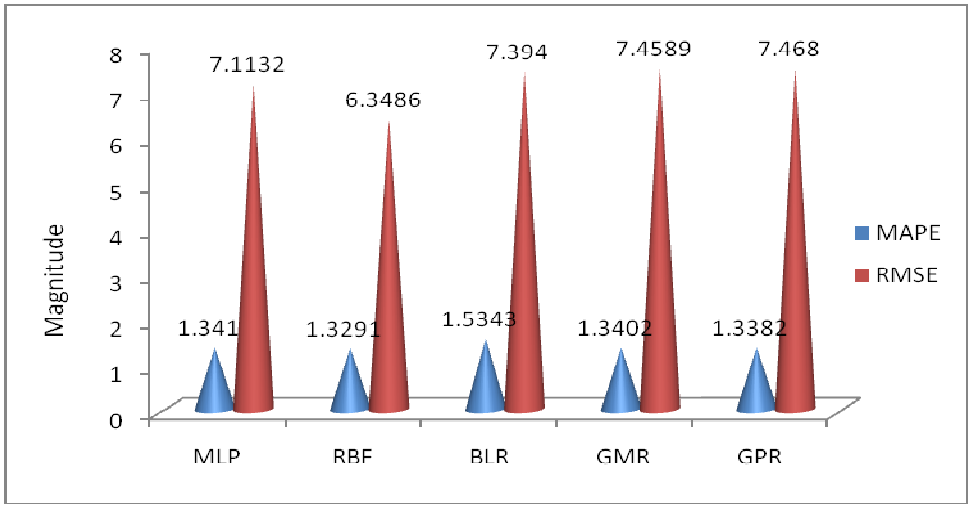

Figure 3.8: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 1 based on the independent samples

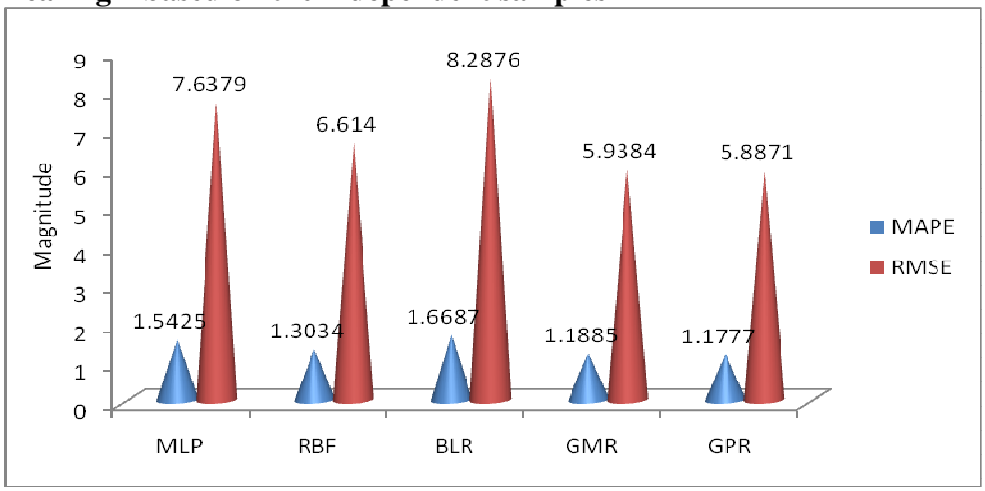

Figure 3.9: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 2 based on the independent samples

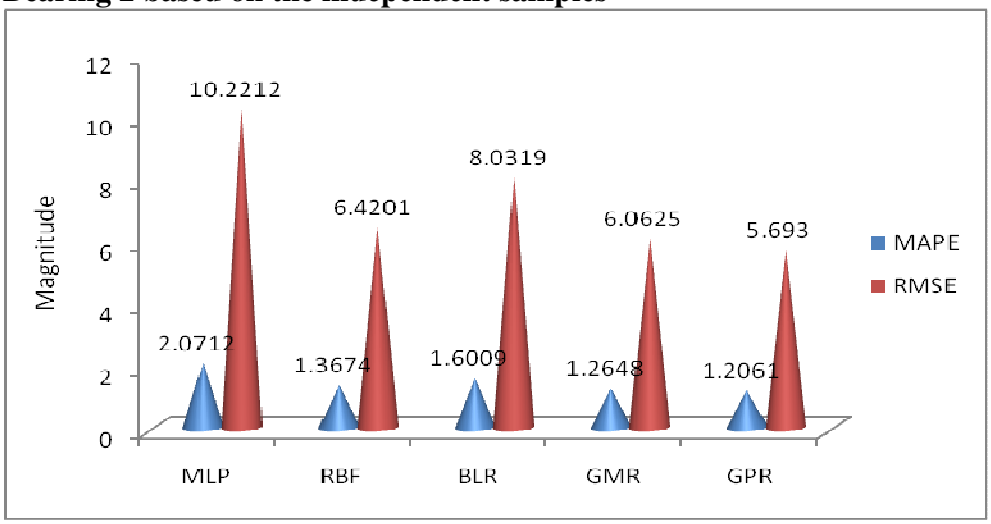

Figure 3.10: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 3 based on the independent samples

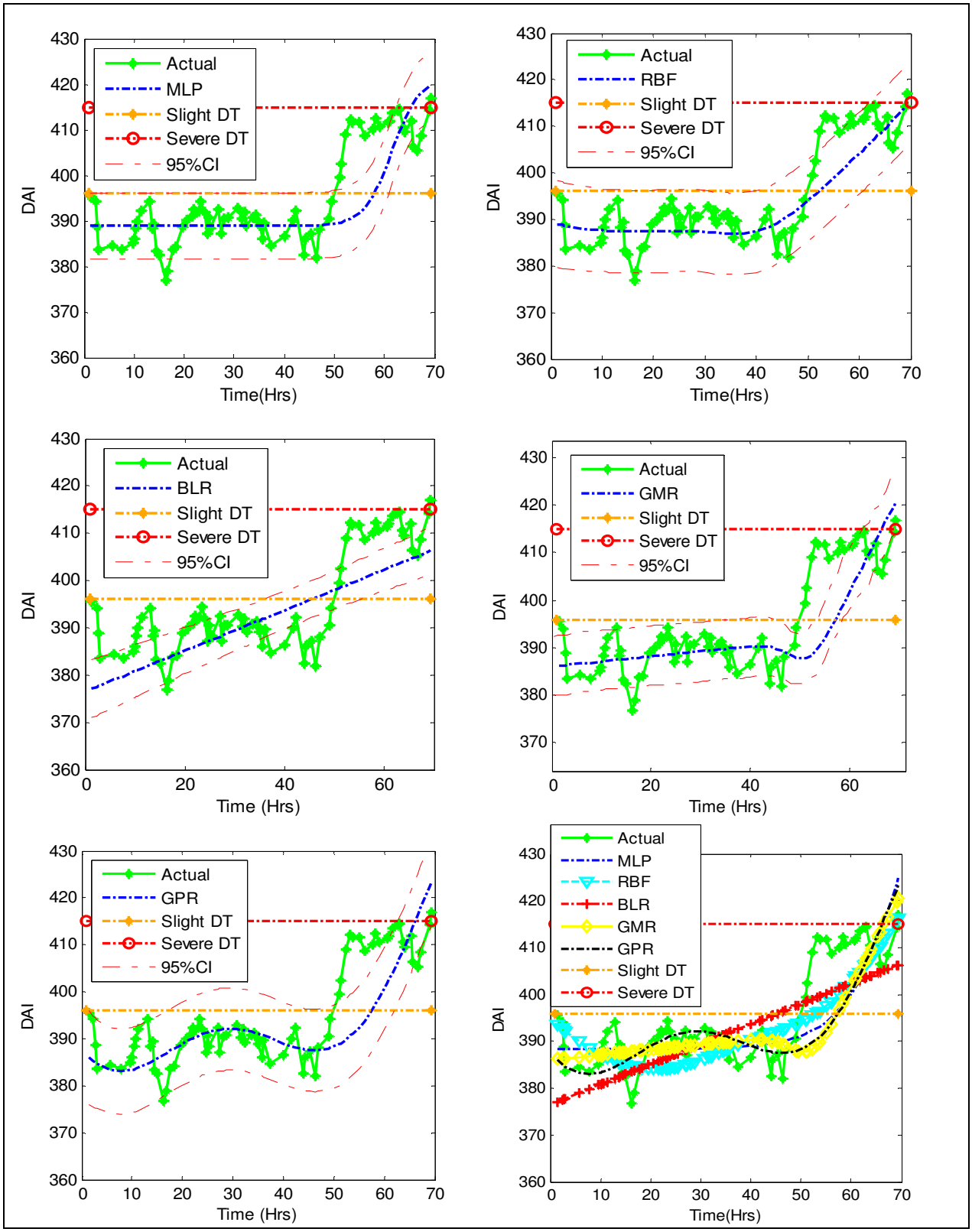

Figure 3.11: Prediction for the whole lifespan using Bearings 2 and 3 as training set and Bearing 1 as test set based on different methodologies and independent samples

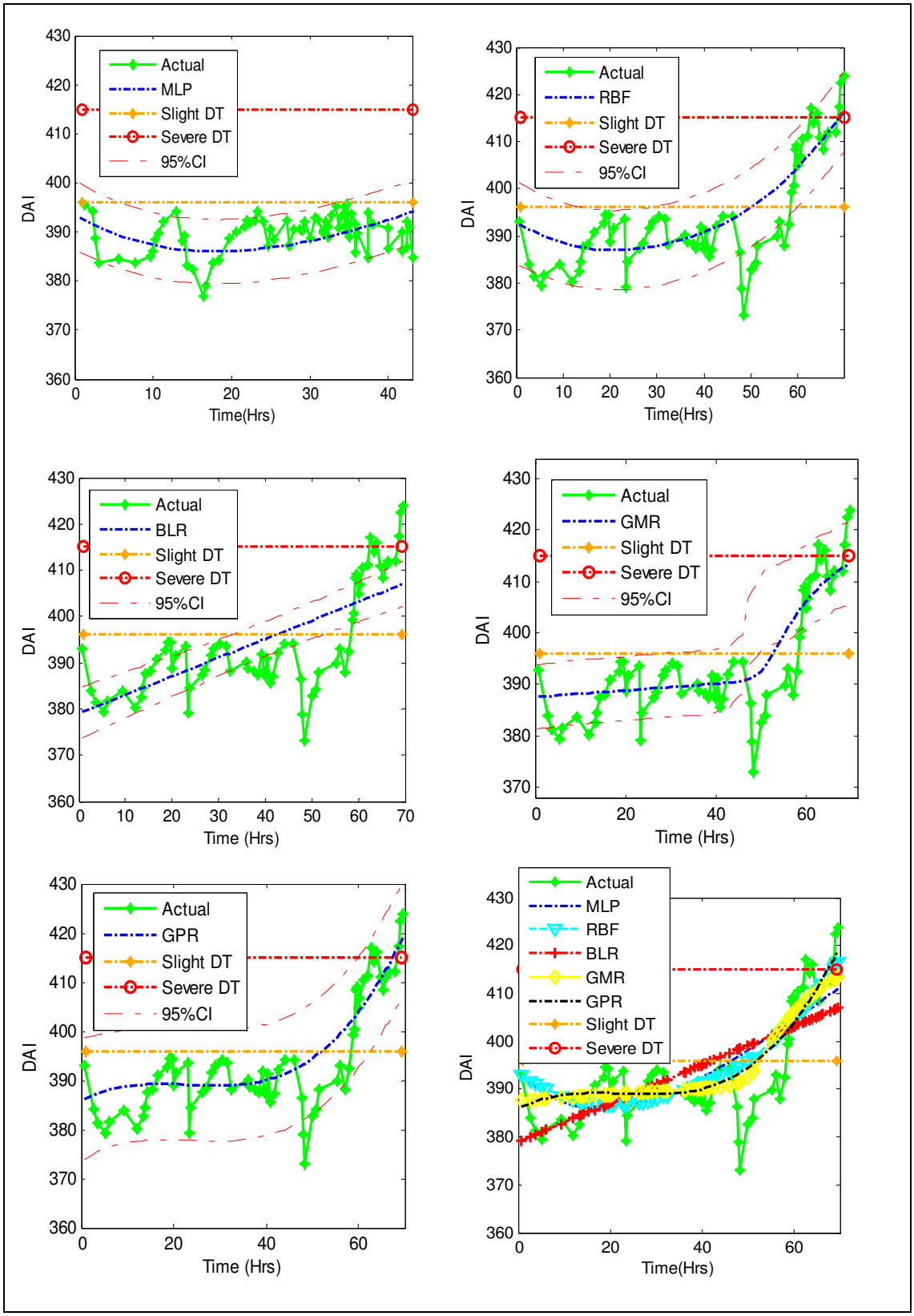

Figure 3.12: Prediction for the whole lifespan using Bearings 1 and 3 as training set and Bearing 2 as test set based on different methodologies and independent samples

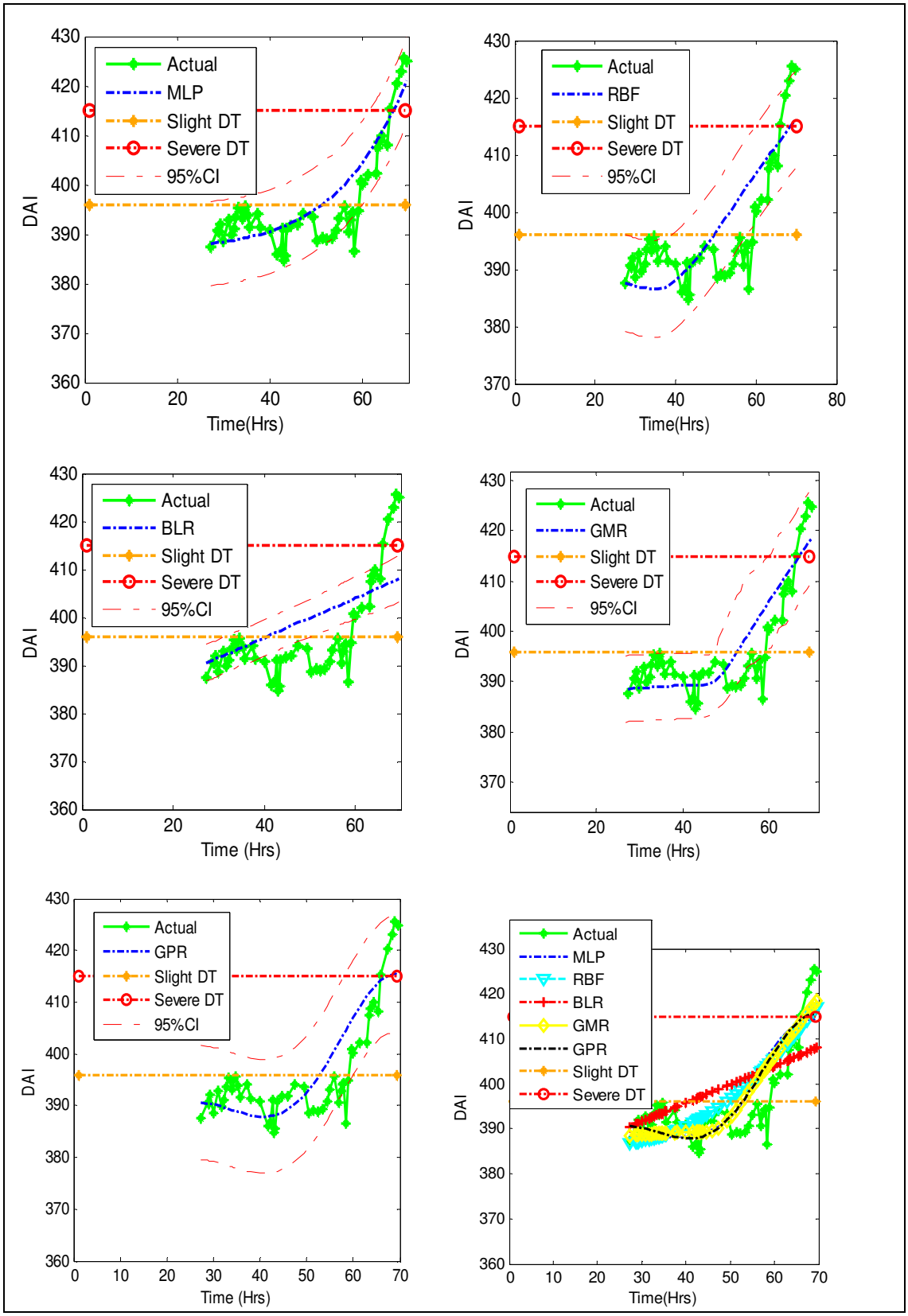

Figure 3.13: Prediction for the whole lifespan using Bearings 1 and 2 as training set and Bearing 3 as test set based on different methodologies and independent samples

3.4.2.2 RUL using RBF regression

The DAI was used as input into the RBF. The RBF was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the predicted and the original DAI and shown in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 respectively. The RBF predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were afterwards plotted as presented at the top right hand corner of Figures 3.11, 3.12 and 3.13 in that order.

3.4.2.3 RUL using BLR

Similarly, the DAI was used as input into the BLR. The BLR was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the predicted and the original DAI and shown in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 respectively. The BLR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as shown in the middle left hand of Figures 3.11, 3.12 and 3.13 respectively.

3.4.2.4 RUL using GMR

Furthermore, the DAI was used as input into the GMR. The GMR was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI and shown in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 respectively. The GMR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were then plotted as presented in the middle right hand of Figures 3.11, 3.12 and 3.13 in turn.

3.4.2.5 RUL using GPR

The DAI was used as input into the GPR. The GPR was then trained with the DAI which had been obtained from the bearing data at dynamic loading conditions. The MAPE and RMSE were again computed between the predicted and the actual bearing DAI and shown in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 respectively. The GPR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were afterwards plotted as shown in the bottom left hand of Figures 3.11, 3.12 and 3.13 in that order.

3.4.2.6 Model evaluation based on independent samples

After the training process using Bearings 2 and 3; Bearings 3 and 1; Bearings 1 and 2 as training dataset; the prediction was done with data points equal to the training data points from Bearings 1, 2 and 3 respectively. The predicted and actual RUL plots of all the models were plotted in the bottom right hand of Figures 3.11, 3.12 and 3.13 respectively.

The MAPE and RMSE between the output and real values are observed as plotted in Figures 3.8, 3.9 and 3.10 for Bearings 1, 2 and 3 for all the five models namely: MLP, RBF, BLR, GMR and GPR. All the models attempted to predict damage and RUL to a great degree.

For Bearing 1, using the independent approach the MAPE from the models were 1.3291, 1.3382, 1.3402, 1.341 and 1.5343 for RBF, GPR, GMR, MLP and BLR respectively for Bearing 1 from the least to the highest PEs. However, the RMSE PEs from the models were 6.3486, 7.1132, 7.394, 7.4589 and 7.468 for RBF, MLP, BLR, GMR and GPR respectively from the least to the highest PEs. The worst prediction was that of the GPR model. The best predictive model for Bearing 1 was the RBF. However, for Bearings 2 and 3 the best predictive model was the GPR.

Overall, the GPR model had the least error for all the three bearings. It was therefore concluded that the GPR model predicts damage and RUL better than the other models.

Table 3.2 reports the actual values of the incipient damage, final failure and RUL as well the prediction accuracy for each test bearing. For Bearing 1, the RBF model predicted the RUL better than all the other models with a prediction accuracy of about 84%. This is followed by GPR and GMR with an RUL prediction accuracy of about 53%. For Bearing 2, the GPR model performed far better than the rest models in predicting the RUL with a prediction accuracy of 80%. For Bearing 3, although the various prediction accuracy values are negative due to the wide difference between the actual incipient and predicted incipient damage, the GPR still produced a better prediction of the RUL than the rest of the models.

Table 3.2: RUL prediction using independent samples

Test set		Actual			Predicted			Accuracy
Bearing 1	Model	Incipient	Failure	**RUL**	Incipient	Failure	**RUL**	
	MLP	50	69	19	56	64	8	42.11
	RBF	50	69	19	52	68	16	84.21
	BLR	50	69	19	42	-	-	-
	GMR	50	69	19	56	66	10	52.63
	GPR	50	69	19	55	65	10	52.63
Bearing 2	MLP	58	68	10	-	-	-	-
	RBF	58	68	10	50	68	18	20.00
	BLR	58	68	10	43	-	-	-
	GMR	58	68	10	52	70	18	20.00
	GPR	58	68	10	54	66	12	80.00
Bearing 3	MLP	33	39	6	25	41	16	-
	RBF	33	39	6	23	43	20	-133.33
	BLR	33	39	6	15	-	-	
	GMR	33	39	6	27	41	14	-33.33
	GPR	33	39	6	28	41	13	-16.67

3.5 Comparison of model performance based on dependent and independent samples

The ranks of each prediction model according to whether the training and tests samples are dependent or independent are presented in Tables 3.3, 3.4 and 3.5 for Bearings 1, 2 and 3 respectively. While some of the models are sensitive to the type of sample, others are not. For example the GPR and BLR models ranked mainly 1st and last respectively in both the dependent and independent samples (see paragraph 1.8.7) whereas RBF and MLP ranked differently. However, the errors obtained from the CV based on independent samples (see Figures 3.8 to 3.10) were relatively larger than those from the dependent samples (see Figures 3.2 to 3.4) which could be an indication that the latter slightly overfitted the models.

Table 3.3: Ranking of models based on dependent and independent samples for Bearing 1

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	1	2	5
GMR	3	3	3	4
RBF	4	4	1	1
MLP	3	2	4	2
BLR	5	5	5	3

Table 3.4: Ranking of models based on dependent and independent samples for Bearing 2

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	2	1	1
GMR	2	1	2	2
RBF	4	4	3	3
MLP	3	3	4	4
BLR	5	5	5	5

Table 3.5: Ranking of models based on dependent and independent samples for Bearing 3

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	1	1	1
GMR	2	2	2	2
RBF	4	4	3	3
MLP	3	3	5	5
BLR	5	5	4	4

3.6 Summary

This chapter proposes a novel approach to damage detection and prediction of RUL of slow rotating bearings. During this investigation, three healthy slow rotating bearings were run to the point of failure. The DAI which was obtained by the integration of PKPCA, GMM and EWMA was used in slow rotating bearing prognostics. The slight and severe DTs are obtained through the use of the KDE technique on the healthy and slightly degraded bearing data respectively.

The DAI is used in the prediction of bearing damage and RUL using the MLP, RBF, BLR, GMR and GPR models respectively. Predictions were obtained using test and training sets from both dependent and independent samples. The models were able to predict damage and RUL of the slow rotating bearing. Overall, the GPR had the least MAPEs and RMSEs in damage prediction in this investigation for the slow rotating bearings. The GPR has also performed better than all the other models with respect to RUL prediction and this is robust to dependent and independent samples under varying operating conditions. Hence, the GPR is chosen as the most efficient model for prediction of RUL of slow rotating bearings. This proposed approach is useful and its application can be extended to the CM of other mechanical and allied systems.

Chapter 4 An integrated GPR for the RUL prediction of slow rotating bearings based on AE

4.1 Introduction

This chapter implements GPRs. The choice of GPR is motivated by a number of findings from previous research. The motivation for investigating an optimal GPR is based on the findings of chapter 3 which showed the standard GPR to have outperformed four other Bayesian models namely: BLR, GMR, MLP neural network and RBF neural network for prognostics of slow rotating bearings.

GPR RUL prediction could be improved by considering the mean and covariance functions simultaneously (see chapter 1, paragraph 1.8.5). This chapter proposes an integrated GPR model for prediction of RUL of slow rotating bearings combining the advantages from the individual mean and covariance functions. Moreover, the proposed technique consists of a construction of composite mean and composite covariance functions. These are respectively formed from two simple mean functions and three covariance functions selected based on acceptable scientific criteria rather than ad hoc choice as the later may lead to model misspecification. This procedure ensures that only the mean and covariance functions which are well-suited for the problem at hand are combined in order to obtain an optimal GPR model for prognostics of slow rotating bearings. This study deviates from previous studies that mostly implement a GPR using only the covariance function while assuming the mean function is equal zero (see chapter 1, paragraph 1.8.5). Therefore, in this chapter a novel integrated GPR-based model, where a scientifically selected composite mean function is simultaneously estimated with composite covariance function enhancing a more flexible and accurate prediction of RUL of slow rotating bearings is implemented. Further, this study is investigated using dependent and independent data samples (see chapter 1, paragraph 1.8.7).

4.2 Methodology

GP is utilized in the modelling of nonlinear mappings of functions to a target space from an input space (see chapter 1, paragraph 1.8.5). The scheme used in showing the overview of the integrated GPR prognostic or RUL procedure for the slow rotating bearing is shown in Figure 4.1. The scheme involves five steps: (1) the estimation of simple GPR models based on simple mean and simple covariance functions; (2) selection of 2 best simple mean and 3 best simple covariance functions using MAPE and RMSE; (3) development of integrated GPR models based on the 2 best mean and 3 best covariance functions; (4) GPR model selection and evaluation which basically involves a comparison of the best simple GPR models and the integrated GPR models and the selection of the most optimal model (5) RUL prediction based on the overall optimal model. The DAI earlier developed in chapter 2 is used as an input into the various GPR(s) in the prognostics of slow rotating bearings.

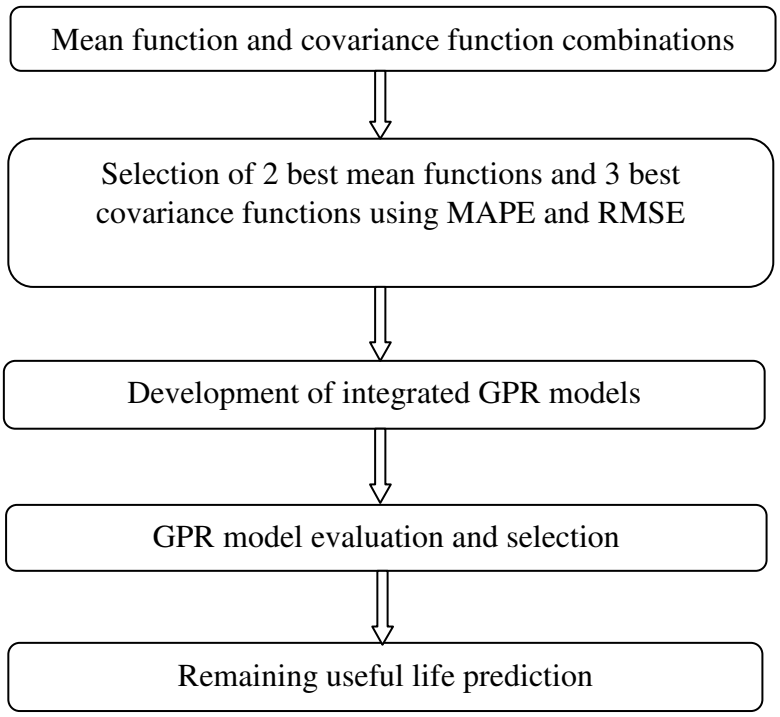

Figure 4.1: Overview of integrated GPR development and RUL prediction

4.2.1 Simple GPR models based on simple mean and simple covariance functions

Let the vector $\mathbf{x}_n$ indicate a particular position in the input space. The set of training input vectors $\mathbf{X}_N \equiv \{\mathbf{x}^n\}_{n=1}^N$ corresponds to the target vector $\mathbf{y}_N \equiv \{y^n\}_{n=1}^N$. For prognostics as in this study, x is the time period while y is a novel DAI for monitoring the health state of slow rotating bearings. A Gaussian process $f(\mathbf{x})$ can be fully described by its mean and covariance (or kernel) function (Rasmussen and Williams, 2006) (see chapter 3, paragraph 3.2.6).

The assumption of a zero mean implies that the mean of the function vanishes away and therefore does not make an impact in the prediction output. In this case, the prediction is only based on the covariance function. However, prediction can be improved by considering the mean and covariance functions simultaneously. Further, the widely used SE covariance function is infinitely differentiable, meaning that the GP with this covariance function has mean square derivatives of all orders, and is therefore very smooth (Rasmussen and Williams, 2006). Hence, Stein (1999) argues that the strong smoothness assumptions in the SE covariance functions are unrealistic for modelling several physical processes, and recommends the Matérn class named after Matérn (Matérn, 1986). Currently, there are quite a number of other mean and covariance functions. Therefore, to arrive at the best mean and covariance function, three simple mean functions are implemented in this study, namely zero mean as defined in Equation (3.35), constant mean and linear.

The constant mean and linear mean functions are respectively given as (Rasmussen and Nickisch, 2013):

$$m(\mathbf{x}) = d \tag{4.1}$$

$$m(\mathbf{x}) = c\mathbf{x} \tag{4.2}$$

Similarly, for the covariance functions, seven different forms in addition to the basic two above were fitted namely the linear covariance function (covLIN), the linear covariance function with ARD covariance function (covLINard), Matérn covariance function with isotropic distance measure (covMaterniso), noise covariance function (covNoise), periodic covariance function (covPeriodic), rational quadratic covariance function with ARD (covRQard), and rational quadratic covariance function with isotropic distance measure (covRQiso). These additional seven covariance functions are respectively defined as (Rasmussen and Nickisch, 2013):

$$k_{LIN}(\mathbf{x},\mathbf{x}') = \mathbf{x}^{\mathrm{T}}\mathbf{x}' \tag{4.3}$$

$$k_{LINard}(\mathbf{x},\mathbf{x}') = \mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x}' \tag{4.4}$$

where $\mathbf{M} = diag(l)$ with a positive ARD otherwise known as the characteristics length-scale, l.

$$k_{Materniso}(\mathbf{x},\mathbf{x}') = \sigma_f^2 f_d(r_d)\exp(-r_d) \quad \text{where} \tag{4.5}$$

$$r_d = \sqrt{\frac{d}{l^2}(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}(\mathbf{x}-\mathbf{x}')}$$

where σ_f^2 is the signal variance, l is the characteristic length-scale, and d relates to the smoothness of the Gaussian process.

$$k_{Noise}(\mathbf{x},\mathbf{x}') = \sigma_n^2\delta(\mathbf{x}-\mathbf{x}') \tag{4.6}$$

where σ_n^2 is the noise variance and δ is the Kronecker delta function which is equal 1 if $\mathbf{x} = \mathbf{x}'$ and zero otherwise.

$$k_{Periodic}(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\delta\left(-\frac{2}{l^2}\sin^2[\frac{w}{2\pi}(\mathbf{x}-\mathbf{x}')]\right) \tag{4.7}$$

where w is the period or angular frequency and the rest hyperparameters which are as previously defined.

$$k_{RQard}(\mathbf{x},\mathbf{x}') = \sigma_f^2\left(1+\frac{1}{2\alpha}(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}M(\mathbf{x}-\mathbf{x}')\right)^{-\alpha} \tag{4.8}$$

where α is the shape parameter for the rational quadratic covariance, and the rest hyperparameters as were previously defined.

$$k_{RQiso}(\mathbf{x},\mathbf{x}') = \sigma_f^2\left(1+\frac{1}{2\alpha l^2}(\mathbf{x}-\mathbf{x}')^{\mathrm{T}}(\mathbf{x}-\mathbf{x}')\right)^{-\alpha} \tag{4.9}$$

All the hyperparameters are unknown and hence needs to be inferred from the training data.

4.2.2 Simple means and simple covariance functions evaluations and selection

Three simple means and nine covariance functions were fitted. These eventually led to the need for the selection of the 2 best mean and 3 best covariance functions. These were selected based on the MAPE and RMSE given as Equations (3.42) and (3.43) respectively.

4.2.3 Development of integrated GPR models

It is possible to construct composite mean and composite covariance functions from the simple mean and covariance functions, respectively. This allows for more flexibility and accurate prediction. In broad terms, a random function of $\mathbf{x}$ and $\mathbf{x}'$ pairs of input may not be a suitable function (Rasmussen and Williams, 2006).

Based on the above three simple mean functions, three composite mean function are proposed. These are derived from the best two out of the three simple mean functions as defined in Equations (3.35), (4.1) and (4.2). The constant and linear mean functions outperformed the zero mean function as is evidenced in the empirical section and therefore were used to construct three composite mean functions namely an affine, a polynomial quadratic and a polynomial cubic mean functions, respectively expressed as:

$$m(\mathbf{x}) = c\mathbf{x} + d \tag{4.10}$$

$$m(\mathbf{x}) = b\mathbf{x}^2 + c\mathbf{x} + d \tag{4.11}$$

$$m(\mathbf{x}) = a\mathbf{x}^3 + b\mathbf{x}^2 + c\mathbf{x} + d \tag{4.12}$$

Further, based on the best three covariance functions which for this study were the rational quadratic with ARD, rational quadratic with isotropic distance measure and the Matérn form with isotropic distance measure covariance functions as evidenced in the empirical section, a composite covariance function was formed. The proposed composite covariance function, k^c formed from the three simple covariance functions defined in Equations (4.5), (4.8) and (4.9) are implicitly expressed as:

$$k^c = k_{RQard}(\mathbf{x}, \mathbf{x}') + k_{RQiso}(\mathbf{x}, \mathbf{x}') + k_{Meterniso}(\mathbf{x}, \mathbf{x}') \tag{4.13}$$

To derive the predictive distribution in this case, the hyperparameters in the mean and covariance functions were optimized using the maximization of the log likelihood and the resulting predictive distribution is now described by the following equations:

Prior:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim N\left(\begin{bmatrix} \mu^c \\ \mu_*^c \end{bmatrix}, \begin{bmatrix} K^c(\mathbf{X}, \mathbf{X}) & K^c(\mathbf{X}, \mathbf{X}_*) \\ K^c(\mathbf{X}_*, \mathbf{X}) & K^c(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \tag{4.14}$$

where the posterior is as given in Equation 3.39.

The equation $\mu^c = m(\mathbf{x})\ i = 1,2,....M$ is used for the training composite means and μ_*^c for the test composite means.

The MAP estimates $\bar{y}_*$ were used as slow rotating bearing remaining useful metrics are expected to improve the flexibility and accuracy of the simple mean and simple covariance functions.

4.2.4 GPR model evaluation and selection

The prediction performance of the different GPR models were evaluated using two widely used forecast evaluation criteria namely the MAPE and the RMSE given as in Equations (3.42) and (3.43) respectively.

4.2.5 RUL prediction

The best performing model was selected and subsequently used in the RUL prediction of the slow rotating bearings.

The detailed proposed integrated GPR methodology is shown diagrammatically in Figure 4.2.

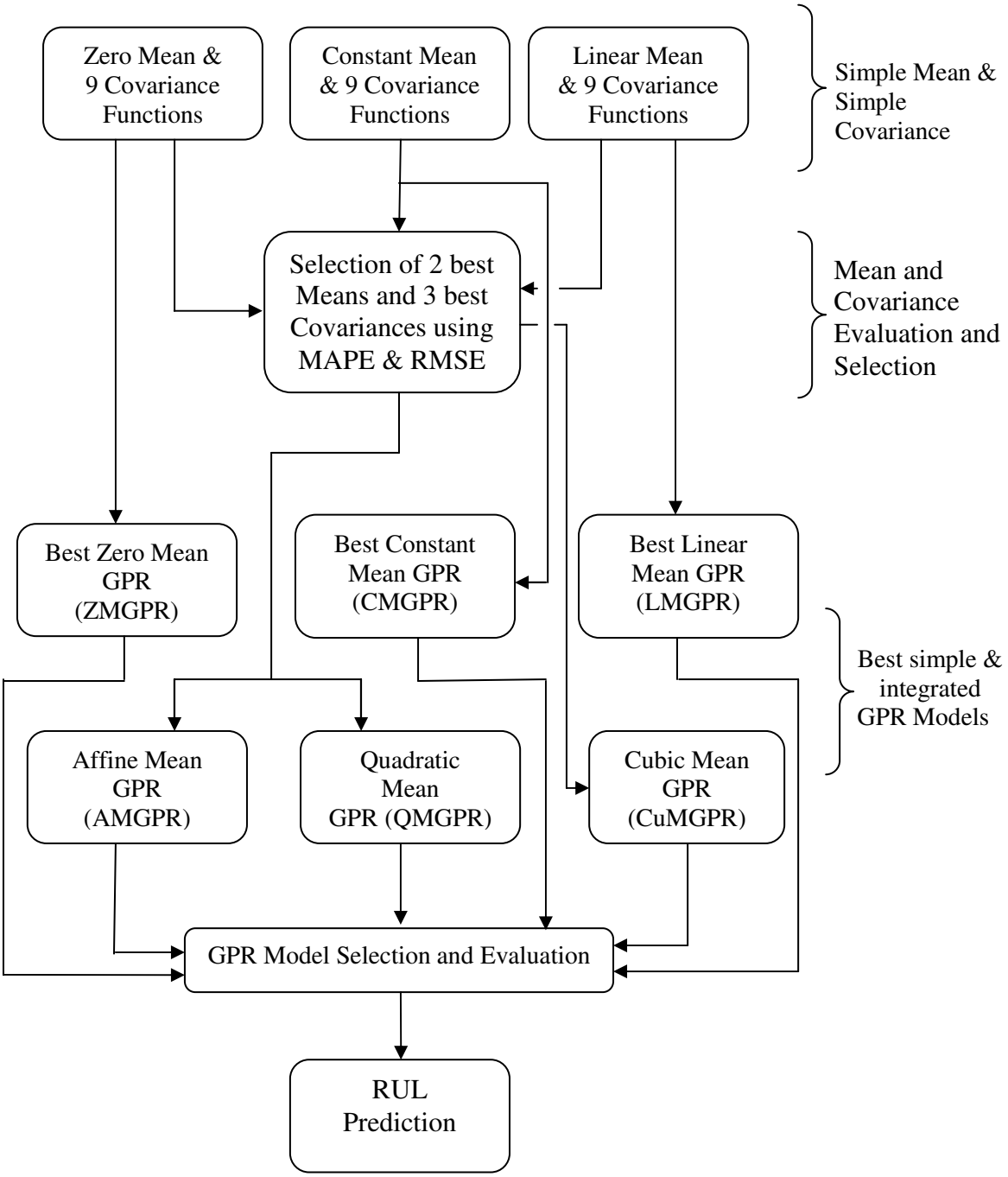

Figure 4.2: Framework for integrated GPR modelling and prediction of RUL

4.3 Results and discussion

4.3.1 Prediction based on dependent samples

Prior to any prediction, it is important to decide on which method to use for obtaining the training and the test sets. The leave-one-out CV was used in this study. The

results and discussion that follow are those obtained using the dependent sample which involves an equal division of the entire sample into training and test sets. The goal of this chapter is to build an optimal GPR model for RUL prediction; hence the input set is the time point while the target set is the DAI discussed in chapter 2.

Obtaining the optimal GPR may take several steps. To implement the GPR-based prognostics, the first and most important step is the selection of the mean and covariance functions. Starting with the simple GPR modes, three simple mean functions and nine simple covariance functions were considered. The three mean functions are the zero mean function, the constant mean function and the linear mean function.

The nine simple covariance functions are the linear covariance function (covLIN), the linear covariance function with ARD covariance function (covLINard), Matérn covariance function with isotropic distance measure (covMaterniso), noise covariance function (covNoise), periodic covariance function (covPeriodic), rational quadratic covariance function with ARD (covRQard), rational quadratic covariance function with isotropic distance measure (covRQiso), the SE covariance function with ARD (covSEard) and SE covariance function with isotropic distance measure (covSEiso).

To implement a specific simple GPR model, one simple mean and one simple covariance functions were simultaneously used. This means that for each of the different operating conditions depicted in Bearings 1, 2 and 3, nine (9) simple GPR models were implemented making a total of 27 simple GPR models per bearing since there are three simple mean and nine covariance functions.

The average MAPE and the average RMSE with respect to the selection of the mean functions are presented in Figures 4.3, 4.4 and 4.5 for Bearings 1, 2 and 3, respectively. Models with smaller MAPE and RMSE are considered to be better than models with larger values.

The results show that the constant mean and linear mean functions have better predictive ability than the zero mean function, however, the constant mean function is the best on average. This finding is robust to the different operating conditions. According to Rasmussen and Nickisch (2013), composite functions can be composed of either a combination of two or more simple functions or a combination of other composite functions, thus allowing for very flexible and interesting structures.

Therefore, rather than proceeding with just the best mean function, the best two are selected and were subsequently used for building three composite mean functions namely an affine mean function, a quadratic mean function and a cubic mean function. This ensures that the advantages from this two are integrated for more accurate predictions of RUL of slow rotating bearings.

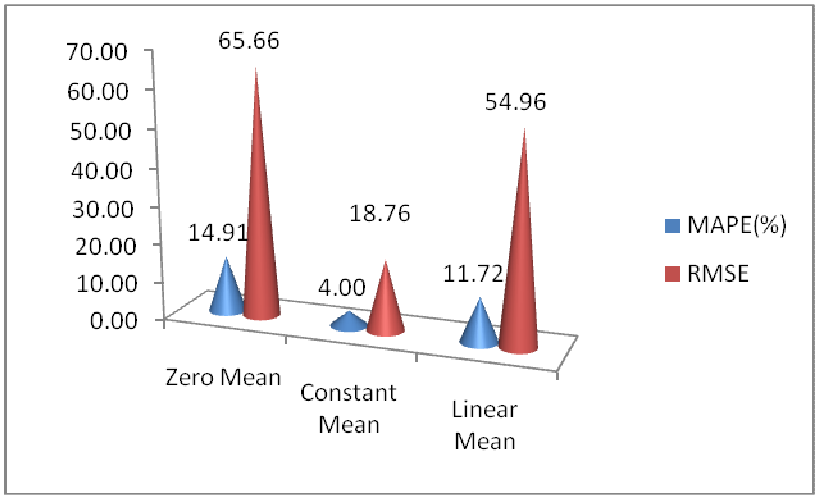

Figure 4.3: Average RMSE and MAPE for GPR with zero, constant and linear mean functions for Bearing 1

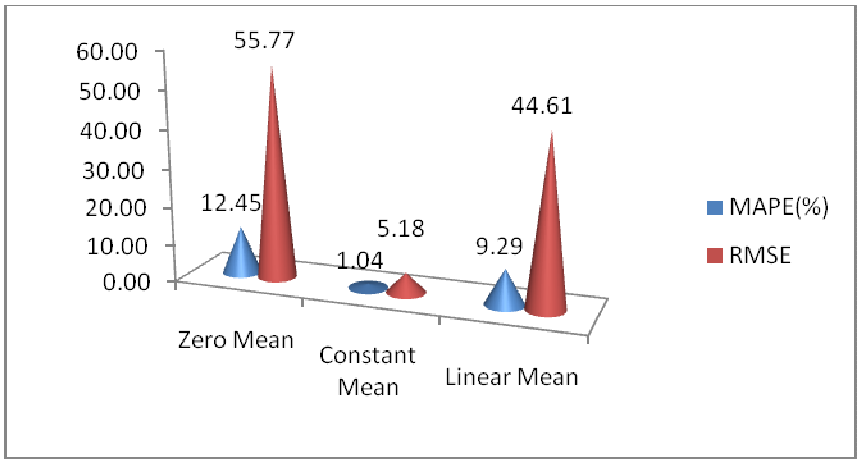

Figure 4.4: Average RMSE and MAPE for GPR with zero, constant and linear mean functions for Bearing 2

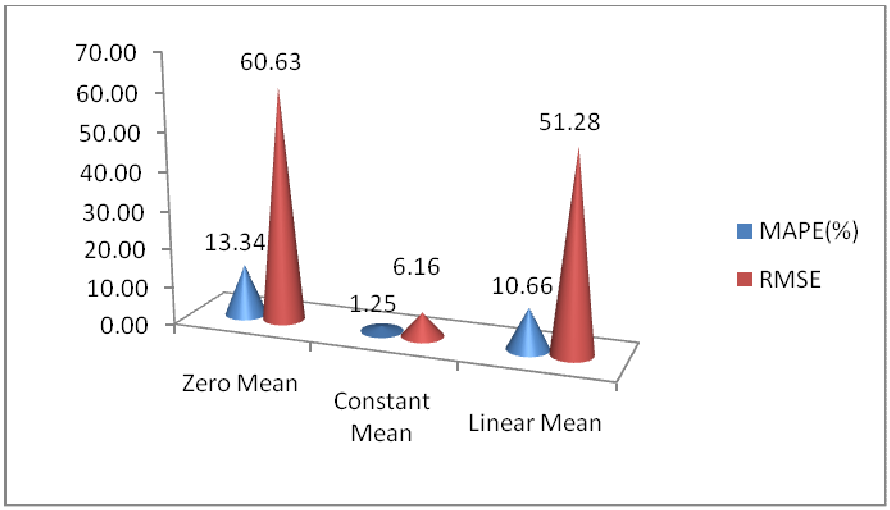

Figure 4.5: Average RMSE and MAPE for GPR with zero, constant and linear mean functions for Bearing 3

Having selected the constant and linear mean functions as the best, the next task is to also select the best covariance functions based on these two mean functions. Only the best three covariance functions are selected to keep the modelling tractable. The results of the nine covariance functions based on the MAPE and RMSE rankings for Bearings 1, 2 and 3 under the constant mean functions are presented in Figures 4.6, 4.7, and 4.8, respectively. Since MAPE and RMSE actually depicts PE of a specific model, low ranks would then mean low values of MAPE and RMSE and hence lower PE. In other words, the lower the rank, the better is the predictive performance of the model. The results indicate that the covRQiso, covRQard and covMaterniso were the

best covariance for bearings one and two. For Bearing 3, the best three covariance functions were covRQiso, covRQard and covPeriodic.

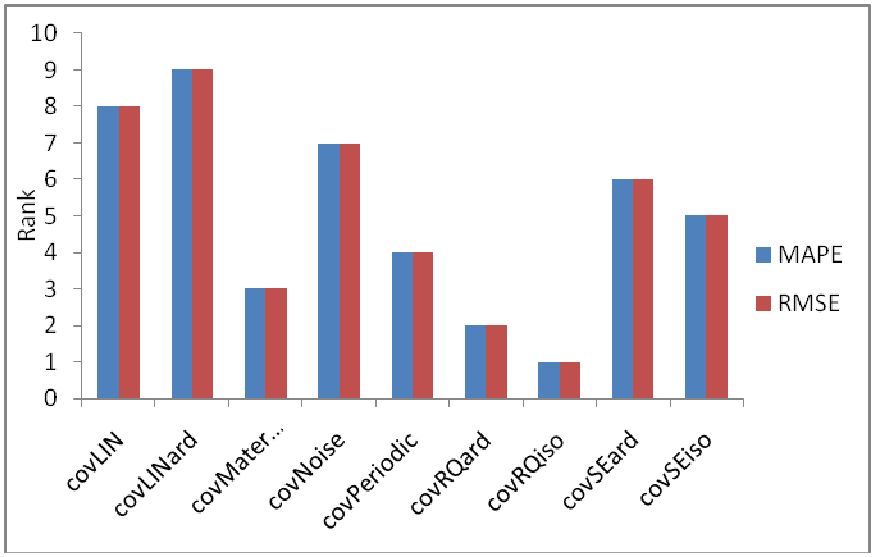

Figure 4.6: Ranking RMSE and MAPE from GPR with constant mean and 9 covariance functions for Bearing 1

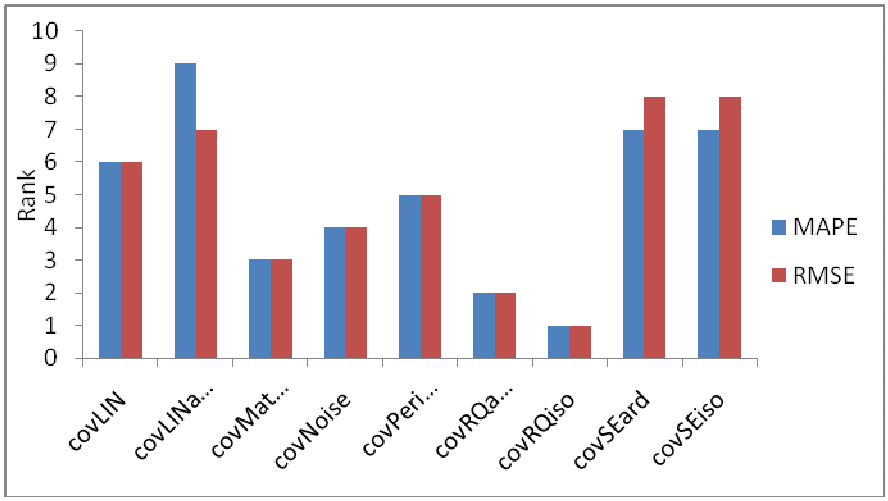

Figure 4.7: Ranking RMSE and MAPE from GPR with constant mean and 9 covariance functions for Bearing 2

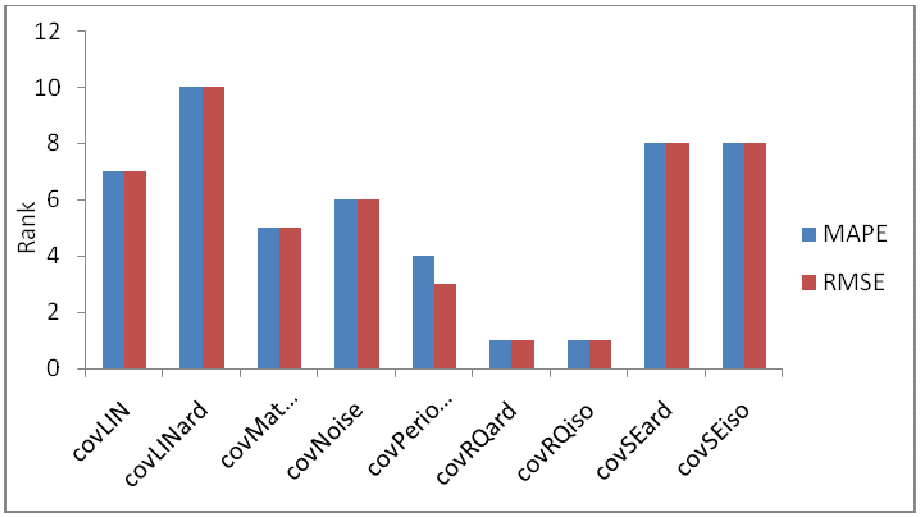

Figure 4.8: Ranking RMSE and MAPE from GPR with constant mean and 9 covariance functions for Bearing 3

Under the linear mean function, the ranking with MAPE and RMSE are presented in Figures 4.9, 4.10 and 4.11 for Bearings 1, 2 and 3, respectively. For all three bearings, the best covariance functions are covRQiso, covRQard and covMaterniso. Although the covPeriodic seems to have performed better than covMaterniso under constant mean for Bearing 3, on aggregate the latter had smaller error than the former when viewed across the three bearings. On this basis therefore, the best three covariance functions that fit the data for this study are covRQiso, covRQard and covMaterniso. These three are then subsequently used in building one composite covariance function. This outcome is quite intuitive. Based on Rasmussen and Williams (2006), the covRQiso and covRQard can effectively model the small and medium irregularities in the data. Further, the smooth rising trend can be effectively modelled using the Matern class covariance function with two hyperparameters controlling the amplitude and characteristic length-scale.

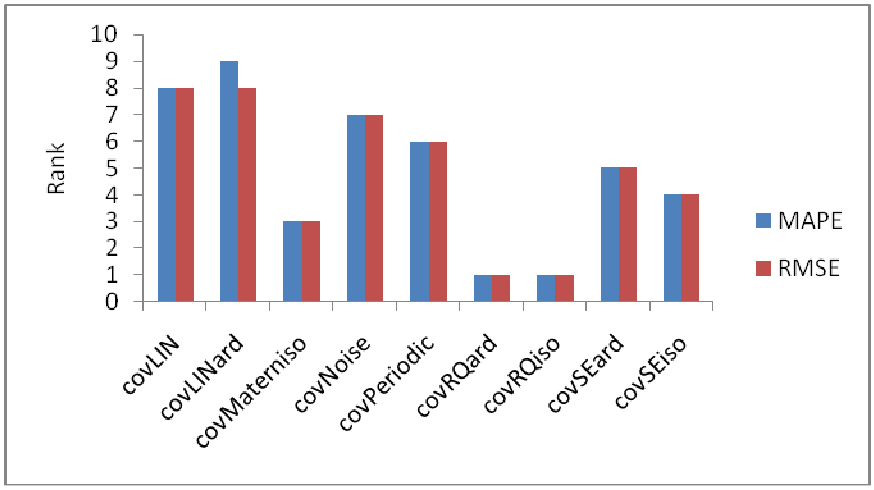

Figure 4.9: Ranking RMSE and MAPE from GPR with linear mean and 9 covariance functions for Bearing 1

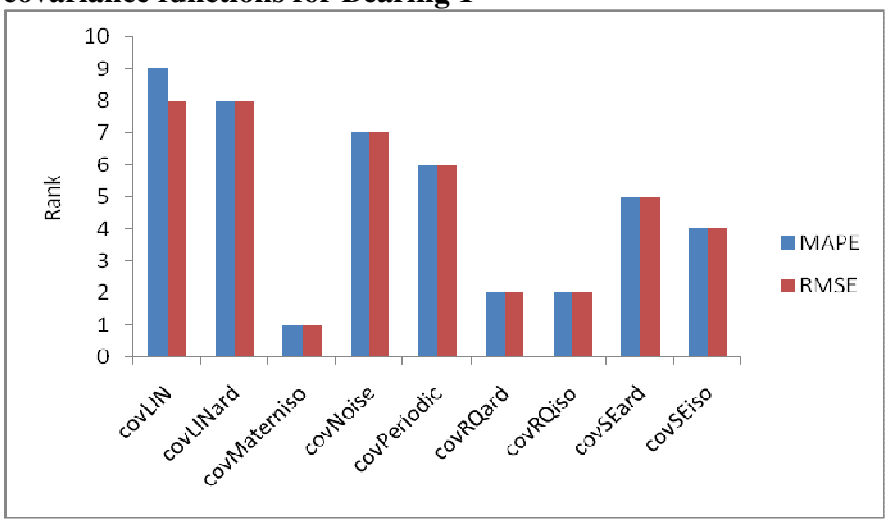

Figure 4.10: Ranking RMSE and MAPE from GPR with linear mean and 9 covariance functions for Bearing 2

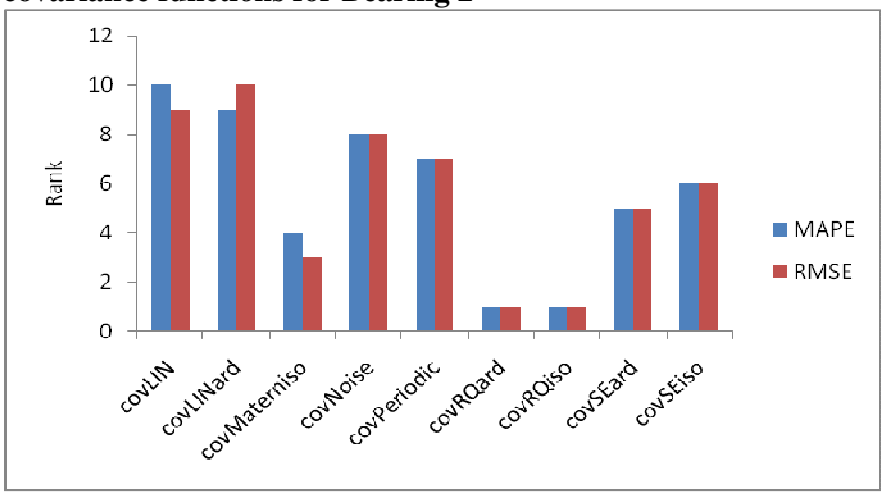

Figure 4.11: Ranking RMSE and MAPE from GPR with linear mean and 9 covariance functions for Bearing 3

The next step in building an optimal GPR-based model is to compare the performance of the different resulting GPR models and then select the very best for inference about the RUL of slow rotating bearings. Each of the three previously constructed composite mean functions and the one composite covariance function are now simultaneously used to obtain three combination GPR models. These models are denoted as affine mean Gaussian process regression (AMGPR), quadratic mean Gaussian process regression (QMGPR), and cubic mean Gaussian process regression (CuMGPR). The results from these three combinations or integrated GPR models are compared with results from the 3 best simple GPRs. This simple GPR models are denoted as zero mean Gaussian process regression (ZMGPR), constant mean Gaussian process regression (CMGPR) and linear mean Gaussian process regression (LMGPR). It is important to note that these 3 best simple GPR models are obtained by estimating the relevant mean function over nine different covariance functions and subsequently the overall best model is selected and compared with the integrated GPR models. Hence, six different GPR models are compared.

The forecast evaluation is based on the MAPE and RMSE values. The results are presented in Table 4.1. For Bearing 1, the QMGPR has the best predictive ability based on the two evaluation criteria. For Bearing 2, the best zero mean GPR (ZMGPR) is the best. However, the margin between this model and the rest of the models barring the cubic mean GPR (CuMGPR) is quite small. For Bearing 3, the performance of the affine mean GPR (AMGPR) is best based on MAPE the quadratic mean GPR (QMGPR) is the best based on RMSE. Viewed across the three bearings, and based on the average, there is basically no difference between the predictive performance of the affine mean and quadratic GPR. The cubic GPR is the worst performing model across all bearings. The model breaks down once estimation exceeds the polynomial of order 2 which in this case is the quadratic mean. This calls for caution in the choice of any mean or covariance function as it had been clearly shown that any arbitrary selection would not always guarantee a valid result.

Table 4.1: Comparison of the proposed integrated GPR with the best of linear, constant and zero mean GPR models based on dependent samples

	Bearing 1		Bearing 2		Bearing 3	
Models	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
ZMGPR	0.2951	1.5041	1.93E-06	1.14E-05	0.0014	0.0083
CMGPR	0.2950	1.5033	0.0074	0.0433	0.0017	0.0099
LMGPR	0.0017	0.0089	0.0001	0.0006	0.0011	0.0061
AMGPR	0.0013	0.0069	0.0004	0.0025	0.0007	0.0040
QMGPR	0.001	0.0052	0.0006	0.0031	0.0008	0.0038
CuMGPR	93.6593	612.8139	108.4839	661.7885	21.9123	154.0565

Given that it is more complex to estimate the QMGPR compared to the affine GPR, it would then be preferable to stick to the later for RUL inferences. We now turn to the AMGPR predictions for Bearings 1, 2 and 3 as presented in Figures 4.12, 4.13 and 4.14 respectively. It can be seen that the AMGPR traced the actual whole life of the bearing quite closely. This is confirmed by the very narrow uncertainty or confidence bands. Concentrating on the RUL which basically starts from the slightly degraded threshold line and ends with the severely degraded threshold line, it is also clear that AMGPR predicted the RUL of the three tested bearings with almost 100% accuracy. This can be clearly seen in the computations presented in Table 4.2 which reports the values for the onset of incipient damage and those of the final failure along with the RUL computed by taking the difference between the two values. These are reported for both actual and predicted RUL. The prediction accuracy computed by comparing the actual RUL and predicted RUL is reported in the last column of Table 4.2 for each test bearing. This results show that irrespective of the speed and loading operating conditions to which slow rotating bearings are subjected, the AMGPR model is capable of predicting almost the exact RUL. This is very important for preventive maintenance of bearings in particular but also other mechanical and allied systems in general.

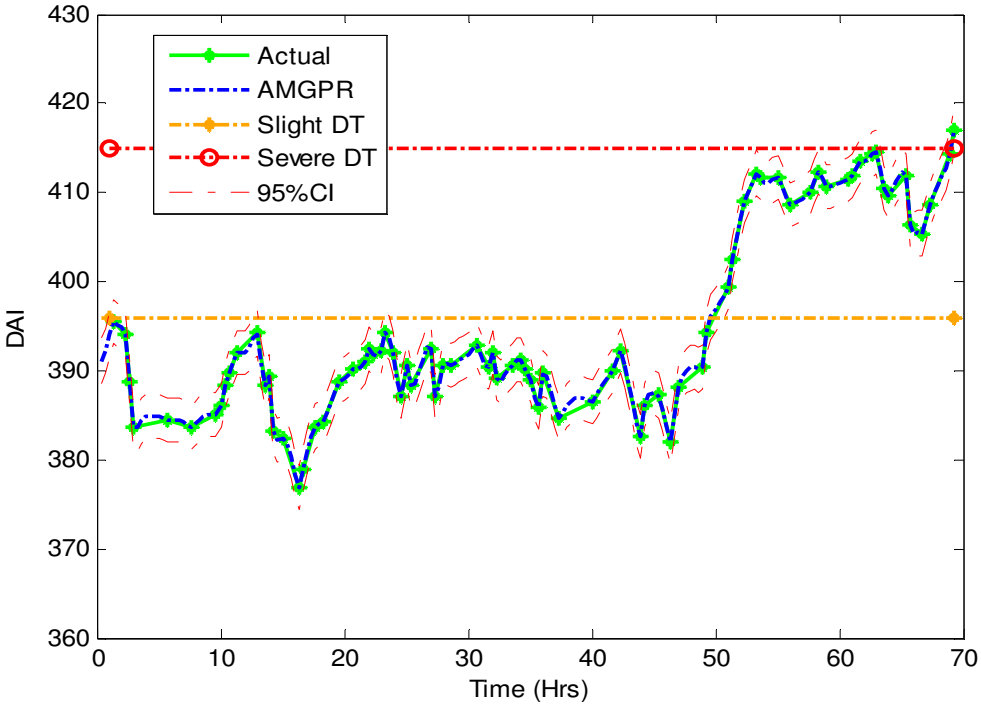

Figure 4.12: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 1 based on dependent samples

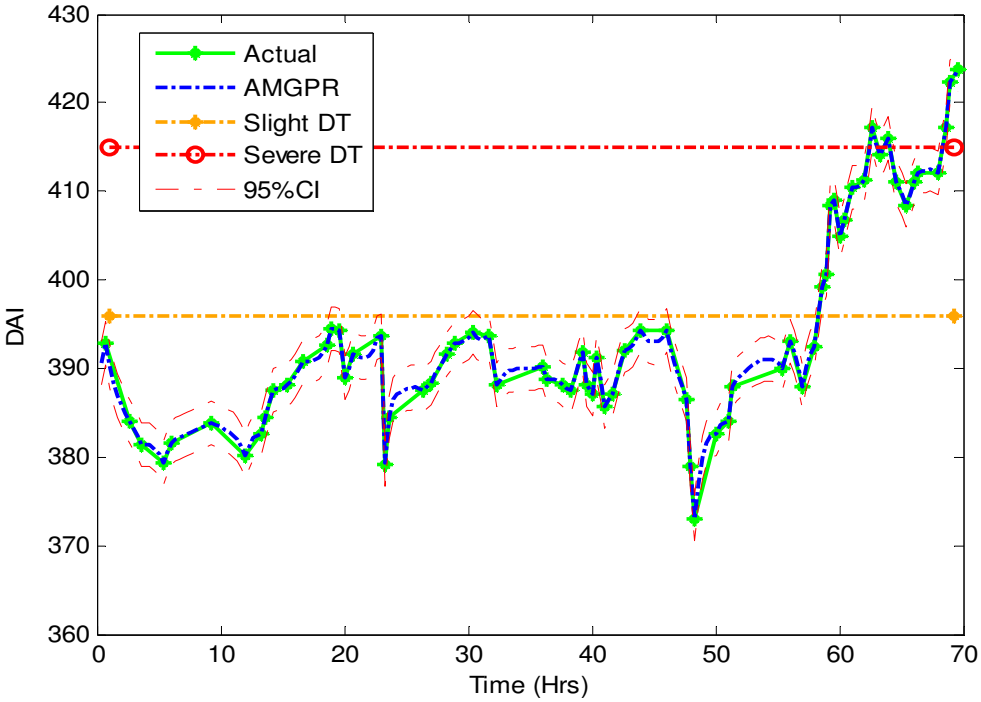

Figure 4.13: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 2 based on dependent samples

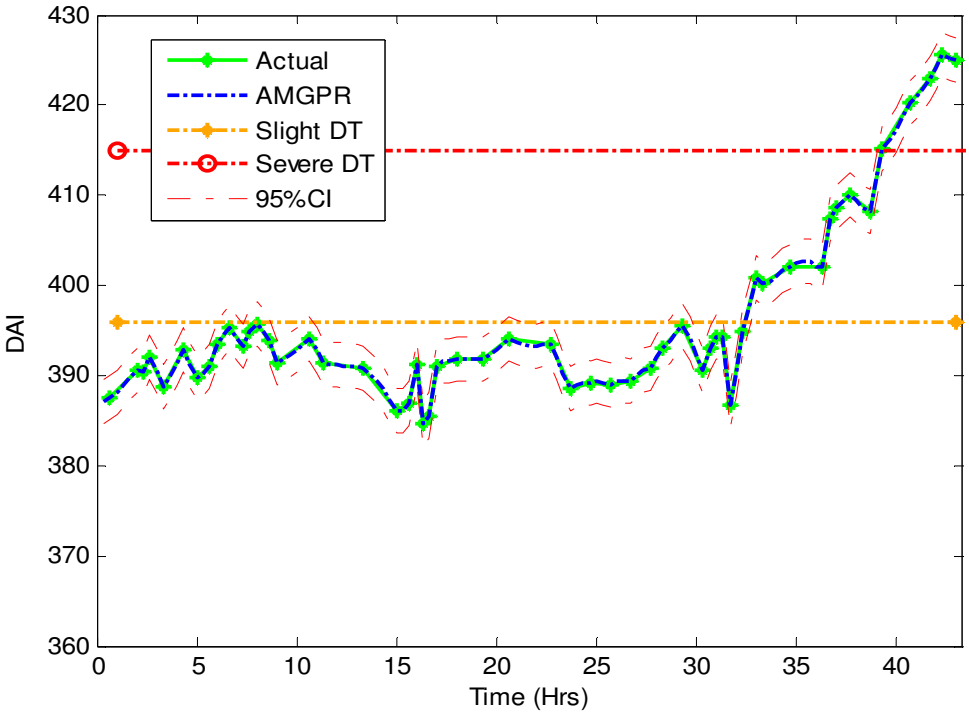

Figure 4.14: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 3 based on dependent samples

The findings here are consistent with Hong and Zhou (2012a) who predicted RUL of bearings based on kurtosis and RMS using a zero mean function and three single covariance functions namely: the Matern class, the neural network and the isotropic SE covariance functions and found that in general, the Matern class performed better than the other two covariance functions. However while their PE for the zero mean, Matern class covariance function range from 0.17% to 0.48%. The result from this study based on the preferred developed integrated GPR model has a near zero PE and thus shows it is more reliable than Hong and Zhou (2012a) results from single GPR model. Also, Hong and Zhou (2012b) used a zero mean function and combination of SE isotropic (covSEiso) and isotropic rational quadratic covariance function (covRQiso) for bearing degradation assessment based on root mean square. Comparing results from each single covariance function and the composite covariance function, Hong and Zhou (2012b) found that at one step, the relative PE for composite covariance function, covSEiso and covRQiso were 0.11%, 0.41%, and 0.33% respectively. A comparison of Hong and Zhou (2012b) findings with the one

from this study clearly shows that it is more rewarding to model both mean and covariance functions simultaneously and better still using the composite of each as this current study obtained almost a zero PE.

The findings here are also consistent with Chen and Ren (2009) who obtained several re-sampled training data from an industrial chemical plant via a bootstrap method and subsequently use each sample to form a GPR model-based on the zero mean and composite covariance function. Combining each GPR prediction using different bagging or combination methods, Chen and Ren (2009) found that the combined GPR models are better suited for RUL prediction than any of the single models. However, while their RMSE from the combined model ranges from 0.09 to 0.2, the RMSE for the developed preferred combined GPR model in this study is approximately zero. Across all scenarios indicating that modelling the mean and covariance functions simultaneously improves the predictive ability of the GPR model.

Table 4.2: AMGPR RUL prediction using dependent samples

Test set	Actual			Predicted			Accuracy
	Incipient	Failure	**RUL**	Incipient	Failure	**RUL**	
Bearing 1	50	69	19	50	69	19	100.00
Bearing 2	58	68	10	58	68	10	100.00
Bearing 3	33	39	6	33	39	6	100.00

4.3.2 Predictions based on independent samples

The results of the three integrated GPR models and the three best single mean single covariance function models are replicated here using the independent sample. To validate the models under this approach, the analysis proceeded as follows: the training sets for Bearing 1 is the DAI on Bearings 2 and 3 while the test set is DAI on Bearing 1. For Bearing 2, the training sets are the DAI on Bearings 1 and 3 while the test set is the DAI on Bearing 2. Similarly, for Bearing 3, the training sets are Bearings 1 and 2 while the test set is Bearing 3. The forecast evaluation is based on the MAPE and RMSE values. The results are presented in Table 4.3. For Bearing 1,

the QMGPR has the best predictive ability based on MAPE while AMGPR is the best based on RMSE. For Bearing 2, AMGPR outperform the rest based on both MAPE and RMSE. For Bearing 3, the CMGPR model produced the least MAPE while AMGPR produced the least RMSE. Viewed across the three bearings, the AMGPR is the best predictive model and hence inferences on RUL are based on this model.

Table 4.3: Comparison of the proposed integrated GPR with the best of linear, constant and zero mean GPR models based on independent samples

	Bearing 1		Bearing 2		Bearing 3	
Models	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
ZMGPR	1.0728	5.6366	1.1052	5.4577	1.1479	5.5342
CMGPR	1.0728	5.6368	1.1331	5.5736	1.0318	5.0105
LMGPR	1.0585	5.7035	1.1692	5.7400	1.1432	5.5102
AMGPR	0.9695	5.1175	1.0345	5.0743	1.0346	4.9982
QMGPR	0.9358	5.2480	1.1650	5.7940	1.1540	5.4758
CuMGPR	1314.6	12172.3	3824.0	31011.0	1940.8	16844.0

The predictions from the AMGPR model are shown in Figures 4.15, 4.16 and 4.17 for Bearings 1, 2 and 3, respectively. The failure times as predicted by the optimal model are slightly different from the actual failure times for each bearing. The accuracy of prediction of failure times are 95.6%, 98.5% and 95.4% for Bearings 1, 2 and 3, respectively. A good feature of a prediction model is its ability to predict failure either at the exact time or close to the actual time. The AMGPR model is able to do this with exception of Bearing 3 where it predicted failure after the actual failure had occurred. Table 4.4 reports the RUL values for the onset of incipient damage and those of the final failure along with the RUL computed by taking the difference between the two values. These are reported for both actual and predicted RUL based on the independent samples. The prediction accuracy computed by comparing the actual RUL and predicted RUL is reported in the last column of Table 4.4 for each test bearing. The results show that AMGPR model predicted the RUL of Bearing 1, Bearing 2 and Bearing 3 with about 42%, 60% and -67% accuracy, respectively.

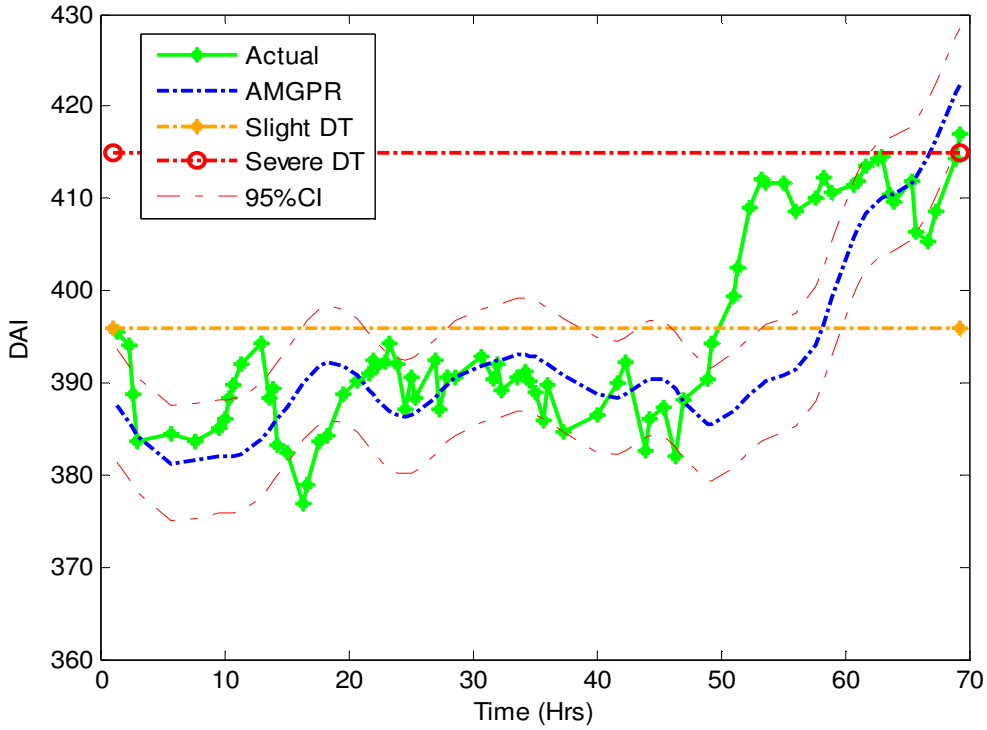

Figure 4.15: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 1 based on independent samples

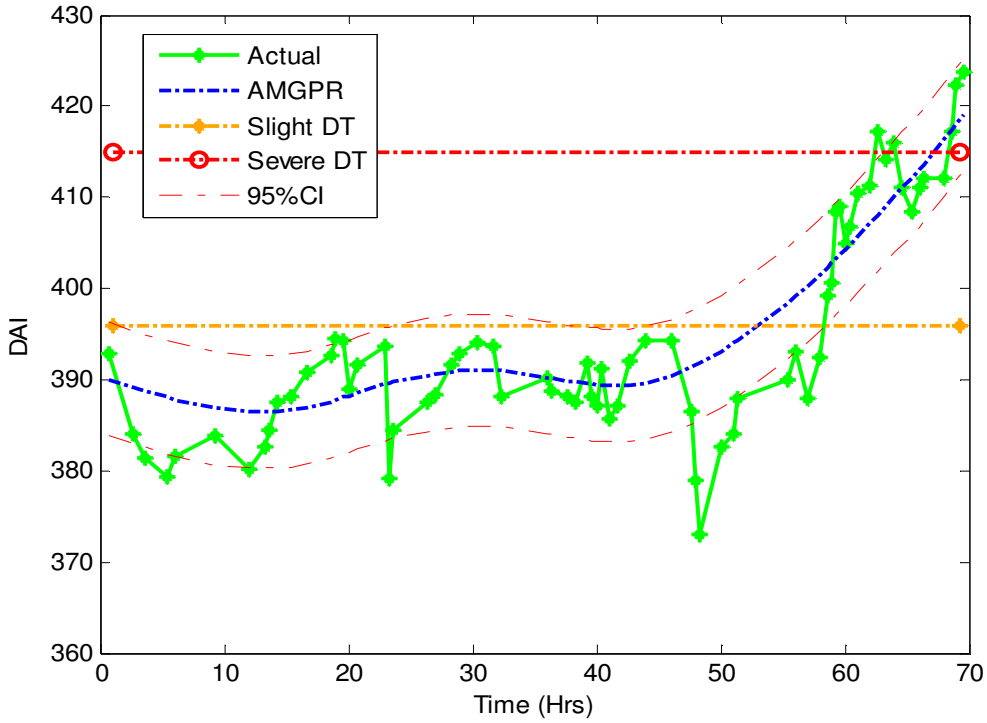

Figure 4.16: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 2 based on independent samples

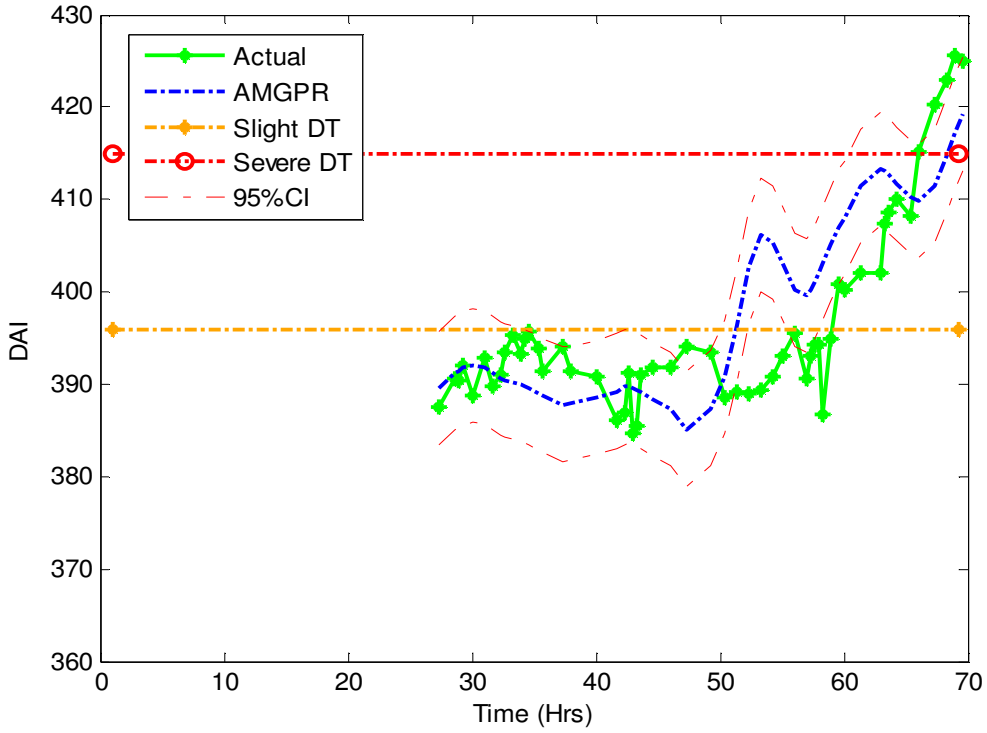

Figure 4.17: Affine GPR prediction of RUL with 95% CI and the actual RUL for Bearing 3 based on independent samples

Table 4.4: AMGPR RUL prediction using independent samples

Test set	Actual			Predicted			Accuracy
	Incipient	Failure	**RUL**	Incipient	Failure	**RUL**	
Bearing 1	50	69	19	58	66	8	42.11
Bearing 2	58	68	10	52	66	14	60.00
Bearing 3	33	39	6	27	43	16	-66.67

Comparing results from estimations based on training and test sets that are dependent and sets that are independent, the results in Tables 4.1 and 4.3 show that predictions based on independent samples yielded larger MAPE and RMSE values than the estimations based on dependent. This might be an indication of overfitting of the later. With respect to prediction of RUL as shown in Tables 4.2 and 4.4, while the optimal model yielded a near perfect RUL prediction in the case of dependent samples, its RUL prediction is slightly different from the actual RUL in the case of

independent samples. This may not be surprising given the potentials of positive high correlation between dependent samples. Also the performances of the models vary depending on whether dependent or independent samples are considered. For instance, for the dependent samples, the AMGPR and QMGPR performed nearly the same way, whereas their performances differ for the independent samples with the former clearly outperforming the latter.

4.4 Summary

The chapter proposes an integrated GPR model for predicting the RUL of bearings. A scientific way of obtaining the optimal GPR model for prediction is demonstrated as well. Further, the study investigated the performance of the models when the training and test sets are obtained from dependent samples and when the samples are independent. The zero mean, constant mean and linear mean functions were simultaneously modelled with nine different covariance functions namely the linear covariance function, the linear covariance function with ARD covariance function, Matérn covariance function with isotropic distance measure, noise covariance function, periodic covariance function, rational quadratic covariance function with ARD, rational quadratic covariance function with isotropic distance measure, the SE covariance function with ARD and SE covariance function with isotropic distance measure.

From these different mean and covariance functions, 9 simple GPR models were estimated for each of the three different operating conditions (as indicated in Bearings 1, 2 and 3) making a total of 27 simple GPR models per bearing. Out of these 27 models, the best two mean functions and the best three covariance functions were selected based on the RMSE and MAPE. These selected functions were subsequently used to form three combination GPR models namely an AMGPR, QMGPR and a CuMGPR. The performance of these three combinations or integrated GPR models were then compared with the best zero mean, best constant mean and best linear mean GPR models and the optimal model was selected based on RMSE and MAPE.

The findings show that in general the composite integrated GPR models perform better than the simple mean simple covariance GPR models, irrespective of whether the training or test sets are dependent or independent. Amongst, the three combination models, the AMGPR and the QMGPR perform pretty much the same way based on dependent samples but their performances varied with independent samples with the former outperforming the latter.

Given that the QMGPR is more complex and the fact it was not better than AMGPR for independent sampling, this study selects the AMGPR as the optimal GPR model for predicting RUL for slow rotating bearings. Results based on dependent samples show that the AMGPR has the capability of predicting RUL with almost an almost 100% accuracy and generalizes well even with limited data. However, with independent samples, the AMGPR model predicted a slightly different RUL than the actual. These findings are robust to varying operating conditions which is intuitive given that in real life bearings and other systems cannot operate under constant operating conditions.

Chapter 5 Conclusions

5.1 Contributions of the research

Bearings, including slow rotating bearings operating at varying speeds, are one of the major machinery component types used in industries like mining, automotive, power generation, railway, shipping, manufacturing, and chemical plants. Failure of bearings is a major source of rotating machinery failure. It has been shown that bearing defects account for about 40% of motor faults. Several causes of bearing failure include fatigue, wear, corrosion, debris contamination and misalignment, faulty installation or improper mounting, blockade, passage of foreign particles, inadequate or improper lubrication, excessive speed and inappropriate temperature, vibration, inefficient seals, and overloading etc. The incipient damage needs to be detected using diagnostics techniques. Also RUL needs to be predicted using prognostics methods. Hence, the contributions of this research are as follows in the paragraphs below.

5.1.1 Development of a novel DAI for diagnostics

This study uses statistical data-driven Bayesian methods to develop a novel DAI for slow rotating bearing fault diagnostics based on AE data obtained from run-to-failure experiments. The proposed model is capable of accounting for data dimensionality reduction and hence for the reduction of high feature correlation, nonlinearities, noise filtering, non-stationarities, uncertainties, time variation (dynamics) and multimodal distribution in the data, under varying operating conditions. A DAI which accounts for these characteristics for the assessment of machine performance is important for effective diagnostics. The proposed novel DAI is the first known attempt to capture the features spelt out above. The early detection of faults through diagnostics and subsequent prediction of RUL through prognostics provides for the planning of the maintenance schedule thereby minimising machinery downtime.

The novel DAI is obtained by the integration of PKPCA, the GMM and the EWMA and used for the diagnostics of slow rotating bearings. The nonlinear PKPCA is used to extract AE reduced features from the large statistical features obtained from the AE data capitalising on the strength of each thereby making it more sensitive and robust for the detection of defects. Next the GMM with high computational efficiency, descriptive ability and robustness is used in modelling the multimodal data. This characteristic makes it very easy for GMM to render a smooth estimation of the AE signal variability of healthy bearings. Finally, the novel DAI is obtained by the use of the EWMA model in the improvement of the sensitivity and dependability of the NLL with regard to the detection of the degradation of the slow rotating bearing.

5.1.2 Prognostics using various approaches

Further, a novel approach by integrating a newly developed DAI is used as an input in several regression models such as the MLP, RBF, BLR, GMR and the GPR for RUL prediction. Secondly, the MAPE and RMSE were used in model evaluation and selection of the best performing model. There is no known study that has used these set of models for the evaluation on the same data set. Thirdly, the RUL prediction of slow rotating bearing is obtained by the utilization of the best performing model.

5.1.3 Prognostics based on an integrated GPR model

An integrated GPR model for prediction of RUL of slow rotating bearings which combines the advantages from the individual mean and covariance functions is proposed. The proposed technique consists of the scientific selection and construction of composite mean made up of two simple mean functions and composite covariance functions made up of three covariance functions so as to avoid model misspecification. The procedure ensures that only well suited mean and covariance functions are combined in order to obtain an optimal GPR model for prognostics of slow rotating bearings.

Previous studies mostly implement a GPR using only the covariance function while assuming the mean function is equal zero. Therefore, the contribution of this investigation consists of the implementation of a novel integrated GPR-based model, where a scientifically selected composite mean function is simultaneously estimated with composite covariance function for a more flexible and accurate prediction of RUL of slow rotating bearings.

Although, there is good reasoning for the use of SE covariance function in many settings (Rasmussen and Williams, 2006), it is still imperative to explore thoroughly which covariance functions are more appropriate for specific data (see chapter 1, paragraph 1.8.5). Further, this study contributes to prognostics by using the dependent and independent modelling approaches in the evaluation of the performance of the proposed integrated GPR models under a leave-out-one CV approach (see paragraph chapter 1, 1.8.7).

5.2 Conclusions

A novel DAI was developed for the CM of slow rotating bearings. The DAI was determined by integrating the PKPC, GMM and EWMA and subsequently using it for the assessment of the damage degradation of the slow rotating bearing from incipient damage to failure. The K-means classification was used in the state classification of the entire bearing data. The NLL was subsequently obtained by using the best-performing PKPCs in the GMM. Finally, the DAI was obtained by EWMA smoothening of the NLL and using the DAI for the assessing bearing damage. The slight DT and severe DT were obtained by the use of KDE method on the bearing data. The DAI's effectiveness was investigated in comparison to other monitoring indexes and it was established that it outperformed them. The DAI was found to be the only model with all of these properties: reducing data redundancy by using fewer kernel principal components from the original features, being more sensitive and reliable in the detection of slight faults via EWMA which incorporates the dynamics of the slow rotating bearings, elimination of the bias as a result of wrong model specifications, accounting for the nonlinearities and multimodal data, noise filtration,

and smooth feature characteristics allowance through both the GMM and EWMA components parts, reduction of data measurement and model parameter estimation uncertainties and external knowledge incorporation via the GMM component. The proposed DAI is found to be effective in the CM of slow rotating bearings under varying operating conditions.

Subsequently, the novel DAI was used in MLP, RBF, BLR, GMR and GPR models respectively in the prognostics and RUL prediction of slow rotating bearings. Three healthy slow rotating bearings were run to failure in the course of the investigation. Predictions were obtained using test and training sets from both dependent and independent samples. All the models predicted incipient damage and RUL of the slow rotating bearings. However, the GPR had the least MAPEs and RMSEs and is robust to dependent and independent samples under varying operating conditions. Therefore, the GPR is chosen as the most efficient model for prediction of RUL of slow rotating bearings.

Finally, an integrated GPR model for predicting the RUL of slow rotating bearings is developed. A novel scientific way of obtaining the optimal GPR model for prediction is demonstrated as well. The zero mean, constant mean and linear mean functions were simultaneously modelled with nine different covariance functions. From these different mean and covariance functions, 9 simple GPR models were estimated for each of the three different operating conditions. Out of these 27 models, the best two mean functions and the best three covariance functions were selected based on the RMSE and MAPE. These selected functions were subsequently used to form three combination GPR models namely an AMGPR, a QMGPR and a CuMGPR. The performance of these three integrated GPR models were then compared with the best zero mean, best constant mean and best linear mean GPR models and the optimal model was selected based on RMSE and MAPE. The findings show that in general the integrated models perform better than the simple mean simple covariance models irrespective of whether the training or test sets are dependent or independent. In this study, the AMGPR is the optimal GPR model for predicting RUL for slow rotating bearings. These findings are robust to varying operating conditions which is intuitive

given that in real life bearings and other systems cannot operate under constant operating conditions. The results obtained in this study are not only relevant for bearings preventive maintenance but also to other mechanical and non-mechanical systems.

5.3 Future work

In this study, improved diagnostic and prognostic models were developed for the CM of slow rotating bearings and by extension other mechanical systems. As prognostics is an evolving technique, it is recommended that more work still needs to be done. The study was done on a type of bearing i.e. the Timken tapered roller bearing HR 30307 J. The novel model developed was applied to the specific bearing. Hence, it is expected that further studies would be carried out on other bearing types where the newly developed model would be applied to ascertain its general applicability to different types of bearings mounted in other structures.

This study applied the novel DAI and optimised GPR RUL prognostics to slow rotating bearings. Several applications in engineering are needed so as to improve on the developed and existing models and to investigate the effectiveness of the proposed model in diagnostics of damage in those mechanical and allied systems.

The standard GMR, BLR, RBF and MLP models were also used in the prognostics of slow rotating bearings. These models could also be modified and optimised for further testing of their effectiveness in predicting RUL of slow rotating bearings more accurately. If the results are found to be plausible, then the new improved models could also be applied to other mechanical and allied systems.

This study can be considered as the first investigative step which considers a single application of the method to the specific type of bearing and to unique specimens and therefore its effectiveness has to be proved with further investigations.

References

Al-Ghamd, A.M. and Mba, D. (2006). A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing*, 20, 1537–1571.

An, D., Choi, J-H. and Kim, N.H. (2012). A comparison study of methods for parameter estimation in the physics-based prognostics. Annual Conference of Prognostics and Health Management Society, Hyatt Regency Minneapolis, Minnesota, USA, 23 –27 September, 2012.

An, D., Kim, N.H. and Choi, J-H. (2013). Options for prognostics methods: A review of data-driven and physics-based prognostics. Annual Conference of the Prognostics and Health Management Society, New Orleans, 14-17 October, 2013.

Arlot, S. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.

Al-Raheem, K.F. and Abdul-Karem, W. (2010). Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis. *International Journal of Engineering, Science and Technology*, 2(6), 278–290.

Artes, M., Del Castillo, L. and Perez, J. (2003). Failure prevention and diagnosis in machine elements using cluster. Proceedings of the Tenth International Congress on Sound and Vibration, Stockholm, Sweden, 1197–1203.

Banjevic, D. and Jardine, A.K.S. (2006). Calculation of reliability function and remaining useful life for a Markov failure time process. *IMA Journal of Management Mathematics*, 17, 115–130.

Barber, D. and Bishop, C.M. (1998). Variational learning in Bayesian neural networks. In: *Neural Networks and Machine Learning*, C. M. Bishop, ed. Springer, New York, 215–237.

Baruah, P. and Chinnam, R.B. (2005). HMM for diagnostics and prognostics in machining processes. *International Journal of Production Research*, 43(6), 1275–1293.

Batko, W. (1984). Prediction method in technical diagnostics. Doctoral Thesis, Cracow Mining Academy.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Clarendon, Oxford.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer, Cambridge, U.K.

Bolander, N., Qiu, H., Eklund, N., Hindle, E. and Rosenfeld, T. (2009). Physics-based remaining useful life prediction for aircraft engine bearing prognosis. Annual Conference of the Prognostics and Health Management Society, San Diego, CA, 27 September – 1October , 2009.

Brotherton, T. (2000). Prognosis of faults in gas turbine engines, in Proceedings of IEEE Aerospace Conference, 6, 163–171.

Buntine, W.L. and Weigend, A.S. (1991). Bayesian back-propagation. *Complex Systems*, 5(6), 603–643.

Cai, H.C., Wang, D.F., Xu, Q.Y. and Shang, Z.D. (2011). The acoustic emission diagnose and analysis of journal bearings inner ring looseness fault by truck wheel. *Applied Mechanics and Materials*, 130–134, 54–57.

Calinon, S. (2009). *Robot Programming by Demonstration: A Probabilistic Approach*. EPFL/CRC Press.

Camci, F. and Chinnam, R.B. (2010). *Process Monitoring, Diagnostics and Prognostics in Machining Processes.* LAP Lambert Academic Publishing: Saarbrücken, Germany.

Camci, F., Medjaher, K., Zerhouni, N. and Nectoux, P. (2012). Feature evaluation for effective bearing prognostics. *Quality and Reliability Engineering International*, 1–15. DOI : 10.1002/qre.1396.

Camci F., Medjaher K., Zerhouni N. and Nectoux P. (2013). Feature Evaluation for Effective Bearing Prognostics, Quality and Reliability Engineering International, 29, 477–486

Capgo Pty Ltd. (2009). Fatigue detection by acoustic emissions, Capgo Pty Ltd, http://www.capgo.com/Resources/ConditionMonitoring/Acoustic.html.

Cempel, C. (1987). Simple condition forecasting techniques in vibro-acoustical diagnostics. *Mechanical Systems and Signal Processing*, 1, 75–82.

Chatzis, S.P., Korkinof, D. and Demiris, Y. (2012). A non parametric Bayesian approach toward robot learning by demonstration. *Robotics and Autonomous Systems*, 60(6), 789–802.

Chebil, J., Noel, G., Mesbah, M. and Deriche, M. (2009). Decomposition for the detection and diagnosis of faults in rolling element bearings. *Jordan Journal of Mechanical and Industrial Engineering Wavelet*, 3 (4), 260 –267.

Chen, C., Zhang, B. and Vachtsevanos, G. (2012). Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms. *IEEE Transactions on Instrumentation and Measurement*, 61(2), 297–306.

Chen, S.L., Craig, M., Wood, R.J.K., Wang, L., Callan, R. and Powrie, H.E.G. (2008). Bearing condition monitoring using multiple sensors and integrated data fusion techniques. Proceedings of the Ninth International Conference in Vibrations in Rotating Machinery, Oxford, UK, Chandos, 586–600.

Chen, T. and Ren, J. (2009). Bagging for Gaussian process regression. *Neurocomputing*, 72(7–9), 1605–1610.

Chua, C.G. and Goh, A.T.C. (2003). A hybrid Bayesian back-propagation neural network approach to multivariate modelling. *International Journal for Numerical and Analytical Methods in Geomechanics*, 27, 651–667.

Di Maio, F., Ng, S.S.Y., Tsui, K.L. and Zio, E. (2011). Naïve Bayesian classifier for on-line remaining useful life prediction of degrading bearings. Proceedings of the 7th International Conference on Mathematical Methods and Reliability (MMR), Beijing, China, 20–24 June, 2011.

Dong, M. and He, D. (2007). A segmental hidden semi-Markov model (HSMM) – based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 21(5), 2248–2266.

Dong, S. and Luo, T. (2013). Bearing degradation process prediction based on the PCA and optimised LS-SVM model. *Measurement*, 46, 3143–3152.

Dube, A.V., Dhamande, L.S. and Kulkarni, P.G. (2013). Vibration based condition assessment of rolling element bearings with localized defects. *International Journal of Scientific and Technology Research*, 2(4), 149–154.

Dyer, D. and Stewart, R.M. (1978). Detection of rolling element bearing damage by statistical vibration analysis. Proceedings of Design Engineering Technical Conference of ASME, Chicago III, 20–30 September, 1978.

Ebden, M. (2008). Gaussian processes for regression: A quick introduction. *Robotics Research Group*, Oxford, UK.

Eker, O.F., Camci, F., Guclu, A., Yilboga, H., Sevkli, M. and Baskan, S. (2011). A simple state-based prognostic model for railway turnout systems. *IEEE Transactions on Industrial Electronics*, 58(5), 1718–1726.

Eker, O.F., Camci, F. and Jennions, I.K. (2012). Major challenges in prognostics: Study on benchmarking prognostics datasets. European Conference of Prognostics and Health Management Society, Minneapolis, Minnesota, USA, 23–27 September, 2012.

Elforjani, M. and Mba, D. (2008). Detecting the onset, propagation and location of non-artificial defects in a slow rotating thrust bearing with acoustic emission, Insight - Non-Destructive Testing and Condition Monitoring, 50(5), 264–268.

Elforjani, M. (2010). Condition monitoring of slow speed rotating machinery using acoustic emission technology. *PhD Thesis, Cranfield University*, UK.

Elforjani, M. and Mba, D. (2011). Condition monitoring of slow-speed shafts and bearings with acoustic emission. *Strain*, 47(2), 350–363.

Enzo, C.C.L. and Ngan, H.W. (2010). Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis. *IEEE Transactions on Instruments and Measurement*, 59, 10, 2683–2690.

Envirocoustics S. A. (2007) Acoustic Emission Theory, Envirocoustics S. A. Publications, http://www.envirocoustics.gr/acoustic_emission_theory_eng.htm.

Ferreira, J.L.A., Balthazar, J.C. and Araujo, A.P.N. (2003). An investigation of rail bearing reliability under real conditions of use. *Engineering Failure Analysis*, 10, 745–758.

Frank P.M., Ding S.X. and Marcu T. (2000). Model-based fault diagnosis in technical processes. Transactions of the Institute of Measurement and Control, 22(1), 57–101.

Garga, A.K., McClintic, K.T., Campbell, R.L., Yang, C.C., Lebold, M.S., Hay, T.A. and Byington, C.S. (2001). Hybrid reasoning for prognostic learning in CBM systems. *IEEE Aerospace Conference Proceedings*, New York, 1–7 March, 2001, 2957–2969.

Gebraeel, N., Lawley, M., Liu, R. and Parmeshwaran, V. (2004). Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Transactions on Industrial Electronics*, 51, 694–700.

Gebraeel, N., Lawley, M., Rong, L.I. and Ryan, J. (2005). Residual life distributions from component degradation signals: a Bayesian approach. *IIE Transactions*, 37, 543–557.

Gebraeel, N.Z. and Lawley, M.A. (2008). A neural network degradation model for computing and updating residual life distributions. *IEEE Transactions on Automation Science and Engineering*, 5(1), 154–163.

Ghahramani, Z. and Jordan, M.I. (1994). Supervised learning from incomplete data via an EM approach. In: Cowan, J.D., Tesauro, G.T. and Alspector, J. (eds), *Advances in Neural Information Processing systems*, San Mateo, CA: Morgan Kaufmann, 6, 120–127.

Goebel, K., Saha, B. and Saxena, A. (2008). A comparison of three data - driven techniques for prognostics. Proceedings of the 62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT), May 6-8,Virginia Beach, VA.

Haykin, S. (1999). *Neural Networks*. Prentice-Hall, Englewood Cliffs, NJ.

Hazan, A., Verleysen, M., Cottrell, M. and Lacaille, J. (2011). Bayesian inference for outlier detection in vibration spectra with small learning dataset, Surveillance 6 International Conference, 25–26 October, 2011.

He, Y., Friswell, M.I. and Zhang, X. (2009). Defect diagnosis for rolling element bearings using acoustic emission. *Journal of Vibration and Acoustics*, 131(6).

Heng, A., Zhang, S., Tan, A.C.C. and Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23, 724–739.

Heng, R.B.W. and Nor, M.J.M. (1998). Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition. *Applied Acoustics*, 53(1–3), 211–226.

Heyns, T., de Villiers, J.P. and Heyns, P.S. (2012a). Consistent haul road condition monitoring by means of vehicle response normalisation with Gaussian processes. *Engineering Applications of Artificial Intelligence*, 25(8), 1752–1760.

Heyns, T., Godsill, S.J., de Villiers, J.P. and Heyns, P.S. (2012b). Statistical gear health analysis which is robust to fluctuating load and operating speeds. *Mechanical Systems and Signal Processing*, 27, 651–666.

Hippert, H.S. and Taylor, J.W. (2010). An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting, *Neural Networks*, 23, 386–395.

Hong, S. and Zhou, Z. (2012a). Remaining useful life prognosis of bearing based on a Gauss process regression. 2012 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012), Chongqing, China.

Hong, S. and Zhou, Z. (2012b). Application of Gaussian process regression for bearing degradation assessment. Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on New Trends, Taipei, 23-25 Oct. 2012, 644–648.

Hort, F., Mazal, P. and Vlasic, F. (2010). Monitoring of acoustic emission signal of loaded axial bearings. Proceedings of European Working Group on Acoustic Emissions (EWGAE2010), Conference, Vienna, Austria, 8-10 September, 2010.

Huang, R.Q., Xi, L.F., Li, X.L., Liu, C.R., Qiu, H. and Lee, J. (2007). Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, 21(1), 193–207.

Jalan, A.K. and Mohanty, A.R. (2009). Model based fault diagnosis of a rotor–bearing system for misalignment and unbalance under steady-state condition, Journal of Sound and Vibration, 327, 604–622

Jammu, N.S. and Kankar, P.K. (2011). A review on prognosis of rolling element bearings. *International Journal of Engineering Science and Technology* (IJEST), 3(10), 7497–7503.

Jamaludin, N. and Mba D. (2002). Monitoring extremely slow rolling element bearings: part I, *NDT & E International*, 35(6), 349–358.

Jantunen, E. (2004). Prognosis of rolling bearing failure based on regression analysis and fuzzy logic. Proceedings of the VETOMAC–3 and ACSIM–2004, New Delhi, India, 836–846.

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

Kacprzynski, G.J., Sarlashkar, A., Roemer, M.J., Hess, A. and Hardman, B. (2004). Predicting remaining life by fusing the physics of failure modeling with diagnostics. *Journal of the Minerals, Metals and Materials Society*, 56(3), 29–35.

Kakishima, H., Nagatomo, T., Ikeda, H., Yoshioka, T. and Korenaga, A. (2000). Measurement of acoustic emission and vibration of rolling bearings with an artificial defect. *Quarterly Report of RTRI*, 41(3), 127–130.

Kazmierczak, K. (1983). Application of autoregressive prognostic techniques in diagnostics. Proceedings of the Vehicle Diagnostics Conference，Tuczno, Poland, 1983.

Kim, E.Y., Tan, A.C.C., Yang, B. and Kosse, V. (2007). Experimental study on condition monitoring of low speed bearings: time domain analysis. Proceedings of the 5th Australasian Congress on Applied Mechanics (ACAM 2007), Brisbane, Australia, 10–12 December, 2007.

Lampinen, J. and Vehtari, A. (2001). Bayesian approach for neural networks – review and case studies. *Neural Networks*, 14, 257–274.

Larson, E.C., Wipf, D.P. and Parker, B.E. (1997). Gear and bearing diagnostics using neural network-based amplitude and phase demodulation. Proceedings of the 51st Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, 511–521.

Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C.E. and Figueiras-Vidal, A.R. (2010). Sparse spectrum Gaussian process regression. *Journal of Machine Learning Research*, 11, 1865–1881.

Lee, J., Abujamra, R., Jardine, A.K.S., Lin, D. and Banjevic, D. (2004). An integrated platform for diagnostics, prognostics and maintenance optimization. The IMS International Conference on Advances in Maintenance and in Modeling, Simulation and Intelligent Monitoring of Degradations, Arles, France.

Lee, J.-M., Yoo, C., Choi, S.W., Vanrolleghem, P.A. and Lee, I.-B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59, 223–234.

Lee J., Wu F., Zhao W., Ghaffari, Liao L. and Siegel D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications *Mechanical Systems and Signal Processing*, 42, 314–334.

Lees, A.W., Quiney, Z., Ganji, A. and Murray, B. (2011). The use of acoustic emission for bearing condition monitoring. Proceedings of the 9th International Conference on Damage Assessment of Structures (DAMAS): Journal of Physics: Conference Series, 305(1), 012074.

Li, B., Chow, M.-Y., Tipsuwan, Y. and Hung, J. (2000). Neural networks based motor rolling bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 47(5), 1060 – 1069.

Li, C.J. and Choi, S. (2002). Spiral gear root fatigue crack prognosis via crack diagnosis and fracture mechanics. Proceedings of the 56th Meeting of the Society of Mechanical Failures Prevention Technology, Virginia Beach, VA, 311–320.

Li, C.J. and Lee, H. (2005). Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. *Mechanical Systems and Signal Processing*, 19(4), 836–846.

Li, C.J. and Li, S.Y. (1995). Acoustic emission analysis for bearing condition monitoring. *Wear*, 185, 67–74.

Li, C.J. and Shin, H. (2004). Tracking bearing spall severity through inverse modelling. Proceedings of the ASME International Mechanical Engineering Congress, Anaheim, California, USA, 13–19 Nov., 2004.

Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S. and Liang, S. (1999). Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing*, 13(1), 103–113.

Liang, F. (2005). Bayesian neural networks for nonlinear time series forecasting. Statistics and Computing, 15, 13–29.

Liu, D., Pang, J., Zhou, J. and Pang, Y. (2012). Data driven prognostics for Lithium-ion battery based on Gaussian process regression. 2012 Prognostics and System Health Management Conference (PHM–2012), Beijing, China.

Loparo K.A., Adams M.L., Lin W., Abdel-Magied M.F., and Afshari N. (2000). Fault Detection and Diagnosis of Rotating Machinery, IEEE Transactions on Industrial Electronics, 47(5), 1005 – 1014.

Lucas, J.M. and Saccucci, M.S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32, 1–29.

Luo, J., Namburu, M., Pattipati, K., Kawamoto, M. and Chigusa, S. (2003). Model-based prognostic techniques [maintenance applications]. Proceedings AUTOTESTCON 2003, IEEE Systems Readiness Technology Conference, 330–340.

Ma, J. (2012). Development of integrated prognostics application to bearing and bevel gear life prediction. PhD thesis, Industrial Engineering and Operations Research, University of Illinois at Chicago.

Mackay, D.J.C. (1992a). Bayesian interpolation. *Neural Computation*, 4, 415–447.

Mackay, D.J.C. (1992b). A practical Bayesian framework for back propagation networks. *Neural Computation*, 4, 448–472.

Malhi, A. and Gao, R.X. (2004). PCA-based feature selection scheme for machine defect Classification. *IEEE Transactions on Instrumentation and Measurement*, 53 (6), 1517–1525.
Marble, S. and Morton, B.P. (2006). Predicting the remaining life of propulsion system bearings. Proceedings of the IEEE Aerospace Conference, Big Sky, MT.

Matérn, B. (1986). *Spatial Variation*, *Lecture Notes in Statistics*, 36, Springer-Verlag, New York.

Martin, E.B. and Morris, A.J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, 6, 349–358.

Martinez, L. and Martinez, A. (2004). *Exploratory Data Analysis with Matlab*. CRS Press, USA.

Mazal, P. (2009). Current possibilities of AE method application in selected areas of industrial practice. The 10th International Conference of the Slovenian Society for Non-Destructive Testing, Application of Contemporary Non-Destructive Testing in Engineering, Ljubljana, Slovenia,1–3 September, 2009, 285–295.

Mba, D. (2003). Acoustic emissions and monitoring bearing health. *Tribology Transactions*, 46(3), 447–451.

Mba, D. (2005). Prognostic opportunities offered by acoustic emission for monitoring bearings and gearboxes. Twelfth International Congress on Sound and Vibration, Lisbon, 11–12 July, 2005.

Mba, D., Bannister, R.H. and Findlay, G.E. (1999). Condition monitoring of low-speed rotating machinery using stress waves: Parts I and II. *Proceedings of the Institution of Mechanical Engineers*, 213(E), 153–185.

Miettinen, J. (2000). The influence of the running parameters on the acoustic emission of grease lubricated rolling bearings. PhD Thesis Tampere University of Technology.

Miettinen, J. and Pataniitty, P. (1999). Acoustic emission in monitoring extremely slowly rotating rolling bearing. Proceedings of COMADEM, Oxford, Coxmoor Publishing Company, 289–297.

Mohamad, A.K., Saon, S. and Hiyama, T. (2010). Predicting remaining useful life of rotating machinery based artificial neural network, *Computers & Mathematics with Applications,* 60, 1078-1087.

Morhain, A. and Mba, D. (2003). Bearing defect diagnosis and acoustic emission. Proceedings of the Institution of Mechanical Engineers, Part J: *Journal of Engineering Tribology*, 217(4), 275–272.

Mosallam A., Medjaher K. and Zerhouni N. (2013). Nonparametric time series modelling for industrial prognostics and health management, International Journal of Advances in Manufacturing Technology, 69, 1685–1699

Mozetič I. (1992). Model-based diagnosis: An overview, Advanced Topics in Artificial Intelligence - Lecture Notes in Computer Science, 617, 419-430.

Nabney, I.T. (2002). *NETLAB Algorithms for Pattern Recognition*. Springer Publication, Great Britain.

Nataraj, C. and Kappaganthu, K. (2011). Vibration-based diagnostics of rolling element bearings: state of the art and challenges. 13th World Congress in Mechanism and Machine Science, Guanajua´to, Mexico, 19–25 June, 2011.

Neal, R.M. (1996). *Bayesian Learning for Neural Networks* (Lecture Notes for Statistics). Springer.

Nienhaus, K., Boos, F.D., Garate K. and Baltes, R. (2012). Development of acoustic emission (AE) based defect parameters for slow rotating roller bearings, 25th

International Congress on Condition Monitoring and Diagnostic Engineering, Journal of Physics: Conference Series, 364, doi:10.1088/1742–6596/364/1/012034.

Oppenheimer, C.H. and Loparo, K.A. (2002). Physically based diagnosis and prognosis of cracked rotor shafts. Proceedings of SPIE Component and Systems Diagnostics, Prognostics, and Health Management, Bellingham, WA, 2002, II, 122–132.

Opsomer, J., Wang, Y. and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16(2), 134–153.

Orchard, M., Wu, B. and Vachtsevanos, G. (2005). A particle filter framework for failure prognosis. Proceedings of the World Tribology Congress, Washington, DC, 2005.

Orsagh, R., Roemer, M. and Sheldon, J. (2004). A comprehensive prognostics approach for predicting gas turbine engine bearing life. Proceedings of the 10th National Turbine Engine High Cycle Fatigue Conference, New Orleans, LO, 2004.

Orsagh, R.F., Sheldon, J. and Klenke, C.J. (2003). Prognostics/diagnostics for gas turbine engine bearings. Proceedings of the IEEE Aerospace Conference, Big Sky, MT, 2003, 7, 3095–3103.

Palm, F.C. and Zellner, A. (1992). To Combine or not to Combine? Issues of Combining Forecasts. *Journal of Forecasting*, 11, 687–701.

Peng Z., Kessissoglou N.J., Cox M. (2005). A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques, *Wear*, 258, 1651–1662.

Penny, W.D. and Roberts, S.J. (1999). Bayesian neural networks for classification: how useful is the evidence framework? *Neural networks*, 12, 877–892.

Phelps, E., Willett, P. and Kirubarajan, T. (2001). A statistical approach to prognostics. In Proceedings of SPIE, Component and Systems Diagnostics, Prognosis, and Health Management, Peter K. W. and Thiagalingam, K. (Eds), 4389, 23–34.

Qi, H., Bo, Y., Lu-ping, L. and Hai, Z. (2008). Experimental study on condition diagnosis of sliding bearings based on acoustic emission detection. *Power System Engineering*.

Qiu, H., Lee, J., Lin, J. and Yu, G. (2003). Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics*, 17, 127–140.

Qiu, J., Set, B.B., Liang, S.Y., and Zhang, C. (2002). Damage mechanics approach for bearing life time prognostics. *Mechanical Systems and Signal Processing*, 16(5), 817–829.

Randall, R.B. (2011). *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. Wiley, West Sussex, UK.

Randall R.B. and Antoni J. (2011). Rolling element bearing diagnostics — A tutorial *Mechanical Systems and Signal Processing*, 25, 485–520.

Rasmussen, C.E. (1996). Evaluation of Gaussian processes and other methods for non-linear regression. PhD Thesis, Department of Computer Science, University of Toronto.

Rasmussen, C.E. and Nickisch, H. (2013). Gaussian processes for machine learning tool box version 3.2.

Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge MA.

Roemer, M.J., Hong, C. and Hesler, S.H. (1996). Machine health monitoring and life management using finite element-based neural networks. *Journal of Engineering for Gas Turbines and Power-Transactions of the ASME*, 118, 830–835.

Romdhani, S., Gong, S. and Psarrou, A. (1999). A multi-view nonlinear active shape model using kernel PCA. Proceedings of BMVC, Nottingham.

Rossi, V. and Vila, J.P. (2006). Nonlinear filtering in discrete time: A particle convolution approach. *Annales de l'I.S.U.P. Publications de l'Institut de Statistique de l'Université de Paris (Ann. I.S.U.P.),* 50(3), 71–102.

Saha, B., Goebel, K., and Christophersen, J. (2009). Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement and Control*, 31(3–4), 293–308.

Saha, S., Saha, B., Saxena, A. and Goebel, K. (2010). Distributed prognostic health management with Gaussian process regression. IEEE Aerospace Conference, 2010, 1–8.

Saxena, A., Celaya, J., Saha, B., Saha, S. and Goebel, K. (2009). On applying the prognostic performance metrics. Annual Conference of the Prognostics and Health Management Society, 2009, 1–13.

Schölkopf, B., Smola, A., and Muller, K.R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1319.

Schölkopf, B., Smola, A. and Muller, K.R. (1999). *Kernel principal component analysis*. In *Advances in Kernel Methods*-Support Vector Learning, MIT Press, 327–352.

Schurmann J. (1996). *Pattern Recognition*: *A Unified View of Statistical And Neural Approaches*. Wiley, New York.

Şengüler, T., Karatoprak, E. and Şeker, S. (2010). A new MLP approach for the detection of the incipient bearing damage. *Advances in Electrical and Computer Engineering*, 10(3), 34–39.

Shao, Y. and Nezu, K. (2000). Prognosis of remaining bearing life using neural networks. *Proceedings of the Institution of Mechanical Engineers, Systems and Control Engineering*, 214(3), 217–230.

Shen, Z., He, Z., Chen, X., Sun, C. and Liu, Z. (2012). A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time. *Sensors*, 12, 10109–10135.

Shi, J.Q. and Wang, B. (2007). Gaussian process functional regression modelling for batch data. *Biometrics*, 63, 714–23.

Shiroishi J., Li Y., Liang S. and Kurfess T. (1997). Bearing condition diagnostics via vibration and acoustic emission measurements. *Mechanical Systems and Signal Processing*, 11(5), 693–705.

Shu-kai Q., Xue-peng F. and Xiao-bo C. (2008). Fault diagnosis method based on the EWMA dynamic kernel principal component analysis. Proceedings of the IEEE 2008 Chinese Control and Decision Conference (CCDC), 463–467.

Si, X.S., Wang, W., Hu, C-H. and Zhou, D-H., (2011). Remaining useful life estimation- A review of the statistical data driven approaches. *European Journal of Operational Research*, 213, 1–14.

Sikorska J.Z., Hodkiewicz, M. and Ma, L., (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25, 1803–1836.

Skabar, A. (2007). Mineral potential mapping using Bayesian learning for multilayer perceptrons. *Mathematical Geology*, 39, 439–451.

Skormin, V.A., Popyack, L.J., Gorodetski, V.I., Araiza, M.L. and Michel, J.D. (1999). Applications of cluster analysis in diagnostics-related problem. Proceedings of the IEEE Aerospace Conference, Snowmass at Aspen, CO, USA, 1999, 3, 161–168.

Stein, M.L. (1999). *Interpolation of Spatial Data*. Springer-Verlag, New York.

Sun, C., Zhang, Z., Cheng, W., He, Z., Shen, Z. Chen, B. and Zhang, L. (2013a). Manifold subspace distance derived from kernel principal angles and its applications to machinery structural damage assessment. *Smart Materials and Structures*, 22, 1–12.

Sun, D., Lu, G., Zhou, H. and Yan, Y. (2013b). Condition monitoring of combustion processes through flame imaging and kernel principal component analysis. *Combustion Science and Technology*, 185(9), 1400–1413.

Tandon, N. and Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32(8), 469–480.

Tandon, N. and Nakra, B.C. (1992a). Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings. *Tribology International*, 25(3), 205–212.

Tandon, N. and Nakra, B.C. (1992b). Vibration and acoustic monitoring techniques for the detection of defects in rolling element bearings – A review. *The Shock and Vibration Digest*, 24(3), 3–11.

Tavakoli, M.S. (1991). Bearing fault detection in acoustic emission frequency range. Proceedings of the 11th National Conference on Noise Control Engineering, Tarrytown, NY, 14–16 July, 1991.

Thodberg, H. H. (1996). A review of Bayesian neural networks with an application to near infrared spectroscopy. *IEEE Transactions on Neural Networks*, 7(1), 56–72.

Titterington, D.M. (2004). Bayesian methods for neural networks and related models. *Statistical Science*, 19(1), 128–139.

Tobon-Mejia D.A., Medjaher K., Zerhouni N. and Tripot G. (2012). A Data-Driven Failure Prognostics Method based on Mixture of Gaussians Hidden Markov Models, *IEEE Transactions on reliability,* 2, 1-16.

Tse, P. and Atherton, D. (1999). Prediction of machine deterioration using vibration based fault trends and recurrent neural networks. *Transactions of the ASME: Vibration and Acoustics*, 121(3), 355–362.

Vila, J-P., Wagner, V., and Neveu, P. (2000). Bayesian nonlinear model selection and neural networks: a conjugate prior approach. *IEEE Transactions on Neural Networks*, 11(2), 265–278.

Vlok, P.-J., Wnek, M. and Zygmunt, M. (2004). Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical Systems and Signal Processing*, 18(4), 833–847.

Verbruggen, T.W.(2003). Wind turbine operation & maintenance based on condition monitoring. ECN-C-03-047. 2003. WT-OMEGA, ECN Wind Energy.

Walker, R., Perinpanayagam, S. and Jennions, I.K. (2013). Rotordynamic faults: Recent advances in diagnosis and prognosis, *International Journal of Rotating Machinery*, 2013, 1-12.

Wang, G., Qian, L. and Guo, Z. (2013). Continuous tool wear prediction based on Gaussian mixture regression model. *International Journal of Advance Manufacturing and Technology*, 66, 1921–1929.

Wang, J.M., Fleet, D.J. and Hertzmann, A. (2008). Gaussian process dynamical models for human motion. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 30(2), 283–298.

Wang, P. and Vachtsevanos, G. (2001). Fault prognostics using dynamic wavelet neural networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 15(4), 349–365.

Wang, Q. (2012). Kernel principal component analysis and its applications in face recognition and active shape models. arXiv:1207.3538v1 [cs.CV] 15 Jul 2012.

Wang, M. and Wang, J. (2012). CHMM for tool condition monitoring and remaining useful life prediction. *International Journal of Advanced Manufacturing Technology,* 59, 463–471.

Wang, W. (2002). A model to predict the residual life of rolling element bearings given monitored condition information to date. *IMA Journal of Management Mathematics*, 13, 3–16.

Wang, W. (2007). An adaptive predictor for dynamic system forecasting. *Mechanical Systems and Signal Processing*, 21, 809–823.

Wang, W.Q., Golnaraghi, M.F., and Ismail, F. (2004). Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18(4), 813 –831.

Warrier, S.G., Jarmon, D.C. and Chin, H.A. (2000). Finite element analysis of the critical flaw size in hybrid silicon nitride bearing ball. Proceedings of ASME Turbo Expo 2000, No. 2000-GT-65, Munich, Germany, 2000.

Weinberger, K.Q., Sha, F. and Saul, L.K. (2004). Learning a kernel matrix for nonlinear dimensionality reduction. Proceedings of the 21st International Conference on Machine Learning (ICML-04), ACM Press, New York, NY, USA, 839–846.

Wemhoff, E., Chin, H. and Begin, M. (2007) Gearbox diagnostics development using dynamic modelling. AHS 63rd Annual Forum, Virginia Beach, VA.

Widodo, A., Kim, E.Y., Son, J., Yang, B., Tan, A.C.C., Gu, D., Choi, B. and Mathew, J. (2009). Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Systems with Applications*, 36, 7252–7261.

Wold, S. (1994). Exponentially weighted moving principal components and projections to latent structures. *Chemometrics and Intelligent Laboratory Systems*, 23, 149–161.

Xi, F., Sun, Q. and Krishnappa, G. (2000). Bearing diagnostics based on pattern recognition of statistical parameters. *Journal of Vibration Control*, 6, 75–92.

Yam, R.C.M., Tse, P.W., Li, L. and Tu, P. (2001). Intelligent predictive decisions support system for CBM. *The International Journal of Advanced Manufacturing Technology*, 17(5), 383–391.

Yoo, C.K. and Lee, I.B. (2006). Nonlinear multivariate filtering and bioprocess monitoring for supervising nonlinear biological processes, *Process Biochemistry*, 41, 1854–1863.

Yoshioka, T. (1992). Detection of rolling contact subsurface fatigue cracks using acoustic emission technique. *Lubrication Engineering,* 49, 303–308.

Yu, J-B. (2011a). Bearing performance degradation assessment using locality preserving projections. *Expert Systems with Applications*, 38, 7440–7450.

Yu, J. (2011b). Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. *Mechanical Systems and Signal Processing*, 25(7), 2573–2588.

Zaidan, M.A., Mills, A.R. and Harrison, R.F. (2011). Towards enhanced prognostics with advanced data-driven modelling, In: Proceedings of The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies. The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 20–22 June 2011, Cardiff.

Zhang, B., Sconyers, C., Byington, C., Patrick, R., Orchard, M.E. and Vachtsevanos, G. (2001). A probabilistic fault detection approach: application to bearing fault detection. *IEEE Transactions on Industrial Electronics,* 58(5), 2011–2018.

Zhang, S. and Ganesan, R. (1997). Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery. Transactions of the ASME, *Journal of Engineering for Gas Turbines and Power*, 119, 378–384.

Zhang, S., Ma, L., Sun, Y. and Mathew, J. (2007). Asset health reliability estimation based on condition data. Proceedings of the 2nd World Congress on Engineering Asset Management, (WCEAM) and the 4th ICCM, Harrogate, UK, 2195–2204.

Zhang, X., Xu, R., Kwan, C., Liang, S.Y., Xie, Q. and Haynes, L. (2005). An integrated approach to bearing fault diagnostics and prognostics. Proceedings of American Control Conference, Portland, OR, 2750–2755.

Zhu, J., Yoon, J.M., He, D., Qu, Y. and Bechhoefer, E. (2013). Lubrication oil condition monitoring and remaining useful life prediction with particle filtering, International Journal of Prognostics and Health Management, 1–15.

Appendix

Peak to peak

This is the value from one peak to the other peak and computed as given in Equation A1 below:

$$peak - to - peak = \frac{1}{2}\left(\max(x(t)) - \min(x(t))\right) \tag{A1}$$

where $x(t)$ is the AE time signal and t is the duration of the measurement.

the mean value of the time signal x (t) having N data points

Root mean square

The root mean square (RMS) is the square root of the average of the squares of the values. The RMS of the AE signals is calculated as given in Equation A2:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i) - \bar{x})^2} \tag{A2}$$

where N is the number of data points and $\bar{x}$ is the mean value of the $x(t)$ AE time signal.

Crest Factor

The crest factor (CF) is the ratio of the peak (maximum RMS) value to the RMS value and is computed as given in Equation A3:

$$CF = \frac{peak}{RMS} \tag{A3}$$

Kurtosis

Kurtosis is computed as given in Equation A4:

$$Kurtosis = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i) - \bar{x})^4}}{RMS^4} \tag{A4}$$