University of Strathclyde

Department of Mechanical &Aerospace Engineering

Data – driven methodologies for bearing vibration analysis and vibration based fault diagnosis

Hussein Razzaq Al-Bugharbee

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2016

Declaration of author's rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Hussein Razzaq Al-Bugharbee

2016

Acknowledgments

First of all, all the praises be to Allah, the Most Beneficent, the Most Merciful for providing me the blessings to complete this work. Then, I am truly indebted and thankful to the following colleagues, friends and family who have contributed their advice and support towards the development of this research. I would like to express my special thanks of gratitude to my supervisor, Dr. Irina Trendafilova, who gave me this great opportunity to discover the machinery condition monitoring world. She continuously encouraged and guided me through my research programme. The Iraqi Ministry of Higher Education and Scientific Research, for funding this research. The technical staff for their expertise, extensive help and assistance in the workshop. Staff at the School of Engineering, University of Strathclyde, for their assistance, help and accommodating my demanding needs on resources throughout my studies.

I would like to express my sincere thanks to all the friends and colleagues, Dr. David Garcia, for offering help whenever I need, Dr Hassan AL-Budairi, Mr.Manaf Al-Taleb, and Dr. Ahmed Al –Waaly, for their spirited support and encouragement has been of great help to me throughout the relentless endurance test of my PhD, Dr. Abbas Juber for his continuous calls from Iraq to provide support.

Most especially to my big family, the soul of my father, my lovely mother, sisters and brothers, and to my small family, my wonderful wife Aqbal and my little two angels, Ruqayah and Zahraa, for their understanding and support. Words alone cannot express

what I owe them for their encouragement and patient love that inspired me to complete this work.

Abstract

Rolling element bearings (REBs) are one of the most critical mechanical components. Their failures can lead for catastrophic failures which might include great loss in economy or even in the lives of people. REBs are inherently dynamic and they demonstrate complex vibration behaviour where conventional vibration –based fault diagnosis methods might not give sensitive indicators of the presence of the defects. This thesis investigates the singular spectrum analysis (SSA) capabilities as completely data-based fault diagnosis method in REBs. The SSA is used to decompose the bearing vibration acceleration signals in a certain number of principal components having the trend, periodical components and structure-less noise. This thesis develops two methodologies to use SSA in different ways and for different purposes. The first methodology uses the SSA (i.e only the decomposition stage) to create a baseline space from healthy bearing vibration signals. Then, any new signals are projected onto this baseline space. From these projections, features are made and used for fault diagnosis purposes. In the second methodology, the SSA contributes to the development of an advanced signal pretreatment that efficiently improves representing the nonstationary bearing vibration signals by linear time invariant autoregressive (LTIVAR) model. Then the coefficients of LTIVAR model are used as features for fault diagnosis purposes.

The two methodologies have been validated by using experimental data obtained from three different bearing test rigs. The data used in the analysis covers different defect locations and different defect severities. The results of both methodologies, in terms of correct classification, were compared to some other recent methodologies. In

comparison, it is shown that both methodologies have a very good performance and they are superior to those methodologies.

The thesis offers simple and efficient methodologies for a complete fault diagnosis in terms of fault detection, identification and severity estimation. Thus, these methodologies have a potential possibility for automation of the entire process of each method.

Publications

Fully referenced international journal papers

- H. Al-Bugharbee, I. Trendafilova, *A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling*, Journal of Sound and Vibration, 369 (2016) 246-265.
- A.A. Tabrizi, H. Al-Bugharbee, I. Trendafilova, L. Garibaldi (2016), *A cointegration-based monitoring method for rolling bearings working in time-varying operational conditions*, Meccanica, 1-17. 10.1007/s11012-016-0451-x
- Hussein. Al-Bugharbee and Irina.Trendafilova (2016).*'' A new methodology for fault detection in rolling element bearings using singular spectrum analysis''* submitted to the International journal of condition monitoring (IJCM) (*submitted under review*)

Fully referenced conference papers

- H.Al_Bugharbee and I.Trendafilova, *Fault diagnosis in roller element bearings by using a linear autoregressive model*, in proceedings of ISMA 2014-International Conference on Uncertainity in Structural Dynamics, 2014, pp.2765-2776
- David Garcia, Irina Trendafilova, Hussein Al-Bugharbee. *Vibration-Based Health Monitoring Approach for Composite Structures Using Multivariate Statistical Analysis*. Le Cam, Vincent and Mevel, Laurent and Schoefs, Franck. EWSHM - 7th European Workshop on Structural Health Monitoring, Jul 2014, Nantes, France. 2014.
- H. Al-Bugharbee, and I. Trendafilova (2015), "*Autoregressive modelling for rolling element bearing fault diagnosis*." Journal of Physics: conference Series, V 628, 012088.

Contents

List of figures

List of tables

Nomenclature

a_i	The ith LTIVAR model coefficient
C	Constant load
C_{ij}	percentage of vectors from class i which are classified as class j
D	A vector included the Mahalanobis distance of the baseline feature vectors
D_i	Mahalanobis distance of the i^{th} feature vector to the baseline feature matrix $\mathbf{F}_{baseline}$
D_{Ki}	Mahalanobis distance of **fv** to the feature matrix **F**k i
$\mathbf{E}_{baseline}$	The mean of the baseline feature matrix rows.
$\mathbf{E}_K$	The mean of the feature matrix rows.
El	An elementary matrix
F	Faulty bearing class
$\mathbf{F}_{baseline}$	baseline feature matrix
f_{ij}	Euclidean norm of the i^{th} observation and the j^{th} principal component.
$\mathbf{F}_K$	feature matrix of the ith a category.
$\mathbf{fv_i}$	ith feature vector
H	Healthy bearing category

K	A length of a principal component
k	is a number of signals corresponding to a baseline condition
L	Embedding window size
L	Large fault class
L_{10}	Estimated bearing life
L_m	Lower bound of the summation in the diagonal averaging technique
μ	Mean of the lognormal probability density function
n_o	A constant to be determined from the bearing type
n	Sub-signal length
N_m	Normalising factor of the diagonal averaging technique
P	Equivalent dynamic load
p	Model order
$p(D_i)$	is lognormal probability density function value at D_i.
$P(D_i)$	Is cumulative lognormal probability density function value at D_i.
$\mathbf{PC_i}$	A principal component vector
p_{max}	maximum optimum order
p_{min}	Minimum optimum order
Q	Number of PCs selected to build the baseline space
$\mathbf{S}^{-1}$	The inverse of the covariance of the baseline feature matrix $\mathbf{F}_{baseline}$
$Thr_{baseline}$	A threshold made from Mahalanobis distance of the baseline feature vectors

U	An Eigen vector
U_m	upper bound of the summation in the diagonal averaging technique
V	the variance of the difference between the predicted and the measured sub-signal
x	A sub-signal of length (n)
X	Trajectory/embedding matrix
$\overline{xr}_{ms}$	the mean value of xr_{ms}
xr^{dif}	Differenced noise cleaned signal
xr_{ms}	the stationary real time measured sub-signal
xr_{pr}	Predicted stationary sub-signal
xr	A reconstructed sub-signal
$\mathbf{X^T}$	A transpose of **X**
α	Risk level
λ	An Eigen value
σ	Standard deviation of the lognormal probability density function
ε	error term which represents the difference between the real and predicted the *m* value of a sub signal
Nr	Number of shaft revolution at a sub-signal length(*n*)
Sp	Shaft rotational speed (RPM)
Sr	Sampling rate (Hz)

Acronyms

AD	Amplitude demodulation
AE	Acoustic emission
ANN	Artificial neural network
AR	Autoregressive model
ARMA	Autoregressive moving average model
BF	Ball fault
BPNN	Back propagation neural network
CBM	Condition based monitoring
CM	Condition monitoring
CRWU	Case reserve western university
DH	Difference histograms
EMD	Empirical mode decomposition
FFNN	feed forward neural network
FFT	Fast Fourier transform
FPE	Final prediction error
FV	A feature vector
Hz	Hertiz
IRF	Inner race fault

KPSS	Kwiatkowski–Phillips–Schmidt–Shin test
LBF	Large ball fault
LIRF	Large inner race fault
LORF	Large outer race fault
LTIVAR	Linear time invariant autoregressive model
M	Medium fault class
MA	Moving average
MBF	Medium ball fault
MD	Mahalanobis distance
MIRF	Medium inner race fault
MORF	Medium outer race fault
MVE	minimum volume ellipsoid
NMSE	normalised mean square error
NN	Nearest neighbour
ORF	Outer race fault
PC	Principal component
PCA	Principal component analysis
REB	Rolling element bearing
REBs	Rolling element bearings
S	Small fault class

SBF	Small ball fault
SHM	Structural health monitoring
SIRF	Small inner race fault
SORF	Small outer race fault
SSA	Singular spectrum analysis
STFT	Short time Fourier transform
SVD	Singular value decomposition
TVAR	Time varying autoregressive
VTM	variance threshold method
WT	Wavelet transform
ZP	Zero padding

Chapter 1

Introduction

1.1 Background

All mechanical and electrical machines are generating signals when they are in an operating process. These signals are characterising those machines' condition. Thus any change in these signals may indicate a change of the machine condition. Condition monitoring (CM) is a term refers to the process of monitoring the change of a parameter or some parameters which are obtained from these characterising signals.

1.2 Importance of condition monitoring

Rotating machines are one of the popular machines in industrial productive stations such as petrochemical stations and power stations. Globally, safety, reliability, reducing the cost of products, and scheduling the time of production are the key aspects for successful productive companies. However, all machines are expected to deteriorate, no matter how well they were designed, because of the nature of severe operating condition that they work in. Failure of machines, particularly the critical ones, is one of main issues that affect negatively the production's cost and time schedules. This is because

failures cause downtime of a machine which means enormous cost for maintenance or replacement and delay in production. It can lead to a costly loss and even human causalities. These failures are usually caused by a developed small defect in one of the machine components. If these defects are not detected at an early stage, and a right decision is not taken, they develop and may lead to a catastrophic failure.

A fundamental solution to reduce the possibility of machine breakdown is the maintenance which ensures keeping the machines up to a satisfactory level of performance. A brief summary on the history of maintenance techniques is given in[1]. The earliest and most classic way of maintenance is called unplanned maintenance (it can also be called breakdown maintenance or run to failure maintenance). This kind of maintenance is applied only when a failure occurs and a machine breaks down. Another kind of maintenance is the so called planned maintenance (i.e preventive maintenance) which implies doing periodic maintenance regardless of the condition of a machine. This includes a periodic stop of the machines and changing some parts and doing other maintenance work. However, these two kinds of maintenance have some potential disadvantages, such as consuming time and spending extra money for the new parts. Furthermore, such kinds of maintenance can cause some mechanical faults throughout to the disassembling and assembling of machine parts during the periodic maintenance. They can also affect the planned schedule of production.

Usually such kinds of maintenance cause loss of money which might be estimated, as an example in some cases, as much as one third of the production costs [2].

A more efficient way of maintenance is called condition-based maintenance (CBM) which can be defined as the maintenance process of a machine based on the information obtained from the machine itself during the condition monitoring process.

The CM of a machine has several advantages; just a few to mention are [3-5] : -

- The maintenance decision based on CM helps avoiding dangerous failures, considerably reducing maintenance costs by eliminating unnecessary maintenance and eventually extending the remaining life of the machine
- CM can provide advanced information about machine defects which means it gives an opportunity to plan for a shorter time of maintenance which necessarily means shorter time of downtime of a system.

1.3 CM for Rolling element bearings (REBs)

Rolling element bearings (REBs) are one of the most critical mechanical components in the rotating machinery such as motors and pumps where more than 90% of rotating machines are using rolling elements bearings [6]. No matter how well the bearings are precisely designed, normally loaded and correctly assembled and properly lubricated they fail due to material fatigue after a certain operating time. This period of operating time is called the bearing fatigue life and it is estimated by the number of revolutions that a bearing can do before it comes to failure. The estimated bearing fatigue life can be calculated according to the formula: -

$$L_{10} = \left(\frac{C}{P}\right)^{n_o}$$

Where P is the equivalent dynamic load and it can be determined depending on the bearing geometry and the radial and thrust load components [7]. C is the constant load. The n_o is a constant which is estimated as n=3 for ball bearings and n=10/3 for roller bearings. It is mentioned in the literature that the majority of the identical test bearings have a fatigue of bearing is around one million revolutions [8]. Defects in REBs shorten their fatigue lives. These defects can be caused by either internal factors such as assembly error or outer factors such as the force from other components in the assembly. The defects of REBs are considered as one of the main reasons of rotating machinery failures. For instance, the literature shows that they form around 40% of reasons for failure in induction machines [9]. Faults in REBs can also be indicators for the existence of another fault in the machines such as the misalignment and unbalance.

The lack of fault detection at an early stage can lead to dangerous consequences. Detection of bearing faults can help to avoid the catastrophic failure of a bearing, and as a result a failure of the machine, when right decision is taken at a due time.

1.4 Bearings CM strategies

Usually, most of bearing defects occur during the operating time. Thus, monitoring the condition of the bearings and detecting such faults without dismounting the bearing or even without stopping the machine is very important. There are various approaches used for detecting defects in bearing. The condition strategies can be classified, based on the type of acquired signals, into acoustic signal analysis, temperature measurement, lubricant analysis, electrical current analysis, and vibration measurement [10]. Several strategies are developed to achieve the goals of condition monitoring. They are mainly

classified as acoustic measurements, machine current and bearing temperature monitoring, lubricants analysis and bearing vibration analysis. A brief explanation for each of these strategies is introduced below: -

1.4.1 Acoustic Emission

Acoustic emission (AE) can be defined as the generation of elastic waves due to a rapid release of strain energy. This strain energy can be induced by a deformation or damage within or on the material surface [11-13]. Although AE based defect techniques are useful, their successful application and accuracy have been partly limited to the processing, and interpreting of the acquired data[14].

1.4.2 Bearing temperature monitoring

Excessive heat can be generated in the rotating components of a bearing due to the presence of distributed defects [15, 16]. Thus monitoring of the temperature of a bearing housing or lubricant is considered as one of the simplest methods for fault detection in rotary machines.

1.4.3 Electrical Motor Current Monitoring

Monitoring the current of a machine such as induction motors, can be also used as a bearing defect detection tool [17]. The changes in electrical current of a machine may be associated with changes of the mechanical components such as a bearing [18].

1.4.4 Lubricants analysis

Some bearing defects can also be detected through the analysis of a bearing lubricant sample. The analysis of a lubricant sample usually looks for the presence of any metallic particles. Such kind of analysis is done using some special sensors [19]. Furthermore, the analysis of the different metallic elements in the lubricant can facilitate the location of the fault.

1.4.5 Bearing vibration analysis

Every machine which has moving parts has a specific vibration signature related to its state. The machine's vibration signal contains information about the global behaviour of the machine and any changes within the machine result in changes in the vibration signal. This change can be used in the detection of incipient defects before the development of the severity of the defects. For a defected bearing, the interaction between the defect and any of other mating bearing components generates a vibration signal which can be analysed for the purpose of fault diagnosis of REBs. The bearing vibration signals are considered more effective and suitable for fault diagnosis among all other bearing data [20]. For this reason, vibration-based strategy is the most popular, among the other different CM strategies, for fault diagnosis of rolling element bearings [21-23].

Other reasons for the popularity of the vibration-based strategy is that it is it is an online, and non-destructive methodology and able to successfully detect 90% of all machinery defects [3]. The vibration strategy is widely used in the industrial applications and there are different portable analysers which are based on mathematical

algorithms and are used for detecting faults in machinery. For these reasons, the work in this thesis is based on analysing bearing vibration signals. Other benefits of the bearing vibration analysis include relatively low cost sensors and a simple set up.

1.5 Research objectives

The primary objective of this research is the development of a data–driven methodology for a complete fault diagnosis in REBs. This methodology should be able to detect, identify the location and estimate the severity of the faults.

In addition, there are some other specific objectives can be defined as:-

A. To study of the capabilities of the singular spectrum analysis for the purposes of rolling element bearing fault diagnosis. This investigation includes the use of singular spectrum analysis in different ways and for different purposes.
B. To investigate a new way of applying SSA by using only the decomposition stage.
C. To demonstrate the importance of signal pre-treatment for the success of machinery diagnosis methods
D. To develop a signal pretreatment that facilitates the efficient use of simple forms of autoregression rather than complicating the structure of the auto-regression function to improve the diagnosis.
E. To validate the methodologies developed in this thesis for a variety of data sets and operational conditions.

1.6 Research motivations and contributions

One of the main interests of industries is in obtaining a relatively simple and complete fault diagnosis method for REB monitoring which is reliable enough and easy to apply. Although many interesting vibration data –based CM techniques have been developed during the last few decades, almost of them are unable to completely and precisely diagnose the present fault. So currently there is still a need for the development of simple methods that can be applied without any specific training of the personnel operating the machinery, and which give complete diagnosis of the existing fault in the sense of detection, localization (qualification) and size/extent estimation (quantification).

One of the main goals of the present research is to utilise the data-measured in the development of a simple and a complete full diagnosis. The term ‘complete’ means that the methodologies developed in this research are able to detect, localize and estimate the severity of the fault.

This research’s goal includes the development of two methodologies for extracting some features which are sensitive to the change of bearing condition due to the presence of a defect. Both of these methodologies use the singular spectrum analysis-based techniques but in different ways and for different purposes. As a result two methodologies, which are fully based on data analysis, have been developed

A new singular spectrum analysis based methodology for fault diagnosis in REBs:

In this method, the SSA is used to build a baseline space (corresponding to healthy bearing) and uses it as a reference helping in the detection of fault corresponding to different locations and different severities. The baseline space is made from the healthy bearing signals category and the decomposition stage only. When the baseline space is made, any new (i.e probably faulty) signal is projected onto this baseline space. The signals that come from the same signal categories will have similar projections while the signals come from different categories will have different projections and this will be helpful in assessing the changing of the bearing condition. The methodology helps not only in distinguishing between the baseline and non-baseline signal categories but also in distinguishing among the non-baseline categories as well. The methodology is validated for three phases of fault diagnosis (i.e detection phase, fault type identification phase and fault severity estimation phase) using data sets obtained from three different bearing test rigs. The methodology shows very good performance when compared to other recent methods in terms of correct classification rates.

Bearing fault diagnosis using a novel signal pretreatment and Linear time invariant auto regressive modelling:

This methodology suggests the use of SSA in an advanced signal pre-treatment before the use of a linear time invariant autoregressive (LTIVAR) model. This signal pre-treatment de-noises the signal by SSA and transforms the non stationary signal to a stationary one using a simple procedure called differencing. This helps the use of the simple form of modelling (i.e LTIVAR) effectively. The feature vectors (FVs) are

extracted using the coefficients of the LTIVAR model. This methodology is validated using three different bearing test rigs for the purpose of fault detection and identification and for fault severity detection. The methodology shows a very good performance when compared to other recent methods in terms of correct classification rates. The methodology and the related results are presented in chapters 6&7. Three publications are developed based on the results of this methodology [24-26].

The first motive for this research is that it is important to detect rolling element bearing defects at different locations and of different severities as this provides more time for corrections that can prevent further damage.

Another motive is the need for simplification of data-based CM techniques.

The following are the main contributions made in the thesis: -

- A very simple and efficient data analysis based method is developed in the present research (i.e in chapter 4). The simplicity of the current approach comes from the fact that only healthy signals are required for subjecting to the decomposition stage and then building the baseline space. Compared to other singular spectrum (SSA) - based techniques (i.e shown in section 3.4.1), which subject signals from different categories to SSA, the current SSA-based methodology is simpler. In this method, only the signal corresponding to the healthy bearing category is used to build a multi-dimensional space where all other signals are projected and then to be compared to the baseline (i.e the healthy state).

This method is used for a complete fault diagnosis in terms of detection, identification of location and severity estimation. The baseline space in all the levels of the fault diagnosis is made by subjecting the training sample, which is only **made from healthy data** to a decomposition stage of the SSA. The features are obtained from the norms of the PCs corresponding to this training sample. A threshold is found from the features corresponding to this training sample. Until this moment, no damage classes have been involved in building the baseline space or determining the threshold, and that is what is meant by referring the use only of healthy category signals.

Any new signal (healthy or faulty) is projected onto this baseline space and from the projection, features are obtained and the Mahalanobis distance is measured between the feature (i.e the feature vector) and the baseline category. This Mahalanobis distance is then compared to the predetermined threshold, and a decision about assigning the new signal to either the baseline or non-baseline category is made. Thus, this part of the methodology can be considered as unsupervised learning because no information from the fault category is involved in the assignment decision.

- To the best of the author's knowledge none of the previous SSA- based methods used only healthy signals to build the baseline space by subjecting them to only the decomposition stage.

- Besides the simplicity and efficiency of the methodology, it is able to distinguish not only between the baseline and non-baseline signal categories but also among the different non-baseline signal categories From the results obtained for this methodology, it is observed that the Mahalanobis distances corresponding to the features from the faulty category are deviate significantly from those corresponding to the baseline category. Moreover, the Mahalanobis distance levels (ranges) change with the change of fault severities and fault locations. This means that the methodology is able to detect faults corresponding to different locations (such as in the inner and outer races of a bearing) and of different severities, from small to large.
- When a higher phase / level of diagnosis is required, a number of feature matrices, which belong to different fault categories, are built. More specifically when a category of a specific feature set (i.e to which the fault location category or severity belongs) needs to be known, a number of signal categories from the training samples are built. In this case this process can be considered to be supervised identification.
- Development of a new and simple data-analysis based signal pretreatment. This new signal pretreatment is suggested to make a nonstationary bearing vibration signal more amenable to analysis by the simplest form of autoregression (i.e by use of the stationary autoregressive (AR) model). Most AR – based methodologies for fault diagnosis in REBs tend to complicate the structure of the AR model to

represent the complex bearing vibration signal. Using such complex AR models requires proposing the shape of evolution of the AR model coefficients. Furthermore, an initial set of AR model coefficients is to be proposed as well. In the current research, the new suggested signal pretreatment facilitates the efficient use of a stationary AR model, which does not require the same assumptions as for nonstationary AR models. Little attention has been given to simplifying the signal structure before subjecting them to stationary autoregressive modelling. Thus one of the contributions offered here is to enrich the research in this direction. The suggested signal pretreatment combines a SSA-based denoising technique, together with a stationarisation technique by means of differencing. To the best of the author's knowledge, both the proposed denoising and stationarising techniques are used here for the first time (i.e in the way that they are used in the current thesis) for fault diagnosis of REBs.

- Development of a methodology for complete fault diagnosis in REBs. The methodology combines the above signal pretreatment, a linear time invariant autoregressive (LTIVAR) model and a nearest neighbour classifier (i.e chapter 6 & chapter 7).
- Both methodologies, which are developed in this thesis, are compared with some recent other methodologies. The similarity of data sets in all the compared methodologies is taken in consideration. Although all other methodologies show a very good performance, the present methodology

has also very good data and performs them sometimes. So one of the contributions of the present research is the very good performance of the methodology in detecting, localising and estimating the fault severity.

1.7 Thesis organisation

This thesis begins with introducing the definition of the basics of condition monitoring. Then, the importance and benefits of condition monitoring are highlighted. Furthermore, the common strategies of bearing condition monitoring are briefly introduced. The motivation and contribution of present research work are also presented in this chapter 1.

Chapter 2 introduces the fundamentals of rolling element bearings and main defects. The chapter starts with presenting the most common types of rolling element bearings with brief description and illustrations. Then a description of the parts of a typical bearing is also given. The modes and reasons of bearing failure are also discussed in this chapter.

Chapter 3 shades the light on the main techniques for vibration based condition monitoring of rolling element bearings. The methods are divided into two main types namely, model –based techniques and data –driven techniques. Some works which using these techniques are discussed.

Chapter 4 introduces the fundamentals of the SSA. Then, it presents the REB fault diagnosis methodology which is based on the SSA method. The two main stages of the method (i.e the building of the baseline space and the fault diagnosis) are explained in

details. In the first stage, the chapter explains the building of the baseline space from the healthy bearing category and using the decomposition stage of the SSA only. It also explains the projection process of other signals and the feature extraction from these projections. In the second stage, the chapter explains the processes of the three phases of the fault diagnosis which are fault detection, fault type identification and fault severity estimation.

Chapter 5 demonstrates some results obtained using the method presented in chapter 4. These results include the correct classification rates of the fault diagnosis (for the three phases, namely fault detection, fault type identification and fault severity estimation). Data sets obtained from three different test rigs are used for the purpose of the methodology validation. A comparison with other fault diagnoses is also presented.

Chapter 6 presents the REB fault diagnosis methodology which uses LTIVAR modelling. This methodology is divided into two main parts: signal pre-treatment and signal diagnosis. In the first part, steps of the advanced signal pre-treatment are explained. In the second part, the steps of building feature vectors and diagnosis process is explained.

Chapter 7 introduces some results obtained from the methodology, which is presented in chapter 6, using three different studies are explained in chapter 6. Demonstrative figures are presented to illustrate the steps of the signal pre-treatment using one of the case studies. In addition, the correct classification rates are also introduced for all the case studies.

Chapter 8 presents some important conclusions drawn from the work and provides some recommendations for future research.

1.8 Summary

This chapter shows the importance of condition monitoring and focuses specifically on condition monitoring for REBs. Then the chapter discusses the bearing CM strategies and more illustration is given to a strategy which is based on the analysis of bearing vibration. The objectives of the current thesis is also introduced, and discussed. The motivation and contributions made in this thesis are also explained in this chapter.

Chapter 2

Fundamentals of rolling element bearings and main defects

2.1 Background

Rolling elements bearings (REBs) are one of the mechanical components which tend to fail due to some operational conditions. It is described in the literature that bearing failures are behind more than half of the failures in motors [27]. In this chapter, the common types of REBs and the parts of a typical REB will be presented. Then the reasons and modes of failures in these REBs are also explained.

2.2 Types of rolling element bearings

There are many different designs of rolling element bearings. They can be classified based on the shape of the rolling elements or based on the way in which the load is supported. According to the shape of the rolling elements, the rolling element bearings can be divided into ball bearing and roller bearing. And each type of these bearing can further be divided into more subgroups according to the direction of load and other design requirements.

2.2.1 Ball bearings

These types of bearings have rolling elements of a ball shape. They are mainly divided, according to the direction of applied load, into: radial, angular contact and thrust bearings.

Radial ball bearing

They can be in different configurations such as single row deep groove and double row deep groove bearings to meet the requirements of different applications. Figure 2.1 shows a typical single and double row deep groove ball bearings. A single row deep groove is the most popular rolling bearing. The curvature radii of the inner and outer raceway grooves are in between 51.5-53 % of the ball diameter. A double row deep groove ball bearing has greater radial load capacity where the load is shared between the rows. Otherwise, both single row and double row deep groove bearings behave in similar way. For further details see [28].

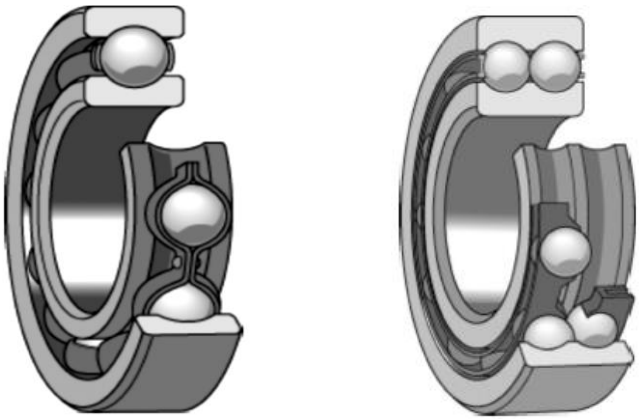


Figure 2.1: A deep grove radial ball bearing. A single row (left) and a double row (right) [28].

Angular contact ball bearings

Figure 2.2 shows a typical single row angular contact ball bearing. Such kinds of ball bearings are designed to support heavy thrust load or combined radial depending on the magnitude of the contact angle. The curvature radii are usually in the range 52-53 % of the ball diameter and the contact angles is no more 40%.

Figure 2.2: An angular-contact ball bearing [28].

Thrust ball bearings

The standard thrust ball bearings have a 90^{o} contact angle as shown in Figure 2.3. Usually, the ball bearings which have contact angle greater than 45^{o} can be classified as thrust bearings. The thrust ball bearings of 90^{o} can only support thrust loads but no radial load. Thrust ball bearings are suitable for operating at high speeds and sometimes they are usually mounted on a spherical seat to achieve some aligning ability.

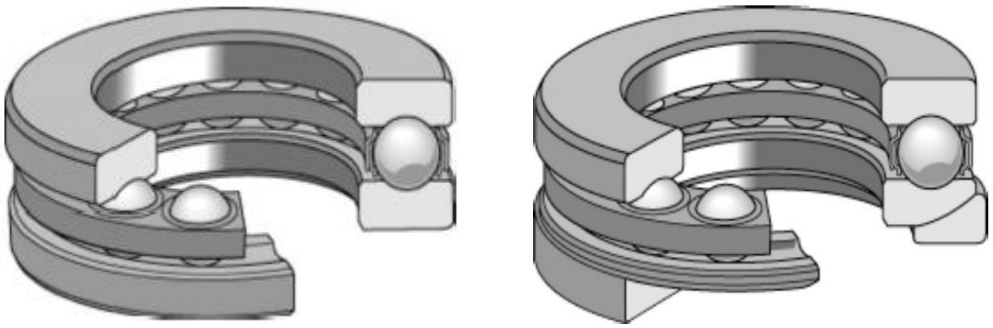


Figure 2.3: A 90^{o} contact-angle thrust ball bearing. The right is with spherical seat [28].

2.2.2 Roller bearings

These types of bearings offer a line contact and usually used for applications which require heavily loadings. However, they cost more in manufacturing when compare to the comparable size ball bearing. They also require more care during mounting than the ball bearing does. They include radial, angular contact and thrust roller bearings.

Radial roller bearings

Radial roller bearings includes cylindrical and needle roller bearings. The former are suitable for high speed operating conditions and high radial load capacity. They usually made of multi rows to support more load rather than to use longer cylindrical rollers see Figure 2.4.

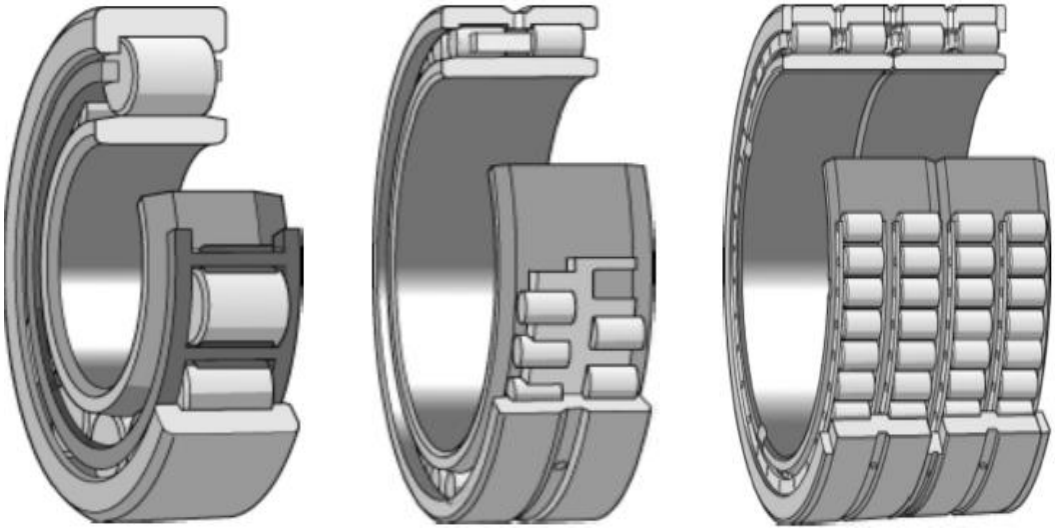


Figure 2.4: A radial cylindrical roller bearing (a single row (left), a double row (middle) and a multi row (right) [28].

The corners of the roller elements are usually crowned to prevent high stresses at the roller edges and to enhance the bearing protection against small misalignments.

Another type of radial roller bearings is called needle roller bearings illustrated in Figure 2.5 in which the length of the rolling element is greater than the roller diameter. They are usually used when there is a need to conserve radial space where they can usually also set directly on shaft. Such bearing are suitable for light loads applications.

Figure 2.5: A needle roller bearing [28].

Tapered roller bearing

Tapered roller bearings illustrated in Figure 2.6 are design to support combined load (radial and thrust loads). However, it is not suitable for high speed applications without considering more attention to lubrication and cooling.

Figure 2.6: A single-row tapered roller bearing showing separable cup and non separable cone, cage and roller [28].

Thrust roller bearings

They can have different configuration based on the type of the rollers. They includes cylindrical and needle roller thrust bearings as shown in Figure 2.7 from [29] .

Figure 2.7: a cylindrical (left) and needle (right) thrust roller bearing [29].

In the next section, more description is given about a typical single row radial bearing.

2.3 Parts of a rolling element bearing

Rolling element bearings are those mechanical parts which used to reduce the friction of rotating parts. Depending on the purpose of use, there are many kinds of bearing such as thrust and radial bearing. A typical radial bearing has four main parts, namely, Inner raceway, rolling elements (i.e balls or rollers), outer race and a cage which separate the bearing rolling elements and prevent friction among them [28]. The Figure2.8 shows a typical single raw, deep groove bearing.

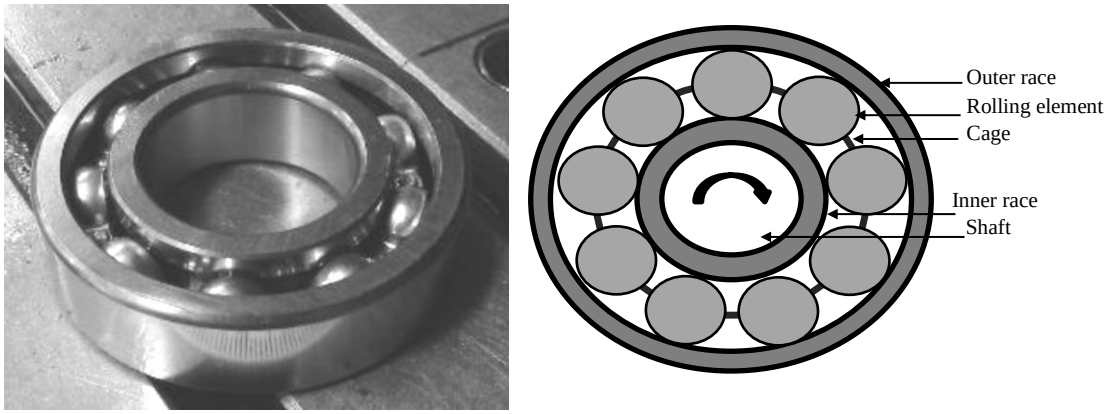


Figure 2.8:- The structure of a typical ball bearing [28].

The inner raceway is usually mounted on the shaft and thus rotates with it. Outer raceway is inserted in a bearing housing and it is usually stationary. The rolling elements can be balls, tapered rollers, cylindrical rollers or some other shapes. They locate between the inner and outer races and they are transferring the load by contact with the raceways and maintaining the motion between the stationary outer race and the rotating raceway (i.e inner raceway). Cage is installed to separate the adjacent rolling

element, provide equal spacing and to provide equal spacing among the rolling elements and to prevent the contact internal strikes due to the collision of them.

2.4 Reasons of bearing failure

Bearing faults occur in different locations and for different reasons. There are several reasons behind the defect occurrences in bearing. These faults prevent the bearings from achieving the designed life span of operating. Less than 1% of bearing achieve their design life span [30]. The popularity of bearing failure defects reasons is shown in Figure 2.9 below.

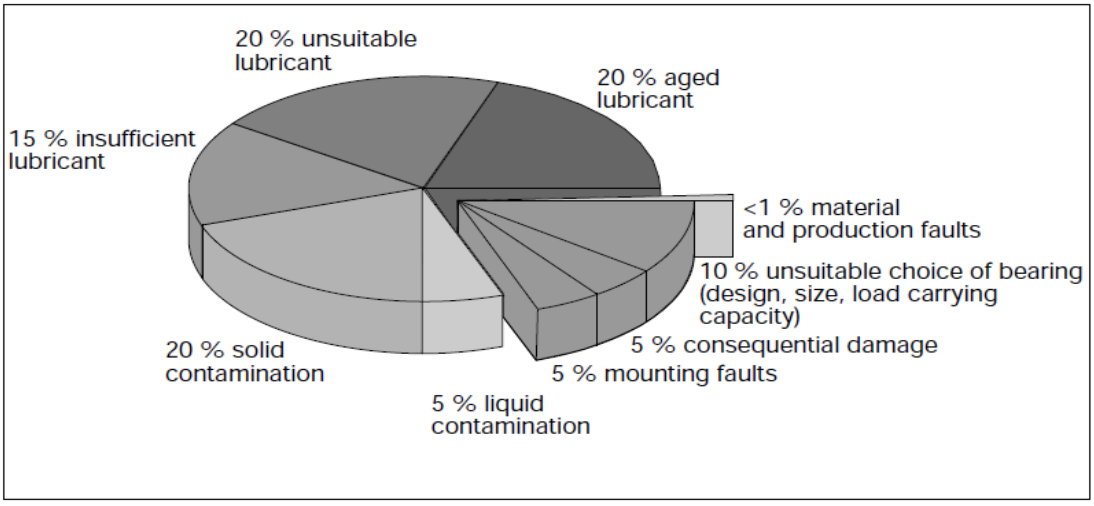


Figure 2.9: Causes of failures in rolling bearings [30].

2.5 Modes of bearing failure

Generally, bearing defects can be classified into localised and distributed defects. Cracks, pits and spalls are typical examples of localised bearing defects which are usually caused by plastic deformation, brinelling, and material fatigue. For the distributed bearing defects, they include waviness, surface roughness, off-size rolling

elements and misaligned races. They usually occur due to wear, improper mounting, design and manufacturing errors and corrosion [31, 32].

Although both of these two classes of bearing defects can cause machinery breakdown, localised bearing defects are more important from the point of bearing healthy condition monitoring. This is because the localised bearing defects are the dominant style of bearings in real industrial applications while many distributed bearing defects are originally started as localised defects [33]. Thus, the work in this research focuses on detecting bearing localised bearing defects.

Experimentally, there are two approaches to generate faults in bearing for the purpose of investigating the vibration signature of bearing. One is the running of bearing until failure under sever conditions such as overloading, overspeeding or lack of sufficient lubricants [34-36]. The other approach is creating defects in bearing using spark erosion, acid etching, scratching or mechanical indentation [37-39]. The presence of these defects in bearing components can considerably shorten the typical designed operating life of a bearing. Some of the common modes of defects are explained below.

2.5.1 Wear

Bearing wear is one of the most common failure modes and can be a result of a fatigue in the bearing material, lubricant contamination and existence of foreign particles such as dust. This kind of failure is started as a localised defect, at early stages, and then it can be distributed for a larger area of the internal bearing surface see Figure 2.10 [40].

Figure 2.10: Wear in roller elements [40]

2.5.2 Plastic Deformation

A plastic deformation is usually a localised damage at contact surface caused by a high amplitude impact, a shock while the bearing a stationary.

2.5.3 Corrosion

Bearing corrosion damage occurs due to a corrosive operating environment such as when a water entering to bearing. The rust particles in a corroded bearing are worn off during the bearing rotation and wear may generate [41].

Figure 2.11: Corrosion on a tapered roller race [40].

2.5.4 Brinelling

It can be defined as the permanent indentation on the surface of the bearing raceways. There are several causes of the occurrences of such indentation such as improper hammering during bearing installation and accidental drops. See Figure 2.12 [42].

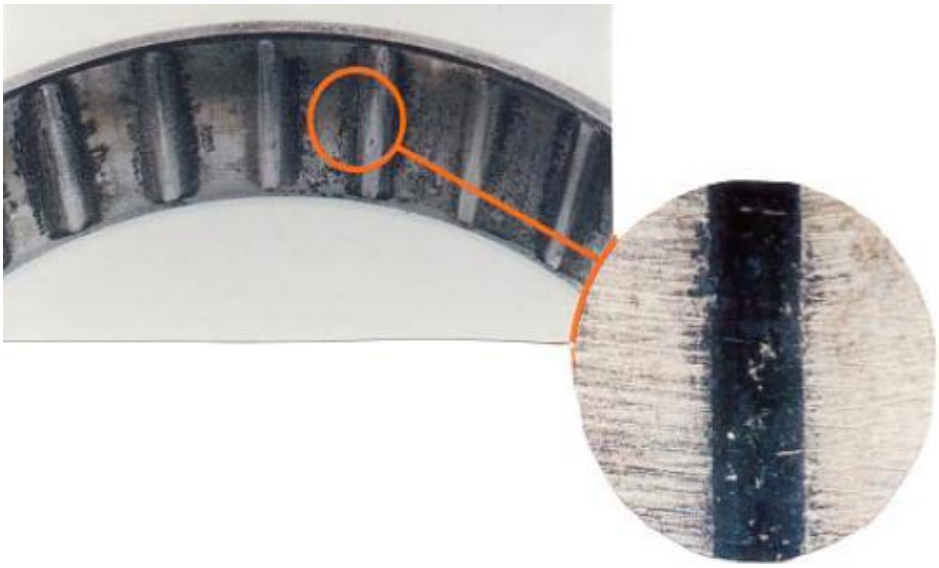


Figure 2.12: Brinelling failure [42].

2.5.5 Improper mounting

The common problem in the bearing mounting process is the excessive preloading because of the improper tolerances. Such problem can be recognised by the appearance of track formation in the bearing raceways. Another types of problem includes the improper hammering and pulling of a bearing during the installation and dismounting processes. More details about these faults is given in [43].

2.6 Summary

The chapter gives a short background about the importance of REBs (section 2.1). It also explored the main categories of the REBs (i.e section 2.2) and illustrated the main components of a typical bearing structure (section 2.3). Sections 2.4 and 2.5 explored the reasons and modes of bearing failures respectively. Compared to distributed defects, localised bearing defects are more important from the point of bearing healthy condition monitoring. This is because the localised bearing defects are the dominant style of bearings in real industrial applications while many distributed bearing defects are originally started as localised defects.

Chapter 3

Literature review

3.1 Introduction

This chapter presents the general procedure for vibration-based in fault diagnosis REBs. Then it presents a short overview on the bearing fault diagnosis vibration -based techniques. These techniques are based on the concept that when a defect occurs in the bearing, a change occurs in the vibration signal obtained from that bearing. In this chapter, the vibration –based techniques is classified into two main groups, depending on how the vibration data are obtained, to model-based techniques and data-driven techniques. The concept of each group is presented briefly and a presentation to some of studies of each group is introduced. More emphasis is given to the data driven techniques as they are more popular in the REB fault diagnosis area.

3.2 A general vibration –based machinery health monitoring procedure

As was mentioned in section 1.4.5, vibration –based techniques are one of the most popular strategies for monitoring the bearing health state. In this strategy, the healthy state can be assessed by analysing and monitoring the changes of the vibration signals obtained from the machine (i.e usually from bearing housing). The typical vibration-based fault diagnosis strategy usually includes three main steps as below:

Data acquisition is the process of collecting useful information (i.e signal) about the bearing. In this step, the number, type, locations and sensitivity of the sensors to collect the signal are determined. This step may be sometimes affected by the economic factor as more sensors and data acquisition cards of high technical specifications cost more money.

Signal analysis is a crucial step and it receives a considerable attention by the researchers. Usually in this step includes subjecting the signal to a certain pretreatment and extracting some compact information ''features'' which are used for monitoring the health statues of the REBs or even for distinguishing among different health conditions of a REB. An enormous number of techniques are developed by the researchers for the purpose of bearing signal analysing and feature extraction. These techniques can be classified according to the signal domain, such as time, frequency and time-frequency domain, where the analysis is carried out. For a successful fault diagnosis process, these features should be high sensitive to defects and convenient to be measured technically and economically.

Diagnosis is the step in which the bearing signal category is assigned to one of the possible bearing health conditions such as healthy and faulty. Depending on the level of diagnosis, classification can be only for two groups such as healthy and faulty or can be extended to classify the bearing vibration signal to one of the signal categories under study. Generally, a model of classification can be learnt using features corresponding to data training sample. Then a new feature/ feature set is / are assigned to one of the classes made by the model. The classification can be carried out using different

techniques such as artificial neural network based techniques (ANN) [44, 45], support vector machines based (SVM) [46, 47] and nearest neighbour based techniques (NN)[48, 49].

3.3 Literature overview of vibration-based REB condition monitoring

Over the last few decades, an enormous number of studies was presented by the researchers in the field of REBs fault diagnosis. These studies aim mainly to develop new methodologies or improve / extend the available methodologies of fault diagnosis. In this section, a short overview for the vibration –based bearing fault diagnosis is presented. This short overview does not mainly compare among the fault diagnosis techniques because each technique has its limitations and advantageous and limitations depending on the case of study. It aims to highlight some of the main techniques and then more discussion is provided to the techniques which are related to the work presented in this thesis. Further details about the vibration-based techniques of REB fault diagnosis can be in [4, 5, 33, 50-52]. The REB fault diagnosis techniques can be classified into two main groups: model- based techniques and data driven techniques depending on the type of the source where the data are obtained. In the former, the techniques use mathematical models to derive a knowledge which can be used to determine the bearing's condition. In the latter, the techniques use collected experimental data to learn about the system and infer it is current state of health.

3.3.1 Model –based techniques

These techniques are based on a mathematical model describing the vibration response of a REB and use the modelled responses for assessment of the system health (state). The complexity of a model structure is increased, but more comprehensive matching is obtained, when more effects such as misalignment and radial clearance are considered. Models can be built for a bearing working under normal condition (i.e healthy) and then experimental data are used to assess the similarity with those obtained from the model and eventually to evaluate the state of the bearing. Models can also be built for different conditions of bearing such as healthy, a fault on a raceway or on a rolling element and then the experimental vibration signal can be compared to the simulated one obtained from these models to evaluate the category of the signal using some intelligent algorithms. One of the first interesting studies on dynamic modelling of a REB, with a single point defect on the inner raceway, was presented in [53]. The model can consider several parameters such as load distribution and shaft rotational speed. The same authors of the work above have extended the model in [54] considering the case of multiple defects. another attempt for dynamic modelling of bearing can be found which model the defect using the forces introduced by the deformation of rolling element bearing of modelling is presented in [55]. In [56] the authors model a localised defect as a smooth pit with a curved shape. Another study models a single point defect bearing in the case of high speed [57].

The study presented in [58] discuss a model considering radial clearance effect. A single defect was also modelled as a sphere and the contact between the ball and

bearing races is modelled as nonlinear spring [59]. More details about these studies can be found in [52, 60] which discuss more studies of dynamic modelling of REB.

One of the advantages of model-based techniques is that they are used generally for system model updating. For model updating, experimental data are used to enhance the accuracy of the model in representing the vibration behaviour of the system. There are advantages to the use of the model-based techniques. However, complex systems are difficult to model accurately, and model updating methods can be affected by the variability of the system experimental measurements [61].

The next section explains the non-model based techniques (e.g data driven techniques) which rely completely on the vibration data measured from the REBs.

3.3.2 Data-driven techniques

These techniques rely totally on the analysis of vibration data which are obtained experimentally from the machine under consideration and use them to infer conclusions on the condition and the health state of the bearing. One of the first data driven techniques are those that deal with the time waveform directly and use rather simple and straightforward calculations. This includes for example extracting some statistical features and assessing their values at different bearing conditions. Kurtosis and crest factor are one of the most common features used in the fault diagnosis problems. Both of these factors characterise the 'peakiness' of a signal. However, according to the literature, kurtosis is more sensitive than the crest factor in indicating the defects [62-64]. Each of the crest factor and the kurtosis has typically a certain value at normal bearing conditions but that value changes when the bearing condition changes. For

example and as was mentioned in[1] , for a vibration signal obtained from an intact bearing, the crest factor and the kurtosis values are 3.5 and 3 respectively. While for a bearing with a defect on its inner race they are 6.35 for a crest factor and 4.75 for a kurtosis .Some other studies basically deal with the frequency domain of the vibration signal (i.e the values of a signal verses frequencies) where the repetitive signal component is shown as a peak at the frequency of repetition. This may include finding the frequency spectrum of the vibration signals and a certain frequency band of interest which can be isolated for further analysis [33]. The fast Fourier transform (FFT) is one of the most popular techniques [65] used for obtaining the frequency spectrum of the signal. There are other techniques which use other spectral expressions such as cepstrum [66] and envelope analysis [67]. Other studies present the analysis of vibration signal in time-frequency domain instantaneously using some particular techniques such as such as short time Fourier transform (STFT) [68], wavelet transform (WT) [39, 69, 70].

In comparison to model-based techniques, there can be no fear of losing information due to lack of model accuracy, as no analytical or numerical model is made of the REBs. Usually these techniques use measurements, acquired under normal conditions for the machine, and then they are used to build the baseline state, against which the measurements from the machine in service are compared. Any deviation of the new observations from the baseline state may indicate the presence of a defect. Compared to model based-techniques, data driven-techniques are considered to be easier to manage. This is because it is not trivial to construct or derive models which can properly

describe machine behaviour, particularly in the presence of several components and nonlinearities [71].

Another group of interesting studies on fault diagnosis in REBs are those that apply the concept of time series analysis. In the next section, an overview of the some of the most important time series analysis techniques will be presented.

3.4 Time series analysis techniques

A time series can be defined as a sequence of measurements (i.e data points) of a variable, .e.g. acceleration or velocity, collected over time. The concept of time series analysis is very popular in climate and financial research fields[72] and many techniques were developed for purposes of analysis or predicting of the future values of a time series. These techniques take into account some aspects of the internal structure in the data. Some of these techniques offer representing a time series by some parametric models such as Autoregressive modelling (AR) [73]. Others, such as singular spectrum analysis (SSA) [74], are used to decomposed a time series into a number of independent components that can have some meaningful interpretations such as trend and periodic components.

The concept of time series analysis has also a wide attention in engineering areas such as structural health monitoring (SHM) and machinery diagnosis. This study contributes to the knowledge of fault diagnosis in REBs by using SSA in two new different ways and for different purposes. 1) A new SSA-based method is developed for a complete fault diagnosis in REBs and 2) SSA combined with differencing is used in developing an advanced signal pretreatment that enhances the goodness of fit of linear AR model.

The next sections introduce the basics of the two time series analysis techniques used in this thesis, namely, singular spectrum analysis (SSA) and autoregression (AR).

3.4.1 Singular spectrum analysis

The SSA is a data analysis method which is popularly used in the climate data analysis [75, 76], biomedical signal analysis [77, 78] and tool wear health monitoring [79, 80]. However, few researches are done on the use of SSA for purposes of fault diagnosis in rolling element bearing. Bubathai [81] has published the first study which uses SSA for classifying signals as healthy or faulty (with a fault on the inner raceway) for detection purposes only. In this study both the vibration acceleration signals (i.e those corresponding to healthy and faulty bearings) were subjected to SSA and the original signals are then decomposed into two time domain main components, namely, the trend and the residuals. Then, only the trend component was considered for further analysis. A number of statistical features such the peak value and the standard deviation were obtained from the trend. These features were used to form the feature vectors which were eventually used as input for a neural network classifier. In [82] the SSA was also used as a multi-decompositional analysis technique. In this study the number of singular values, which preserves a specific predetermined variance percentage, is used as an indicator for a fault presence.

Two different feature sets were obtained from the application of SSA in [83] and they are used as feature vectors. The first feature vector was made of the singular values and the second feature vector was made from the energy of the first time domain principal components. The accuracy of the bearing condition classification was investigated by

using those feature vectors as input to the back propagation neural network (BPNN) classifier.

In the study presented in this thesis, a new methodology using only the decomposition stage in SSA is proposed[1]. This makes the suggested method much simpler as compared to the previously developed ones.

In this thesis, the SSA (more specifically only the decomposition stage) is used to build a baseline space corresponding to the healthy condition. Afterward the lagged versions of the new signals are projected onto the baseline space. To the best of the author's knowledge none of the previous SSA-based methods used only healthy signals to build the baseline space by subjecting them only to a decomposition stage. This explains the simplicity of the current methodology. The projections of the other signals onto the baseline space are then used to make the FVs which are then used in the diagnosis stage. Besides the simplicity and efficiency of the method, it is able to distinguish not only between the baseline and non-baseline signal categories but also among the different non-baseline signal categories, which include different fault locations and different fault severities categories.

The method was validated using data sets obtained from three different bearing test rigs. The results prove the very good performance of the methodology in terms of correct classification rates.

[1] Part of the work is submitted to one of the standard journals and it is under review

3.4.2 Autoregressive modelling (AR)

Another direction of studies, which has been attracted the interest of many researchers, is the use of the parametric modelling concept for fault diagnosis in rolling element bearings. According to this concept, the bearing vibration signals are considered as time series and represented by a suitable model with few parameters [84]. There are three main types of parametric modelling and they are: Autoregressive (AR) model, Moving Average (MA) model and Autoregressive Moving Average (ARMA) model [85]. Among these models, the AR models receive more attention by the researchers in developing techniques for the REBs fault diagnosis. One of the reasons for the popularity of AR use in the fault diagnosis of REBs is that the sensitivity of the AR parameters to the variation of the system where the AR model is used to describe its response[86]. Generally, the AR model used to represent the bearing vibration signals and the parameters (prediction error and the AR coefficients) are used for fault detection and identification. Faults can be also diagnosed when by constructing a model for each known bearing condition. Then for any new bearing vibration observation, it (i.e the new observation) can be assigned to the model which represents it more precisely.

There are several challenges accompanied with the use of the AR model to represent the bearing vibration signals. One of the most important challenges is the complexity of the bearing vibration signal. Bearing vibration signals are almost always non-stationary because they are inherently dynamic. The definition of stationarity of signal is the condition of a signal where all the first four statistical moments are constant with time. However, the definition is used for the so-called strict stationarity which occur rarely.

Instead, the term of 'weak stationarity' is used for signals which the first and second statistical moments (i.e mean and standard deviation). The non-stationarities can be found in the benign bearings and might come from the clearance between the bearing components and sliding of rolling element, just to mention a few [87]. Non-stationarities can also come, for example, from the impact between the damage and non-damaged parts of the bearing [88]. Therefore, compared with gear system analysis, bearing fault diagnosis is an even more challenging task in condition monitoring, especially when the machine is operating in a noisy environment.

Some researchers suggest the use of time-varying autoregressive (TVAR) models which take into account the presence of non-stationarities in the signal. In [89, 90], the investigation for three different algorithms for the time varying autoregressive model coefficients estimation was introduced. These algorithms are namely, Kalman, extended Kalman and modified extended Kalman filter. A parametric time-frequency using a time varying autoregressive model is introduced in [91]. In that study, the time-frequency array is subjected to the singular value decomposition (SVD) to form features vectors which are used for the classification based on the radial basis (RB) neural network. In all the above studies, the TVAR model coefficients were assumed to change over time in a pre-proposed way to facilitate the representation of the non-stationary bearing vibration signal. However, there are some challenges implicit in obtaining these coefficients. One of these challenges is the setting of a proper assumption for the way that the model coefficients evolve with time. Another challenge is the need to assume a proper initial model coefficient to start the evolution over the time. Thus, the entire process of building an accurate model might be jeopardised by an improper assumption

for the initial coefficient set and/or an inappropriate assumption for the evolution of the coefficients.

Other forms of autoregressive modelling, which assuming that data points of a signal are related in nonlinear way, is also investigated in [92, 93]. Such kinds of models are also challengeable in terms of the requirement of proper assumption of the nonlinear relationship form.

Another form of autoregressive model is proposing that the model coefficients have a periodic time –varying behaviour [94]. In this kind of autoregressive model, it is assumed as that bearing vibration signal has cyclo-stationary behaviour.

In terms of complexity of the forms of autoregressive model above, linear autoregressive models with time invariant coefficients are mainly simpler. as compared to other forms of representing signals. Nevertheless, these models are suitable for stationary signals, which makes their use in the fault detection in rolling element bearings is limited [95].

Some studies are presented to investigate the application of some signal pre-treatment such as empirical mode decomposition (EMD) [86], and amplitude demodulation (AD) [93] to enable the use of the LTIVAR model in the fault detection of rolling element bearings.

One of the main contributions in the study presented in this thesis is the use of a simple LTIVAR model, in combination with the suggested signal pretreatment, in a precise and successful fault diagnosis. This is achieved by the development of a new signal pre-

treatment. This signal pre-treatment aims covering the limitation associated to the use of such stationary model in modelling the bearing vibration signal, which is almost always non stationary. By using the LTIVAR, neither the assumption of an initial set of model coefficients nor the assumption of the shape of coefficient evolution over time are required. The pretreamtent includes subjecting the signals to SSA –based de-nosing step and stationarising the non stationary signals by stabilising their means over the time using a simple procedure called as differencing. Eventually the goodness of fit of the LTIVAR model is enhanced and a very good fault diagnosis results were obtained.

3.5 Summary

The chapter explains the general vibration –based machinery health monitoring procedure. Then it discusses model –based techniques and data –driven techniques used in vibration-based bearing condition monitoring and discusses their advantages and disadvantages. More focus is also given to the time series analysis techniques, particularly the SSA-based techniques and the AR-based techniques. A critical review was conducted for each two approaches, which are developed in the present thesis, with other SSA-based technique and AR-based techniques in the literature. The contributions behind using these two approaches are also explained.

Chapter 4

A new SSA based methodology for fault diagnosis in REBs

4.1 Introduction

In this chapter a new methodology for the fault diagnosis in rolling element bearing is suggested. The methodology is based on the use of singular spectrum analysis (more specifically the decomposition stage only). The methodology assumes the building of a reference space from the healthy bearing signals and then to project any new signal in that baseline space. From these projections feature vectors are made and used in different phases of fault diagnosis. The fault diagnosis's phases investigated in this method are: fault detection phase where only the presence of a fault is detected, fault type identification where the type (i.e location) of the fault is detected and fault size estimation where the different faults severity is detected. The similarity of the new signal to the reference category is assessed by measuring the Mahalanobis distance. The steps of the methodology, for each fault diagnosis phase, are presented in this chapter.

The results regarding the performance of the methodology in terms of the correct fault diagnosis rates are presented as well using more than test rig data.

4.2 The fundamentals of SSA

SSA is a statistical procedure which has been used extensively for climate analysis [75, 76], biomedical analysis [77, 78] and meteorology analysis, but has not yet gained popularity for machinery analysis. It is simply principal components analysis applied to the lagged components of a time series.

SSA is used to decompose the original signal into a number of independent components; the principal components (PCs). The initial time series can be then reconstructed by using a number of PCs.

The primary aim of SSA is to uncover the trend in a signal, particularly its oscillatory patterns. SSA can, however, also be used as a noise-cleaning procedure; it is known to clean structure-less noise by transforming it into low singular value components [26, 74]. The SSA has two main stages: decomposition and reconstruction. The fundamental of each stage will be illustrated in this section. Further details about the method is given in [96].

4.2.1 Decomposition stage

In the decomposition stage, a sub-signal **x** of length n , $x(1),x(2),\ldots,x(n),$is mapped onto a window of length (L) to form the so-called trajectory matrix **X** (LxK) where $K=n-L+1$ (see Eq. (4.1)).

$$\mathbf{X} = \begin{bmatrix} x(1) & x(2) & x(3) & \dots & x(K) \\ x(2) & x(3) & x(4) & \dots & x(K+1) \\ x(3) & x(4) & x(5) & \dots & x(K+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(L) & x(L+1) & x(L+2) & \dots & x(n) \end{bmatrix} \tag{4.1}$$

Then, the covariance matrix ($\mathbf{C_X}$) of $\mathbf{X}$ is calculated as in Eq.(4.2).

$$\mathbf{C_X} = \frac{\mathbf{X}\mathbf{X}'}{L} \tag{4.2}$$

The $\mathbf{C_X}$ has a dimension (*KxK*) and it defines the covariance between signal realisations. The $\mathbf{C_X}$ is then subjected to Eigen decomposition by singular value decomposition (SVD) to obtain L eigenvectors ($\mathbf{U}_i, i = 1,2..,L$) and L eigenvalues ($\lambda_i, i = 1,2..,L$) by solving the following expression (see Eq.(4.3))

$$\mathbf{C_X}\mathbf{U}_i = \lambda_i \mathbf{U}_i \tag{4.3}$$

Each λ_i represents the partial variance of the original time series in the direction of the $\mathbf{U_i}$. Projecting the trajectory matrix onto each eigenvector provides the corresponding principal components ($\mathbf{PC_i}$.):

$$PC_i(m) = \sum_{j=1}^{L} X^T(m+j-1) * U_i(m) \tag{4.4}$$

where

i=1, 2….L;

m=1, 2…n;

j=1, 2…L.

Then, L elementary matrices ($\mathbf{El}_i = \mathbf{U}_i\mathbf{PC}_i'$ *where* $i = 1,2,\dots L$ *and the prime means transpose*) can be created by the projection of the PCs on the Eigenvectors **U**.

The contribution of these elementary matrix norms to the original trajectory matrix norm follows the trend of the singular values, which is the first matrices have the highest contribution while the last ones have the lowest.

4.2.2 Reconstruction stage

As was mentioned above, the signals can be reconstructed by a linear combination of all or a number of the PCs. Different criteria can be used to select the number of PCs [82]. The reconstruction process is done by the diagonal averaging technique, which is described below, to produce the reconstructed signal (**xr**)[97]:

$$xr(m) = \frac{1}{N_m}\sum_{i\in w}\sum_{j=Lm}^{Um} PC_i\,(m-j+1) * U_i(m) \quad ,m = 1,2,,,,n-1 \qquad (4.5)$$

The Normalisation factor (N_m) and the lower (Lm) and upper (Um) bounds of sums differ for the edges and the centre of the signal. They are defined as follows:

$$\left(\frac{1}{N_m}, Lm, Um\right) = \begin{cases} \left(\frac{1}{m}, 1, m\right), & for \quad 1 \le m \le L-1 \\ \left(\frac{1}{L}, 1, K\right), & for \quad L \le m \le K \\ \left(\frac{1}{n-m+1}, m-n+L, L\right), & for \quad K+1 \le m \le n \end{cases} \qquad (4.6)$$

The *w* refers to the number of the PCs used in the reconstruction of the signal.

4.3 The SSA- based methodology

It was mentioned in section 1.6 in the present thesis that the SSA is used in two different ways and for different purposes and therefore two SSA-based methods results. In this chapter, the first method is presented where the SSA, more specifically only the decomposition stage, is used for developing a methodology for fault diagnosis in REBs. The methodology presented in this chapter includes the use of the decomposition stage only to build a baseline space and project other signal on it. The projections on to the baseline space are used to form the feature vectors (FV) which are used in the fault diagnosis process. Three fault diagnosis phases are investigated in this chapter; namely, fault detection, fault type identification and fault severity estimation. In the fault detection phase, FVs are classified into baseline and non-baseline categories. In the fault type identification, the FVs are assigned to one of fault location categories such inner race fault (IRF) and outer race fault (ORF). In the fault severity estimation, the method is use for the detection of different fault severity.

The method has two main steps: building baseline space using a healthy signal only and fault diagnosis.

4.3.1 Building baseline space

As was introduced in the last section, the building of the baseline space is based on the decomposition stage of the SSA and using the signals obtained from healthy bearing

only. First a signal **x**=[*x (1), x (2)… x (n)*] is lagged (i.e embedded) on window (*L*) to obtain the trajectory matrix (**X**) as in Eq. (4.1).

Then, the covariance matrix of **X** is obtained (Eq.(4.2) and subjected to the singular value decomposition (Eq.(4.3) to obtain L eigenvectors $(\mathbf{U}_i, i = 1,2..,L)$ and L eigenvalues $(\lambda_i, i = 1,2..,L)$. Each λ_i represents a partial variance proportion of the original signal in the direction of the corresponding $\mathbf{U_i}$. The $(\lambda_i : i = 1:L)$ and their corresponding (**U*i***: *i=1: L*) are arranged ascendingly. The plots which show eigenvalues $(\lambda_i : i = 1:L)$ versus the principal components (PCs) are usually called scree plots.

When the eigenvalue obtained, all or a number of the first **U***i=1:L* corresponding to the healthy state can be used for building the reference space. .

$\mathbf{R}^{L} = [\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_L]$

There are number of criteria mentioned in [82] which can be used as guidance for the selection of the number of the PCs (i.e consequently means the number of λ and **U**). In this study we investigate the effect of increasing the PC number on the fault diagnosis results. However, for the purposes of visualisation we sometime use the first three PCs. The next section will explain the fault detection phase where the signals are assigned to a healthy or faulty category.

4.3.2 Fault diagnosis methodology : fault detection phase

When a reference space is built, the fault detection phase can then be conducted. It has two main parts: feature extraction and fault detection process

Feature extraction

First, the signals corresponding to healthy bearing are divided into a training and testing samples, whereas all the signals corresponding to the faulty bearing conditions are used as testing sample. For all the signals corresponding to the training or testing, the FVs are obtained in the same way as will be described later in this section. For any signal, the trajectory matrix **X** of a dimension (*KxL*) is obtained and projected (i.e multiplied by) onto the *L* dimensional baseline space to obtain *L* of PCs (see Eq. (4.4)). Supposing a (*Q*) of the first PCs are selected for the extraction of features. Then, the Euclidean norm (f_{ij}) of each of PC, is calculated according to Eq.(4.7)

$$f_{ij} = \sum_{m=1}^{K} \left(PC_{ij}(m)\right)^2 \quad , j = 1,2,3 \dots Q \quad i = 1,2,3 \dots k \tag{4.7}$$

Where

f_{ij} is a norm of the j^{th} **PC** obtained from i^{th} signal.

K is the length of a **PC** and equals *(n-L+1***)**

k is a number of signals corresponding to a baseline condition

$PC_{ij}(m)$ is the m^{th} element of the j^{th} **PC** obtained from i^{th} signal

The Euclidean norms obtained from the *Q* of PCs, which corresponds to i^{th} signal, are used to form a *Q* dimensional feature vector (FV) see Eq.(4.8) below.

$$\mathbf{fv}_i = [f_{i1} \ f_{i2} \ f_{i3} \dots f_{iQ}] \tag{4.8}$$

As was mentioned previously, a special case of 3 dimensional feature vectors is used for the purpose of visualisation (see Figure 4.1).

On the basis of the Eq. (4.7) which illustrated that the Euclidian norm is simply the summation of squared values, all the features have positive values and when they are projected on the feature space they are located in the first quadrat (i.e for the case of a 3-dimensional feature vector).

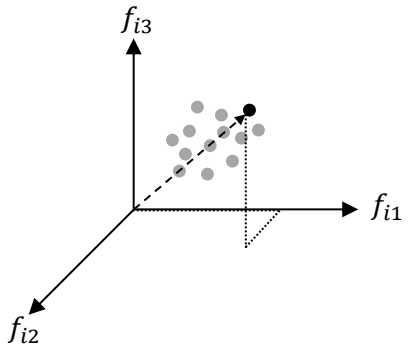


Figure 4.1: A 3D visualisation of a $\mathbf{fv}_i$ in the feature space

The feature vectors obtained from the training sample corresponding to the healthy bearing condition are used to make the baseline feature matrix ($\mathbf{F}_{baseline}$). By arranging the baseline FVs in rows (Eq. (4.9))

$$\mathbf{F}_{baseline} = \begin{bmatrix} f_{11} & f_{12} & \cdot & f_{1Q} \\ f_{21} & f_{22} & \cdot & f_{2Q} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ f_{k1} & f_{k2} & \cdot & f_{kQ} \end{bmatrix} \tag{4.9}$$

Following the steps described above by Eqs. (4.4),(4.7),(4.8), the FV for any other/new signals (i.e testing sample) can be obtained .

Fault detection process

In this section, the details of the fault detection process are introduced. In the current study, the classification of the baseline and non-baseline bearing conditions is made on the basis of a threshold. In this regard the Mahalanobis distance is calculated and used to measure the similarity of a testing FV to the training feature matrix corresponding to the baseline condition. Then, it is compared to a pre-determined threshold corresponding to the baseline condition. Thus all the FVs whose distance is equal to or less than the threshold are classified as healthy and all the ones with distance greater than the threshold are considered as faulty. The fault detection process contains two main parts 1) setting a threshold of the baseline condition and 2) the comparison of the Mahalanobis distance, corresponding to a new testing FV, to the threshold.

Setting a threshold for the baseline condition

When the feature matrix of the baseline condition $\mathbf{F}_{baseline}$ is obtained (i.e as in Eq.(4.9), the Mahalanobis distance of each of the baseline $\mathbf{fv}_i$ to the $\mathbf{F}_{baseline}$ is calculated as shown below (Eq. (4.10))

$$D_i = \sqrt{(\mathbf{fv}_i - \mathbf{E}_{baseline}).\mathbf{S}^{-1}.(\mathbf{fv}_i - \mathbf{E}_{baseline})'} \quad (4.10)$$

where D_i is Mahalanobis distance of a $\mathbf{fv}_i$ to the to the matrix $\mathbf{F}_{baseline}$

$\mathbf{S}^{-1}$ is the inverse of the covariance matrix of $\mathbf{F}_{baseline}$.

The prime ($'$) denotes the transpose of the vector$(\mathbf{fv}_i - \mathbf{E}_{baseline})$.

$\mathbf{E}_{baseline}$ is the mean of $\mathbf{F}_{baseline}$rows ; which is given by Eq.(4.11)

$$\mathbf{E}_{baseline} = \frac{\sum_{i=1}^{k}\begin{pmatrix} f_{i1} & f_{i2} & \cdots & f_{iQ} \end{pmatrix}}{k} \ , k = \ number\ of\ the\ \mathbf{F}_{baseline} rows \quad (4.11)$$

Let **D** be the vector which has the k of D_i corresponding to the baseline condition. A threshold is obtained based on the probability distribution of the elements of the **D**. In fact, most research has assuming a normal distribution as it is the most likely distribution to be used in practice. However, in this work, the author has questioned this assumption and the selection of a suitable probability distribution can be made on the basis of the similarity between the shape of the histogram of the **D** and a selected probability distribution function. In this study, as the values of the **D** elements are positive and skewed, the lognormal distribution is selected for the purpose of determining the threshold of the baseline condition. The similarity of the lognormal distribution to the histogram of the MD is checked visually and it is shown that they are very similar (see Figure 5-B.1 in appendix B page 190). The formula of the lognormal probability distribution function is given by Eq.(4.12) below

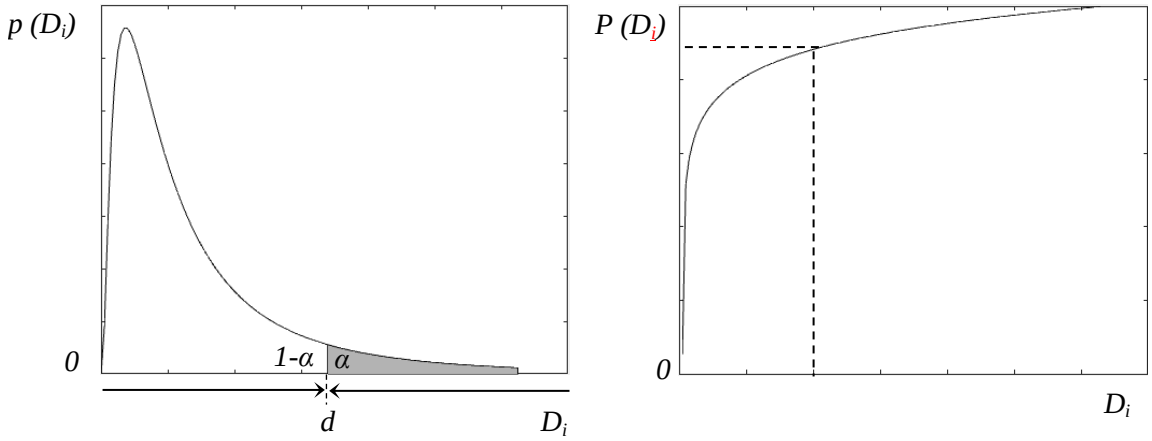
$$p(D_i)|_{\mu,\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{(\ln(D_i)-\mu)^2/(2\sigma^2)} \quad (4.12)$$

where

$p(D_i)$ is the lognormal probability density function value at D_i.

μ, σ are the mean and standard deviation of the $p(D_i)$ $i = 1,2, \ldots . k$

Figure 4.2 shows the standard shape of both the probability density function and the cumulative density function of the lognormal probability distribution.



Figures 4.2: Statistical hypothesis threshold based on lognormal distribution (only single tail)

The figure to the left shows the lognormal probability density function. The x-axis represents a range of values (i.e in our case the elements of **D**) while the y-axis represents the relative likelihood that a random number to take a given number on x-axis (i.e $p(D_i)$). The figure to the right represents the cumulative distribution function (i.e $P(D_i)$) which is related to the area under the curve corresponding to the probability distribution function (i.e $p(D_i)$).

To set up a threshold, a certain value of *d* is determined such that α% area under the curve locates to the right and 1- α % locates to the left. In classification problems, misclassifications are divided into two main groups: 1) False-positives which indicate the presence of a fault when there is not one, and 2) False-negatives which indicate no fault when one is actually present. In the current study, a statistically based threshold is determined such that these false alarms are kept to a minimum during the fault detection process.

The author has checked the misclassification rates at different cumulative probability

values (i.e 0.97 and 0.99). Eventually, d is selected such that the area cumulative probability equals 0.99, which gives minimum misclassification rates.

Figure 4.3 illustrates the possible distribution of FVs and the $Thr_{baseline}$ (i.e dashed line) corresponding to the baseline condition. The x- axis represents the number of FV whereas the y-axis represents the Mahalanobis distance of FVs to the $\mathbf{F}_{\text{baseline}}$ as was described in Eq.(4.10).

Depending on the value of $Thr_{baseline}$ a considerable number of D_i (corresponding to baseline FVs) will be below this threshold and all the rest will be above.

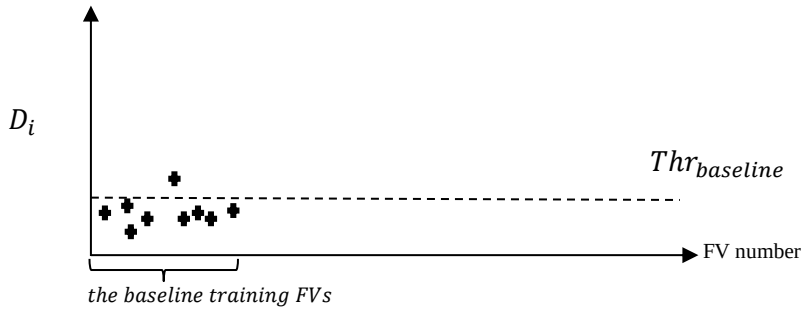


Figure 4.3: Illustration of $Thr_{baseline}$ made from the baseline training FVs

Comparison of a new testing D_i to the threshold.

For any new signal from the testing sample, the FV is obtained as described in the feature extraction section. Then, the Mahalanobis distance of the FV to $\mathbf{F}_{baseline}$ is calculated as described in Eq.(4.10) and finally it is compared to the chosen threshold $Thr_{baseline}$. If the Mahalanobis distance of the new FV is less than or equal to $Thr_{baseline,}$ the FV is assigned to the baseline condition and *vice versa* (see Eq.(4.13)).

$$\left.\begin{array}{ll} D_i > Thr_{baseline} & a\ FV_i\ is\ assigned\ to\ faulty\ category \\ D_i \leq Thr_{baseline} & a\ FV_i\ is\ assigned\ to\ baseline\ category \end{array}\right\} \quad (4.13)$$

The performance of the methodology in terms of the correct assignment of the FVs to their actual category (baseline & non baseline condition) is evaluated on the basis how many FVs is classified to their actual classes.by using the so-called confusion matrix.

The confusion matrix is a square (*K* x *K*) matrix, where *K* is the number of sub-signal categories. The columns represent the predicted classes, while the rows represent the actual classes. Thus, the main diagonal represents the correctly categorised signals, while all the other elements represent the miscategorised signals. The name 'confusion' stems from the fact that this matrix makes it easy to evaluate whether the proposed methodology confuses two or more classes (i.e. mislabelling one category as another). Table 4.1 shows the structure of a confusion matrix:

Actual class/predicted class	***1***	***2***	***....`..***	***K***
1	*C11%*	*C12%*	*……*	*C1Kk%*
2	*C21%*	*C22%*	*……*	*C2Kk%*
.	.	.	.	.
.	.	.	.	.
K	*CKk1%*	*CKk2%*	*……*	*CKkKk%*

`Table 4.1: An example of a confusion matrix

Cij (*i* and *j* =1,2…K_k) refers to the percentage of vectors from class *i* which are classified as class *j*. It is clear that if *i=j*, then *Cij* represents the percentage of correct classification. If *i≠j*, then *Cij* represents the percentage ratio of the misclassification.

A flowchart showing the steps of the methodology- fault detection phase is presented in Figure 4.4. The flowchart has three main blocks given in the dash-line boxes:

1. **Building a baseline space**: Illustrating the building of Q-dimensional baseline space from training sample which corresponding to healthy bearing category.
2. **Extraction of FVs**: Projecting the lagged versions of the new signals from onto the baseline space to obtain the PCs and then creating FVs.
3. **Fault detection phase**: Measuring the Mahalanobis distance (D_i), setting a threshold and comparing the (D_i) of the FVs corresponding to the testing sample with threshold.

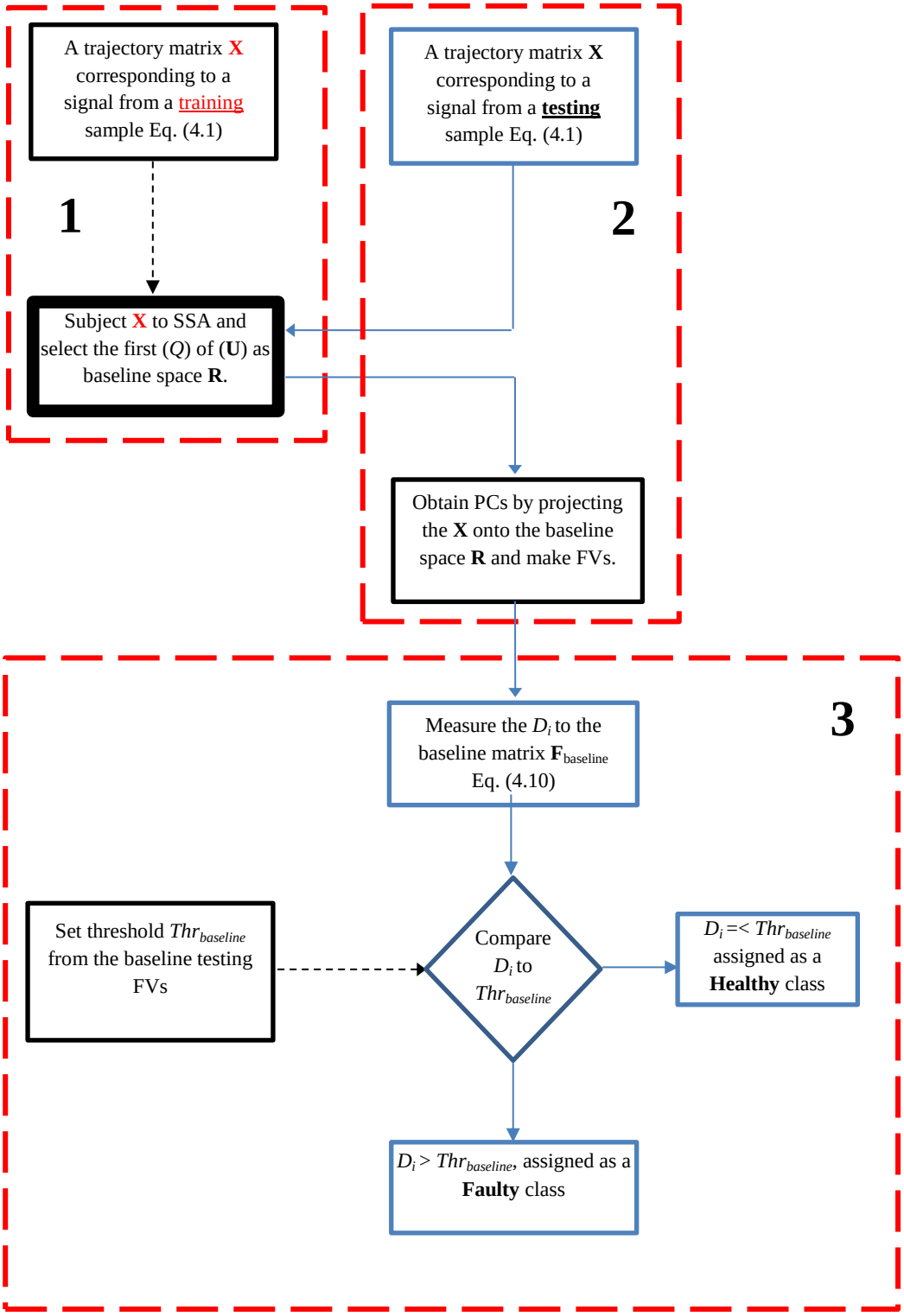


Figure 4.4: A flowchart shows the steps of the fault detection phase in the methodology.

Using all elements versus norms of *Q*- PCs: visual interpretation

As was mentioned previously, the 1st three PCs will be used for the visualisation purposes are used in the fault detection process. In this section, it will be visually illustrated how the use the norms of the PCs provides easier distinguishing between the baseline and non baseline signal categories than the use of all the elements of these PCs. An example from the data sets used in this study is shown to illustrate the improvement of the use of norms of the 1st three PCs as features to distinguish between two different conditions (i.e **H**-Healthy and **F**-Faulty bearing).

The Figure 4.5 shows the 3D visualisation and its 2D projections of using all the elements of the 1st three PCs. In this figure the 1st three PCs corresponding to 30 signals from baseline condition (at 1772 RPM) and another 30 from non-baseline conditions (i.e IRF at 1772 RPM) are plotted. It can be seen that the two conditions are mixed and there is no clear separation between them.

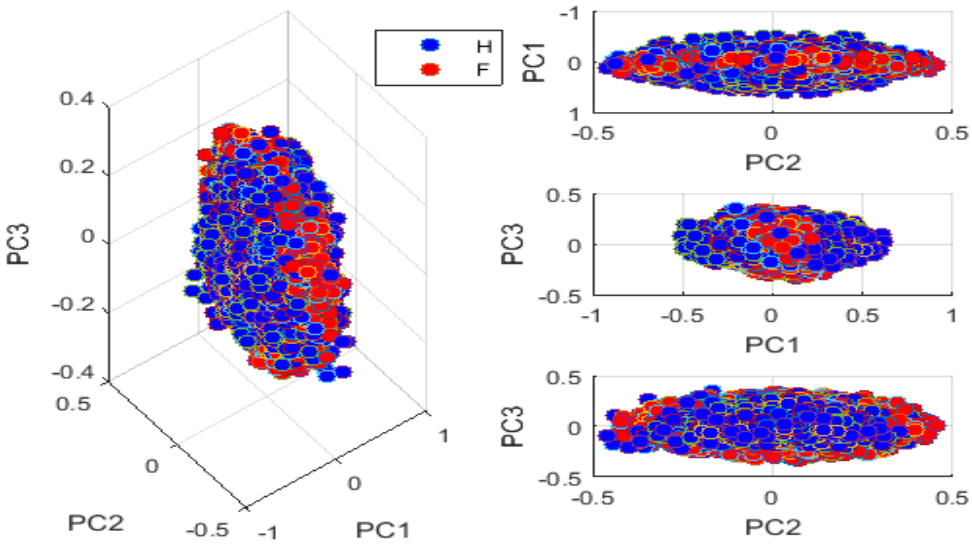


Figure 4.5: 2D&3D visualisation for the case of using all the elements corresponding to the first three PCs. H and IRF – CS1 at 1772 RPM.

Thus, it is suggested in this study that each signal is to be represented by a single feature which is the Euclidean norm of its *Q*- PCs instead of using all the elements of these PCs. This can help by reducing the dimension of the FV used to represent a signal and improves the separation between the two signal categories (i.e baseline and non baseline).

In the Figure 4.6, the 3D visualisation and the three 2D projections of using the norms of the first three PCs are shown. It is clearly seen that the two conditions are well separated.

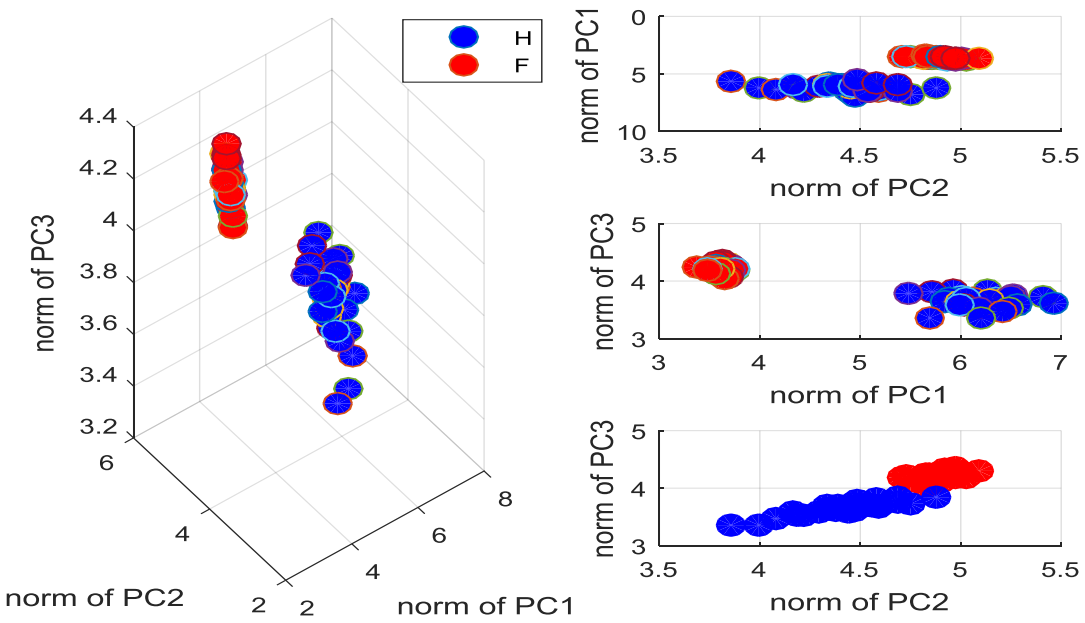


Figure 4.6: 2D& 3D visualisation for the case of using norms corresponding to the first three PCs . H and IRF – CS1 at 1772 RPM.

4.3.3 Fault diagnosis methodology- Fault type identification phase

As was mentioned above, all the FVs are extracted from the norms of the PCs which are obtained from projecting the trajectory matrices on to the baseline space. For the

purpose of fault type identification (i.e classification of the fault to an inner race fault (IRF), ball fault (BF) and outer race fault (ORF)), the FVs of each bearing condition category are divided equally into a training sample and testing sample. From the training sample, the FVs (i.e see Eq.(4.8)) are arranged in rows to form the so called feature matrices $\mathbf{F}_{\boldsymbol{K}}$ see Eq.(4.14)**.**

$$\mathbf{F}_{Ki} = \begin{bmatrix} f_{K_i 11} & f_{K_i 12} & \cdot & f_{K_i 1Q} \\ f_{K_i 21} & f_{K_i 22} & \cdot & f_{K_i 2Q} \\ \cdot & \cdot & \cdot & \cdot \\ f_{K_i N1} & f_{K_i N2} & \cdot & f_{K_i NQ} \end{bmatrix} \tag{4.14}$$

where

$_{Ki}$ number of the signal categories considered in the analysis.

N number of training FVs per category.

Q a feature vector dimension.

Then for any new FV, the Mahalanobis distance (D_i) to each of the feature matrices is measured. Eventually, the testing FV is assigned to the category which has a minimum D_i. The Figure 4.7 illustrates the flow chart of the fault type identification process. The figure has three main blocks, 1) building a baseline space process, 2) Extraction of FVs and 3) Fault type identification phase in which each testing FV is assigned to a specific category based on the minimum Mahalanobis distance (D_i).

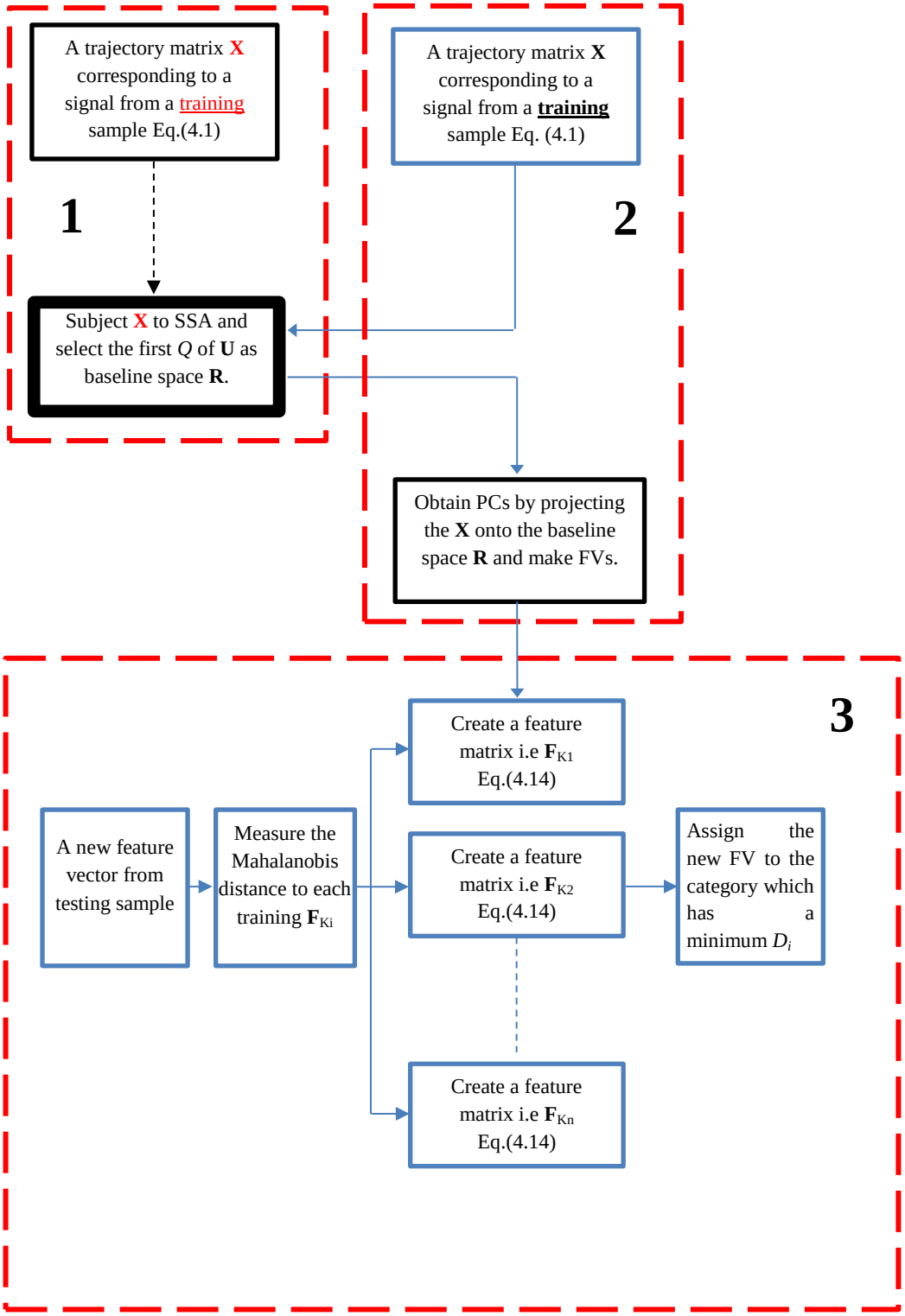


Figure 4.7: A flowchart shows the steps of the fault type identification.

4.3.4 Fault diagnosis methodology-Fault severity estimation phase

The method is also used to investigate the capabilities of fault severity estimation. This includes projecting signals corresponding to different fault severities onto the baseline space and making FVs from the resulting projections. As in the fault type identification phase, the FVs are divided into a training sample and testing sample. From a training sample, a number of feature matrices corresponding to different fault severities are made. Then, the Mahalanobis distances (D_i) of any new FV from testing sample are measured to each of the feature matrices made from the training sample. Eventually, the new FV is assigned to the category which has a minimum D_i.

In this study, the use of D_i as a fault index is also investigated. For this purpose, the D_i of the different fault severities to the baseline space is measured and used as index for fault severity change.

4.4 Experimental data

For the validation purpose of the methodology, data sets from different bearing test rigs are used. The description of the test rigs and data sets are given below:

4.4.1 Case study 1(CS1)

The bearing vibration data were obtained from the test rig of Case Western Reserve University (CWRU). The data-bearing centre [98] shown in Figure 4.8 consists of a 3 HP three-phase induction motor: a dynamometer. The drive end bearing (SKF 6025 deep grove ball bearing) data were used in this analysis. An electrical discharge machine (EDM) was used to introduce single point faults in the bearing raceways and

ball elements of different bearings with fault diameters of 0.007, 0.014, and 0.021 inches and a depth of 0.011 inches. The bearing vibration data sets were obtained at a sampling rate of 12 kHz for different fault sizes and at speed varying from 1797 rpm (0 HP) to 1730 rpm (3 HP). The data for the outer race fault were taken with the fault position centred at the 6 o'clock position with respect to the load zone.

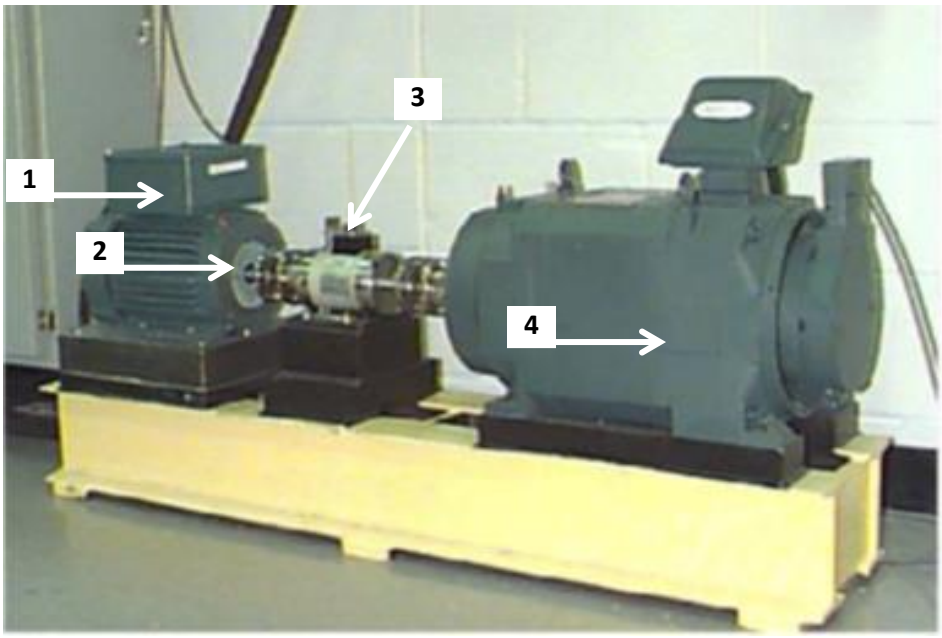


Figure 4.8: The bearing test rig of CWRU [98]. **1.** Induction motor. **2**. Accelerometer position.**3**. Torque transducer. **4**. dynamometer

The data sets obtained from this test rig and used in this study are shown in Table 4.2. The notations in the table can be defined as follow: H for healthy category, IRF for inner race fault category, BF for ball fault category, ORF for outer race fault category, S refers to the small fault size (i.e 0.007 inch), M refers to medium fault size (i.e 0.014 inch and L refers to large fault size (i.e 0.021 inch). The first four cases (i.e CW1-CW4) represent the data sets corresponding to healthy and different small fault locations categories obtained at different rotational speed (i.e 1730, 1750, 1772 and 1797 RPM).

The remaining cases (i.e CW 5-CW16) represent the data sets corresponding to Healthy and different fault severities.

Case no.	speed(rpm)	Signal Category
CW1	1730	Healthy, (IRF, BF and ORF) small fault 0.007''
CW2	1750	Healthy, (IRF, BF and ORF) small fault 0.007''
CW3	1772	Healthy, (IRF, BF and ORF) small fault 0.007''
CW4	1797	Healthy, (IRF, BF and ORF) small fault 0.007''
CW5	1730	H&IRF (S,M and L)
CW6	1750	H&IRF (S,M and L)
CW7	1772	H&IRF (S,M and L)
CW8	1797	H&IRF (S,M and L)
CW9	1730	H&BF (S,M and L)
CW10	1750	H&BF (S,M and L)
CW11	1772	H&BF (S,M and L)
CW12	1797	H&BF (S,M and L)
CW13	1730	H&ORF (S,M and L)
CW14	1750	H&ORF (S,M and L)
CW15	1772	H&ORF (S,M and L)
CW16	1797	H&ORF (S,M and L)

Table 4.2: The drive end bearing vibration datasets obtained from case study 1 used in the bearing condition diagnosis

The method was investigated using different sub-signal lengths (i.e the long signals are segmented into shorter sub-signals) where the method is applied using each of the sub-

signal length (512, 1024, 2048 and 4096 data points). The number of the sub-signals obtained at for a different sub-signal length is shown, for all the Case studies used in the analysis, in Tables 4-A.1, 4-A.2& 4-A.3 (i.e *in appendix 4-A attached at the end of the thesis*). The number of shaft revolutions included in a certain sub- signal length (*n*) can be calculated from the following Eq. 4.15.

$$Nr = \frac{n * Sp}{60 * Sr} \tag{4.15}$$

where

Nr the number of shaft bearing inner race revolutions in a sub-signal length (*n*).

n a sub-signal length.

Sp the rotational speed of the inner race of the bearing (RPM)

Sr the sampling rate (Hz)

For example, for a sub-signal of 2048 data points length obtained at 12 kHz and a rotational speed of 1797 RPM, the number of revolution will be (2048*1797/(60*12000))=5.11 revolutions.

4.4.2 Cases study 2 (CS2)

The experimental test rig was designed and set up ,and the measurements were taken by the author in the laboratories at the Department of Mechanical and Aerospace Engineering and the University of Strathclyde. The experimental test set-up is shown in Figure 4.9. The test rig consists of a 1 HP shunt DC motor, a bearing assembly and a

mechanical loading system. The bearings used in the experiment are SKF 6308 single row deep groove. The motor shaft's torque is transmitted to the test bearing assembly by a pinion-toothed belt mechanism. As the pinions at the motor and the entrance of the bearing assembly are of different diameters, the rotational speed of the test bearing shaft is measured by a contactless tachometer. Faults were introduced using an electrical discharge machine on the inner raceway, a rolling element and outer raceway using different bearings with a fault diameter of 0.05 inches in approximate. This fault size was deliberately made small by the author to challenge the performance of the methodology in detecting small faults. Due to the limitations of the EDM tool size at the time of conducting the experiments, a smaller fault size than this was not possible. However, cases of smaller fault sizes were covered in the other case studies (i.e CS1 and CS3). The bearing vibration data were obtained for healthy, inner raceway, a ball and outer raceway fault conditions at three rotational speeds (250, 750 and 1250 RPM). Signals were obtained at a 12 kHz sampling rate.

A healthy bearing is mounted on the shaft and inside the bearing housing, and then data are acquired for acceleration vibration signals at a specific speed, such as 250 RPM. Then the motor speed is increased and another five signals are obtained at the new speed and the same is done for the speed considered in the analysis. When the measurement from the healthy signals is finished, a faulty bearing, such as a bearing with a fault in its inner raceway, is mounted in the test rig and the author follows the same measurement procedure. The measurements are repeated for the other faulty bearings.

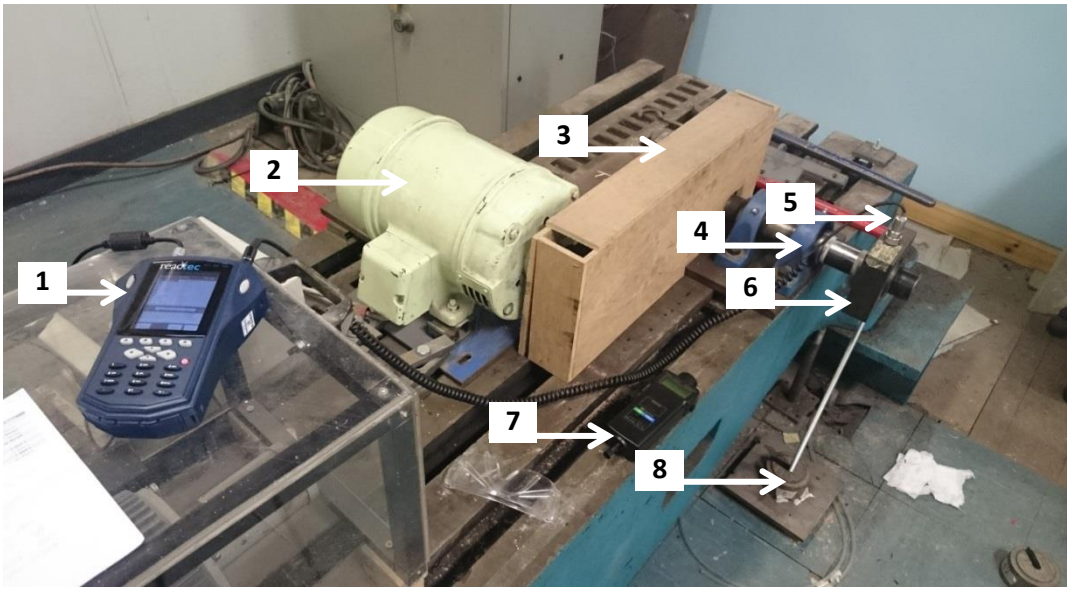


Figure 4.9: Bearing test rig CS2. **1**- Portable vibration analyser. **2**- DC motor. **3**- Wooden shield to cover the toothed belt and pinion. **4**- Supporting journal bearing. **5**-accelerometer. **6**-test bearing housing**. 7**-wireless tachometer. **8.**weights.

The data sets obtained from this test rig and used in this study are shown in Table 4.3 where the ST denotes Strathclyde University where the bearing test rig is.

Case no.	Motor speed(rpm)	Signal Category
ST1	250	Healthy, (IRF, BF and ORF) small fault 0.05''
ST2	750	Healthy, (IRF, BF and ORF) small fault 0.05''
ST3	1250	Healthy, (IRF, BF and ORF) small fault 0.05''

Table 4.3: The bearing vibration datasets obtained from case study 2 used in the bearing condition diagnosis

4.4.3 Case study 3 (CS3)

The data were acquired from a test rig assembled at the Department of Mechanical and Aerospace Engineering of Politecnico di Torino by the Dynamics & Identification

Research Group (DIRG) (Figure 4.10) [99]. The author acknowledges the help of colleagues from the University of Torino in providing the data. The bearing is a roller bearing supplied by the SKF according the specification required by the teamwork. The signals were acquired at a 102.4 kHz sampling frequency for both healthy, defective inner raceway and defective roller elements at 18000, 24000 and 30000 RPM shaft speed and 1.4 kN load. For the cases considered in this thesis, the defects are made by Rockwell type indentation. The have three different diameters namely 0.006, 0.0098 and 0.0178 inches. The data sets obtained from this test rig and used in this study are shown in Table 4.4, where PT denotes Politecnico Torino.

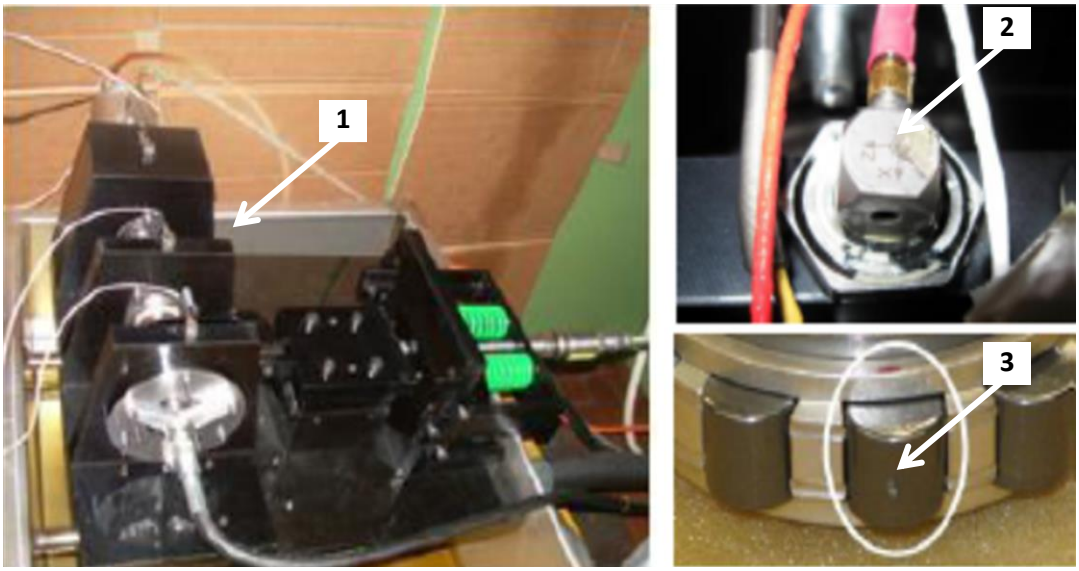


Figure 4.10: DIRG test rig, the triaxial accelerometers (X, Y, Z) and the damaged roller used in the tests [99]. **1.** Bearing assembly. **2.** Three axes accelerometer. **3**. Indention on a roller element

Case no.	Motor speed (rpm)	Signal Category
PT1	18000	(H, IRF and BF) small fault 0.006 inch diameter
PT2	24000	(H, IRF and BF) small fault 0.006 inch diameter
PT3	30000	(H, IRF and BF) small fault 0.006 inch diameter
PT4	18000	H&IRF (S, M and L)
PT5	24000	H&IRF (S, M and L)
PT6	30000	H&IRF (S, M and L)
PT7	18000	H&BF (S, M and L)
PT8	24000	H&BF (S, M and L)
PT9	30000	H&BF (S, M and L)

Table 4.4: The bearing vibration datasets obtained from case study 3 used for the bearing condition diagnosis

4.5 Variety of the data sets obtained from the three bearing test rigs

From Tables 4.2, 4.3 and 4.4 it can be seen that a wide range of data sets were used for validation of the present methodologies. The data sets cover a variety of bearing rotational speed ranges. For the test bearing at Case Western Reserve University the speed ranges used were from 1730 RPM to 1797 RPM. For the test bearing at Strathclyde university the speed ranges went from 250 RPM (which is a relatively low speed) up to 1250 RPM. For the test bearing at the Politecnico di Torino the speeds are very high, and range from 18000 RPM to 30000 RPM. For all of these speeds the methodology performs very well and shows a 100% correct classification for most of the cases.

The test bearings cover various fault severities which are detected and diagnosed by the current methodologies. Some of these fault severities were very small such as the one used in the Case Reserve Western university (i.e CS1) which is 0.007 inch, the one used at Strathclyde university (i.e CS2) which is 0.050 inch and the one used in the Politecnico di Torino (i.e CS3) which is 0.006 inch in diameter (around one tenth of a millimetre).

The structure of the test bearings was not the same and the bearings used were also different. For example, the bearing used in the test rig at Politecnico di Torino is a specifically made rolling element bearing of a special dimension, while the other test rigs used different types of ball bearing.

The sampling rate of the data taken from Politecnico di Torino was different to that which was used in the other test bearing rigs. The loads applied to the bearing were different for the entire test bearing rigs. The author believes that the three case studies cover a wide range of speeds, a good range of bearing types, a good range of fault extensions and severities from very small ones to those that are rather large, a good variety of fault types covering most of the faults experienced in bearings, and a good range of operational conditions. In conclusion the author is confident that the methodology developed is general enough to be applied for most types of machines.

4.6 Summary

The chapter introduces the fundamentals of the singular spectrum analysis. It first explains both stages of the SSA namely; the decomposition stage and the reconstruction

stage. Then it introduces the steps of the SSA-based methodology in detail. For validation of the methodologies developed in this thesis, the description of a three bearing test rig is given in this chapter. These bearing test rigs are different in terms of the bearing specifications, the fault severities and the rotational transmission mechanism. The obtained acceleration vibration data sets cover different fault locations, different fault severities and different shaft rotational speeds. The variety in the data set specifications gives good support for the generalisation of the application of the methodologies.

Chapter 5

Results and discussion – a new SSA-based methodology for fault diagnosis in REBs

5.1 Introduction

The section of the results and discussion has two parts. The first part showing a demonstration for the method presented in this chapter using a data set from the CS1. The second part illustrates the results obtained for the three phases of the fault diagnosis (i.e fault detection, fault type identification and fault severity estimation).

5.1.1 Part 1 - Methodology demonstration based on CS1

The demonstration of the methodology steps is presented in this section using a data set from CS1. The selected data set for demonstration contains the signals obtained at 1730 RPM for a healthy bearing (H), a bearing with inner race fault (IRF), a bearing with a ball fault (BF) and a bearing with outer race fault (ORF) with a fault size of 0.007 inch in diameter (i.e CW1 in Table 4.2). Figure 5.1 presents a sub- signal corresponding to the healthy bearing category in the time domain at 1730 RPM. The x-axis represents the number of the data points while the y-axis is the acceleration of the signal vibration in m/sec^2. Although recording the signal in the time domain provides rich information, no

direct information can be easily obtained from presenting it in its time domain. The next step is subjecting the lagged version of the sub-signal to SSA in order to unfold it through the decomposition stage and extract more information.

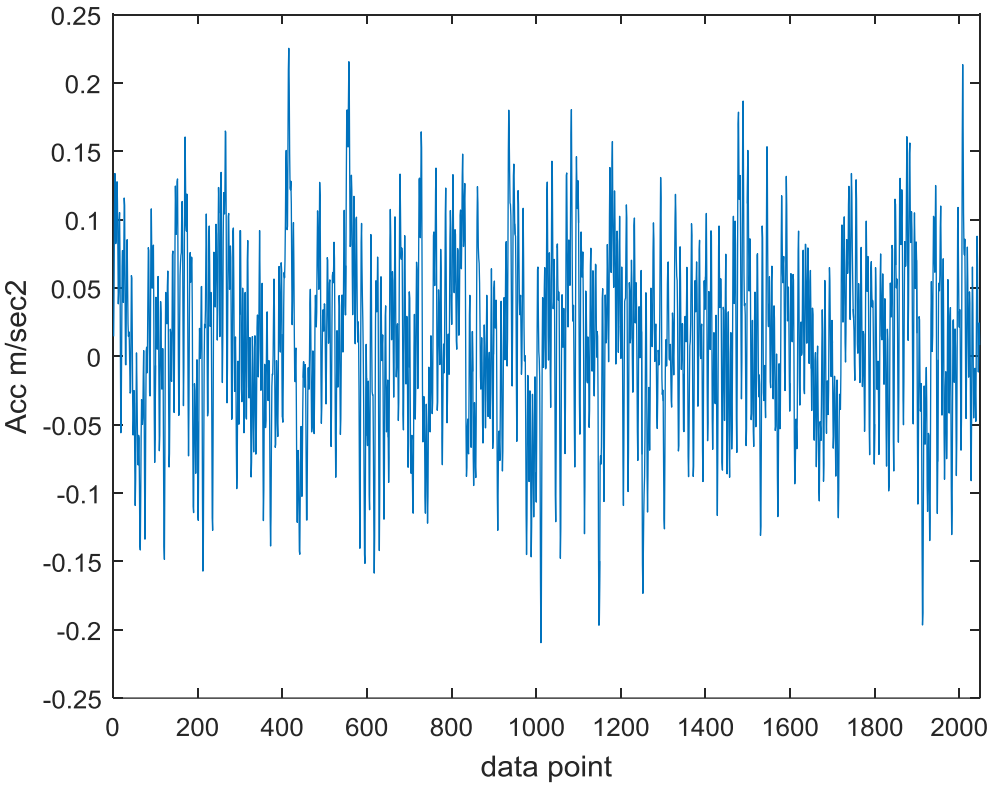


Figure 5.1: A sub- signal of 2048 data points corresponding to a healthy bearing at 1730 RPM. CS1-CW1

The subjection of the sub-signal above to the 1st stage of SSA, leads to decomposition into a number of PCs equal to the number of the embedding window, which is 10 in this case. Each PC contributes in a certain ratio of the original sub-signal variance. The distribution of this contribution versus the number of PCs is shown in the so-called scree plot as in Figure 5.2. The plot displays the normalised eigenvalues versus the number of PCs for a window L =10. The y axis represents the normalised eigenvalues

(i.e $normalised\ \lambda_i = \lambda_i / \sum_{i=1}^{L} \lambda_i$) which explains the percentage of variance portions explained by each of the PCs. It is clearly seen that the first PCs have the highest variance portions when compared to the last ones. Larger (but should not be larger than half of a sub-signal length) or smaller embedding windows can also be used but in this study selecting the embedding window as 10 was convenient. When a larger embedding window is used, the variance portions contained in the first PCs decreased and the variance portion in the far PCs increase and vice versa. As was mentioned in section 4.3.2, any number of PCs can be selected and used in the analysis to form the FVs. In this study, the methodology's correct classification rate has been checked using different PC numbers (i.e from 1 to 10) and eventually the minimum number of PCs to achieve the maximum correct classification rates. However for the purposes of visualisation only the 1st three PCs are used. In Figure 5.2, the variance proportion accounted for in the first three PCs is at least 75% of the variance of the original signal and it meets one of the criteria mentioned in [82].

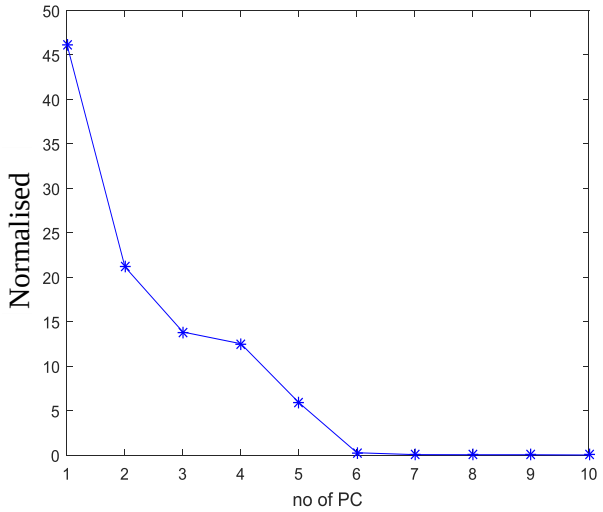


Figure 5.2: Scree plot of a healthy signal at 1730 RPM

As mentioned in section 4.3.1, that baseline space is made from the eigenvectors corresponding to a healthy bearing category. The lagged version of any new sub-signal is projected on that baseline space. The resulted projections are made to obtain the FVs. Figure 5.3 represents a 3D visualisation of 3D feature space. This feature space represents the norms of the projections of the FVs on the baseline space. The horizontal axes represent the values of the 1st and 2nd features of the FV respectively (i.e norms of the projections on to the $\mathbf{U_1}$ & $\mathbf{U_2}$) while the perpendicular axis represents the value of the 3rd features (i.e the norms of the projection on to the $\mathbf{U_3}$). The number of the FVs shown in this figure is 30 FVs corresponding to the baseline training sample (H_{tr}) category.

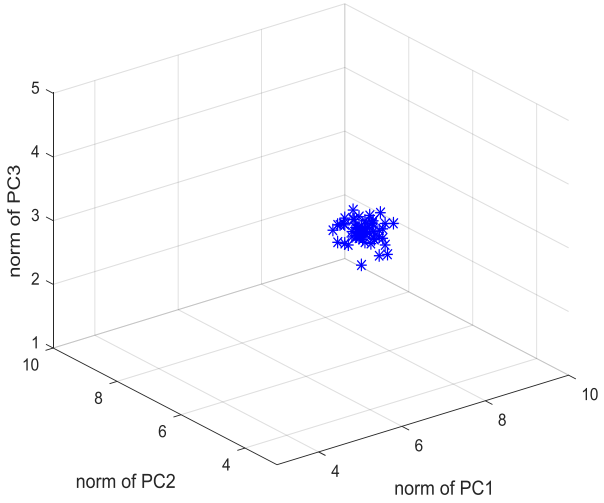


Figure 5.3: 3D visualisation of feature space corresponding o H_{tr} category (baseline condition) at 1730 RPM-CS1-CW1

Since the baseline eigenvectors are corresponding to a healthy bearing, it is expected that FVs corresponding to non baseline bearing condition will differ from the baseline the FVs. The Figure 5.4 shows a 3D visualisation and 2D views of 3D feature space.

The 3D- FVs shown in the figure are corresponding to Healthy (H) and different faulty categories (F). The total number of FVs is 240 (60 for H and 180 for F) corresponding to a sub-signal length of 2048 and a rotational speed 1730 RPM. As was mentioned in Section 4.3.2, the trajectory matrix of each testing sample sub- signals is projected on to the baseline space (i.e multiplied by baseline eigenvectors). Figure 5.4-**A** shows the 3D visualisation of the FVs corresponding to the baseline condition in blue and the other different faulty bearing conditions in red for IRF, green for BF and black for ORF. It can be seen that FVs corresponding to the faulty classes are well separated from the baseline conditions and they are also separated from each other. Figures 5.4-**B**, 5.4-**C** and 5.4-**D** show the distribution of the FVs from different 2D views. Figure 5.4 shows that selecting three PCs to form the FVs improves the separation not only between the baseline and non-baseline categories but also among the non-baseline categories as well.

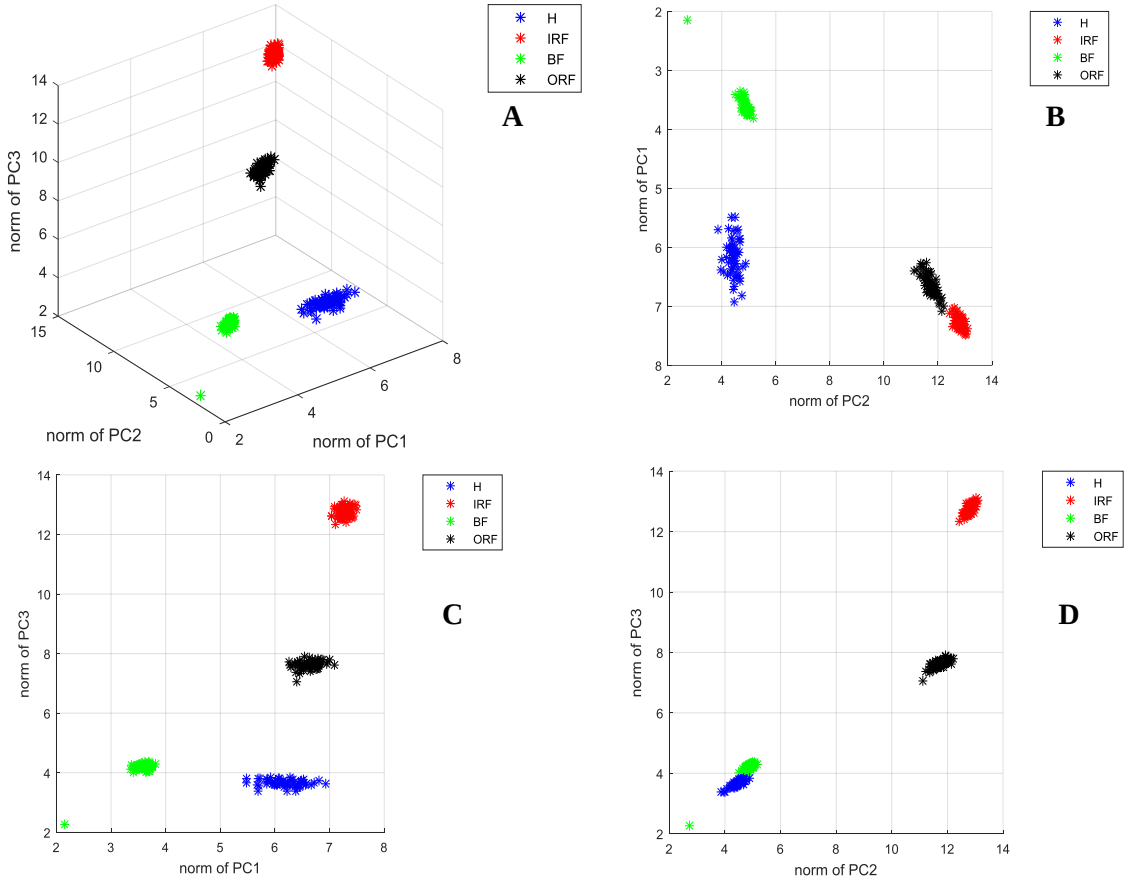


Figure 5.4: The 2D & 3D visualisation of the FVs corresponding to different bearing conditions obtained at 1730 RPM, sub-signal length (2048).**A.**3D view, **B.**PC1&PC2 view**, C.**PC1&PC3 view. **D.** PC2&PC3 view.

The separation of the non-baseline FVs from the baseline one can be assessed by calculating the Mahalanobis distance (*D*) (see Eq. (4.10)) and comparing it to a predetermined threshold. Figure 5.5 presents the *D* of the testing and the training FVs to the baseline FVs. From the *D* corresponding to the healthy training sample a threshold ($Thr_{baseline}$) is determined based on the lognormal probability density function as was described in Section 4.3.2 (see Figure 5-B.1 in appendix 5-B for comparison of lognormal probability distribution with the *D* histograms corresponding to healthy training sample). As it is seen from the figure, there are five regions /parts of *D*

separated by vertical dashed lines. The most left part represents the *D* of the ‘’Training baseline FVs” which represents the *D* of each training FV measured to $\mathbf{F}_{\text{baseline}}$. This is the part where the threshold is made. The threshold is represented by the horizontal dashed line in the figure. The second left part represents the *D* of the ‘’Testing baseline FVs’’ measured to the baseline FVs. The other parts (i.e the last three) represent the *D* of the testing FVs corresponding to different fault locations. It is clear that all the *D* corresponding to training baseline FVs are below the $Thr_{baseline}$ and most of the ones corresponding testing healthy sample are also still below (i.e only two out of 30 locate a slightly above the $Thr_{baseline}$ value). However, all the other distances corresponding to the faulty condition are quite far from the $Thr_{baseline}$ and the healthy bearing category. The y-axis is in a log scale in order to show how well the non-baseline FVs are separated from the baseline ones.

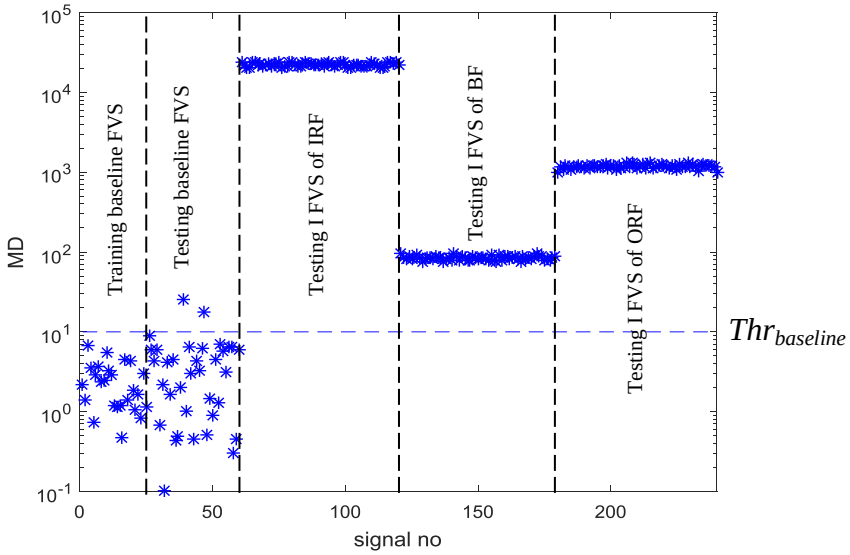


Figure 5.5: Shows the *D* as an indicator of the fault detection in the bearing 1730 RPM.

The total FVs used in this example is 240 (4 categories (i.e H, IRF, BF and ORF) * 60 (FV for each category)). Half (i.e 30) of the (H) category *D* values is used to determine the baseline threshold while the other 30 as well as the other 180 *D* values corresponding to the faulty (F) category are used as testing sample. The confusion matrix corresponding to the classification of testing FVs obtained at 1730 RPM is shown in Table 5.1. It is clearly shown that the positive false alarm (i.e FVs corresponding to the healthy class but assigned as faulty) is 6.7% as only 2 out of 30 is misclassified while there is no negative false alarm (i.e none of the FV corresponding to faulty bearing category is assigned to the healthy category).

Actual class/predicted classes	H	F
H	93.3%	6.7%
F	0%	100%

Table 5.1: Confusion matrix of testing FVs, 1730 RPM.

It is important to mention that the present method can be successfully applied to a real complex machine. This is because the baseline space will be created based on the real data of the machine itself (not from a simulated model) using the healthy condition. Then, a threshold will be set up from this healthy category data and any anomaly will be detected and classified as an observation different to the baseline space.

5.1.2 Part2- Fault diagnosis

In this part, results are presented for all the three phases of fault diagnosis; 1-Fault detection, 2: Fault type identification and 3. Fault severity estimation (see section 4.3).

Fault detection

This part of the results shows the methodology performance in distinguishing between healthy and faulty categories (i.e baseline and non-baseline). As was mentioned above (section 4.3.2), the FVs corresponding to the healthy category are equally divided into a training sample and a testing sample. The training sample is used to build the baseline space and setting a threshold while the testing sample is used to validate the methodology. All the FV corresponding to a fault category are used as a testing sample. The methodology is applied for different FV dimensions (i.e different baseline space dimensions) and different sub-signal lengths. The results for the correct classification rates are shown in the following sections:-

A study of FV dimension and sub signal length effect on fault detection accuracy rate

The method is applied for different FV dimensions (i.e from 1D to 10 D) in order to investigate the effect of the FV dimension on the fault detection accuracy. In the case of 1D, only the first PC is used to make the FV and in the case of 2D only the first two PCs are used for the FV and the same process for other FV dimensions. Figure 5.6 shows the *D* (i.e Mahalanobis distance of testing FV to the baseline) at different FV dimensions and for testing FVs corresponding to the H category and IRF category. These two categories are selected for the simplicity of showing the effect of baseline space dimension on the D. The x-axis represents the number of the testing FV used in the classification process while the y-axis represents the *D* of the testing sample FVs to the baseline feature space. It can be seen that the *D* values increases with the increase of the dimension of the FVs. At 1D, there is no considerable deviation between the

healthy and faulty categories but the deviation becomes clear when the FV dimension increases. This can be explained as considering more dimensions means including more PCs and more information about the fault. The figure shows the increase in *D* using the FV of one dimension (1D), where only the first PC is used in extracting the features, up to the use of a 5 dimensional FV (5D), where the first five PCs are used in the extracting the features.

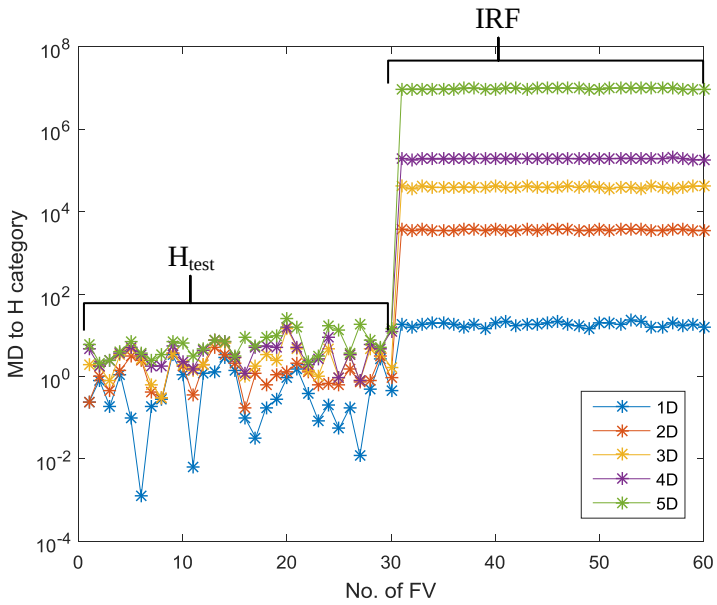


Figure 5.6: Effect of increasing the FV dimension of on *D*

Figure 5.7 **A-D** shows the effect of increase of FV dimension on the correct classification rate of the proposed methodology. The x-axis represents the dimension of the FV while y-axis represents the correct classification rate in percent. The red line corresponds to the FV from faulty bearing class whereas the blue line corresponds to the FV from the healthy testing sample.

Generally, there are several notes that can be interpreted from the figure. First, a general phenomenon can be seen in the FV dimensionality versus the correct classification rates. It can be easily seen that the distinguishability of the FVs corresponding to the faulty category is improved with the increase of FV dimension. This can be interpreted as an increase of FV dimension (i.e when more PC components are used) meaning the involvement of more information about the bearing fault. Second, the correct classification of FV corresponding to the H testing sample decreases with the increase of the FV dimension. And this decrease is much clearer at larger sub-signal length particularly at 4096 (i.e Figure 5.7-**D**).

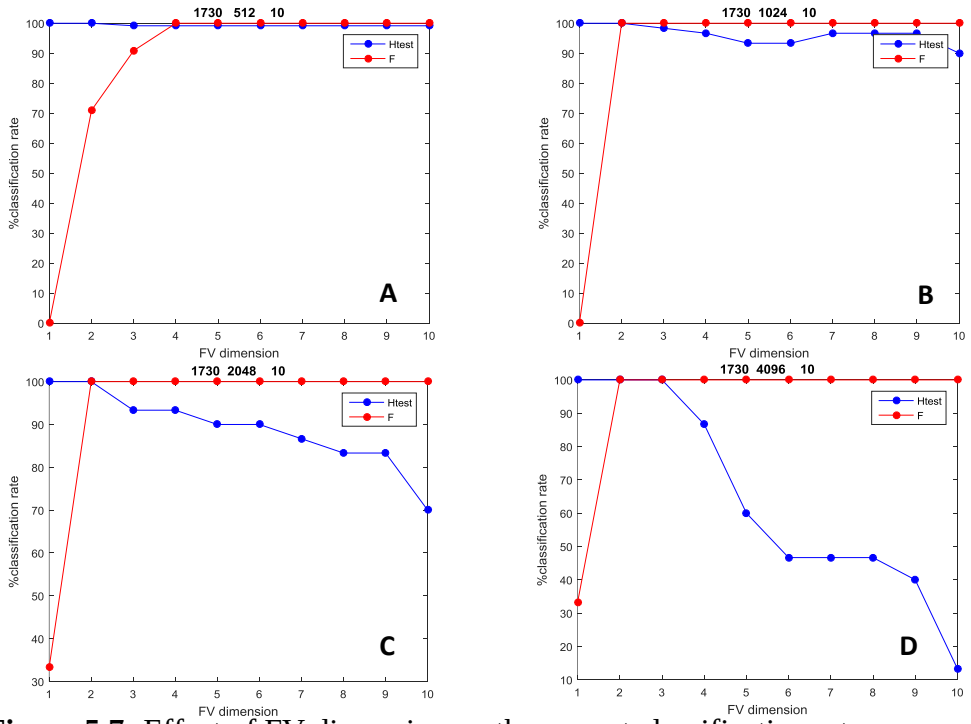


Figure 5.7: Effect of FV dimension on the correct classification rates. CW1-1730 RPM at different sub-signal length **A**-512, **B**-1024, **C**-2048 and **D**-4096

However, it can be seen that at some certain FV dimension the correct classification rates reach a maximum and there is no need to add more PCs. The same notes can be

made for the CS-2&3 as shown in Figures 5.8-5.9 respectively. Figure 5.8 shows the correct classification rates at different FV dimensions and at different sub-signal length corresponding to CS2. Figure 5.9 shows the correct classification rates at different FV dimensions and at different sub-signal length corresponding to CS2. In all these Figures it can be seen when the sub-signal length increased, the FV dimension to achieve the maximum correct classification rate decreases.

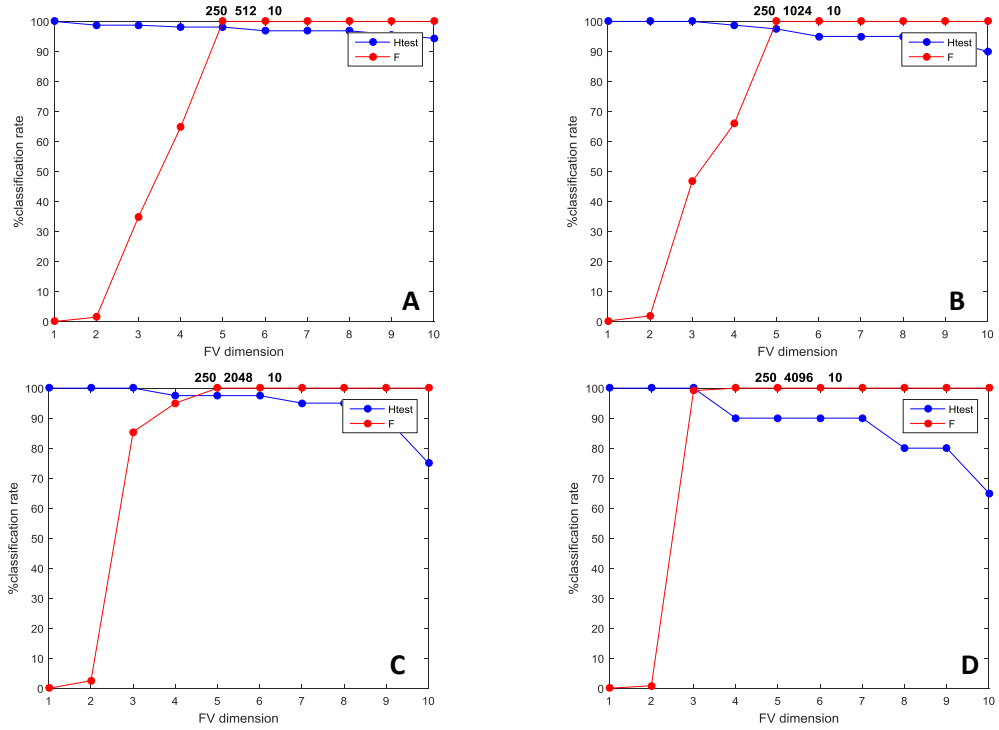


Figure 5.8: Effect of FV dimension on the correct classification rates. ST1-250 RPM at different sub-signal length **A**-512, **B**-1024, **C**-2048 and **D**-4096

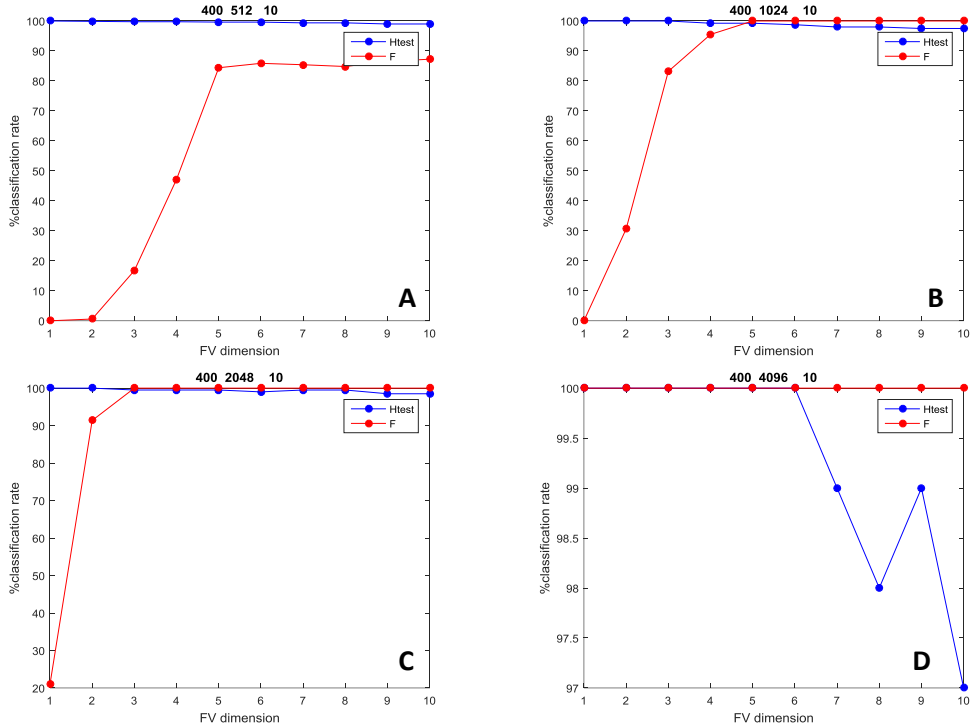


Figure 5.9: Effect of FV dimension on the correct classification rates. PT4-24000 RPM at different sub-signal length **A**-512, **B**-1024, **C**-2048 and **D**-4096

Minimum baseline space dimension to achieve the highest classification rate.

Recalling the Figures 5.7-5.9 above, which are showing the effect of FV dimension (i.e baseline space dimension) on the correct classification rates, It can be seen that the minimum baseline space dimension to achieve the highest correct classification rates of FV corresponding to fault bearing category is case dependent (i.e it is not a unique for all the cases). For example in the Figure 5.7-**A**, using the 3D baseline space is sufficient to achieve 100% of fault detection rate for sub-signal length of 512 data points. However when a longer sub-signal is used (such as 1024, 2048, 4096 in the Figures 5.7 **B, C, D** respectively), lower baseline space dimension is required (i.e only 2). This can be interpreted as incorporation of more data points in the sub-signal means

incorporation of more information about the fault in the sub-signal. Consequently more information about the fault will be involved in the first PCs.

Table 5.2 shows the minimum baseline space dimension required to achieve maximum fault detection rates for the data sets of CS1 for different speeds and different sub-signal lengths. The correct classification rates are shown in the form '' **minimum baseline space dimension** [%classification rate of H_{test}, %classification rate of F]''. The numbers out of the brackets represents the minimum baseline space dimension required to achieve the maximum fault correct classification, while the numbers inside the square brackets refer to correct classification rate of the FVs corresponding to the healthy testing sample and the correct classification rate of the FVs corresponding to the faulty testing sample respectively in percent.

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
1730	**4**[99.2,100]	**2**[100,100]	**2**[100,100]	**2**[100,100]
1750	**3**[97.5,100]	**2**[100,100]	**2**[100,100]	**2**[100,100]
1772	**4**[96.5,100]	**2**[100,100]	**2**[100,100]	**2**[100,100]
1797	**3**[100,100]	**2**[100,100]	**2**[100,100]	**1**[100,100]

Table 5.2: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS1-fault detection phase

However, it can be seen that, in general, the minimum baseline dimension tends to decrease with the increase of sub-signal length. For example, at 1797 RPM it is three at sub-signal length of 512 but it decreases to one when sub-signal length is 4096. Results of correct classification rates for CS2 and CS3 are shown respectively in the Tables 5-C.1 and 5-C.2 (Appendix 5-C).

Fault type identification

In the previous section, the classification of FVs into only a baseline and non-baseline categories are discussed. In this section, the results of the correct classification rates of the FVs to their actual categories are presented (i.e to baseline and different non-baseline categories such as inner race fault, ball fault and outer race fault). Figure 5.10 shows the average and lowest correct classification rates versus FV-dimension for CS1 at 1730 RPM. The average correct classification rate is the mean value of the main diagonal of the corresponding confusion matrix while the lowest correct classification rate is the lowest value of main diagonal of the same confusion matrix. Generally, it can be seen that increasing the baseline space dimensions improves the correct classification rates. As seen in the detection phase, the average correct classification rates reach to a maximum value at a certain baseline space dimension. This means that it is unnecessary to increase the baseline space dimension. It can be seen also that the minimum baseline space dimension, which is required to obtain the highest average correct classification, decreases with the increase of the sub-signal length. For example, in the Figure 5.10, which shows the classification rates for FV made at different sub-signal length (**A**-512, **B**-1024, **C**-2048 and **D**-4096) the minimum baseline space dimension (i.e minimum FV dimension) is (2 at 512, 2 at 1024, 1 at 2048 and 1 at 4096). After these minimum dimensions it is unnecessary to increase the baseline space dimension because the average correct classification rates reach maximum.

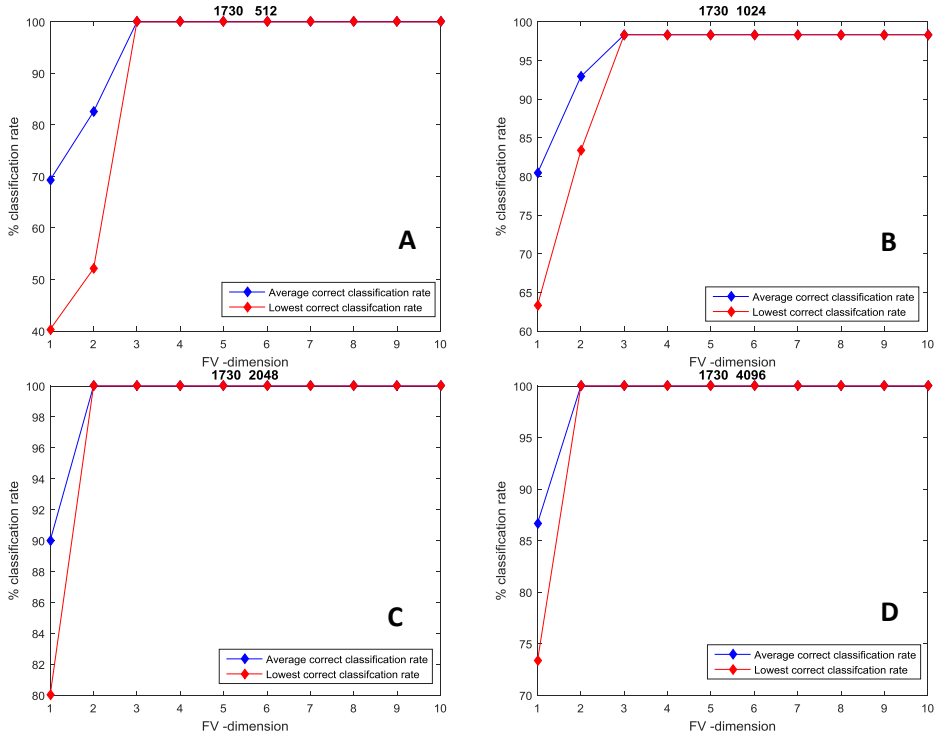


Figure 5.10: Average and lowest correct classification rate versus FV-dimension CS1, 1730 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

Figures 5.11 & 5.12 show the average and lowest correct classification rates for CS2 and CS3 respectively. For CS2, the method shows poor performance for the case where the sub-signal length is 512 data points while it shows a good performance at the higher sub-signal lengths. The trend of classification rates improvement shows an increase with the increase of baseline space dimension. Another note can be seen that increasing the sub-signal length decreases the baseline space dimension required for better correct classification rates.

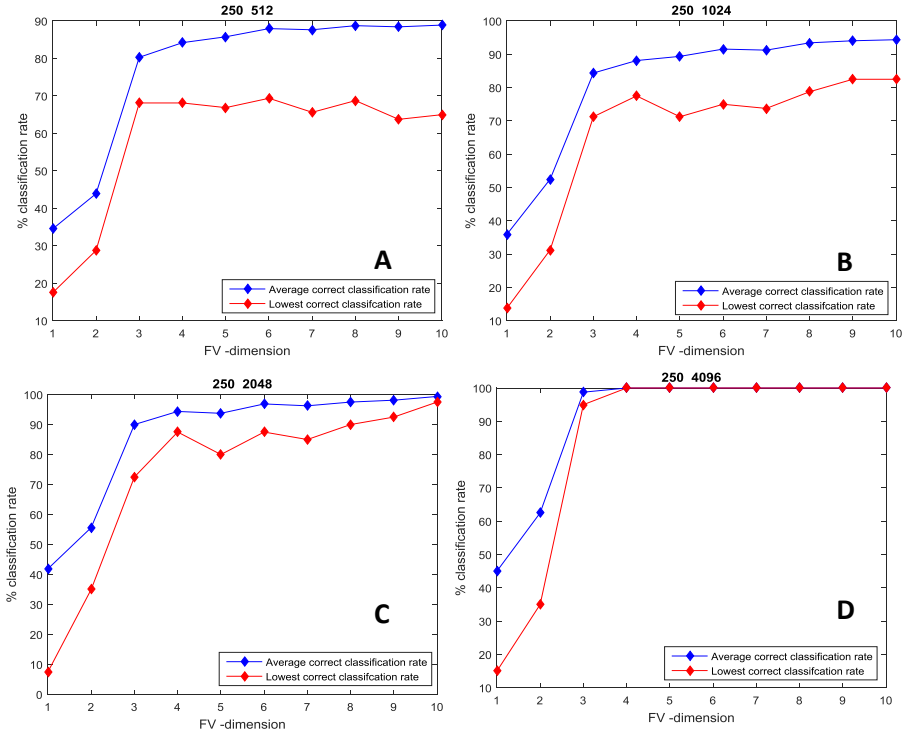


Figure 5.11: Average and lowest correct classification rate versus FV-dimension CS 2, 250 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

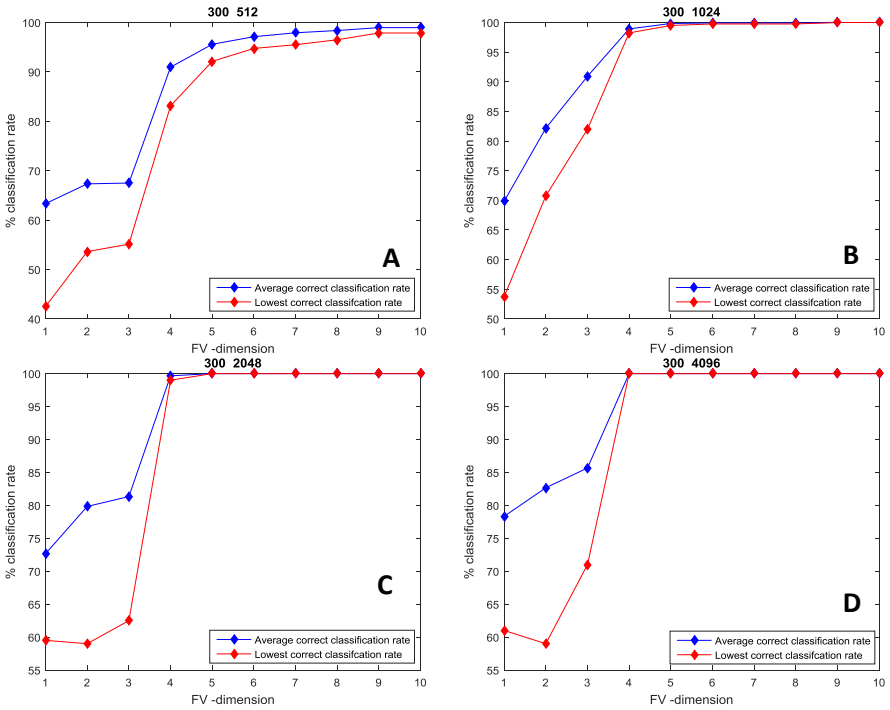


Figure 5.12: average and lowest correct classification rate versus FV-dimension CS3, 18000 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

For the data sets corresponding to CS3, the average and lowest correct classification rates has the same trend as in CS2.

Table 5.3 shows correct fault type classification rate in percent for the data sets of CS1. The numbers beside the square brackets defines the minimum baseline space dimension while the numbers inside the square brackets show the average and lowest correct classification rates respectively. As mentioned previously in section 4.3.3 the FVs are divided into a training sample and a testing sample. From the training sample a number of feature matrices are made. The testing sample is used for validating the methodology. It can be seen from the table, the method shows a very good performance for all the data sets considered in the CS1 and for the different sub-signal length. The number of FVs used in the fault type identification phase is 960 (4category *240 FVs for each category) for sub-signals of 512, 480 FVs for sub-signal length of 1024, 240 FVs for sub-signal lengths of 2048 and 120 FVs for sub-signal length of 4096. The results for other case studies CS2&CS3 can be seen in Table 5-C.3 and 5-C.4 in appendix5- C

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
1730	**3**[100%,100%]	**3**[100%,100%]	**2**[100%,100%]	**2**[100%,100%]
1750	**3**[100%,100%]	**2**[100%,100%]	**2**[100%,100%]	**2**[100%,100%]
1772	**3**[100%,100%]	**2**[100%,100%]	**2**[100%,100%]	**2**[100%,100%]
1797	**2**[100%,100%]	**1**[100%,100%]	**1**[100%,100%]	**1**[100%,100%]

Table 5.3: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS1-fault type identification phase.

Fault severity estimation

The methodology is also investigated for the detection of different fault severities. Figures 5.13, 5.14 and 5.15 show a 3D visualisation of 3D feature space corresponding to healthy and different fault severities categories for CS1. Figure 5.13 shows an example of the clustering of FVs corresponding to healthy and different fault severity at inner raceway at sub-signal length of 2048. The total number of FVs used in this example is 240 (4 categories * 60 FV foe each category). The SIRF, MIRF and LIRF refer to the small (0.007 inch), medium (0.014 inch) and large (0.021 inch) fault size at the inner raceway respectively. The figure shows that the methodology is able to distinguish not only between the baseline and non-baseline categories but also among the different fault size non-baseline categories. Figure 5.14 shows the visualisation of the 3D feature space of FVs corresponding to healthy and different fault severities at a ball element. Although the FVs corresponding to MBF (i.e medium ball fault size) are not very close each to other but they are distinguished from other categories. Figure 5.15 shows the visualisation of the 3D feature space of FVs of FVs corresponding to healthy and different fault severities at outer raceway (ORF). It shows a very good distinguishing of different fault severities signal categories from the baseline (i.e healthy) and among themselves as well which indicates the robustness of the present method.

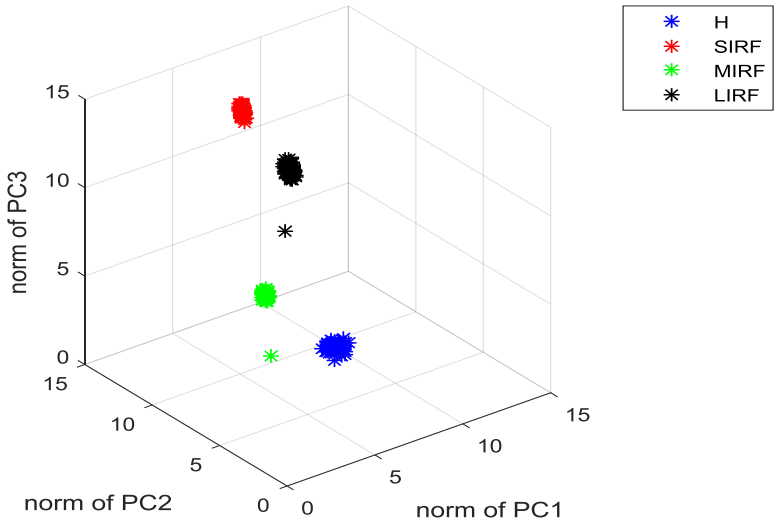


Figure 5.13: Clustering of FVs which corresponding to healthy and different inner race fault (IRF) severities at 1730 RPM and sub-signal length 2048

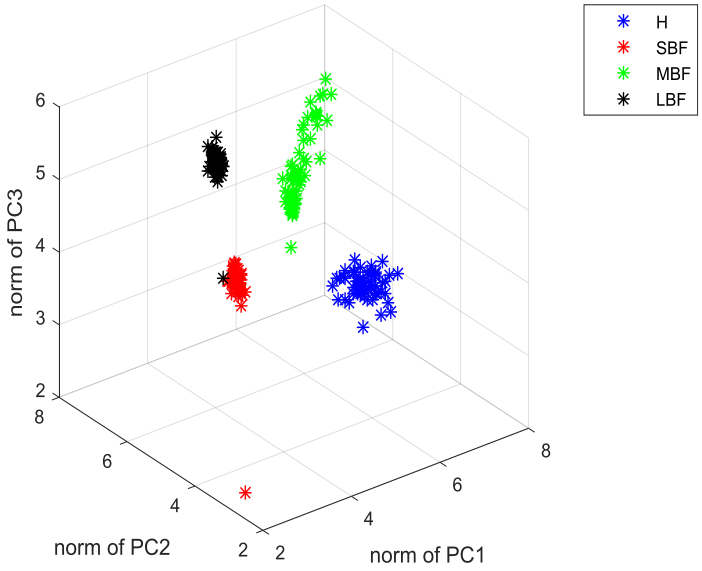


Figure 5.14: Clustering of FVs which corresponding to healthy and different ball fault (BF) severities at 1730 RPM and sub-signal length 2048

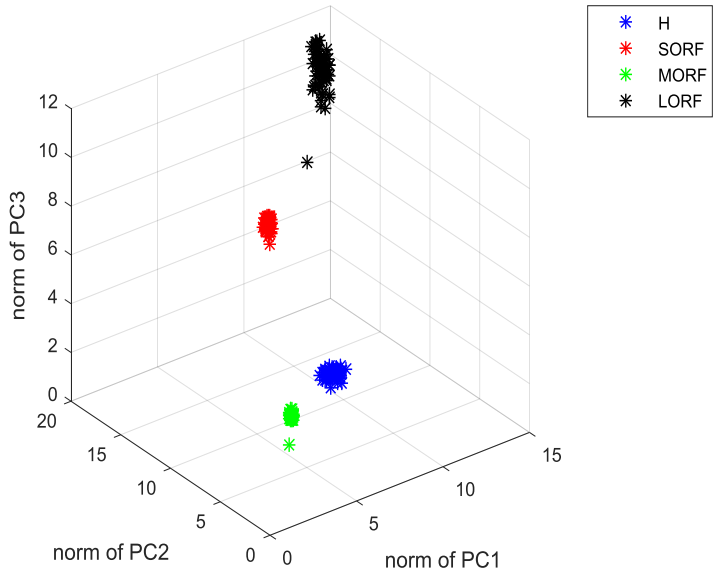


Figure 5.15: Clustering of FVs which corresponding to healthy and different outer race fault (ORF) severities at 1730 RPM and sub-signal length 2048.

A study of FV dimension and sub signal length effect on detection accuracy rate of different fault severities

The method performance in terms of detecting different fault severities is also investigated. First the capability of the method in distinguishing the healthy and different fault severities FVs into a baseline and non-baseline is firstly investigated. As described previously, from the Mahalanobis distances (D) of the healthy training sample to the baseline a threshold is made. The D values corresponding to the testing sample FVs are compared to the $Thre_{basline}$ and used to validate the method. Figure 5.16 shows the average correct classification rates of the testing FVs versus different FVs dimension (i.e different baseline space dimension). The F and H_{test} notations in the figure refer respectively to the testing sample of Faulty and Healthy FVs categories. The

average correct classification rates in the figure shows how well the method is able to distinguish the FVs corresponding to different fault severities into a baseline and none-baseline categories. The same general phenomenon, which can be seen in the case of detection of fault locations (i.e see page 76), can also be noticed here. The increase of FV dimensionality (i.e involving more PCs in building the space lines) helps in incorporating more information about the fault presence. This leads to an improvement in the average correct classification rates of the FVs corresponding to the F category. On the other hand, the average correct classification rates for the FVs corresponding to the H_{tes} decreases with the increase of FV dimensionality.

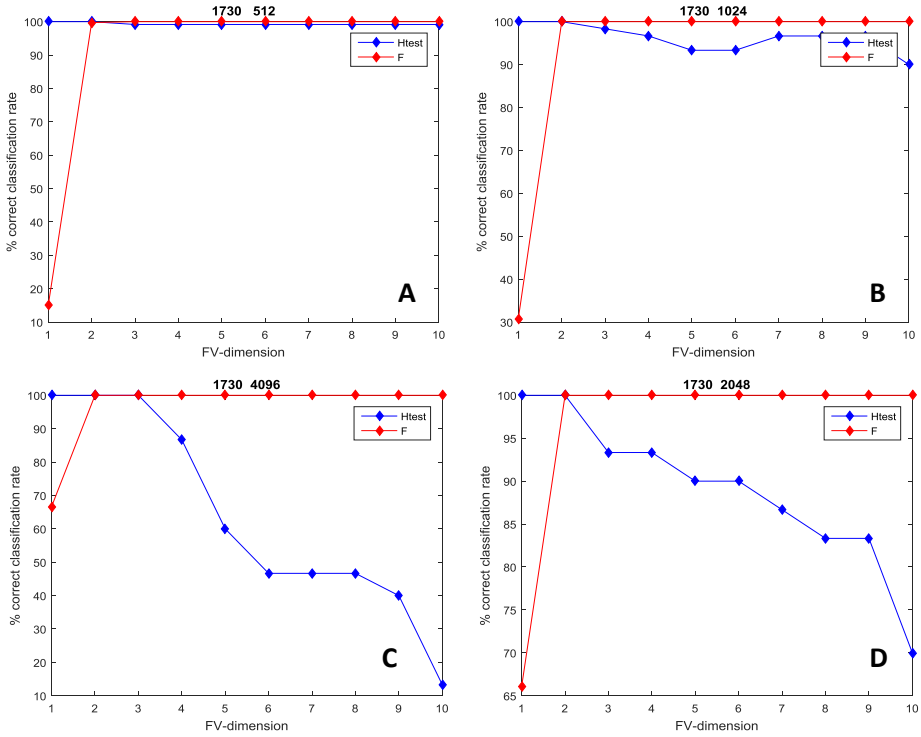


Figure 5.16: Average correct classification rate versus FV-dimension CS 1-different fault severities at inner race, 1730 RPM. RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

However, it can be seen clearly that the average correct classification rates of FVs corresponding to F category reach maximum when the baseline space dimension is two

and there is no need to use a higher space dimension. Furthermore, increasing the baseline space dimension can deteriorate on the average correct classification rates of the FV corresponding to H_{test} category.

Minimum baseline space dimension to achieve the highest classification rate.

Tables 5.4-5.6 show the minimum baseline space that gives best average correct classifications of FV of F and H_{test} categories. In each of these tables, the first row represents the sub-signal length which the FVs are made from. The first column represents the shaft rotational speeds. The values in the internal cells of the tables represent the following expressions

‘’**baseline space dimension** $_{minimum}$ [%average correct classification rate of H_{test} category %average correct classification rate of F category]’’

Generally, a number of notes can be seen for all the cases shown in the Tables 5.4-5.6. The minimum dimension of the baseline space decreases when sub-signal length increases. For example, in Table 5.4, the data corresponding to a speed 1772 RPM, the minimum baseline space dimension increase from 4 to 2 when the sub-signal length increases from 512 to 4096 data points. This might be interpreted as incorporation of more data points in the sub-signal (i.e increasing its length) means incorporation of more information about the fault in the sub-signal. The minimum dimension of the baseline space to achieve highest average correct classification rates is case dependent (it is not the same for all the cases).

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
1730	**3**[99.2 100]	**2**[100 100]	**2** [100 100]	**2** [100 100]
1750	**4**[97.5 100]	**3**[96.7 100]	**2** [100 100]	**2** [100 100]
1772	**4**[96.6 100]	**3**[95 100]	**2**[100 100]	**2**[100 100]
1797	**3**[100 100]	**3**[100 100]	**1**[100 100]	**1**[100 100]

Table 5.4: Minimum dimension of the baseline space , average correct classification rate of H_{test} category and average correct classification rates of F category of different IRF severity category FVs. CS1

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
1730	**5**[99.2 100]	**3**[98.3 100]	**2**[100 100]	**2**[100 100]
1750	**5**[96.6 100]	**5**[96.7 100]	**4**[80 100]	**4**[53.3 100]
1772	**5**[96.6 100]	**4**[95 100]	**4**[80 100]	**3**[86.7 100]
1797	**4**[100 100]	**4**[100 100]	**2**[100 100]	**2**[100 100]

Table 5.5: : Minimum dimension of the baseline space , average correct classification rate of H_{test} category and average correct classification rates of F category of different BF severity category FVs. CS1

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
1730	**3**[99.2 100]	**3**[98.3 100]	**2**[100 100]	**2**[100 100]
1750	**3**[97.5 100]	**3**[96.7 100]	**2**[100 100]	**2**[100 100]
1772	**3**[96.7 100]	**2**[100 100]	**2**[100 100]	**2**[100 100]
1797	**3**[100 100]	**2**[100 100]	**2**[100 100]	**1**[100 100]

Table 5.6: : Minimum dimension of the baseline space , average correct classification rate of H_{test} category and average correct classification rates of F category of different ORF severity category FVs. CS1

The Tables 5.7 and 5.8 show the minimum dimension of baseline space that gives best average correct classifications of FV of F and H_{test} categories corresponding to CS3. The same notes obtained for CS1 can also be seen for this case study where the

minimum dimension of the baseline space decreases when the sub-signal length increases.

	sub-signal length			
RPM	**512**	**1024**	**2048**	**4096**
18000	**7**[99.4 93.7]	**4**[99.5 100]	**4**[99.5 100]	**3**[100 100]
24000	**5**[99.6 99]	**5**[99.3 100]	**3**[99.5 100]	**1**[100 100]
30000	**10**[98.8 98.7]	**7**[98.3 100]	**5**[99 100]	**2**[100 100]

Table 5.7: Minimum dimension of the baseline space , average correct classification rate of H_{test} category and average correct classification rates of F category of different IRF severity category FVs. CS3

	sub-signal length			
RPM	**512**	**1024**	**2048**	**4096**
18000	**9**[99.1 98.8]	**5**[99.3 100]	**4**[99.5 100]	**3**[100 100]
24000	**10** [99 92.4]	**5**[99.3 100]	**4**[99.5 100]	**3**[100 100]
30000	**8**[99.3 79.6]	**7**[98.3 99.3]	**6**[97.5 100]	**4**[99 100]

Table 5.8: Minimum dimension of the baseline space , average correct classification rate of H_{test} category and average correct classification rates of F category of different BF severity category FVs. CS3

A study of FV dimension and sub signal length effect on detection accuracy rate of identification of different fault severities

In the previous section of the results, it is shown that the methodology has a very good performance in distinguishing the FVs corresponding to H and different fault severities into a baseline and non-baseline categories. The current section of results discusses the method performance in distinguishing not only to a baseline and non-baseline categories but among the non-baseline categories as well. As explained in section 4.3.4 the FVs are divided into a training sample and testing sample. The training sample is used to build the feature matrices corresponding to the healthy and the different fault severities such as small, medium and large inner race fault. The testing sample is used to validate the

method performance. The method performance is investigated in terms of how many FVs are assigned to their actual categories. The method was applied using FVs obtained at different sub-signal length, different rotational speed and different baseline space dimensions.

Figure 5.17 **A**-**D** illustrates the average and lowest correct classification rates for a data sets corresponding to different fault severities at inner race from CS1 (i.e CW5 in Table 4.2). In Figure 5.17**A**&**B**), it can be seen that selecting a two dimensional baseline space was sufficient to achieve the maximum average correct classification rate and there is no need to increase the baseline space dimension. In Figure 5.17-**A**, the FVs were obtained from the projections corresponding to the sub-signals of a length 512 data points. When only the projection on the $\mathbf{U_1}$ is considered, the average correct classification rate was 89.3%. However, when the projections on the $\mathbf{U_1}$ and $\mathbf{U_2}$ are also included, the average correct classification is improved to be 100%. The same note can be seen for Figure 5.17-**B**, where using projections on $\mathbf{U}_1$ and $\mathbf{U}_2$ improve the average correct classifications form 95.4% (i.e when using projections on the $\mathbf{U}_1$ only) to 98.3% (i.e when using projections on the $\mathbf{U}_1$ and $\mathbf{U}_2$ together). In Figures 5.17 **C**&**D**, it can be seen that using FVs obtained from the projections on the $\mathbf{U}_1$ was sufficient to achieve 100% classification where all the testing FVs are classified correctly to their actual category.

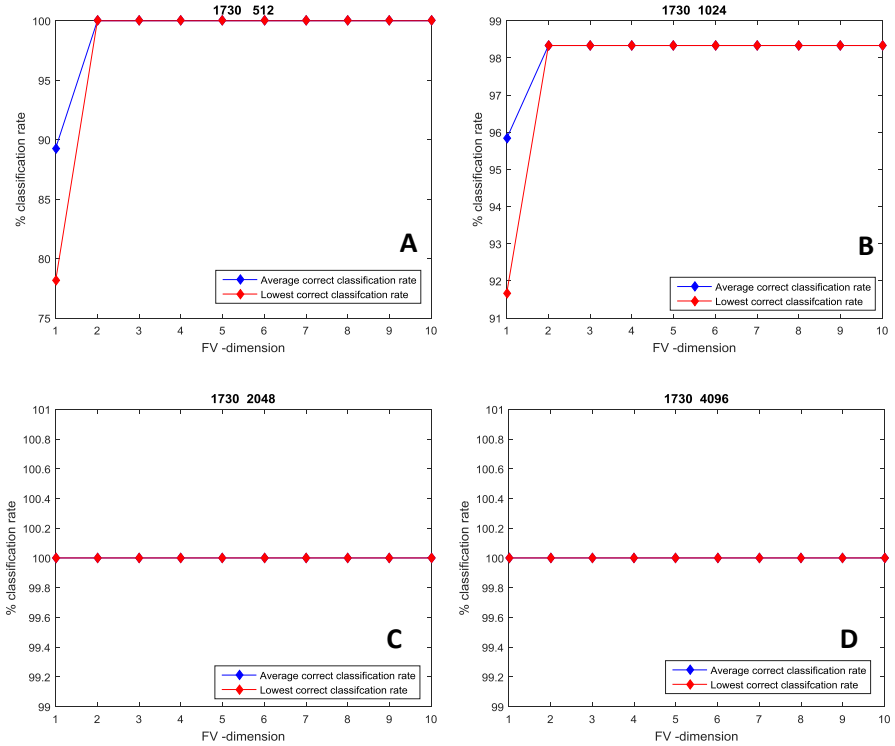


Figure 5.17: Average and lowest correct classification rate versus FV-dimension CS1,CW5 1730 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

Figure 5.18 **A**-**D** represents the average and lowest correct classification rates for the data sets corresponding to healthy and different fault severities at a ball element (i.e CW9 in Table 4.2). As was seen before the correct classification rates improve with the increase of baseline space dimension. For example, for a sub-signal of 512 data points (i.e Figure 5.18 **A**), the average correct classification rates improves from 62.4% to 100% when the baseline space dimension increased from 1D to 3D.

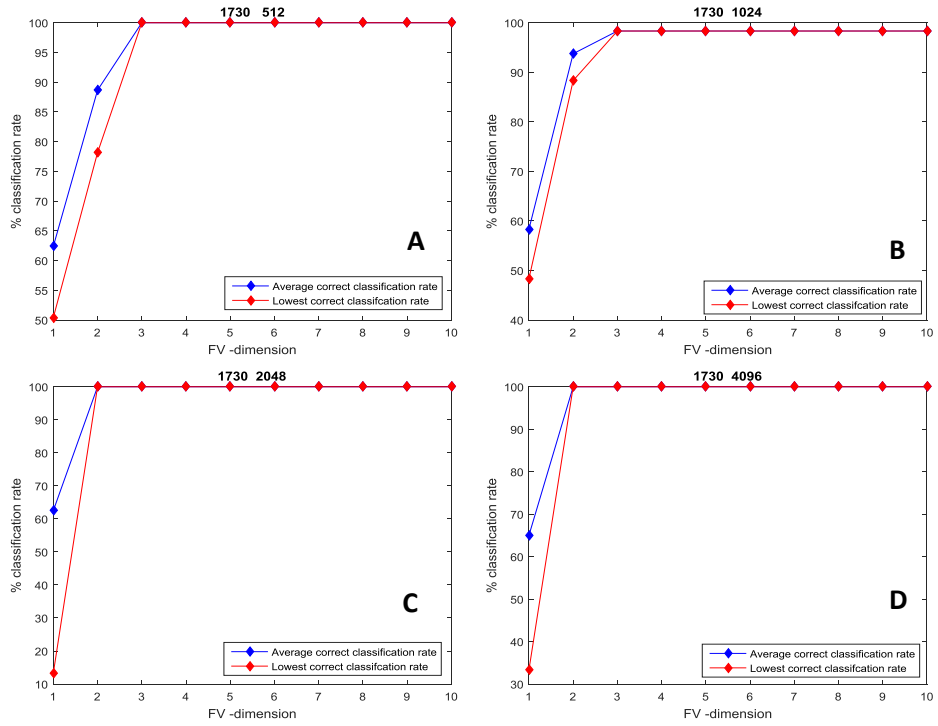


Figure 5.18: Average and lowest correct classification rate versus FV-dimension CS1, CW9 1730 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

Figure 5.19 **A**-**D** represents the average and lowest correct classification rates for the data sets corresponding to healthy and different fault severities at outer race (i.e CW13 in Table 4.2). As was seen before that the correct classification rates improve with the increase of baseline space dimension. For example, for a sub-signal of 512 data points (i.e Figure 5.19 **A**), the average correct classification rates improves from 85.8% to 98.8% when the baseline space dimension increased from 1D to 3D.

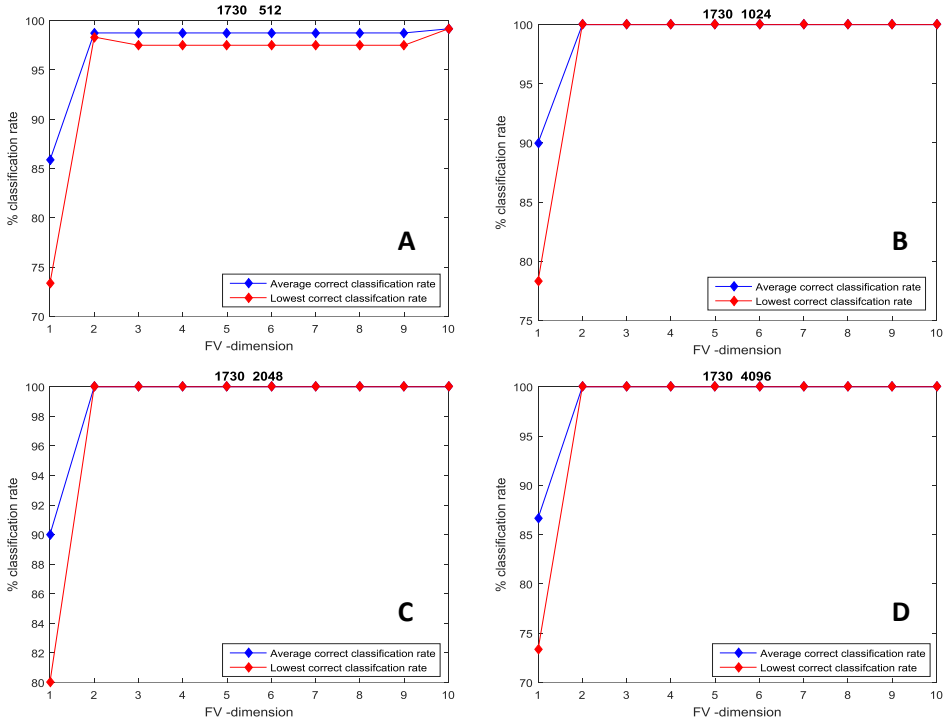


Figure 5.19: Average and lowest correct classification rate versus FV-dimension CS1, CW13 1730 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

Figure 5.20 **A**-**D** and Figure 5.21 **A**-**D** represent the average and lowest correct classification rates for the data sets corresponding to PT5 and PT8 of CS3 (see Table 4.2). Figure 5.20 shows the average and lowest correct classifications of FVs corresponding to H and different inner race fault severities of CS3. The minimum dimension of baseline space decreases with the increase of the sub-signal length. Furthermore, the correct classification rates improve when the dimension of the baseline space increased. This improvement is clearly seen when incorporating few first PCs while there is none or slight improvement when more of the last PCs are incorporated. Similar notes can be seen for correct classification rates for the FVs corresponding to healthy and different roller element fault severities categories (CS3) as in Figure 5.21.

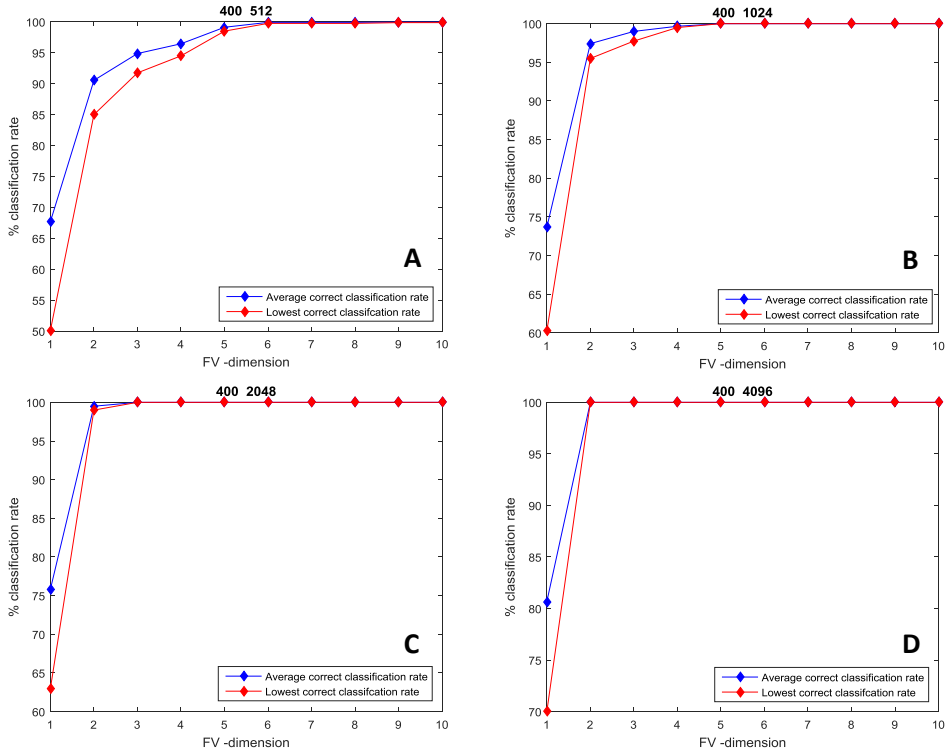


Figure 5.20: Average and lowest correct classification rate versus FV-dimension CS3, PT5, 24000 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

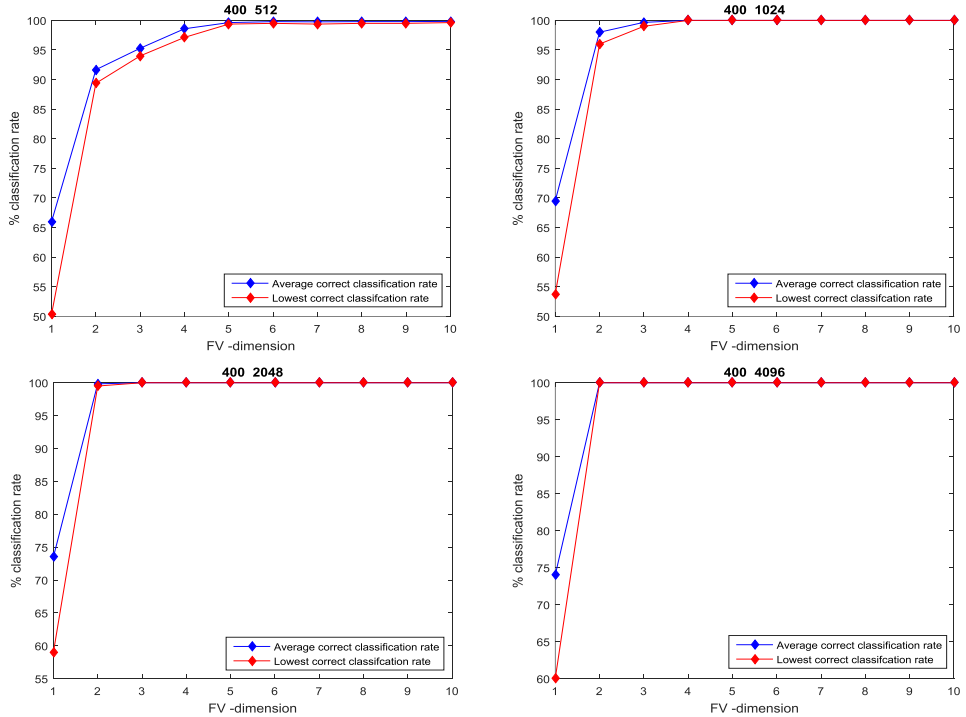


Figure 5.21: Average and lowest correct classification rate versus FV-dimension CS3, PT8,24000 RPM, **A**-512, **B**-1024, **C**-2048 & **D**-4096

Minimum baseline space dimension to achieve the highest classification rate.

Recalling Figures 5.17-5.19 above, which are showing the effect of FV dimension (i.e baseline space dimension) on the classification rates of a baseline and a different fault severities categories., It can be seen that minimum baseline space dimension to achieve the highest classification rates of FV corresponding to fault bearing category is also a case dependent (i.e it is not a unique for all the cases). For example in Figure 5.17-**A**, using 2D baseline space is sufficient to achieve 100% of fault estimation rate for sub-signal length of 512 data points. When a longer sub-signal is used (such as 2048, 4096 in the Figures 5.17 **C**, **D** respectively), the minimum baseline space dimension to achieve 100% correct fault estimation rate is only one. This can be interpreted as more data points in the sub-signal means more information about the fault might be included in the sub-signal and consequently more information about the fault will be involved in the first PCs.

Table 5.9-5.11 show the minimum baseline space dimension required to achieve maximum fault detection rates for the data sets of CS1 for different speeds and different sub-signal lengths. The correct classification rates are shown in the form ‘’ **minimum baseline space dimension** [%average correct classification rate, % lowest correct classification rate]’’. The numbers out of the brackets represents the minimum baseline space dimension required to achieve the maximum average correct fault estimation rates, while the numbers inside the square brackets refer to average correct fault estimation rate and lowest correct fault estimation rate respectively. Generally, it can be also seen that the minimum baseline space dimension decreases when the sub-signals length, which used to obtain the FVs, increases.

	sub-signal length			
CRWU	**512**	**1024**	**2048**	**4096**
1730	**2**[100 100]	**2**[98.3 98.3]	**1**[100 100]	**1**[100 100]
1750	**2**[100 100]	**2**[98.3 98.3]	**2**[100 100]	**1**[100 100]
1772	**3**[100 100]	**2**[98.3 98.3]	**2**[100 100]	**2**[100 100]
1797	**3**[99.299.2]	**2**[98.3 98.3]	**2**[100 100]	**1**[100 100]

Table 5.9 : Minimum baseline space dimension, average correct fault estimation rates of FVs corresponding to H and different IRF severities CS1

	sub-signal length			
CRWU	**512**	**1024**	**2048**	**4096**
1730	**3**[100 100]	**3**[98.3 98.3]	**2**[100 100]	**2**[100 100]
1750	**5**[100 100]	**5**[98.3 98.3]	**5**[100 100]	**2**[100 100]
1772	**7**[100 100]	**8**[98.3 98.3]	**6**[100 100]	**3**[100 100]
1797	**9**[99.4 97.5]	**8**[98.3 98.3]	**7**[100 100]	**4**[100 100]

Table 5.10 : Minimum baseline space dimension, average correct fault estimation rates of FVs corresponding to H and different BF severities CS1

	sub-signal length			
CRWU	**512**	**1024**	**2048**	**4096**
1730	**2**[98.8 98.3]	**2**[100 100]	**2**[100 100]	**2**[100 100]
1750	**2**[100 100]	**2**[98.3 98.3]	**2**[100 100]	**2**[100 100]
1772	**2**[99.2 99.2]	**2**[100 100]	**2**[100 100]	**2**[100 100]
1797	**3**[100 100]	**1**[98.3 98.3]	**1**[100 100]	**1**[100 100]

Table 5.11 : Minimum baseline space dimension, average correct fault estimation rates of FVs corresponding to H and different ORF severities CS1

Fault severity index

In this research, the use of the Mahalanobis distance (*D*) of testing FVs measured to the baseline space as a fault severity index is investigated. Figure 5.22 show the *D* values of the FVs corresponding to healthy and different fault severities at inner raceway (CS1) measured to the baseline space. The dimensions of the baseline space are selected from Table 5.9. It can be seen that fault category's FV are correctly distinguished not only

from the baseline category but among themselves. Although *D* changes with the fault severity changes, it does not always follow a monotonic function.

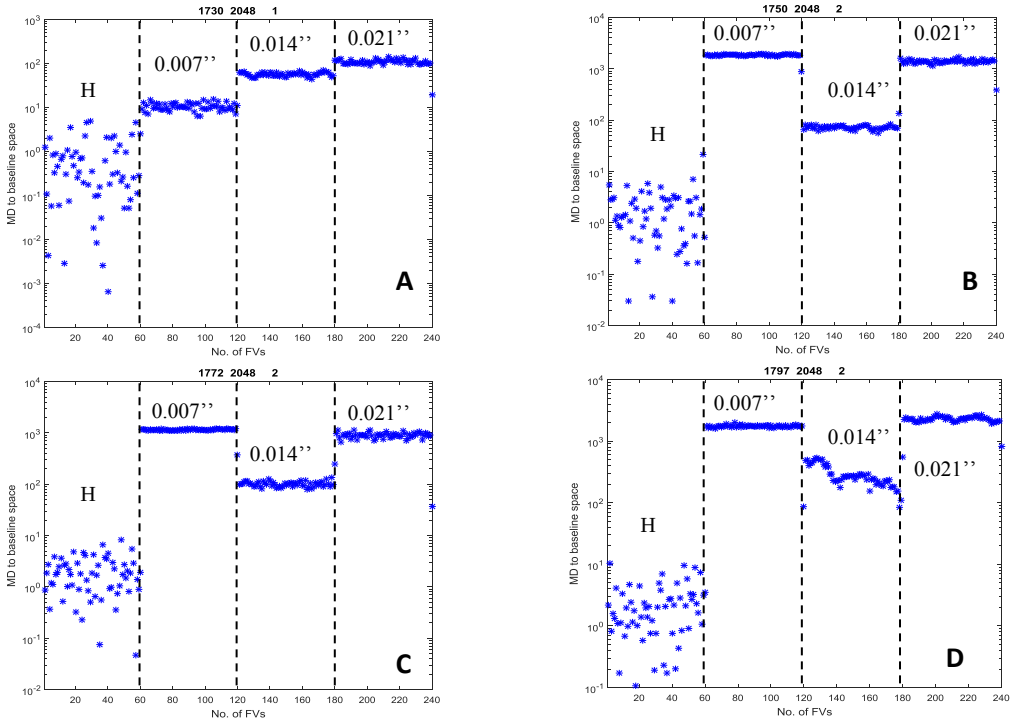


Figure 5.22: The Mahalanobis distance (*D*) of FVs corresponding to H and different inner race fault severities categories measured to baseline conditions sub-signal length of 2048 data points – CS1. **A**-1730 RPM. **B.** 1750RPM **C.** 1772RPM **D.** 1797RPM

In Figure 5.23, shows the *D* of the FVs corresponding to H and different inner race fault severities categories for the data set of CS3 of sub-signal length 2048 data points. It can be seen that the *D* is generally following a monotonic function as it increases with the increase of the fault severity, especially for the data sets obtained at 18000 & 30000 RPM. However, at speed 24000 RPM the *D* level changes when the fault severity becomes large.

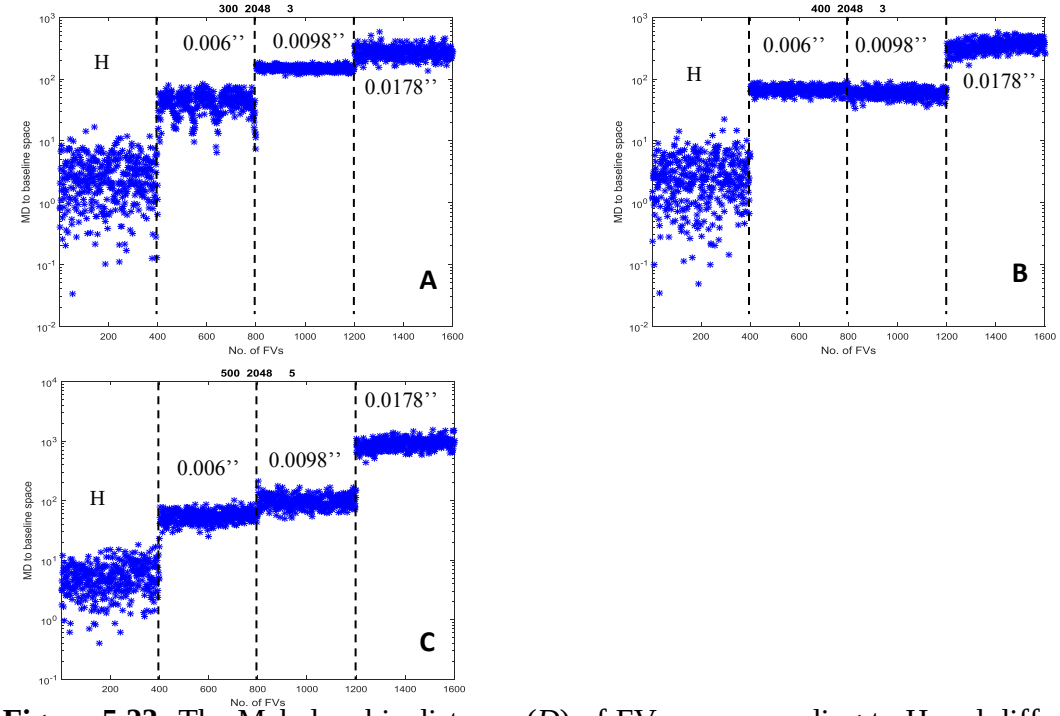


Figure 5.23: The Mahalanobis distance (*D*) of FVs corresponding to H and different inner race fault severities categories measured to baseline conditions sub-signal length of 2048 data points – CS3. **A**-18000 RPM. **B.** 24000 RPM **C.** 30000 RPM**.**

5.2 Results comparison to some recent published work

Table 5.12 shows the precision of the present method as compared to some other recent methods which are based on different time series analysis techniques but they use the same part of datasets of CWRU. Information regarding the datasets details, number of testing and training FVs and average correct classification rates are presented in Table 5.12 for these methods and for the method suggested here.

The methods are listed below:

1) Difference histograms (DHs) and feed forward neural network (FFNN)[100]: The difference histograms (DHs) based method includes the formation of scaled

representation for histograms of increased sub-signal lengths. Some of the first histogram bins are used as inputs to a FFNN for classification purpose of faults.

2) SSA and back propagation neural network (BPNN) [83]:In this study two sets of feature vectors are developed using SSA. The first FV includes the singular values of some of the first several principal components and the other FV uses the energy of the time domain of sub-signal components corresponding to these principal components. These feature vectors are used as input to the BPNN classifier.

3) In [101], signals from two different accelerometers are used to create a two dimensional representation of the bearing condition. The minimum volume ellipsoid (MVE) method is used to extract the features. Principal component analysis (PCA) is used for selecting the most important features. The last step is to input the selected features to the nonlinear nearest neighbour classifier.

It can be seen from Table 5.12 that all the compared methods achieve a rather good classification rate. From all the four methods compared it can be seen that the method suggested here demonstrates the best classification rate, which is between 98.3-100%

Method	Data set	Training and testing FVs	Conditions classified	Average correct classification rates (%)	Fault features
1)DH and FFNN[95]	0.18, 0.36,0.53mm; 0–3 HP load;30000 data points	Both 144	(IF, OF and BF classification only)	92-95	First 6 histogram bins
2)SSA and BPNN [65]	0.18, 0.36, 0.53, 0.71 mm; 0– 3 HP load;6100 data points	Train—336 test—144	H, IF,OF and BF	96.53–100 95–100	4 singular values 3 energy features
3)MVE, PCA and nonlinear neighbour classifier [96]	0.18, 0.36,0.53mm , 0-3 HP, 2000 data points	-	Healthy and faulty. It was reported by the author that there is some	94.68-99.98	-

			overlapping among some fault classes		
4)The methodology presented in his chapter 4	0.18, 0.36,0.53 mm; 0-3 HP loads; 2048 data points	Train-Test-	H, IF, OF and BF	98.3-100	Norms of the 1st three projections on to the baseline space

Table 5.12: A comparison of the performance of the present methodology with other published work using the same parts of CWRU data

5.3 Conclusions and discussion

This chapter suggests a rather simple and easy to apply but accurate method for fault diagnosis in rolling element bearings based on singular spectrum analysis. The simplicity of the method can be explained in terms of its application as it uses solely the signals from the healthy bearing state to build the so-called baseline space. Furthermore, it uses only the decomposition stage of the SSA. It is applied to investigate the capabilities of the SSA for different fault diagnosis phases namely the detection of fault presence, the identification of fault type and the estimation of fault severity in REBs.

In all the levels of fault diagnosis, the baseline space is made from subjecting only the measurements from the healthy bearing category to the decomposition stage, and all other signals are projected on to it. In such a sense the transformations applied to the measured signals are minimal and very simple. Then, features are made from the Euclidean norm of these projections. As illustrated in Figures 4.5 and 4.6, using the Euclidean norm of the projections improves the distinguishability of FVs corresponding to different categories.

For fault detection phase, the method does not require any previous measurements corresponding to faulty/anomalous conditions. A threshold based on the statistical hypothesis was set for this level of fault diagnosis. The assumptions made for determining the threshold have been questioned. The minimizing of positive and negative misclassification was taken in consideration when the threshold was made. The classification rule, which is based on a threshold of the Mahalanobis distance, is also a robust and simple one. As a result of these points, the methodology holds considerable

potential for automatization and practical implementation. The results obtained for this level of fault diagnosis show that the method is able to distinguish not only between baseline and non-baseline categories but also amongst non-baseline signal categories themselves. Since only the signals corresponding to healthy signals are used for building the baseline space, and none of the fault signals are used as a training sample, this level of fault diagnosis is considered to be a form of unsupervised learning.

For fault type identification, the method shows a very good performance in distinguishing among different fault location signal categories. As in the fault detection phase, the baseline reference is build using a number of signals from training samples of healthy bearing categories. The other training sample signals which correspond to the other categories are all projected on the baseline space and a number of feature matrices are made from these projections as was explained in Section 4.3.3. The classification rule is based on the minimum Mahalanobis distance which a new FV from testing sample has with one of the feature matrices.

For fault severity estimation, the signals corresponding to fault categories of different severities were projected on to the baseline space which was made from the healthy bearing category. The FVs were made from these projections and the Mahalanobis distance of these FVs to the baseline space is measured and used as indicator for fault size change.

The directions of the principal components (i.e eigenvectors) are used to build the reference space. And the classification/fault diagnosis is done on the basis of projecting the data on these directions. In such a sense it can be argued that the PCs obviously contain information regarding the state of the bearing. The effect of using more

directions (i.e more eigenvectors), in building the baseline space, on the performance of the method in terms of classification accuracy is also investigated.

In machine learning it is preferable and easier to work with fewer features (i.e short FV lengths). In the current study, and for all levels of fault diagnosis, a clear phenomenon is seen in the FV dimensionality (i.e the number of PCs) with respect to classification accuracy. It is noticed generally that increasing FV length affects positively the correct classification rates. This can be interpreted that an increase in FV length means the incorporation of more PCs, and then more information about the fault present is included.

It should be noted as well that the methodology is rather general and can be applied to any measured signals regardless of their stationarity. The 3D visualisation showed that not only baseline and non-baseline signal categories can be distinguished but in some cases the method can be used to separate different faulty categories. This is still a subject of further research and in such a sense the interpretation of the PCs/the principal directions will be very helpful.

The Mahalanobis distances of different fault severities, which are measured to the baseline category, are also used as a fault index. Although the Mahalanobis distance level changes when fault size changes, sometimes (as in CS1) its trend does not follow a monotonic path. In fact, the reasons behind the non-monotonic path followed by the data of CS1, are not investigated in the current study. This might be explained by the fact that the feature selected are not always capable of behaving linearly with the

change of fault severity. A further future study might help in further understanding the physical meaning of the features.

Based on the obtained results, it has been shown for the case studies considered that SSA in the way that it is used in this chapter, is capable of extracting essential information regarding the fault detection, fault type identification and fault severity estimation.

Although there were not any opportunities to validate the method using data sets from a real machine in service, the author is confident that it also would perform well in more challenging scenarios, and below are some reasons for this assertion: -

- First of all, the method suggested is one which is purely based on data/signal analysis. SSA and the method developed in this thesis use a sample of signals measured for the healthy condition in order to build the space corresponding to the healthy bearings signals. Thus this space should incorporate changes stemming from noise or other deviations within the healthy signals sample. And in effect the method will detect the changes due to the bearing fault rather than these deviations.

- It should also be mentioned that structure-less noise usually affects the higher order PCs [26, 74], while the method only uses the first few PCs.

- SSA has been extensively used for analysis of biomedical signals and climatological data where the measurements come from different categories such as seasonal data. In all such cases one is dealing with real data which on most occasions will be noisy and error-contaminated, but SSA has the capability to distinguish between data

obtained from the different categories. This is because for the case of two or more categories of observations, the distances amongst the PCs from the same category tends to decrease, whereas the distances among the PCs from different categories tend to increase [102, 103].

- The author also would like to stress the fact that the present method has a good capability for being used for fault detection in complex systems because it simplifies complex signals by projecting them onto the higher dimensional baseline space, and thus unfolds the information contained in these signals. So, assessing the similarity of these new projections to the projections obtained from baseline conditions can give us an indication of the presence of a defect when these projections are different to the baseline one.

The author would like to clarify that the current SSA-based methodology (i.e chapter 4) has been able to achieve 100% correct classification by using only the first PC. Thus it is noticed that in comparison to other SSA-based techniques, as discussed in section 3.4.1, the way that SSA is applied in the current methodology reduces the necessary number of PCs required for achieving maximum correct classification.

5.4 Summary

In this chapter the results for the methodology, presented in chapter 4, are demonstrated. The results include two main parts 1) demonstration of the methodology using data from CS1 and 2) fault diagnosis. In the first part, the building of a baseline space from healthy signal category, projections of other signals onto the baseline space are

demonstrated. In the second part, the fault diagnosis results corresponding to fault detection, fault type identification and fault severity estimation are presented. The methodology performance is compared to some other recent published methodologies and it is shown that the present methodology is superior to those methodologies.

Chapter 6

Bearing fault diagnosis using a novel signal pretreatment and LTIVAR modelling

6.1 Introduction

As was introduced in section 2.4.2, a novel signal pretreatment based on singular spectrum analysis and differencing techniques is suggested to facilitate the representation of bearing vibration signals using a LTIVAR model. In this chapter the steps and mathematical formulas are presented.

The methodology presented in this chapter has two main stages: -

1. signal pretreatment.
2. signal diagnosis.

In the first stage, the non-stationary bearing vibration signal is firstly transformed to a stationary one through several steps and then subjected to the modelling with the linear time invariant autoregressive model to obtain the model coefficients.

In the second stage, the coefficients of the LTIVAR model are used to make the feature vectors. These feature vectors are used to form feature matrices that belongs to the different signal classes used in the analysis. The fault diagnosis process is achieved by

using the pattern recognition approach. This contains the classification of any new feature vector using the nearest neighbour (NN) rule based on the Mahalanobis distance.

6.2 Signal pretreatment

6.2.1 Signal Segmentation

The acquired relatively long signals are divided into a number of equal-length and non-overlapping sub-signals (i.e segments). This segmentation is useful into providing more signal samples and saving time and overcoming the difficulty of repeating measurements from a machine. The length of the sub-signal can be estimated based on the type of analysis that will be done on the sub-signals. For example, it is recommended in [104] that when autoregressive modelling is used, the sub-signal length should follow the ratio $p/n < 0.1$, where p is the model order and n is the sub-signal length.

6.2.2 SSA-based noise cancellation

As was mentioned in section 4.2, the SSA has not yet obtained popularity for fault machinery diagnosis. When a lagged version of the original sub-signal (see Eq. (4.1)) is subjected to the SSA, a number of independent principal components (PCs) can be obtained following Eq. (4.2) to Eq. (4.4). Then the initial sub-signal can be then reconstructed using a number of these PCs and by using the diagonal averaging technique based on Eq. (4.5). In this method, the SSA is used to clean the sub-signal from the structure-less noise. This noise cancellation is applicable as the structure-less noise can be transformed through the SSA steps into low singular value components

[26, 74]. That means only the PCs that explain a large portion of original sub-signal variance will be selected in the reconstruction stage. Such kind of noise cancellation resulted in a considerable improvement in the LTIVAR model accuracy of sub-signal representation. The details of SSA stages (decomposition and reconstruction were explained in details into Chapter 4-section 4.2).

6.2.3 Stationarization

The definition of a stationary signal is basically that the consistency of the first four statistical moments of a signal over the time. Conversely, non stationarity occurs when these statistical moments are changing with time. Nevertheless, the definition above is used for the so-called ‘’strict stationarity’’ and it is difficult to achieve a signal in which its first four statistical moments are constant over the time. Thus, a second order or weak stationarity is usually meant by the term ‘’stationary’’. And it refers to the case when the first two statistical moments are constant over the time [105].

As was mentioned that LITVAR model is primarily developed for the stationary time series, it is necessary to bring the signals to stationarity before subjecting them to the model. There are several ways to transfer the non-stationary signals to stationary one such as finding their logs, square roots and differencing them. The suggested transformation in this thesis is to use the differencing technique as it is the simplest form of stationarisation. Although differencing brings the nonstationary signals to a weaker form of the stationary ones, it has been a sufficient contribution to the improvement of the goodness of fit of the LTIVAR. However, it should be taken in consideration that applying differencing several times can introduce considerable

amounts of high frequency noise. It is suggested firstly to check the sub-signal for stationarity and then apply the differencing to the nonstationary sub-signals just once. There are a couple of tests which have been developed in the literature for testing the stationarity of signals. In this research, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test [106] is used to test the stationarity of each sub-signal. In this test, the null hypothesis assumes that the sub-signal is stationary around a deterministic trend. If the sub-signal is recognised as non-stationary then it is subjected to differencing technique. The differencing technique be described using the following equation[105]:

$$xr^{dif}(m) = xr(m) - xr(m-1), \text{where } m = 2,3,\dots n \tag{6.1}$$

It is clear that the length of the new transformed sub-signal xr^{dif} is shorter than the original sub-length by one. The primary aim of differencing is to remove the stochastic trends responsible for the sub-signal non-stationarity (i.e. to stabilise the mean of a non-stationary sub- signal[107]. This is the simplest method that can be used for the purposes of stationarisation.

Generally, the number of differencing (i.e order of differencing) relates to the complexity of the non-stationary signals. If, for example, a signal lacks the tendency to return to its mean value, higher order of differencing might be applied. In general the differencing method is a rather simple but powerful process so on most occasions quite high stationarity is achieved through just one application. In most studies which apply differencing for other purposes not related to fault diagnosis and machinery dynamics the process is applied just once[108-110] .

In this work, one application of differencing was sufficient to bring the sub-signals to stationary ones.

The differencing technique has been used for purposes of stationarisation in climate research [111]; to the knowledge of the authors, however, it has not been used in the field of machinery fault diagnosis.

6.2.4 Linear time invariant autoregressive (LTIVAR) modelling

Once the stationarity of a sub-signal is achieved, the sub-signal is subjected to LITVAR modelling. The mathematical structure of the LITVAR is given Eq.(6.2)[112].

$$xr(m) = a_o + \sum_{m=1}^{p} a_m \, . \, xr(m-i) + \varepsilon(m) \qquad (6.2)$$

where

$xr(m)$ is the predicted value of a sub-signal at time m ;

p is the order of the LTIVAR model;

a_m ($m = 0,1,2\ldots,p$) are the LITVAR model coefficients;

$\varepsilon(m)$ is an error term which represents the difference between the real and predicted the m value of a sub signal

There are a number of algorithms to estimate the LTIVAR model coefficients detailed in [112]. In this study, the least square algorithm is used. When the coefficients of the

LTIVAR model are estimated, it is necessary to check the accuracy of the model in representing the sub-signal (i.e. goodness of fit). The accuracy of the LTIVAR model in representing the real sub-signal can be evaluated by the normalised mean square error. There are several way to measure the goodness of fit of a model such normalised mean square error which given by the equation (6.3).

$$Goodness\ of\ Fit|_{NMSE} = \left(1 - \frac{\sqrt{\sum_{i=1}^{i=n}\left(xr_{pr}(i) - xr_{ms}(i)\right)^2}}{\sum_{i=1}^{i=n} xr_{pr}(i) - \overline{xr}_{pr}}\right) * 100\% \qquad (6.3)$$

where:

xr_{pr} is predicted stationary sub-signal;

xr_{ms} is the stationary real time measured sub-signal;

$\overline{xr}_{ms}$ is the mean value of xr_{ms};

n is the sub-signal length.

The calculations for the least square algorithm and the NMSE values mentioned above were carried out using MATLAB.

6.2.5 Model order selection

As was illustrated in Eq (6.2) in the LTIVAR model each point is linearly related to a number of past points (i.e model order). The model order is not selected arbitrarily because using improper model order can affect negatively on the model goodness of fit

and may lead to the problems of either under or over fitting [89, 113]**.** Thus, a proper selection of model order can considerably reduce the consequences of the over-fitting or under-fitting. There are a number of approaches which are developed for estimating the optimum model order. These techniques are generally based on calculating a certain loss function value over the increase of model order and then to estimate the optimum order which corresponds to the minimum error criterion function. These error criterion functions are based mainly on the calculations of the sum of squared differences between the predicted and measured data. In this study, the Akaike's final prediction error (FPE) approach is used[114]. This FPE approach is given in the equation (6.4) below:

$$FPE = V.(1 + p/n)/(1 - p/n) \tag{6.4}$$

where

V is the variance of the difference between the predicted and the measured sub-signal;

According to this approach, the optimum model order is defined as that the model order corresponding to the minimum FPE.

6.2.6 Dealing with different model orders

It was mentioned in section 6.2.5 that the optimum model orders is that order corresponding to the minimum information criterion (i.e in this study it is the FPE). It is worth mentioning that all the criteria of model order estimation suggest that the optimum model order corresponds to the minimum value of a loss function. However, it

is not always easy to find a clear minimum value of the appropriate loss function. Therefore, a criterion is used for guidance only and the model order is estimated. In some cases, different optimum model orders might be estimated for different signals and signal classes. In the current study, different optimum model orders means different FV lengths. The equalising of FV lengths is necessary when a NN rule-based classifier (i.e one more specifically based on the Mahalanobis distance) is used. This is because both the training and testing FVs must have the same length when such a classifier is used. So, the equalisation of FV lengths is applied if there is a difference in the optimum model order.

Four different possibilities for equalising FV lengths are investigated to see whether the way of equalising can affect the performance of the methodology. Assuming that the minimum and maximum optimum orders for several signals/categories are (p_{min}) and (p_{max}) respectively, the equalising possibilities are:-

Zero padding. One possibility for equalising different length FVs is to add zero elements to the short length FVs. That is, the FVs of a lower order (i.e $(l < p_{max})$ are transformed by adding $(p_{max} - l)$ zero elements at the ends (see Figure 6.1). Eventually all the feature vectors acquire the maximum length p_{max}..

$$[a_o\ a_1\ a_2\ \ldots\ a_l\ \overbrace{0\ 0\ 0\ 0}^{(p_{max}-l)\text{ elements}}]$$

$$[a_o\ a_1\ a_2\ \ldots\ \ldots\ a_{p_{max}}]$$

Figure 6.1: Illustration of zero padding.

Trimming. It is proposed to trim/cut the features vectors of length *l*, where $l > p_{min}$, to length (p_{min}). If $[a_o\ a_1\ a_2 \ldots\ldots a_l]$ is a longer feature vector (i.e. $l > p_{min}$), then its last $l-p_{min}$ components are removed and the new vector is $[a_o\ a_1\ a_2 \ldots\ldots a_{Pmin}]$ (see Figure 6.2).

$$[a_o\ a_1\ a_2\ \ \ldots\ \ a_{l-3}\ \overbrace{a_{l-2}\ a_{l-1}\ a_l}^{cut\ (l\text{-}p_{min})\ \text{elements}}$$

$$[a_o\ a_1\ a_2 \ldots\ \ a_{p_{min}}]$$

Figure 6.2: Illustration of trimming.

Maximizing. It is proposed to expand the shorter feature vectors up to the maximum one. This time is not by adding zeros but by increasing the smaller model order up to the largest one. Eventually, all the feature vectors will have the same lengths.

Variance Threshold Method: In this method, it is proposed to select the model coefficients which have the highest variance contribution among the feature vectors. This can be done using the variance threshold method (VTM), which proposes to select only the model coefficients with a variance higher than the mean value of all the features vectors variances.

In this research, all the four possibilities above were applied during the detection and the fault identification stage. The best one in terms of highest correct classification rate was selected and used further for the next stage of fault severity estimation.

6.3 Fault diagnosis method

In this section, the method of fault diagnosis is explained. The method is divided into two parts; construction of feature vectors and diagnosis methods.

6.3.1 Construction of feature vectors

When the coefficients of the LTIVAR model are obtained as described in section 6.2.5, they are ordered to make the feature vectors. The different lengths of the feature vectors are equalised by one of the techniques mentioned in section 6.2.6. Thus a feature vector of a sub-signal is formed from the LTIVAR model coefficients corresponding to this sub-signal.

6.3.2 NN rule based fault diagnosis

The fault diagnosis presented here is implemented into two stages 1) fault detection and identification and 2) fault severity estimation. In the first stage, a feature vector (i.e a sub-signal) is assigned to a healthy bearing class or to a bearing with a specific located fault class. In the second stage, the FV which is assigned to one of the faulty classes in the first stage is further assigned to one of the different fault severity levels. At first the FVs are divided into training and testing samples. The training FVs are used to build the feature matrices which defines the classes in each fault diagnosis stages. The testing FVs are used as inputs for the classification process. The diagnosis of the severity condition can be considered as a classification problem or a regression problem. It might be desirable to use a regression model, however this might complicate the methodology as it requires determination of the coefficients of the regressor. In the

current study we dealt with the quantification of the bearing fault severity as a classification problem assuming that a faulty bearing signal can be assigned to one of fault severity categories (i.e small, medium and large).

Building feature matrices

As was mentioned in the section above the FVs were used to build the feature matrices and to define the classes for each fault diagnosis stages. For the detection and identification of fault stage, K_a feature matrices corresponding to healthy bearings (H), Inner race fault condition (IRF), ball bearing fault condition (BF) and Outer race fault condition (ORF) are made. For the fault estimation stage, K_b feature matrices are made corresponding to small fault class (S), medium fault class (M) and large fault class (L). For both fault diagnosis stages, the feature matricies are made from arranging the training FVs in rows as described in Eq. (6.5):

$$\mathbf{F}_{K_i} = \begin{bmatrix} a_{K_i11} & a_{K_i12} & . & a_{K_i1p} \\ a_{K_i21} & a_{K_i22} & . & a_{K_i2p} \\ . & . & . & . \\ a_{K_iN1} & a_{K_iN2} & . & a_{K_iNp} \end{bmatrix} \tag{6.5}$$

For each matrix, the number of rows (e.g N) equals the number of the training FVs corresponding to the *K* class while the number of the columns (*p*) equals the length of a training FV (i.e model optimum order) corresponding to the class (*K*).

where K_i corresponds to one of the categories in the diagnosis stages (i.e. fault identification $K_i \in K_a$ and quantification). $K_i \in K_b$

Nearest neighbour rule

In both fault diagnosis stages, the assignment of any new testing FV to one of the K_a or K_b classes is based on the use of NN rule[115] . According to this rule, the distance (e.g Mahalanobis distance) of any new testing FV is measured to each of the feature matrices (i.e $\mathbf{F}_{KI}$ in Eq. (6.5)). In this study, the Mahalanobis distance (D_{Ki}) is used to measure the distance between a testing FV to the matrix $\mathbf{F}_k$ as defined by the following Eq. (6.6) in below:

$$D_{K_i}(\mathbf{fv}, \mathbf{E}_{K_i}) = \sqrt{(\mathbf{fv} - \mathbf{E}_{K_i}).\mathbf{S}^{-1}.(\mathbf{fv} - \mathbf{E}_{K_i})'} \qquad (6.6)$$

where

D_{Ki} is Mahalanobis distance of $\mathbf{fv}$ to the feature matrix $\mathbf{F_{Ki}}$. ;

$\mathbf{fv}$ is a testing FV;

$\mathbf{S}^{-1}$ is the inverse of the covariance of the feature matrix $\mathbf{F_{Ki}}$;

The prime ($'$) denotes the transpose of the vector $(\mathbf{fv} - \mathbf{E}_{K_i})$.

The mean of the row feature vectors for each of the matrices $\mathbf{F}_{K_i}$ is calculated as follows:

$$E_{K_i} = \frac{\sum_{j=1}^{N}(a_{K_i j1} \ldots\ldots\ldots\ldots a_{K_i jp})}{N} \qquad (6.7)$$

The lengths of the$\left(\mathbf{fv}, \mathbf{E}_{K_i}\right)$, in which the Mahalanobis distance is measured in between, should be equalised as discussed in section 6.2.6.

After obtaining the Mahalanobis distances of the new testing FV to each of the feature matrices, the new vector is assigned to the class for the Di has a minimum all over the distances.

Figure 6.3 shows the stages of the diagnosis procedure. It is clear from the figure that in a case in which the new feature vector is assigned as healthy, it will not go further for the second stage of diagnosis.

Figure 6.3 shows the block diagram of the whole process of the methodology. There are three main blocks in the flow chart (1, 2 and 3), each bordered by a dashed rectangle.

- Block 1: the stages of signal pretreatment.
- Block 2: Fault detection and identification.
- Block 3: Fault severity estimation.

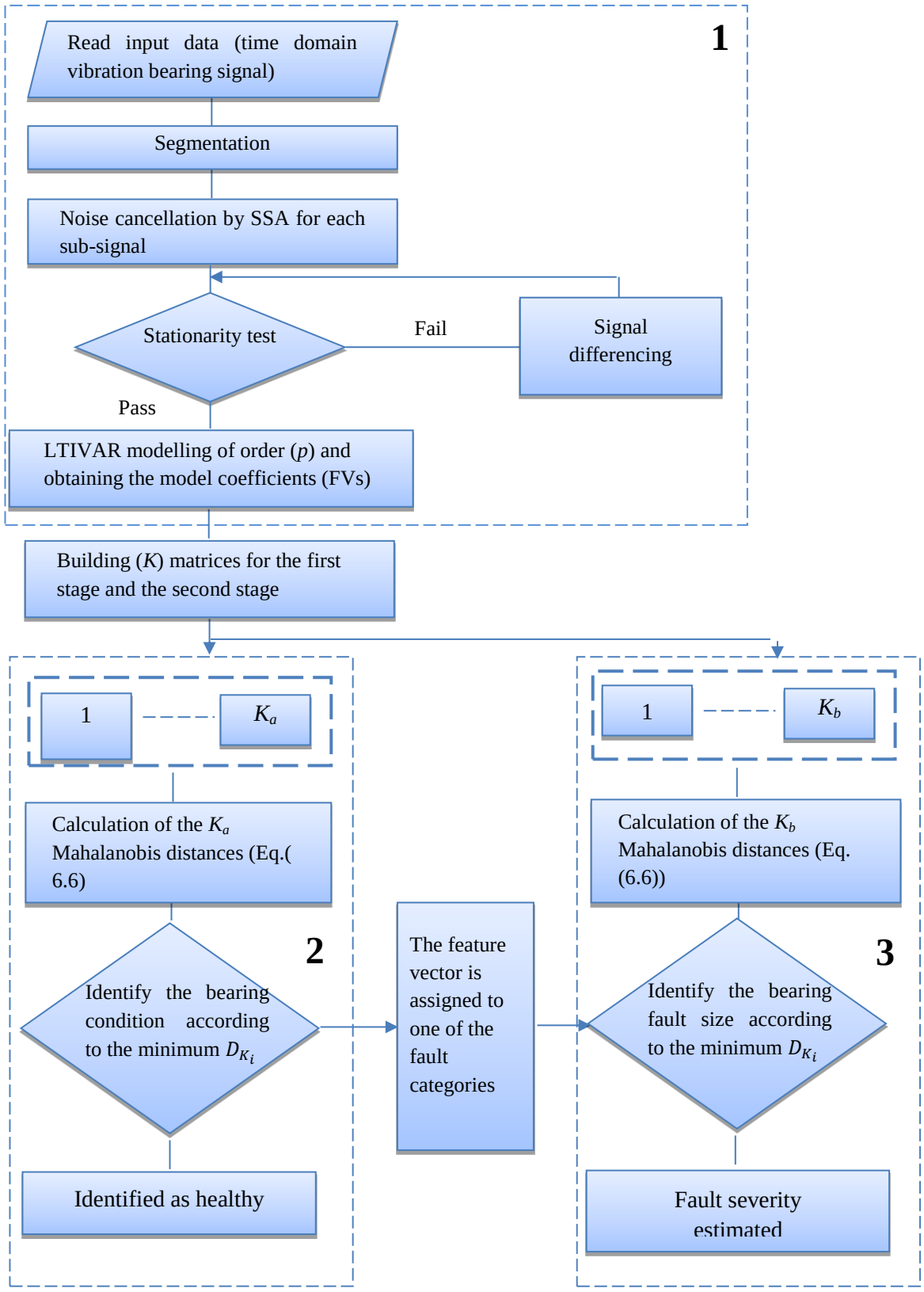


Figure 6.3: A block diagram illustrating the proposed classification method

The performance of the methodology was estimated by using the so-called confusion matrix described in section 4.3.2.

6.4 Summary

The method is based on developing an SSA and differencing –based advanced signal pretreatment technique that facilitate the use of the LTIVAR model for fault diagnosis in bearing. The method has two main stages:- signal pretareatment and signal diagnosis. The fundamentals and steps of the first main stage are introduced in section 6.2 while the fundamentals and steps of the second main stage are introduced in section 6.3. The method aims to provide an advanced signal pretreatment that denoise and stationaries the signal before subjecting them to LTIVAR modelling. This signal pretreatment will simplify the signal and improves the goodness of fit of the LTIVAR. Then, the coefficients of the LTIVAR are used as features in the classification using NN rule. The method was validated using three different data sets obtained from different bearing test rigs.

Chapter 7

Results and discussion- Bearing fault diagnosis using a signal pretreatment and LTIVAR model

7.1 Introduction

In the next sections, results will be presented for both methodology stages (i.e signal pretreatment and signal diagnosis). The first stage of the methodology is demonstrated using data sets obtained from case study 1. A couple of figures and tables are presented to demonstrate the steps of the signal pretreatment. For the second stage, results are presented for the averaged correct classification rates for both 1) the fault detection and identification and 2) the fault estimation using case study 1, 2 and 3. The effect of sub-signal lengths, equalisation of FV lengths is also discussed.

7.2 Part1- methodology demonstration based on CS1

7.2.1 Signal pretreatment

In this section, the illustration of each step of the signal pretreatment stage is introduced. The Figure 7.1 shows a bearing vibrational acceleration sub-signal, of length 2048 data points, acquired from the drive end bearing at different bearing conditions namely, H, IRF, BF and ORF for a rotational speed 1772 RPM. The sub-signals are shown in their time domain. The x-axis represents the data point number while the y-axis represents the acceleration amplitude in (m/sec^2). It can be seen that

for the sub-signals corresponding to the faulty raceways, there are several peaks related to the strikes of the rolling element with the raceways. However, these peaks may not be seen in some cases of noisy signals.

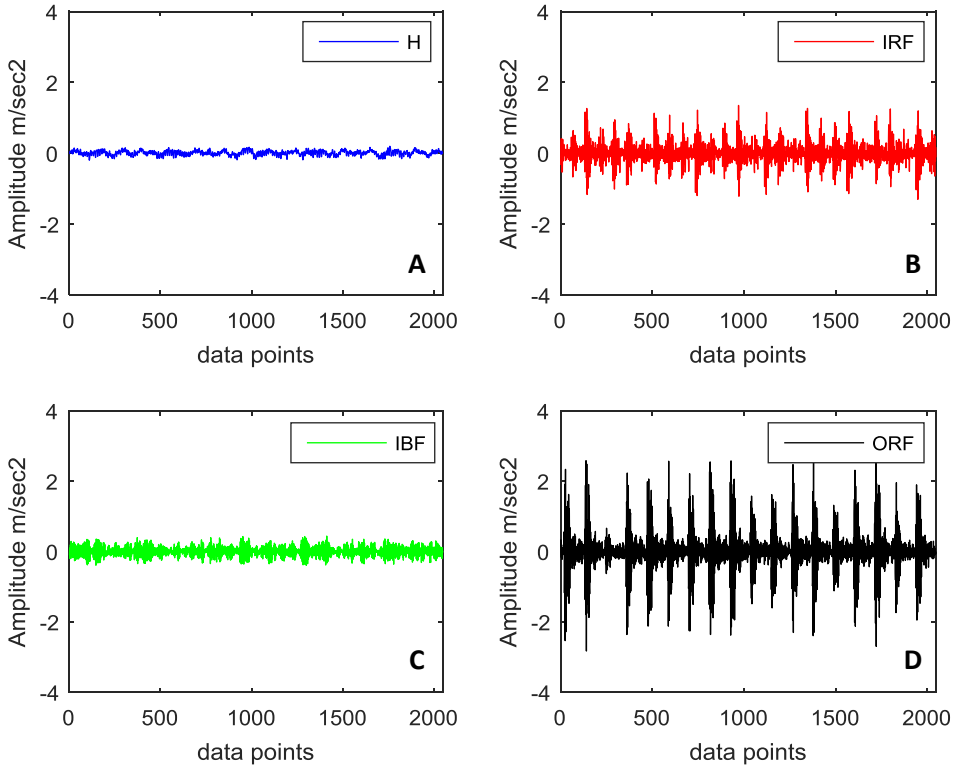


Figure 7.1: Raw sub-signals for different bearing conditions at 1772 RPM-CS1-CW3.**A**.Healthy category. **B**. Inner race fault category .**C**. Ball fault category .**D**. Outer race fault category

Signal denoising

Each sub-signal is subjected to the SSA and decomposed into a number of PCs where each one contributes to the variance of the original sub-signal. The maximum window size of the SSA should not be more than half of the sub-signal length (*n*) under study. In the current study, the SSA window size is selected as 140 for all the cases. The Figure 7.2-**A** illustrates the normal scree plot in which the normalised eigenvalues per each PC.

The Figure 7.2-**B** shows the cumulative scree plot where cumulative values of the normalised eigenvalues are shown for the four bearing conditions. It can be easily seen that most variance of the original sub-signal is conserved in the first components while the last components have the lowest portion of variance where scree plot becomes flat.

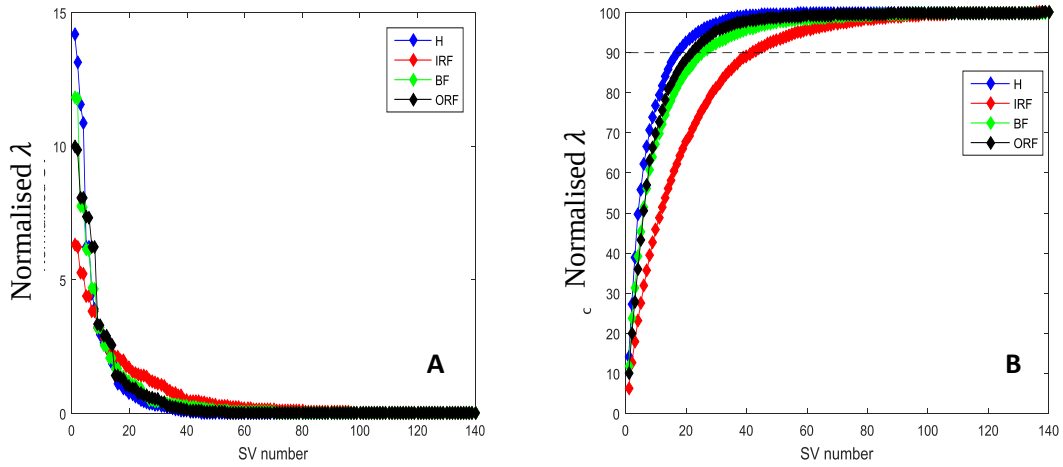


Figure 7.2: A-Normal scree plot and **B**- Cumulative scree plot for different bearing conditions at 1772 RPM

The new denoised sub-signal can be reconstructed using a number of PCs. The Figures 7.3**A** -7.3**D**, show the comparison between the real sub-signal and a reconstructed and denoised sub-signal at different number of PCs. The x-axis represents the number of data points while the y axis represents the vibration acceleration amplitude in m/sec^2. Figure 7.3**A**, compares the real sub-signal versus a reconstructed sub-signal made using the 1st PC only. It can be seen that the reconstructed represented mainly the trend of the real sub-signal. When the number of PCs is increased, the reconstructed sub-signal will contain more information from the original sub-signal and becomes more similar to the original one (see the Figure 7.3**B** for 3 PCs, Figure 7.3**C** for 5 and Figure 7.3**D** for 18). All these number of PCs are out of 140, the total number of the PCs which the sub-

signals were originally decomposed. The remaining of the PCs was rejected as they are corresponding to the structure-less noise. In this study, the number of the PCs used in the reconstraction stage is selected such that the percentage of the recovered variance is 90% of the original variance of the original sub-signal.

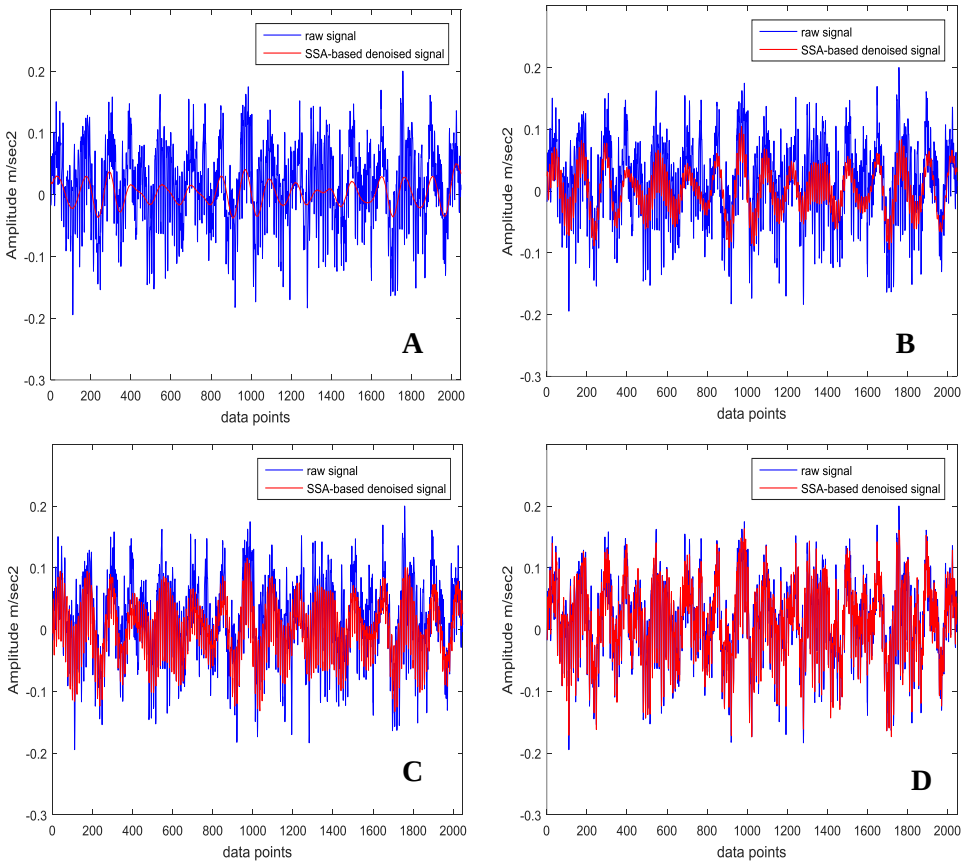


Figure 7.3: The real raw signal versus SSA-based denoised signal. **A.** 1PC,**B.**3PCs**. C.** 5PCs. **D.**18 PCs

For the case used in the demonstration the number of PCs used in the reconstruction signals are shown in Table 7.1. The number in bold represents the number of the PCs selected for the reconstruction of the signal. The percentage inside the square brackets represent the percentage of the variance recovered using these numbers of PCs. It can be seen that percentages are very close to 90%. For all the data sets considered in the

analysis, the number of the selected PCs and the recovered percentage of original signal variance can be seen in Tables 7-A.1 to 7-A.3 (appendix 7-A).

H	IRF	BF	ORF
18[90.88%]	**42**[90.1%]	**26**[90.38%]	**22**[90.176%]

Table 7.1: the number of PCs selected for reconstruction purpose and the corresponding variance percentages for CS1-CW3

Signal differencing

As it was mentioned in section 6.2, a noise free sub-signal is subjected to a stationarity test and then if it is identified as a non-stationary, it is subjected to differencing in order to transform it to stationary by stabilising its mean. The differencing helps in stabilising the mean of the sub-signal over the time and eventually stationariese the sub-signal. Figure 7.4 shows the sub-signal before and after the 1^{st} order differencing process of the non-stationary sub-signal. The figure compares the fluctuation of mean of the non-stationary sub-signal over the time before and after 1^{st} order differencing. For both versions of sub-signal (before and after differencing), the sub-signal is divided into a number of smaller sub-sections each of 100 data points and then the averages of each of these 100 data points are found. The x-axis represents the number of the sub-sections while the y-axis represents the average of these subsections data points. It is clear from the figure, that mean of the sub-signal is considerably stabilised which indicates the achievement of the stationarity.

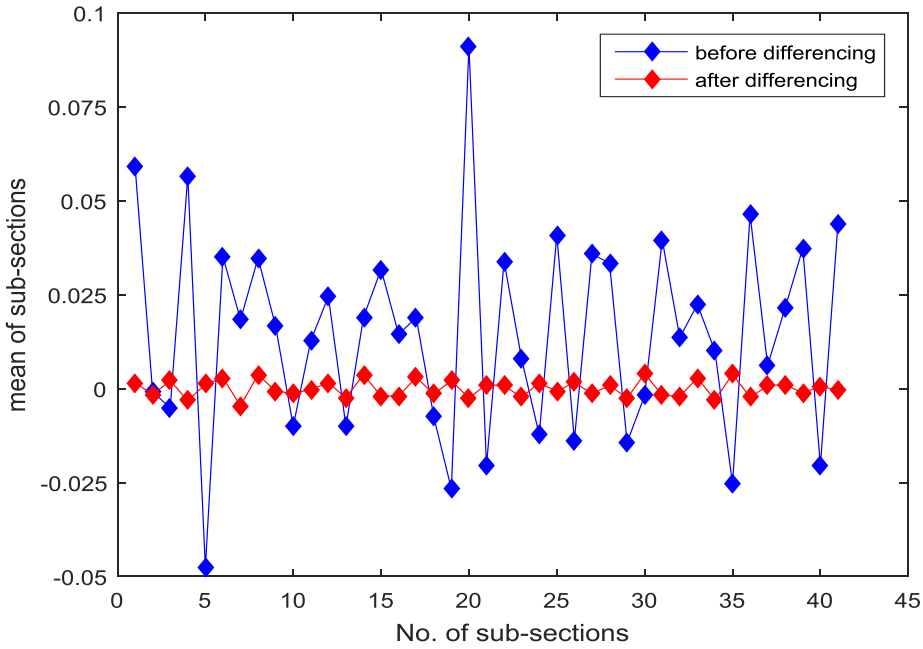


Figure 7.4: stabilisation of a non- stationary sub-signal mean by 1st order differencing

LTIVAR model

At the next step of the analysis, the stationarized sub-signals are subjected to the LTIVAR modelling (Eq. 6.2). Before subjecting the sub-signals to the modelling, it is necessary to estimate the optimum model order in order to avoid the problems of the under fitting and overfitting which arise from the use of improper model order. As it was mentioned in section 6.2.5, the FPE function is used for the purpose of estimation of the optimum model order. Figure 7.5 shows the FPE values versus the model order for the four bearing conditions H, IRF, BF and ORF. For all the conditions in the figure, the value of the FPE decreases considerably and at some point it has a flat trend. According to the criterion of the optimum model order selection, the first point, where FPE curve shows no more considerable decrease, is selected to determine the optimum model order.

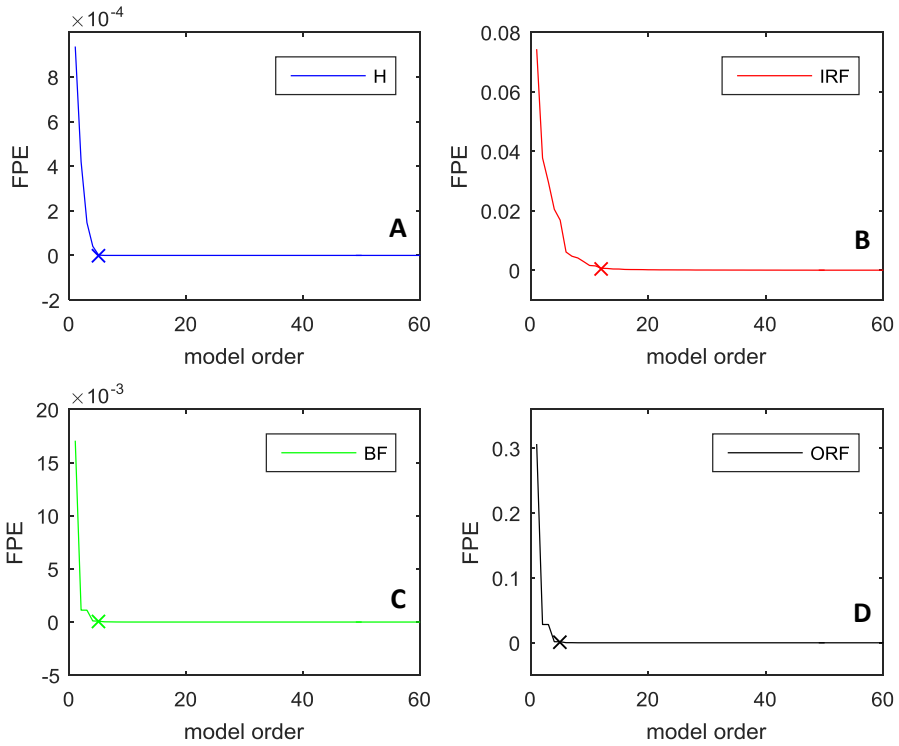


Figure 7.5: FPE values versus different model order.**A**.Healthy category.**B**. Inner race fault category.**C**.Ball fault category.**D**.Outer race fault category.

For the case considered in the demonstration (i.e CW3 in Table 4.2) the optimum model orders are shown in Table 7.2. The model optimum orders and NMSE are obtained according to the Eqs. (6.4) & (6.3) respectively. For all the data sets considered in the analysis, the optimum orders for the other case studies can be seen in Table 7-A.4 to 7-A.6 in appendix 7-A.

H	IRF	BF	ORF
5[99.96 %]	**12**[99.07%]	**5**[99.47%]	**5**[99.35 %]

Table 7.2: The optimum model order of the LTIVAR (bold) and the NMSE (percentages inside the square brackets) for the CW3.

It is important to mention that the NMSE values shown in the Table 7.2 and any other tables are normalised values (i.e subtracted from one and then multiplied by 100 see Eq.(6.3))It can be seen from the Table 7.2, the optimum order of the LTIVAR are the same for the H, BF, and ORF categories while it is different for the IRF category. For this reason, the change of the optimum order of the model cannot be always used as an indicator for the fault detection. In the next step, the stationarised bearing vibrations acceleration sub-signals are subjected for representing by the LITVAR model and the model coefficients are obtained by least square algorithm. The Figure 7.6 compares an example the stationarised real time sub-signal with the modelled one. It can be seen that the model very precisely represents the sub-signal with a goodness of fit 99.97% calculated according (Eq. (6.3)). It is very difficuilt to distinguish among them. And a part of the sub-signal is magnified largely to show how the two signals (real and simulated) are very similar.

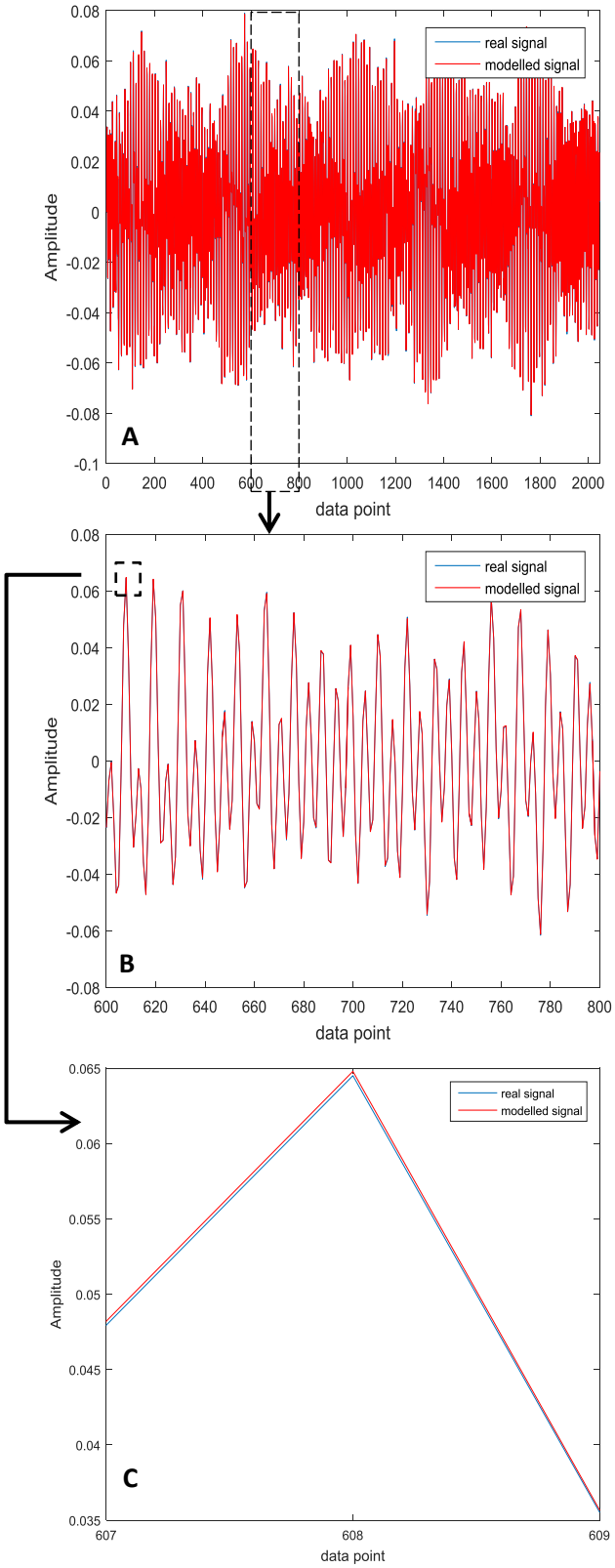


Figure 7.6: Comparison of a modelled and a real time denoised sub-signals. **A.** whole length of the signals**. B.** zoomed part (600^{th} -800^{th}) data points.. **C**. zoomed part (607-609) data points.

Improvement of Goodness of Fit of LTIVAR model

The signal pretreatment proposed in the 2nd methodology is aimed to improve the goodness of fit of the model particularly for the faulty signals which are considered more complex than healthy signals. Table 7.3 shows the improvement in LTIVAR goodness of fit (i.e NMSE values) due to signal pretreatment. In the Table 7.3, the NMSE is shown in normalised and percentage form as was calculated in Eq.(6.3). The first and second columns represent the rotational shaft speed and the bearing condition respectively. The third and fourth columns show the normalised NMSE in percentages (before and after NMSE). The last columns represent the improvement achievement in the LTIVAR goodness of fit. This improvement is calculated by finding the difference between the NMSE of before and after applying the signal pretreatment. For all the data sets considered in the analysis, the improvement achieved in the NMSE can be seen in the Tables 7-A.7 &7-A.8 in appendix 7-A.

Shaft speed (RPM)	Bearing condition	NMSE -before signal pretreatment	NMSE -after signal pretreatment	achieved improvement
1730	**H**	94.6%	99.99 %	5.71%
	IRF	86.14%	99.54%	15.55%
	BF	90.97 %	99.91 %	9.83%
	ORF	83.04 %	99.93%	20.34%
1750	**H**	93.45 %	99.96 %	6.97%
	IRF	81.96 %	99.24%	21.08%
	BF	88.94%	99.97 %	12.41%

1772	**ORF**	85.48%	99.93 %	16.19%
	H	94%	99.96 %	6.34%
	IRF	78.39%	99.07%	26.37%
	BF	87.6%	99.47%	13.55%
1797	**ORF**	83.66%	99.35 %	18.75%
	H	95.27 %	99.97%	4.93%
	IRF	76.36 %	98.99%	29.62%
	BF	88.51%	99.71%	12.66%
	ORF	87.45 %	99.46%	13.73%

Table 7.3: Improvement of NMSE due to signal pretreatment CS1-CW3

7.2.2 Bearing signal diagnosis

In this section the results of the sub-signals classification are introduced. This section will be divided into two parts namely, 1) fault detection and identification and 2) fault severity estimation. For both stages of diagnosis, the classification of sub-signals is based on the minimum Mahalanobis distance as explained in Eq (6.6).In the fault detection and identification part results will be introduced for the average correct classification rates for FVs based on the fault locations. In the fault estimation part, results will be introduced for the average correct classification rates for FVs based on the fault severity.

Fault identification

As was illustrated in section 6.3.2, the Mahalanobis distances are measured for any new testing FV to each of the feature matrices corresponding to the training sample. The Figures 7.7**A**-7.7**D** show the Mahalanobis distances (D_{Ki}) of the testing FVs to the feature matrices corresponding to the training sample (see Eq.(6.5). In each figure the number of the testing FVs is 30. The D_{Ki} of Each of these 30 FVs is measured to each of the four feature matrices corresponding to H, IRF, BF and ORF as was explained in Section 6.3.2. Thus in each of the Figures 7.7**A**-7.7**D**, the x-axis represents the 120 D_{Ki} values (i.e D_{Ki} of 30 FVs to 4 feature matrices). Figure 7.7**A** represents the D_{Ki} of the testing FV corresponding to H bearing condition. Figure 7.7**B** represents the D_{Ki} of the testing FV corresponding to IRF bearing condition. Figure 7.8**C** represents the D_{Ki} of the testing FV corresponding to BF bearing condition. Figure 7.8**D** represents the D_{Ki} of the testing FV corresponding to ORF bearing condition. In these figures, the blue points represent the D_{Ki} of the testing FVs to the training feature matrix corresponding to H bearing condition. The red points represent the D_{Ki} to the training feature matrix corresponding to IRF bearing condition. The green points represent the D_{Ki} to the training feature matrix corresponding to IRF bearing condition. The black points represent the D_{Ki} to the training feature matrix corresponding to IRF bearing condition. From Figure 7.7**A**, it is clear that testing FV corresponding to the H bearing class have the minimum distances to the feature matrix corresponding to the H bearing conditions. This means that all the testing FV from H signal class are correctly classified. The same notes can be seen for the other figures (i.e **B**, **C** and **D**).

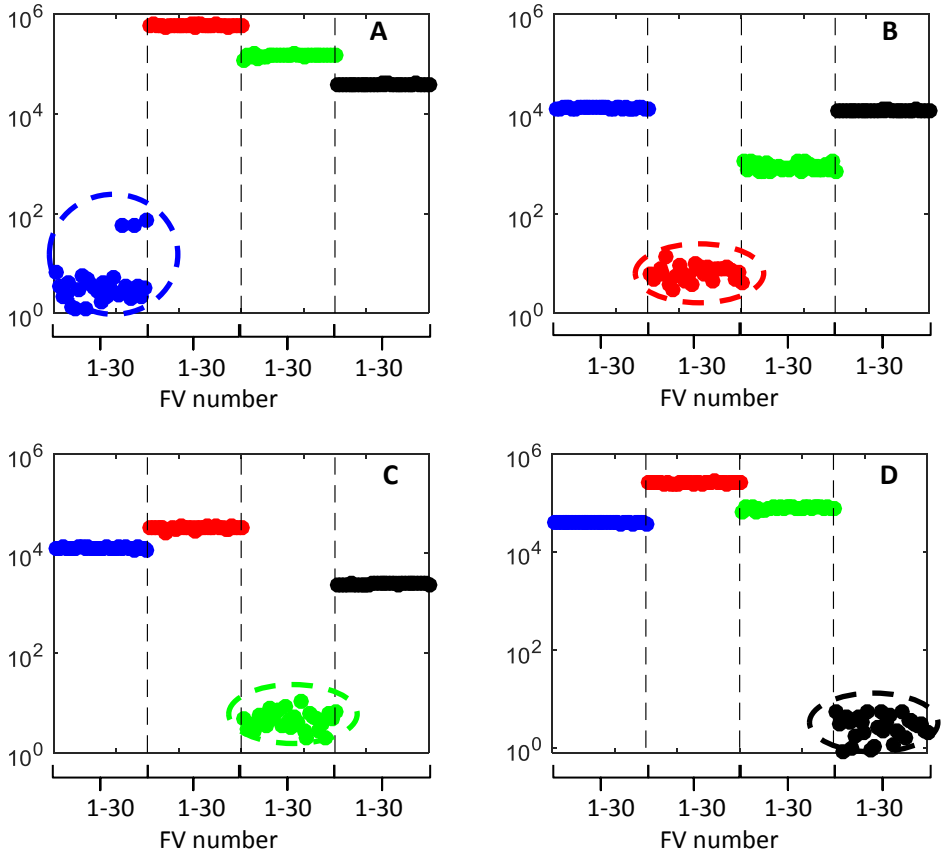


Figure 7.7: Mahlanobis distances (D_{Ki}) of the testing FVss corresponding to, **A**. Healthy category, **B**.Inner race fault category. **C**.Ball fault category and **D**. outer race fault category, to the different training feature matrices.

A confusion matrix in Table 7.4 can be made based on the results shown in Figure 7.7. The Table 7.4 represents the correct classification percentage rates of classifying 120 FVs (4 categories * 30 testing FVs of each category). It is clearly seen that all the testing FV are correctly classified to their original classes.

		Prediction			
		H	**IRF**	**BF**	**ORF**
Actual	**H**	100%	0%	0%	0%
	IRF	0%	100%	0%	0%
	BF	0%	0%	100%	0%
	ORF	0%	0%	0%	100%

Table 7.4 : The confusion matrix using feature vectors modified using zero padding: 120 feature vectors at 1772 RPM

Effect of sub-signal length and length equalisation

The performance of methodology in terms of correct classification rates is also investigated using different sub-signal lengths, typically 512, 1024, 2048 and 4096, and different possibilities of equalisation of FVs (see section 6.2.6). In this investigation, all the method steps are repeated using different sub-signal lengths and using different procedures of FV lengths equalisation. Figure 7.8 **A**-**D** shows the percentage of the average correct classification rates for using different sub-signal lengths for different FV lengths equalisation procedures. In each figure, the x-axis represents the sub-signal length (*n*) where the methodology is applied. The y axis represents the average correct classification rate in percent. The percentages of the average correct classification rates are shown for different shaft rotational speed (i.e 1730, 1750, 1772 and 1797 RPM).

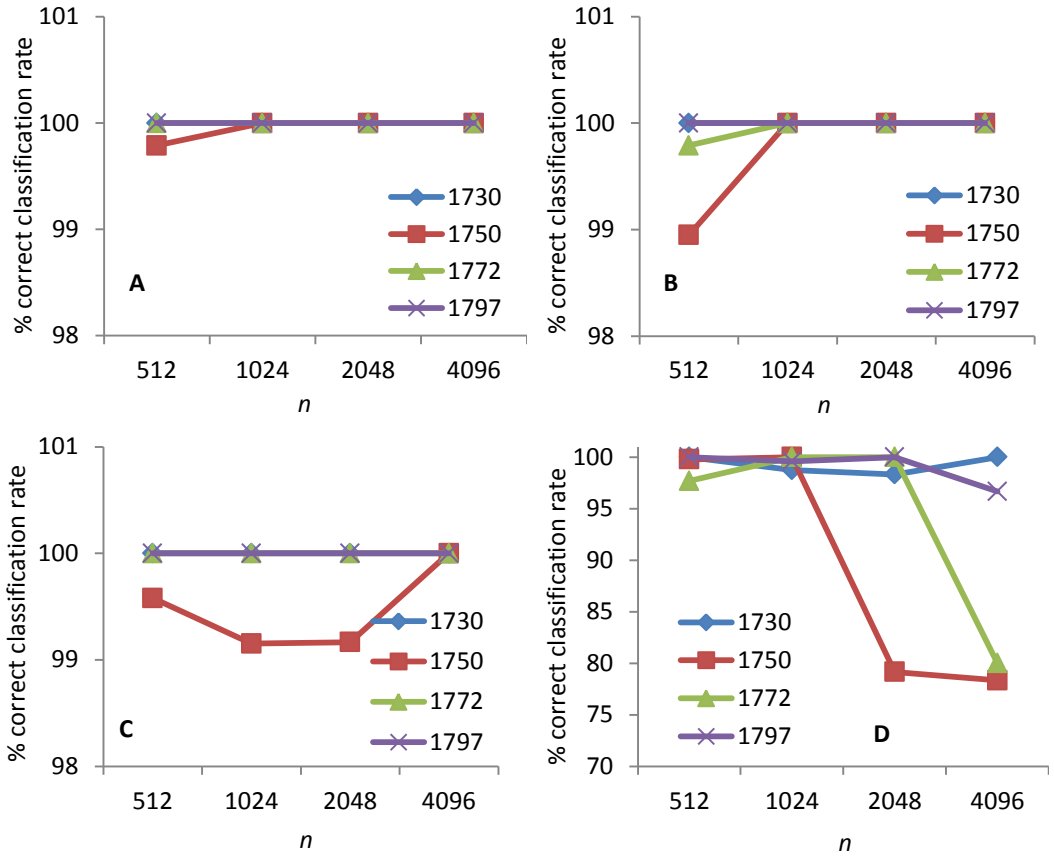


Figure 7.8: The percentages of averaged correct classification rates for different sub-signal lengths and different FV equalisation procedures. **A**-ZP, **B**-Trim, **C**-Max and **D**-VTM.

The plots show the percentages of the average correct classification rates which are obtained from finding the mean value of the main diagonal of the confusion matrix of each case. The following paragraph will explain the obtaining of the percentage of the average correct classification rate.

For example, in the Figure 7.8-**B** which represents the percentage of the average correct classification rate data obtained from case study CW3 (see Table 4.2) for the case of equalisation of FVs by trimming. For a sub-signal length of 512 data points, the corresponding confusion matrix is shown in the Table 7.5.

		Prediction			
		H	**IRF**	**BF**	**ORF**
Actual	**H**	100%	0	0	0
	IRF	0	100%	0%	0%
	BF	0	0%	100%	0%
	ORF	0	0	0.83%	99.17%

Table 7.5: The confusion matrix of CW3, sub-signal of 512 (data points), equalisation by trimming

The main diagonal of the confusion matrix shows the percentages of the correct classification rate for 480 FVs from the testing sample. Only one FVs corresponding to ORF category was misclassified as BF category which gives (100* 119/120%=99.17%) percentage of correct classification rate. The percentage of the correct classification rate is calculated by finding the mean value of the main diagonal of the confusion matrix. Thus, the percentage of the average correct classification rate related to this confusion matrix is (100%+100%+100%+99.17%)/4=99.79%. The same procedure is followed for all other cases.

The Table 7.6 shows the average correct classification rate for the case study 1.In terms of equalisation of FV lengths, it can be seen that methodology has a poor performance when the VTM procedure is used (see scion 5.2.6). Furthermore, the methodology performs well when the other equalisation procedures (i.e ZP, trimming and MAX) are used.

Case no.	Speed (RPM)	No. of FVs	ZP	TRIM	MAX	VTM
				***n*=512**		
CW1	1730	960	99.79%	100%	100%	100%
CW2	1750	960	99.79%	98.95%	99.58%	99.79%
CW3	1772	960	100%	99.79%	100%	97.69%
CW4	1797	960	100%	100%	100%	100%
				***n*=1024**		
CW1	1730	480	100%	100	100	98.75%
CW2	1750	480	100%	100	99.15	100%
CW3	1772	480	100%	100	100	100%
CW4	1797	480	100%	100	100	99.58%
				***n*=2048**		
CW1	1730	240	100%	100%	100%	98.33%
CW2	1750	240	100%	100%	99.17%	79.16%
CW3	1772	240	100%	100%	100%	100%
CW4	1797	240	100%	100%	100%	100%
				***n*=4096**		
CW1	1730	120	100%	100%	100%	100%
CW2	1750	120	100%	100%	100%	78.33%
CW3	1772	120	100%	100%	100%	80%
CW4	1797	120	100%	100%	100%	96.66%

Table 7.6: The average correct classification rates for CS1

Table 7.7 shows the average correct classification rate for the data sets obtained for CS2 (see Table 4.3). The classification rates are shown for different possibilities of equalisation of feature vector lengths. In terms of the procedures of equalisation of feature vector lengths, it can be notices that classification rates, when ZP or TRIM, are generally better than when MAX and VTM are used (see for example ST2 at *n*=512, ST3 at *n*=2048). Furthermore, the correct classification rates when the VTM procedure use are generally lowest among all of those corresponding to different procedures of equalisation of feature vector lengths.

Case no.	Speed (RPM)	No. of FVs	ZP	TRIM	MAX	VTM
				***n*=512**		
ST1	250	1280	94.37%	94.84%	88.12%	94.84%
ST2	750	1280	93.43%	94.06%	78.43%	92.34%

ST3	1250	1280	84.68%	84.06%	84.68%	78.75%
				***n*=1024**		
ST1	250	640	92.5%	87.5%	92.5%	85.62%
ST2	750	640	85.62%	86.87%	85.62%	82.18%
ST3	1250	640	93.75%	91.25%	93.75%	84.06%
				***n*=2048**		
ST1	250	320	98.75%	95.62%	98.75%	89.37%
ST2	750	320	90.62%	90%	90.62%	80.62%
ST3	1250	320	100%	99.37%	98.75%	86.87%
				***n*=4096**		
ST1	250	160	100%	100%	98.75%	100%
ST2	750	160	97.5%	98.75%	97.5%	97.5%
ST3	1250	160	100%	100%	96.25%	98.75%

Table 7.7: The average correct classification rates for CS2

In terms of sub-signal lengths, it can be seen that increasing the sub-signal lengths relatively improve the correct classification rates. However the improvement of the correct classification rates with the increase of the sub-signal lengths is not completely linear. For example, under the column (ZP) the classification rates is 94.37% at *n*= 512 and it slightly decreases to 92.5% when *n* increases to 1024 then it eventually increases to 98.75% and to 100% when *n* becomes 2048 and 4096 respectively.

In terms of sub-signal lengths, it can be seen that increasing the sub-signal lengths relatively improve the correct classification rates. However the improvement of the correct classification rates with the increase of the sub-signal lengths is not completely linear. For example, under the column (ZP) the classification rates is 94.37% at *n*= 512 and it slightly decreases to 92.5% when *n* increases to 1024 then it eventually increases to 98.75% and to 100% when *n* becomes 2048 and 4096 respectively.

In the Table 7.8, the results of the average correct classification rates are presented for the data sets corresponding to CS3. It can be seen that correct classification rates, when

the VTM procedure used for equalisation of feature vector lengths, are lowest when compared to the classification rates at different procedures (i.e ZP, TRIM and MAX).

Case no.	Speed (RPM)	No. of FVs	ZP	TRIM	MAX	VTM
				***n*=512**		
PT1	18000	4800	99.45%	99.45%	99.45%	94.54%
PT2	24000	4800	98.25%	96.87%	98.29%	80.45%
PT3	30000	4800	97.54%	97.54%	97.54%	92.5%
				***n*=1024**		
PT1	18000	2400	99.91%	99.83%	99.91%	91.08%
PT2	24000	2400	100%	100%	99.83%	98.33%
PT3	30000	2400	99.5%	99.5%	99.5%	94.83%
				***n*=2048**		
PT1	18000	1200	100%	100%	100%	99%
PT2	24000	1200	100%	99.83%	100%	98.5%
PT3	30000	1200	99.33%	99.16%	99.33%	98.16%
				***n*=4096**		
PT1	18000	600	99.66%	99.66%	99.66%	97.33%
PT2	24000	600	100%	100%	100%	100%
PT3	30000	600	100%	100%	100%	95%

Table 7.8: The average correct classification rates for CS3

It can be seen also that the method performs well at all the sub-signal lengths considered in the analysis for this case study. It performs well for different speeds. In conclusion, the method performs well for detecting fault different locations of fault at different speeds and different sub-signals lengths.

Fault severity estimation

Figures 7.9-7.11 show the D_{Ki} (i.e Mahalanobis distance) of the testing FV to the training feature matrices for the cases of different fault severity (sub-signal length n=2048). The y-axis represents the D_{Ki} to the feature matrices which are made from the training FVs. The x-axis represents the number of training FVs. For the case of n =2048, the number of the total FVs is 240 as shown in Table 7.6 and they were divided

equally into 120 training sample (i.e 30 FVs *4 categories) and 120 testing sample(i.e 30 FVs *4 categories). In each figure, the D_{Ki} of the 30 training FVs measured to the 4 testing sample categories are shown. At each x-axis of the figures, 4 groups of 1-30 training FVs are demonstrated. In these figures, the blue points represent the D_{Ki} of the testing FVs to the training feature matrix corresponding to H bearing condition. The red points represent the D_{Ki} to the training feature matrix corresponding to small size fault signal category (i.e 0.007 inch). The green points represent the D_{Ki} to the training feature matrix corresponding to medium size fault (i.e 0.014 inch) signal category. The black points represent the D_{Ki} to the training feature matrix corresponding to the large size (i.e 0.021 inch) signal category.

For the Figure 7.9 which shows the D_{Ki} corresponding to the IRF signal category, it can be seen that all the testing FV has a minimum distances with their actual feature matrix. This means they are all successfully classified to their actual category. The same notice can be made for the other sub-signal categories as in Figure 7.10 for detection of different fault severity in BF and Figure 7.11 for detection of different fault severity in ORF.

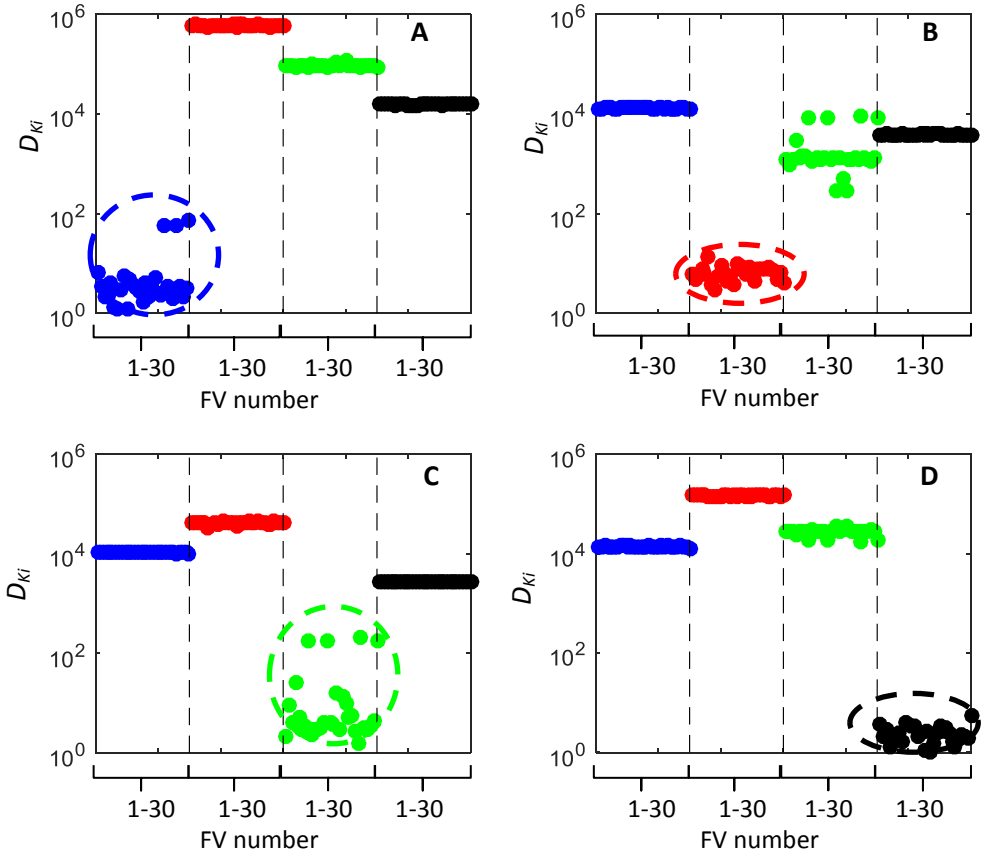


Figure 7.9: Mahalanobis distances (D_{Ki}) of the testing FVs corresponding to, **A**. Healthy category, **B.**small size inner race fault category. **C**. medium size inner race fault category and **D**. large size inner race fault category, to the different training feature matrices.

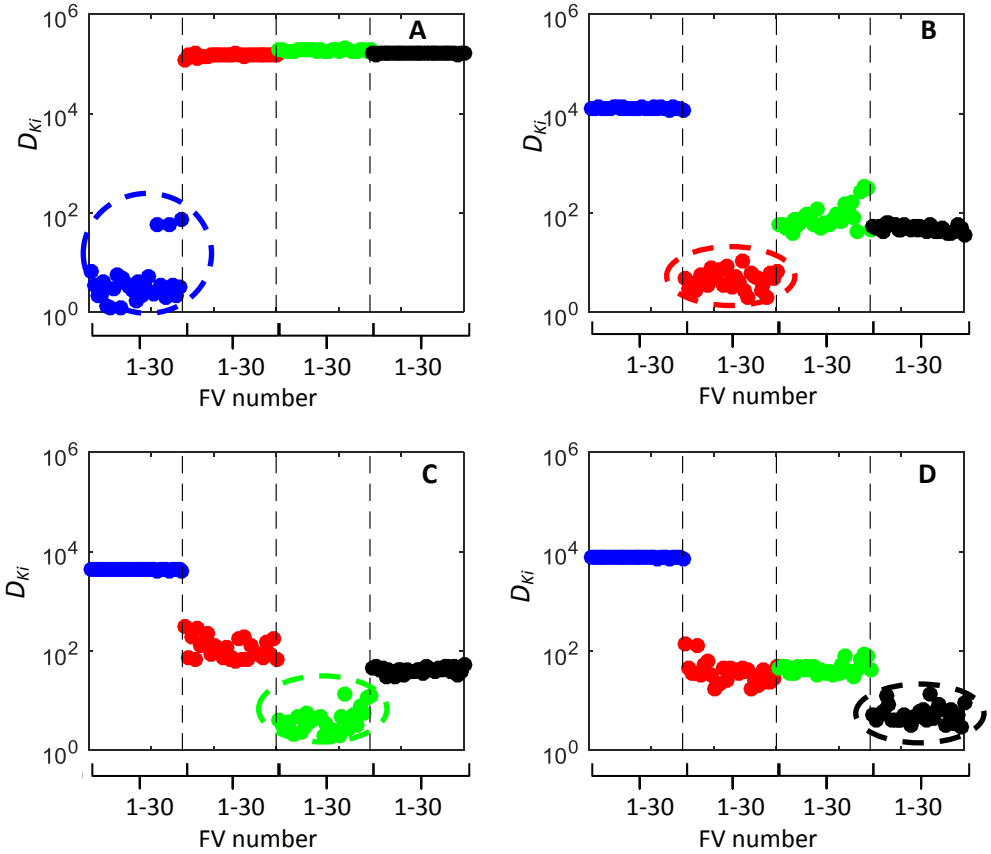


Figure 7.10: Mahalanobis distances (D_{Ki}) of the testing FVs corresponding to, **A**. Healthy category, **B.**small size ball fault category. **C**. medium size ball fault category and **D**. large size ball fault category, to the different training feature matrices.

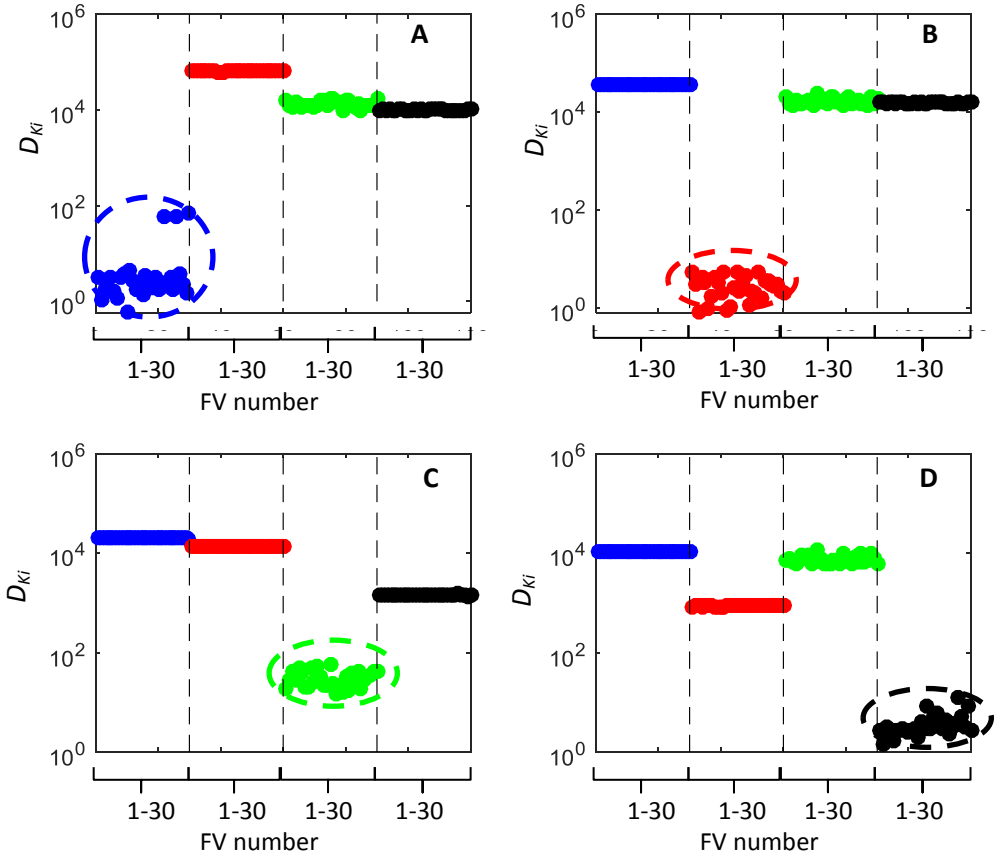


Figure 7.11: Mahalanobis distances (D_{Ki}) of the testing FVs corresponding to, **A**. Healthy category, **B.**small size outer race fault category. **C**. medium size outer race fault category and **D**. large size outer race fault category, to the different training feature matrices.

The rates of averaged correct classification are shown in Table 7.9. It can be seen that all the testing FVs are correctly classified to their actual categories. However, for the case CW12 , which refers to the data sets corresponding to healthy and different fault severities at a ball element obtained at 1797 RPM, the correct classification rate is 95% . For this case, 6 out of the 30 testing FVs corresponding to small (0.007 inch) ball fault size category are misclassified as medium (0.014 inch) ball fault size category.

Case no.	Motor speed (rpm)	Signal Category	No of testing FVs	Average correct classification rate
CW5	1730	Healthy & IRF (S,M and L)	120	100%
CW6	1750	Healthy & IRF (S,M and L)	120	100%
CW7	1772	Healthy & IRF (S,M and L)	120	100%
CW8	1797	Healthy & IRF (S,M and L)	120	100%
CW9	1730	Healthy & BF (S,M and L)	120	100%
CW10	1750	Healthy & BF (S,M and L)	120	100%
CW11	1772	Healthy & BF (S,M and L)	120	100%
CW12	1797	Healthy & BF (S,M and L)	120	95%
CW13	1730	Healthy & ORF(S,M and L)	120	100%
CW14	1750	Healthy & ORF(S,M and L)	120	100%
CW15	1772	Healthy & ORF(S,M and L)	120	100%
CW16	1797	Healthy & ORF(S,M and L)	120	100%

Table 7.9: The average correct classification rates based on zero-padding as equalisation of feature vectors for fault identification CS1.

The Table 7.10 shows the average correct classification rates are presented for the CS2 (i.e PT4-PT9) which presents the precision of the methodology in distinguishing among the different fault severities. The first three cases (PT4, PT5 and PT6) represent the methodology preciseness in distinguishing among different fault severities at inner raceway and at different rotational speeds. It can be seen that all the testing FVs were correctly classified to their actual categories.

The cases (PT7, PT8 and PT9) represent the methodology preciseness in distinguishing among different fault severities at an element raceway and at different rotational speeds.

It can be seen that only a few testing FVs are misclassified in the cases PT7 and PT9 while all the testing FVs corresponding to PT8 are correctly classified. For PT7, there was only one testing FVs corresponding to large (0.0178 inch) roller fault size is misclassified as a medium (0.0098 inch). For PT8, a testing FVs corresponding to healthy bearing category is misclassified as a small (0.0006 inch) and 3 testing FVs corresponding to small (0.0006inch) roller fault size category. And one testing FV corresponding to medium (0.0098 inch) roller fault size is misclassified as a small (0.0006 inch) roller fault size.

Case no.	Motor speed (rpm)	Signal Category	No of testing FVs	Average correct classification rate
PT4	18000	IRF (S, M and L)	1200	100%
PT5	24000	IRF (S, M and L)	1200	100%
PT6	30000	IRF (S, M and L)	1200	100%
PT7	18000	BF (S, M and L)	1200	99.88%
PT8	24000	BF (S, M and L)	1200	100%
PT9	30000	BF (S, M and L)	1200	99.38%

Table 7.10: The average correct classification rates based on ZP as equalisation of feature vectors for CS3 fault estimation.

Fault severity index

In this research, the use of the Mahalanobis distance of testing FVs measured to the healthy bearing category ($D_{to\ healthy\ category}$) as a fault severity index is also investigated. The Figure 7.12 shows the $D_{to\ healthy\ category}$ of the training FVs corresponding to healthy and different inner race fault severities measured for CS1. The figure shows that the methodology is capable of detecting different inner fault severities for all the speed

considered in the analysis. However, distinguishing among the different inner fault severity categories is not clear. In Figure 7.13, distinguishing among the different ball severity categories is better than the case shown in the Figure 7.12. For Figure 7.13**A**, all the testing FVs corresponding to the fault categories are detected and the $D_{to\ healthy\ category}$ show some changes when the severity level of the ball fault changes. Figure 7.13**B**, the change of $D_{to\ healthy\ category}$ versus the change of the fault severity level is clearer although it does not follow a monotonic function. Figure 7.14, shows the $D_{to\ healthy\ category}$ of the training FVs corresponding to healthy and different outer race fault severities. It can be seen clearly that the $D_{to\ healthy\ category}$ decreases when the fault severity increases at all rotational speeds considered in the analysis.

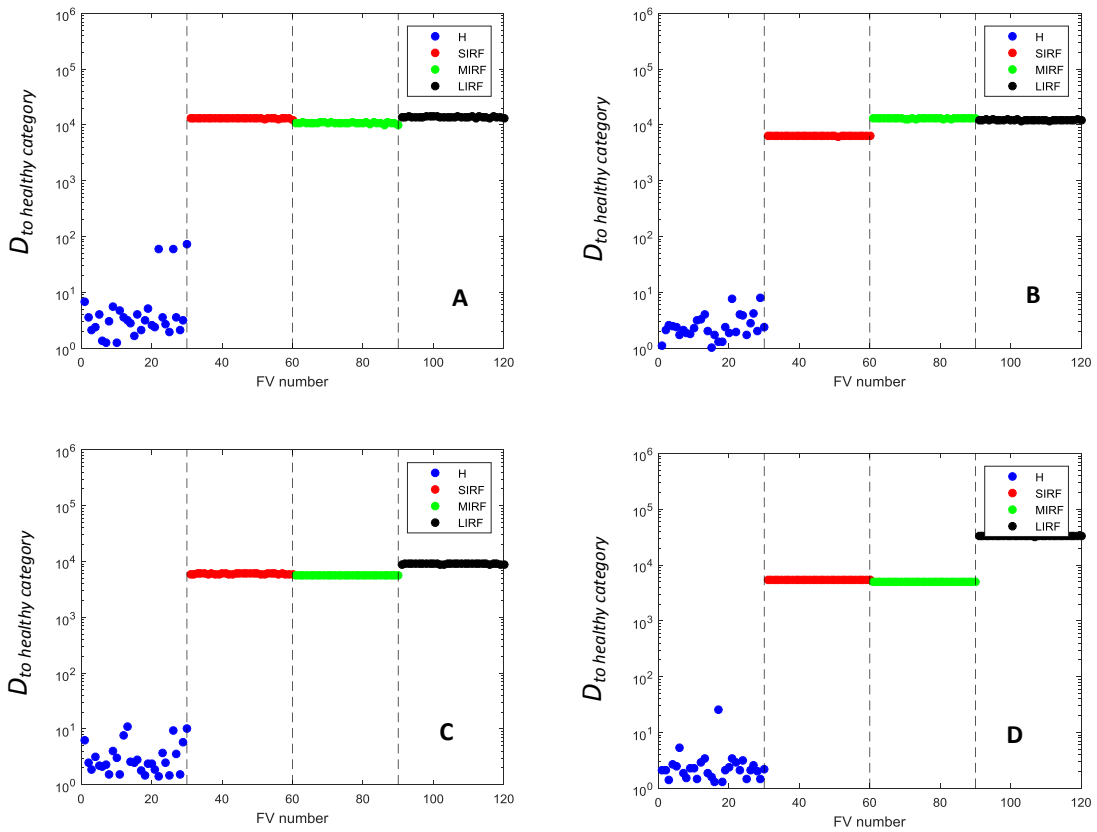


Figure 7.12: The Mahalanobis distance to of IRF testing FVS to the healthy training feature matrix.**A**. 1730RPM.**B**. 1750RPM.**C**. 1772RPM.**D**.1797RPM.

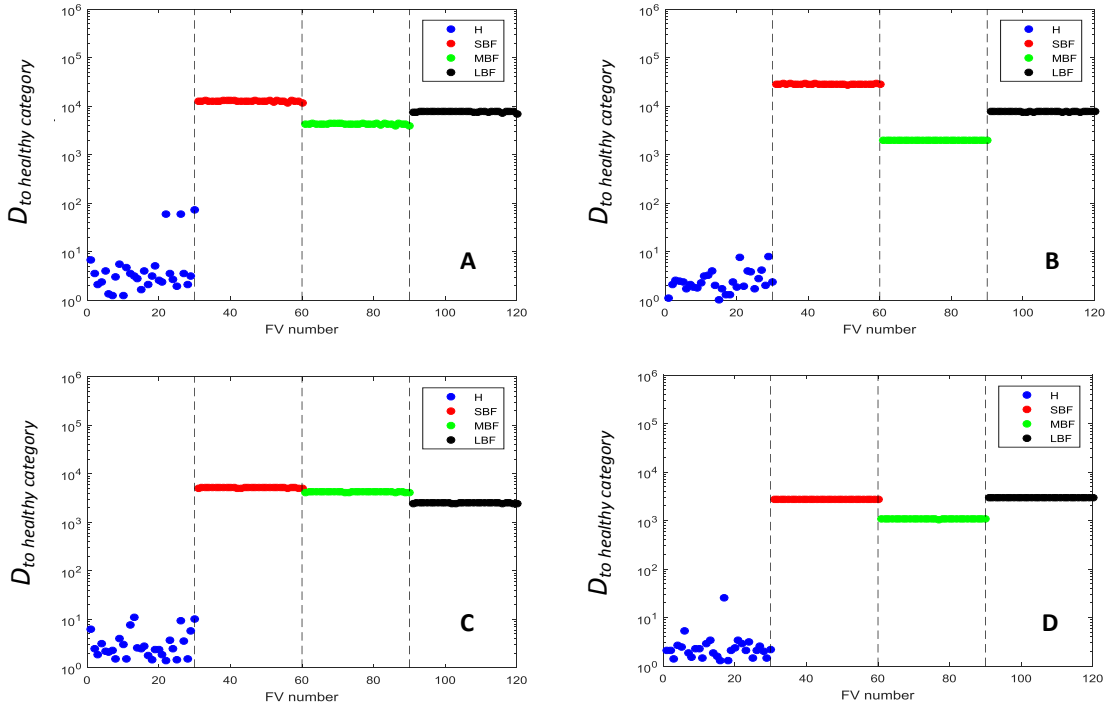


Figure 7.13: The Mahalanobis distance of BF testing FVS to the healthy training feature matrix.**A**. 1730RPM.**B**. 1750RPM.**C**. 1772RPM.**D**.1797RPM.

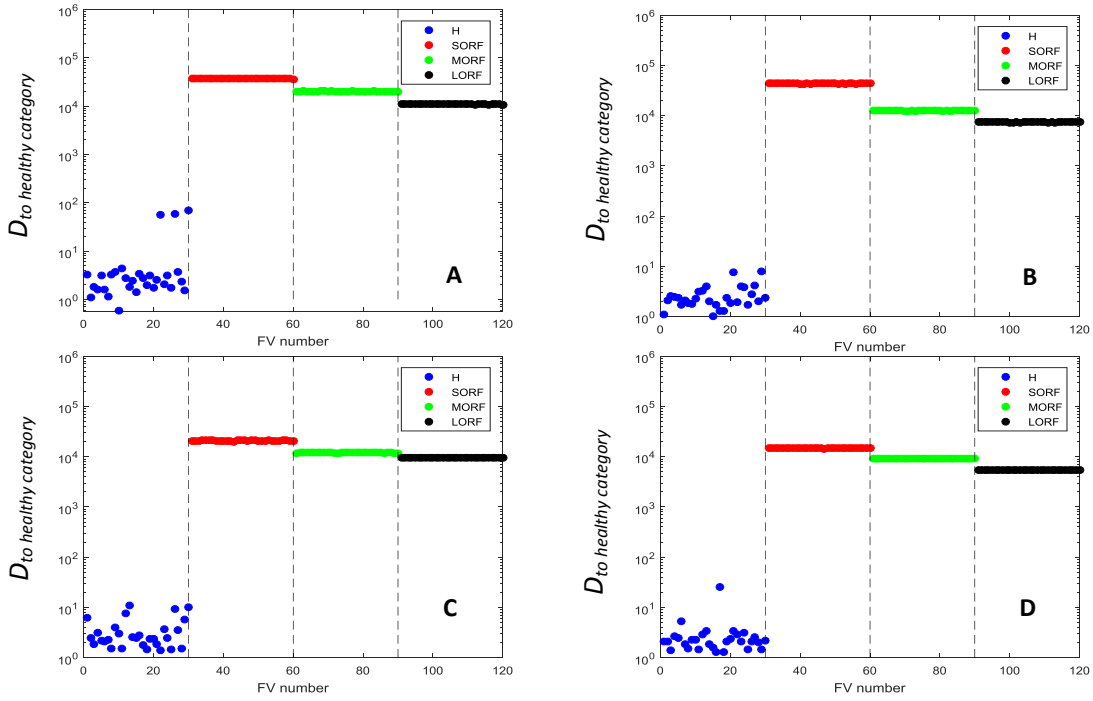


Figure 7.14: The Mahalanobis distance of ORF testing FVS to the healthy training feature matrix.**A**. 1730RPM.**B**. 1750RPM.**C**. 1772RPM.**D**.1797RPM.

7.3 Results comparison to some published work

The performance of the current methodology, which is presented in chapter 6, is also compared to some of recent published work as in the Table 7.11. The details of the compared methods are already shown in (Chapter 5- Table 5.12).

It can be seen from Table 7.11 that all the compared methods achieve a rather good classification rate. However, the methodology presented in this chapter (as well as the methodology presented in chapter 4) over-performs the compared methods in terms of the average correct classification rate, which is between 99.8-100%

Method	Data set	Training and testing FVs	Conditions classified	Average correct classification rates (%)	Fault features
1)DH and FFNN[95]	0.18, 0.36,0.53mm; 0–3 HP load;30000 data points	Both 144	(IF, OF and BF classification only)	92-95	First 6 histogram bins
2)SSA and BPNN [65]	0.18, 0.36, 0.53, 0.71 mm; 0– 3 HP load;6100 data points	Train—336 test—144	H, IF,OF and BF	96.53–100 95–100	4 singular values 3 energy features
3)MVE, PCA and nonlinear neighbour	0.18, 0.36,0.53mm , 0-	-	Healthy and faulty.	94.68-99.98	-

classifier [96]	3 HP, 2000 data points		It was reported by the author that there is some overlapping among some fault classes		
4)The methodology presented in chapter 6	0.18, 0.36,0.53 mm; 0-3 HP loads; 2048 data points	Train- 480 Test- 480	H, IF, OF and BF	99.8-100	coefficients of LTIVAR model

Table 7.11: A comparison of the performance of the present methodology with other published work using the same parts of CWRU data

7.4 Conclusions

The present work suggests a new methodology for rolling element bearing fault diagnosis based on linear time invariant autoregressive modelling and pattern recognition. The suggested method is relatively simple in the sense that it uses a combination of simple processes to first transform the signal and then determine the condition of the bearing. A new signal pretreatment process is applied before subjecting the signals to modelling. This process includes noise cleaning, using singular spectrum analysis (SSA), and stationarisation of the bearing vibration signal by the differencing procedures. The methodology aims to transform the signal to bring it close to a stationary one, rather than complicating the model to bring it closer to the signal. The signal pretreatment proposed enhances the precision of the model prediction, which is influenced by the presence of noise and non-stationary parts in the signal. The LTIVAR model coefficients are extracted using the least squares method and used as FVs for signal classification purpose. The FVs are then presented to the 1-NN algorithm based on the Mahalanobis distance. The signals were assigned to the category of their nearest neighbour. The performance of the methodology is then assessed on the basis of confusion matrices, which provide the percentage of correctly- and incorrectly-classified signals.

The results presented in this chapter have shown the effectiveness of the suggested signal pretreatment in improving the goodness of fit of the LTIVAR model. The combination of the suggested signal pretreatment and the LTIVAR provides an accurate and complete fault diagnosis method.

The current approach also introduces a fresh direction in simplifying non-stationary signals and then representing them by use of a stationary autoregressive model. Compared to other autoregression-based signals, the current methodology is much simpler. This is because the other forms of autoregression that researchers usually use in the literature are complex and require several assumptions. When a LTIVAR model is used, neither the shape of the model coefficient evolution over time, nor the initial set of model coefficients, has to be assumed.

The current approach also shows that equalisation of the FV might also effect the methodology performance in terms of the correct classification rates. It is shown that a zero padding procedure is generally performing well against the other procedures considered in the analysis. The equalisation of the different FV lengths facilitates the classification of different FV lengths based on the Mahalanobis distance classifier.

From the results obtained in chapter 7, it is clear that model order cannot always selected as a feature for the purpose of classification. This is because, for some of the cases considered in the current analysis, we obtain same the model orders (but the model coefficients are different) for some categories but a different one for just one other category. Therefore, a model's optimum order is not always sensitive to changes in the bearing condition.

The current methodology is compared to some other published recent work in terms of correct classification rates, which indicates the accuracy of the methodology. The similarity of the data set sources and specifications for the compared method is

considered. It is shown that the present methodology out-performs those methods and its average correct classification rate is between 99.8%-100%.

The methodology suggested encompasses several relatively simple procedures, which facilitate its potential practical application and its possible automation.

7.5 Summary

In this chapter the results for the methodology, presented in chapter 6, are demonstrated. The methodology was validated using data sets obtained from the three bearing test rigs explained in chapter 4. The results include two main parts 1) demonstration of the methodology using data from CS1 and 2) fault diagnosis. In the first part, the steps of the signal pretreatment including signal denoising, and differencing and modelling by LTIVAR are illustrated. The improvement in the goodness of fit due to signal pretreatment is also explained. In the second part, the fault diagnosis results corresponding to fault detection, fault type identification and fault severity estimation are presented. The correct classification rates are shown and compared to some other recent published methodologies, which use the same data sets. It is shown that the present methodology is superior to those methodologies and that it shows generally higher classification rates

Chapter 8

Conclusions, limitations and future works

8.1 Introduction

This chapter presents the summary of the work done and briefly describes the methodologies developed in this thesis in section 8.2. In section 8.3 it discusses to what extent the objectives, specified in the introductory chapter, are achieved and draws some conclusions made from the work. Then it addresses the limitation challenges that emerged during the work done, and suggests some future work, which could be continued to overcome these limitations and expand the generality of the present methodologies.

8.2 Summary of the work

The thesis covers important areas of bearing conditions, which includes fault detection, fault location identification and fault severity estimation. Two singular spectrum analysis-based methodologies were presented in this thesis. In these two methodologies the singular spectrum analysis method was used in different ways and for different purposes. The work carried out in each method is presented briefly as below: -

8.2.1 A new methodology for fault diagnosis in rolling element bearings using singular spectrum analysis

The SSA is used to create a baseline space from the eigenvectors of the healthy bearing vibration signals. Then the embedded version (i.e the lagged version) of any new signal will be projected on to the baseline space. The feature vectors of the baseline and the new signals are obtained from the norms of the projection as the norm represents the length of the projection vectors. This method is used for the purpose of fault detection, location identification and fault severity estimation.

In the fault detection process, a threshold is determined from the Mahalanobis distances of the training feature vectors corresponding to a healthy bearing category. In the current study, a statistically based threshold is proposed and the necessary assumptions were questioned in chapter 4. Then, the assignment of the testing feature vectors to a healthy or faulty bearing category is based on the comparison to the threshold. From the result obtained for this level of fault diagnosis, the methodology shows a very good performance in terms of correct classification rates.

In the fault location identification, the feature vectors of each bearing signal category (i.e healthy, IRF, BF and ORF) are divided into two samples (training sample and testing sample). The training sample is used for building the feature matrices and then the Mahalanobis distance of each testing FV is measured to each of those feature matrices. Eventually, the testing FV is assigned to the category where the feature matrix has a minimum.

In the fault severity estimation, the embedded version of the signals corresponding to the different fault severities is projected to the baseline space. From these projections, the feature vectors are obtained. The feature vectors of the baseline and each fault severity category (i.e small, medium and large) are divided into two samples (the training sample and the testing sample). The training sample is used for building the feature matrices and then the Mahalanobis distance of each testing FV is measured to each of those feature matrices. Eventually, the testing FV is assigned to the category where the feature matrix has a minimum. As discussed before it is desirable to use a regression model for the diagnosis of the severity condition. In this study, the regression model is not used as it might complicate the methodology, as it requires the determination of the coefficients of the regression model.

The Mahalanobis distances of different fault severities, which are measured to the baseline category, are also used as a fault index, which tracks the changes in fault severities.

The method and the results in terms of the correct classification rates are introduced in chapters 4 and 5.

8.2.2 Bearing fault diagnosis using a novel signal pretreatment and LTIVAR modelling

The SSA combined with a simple procedure of signal stationarization (i.e differencing) and contributes to an advanced signal pretreatment, which makes bearing non-stationary signals more amenable to efficient analysis by the stationary LTIVAR model. The SSA is used for the purpose of noise cancellation, where the signals are decomposed in a

number of PCs each having a certain portion of the original variance. Then, some of the PCs with the low portion of variance were excluded from the reconstruction of the bearing vibration signal. The coefficients of the LTIVAR model are used to make the feature vectors, which used for fault diagnosis purposes. In this method, four different procedures were proposed for the purpose of the equalisation of the feature vectors. The method is illustrated in chapter 5 while the results and discussion are presented in chapter 5.

8.3 Conclusions

As mentioned in the introductory chapter, there were a number of objectives to be achieved in the present thesis. This section discusses to what extent these objectives have been achieved. Some of the main conclusions are stated below: -

8.3.1 General Conclusions

Development of a data–driven methodology for a complete fault diagnosis in REBs (the primary objective): - As mentioned in Section 1.5, the primary objective of this research is to develop a data–driven methodology for a complete fault diagnosis in REBs. This methodology should be able to detect, identify the location, and then estimate the severity of the faults. The author believes that this primary objective has been achieved. This is because two methodologies were developed in this thesis and both of them were validated using the wide range of experimental data obtained from different test rigs and under different operational conditions. The performance of the methodologies was assessed in terms of fault detection and fault location and fault

estimation and the results obtained showed a very high rate of accurate fault diagnosis (invariably 100%). The methodologies were able to detect the changes in fault location and changes in the fault severities. Both the methodologies were compared to some recent published works, which used other time series analysis techniques. The comparisons were introduced in sections 5.2 & 7.3 and they showed that both methodologies (developed in this thesis) are superior to other methods used in the comparisons. It is believed by the author that the methodologies developed in this thesis contribute fundamentally to the knowledge of fault diagnosis in REBs.

Capabilities of the singular spectrum analysis (specific objective A): - It is shown from the results, presented in chapters 5 & 7, that the singular spectrum has a very good capability for the purposes of rolling element bearing fault diagnosis. It can be used effectively for generating features very suitable for bearing CM (see chapters 4 & 5). It is also used as part of a pre-processing stage for CM purposes (see chapters 6 & 7).

Performance of the methodologies (specific object F): - The methodology performs very well for a wide range of data sets. These data sets were taken from three different bearing test rigs. The data sets covers different shaft rotational speed, different bearings with different fault types and different fault severities. The methodology can accurately detect, identify and quantify the different faults.

More conclusions are specifically made from each methodology and explained in the following sections.

8.3.2 A new methodology for fault diagnosis in rolling element bearings using singular spectrum analysis

- *Specific objective B:* In chapters 4 and 5 a new way of applying SSA by using only the decomposition stage was investigated (see section 1.7). The methodology, presented in chapter 4, offers a new and simple procedure for detection of faults because only signals from healthy bearing category are required and subjected decomposition stage in SSA. All other new signals are only embedded and projected on to the reference space. The simplicity of the methodology can be taken from the point that it uses only signals from the healthy bearing category to build the baseline space by subjecting them only to the decomposition stage. The way that the SSA is used makes it practical. When compared to other SSA based fault diagnosis techniques, which use signals from both healthy and fault bearing categories and then subject each of them to SSA, the present approach is simpler. The simplicity of the procedure in detection of signals has the potential capability for automation.
- From the results presented in chapter 5, projecting the new signals on to the baseline space does not only distinguish between the healthy and faulty categories but also among the different faulty categories as well. This was concluded when it was confirmed that the Mahalanobis distance levels (their range) were changing with the change of fault severity and location. In other words, the methodology is able to detect faults corresponding to different locations (such as those of the inner and outer races) and of different severities

such as those classed as small and large. This can be used as a stepping-stone for the case where information about other bearing conditions is not available.

- A phenomenology is noticed in the FV dimensionality with respect to the correct classification rates for all the levels of fault diagnosis. The increase of FV dimension (an increase in the number of projections involved in the building the baseline space) improves the classification accuracy. This can be explained as incorporating more PCs means involving more information about faults because the PCs contain portions of the original variances.

- The methodology showed a very good performance in terms of fault detection, fault type identification and fault severity estimation. It is shown, in comparison with the some recent published works, that the methodology is superior. It has a 98.3 -100% average correct classification rates.

- It is shown that the Mahalanobis distance level changes when the fault size changes. However, sometimes (as in CS1) the trend of such a change does not follow a monotonic path. In fact, the reasons behind the non-monotonic path followed by the data of CS1, are not investigated in the current study. This might be explained by the fact that the features selected are not able always to behave linearly with the change of fault severity. A further future study might help in further understanding the physical meaning of this.

8.3.3 Bearing fault diagnosis using a novel signal pretreatment and LTIVAR modelling

- *Specific objective C:* - In chapter 7, it is shown that the method proves the importance of the signal pretreatment for the purposes of REB fault diagnosis and for making signals more amenable to analysis by LTIVAR. The developed signal pretreatment improves significantly the precision of the model in representing bearing vibration signals.

- *Specific objective D:* - The method puts stress onto the pre-treatment process and as a result succeeds in developing a rather simple procedure based on simplest form of autoregression (i.e LTIVAR) for the purposes of fault diagnosis. This is considered as a fresh direction in representing non-stationary signals with a stationary autoregressive model (see the 4^{th} *objective*).

- As a result, the combination of the suggested signal pretreatment and the LTIVAR provides a new methodology for complete bearing fault diagnosis. The methodology performs well for all levels of fault diagnosis, namely fault detection, type identification, and severity estimation.

- Several procedures of length equalisation for the feature vectors were tested. Just one of the above procedures had a negative effect on the average correct classification rates. All the other procedures for feature vectors length equalisation result in quite high precision of the diagnosis. However, the zero padding provides the best results.

- The methodology showed a very good performance in terms of fault detection, fault type identification and fault severity estimation. It is shown, in comparison with the some recent published works, that the methodology is superior. It has a 99.8 -100% average correct classification rates.

8.4 Limitations and future works

The thesis is one of the few works, which uses singular spectrum analysis for fault diagnosis in rolling element bearings. Although the approaches developed in this thesis are validated using a wide range of data sets, there are some limitations which can be covered in further future studies. The following summarise some interesting future studies, which could cover such limitations and / or expand the results already achieved.

- *Considering variation of shaft rotational speed:* - In the current thesis, the rotational shaft speed was constant during the acquisition of the signals so investigating the variation of the operating speed would be an interesting topic. This is important as there are many industrial application where the rotational shaft speed are varying during the its operation.

- *Considering other types of fault and catering for more types of bearing:* Another example of a further future direction would be in using data from other types of bearings such as double row groove bearings, and other types of faults such as multiple faults. The bearings considered in this analysis were all single row types so application of the methodologies for cases where bearings are multi-row bearings would potentially be useful. These bearings are usually used

in heavy load applications. The faults considered in the current investigation have all been single faults. However, there are some instances in practice where multiple faults might develop, and which could then be problematic.

- In the current study it was difficult to obtain a real time vibration signal, and the validation was limited to data sets taken from bearing test rigs built in laboratories. Although the author believes that the methodology will work for real time applications, validating methodologies using real industrial environments will still be very useful and will support the generality of the methodologies.

- Further capabilities of SSA can be investigated such as using the SSA for analysing signals in the time-frequency domain. Each PC component can cover or participate in a certain band of frequency. The width of the frequency bands covered by each PC is affected by the size of embedding window. This kind of analysis can help in building a 3D screen for the signal (frequency, amplitude, number of PCs). Then, the PCs, related to a frequency band, can be chosen and used for further analysis.

- Following to the point above, another direction of further research can be the investigation of the SSA window size variations. In this direction, the possible effect of SSA window size on the frequency bands incorporated in each PC can be investigated. In this suggested future work, selection of proper window size can be investigated in more depth, leading towards the possibility of automation.

- The coefficients of the LTIVAR model can be further analysed by factorial analysis. Such kind of analysis can produces some inference about the relations between the bearing faults and some certain coefficients of the LTIVAR model. In such analysis the bearing condition can be indexed as a latent variable and then model coefficients, which have highest loading, can be used as indicators for the presence of that bearing condition.

- Further analysis can be implemented for automation of the selection of the SSA window size and the number of the PCs selected for the reconstruction of signals.

References

[1] A.K. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical systems and signal processing, 20 (2006) 1483-1510.

[2] H.-H. Huang, H. Wang, An integrated monitoring and diagnostic system for roller bearings, The International Journal of Advanced Manufacturing Technology, 12 (1996) 37-46.

[3] R.B. Randall, Vibration-based condition monitoring: industrial, aerospace and automotive applications, John Wiley & Sons, 2011.

[4] R.B. Randall, State of the art in monitoring rotating machinery-part 1, Sound and vibration, 38 (2004) 14-21.

[5] R.B. Randall, State of the art in monitoring rotating machinery-part 2, Sound and Vibration, 38 (2004) 10-17.

[6] B.P. Graney, K. Starry, Rolling element bearing analysis, Materials Evaluation, 70 (2012) 78.

[7] J.E. Shigley, R.G. Budynas, C.R. Mischke, Mechanical engineering design, (2004).

[8] A.J. Kurup, Bearing Life Analysis Along With Sustainability Of Seals.

[9] H. Çalis, A. Çakir, E. Dandil, Artificial immunity-based induction motor bearing fault diagnosis, Turkish Journal of Electrical Engineering & Computer Sciences, 21 (2013).

[10] J.R. Stack, T.G. Habetler, R.G. Harley, Fault-signature modeling and detection of inner-race bearing faults, Industry Applications, IEEE Transactions on, 42 (2006) 61-68.

[11] D. Mba, R.B. Rao, Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating Structures, (2006).

[12] Y.H. Pao, R.R. Gajewski, A.N. Ceranoglu, Acoustic emission and transient waves in an elastic plate, The Journal of the Acoustical Society of America, 65 (1979) 96-105.

[13] D. Mba, Acoustic Emissions and monitoring bearing health, Tribology Transactions, 46 (2003) 447-451.

[14] A. Morhain, D. Mba, Bearing defect diagnosis and acoustic emission, Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 217 (2003) 257-272.

[15] Y. Jia, J. Henao-Sepulveda, M. Toledo-Quinones, Wireless temperature sensor for bearing health monitoring, Smart Structures and Materials, (2004) 368-376.

[16] T. Harvey, R. Wood, H. Powrie, Electrostatic wear monitoring of rolling element bearings, Wear, 263 (2007) 1492-1501.

[17] R.R. Schoen, T.G. Habetler, F. Kamran, R. Bartfield, Motor bearing damage detection using stator current monitoring, Industry Applications, IEEE Transactions on, 31 (1995) 1274-1279.

[18] M. Blödt, P. Granjon, B. Raison, G. Rostaing, Models for bearing damage detection in induction motors using stator current monitoring, Industrial Electronics, IEEE Transactions on, 55 (2008) 1813-1822.

[19] Y. Ohga, K. Moriguchi, S. Honda, H. NAKAGAWA, Fault Diagnosis System for Hydraulic Turbine Generator The transactions of the Institute of Electrical Engineers of Japan. B, A publication of Power and Energy Society, 122 (2002) 492-497.

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

[21] Y.-T. Su, S.-J. Lin, On initial fault detection of a tapered roller bearing: frequency domain analysis, Journal of Sound and Vibration, 155 (1992) 75-84.

[22] N. Tandon, A comparison of some vibration parameters for the condition monitoring of rolling element bearings, Measurement, 12 (1994) 285-289.

[23] B. Samanta, K. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mechanical systems and signal processing, 17 (2003) 317-328.

[24] H. Al-Bugharbee, I. Trendafilova, Fault diagnosis in roller element bearings by using a linear autoregressive model, Proceedings of ISMA 2014 - International Conference on Noise and Vibration Engineering and USD 2014 - International Conference on Uncertainty in Structural Dynamics, (2014) 2765-2776.

[25] H. Al-Bugharbee, I. Trendafilova, Autoregressive modelling for rolling element bearing fault diagnosis, Journal of Physics: Conference Series, 628 (2015) 012088.

[26] H. Al-Bugharbee, I. Trendafilova, A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling, Journal of Sound and Vibration, 369 (2016) 246-265.

[27] I.Y. Onel, M. El Hachemi Benbouzid, Induction motor bearing failure detection and diagnosis: Park and Concordia transform approaches comparative study, Mechatronics, IEEE/ASME Transactions on, 13 (2008) 257-262.

[28] T.A. Harris, M.N. Kotzalas, Essential concepts of bearing technology, CRC press, 2006.

[29] A. Fasteners. "http://www.alexfasteners.com;" (05May).

[30] F.K. AG, Rolling Bearing Damage Recognition of damage and bearing inspection, Publ. No. WL, 82 (2003) 2.

[31] A. Choudhury, N. Tandon, A theoretical model to predict vibration response of rolling bearings to distributed defects under radial load, Journal of vibration and acoustics, 120 (1998) 214-220.

[32] S.M. Pise, S.S. Kulkarni, A Review on Investigation of Effect of Surface Defect on Ball Bearing Vibration Response, (2016).

[33] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, Tribology international, 32 (1999) 469-480.

[34] M.S. Lai, Detection of developing bearing failures by means of vibration analysis, (1990).

[35] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics, Journal of sound and vibration, 302 (2007) 951-961.

[36] T. Williams, X. Ribadeneira, S. Billington, T. Kurfess, Rolling element bearing diagnostics in run-to-failure lifetime testing, Mechanical Systems and Signal Processing, 15 (2001) 979-993.

[37] F. Immovilli, M. Cocconcelli, A. Bellini, R. Rubini, Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals, Industrial Electronics, IEEE Transactions on, 56 (2009) 4710-4717.

[38] N. Jamaludin, D. Mba, Monitoring extremely slow rolling element bearings: part I, NDT & E International, 35 (2002) 349-358.

[39] S. Prabhakar, A. Mohanty, A. Sekhar, Application of discrete wavelet transform for detection of ball bearing race faults, Tribology International, 35 (2002) 793-800.

[40] A.A.T. Zarringhabaei, Development of new fault detection methods for rotating machines (roller bearings), in, Politecnico di Torino, 2015.

[41] R.D. Evans, Classic bearing damage modes, National renewable energy laboratory wind turbine tribology seminar, (2011).

[42] M.N. Kotzalas, G.L. Doll, Tribological advancements for reliable wind turbine performance, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 368 (2010) 4829-4850.

[43] R. Widner, W. Littmann, Bearing damage analysis, Proceedings of the Mechanical Failures Prevention Group, (1976) 67-84.

[44] I.E. Alguindigue, A. Loskiewicz-Buczak, R.E. Uhrig, Monitoring and diagnosis of rolling element bearings using artificial neural networks, Industrial Electronics, IEEE Transactions on, 40 (1993) 209-217.

[45] B. Li, M.-Y. Chow, Y. Tipsuwan, J.C. Hung, Neural-network-based motor rolling bearing fault diagnosis, Industrial Electronics, IEEE Transactions on, 47 (2000) 1060-1069.

[46] M. Li, P. Zhao, The application of wavelet packet and SVM in rolling bearing fault diagnosis, Mechatronics and Automation, 2008. ICMA 2008. IEEE International Conference on, (2008) 505-508.

[47] Z. Zhang, W. Lv, M. Shen, Active learning of support vector machine for fault diagnosis of bearings, in: Advances in Neural Networks-ISNN 2006, Springer, 2006, pp. 390-395.

[48] C. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearings Part II: The use of nearest neighbour classification, Mechanical Systems and Signal Processing, 6 (1992) 309-316.

[49] R. Casimir, E. Boutleux, G. Clerc, A. Yahoui, The use of features selection and nearest neighbors rule for faults diagnostic in induction motors, Engineering Applications of Artificial Intelligence, 19 (2006) 169-177.

[50] I. Howard, A Review of Rolling Element Bearing Vibration:Detection, Diagnosis and Prognosis', in, DTIC Document, 1994.

[51] X.X. Qi, J.W. Ji, X.W. Han, Fault Diagnosis Methods of Rolling Bearing: A General Review, Key Engineering Materials, 480 (2011) 986-992.

[52] C. Nataraj, K. Kappaganthu, Vibration-based diagnostics of rolling element bearings: state of the art and challenges, 13th World Congress in Mechanism and Machine Science, Guanajuato, Mexico, (2011) 19-25.

[53] P. McFadden, J. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, Journal of sound and vibration, 96 (1984) 69-82.

[54] P. McFadden, J. Smith, The vibration produced by multiple point defects in a rolling element bearing, Journal of sound and vibration, 98 (1985) 263-273.

[55] I. Epps, An investigation into vibrations excited by discrete faults in rolling element bearings, (1991).

[56] N. Sawalhi, R. Randall, Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults, Mechanical Systems and Signal Processing, 22 (2008) 1924-1951.

[57] A. Utpat, R. Ingle, M. Nandgaonkar, A Model for Study of the Defects in Rolling Element Bearings at Higher Speed by Vibration Signature Analysis, (2009).

[58] A. Rafsanjani, S. Abbasion, A. Farshidianfar, H. Moeenfard, Nonlinear dynamic modeling of surface defects in rolling element bearing systems, Journal of Sound and Vibration, 319 (2009) 1150-1174.

[59] M. Patil, J. Mathew, P. Rajendrakumar, S. Desai, A theoretical model to predict the effect of localized defect on vibrations associated with ball bearing, International Journal of Mechanical Sciences, 52 (2010) 1193-1201.

[60] D.S. Shah, V.N. Patel, A review of dynamic modeling and fault identifications methods for rolling element bearing, Procedia Technology, 14 (2014) 447-456.

[61] M.I. Friswell, J.E. Mottershead, H. Ahmadian, Finite–element model updating using experimental test data: parametrization and regularization, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 359 (2001) 169-186.

[62] C. Pachaud, R. Salvetat, C. Fray, Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces, Mechanical Systems and signal processing, 11 (1997) 903-916.

[63] J. Dron, F. Bolaers, Improvement of the sensitivity of the scalar indicators (crest factor, kurtosis) using a de-noising method by spectral subtraction: application to the detection of defects in ball bearings, Journal of Sound and Vibration, 270 (2004) 61-73.

[64] P. Jigar, P. Vaishali, P. Amit, Fault diagnostics of rolling bearing based on improve time and frequency domain features using artificial neural networks, Int. J. Sci. Res. Dev, 1 (2013) 781-788.

[65] T. Igarashi, H. Hamada, Studies on the vibration and sound of defective rolling bearings: First report: Vibration of ball bearings with one defect, Bulletin of JSME, 25 (1982) 994-1001.

[66] Y.-C. Choi, Y.-H. Kim, Fault detection in a ball bearing system using minimum variance cepstrum, Measurement Science and Technology, 18 (2007) 1433.

[67] R. Rubini, U. Meneghetti, Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings, Mechanical systems and signal processing, 15 (2001) 287-302.

[68] M. Safizadeh, A. Lakis, M. Thomas, Using short-time fourier transform in machinery diagnosis, Proceedings of the 4th WSEAS International Conference on Electronic, Signal Processing and Control (ESPOCO'05), (2005) 494-200.

[69] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjöberg, J.-O. Strömberg, Towards automatic detection of local bearing defects in rotating machines, Mechanical systems and signal processing, 19 (2005) 509-535.

[70] K. Mori, N. Kasashima, T. Yoshioka, Y. Ueno, Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals, Wear, 195 (1996) 162-168.

[71] K. Medjaher, N. Zerhouni, J. Baklouti, Data-driven prognostics based on health indicator construction: Application to PRONOSTIA's data, Control Conference (ECC), 2013 European, (2013) 1451-1456.

[72] S.J. Taylor, Modelling financial time series, (2007).

[73] W.W.-S. Wei, Time series analysis, Addison-Wesley publ Reading, 1994.

[74] H. Hassani, Singular spectrum analysis: methodology and comparison, (2007).

[75] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Physica D: Nonlinear Phenomena, 35 (1989) 395-424.

[76] R. Vautard, P. Yiou, M. Ghil, Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D: Nonlinear Phenomena, 58 (1992) 95-126.

[77] S. Sanei, H. Hassani, Singular spectrum analysis of biomedical signals, CRC Press, 2015.

[78] W.C. Pereira, S.L. Bridal, A. Coron, P. Laugier, Singular spectrum analysis applied to backscattered ultrasound signals from in vitro human cancellous bone specimens, Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on, 51 (2004) 302-312.

[79] D. Salgado, F. Alonso, Tool wear detection in turning operations using singular spectrum analysis, Journal of Materials Processing Technology, 171 (2006) 451-458.

[80] B. Kilundu, P. Dehombreux, X. Chiementin, Tool wear monitoring by machine learning techniques and singular spectrum analysis, Mechanical Systems and Signal Processing, 25 (2011) 400-415.

[81] B. Muruganatham, M. Sanjith, B.K. Kumar, S. Murty, P. Swaminathan, Inner race bearing fault detection using Singular Spectrum Analysis, Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference on, (2010) 573-579.

[82] B. Kilundu, X. Chiementin, P. Dehombreux, Singular spectrum analysis for bearing defect detection, Journal of Vibration and Acoustics, 133 (2011) 051007.

[83] B. Muruganatham, M. Sanjith, B. Krishnakumar, S. Satya Murty, Roller element bearing fault diagnosis using singular spectrum analysis, Mechanical Systems and Signal Processing, 35 (2013) 150-166.

[84] M. Arnold, X. Milner, H. Witte, R. Bauer, C. Braun, Adaptive AR modeling of nonstationary time series by means of Kalman filtering, Biomedical Engineering, IEEE Transactions on, 45 (1998) 553-562.

[85] S.M. Kay, S.L. Marple Jr, Spectrum analysis—a modern perspective, Proceedings of the IEEE, 69 (1981) 1380-1419.

[86] C. Junsheng, Y. Dejie, Y. Yu, A fault diagnosis approach for roller bearings based on EMD method and AR model, Mechanical Systems and Signal Processing, 20 (2006) 350-362.

[87] J. Zheng, J. Cheng, Y. Yang, Multiscale Permutation Entropy Based Rolling Bearing Fault Diagnosis, Shock and Vibration, 2014 (2014).

[88] I. Howard, A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis', in, DTIC Document, 1994.

[89] Y. Zhan, A. Jardine, Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. Part 1: modeling, Journal of Sound and Vibration, 286 (2005) 429-450.

[90] Y. Zhan, A. Jardine, Adaptive autoregressive modeling of non-stationary vibration signals under distinct gear states. Part 2: experimental analysis, Journal of Sound and Vibration, 286 (2005) 451-476.

[91] G. Wang, Z. Luo, X. Qin, Y. Leng, T. Wang, Fault identification and classification of rolling element bearing based on time-varying autoregressive spectrum, Mechanical Systems and Signal Processing, 22 (2008) 934-947.

[92] B. Harrap, M. Jason, Fault diagnosis of bearings using short data lengths, Proc. of Condition Monitoring and Diagnostic Engineering Management Conference, (1994) 410-417.

[93] D. Baillie, J. Mathew, A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings, Mechanical Systems and Signal Processing, 10 (1996) 1-17.

[94] A. McCormick, A. Nandi, L. Jack, Application of periodic time-varying autoregressive models to the detection of bearing faults, Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 212 (1998) 417-428.

[95] N. Doulamis, A. Doulamis, T. Varvarigou, Adaptable neural networks for modeling recursive non-linear systems, Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on, 2 (2002) 1191-1194.

[96] N. Golyandina, V. Nekrutkin, A.A. Zhigljavsky, Analysis of time series structure: SSA and related techniques, CRC Press, 2010.

[97] M. Ghil, M. Allen, M. Dettinger, K. Ide, D. Kondrashov, M. Mann, A.W. Robertson, A. Saunders, Y. Tian, F. Varadi, Advanced spectral methods for climatic time series, Reviews of geophysics, 40 (2002) 3-1-3-41.

[98] C.W.R.U.B.D.C. Website. (07-03-2014), http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.

[99] A.A. Tabrizi, L. Garibaldi, A. Fasana, S. Marchesielo, Ensemble Empirical Mode Decomposition (EEMD) and Teager-Kaiser Energy Operator (TKEO) Based Damage Identification of Roller Bearings Using One-Class Support Vector Machine, EWSHM-7th European Workshop on Structural Health Monitoring, (2014).

[100] B.J. van Wyk, M.A. van Wyk, G. Qi, Difference histograms: A new tool for time series analysis applied to bearing fault diagnosis, Pattern recognition letters, 30 (2009) 595-599.

[101] M.O. Mustafa, G. Georgoulas, G. Nikolakopoulos, Bearing fault classification based on Minimum Volume Ellipsoid feature extraction, Control Applications (CCA), 2013 IEEE International Conference on, (2013) 1177-1182.

[102] I. Trendafilova, An inverse vibration-based approach towards modelling and damage identification in nonlinearly vibrating structures. Application for delamination detection in a composite beam, Journal of Physics: Conference Series, 382 (2012) 012030.

[103] D. Garcia, R. Palazzetti, I. Trendafilova, C. Fiorini, A. Zucchelli, Vibration-based delamination diagnosis and modelling for composite laminate plates, Composite Structures, 130 (2015) 155-162.

[104] P.M. Broersen, Facts and fiction in spectral analysis, Instrumentation and Measurement, IEEE Transactions on, 49 (2000) 766-772.

[105] H. Mader, Statistics in volcanology, Geological Society of London, London, 2006.

[106] D. Kwiatkowski, P.C. Phillips, P. Schmidt, Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, Journal of econometrics, 54 (1992) 159-178.

[107] R.J. Hyndman, G. Athanasopoulos, Forecasting: principles and practice, OTexts, 2014.

[108] A. Babu, S. Reddy, Exchange Rate Forecasting using ARIMA, Neural Network and Fuzzy Neuron, Journal of Stock & Forex Trading, 2015 (2015).

[109] L. Mangin, M.-N. Fiamma, C. Straus, J.-P. Derenne, M. Zelter, C. Clerici, T. Similowski, Source of human ventilatory chaos: lessons from switching controlled mechanical ventilation to inspiratory pressure support in critically ill patients, Respiratory physiology & neurobiology, 161 (2008) 189-196.

[110] B. Cernat-Gruici, Some empirical evidence on the relationship between acquisitions, disposals and the stock market, Romanian Economic Journal, 12 (2009) 3-22.

[111] X.H. Cai, Time Series Analysis of Air Pollution CO in California South Coast Area with Seasonal ARIMA model and VAR model, in: Statistics, University of California

Los Angeles, 2008.

[112] G.E. Box, G.M. Jenkins, G.C. Reinsel, Time series analysis: forecasting and control, John Wiley & Sons, 2013.

[113] C.-C. Wang, Y. Kang, P.-C. Shen, Y.-P. Chang, Y.-L. Chung, Applications of fault diagnosis in rotating machinery by using time series analysis with neural network, Expert Systems with Applications, 37 (2010) 1696-1702.

[114] H. Akaike, Statistical predictor identification, Annals of the Institute of Statistical Mathematics, 22 (1970) 203-217.

[115] R.C.G. J.T.Tou, Pattern Recognition Principles, Addison-Wesley Publishing Company, Reading, Massachusetts, 1974.

Appendix 4-A

Tables of number of sub-signals used in the analysis

Table shows the number of sub-signals (i.e number of FVs) and the number of revolution at each sub-signal length for all the case studies used in the analysis. The Nr represents the number of the revolutions in a specific sub-signal length.

Speed (rpm)	Sub signal length							
	512		1024		2048		4096	
	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*
1730	960	1.23	480	2.46	240	4.92	120	9.84
1750	960	1.24	480	2.48	240	4.97	120	9.95
1772	960	1.26	480	2.52	240	5.04	120	10.08
1797	960	1.27	480	2.55	240	5.11	120	10.22

Table 4-A.1: The number of the FVs and the shaft revolutions contained in each sub-signal length CS1.

Speed (rpm)	Sub signal length							
	512		1024		2048		4096	
	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*
250	1280	0.17	640	0.35	320	0.71	160	1.42
750	1280	0.53	640	1.06	320	2.13	160	4.26
1250	1280	0.88	640	1.7	320	3.55	160	7.11

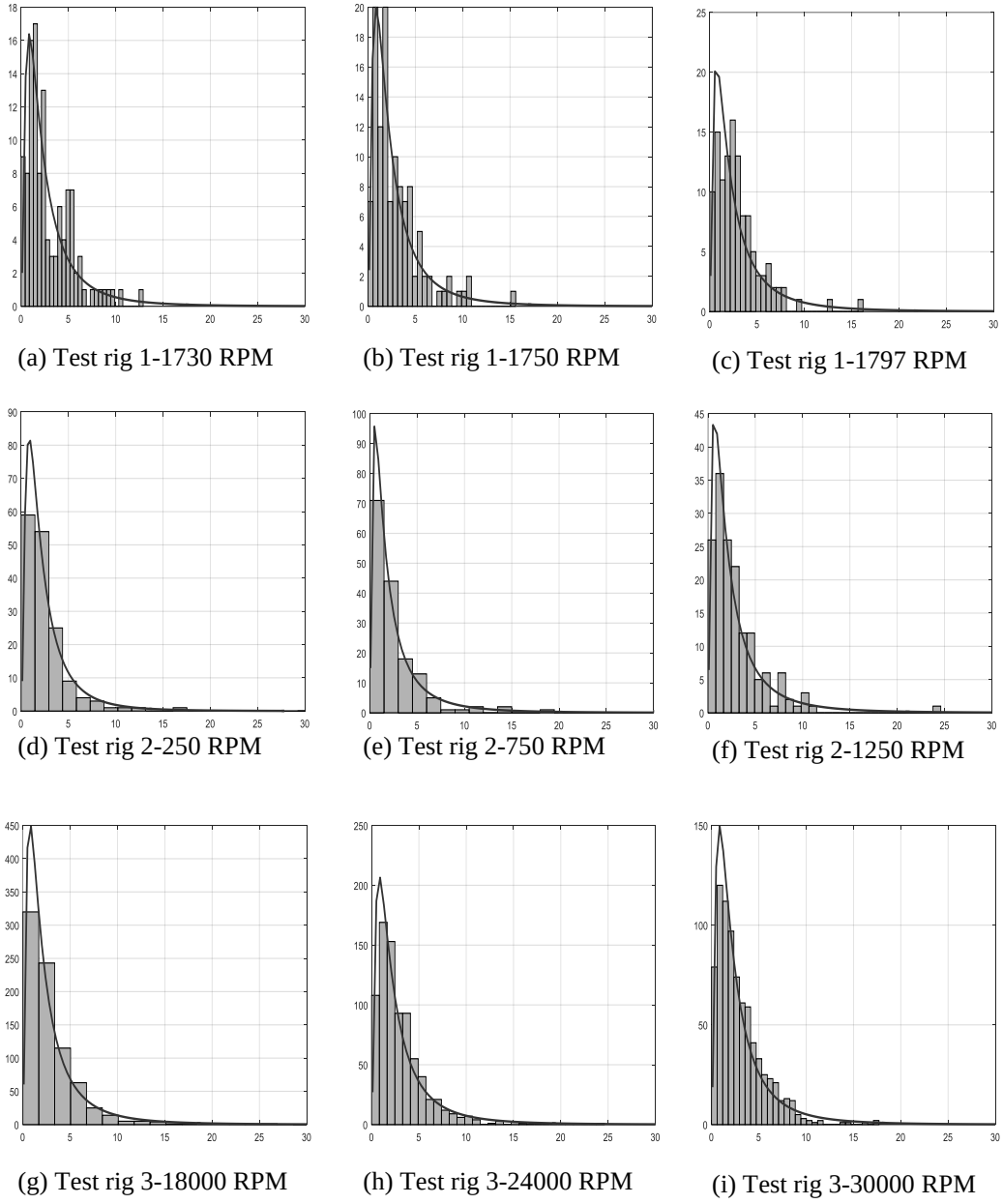
Table 4-A.2 : The number of the FVs and the shaft revolutions contained in each sub-signal length CS2

Speed (rpm)	Sub signal length							
	512		1024		2048		4096	
	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*	No. of FVs	*Nr*
18000	4800	1.5	2400	3	1200	6	600	12
24000	4800	2	2400	4	1200	8	600	16
30000	4800	2.5	2400	5	1200	10	600	20

Table 4-A.3 : The number of the FVs and the shaft revolutions contained in each sub-signal length CS3

Appendix 5-B

Histograms and Lognormal distribution density functions for the healthy training samples data of the test rigs- fault detection phase



(a) Test rig 1-1730 RPM (b) Test rig 1-1750 RPM (c) Test rig 1-1797 RPM

(d) Test rig 2-250 RPM (e) Test rig 2-750 RPM (f) Test rig 2-1250 RPM

(g) Test rig 3-18000 RPM (h) Test rig 3-24000 RPM (i) Test rig 3-30000 RPM

Figure 5-B.1: Examples of the histograms and lognormal distribution density functions for the healthy training samples data of the test rigs

Appendix 5-C

Tables of percentages of correct classification rates for the fault diagnosis – detection and fault type identification phases

Appendix 5-C Tables of percentages of correct classification rates for the fault diagnosis – detection and fault type identification phases

Average and lowest correct classification rates- detection phase – CS2 and CS3

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
250	**5**[98,100]	**5**[97.5,100]	**5**[97.5,100]	**4**[90,100]
750	**5**[100,100]	**5**[98.8,100]	**2**[100,100]	**1**[100,100]
1250	**3**[100,100]	**2**[100,100]	**2**[100,100]	**1**[100,100]

Table 5-C.1: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS2- fault detection phase

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
18000	**7**[99.5,87.5]	**4**[99.2,100]	**4**[100,100]	**3**[100,100]
24000	**10**[99,87.3]	**5**[99.5,100]	**3**[99.5,100]	**1**[100,100]
30000	**10**[98.8 67.3]	**7**[98.5,99]	**6**[97.5,100]	**3**[100,100]

Table 5-C.2: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS3-fault detection phase

Average and lowest correct classification rates- fault identification phase – CS2 and CS3

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
250	**10**(88.9, 65]	**10**[94.4,82.5]	**10**[99.4,97.5]	**3**[98.75,95]
750	**10**[93.8, 80.6]	**6**[97.8,92.5]	**6**[100,100]	**4**[100,100]
1250	**3**[94.8, 88.8]	**7**[99.1,97.5]	**2**[100,100]	**2**[100,100]

Table 5-C.3: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS2- fault type identification phase

	Sub-signal length			
Speed RPM	**512**	**1024**	**2048**	**4096**
18000	**10**[98.9,97.9]	**9**[100,100]	**5**[100,100]	**4**[100,100]
24000	**10**[98.9,97.9]	**6**[99.7,99.5]	**6**[100,100]	**3**[100,100]
30000	**10**[98.7,96.6]	**10**[100,100]	**10**[100,100]	**4**[100,100]

Table 5-C.4: The minimum baseline space dimension and corresponding average and lowest correct classification rates CS3-fault type identification phase

Appendix 7-A

Additional results of Chapter 7- Bearing fault diagnosis using a novel signal pretreatment an LTIVAR modelling

Appendix 7-A Tables PCs number selected for reconstruction of the signals

Number of PCs to achieve 90% of original sub-signal variance to be recovered in the reconstructed sub-signal

Speed (RPM)			***n*=512**	
	H	**IRF**	**BF**	**ORF**
1730	19	36	26	22
1750	18	40	24	23
1772	17	51	25	21
1797	15	43	28	21
			***n*=1024**	
1730	20	37	27	23
1750	19	40	25	23
1772	17	42	26	22
1797	15	44	29	22
			***n*=2048**	
1730	21	37	27	23
1750	19	40	25	23
1772	18	42	26	22
1797	16	44	29	22
			***n*=4096**	
1730	21	37	27	23
1750	19	40	25	23
1772	18	42	26	22
1797	16	44	29	22

Table 7-A.1: the number of PCs selected for the reconstruction stage CS1

Speed (RPM)			***n*=512**	
	H	**IRF**	**BF**	**ORF**
250	22	26	21	26
750	25	28	32	34
1250	20	32	34	30
			***n*=1024**	
250	24	28	23	28
750	27	31	36	37
1250	21	33	36	37
			***n*=2048**	
250	24	30	24	29
750	28	32	37	38
1250	21	34	39	32
			***n*=4096**	
250	24	31	25	29
750	28	32	37	39
1250	21	34	40	33

Table 7-A.2: the number of PCs selected for the reconstruction stage CS2

Speed (RPM)		***n*=512**	
	H	**IRF**	**BF**
18000	62	61	62
24000	59	50	58
30000	54	48	60
		***n*=1024**	
18000	68	64	67
24000	64	53	63
30000	59	51	65
		***n*=2048**	
18000	70	65	68
24000	66	54	65
30000	61	52	67
		***n*=4096**	
18000	71	66	69
24000	68	54	66
30000	62	52	68

Table 7-A.3: the number of PCs selected for the reconstruction stage CS3

The LTIVAR mode optimum order of the reconstructed sub-signals

Speed (RPM)			***n*=512**	
	H	**IRF**	**BF**	**ORF**
1730	5	12	10	6
1750	5	12	8	6
1772	5	12	8	6
1797	5	12	10	5
			***n*=1024**	
1730	5	12	10	8
1750	5	12	8	5
1772	5	12	10	5
1797	5	12	10	5
			***n*=2048**	
1730	6	15	10	6
1750	5	12	8	6
1772	5	12	5	5
1797	5	12	10	5
			***n*=4096**	
1730	5	12	8	5
1750	5	12	5	5
1772	5	12	5	5
1797	5	12	10	5

Table 7-A.4: the number of PCs selected for the reconstruction stage CS1

Speed (RPM)			***n*=512**	
	H	**IRF**	**BF**	**ORF**
250	8	10	7	10
750	5	13	10	10
1250	8	10	10	10
			***n*=1024**	
250	8	10	10	10
750	5	10	10	10
1250	6	10	10	10
			***n*=2048**	
250	5	10	10	10
750	5	10	10	10
1250	5	15	10	10
			***n*=4096**	
250	10	8	8	10
750	5	10	10	10
1250	7	13	10	10

Table 7-A.5: the number of PCs selected for the reconstruction stage CS2

Appendix 7-A Tables PCs number selected for reconstruction of the signals

Speed (RPM)		***n*=512**	
	H	**IRF**	**BF**
18000	25	25	25
24000	30	20	30
30000	20	20	20
		***n*=1024**	
18000	25	20	25
24000	30	20	20
30000	20	20	20
		***n*=2048**	
18000	30	30	30
24000	30	22	30
30000	30	20	30
		***n*=4096**	
18000	28	28	28
24000	30	20	20
30000	24	20	22

Table 7-A.6: the number of PCs selected for the reconstruction stage CS3

NMSE improvement

Speed (RPM)	**Bearing condition**	**Sub-signal length(*n*)**			
		512	**1024**	**2048**	**4096**
250	**H**	11.89%	14.38%	16.94%	13.1%
	IRF	5.23%	5.5%	5.57%	6.2%
	BF	6.74%	6.78%	7.67%	7.62%
	ORF	5.81%	6.47%	6.51%	6.51%
750	**H**	7.1%	7.73%	7.92%	6.98%
	IRF	6.28%	5.17%	5.42%	5.96%
	BF	8.83%	9.73%	9.07%	8.77%
	ORF	7.8%	8.72%	9.13%	9.69%

1250	**H**	2.32%	3.46%	4.2%	2%
	IRF	5.18%	5.08%	6.24%	7.02%
	BF	7%	7.71%	7.96%	8.06%
	ORF	7.27%	6.9%	7.71%	11.73%

Table 7-A.7: Improvement of NMSE due to signal pretreatment CS2

Speed (RPM)	**Bearing condition**	**Sub-signal length(*n*)**			
		512	**1024**	**2048**	**4096**
18000	**H**	54.29%	37.29%	37.21%	61.44%
	IRF	27.5%	26.69%	44.85%	41.16%
	BF	52.63%	36.98%	42.12%	63.19%
24000	**H**	38.24%	37.13%	64.35%	40.05%
	IRF	27.16%	49.17%	66.36%	34.47%
	BF	35.97%	44.53%	71.56%	43.5%
30000	**H**	38.21%	68.18%	39.31%	40.96%
	IRF	50.41%	68.28%	54.89%	48.21%
	BF	54.09%	74.13%	40.19%	47.89%

Table 7-A.8: Improvement of NMSE due to signal pretreatment CS3