Acoustic Emission Monitoring of Wind Turbine Bearings

Jack R. Naumann

A thesis submitted to the University of Sheffield for the degree of Doctor of Philosophy in the faculty of Engineering

Leonardo Centre for Tribology
Department of Mechanical Engineering
The University of Sheffield

January 2016

Abstract

Climate change, national and international targets and a possible impending fuel scarcity is driving the need for a clean, cheap and sustainable energy source. Although onshore wind is currently the most economically viable source of renewable energy, high failure rates often occurring many years prior to their design life, are increasing the cost through additional maintenance and downtime. The gearbox, and in particular the bearings within the gearbox are the components responsible for the largest proportion of downtime. Field studies have shown damage to the inner race of the planetary support bearings found in the epicyclic stage of the gearbox is restricted to an arc of approximately 40° centered on the point of maximum load. Engineers at Ricardo have designed an actuation system to overcome this problem which allows the raceway to be rotated periodically thus distributing damage and increasing bearing life.

The monitoring of planetary support bearings typically found in the epicyclic stage of wind turbine gearboxes has been investigated in this thesis using acoustic emission technology due to its reported increase in sensitivity in detecting damage at low speeds compared to vibration analysis in addition to its ability to locate damage. Primarily, accelerated life tests were performed on a rolling element bearing seeded with a defect mounted in a bespoke full scale test rig designed to mimic loading conditions experienced by the planetary support bearings. In addition, data was recorded and analysed from a split bearing test rig and the high speed shaft bearing of a recommissioned 600kW wind turbine gearbox.

Initial experiments considered the influence of the lubrication regime on the measured acoustic emission signal. It was found that as the oil film reduces, asperity contact, typical of mixed or boundary lubrication, manifests itself as high amplitude transient events. Typical measures of bearing health in the time domain, such as peak values or kurtosis, become unreliable and demonstrates a need for a novel approach. Previous investigations into the use of acoustic emission for the purpose of bearing condition monitoring has focused on instances where full separation between bearing components occurred whereas this work considers a mixed lubrication regime. To overcome the drawbacks of the traditional measures, this work has investigated a process employing wavelet packet decomposition, autocorrelation and cepstrum to reduce the noise and boost the periodicity of a signal from a defected bearing. Outlier analysis was shown to be able to determine the presence of a seeded defect and indicate which bearing component is defected. Such approach was shown to provide a more robust measure than time domain methods. In contrast, this approach was compared to one employing time domain measures for a fully lubricated split bearing. In this case, a time domain approach was more successful at determining the presence of damage than the approach taken for the partially lubricated bearing.

An attempt was made to improve the localisation of defects on a bearing which had, until now, relied on analytical time-of-flight methods. In this work artificial sources, rather than those resulting from a rolling element impinging on a defect, were generated by a standardised pencil lead break and used as training data for two methods namely Delta T mapping and neural networks. The neural networks in particular were shown to reduced the average error from 42mm to 17mm however given the time consuming nature of generating the training data a decision must be made regarding the relative importance of accuracy and ease of implementation.

Contents

Acknowledgements

First and foremost I would like to thank my supervisors, Professor Robert Dwyer-Joyce and Dr. Matthew Marshall, for providing me with the opportunity to work on this project.

This project would have never existed had it not been for the innovation and ingenuity shown by Professor Jonathan Wheals in designing the Multilife™ bearing. Further thanks must be extended to Pierre Guern, who designed the Multilife™ test rig, Hugh Freeman, who was always on hand when things were not going as planned and everyone else at Ricardo was involved on the project throughout the years.

I would also like to thank all of those who have selflessly given their time to helping this project in the last four years in particular to Dave Butcher, Dr. Robin Mills, Chris Grigson and Mike Rennison.

Finally, I reserve my biggest thanks to my colleague and good friend, Tom Howard, whose clear thinking, hard work, organisation and company made the project a lot more successful than it would have been in his absence.

List of Figures

List of Tables

List of Abbreviations

AE	Acoustic Emission
AIC	Aikaike Information Criterion
ALE	Adaptive Line Enhancer
ANN	Artificial Neural Network
ASME	American Society of Mechanical Engineers
CCGT	Combined Cycle Gas Turbine
CDF	Cumulative Distribution Function
CWT	Continuous Wavelet Transform
DEA	Danish Energy Agency
DWT	Discrete Wavelet Transform
ECN	Energy Centre of the Netherlands
EDM	Electrodischarge Machining
EHL	Elastohydrodynamic Lubrication
EI	Energy Index
EU	European Union
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FRP	Fibre Reinforced Plastic
FTC	First Threshold Crossing

H-N	Hsu-Nielsen
HRFT	High Frequency Resonance Technique
IE	Information Entropy
ISO	International Organization for Standardization
K-S Test	Kolmogorov-Smirnov Test
LCOE	Levelised Cost of Energy
MARSE	Measured Area of the Rectified Signal Envelope
MRET	Mandatory Renewable Energy Targets
MSE	Mean Squared Error
O & M	Operation and Maintenance
NDT	Non-Destructive Testing
PAC	Physical Acoustics Corporation
PCI	Peripheral Component Interconnect
PSB	Persistent Slip Band
PV	Photovoltaic
RMS	Root Mean Square
RUL	Remaining Useful Life
RVM	Relevance Vector Machine
SNR	Signal to Noise Ratio
SPM	Shock Pulse Metering
STEF	Short Time Energy Function
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TDB	Top Dead Bottom
TOA	Time of Arrival

WEA	White Etching Area
WPD	Wavelet Packet Decomposition
WSF	White Structure Flaking

Chapter 1

Introduction

1.1 Background

By the end of 2011, global installations of wind turbines had reached a total capacity of 238GW [1] representing an annual market growth of 20% over the previous year. Despite the fact that annual growth over the prior ten years averaged 28% and projected annual growth in the period 2012-2016 is expected to fall further to 16% [1], the wind industry continues to be a growing sector amidst a global economic crisis as shown in Figure 1.1. Such growth can be attributed to an increase in global demand for electricity, an impending scarcity of fossil fuels, environmental concern, regional targets for renewables production, renewed reservations about the safety of nuclear power and most significantly to investors: falling costs and increased revenue.

Estimates have suggested the global demand for energy in 2040 is likely to be 30% higher than in 2010 [2]. Any increases in efficiency and investment in energy efficient equipment will be more than offset by population growth, rising disposable incomes and a population shift towards warmer climates resulting in an increased use of cooling equipment [3]. Such increase in demand is likely to coincidence with peak oil and peak gas, although when this is expected to occur is the source of much debate and uncertainty. Some studies believe oil production could peak before 2030 [4] whereas others suggest global demand will continue to be met well into the 21st century [5, 6]. Mohr investigated a number of possible scenarios and reported that natural gas production will peak between 2025 and 2066 [7]. Despite the lack of

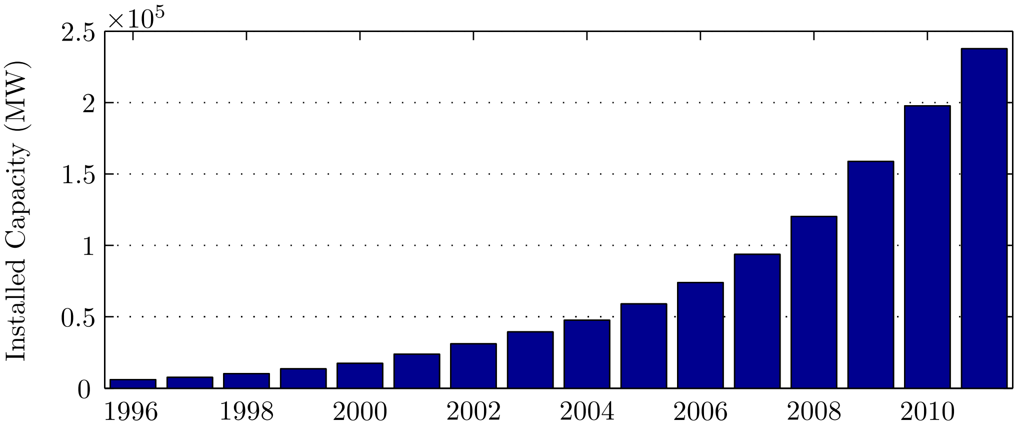

Figure 1.1: Global Cumulative Installed Wind Capacity (1996-2011) [1]

consensus regarding when peak oil and gas will occur, there is a clear agreement that resources are finite. This has led to a number of countries and regions imposing mandatory renewable energy targets (MRETs).

The 2009 European Union (EU) Renewables Directive requires 20% of the EU member states' and 15% of the UK's gross energy consumption to be produced from renewable sources by 2020 [8]. This would require the total renewable capacity in the UK to increase from 6.45GW as it stood at the end of 2011 [9] to roughly 33GW: some 35-45% of the UK's total electricity production. The vast majority of such an increase would be acheived through the installation of onshore and offshore wind. Similar MRETs have been implemented in, amongst other countries, Australia [10], Japan [11] and certain states in the USA [12]. The likelihood of these targets being met is somewhat dependent upon the cost of wind energy in the coming years.

Between the early 1980s and the early 2000s, the capital cost of wind energy declined significantly. The average capital costs in the United States and Denmark in 2004 were approximately 65% and 55% lower than costs during the early 1980s respectively [13, 14]. The combination of falling capital costs and an increase in turbine performance resulted in a dramatic reduction of the levelised cost of energy (LCOE) over the same period. The LCOE, usually measured in cents/kWh, is defined as the cost required for a project to break even over its lifetime and therefore includes the initial capital, cost of operation and maintenance and fuel in addition to discount rates. A report published by the Danish Energy Agency (DEA) estimates the LCOE of onshore wind had fallen by a factor of four by the late 1990s [15]. With the exception of a period between 2004 and 2009 during which an increase in cost occured due to turbine upscaling and increases in material prices, energy prices and labour [16], the downward trend has since continued and is expected to fall further. Lantz et al [17] collated a number of recent studies covering 18 possible scenarios estimating cost reductions through to 2030. Each scenario is shown in Figure 1.2 where each line represents an estimate of the LCOE of wind energy over time. All but one scenario estimated a decrease to less than 88% of the LCOE in 2010, the majority of which estimating a reduction between 20 and 30%. A number of factors, including manufacturing improvements, reduced component loading and cost effective scaling, will determine whether such reductions are achievable.

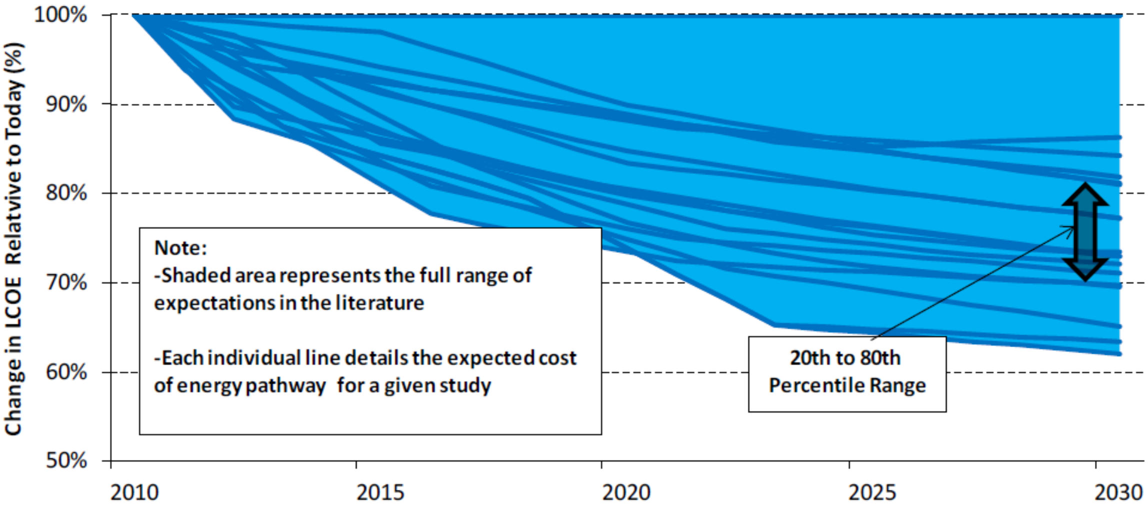

Figure 1.2: Estimated Wind LCOE Projections [17]

Capital expenditure dominates the overall cost of both onshore and offshore wind but operation and maintenance (O&M) costs are not insignificant and are likely to influence future investment. It has been reported that O&M costs in the UK account for approximately 15% of the LCOE for both onshore and offshore wind [18] and 10% in the US [19]. However this is subject to a high degree of uncertainty, particularly for offshore wind since there are few large offshore wind farms that have been in operation for a significant period of time [20]. Large maintenance tasks requiring access to the nacelle of the turbine, such as replacing a gearbox, are both time consuming and expensive onshore but the cost is much higher in hostile offshore environments. The Energy Centre of the Netherlands (ECN) reports that major lifting operations can increase cost up to five to ten times when performed offshore compared to onshore [20]. There is therefore a compound effect of these costs plus the additional downtime due to reduced accessibility which can be as low as 15% in the most hostile offshore environments. Figure 1.3 shows the LCOE for a range of renewable and conventional energy sources taken from a report published by the Fraunhofer Institute [21]. It can be seen that the cheapest onshore wind at 0.045€/kWh, usually from turbines situated in the best wind locations, is marginally more expensive than the cheapest brown coal at 0.038€/kWh. Any modifications that can improve reliability and reduce the O&M costs will make wind energy a more attractive proposition for investors and potentially the cheapest source of energy in the case of onshore wind.

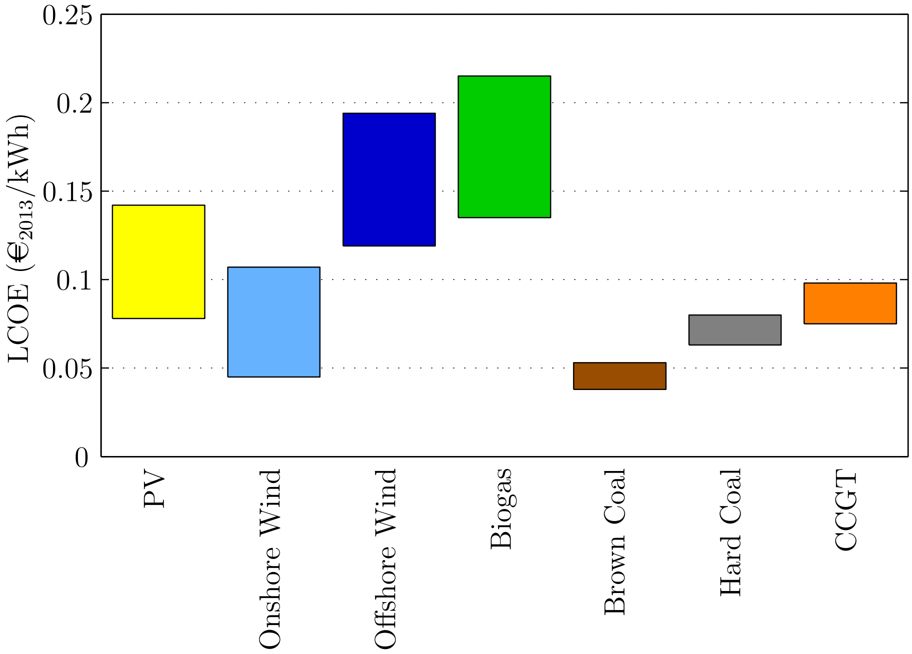

Figure 1.3: LCOE of Renewable and Conventional Energy Sources [21]

1.2 Wind Turbine Failures

In most cases, wind turbines are designed to operate for 20 years [22]. However, since wind energy is a relatively new technology it has not yet been possible to provide an accurate life expectancy for larger turbines which have, in general, accumulated fewer operational years. Reliability is often less than expected since operational conditions and loading patterns are not yet fully understood. Since the 1980s the average capacity of wind turbines has gradually increased. By means of illustration, the average capacity of a single installed turbine in the US doubled from 0.9MW to nearly 1.8MW between 2001 and 2010 (Figure 1.4) [16] while the current largest turbines are rated at 8MW. If this trend continues, failure rates may also increase before the technology has sufficient time to mature. A study of failure rates of more than 1500 turbines from the German '250MW Wind' Program found that larger turbines (>1MW capacity) required significantly more unplanned maintenance activities compared to smaller turbines (Figure 1.5) [22]. Similarly, Spinato et al. [23] reported higher failure rates for larger turbines. For example, in the period 1993-2004, the Vestas V27 turbines with a rated power of 225kW experienced an average failure frequency of 1yr^{-1} whereas the failure frequency of Tacke TW 1.5s turbines rated at 1.5MW increased to approximately 3.5yr^{-1}.

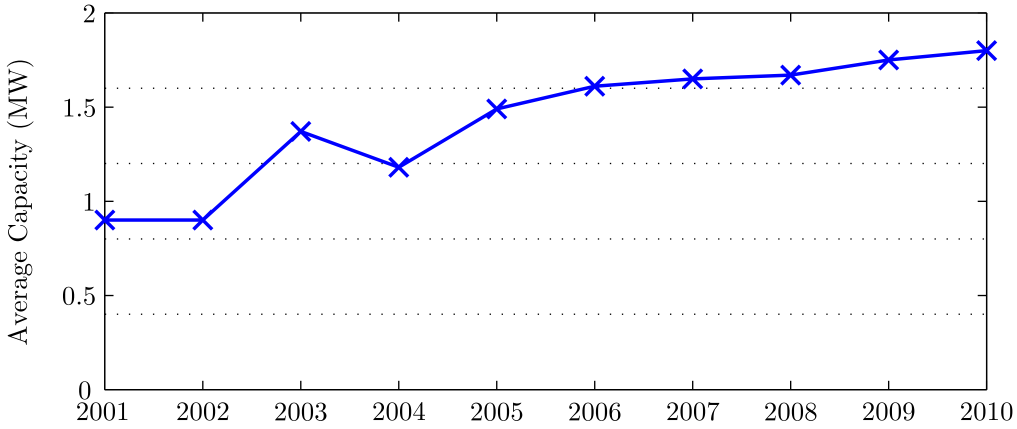

Figure 1.4: Average Turbine Capacity (2001-2010) [16]

A study of wind turbine failures in Sweden between 1997 and 2005 categorised failures by sub-assembly [24]. Figure 1.6 shows the percentage each component contributes to the total number of failures and the corresponding proportion of total downtime. Whilst the gearbox experiences comparatively moderate failure rates the downtime required to carry out a repair or replacement is higher than that of any other single subsystem. More recently, GCube, the leading provider of renewable energy insurance published a report into the most common failures, reasons for failure and claim costs covering all reported wind turbine claims from the US in 2012 [25]. It was found that the 35.1% of claims related to the gearbox, second only to the turbine blades. The success and viability of wind energy in the years to come will be somewhat dependent upon increasing gearbox reliability and this challenge forms the basis and motivation for the following thesis.

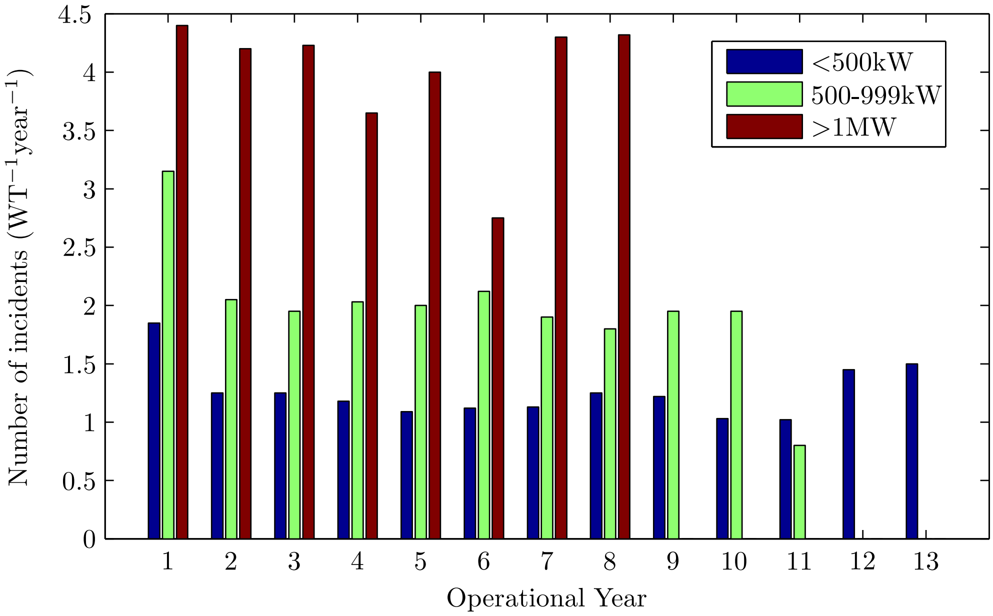

Figure 1.5: Annual Wind Turbine Incidents by Rated Power [22]

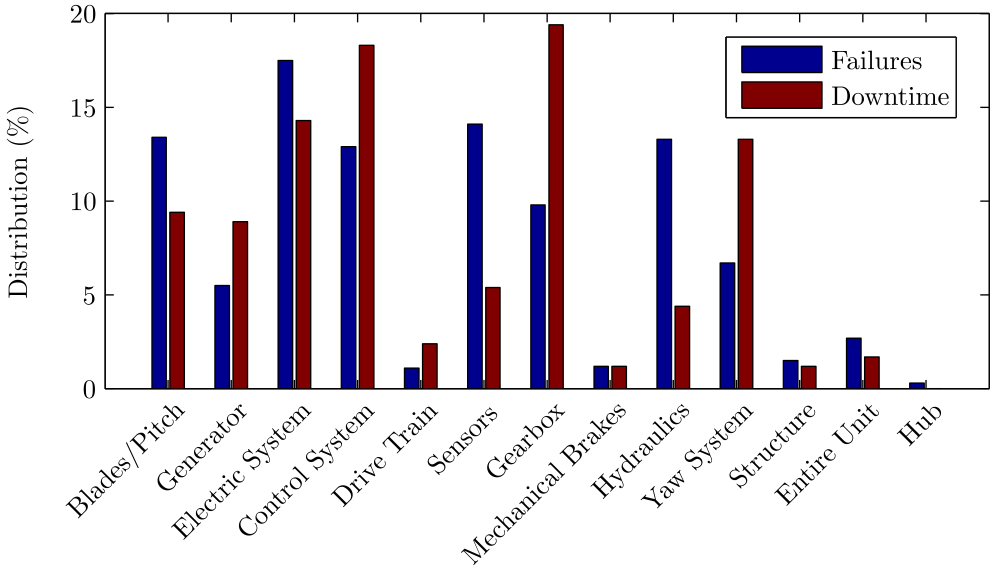

Figure 1.6: Distribution of Failures and Downtime by Sub-assembly [24]

1.3 Wind Turbine Gearboxes

The compound effect of higher than expected failure rates, long periods of downtime and the fact that the gearbox is one of the most expensive components in a turbine is adding to the LCOE. Additionally, manufacturers are adding contingencies to sale prices in order to cover warranties protecting against gearbox failure [26]. It has been reported that German insurers, Allianz received around 1000 claims for damaged gearboxes in 2006 alone [27] while GCube report that the average gearbox claim in 2012 stood at $240 000 [25]. In order to help bring the cost of wind energy down, the causes of failure must be understood so remedial action can be taken. A number of observations have been presented by Musial et al [26] regarding this problem. These are summarised as follows:

- The majority of reported issues are common to all models and manufacturers. As the industry has matured and reliability increased, gearbox designs have converged to a similar architecture indicating that failures generally relate to the generic design rather than manufacturing or quality issues.
- Most wind turbine gearbox failures are due to failures of particular bearings in spite of being designed to the best bearing practices available. The resultant debris or increased clearances may then cause secondary damage to gear teeth.
- Approximately 10% of failures relate to manufacturing or quality issues supporting the evidence that this is not the primary cause of the problem.

Table 1.1 provides a breakdown of gearbox failures by sub-component of nearly all Swedish wind turbines during the period 1997-2005 [24]. Bearing failures account for nearly half of the total failures. Approximately half of the cases where the component was not specified, a serious failure had occurred resulting in a complete gearbox replacement. Many of these failures may have been initiated in the bearings. The average downtime for a bearing failure was 562 hours representing a significant period of lost revenue.

According to Musial et al [26], there are three critical bearing locations which experience high failure rates irrespective of gearbox manufacturer, model or size:

- Planetary bearings
- Intermediate shaft-locating bearings

Table 1.1: Gearbox Subcomponent Failures [24]

Component	Number of Failures	Average Downtime (hrs)
Bearings	41	562
Gearwheels	3	272
Shaft	0	0
Sealing	8	52
Oil System	13	26
Not Specified	44	230

- High speed shaft bearing

The location of these bearings is shown in Figure 1.7 taken from [28]. The rotor shaft rotates the low speed epicyclic planetary gears between 5 and 22rpm. The central sun pinion drives the intermediate shaft which in turn increases the rotational speed of the high speed shaft via two parallel gears to 1000-1600rpm required for the generator.

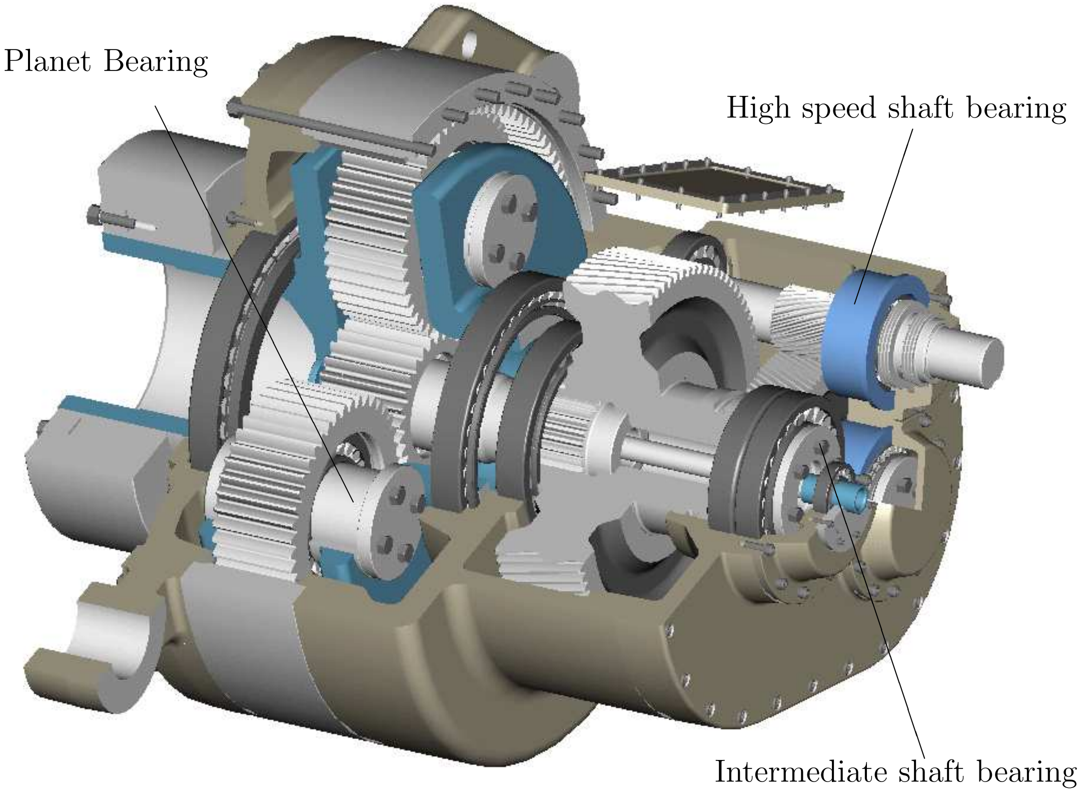

Figure 1.7: Typical Wind Turbine Gearbox [28]

On closer inspection of planetary bearing failures, it has been found that damage is often limited to a localised arc on the inner raceway [29]. This is believed to be due to the uni-directional loading and non-conformal contact which is exacerbated by overload events occurring during emergency stops or periods of high turbulence. It is believed that this leads to a series of indentations on the raceway at the spacing of the rollers which act as stress raisers, thus initiating fatigue damage across the loaded zone of the bearing. Such damage can be seen on the inner raceway of the bearing shown in Figure and 1.8 taken from the epicyclic stage of a decommissioned V42 Scottish Power gearbox.

Figure 1.8: Damaged planetary inner race from a decommissioned V42 gearbox

1.4 Motivation

The nature of the localised damage on the inner raceway of the planet bearings intuitively calls for an indexing solution which can rotate the damage away from the loaded region. Engineers at Ricardo have designed such a mechanism which allows damage to be distributed around the inner raceway [29]. The MultiLife™ bearing is a series of hydraulic actuators which can be fitted onto the stationary shaft of a planet bearing allowing the race to be rotated. Schematics of the first prototype design are shown in Figure 1.9 and Figure 1.10. To give an indication of the size of the device, the diameter of the outside surface of the outer raceway is 400mm and the bore diameter is 220mm.

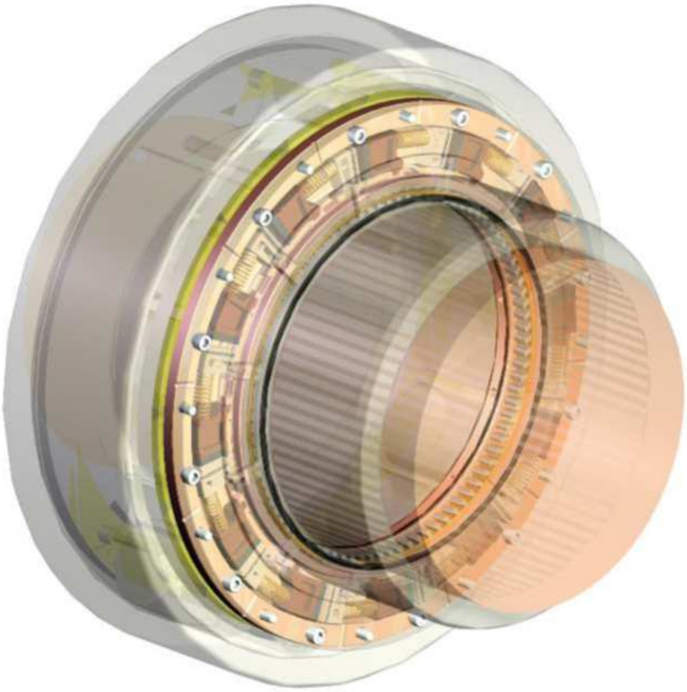

Figure 1.9: MultiLife™ Assembly [29]

Twelve chambers, each housing a piston contributing to the overall torque, are located within a casing fixed to the inner race. When the chambers are pressurised the pistons rotate the casing and raceway relative to the shaft in 5° steps. A 10bar oil supply is required to generate the 500Nm torque necessary to rotate the raceway. The mechanism has been designed in order allow retrofitting to existing gearboxes and can index a number of raceways on a single shaft assuming accessibility allows them to be connected.

The success of the MultiLife™ concept will be largely dependent upon the operating strategy employed; that is, how often and to what extent indexing will take place. Such a strategy must be optimised to ensure damage is rotated away from the loaded region before reaching a critical level and a maximum increase in bearing life is attained. Prior to experimental testing, an analytical model was developed to provide initial estimations of potential increases in life [29]. The model applied the ISO standard L_{10} bearing life equation for a bearing with a dynamic load rating C subjected to a combined radial and axial load P (Equation (1.1)) on an elemental basis by dividing

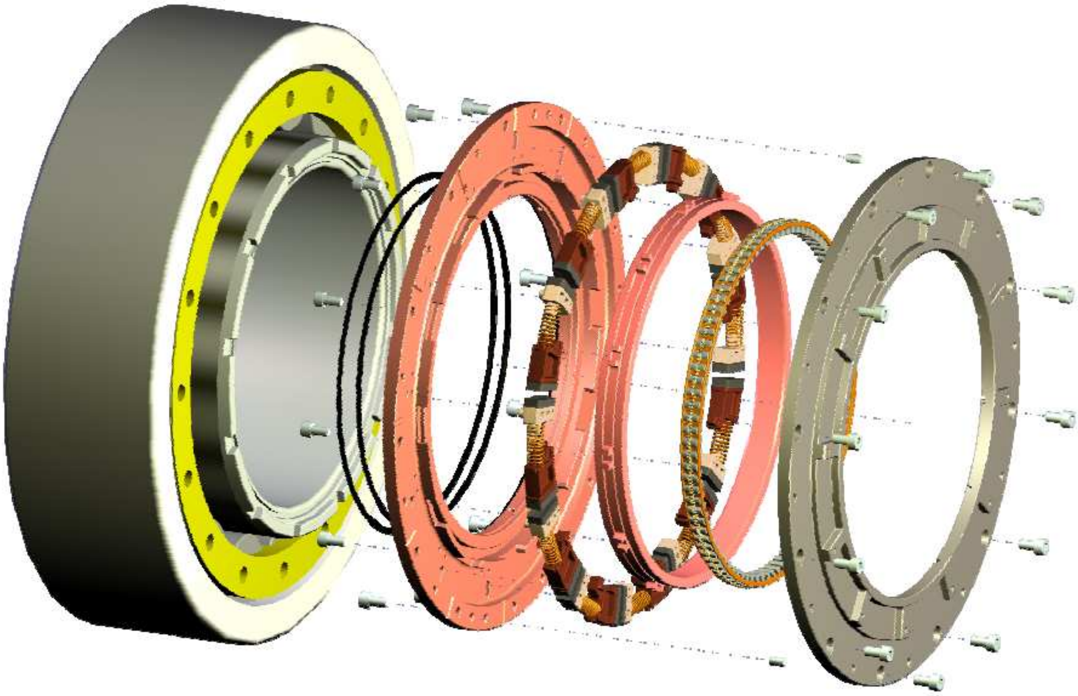

Figure 1.10: MultiLife™ Components [29]

the inner race into 360 arcs of equal width (Equation (1.2)). Equations 1.1 and 1.2 give the life in millions of revolutions after which 10% of bearings are expected to fail. The exponent n is 10/3 for a cylindrical rolling element bearing.

$$L_{10} = \left(\frac{C}{P}\right)^n \tag{1.1}$$

$$L_{10_\theta} = \left(\frac{C}{P_\theta}\right)^n \tag{1.2}$$

The fatigue life therefore becomes a function of the circumferential position of the element. This allows a number of potential of strategies to be investigated. Figure 1.11 shows the predicted L_{10} life for a range of index angles and frequencies. It shows there are a number of possible strategies that would maximise bearing life. The first distinct line from the left, the border between the red and blue zone, represents a single rotation of the inner race and subsequent lines represent additional full rotations. Despite the increase in potential bearing life being the same for each full rotation, there is some benefit in employing a strategy whereby the raceway is indexed multiple times. A single rotation is a risky strategy since there is potential to 'fall' into the red zone which would reduce bearing life significantly.

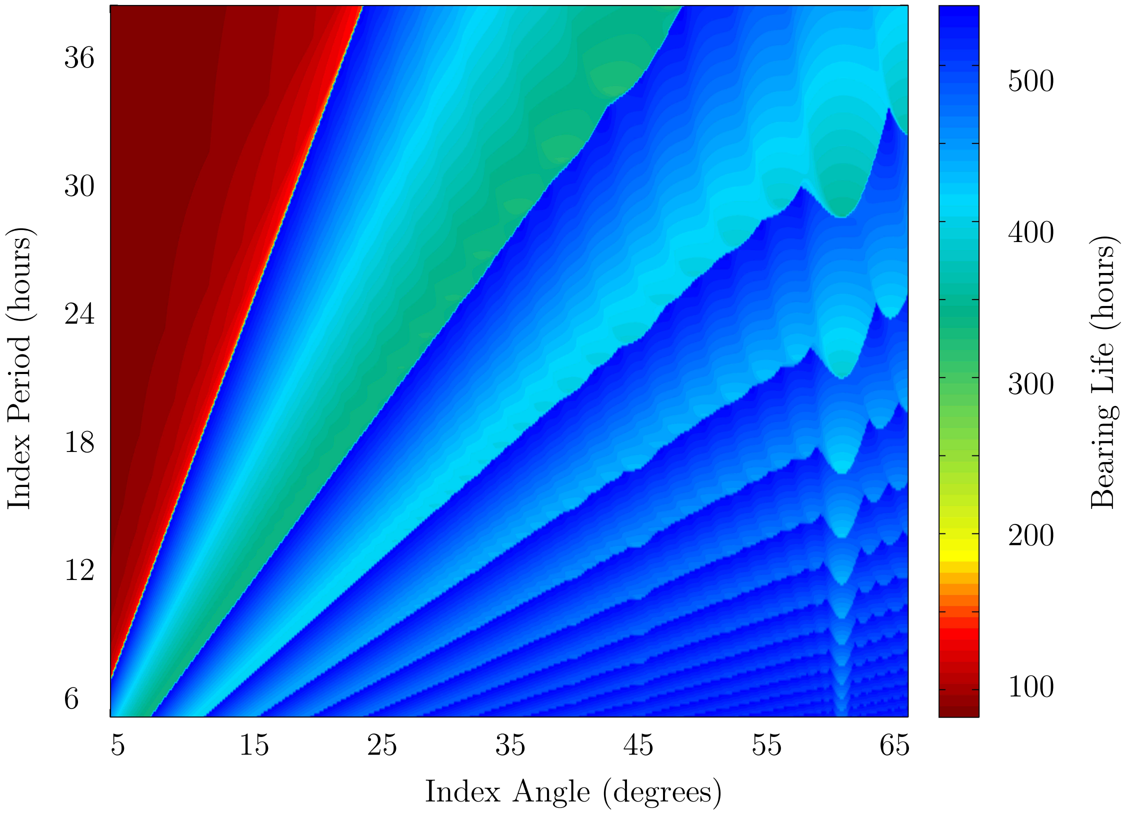

Figure 1.11: Predicted Life for a Range of Indexing Strategies [29]

This model provides an initial estimate that bearing life could be increased by a factor of five however the model has its limitations. Operating conditions are assumed constant whereas in reality speeds and loads may fluctuate dramatically. Furthermore, the model assumes failure by high cycle fatigue whereas field experience suggests the majority of failures occur due to overload events creating stress concentrations. An additional model, taking into account the transient forces experienced by an operating gearbox, is currently being developed while experimental work will validate potential increases in bearing life.

A full-scale bearing test rig, described in greater detail in Section 4.1, has been designed to run accelerated life tests of planet bearings and to validate the use of acoustic emissions (AE) and ultrasound as effective monitoring techniques. The rig will be used to test the concept behind the MultiLife™ device and assess a number of indexing strategies.

This work forms part of a larger project which is looking to prove the Multilife™ concept and potentially adopt an intelligent indexing strategy based on the bearing condition. This thesis will focus on the measuring and the processing of acoustic emission data with the aim of determining bearing condition.

1.5 Aims and Objectives

The overall aim of this work is to develop a means of assessing the condition of bearings typically found in wind turbines using acoustic emission signals. In order to achieve this, the following objectives must be met.

Determine a method for reducing signal noise and dimensionality

As will be demonstrated later, the oil film thickness is expected to be of a similar magnitude to the composite roughness of the contact. This prevents full separation between the raceway and rolling element causing asperity contact and high amplitude, transient features in the acoustic emission signal. The defect signals are therefore unlikely to be distinguishable from the noise. Previous studies regarding the acoustic emission monitoring of bearings have only considered instances of full film separation. The first challenge is to reduce the noise so that smaller defects are detectable. Additionally, since large quantities of data will be produced due to the high sampling frequency, the relevant features of the raw data should be extracted with a much smaller dimension for ease of storage and further processing.

Derive a suitable measure of bearing degradation.

Such a measure should be at least as sensitive to incipient damage as current state of the art methods. The suitability of the methods will be quantified and judged according to the degree of consistency, sensitivity and whether the measure exhibits monotonic behaviour.

Investigate suitable algorithms to estimate the position of a defect.

An accurate method for locating the source of emissions and therefore the damage position is deemed to be vital for the efficient operation of the Multilife™ bearing.

Previous work has focused on using time of flight calculations to estimate the source position. This work will investigate two other data driven methods in an attempt to improve accuracy. A recommendation will be made based upon the accuracy, suitability and computational requirements to carry out such a task.

1.6 Thesis Outline

The remainder of this thesis is organised into seven additional chapters which are summarised as follows.

- Chapter 2 introduces AE technology and provides a brief overview of the history and typical applications of AE. Various mechanisms by which AE sources are generated are described followed by a summary of some parameters which are commonly monitored during a test. The instrumentation and hardware required for AE monitoring are also discussed.
- Chapter 3 summarises the current techniques used for monitoring rotating machinery and in particular bearings, most of which have been applied to vibration data. The chapter concludes with a review of previous studies investigating AE monitoring of rolling element bearings.
- The bearing test rigs and acoustic emission instrumentation used throughout this work is described and shown in Chapter 4.
- Chapter 5 describes a proposed method for the defect detection of a bearing, illustrated using simulated data.
- Chapter 6 introduces various means of locating the source of an emission and proceeds to test the suitability of the most relevant methods on a bearing raceway using artificially generated emissions. The methods are confirmed using a bearing seeded with a real defect.
- Chapter 7 brings together the theory introduced in previous chapters and presents the results and main findings using data measured from two run-to-failure bearing tests. Further data is presented from a instrumented gearbox and a split race bearing test rig.

- Chapter 8 discusses the findings of this thesis and makes recommendations for future work in this area.

Chapter 2

Acoustic Emissions

This chapter provides an overview of acoustic emission technology. A brief history outlining the origins of AE testing is presented. Four categories of AE sources, namely primary, secondary, noise and artificial, are identified alongside the factors affecting their output followed by a description of features and parameters commonly observed during a typical AE test. The chapter concludes with an overview of wave propagation characteristics and the instrumentation required for AE monitoring.

2.1 Definition

Acoustic Emission (AE) technology is a non-destructive testing (NDT) method used in a wide range of industrial applications to detect and locate defects in mechanically loaded specimens. An acoustic emission has been formally defined as "the class of phenomena where transient elastic waves are generated by the rapid release of energy from localized sources within a material, or transient elastic waves so generated" [30]. Such waves usually take the form of high frequency broadband stress waves which are received and converted into electrical signals by piezoelectric transducers. Although the set-up and hardware of AE systems may vary upon application there are four common components required to obtain AE data; a source, a structure through which the source can propagate, the sensor (transducer) and electronics to record and process the data. Later sections will provide more information on each of these stages.

2.2 A Short History of AE

Whilst audible sounds have been used since the beginning of mankind to judge the structural integrity of materials, technical investigations into acoustic emissions from the deformation of materials were first carried out during the early 20th century. In 1917, Czochralski reported the ‘cry’ of tin and zinc crystals as twinning occurs when the materials are subjecting to a bending action [31]. Six years later, Portevin and Le Chatelier noted high frequency audible sounds during the deformation of an alloy of aluminium, manganese and copper. Further progress took place in 1933 when Kishinouye first developed AE instrumentation to monitor deformation in wood in order to understand and develop methods for studying the fracture of the earths crust. The system employed by Kishinouye used a phonograph pick-up with a steel needle to monitor stress waves [32]. In 1936, Forster and Scheil had independently built instrumentation to monitor acoustic emissions generated during martensitic transformations [30]. In 1950 Kaiser, as a part of his doctorate research, undertook the first comprehensive investigation into acoustic emissions. The phenomenon of irreversibility, the Kaiser effect, described in Section 2.6.1 was his most significant discovery [33]. Until the mid-1950s acoustic emission monitoring was limited to frequencies of around 60kHz. This restricted the practical applications of AE since the presence of external noise required experimentation to take place in soundproof facilities in order to achieve meaningful results. In 1964, Dunegan, Tatro and Harris extended the range of frequencies up to 1MHz [34]. This development bought about a rapid increase in applications in areas as broad as materials research, non-destructive testing (NDT), structural evaluation and even cavity detection in fluid systems. In the USA, Greene, Lockman and Steele used AE to assess the structural integrity of rocket motor cases [35] while Dunegan proposed the use of AE to monitor pressure vessels. In 1969, he and Knauss founded the first company producing AE instrumentation: the Dunegan Research Corporation. Independently several research institutions in the former Soviet Union were developing techniques to be used in mostly military applications. Today, applications can be found in almost all industries.

2.3 AE Sources

Acoustic emission sources arising during structural monitoring can be broadly classified into three classes; primary activity, secondary activity, and noise. Artificial sources may also be generated for calibration purposes. The following section describes the most common mechanisms for each of these sources. It is not expected that all of these sources will be detected in this work.

In general, AE sources emit a broadband signal, which can range from sub-audible frequencies during seismic events to frequencies of the order of MHz under certain conditions. However, most commercial AE sensors operate with a frequency response within the range 20kHz-1MHz. At lower frequencies, mechanical and external noise will often dominate the signal whereas at higher frequencies attenuation reduces the applicability and effectiveness of AE as an NDT method with applications usually restricted to the laboratory.

2.3.1 Primary Sources

Primary activity occurs when the original material undergoes new, permanent changes relating to deformation and fracture development arising due to local stresses being higher than the original stress levels. Stress concentrations occur at changes in section, weld details, cracks and structural discontinuities in general. There are a number of types of permanent deformation. The amplitude of source events can vary enormously depending upon both the deformation mechanism taking place and the properties of the material. Factors affecting the amplitude of an AE event are listed in Table 2.1 [31].

Dislocations

All AE processes begin with a material undergoing stress. The structure responds to the stress by changing its shape, or in other words, it experiences strain. If the stress is high enough to exceed the yield stress of the material, plastic deformation will occur whereby the relative position of atoms in the material are changed permanently. Irregularities in the crystalline structure of a material known as dislocations enable

Table 2.1: Factors Affecting AE Response [31]

Factors increasing AE amplitude	Factors decreasing AE amplitude
High strength	Low strength
High strain rate	Low strain rate
Low temperature	High temperature
Anisotropy	Isotropy
Non-homogeneity	Homogeneity
Thick sections	Thin sections
Brittle failure (cleavage)	Ductile fracture (shear)
Materials containing discontinuities	Materials without discontinuities
Martensitic phase transformations	Diffusion controlled phase transformation
Crack propagation	Plastic deformation
Cast materials	Wrought materials
Large grain size	Small grain size
Mechanically induced twinning	Thermally induced twinning

the mechanism of atomic planes to slide over one another resulting in acoustic emissions. This phenomenon was first investigated by Mason et al [36] who suggested that dislocations in fine tin structures produced high frequency vibrations. Single isolated dislocations tend not to produce detectable emissions. However, as a material yields many dislocations occur simultaneously and the stress waves interact and superimpose to form a detectable signal. In fact, during tensile tests of low carbon steel, Akbari and Ahmadi [37] claim that the main source of AE at all stages of deformation is due to massive dislocation motion.

Precipitates and Inclusions

On a microscopic scale, specks of carbide, sulphide and oxide, amongst other non-metallic materials, are scattered through steel. Specks of carbides are usually very small, to the order of a few hundred atoms, and are more often referred to as precipitates. The breakage of precipitates and larger inclusions, such as manganese sulphide particles which are formed during the manufacturing process, is the main source of acoustic emission when crack free metals are deformed. All these non-metallic components are more brittle than the surrounding metallic matrix so break more easily when the material is strained. Again, the amplitude of such emissions are generally quite small and do not represent a significant threat to the structural integrity.

Crack Propagation

Dislocations within a specimen subjected to a large number of loading cycles will tend to pile up resulting in the movement of material along slip planes. Consequently material can rise above or fall below the surface creating what is called a persistent slip band (PSB). This discontinuity at the edge of a PSB acts as a stress concentrator and an initiation point for cracks. Crack nucleation and propagation represents a significant threat to the structural integrity of a material and the detection of such processes is the most common goal of not only acoustic emission testing but indeed the majority of non-destructive evaluation techniques. Emissions from crack propagation or initiation tend to produce signals of higher amplitude than the breakage of inclusions.

2.3.2 Secondary Sources from Repetitive Mechanisms

Acoustic emissions can arise due to repetitive processes such as friction, impact or high pressure leaks. These processes and therefore the emissions they produce do not, in contrast to primary sources, cause permanent changes to the original material. Instead, they may be used to provide an indication of the structure's condition. It is important to differentiate this class of activity from noise which is always seen as a nuisance during acoustic emission inspection. In addition to the emissions produced by the release of energy during the development of cracks, emissions of lower amplitude can occur from the rubbing or fretting of the crack surfaces as they open and close under an oscillating load.

2.3.3 Non-Material Secondary Sources

Secondary activity can arise from material that is not a part of the original fabrication such as the cracking of corrosive products which form on the surface. These tend to cause small amplitude AE waves compared to those arising from primary sources. The emissions can be enhanced through the formation of corrosive products on the surface as discussed in Section 2.3.3 which make the crack more emissive.

2.3.4 Noise

In terms of AE testing, noise refers to any unwanted components of the signal and is a major drawback of the technique. The two main sources of noise arising from environmental conditions are friction and impact. Frictional sources can, for example, be a result of connections or bolts working themselves loose under a changing structural load. Impact sources include rain, wind driven dust or other flying objects. However, as will be shown later, emissions arising from these sources are not always unwanted and can be used to determine the condition of a component. Electromagnetic interference is another significant source of noise which can be minimised using an integrated pre-amplifier sensor combined with careful grounding and cabling. Understanding and eliminating, or at least reducing, noise is a crucial part of any AE testing system enabling the user to focus on the relevant information within the signal.

2.3.5 Artificial Sources

Artificial sources are used in AE testing for a number of reasons. By assessing the attenuation properties of a structure prior to loading the user is provided with valuable information when deciding upon the placement of the sensors. Often, it is desirable to use the minimum number of sensors without losing relevant information due to attenuation. Secondly, a strong response indicates the sensors are in good acoustic contact with the surface of the structure being monitored and can ensure a similar response between experiments. Finally, artificial sources can be used to verify the accuracy of the source location methods, a technique employed in Chapter 6. This can indirectly lead to the calculation of the wave propagation speed. The Hsu-Nielsen source has been used in this work for speed and simplicity however there are other more advanced methods of generating an AE source such as pulsing a transducer or by thermoelastic expansion by laser.

Hsu-Nielsen Source

The Hsu-Nielsen (H-N) source [38, 39] generates an AE event by breaking a brittle 0.5mm 2H graphite pencil lead against a specimen. A Teflon seat fitted to a retractable pencil, as shown in Figure 2.1 ensures the pencil is in the same position every time a source is generated. The fracture produces a monopole, point, step source and is widely used by AE researchers being both cheap and portable [40]. Care must be taken since reproducibility of this method is poor due to the variance in human use.

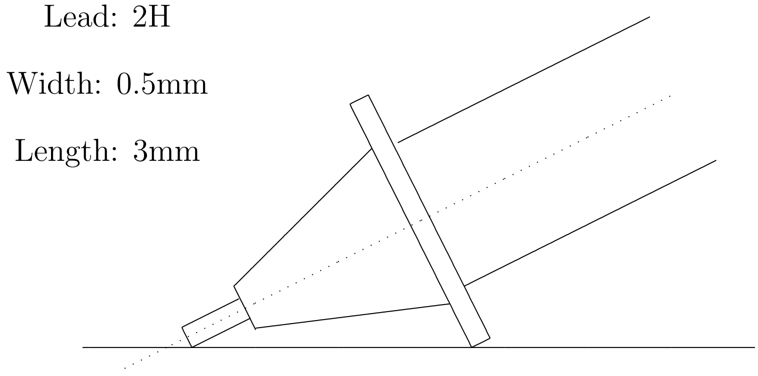

Figure 2.1: Hsu-Nielsen Source

2.4 AE Modes

Acoustic emission signals can be characterised as continuous, burst or a combination of both sometimes referred to as mixed emission; a distinction first made by Kaiser in his seminal work on acoustic emissions [33].

2.4.1 Continuous Emission

Continuous emissions are typically low energy emissions which occur when friction mechanisms or dislocation movements occur on or within a structure and do not display any visually obvious features in the time domain. The amplitude of the emission increases with applied load [41]. Noise can also be a component attributing to a signal exhibiting continuous emission. The monitoring of continuous emission is suited to the detection and analysis of signals produced by plastic deformation in ductile materials, stress corrosion cracking and creep. It has been suggested that this type of emission is best measured with RMS or energy rate measuring circuitry; parameters which are calculated over the entire signal for a predefined period [31].

2.4.2 Burst Emission

Burst emission signals have definite start and end points with distinct peaks of significantly larger amplitude than the background noise. Such bursts tend to be no longer than a few hundredths of a second long. Deformation mechanisms which tend to exhibit burst type emissions include the fracture of non-metallic inclusions, breakage of corrosive products or crack propagation in brittle materials. Impact, which is particularly relevant when monitoring bearing damage, can also be a source of burst emissions. Situations may occur where burst and continuous emissions are present creating a mixed emission. The following section describes the parameters most commonly used to measure burst type emissions.

2.5 AE Signal Features

A transient AE burst occurs when the received signal exceeds a given threshold. The threshold, set by the user, is chosen to be high enough to eliminate any events which may arise due to the presence of low amplitude noise whilst simultaneously capturing genuine AE events. Mba [42] suggests that the relationship between the mechanical integrity of the bearing and AE counts is independent of the chosen threshold level. Some AE systems use a floating threshold which varies with time as a function of background noise. Such thresholds are used to distinguish between noise and genuine events in variable operating conditions or noise. Other methods of determining the presence of an event have also been investigated. For example, Hensman et al [43] noted that the use of a standard threshold crossing would overestimate the start of an event, an error which would increase with distance the wave has travelled due to dispersion. To remedy this error, the use of the Aikaike Information Criterion (AIC) is proposed to detect a more accurate onset of the waveform. This method divides the signal into two and finds the point where the difference in entropy between the first and second section is at a maximum.

Bursts may be characterised by a number of parameters, the most common of which are described in the following sections. These parameters are shown in Figure 2.2. This list is far from comprehensive and further parameters can be derived from those mentioned. The average frequency, for example, is defined as the ring down counts divided by burst duration.

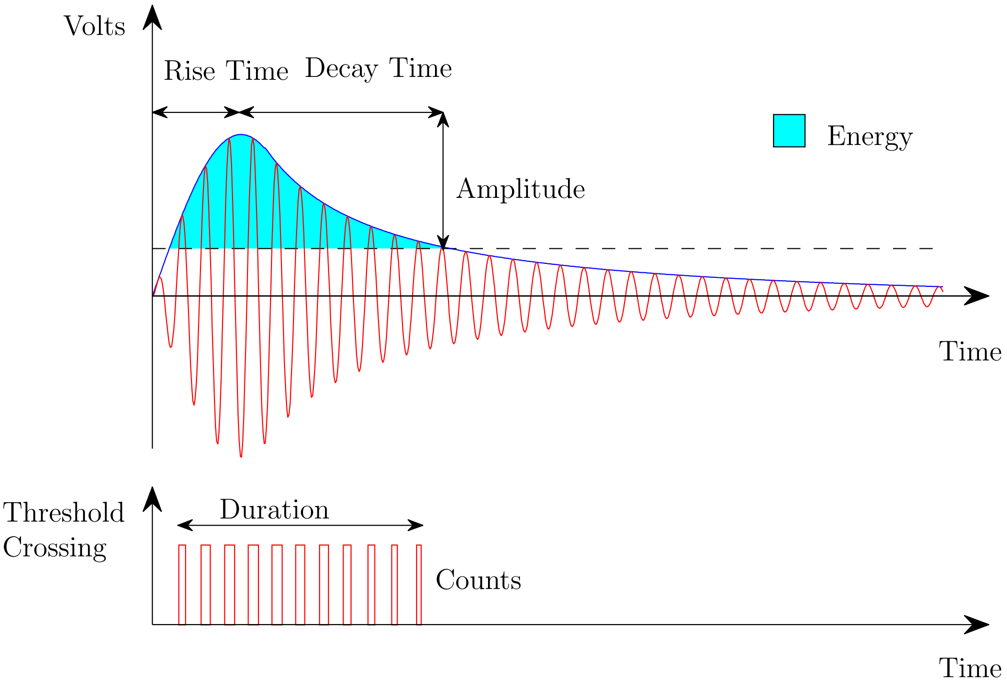

Figure 2.2: Features of an AE Burst

2.5.1 Amplitude

The amplitude is the greatest measured voltage of the signal, corresponding to the greatest surface displacement, and is closely related to the magnitude of the source event. Evidently, only events with an amplitude greater than the threshold will be recorded and subsequently analysed.

2.5.2 Ring Down Counts

Ring down counts (sometimes simply referred to simply as counts) are a count of the number of the times during a single event the signal exceeds the threshold. Counts are often combined with another parameter such as duration or amplitude to provide information about the shape of the signal. The number of counts depends upon the transducer frequency, the damping characteristics of the transducer, the damping characteristics of the structure and the pre-defined threshold [31]. Counts can be used in conjunction with the amplitude to act as a filter against noise, the idea being that amplitude is likely to correlate with counts and high amplitude events with a low number of counts is more likely to be a result of electromagnetic noise.

2.5.3 MARSE

The Measured Area of the Rectified Signal Envelope (MARSE) is a regularly used parameter in acoustic emission testing. The MARSE represents the energy content of a signal although different system suppliers use variations in the method of calculation. MARSE is often preferred over parameters such as counts since it is less dependent on operating frequency and the threshold but is sensitive to amplitude and duration.

2.5.4 Duration

Duration is measured as the period between the first and last threshold crossings. In a similar manner to counts, the duration can be used to differentiate between sources of acoustic emissions and therefore act as a filter. Short duration events, for example, are more likely to be a result of electromagnetic noise rather than genuine event whereas long duration waveforms may, for example, be a result of delamination in composite materials. Duration is highly dependent upon the acoustic properties of the material and, like counts, the threshold. The duration is usually expressed in microseconds.

2.5.5 Rise Time

The rise time is a measure of the time difference between the first threshold crossing and the peak amplitude. It can be used to classify the type of fracture or, as with counts and the duration, eliminate noise signals.

2.5.6 Decay Times

The decay time is a measure of the time difference between the peak amplitude and the final threshold crossing.

2.6 Structural Loading

2.6.1 Kaiser Effect

This effect was first investigated by Joseph Kaiser in the 1950s. When permanent deformation such as that caused by the mechanisms described in Section 2.3.1 occurs, energy is released through stress waves (acoustic emissions) and heat in order to relieve elevated local stresses. This has a stabilising effect since the load is transferred to another part of the structure. If the structure was to be unloaded and reloaded to the same level, the regions which deformed on the first loading will tend to be stable the second time around. No further deformation would take place and no stress waves would be emitted. This principle is called the Kaiser effect and is shown graphically in Figure 2.3. Load is plotted against cumulative AE counts. As the structure is initially loaded (AB) the number of AE events generated increases with load. At point B, the load is partially removed and under reloading (CB) no further activity takes place until the previous maximum load has been exceeded at point C. Structures, and therefore the flaws, which exhibit this behaviour tend to be insignificant and are unlikely to be a threat to structural integrity. The Kaiser effect is most likely to be exhibited in homogeneous materials such as metals whereas it does not occur when deformation is controlled by time dependent mechanisms such as rheological flow or relaxation of the matrix in highly stressed composites, corrosion and hydrogen embrittlement. Other source mechanisms, such as friction in fatigue cracks or between surfaces in composites, can also contravene the Kaiser effect but are still useful for damage detection [31].

The principle of the Kaiser effect naturally leads to the Dunegan corollary; a diagnostic technique developed by Dunegan and Harris used in the field testing of pressurised systems . This approach is based upon the idea that if an operating structure is subjected to periodic proof loads (which are higher than the normal operational loads) emissions will only occur if damage has taken place in the intervening period due to the introduction of stress concentrations [44]. In other words, diagnoses can be made by repeated loads rather than by ever increasing loads.

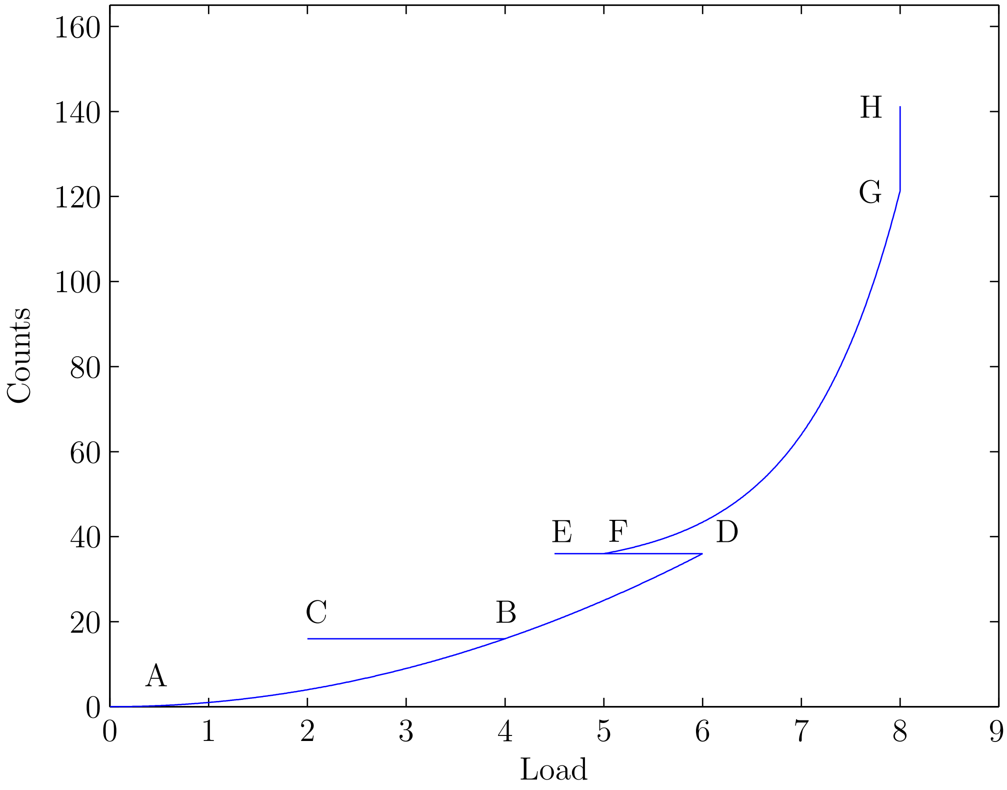

Figure 2.3: AE Activity due to Structural Loading

2.6.2 Felicity Effect

In contrast to the Kaiser effect, whereby a structure will only emit stress waves once a previous maximum load has been exceeded, a structure exhibiting the Felicity effect will become active before a previous maximum load has been reached. With reference to Figure 2.3, the structure will continue to be a source of AE activity as it is loaded to point D. The applied load is reduced (DE) and will begin to re-emit at point F before the previous maximum is reached. The effect can be quantified by dividing the load at which significant AE activity resumes by the maximum applied load (F/D). This quantity is the Felicity Ratio. Structures exhibiting this behaviour infer structurally significant flaws. Likewise, structures which emit during a load hold (GH) also indicate instability. The Felicity effect is more likely to occur in composite materials with inhomogeneous structures such as fibre reinforced plastics (FRP) and concrete. According to American Society of Mechanical Engineers (ASME) Article II, FRP tanks or pressure vessels are rejected when the Felicity Ratio is less than 0.95 [30].

2.6.3 Structural Fatigue

Extensive investigations have been carried out on the AE monitoring of structural fatigue. It has been found under laboratory conditions there are three distinct stages to a fatigue test. These are shown in Figure 2.4. The stress ratio R is defined in Equation 2.1 as the ratio of minimum and maximum applied stresses.

$$R = \frac{\sigma_{min}}{\sigma_{max}} \tag{2.1}$$

Stage I relates to the first few cycles where dislocation movement and cyclic hardening or softening are the primary sources of the AE activity. Very few emissions take place in stage II. During these cycles steady state dislocation motion occurs which will eventually initiate microcracks as described in Section 2.3.1. During stage III, AE activity increases significantly as the crack propagates, the principle sources during this stage are crack tip movement, fracture of inclusions, microcrack coalescence, and fracture along grain boundaries [45]. The number of cycles to failure and therefore AE activity depend greatly upon the stress range applied to the structure as indicated by the different curves in Figure 2.4.

2.7 Wave Propagation

The signal received at an AE sensor is significantly different to that which occurs at the source. This is due to the propagation of the wave through the structure. This propagation largely determines the signal shape and size. A typical AE event generated by a Hsu-Nielsen source is shown in Figure 2.5. It is characterised by a fast rise time and a relatively slow decay. The first part of the waveform is from waves propagating directly from the source whereas the later part is formed from waves reflected before reaching the sensor. The rising part comprises of strong reflections from surfaces close to the sensor whilst the decaying part is shaped by reflections from distant surfaces and acoustic damping.

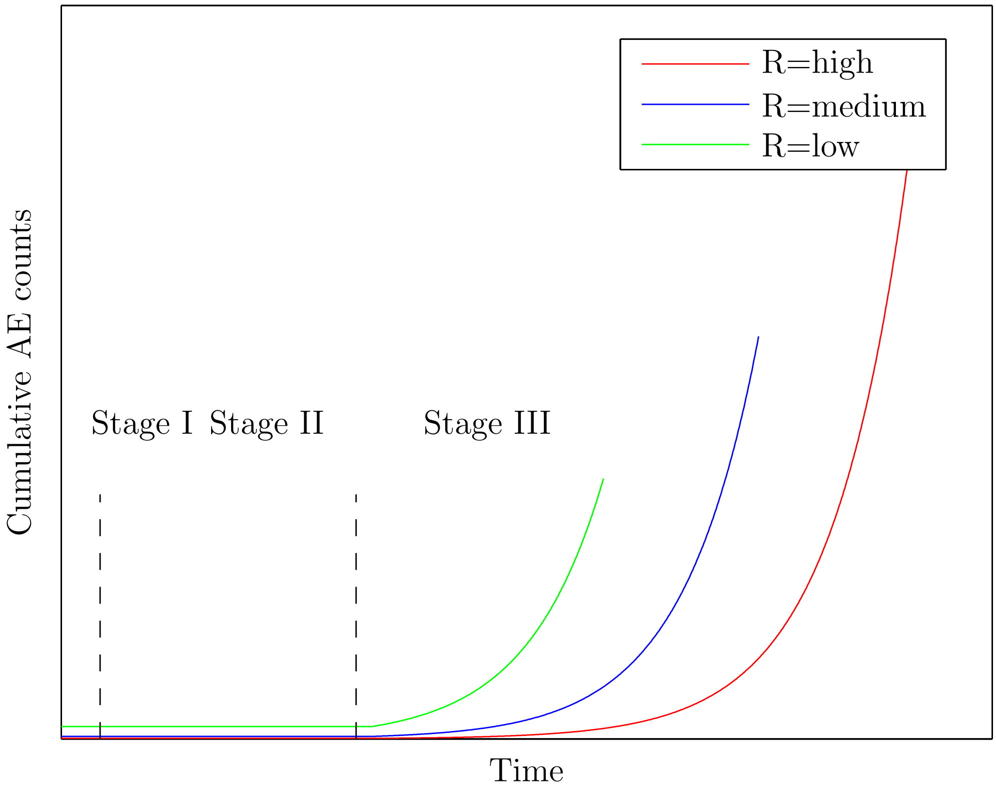

Figure 2.4: Cumulative AE counts for specimen subjected to fatigue loading

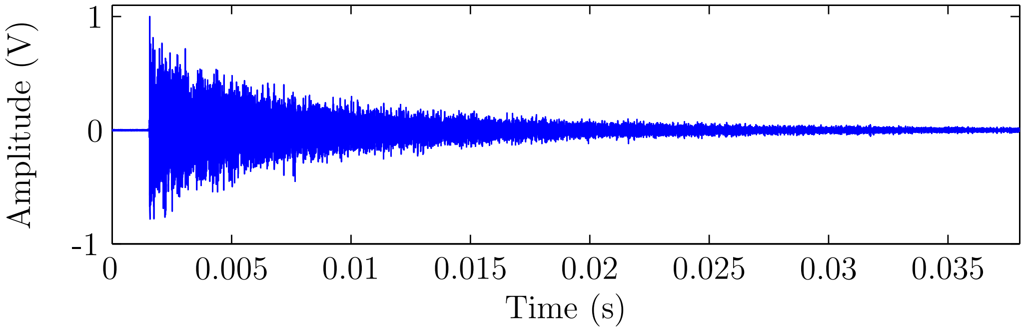

Figure 2.5: Typical AE Event

Waves of different types propagate in different directions with different velocities. According to acoustic theory, wave modes propagate independently and do not interfere with one another. The motion at any point in a structure is the sum of all modes which are present. Source events, such as crack jumps, can last less than a microsecond whereas the time it takes for the wave to reach the sensor and for all reflections to die away can range from 100 microseconds for highly damped non-metallic structures to tens of milliseconds in a lightly damped metallic material [46]. The following factors affect how emissions propagate.

2.7.1 Attenuation

Attenuation of AE waves occur through three mechanisms; geometric spreading, scatter and absorption. Geometric spreading relates to a decrease in amplitude as the wave travels away from the source as the wave energy must be spread over an ever increasing larger wave front. The rules of conservation of energy show that the degree to which attenuation through geometric spreading occurs is dependent upon the structure through which the wave is passing. For example, when propagating through a 2D surface, the amplitude will decrease by 30% every time the distance from the source is doubled whereas this figure increases to 50% in 3D structures [46]. Rods, on the other hand, which do not allow the wave to spread, have much lower rates of attenuation. Geometric spreading is most influential when the sensors are position close to the source.

Scatter is the result of reflections which occur at discontinuities, such as non-metallic inclusions and grain boundaries, and structural boundaries [31]. Scatter becomes increasingly important in complex geometries where there are a number of discontinuities between the source and sensor.

Absorption is the process of kinetic and acoustic energy converting into heat due to internal friction. The rate of absorption is greater at higher frequencies owing to the shorter wavelengths. In contrast to geometric spreading where the rate of attenuation is a function of the distance from source, amplitude reduction due to absorption occurs at a constant rate. Absorption, therefore, becomes relatively more important further from the source. At typical AE frequencies steel absorbs very little energy as heat however non-metallic materials such as paint tend to absorb significantly more.

Attenuation can be measured easily by simulating an AE source. This can be done using a Hsu-Nielson source; a technique which involves breaking a pencil lead of a given size against the material at a set angle (Section 2.3.5). The stress waves produced from the deformation of the lead are fairly reproducible and are of similar amplitude and frequency content to a typical crack source. An attenuation graph can be created by carrying out this technique at a number of distances from a sensor and plotting the peak amplitude against the distance. Such an exercise can aide a user when deciding where to place AE sensors to ensure all relevant events will be captured.

2.7.2 Wave Velocity

Wave velocity is an important characteristic of acoustic emission testing, especially for the purpose of locating source events. A typical event may excite a number of types of waves. The source mechanism, material properties, and geometry influence which type and the degree to which different waves are excited. Sound waves can propagate in four principle modes; as longitudinal, shear, surface or plate waves. These are discussed in the following sections. Anelastic waves can also be generated however these tend to attenuate at short distances from the source so are rarely detected by AE transducers and will not be discussed here.

2.7.3 Longitudinal Waves

Longitudinal waves (otherwise known as dilatational, compressional or P-waves) are waves in which the direction of oscillation occurs in the same direction as the wave propagation. Due to their compressional nature, longitudinal waves can be transmitted through fluids as well as solids.

2.7.4 Shear Waves

Shear waves (otherwise known as transverse, distortional or S-waves) are waves in which the particle oscillations occur perpendicular to the direction of wave propagation. Unlike longitudinal waves, shear waves cannot propagate through liquids since there are no strong bonds to support shear forces. Shear waves are usually weaker compared to longitudinal waves.

2.7.5 Surface Waves

Surface waves (or Rayleigh waves) propagate along the surface of thick materials. Particle motion occurs to about the depth of one wavelength. Surface waves are essentially a combination of shear and longitudinal waves which results in the particles experiencing elliptical motion. Surface waves are useful to many NDT techniques since they can follow the surface of curves and are very sensitive to surface defects.

2.7.6 Lamb Waves

Lamb waves only propagate through thin sections; the maximum thickness of which is to the order of a few wavelengths. There are two families of modes of Lamb waves, symmetrical and asymmetrical. The motions of which are shown in Figure 2.6. The parent modes of these families are a_0 and s_0 respectively. The symmetric mode s_0 is always symmetrical about the median of the plate. The rippling effect is a result of the plate alternately stretching and compressing. This mode requires an exciting force applied parallel to the plate. This can be achieved by a sudden release of in-plane tension.

The motion which occurs in the asymmetric mode a_0 is perpendicular to the plate. Both surfaces move in the same direction. In contrast to the s_0 mode, the exciting force must be perpendicular to the plate. The s_0 mode is important in the acoustic emission testing of thin plates. They usually propagate faster than the a_0 mode so, in most structures, reach the sensor first. The difference in arrival times between the s_0 and a_0 mode has previously been used as a means of source location [47]. The amplitude produced by the a_0 mode tends to be greater than the s_0 mode.

Both families include an infinite number of modes (a_1, s_1, etc) which can travel faster than both the a_0 and s_0 mode. These are of little importance in AE testing because the amplitudes of which are small in comparison to the zero modes. Lamb waves are dispersive meaning their velocity is dependent on frequency. The speeds of each mode is a function of the frequency thickness product (frequency of wave multiplied by plate thickness).

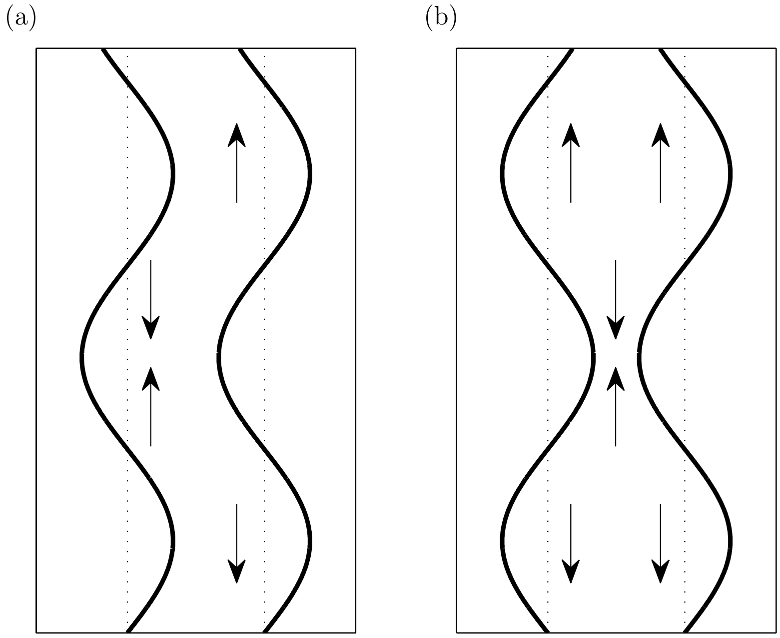

Figure 2.6: Lamb wave modes: (a) Asymmetric and (b) Symmetric

2.8 AE Instrumentation

A schematic of a typical AE system is shown later in Chapter 4, Figure 4.9. The sensor (also discussed later in Section 2.8.1) produces an electrical signal proportional to the displacement at the sensors surface. Preamplifiers amplify the initial signal to reduce noise. Typical values of gain are 20, 40 or 60dB. The signal then undergoes conditioning; this usually includes at least a band pass filter to remove low frequency mechanical noise.

2.8.1 Sensors

The vast majority of AE transducers are made from piezoelectric elements. Capacitive transducers and laser interferometers have been used but have found limited application outside the laboratory. A schematic of the typical construction of an AE sensor in contact with a test object is shown in Figure 2.7.

Piezoelectric elements produce an electrical voltage proportional to the deformation it experiences. The deformation is produced by the motion of the stress waves propagating in the material. The ratio of output voltage amplitude to input motion

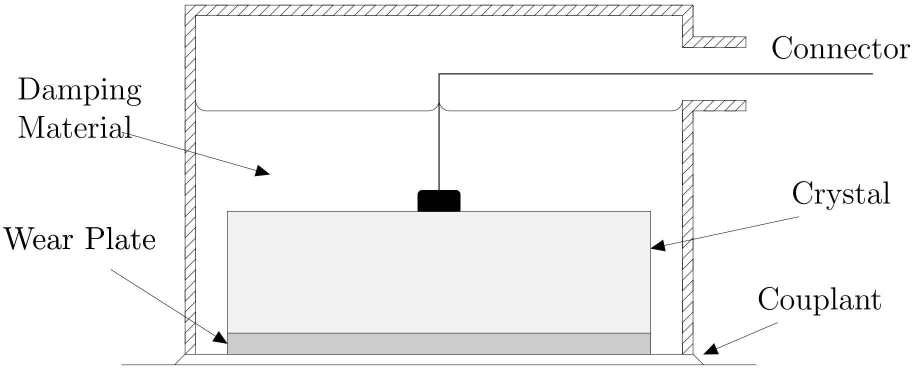

Figure 2.7: Schematic of Typical AE Sensor

amplitude is a measure of the sensitivity of the sensor. Sensitivity is highly dependent upon frequency and the direction of motion. The sensitivity of a sensor is greatest at its resonant frequency. The smallest electrical signal that can be detected is about 10mV, this corresponds to a surface displacement of about 25pm [48]. The thickness of the piezoelectric element defines the sensors resonant frequency and the bandwidth is determined by the properties of the damping material. A calibration certificate with the frequency response is usually supplied with each individual sensor.

There are two principle types of AE sensor; resonant and broadband sensors. Resonant (or narrowband) sensors have a non-linear response; the sensitivity being greatest at the centre or resonant frequency. The resonant frequency of most commonly used commercial sensors is between 100 and 300 kHz. Other operating frequencies can be used but the limitations of which are significant. At higher frequencies attenuation occurs more rapidly therefore a greater number of sensors are required to monitor the same structure. Coupled with the increased sampling rate required, data storage and processing power become the limiting factors. At lower frequencies, mechanical noise becomes prevalent.

Broadband sensors, on the other hand, have a linear response over a range of frequencies but also generally a lower sensitivity. Broadband sensors have been used by researchers to determine the source waveform through detailed (and computationally expensive) analysis of the first part of the received signal characterised by the direct component and early reflections of the wave. In contrast, the majority of practical applications use narrowband sensors to measure broad statistical features of AE activity. Since narrowband sensors measure only a few signal features, hundreds of events can be processed every second, the trend of which can be monitored over time [48].

Desirable characteristics that manufacturers of AE sensors aim for are; high sensitivity, consistent frequency response, robust performance and immunity to noise. In order to achieve a high immunity to noise, some manufacturers have introduced an integrated pre-amplifier to reduce the relative effect of electromagnetic interference in the cabling.

In order to detect the surface motion and produce a strong signal the sensor must be in good acoustic contact with the surface. A couplant must be used to remove any air pockets from the interface. The degree of transmission of a signal across the boundary of two materials is dependent on the similarity of acoustic impedance between the two materials. The acoustic impedance of air is much less than that of the sensor and the surface which would cause large reflections at the interface and therefore a significant loss in transmission. Suitable couplants usually take the form of adhesives, viscous liquids or greases. The surface may require preparation since the surface to which the sensor is to be mounted must be smooth and clean. The couplant and mounting method is dependent upon application and environmental conditions.

The temperature of the structure to which the sensors is mounted must be taken into account when selecting a suitable sensor. A property of piezoelectric elements called the Curie Point is defined as the temperature at which piezoelectric elements permanently lose their piezoelectric properties. This is usually between 125-400℃.

2.9 Other NDT techniques

There are many different applications and components that require condition monitoring. The following section gives a brief overview of the techniques that are most applicable to monitoring rolling element bearings.

2.9.1 Vibration Analysis

Vibration analysis is usually performed by attaching a seismic or piezo-electric accelerometer to the bearing casing. The condition of the bearing can be estimated by comparing vibration levels to historic baseline data. In most cases, the data is analysed in the frequency domain courtesy of a fast Fourier transform (FFT) whereby the examination of prominent frequencies and harmonics can provide an indication of the location, type and possibly the root cause of any defects. Vibration is widely used in industry however the defect signal can often be obscured in the spectrum since the relevant frequencies are often of a similar order to mechanical noise which is present.

2.9.2 Lubricant Analysis

Lubricant analysis is the process of examining a lubricant's properties including the base oil and its additives, suspended contaminants and wear debris from the machinery. Lubricant samples can be taken at regular intervals to provide information about the lubricant and machine condition and can inform maintenance decisions. The analysis is usually performed by off-site laboratories although on-site laboratories and portable analysis tools are becoming more common with the benefit of faster results and earlier maintenance. There are many analytical techniques that can be employed when performing lubricant analysis including spectroscopy, particle counts, infrared, gas chromatography amongst many others. The decision to use a technique or not is application dependent, rotating machinery, for example, tends to generate large wear particles for which rotrode filtration spectroscopy is suitable whereas reciprocating engines tend to generate fine metal elements for which rotating disc electrode spectroscopy or inductive coupling plasma spectroscopy is more suitable [49].

2.9.3 Infrared Thermography

Infrared thermography is method whereby the infrared energy emitted by a surface is detected, converted into a temperature and displayed as a temperature distribution. The use of infrared thermography as an NDT technique is based upon the principle that most components will get hotter or cooler prior to failure. Most commercial cameras are able to capture temperatures ranging from -20 to 500°C. Since contact with the component is not required, it allows moving or dangerous objects to be easily monitored however since direct line of sight to the surface of the object is required it becomes unsuitable for many bearing application which are often enclosed by a housing unit. Once the temperature of the housing increases, it is likely that significant damage will have accrued on the bearing.

2.9.4 Motor Current Signature Analysis

Motor current signature analysis employs the spectral analysis of the stator current of an induction motor to detect both electrical and mechanical defects. Defected rolling element bearings which support the rotor will cause a radial motion between the rotor and the stator. This displacement causes the air gap to vary, resulting in anomalies in the air-gap flux density. The current therefore generates frequencies related to the characteristic defect frequency of the bearing (see Section 3.2.2) which can be monitored over time [50].

2.9.5 Eddy Current Transducers

Eddy current transducers use electromagnetic induction to detect defects in conductive components. The most common method involves placing a circular coil carrying current placed near to the specimen. An alternating current in the coil generates a changing magnetic field which in turn produces eddy current. variations of the phase and magnitude of these currents can be monitored by either a secondary receiver coil or by the primary excitation coil. A change in the phase and magnitude of the eddy current can be a results of changes in conductivity, permeability or due to the presence of flaws. There are a number of limitations of eddy-current monitoring including: only conductive materials may be tested, the surface of the component must be accessible, flaws which lay parallel to the probe may go undetected and the depth to which a defect may be detected is limited by the conductivity of the material.

2.10 Advantages & Disadvantages

The benefits and shortcomings of acoustic emission monitoring compared to other common NDT techniques are summarised as follows.

Advantages

- Ability to determine location of defects since emissions from source emanate in all directions.

- Able to monitor entire structures with a sparse array of transducers but care must be taken that sensors are not so far apart that attenuation becomes significant.
- High signal to noise ratio since mechanical noise tends to occur at lower frequencies.

Disadvantages

- Given the passive nature of acoustic emission technology, it is only possible to detect events such as crack growth or propagation when the system is recording and the monitored component is under load. However, there are instances such as during bearing monitoring where repeated impacts or the testing of pressure vessels for leaks where secondary sources can provide an indication of the component's condition.
- A large amount of data can be acquired which requires storage and computational power to process. This becomes more significant when monitoring low speed machinery since longer samples need to be taken in order to record at least a full rotation of the bearing.
- Due to high frequency, attenuation is significant. Care must taken when placing sensors that the whole structure is monitored.
- The voltages induced in the transducer are very small. Care must be taken to ensure noise does not corrupt the signal before it is amplified.

Chapter 3

Rolling Element Bearing Monitoring

Given the importance of rolling element bearings in virtually all rotating machinery, condition monitoring of bearings has become a well developed technology. Although vibration analysis is the most common method of determining the condition of a bearing, a number of other techniques may also be employed. These include oil debris analysis, temperature monitoring and, in recent years, AE analysis. This chapter begins by examining the common failure modes experienced by rolling element bearings followed by a description of typical parameters and methods used in vibration analysis to diagnose a bearing, all of which are equally applicable to AE analysis. The chapter concludes with a review of the previous studies and diagnostic capabilities of AE technology in relation to rolling element bearing analysis.

3.1 Bearing Failures

3.1.1 Fatigue

A bearing operating under optimum conditions will tend to fail by rolling element fatigue after a given period. The distribution of such failures will follow a Weibull distribution given a sufficient number of repeats. The L_{10} life, introduced in Section 1.4, gives the life after which 10% of a significantly large sample of bearings will have expected to fail. Fatigue is attributed to the Hertzian contact stresses, which are at a maximum below the surface, developed under cyclic loading. This initiates the formation of subsurface cracks which propagate and interact as loading continues. As the cracks reach the surface, a segment of material can break loose creating a spall as shown in Figure 3.1. Damage will spread due to the increase in localised stress around the defect and will severely impact on bearing performance before eventually leading to catastrophic failure unless remedial action is taken. Having said this, often bearings operate under sub-optimal conditions and other failure modes become more prominent.

Figure 3.1: Example of advanced fatigue damage. Copyright Schaeffler Group.

3.1.2 False Brinelling

False brinelling is caused by excessive external vibrations while the bearing is stationary. Since the rolling elements are not rotating no lubricant is entrained into the contact, preventing the formation of an oil film. The small relative motion of the elements and the raceway creates wear particles which oxidise and accelerate the wear process. The presence of bright wear marks surrounded by brown rings at the spacing of the rolling elements indicate that false brinelling has taken place.

3.1.3 True Brinelling

True brinelling occurs when a static load or impact exceeds the elastic limit of the raceway material. This causes a series of indentations on the raceway which increases the vibration within the bearing leading to the rapid onset of fatigue damage.

3.1.4 Overheating

Temperatures in excess of 200°C can anneal the raceway and rolling element materials. The resulting loss of hardness reduces the load capacity of the bearing causing early failures. At even higher temperatures, the bearing components can deform. Additional damage can occur due to the high temperature degrading or destroying the lubricant. The rolling elements, raceways and cage will discolour, turning gold or blue, in subjected to excessive temperatures.

3.1.5 Corrosion

The presence of water, acids or other contaminants in the bearings for prolonged periods of time causes corrosion damage. Rust on the bearing surfaces produces uneven operation which can also lead to abrasive wear. Rust pits can also act as localised stress concentrators leading to the formation of further spalls. Typical corrosion damage is shown in Figure 3.2.

3.1.6 Lubricant Failure

Rolling element bearings require the presence of a thin lubricant film to prevent the rolling elements coming into contact with the raceways. Insufficient lubrication or excessive temperatures which degrades lubricant properties will result in asperity contacts which will weld together under stress and be torn apart as the rolling element passes. Initially, the bearing will become discoloured before excessive wear will result in overheating and catastrophic failure.

Figure 3.2: Example of corrosion damage. Copyright Schaeffler Group.

3.1.7 Micropitting

Micropitting is a prominent failure mode in wind turbine gearboxes, affecting both bearings and gear teeth. It has been reported that micropitting is a fatigue phenomenon that occurs under mixed film elastohydrodynamic lubrication (EHL) where the lubricant film thickness is of the same order of magnitude as the surface roughness (R_a) [51]. Under such conditions a proportion of the load is supported by the asperities in contact resulting in elastic and plastic deformation and the initiation of small cracks. Micropits are formed when these cracks grow and interact resulting in small bits of materials, micrometers in size, breaking off. This process changes the geometry and internal clearance of the bearing, increasing stress concentrations leading to secondary damage mechanisms such as macropitting, scuffing, bending fatigue and polishing [51].

3.1.8 White Structure Flaking

White structure flaking (WSF) is another prominent failure mode of wind turbine gearboxes and is thought to be the principal cause of wind turbine bearing failures that occur within the first 3 years of operation, at 1-10% of their rated L_{10} life [52]. Conditions which are often found in wind turbines and other large industrial applications can alter the subsurface microstructure of a raceway. These areas, which are more brittle than the surrounding material, appear white when etched with nital solution and viewed under reflective light and are hence called white etching areas (WEA) [53]. Subsurface defects, such as inclusions and voids, in WEAs act as stress concentrators and act as initiation points for cracks to form and grow. Cracks which have formed close to the surface can weaken the region and cause spalling as larger bits of material break away. An example of such a crack, which has propagated towards the rolling surface, in a bearing taken from a 600kW wind turbine gearbox is shown in Figure 3.3. Although there is currently no consensus regarding the mechanisms through which WSF occurs, it has been suggested that a complex interaction of a number a factors could be the root cause [52]. These factors include but are not limited to stress levels, impacts, lubricant decomposition, percentage of retained austenite, hydrogen embrittlement and the presence of inclusions, vacancies, voids or carbides.

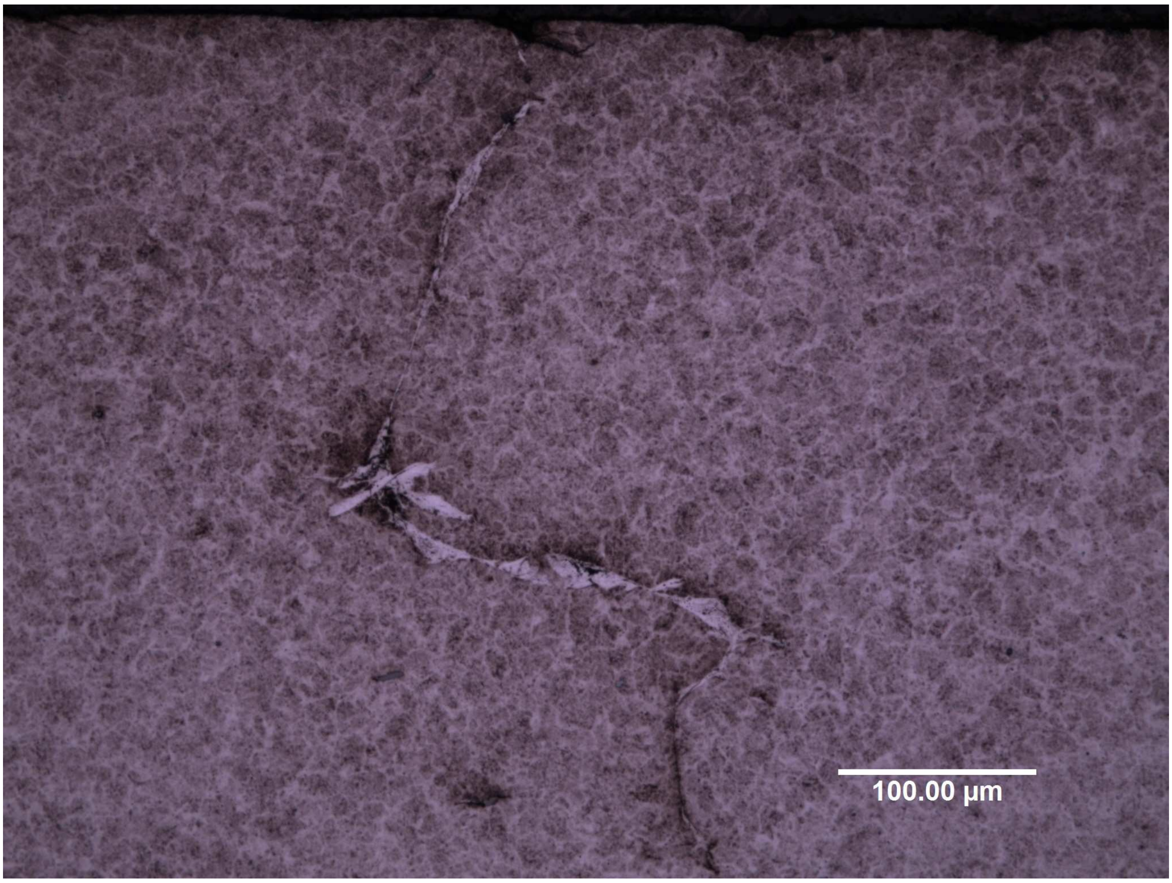

Figure 3.3: Example of White Etching Crack. Copyright Tom Bruce.

3.2 Signal Processing Techniques for Bearing Monitoring

The following section introduces some of the common techniques employed to monitor rolling element bearings. The methods and techniques presented here can be applied to both conventional vibration analysis and AE. Other methods, such as oil debris and temperature monitoring, can also be used to monitor bearing health but have not been included here for the sake of brevity.

3.2.1 Time Domain

The simplest approach when monitoring bearing damage is in the time domain whereby the measured data is displayed or analysed as a function of time [54]. The most common parameters are described in this section. The trend of each parameter can be monitored over time and any significant changes can be either attributed to changes in operational conditions or are indicative of damage. Most of these techniques have been developed for the analysis of traditional vibration signals however they are equally applicable to AE signals.

RMS

The root mean square (RMS) provides a measure of the overall energy of a signal and is commonly used in industry due to its simplicity and ability to be run online. The RMS is given by Equation 3.1

$$RMS = \sqrt{\frac{1}{N}\sum_{t=1}^{N} x(t)^2} \tag{3.1}$$

where $x(t)$ is a discrete time series signal composed of N data points. The RMS is a robust measure of bearing health when operating conditions are constant and can indicate the presence of significant defects. However, since the RMS is calculated over a pre-defined time constant it fails to take into account the transient nature of a defect signal. The interval between impulses arising due to a rolling element impinging on the raceway is usually large in comparison to the duration of the impulses themselves, this is especially true for small defects on slow rotating bearings. Transient bursts with an amplitude significantly smaller than the noise, such as those arising from incipient damage will therefore form a minimal contribution to the overall energy. Damage will have often progressed significantly before it can be detected by the RMS.

Peak Value

The peak of the signal can be defined as the mean of the absolute maximum and minimum of a signal and is therefore given by Equation 3.2. The idea being that larger defects will produce higher amplitude events however since the peak value is not a statistical measure calculated using all data points, it is not seen as a reliable indicator of bearing health as noise and spurious data may skew results.

$$Peak = \frac{1}{2}(max(x(t)) - min(x(t)) \tag{3.2}$$

Kurtosis

Kurtosis k, the fourth standardised moment, measures the degree of 'peakedness' of a signal and is given by Equation 3.3.

$$k = \frac{N \sum_{t=1}^{N}(x(t) - \bar{x})^4}{(\sum_{t=1}^{N}(x(t) - \bar{x})^2)^2} \tag{3.3}$$

The kurtosis of a Gaussian distribution, as is typical of an undamaged bearing, is 3. When damage progresses, the amplitude of the impulses due to the defect will increase along with the kurtosis. Unfortunately, kurtosis is sensitive to spurious vibrations and noise; the effects of which can outweigh its benefit as a degradation indicator [55]. As these defects interact and grow, the impulses merge and the signal reverts back towards a Gaussian distribution resulting in a subsequent decrease in kurtosis. It has been reported that kurtosis is only effective when monitoring for localised defects since distributed damage is similar to asperities but on a larger scale, therefore also producing a Gaussian distribution [56].

Other statistical moments, such as skewness, have been used for condition monitoring of rotating machinery and are based upon the same principle; that a deviation from a Gaussian distribution is indicative of damage. Since these moments provide a measure of the distribution or 'shape' of the data they are not dependent upon operating conditions.

Crest Factor

The crest factor is the ratio of the peak value to the RMS of a signal given by Equation 3.4.

$$Crest\ Factor = \frac{Peak}{RMS} \tag{3.4}$$

The crest factor will increase as discrete defects cause sparse but high amplitude peaks in the signal but as these defects grow and form distributed damage around the raceway the RMS will increase, decreasing the crest factor. Thus, whilst able to indicate the presence of an early defect the measure is not a suitable measure to characterise bearing degradation.

K-S Test

The Kolmogorov-Smirnov (K-S) test provides a measure of similarity between the distribution of a sample data set and that of a reference set. The K-S statistic D quantifies the distance between the cumulative distribution functions (CDF) of two distributions and is given by

$$D = sup_x|F_n(x) - F(x)| \tag{3.5}$$

where F_n and F are the CDFs of test and reference samples respectively and sup_x is the supremum of x. The double ended arrow in Figure 3.4 shows D graphically. The reference distribution may be either empirical (two-sample K-S test), for example, from an undamaged bearing or that of a known distribution, typically Gaussian. A sample taken from the time series data of an undamaged bearing after the running-in period is normally taken to form the reference CDF. The idea being that as a defect grows, a greater proportion of the samples will be of a larger amplitude thus increasing the K-S statistic.

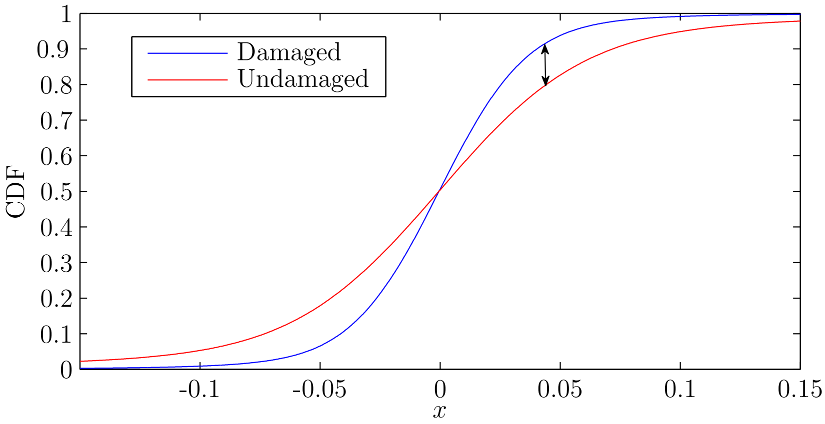

Figure 3.4: K-S Test Example

3.2.2 Frequency Domain

The main purpose of analysing data in the frequency domain is to identify the presence of peaks in the spectrum which relate to specific failure modes or certain damaged components. Degradation can be monitored by tracking the trend of these amplitudes over time. Any stationary signal, that is a signal whose frequency components do not vary with time, can be expressed by a Fourier series, the addition of any number of sine waves of different amplitude and frequency. The Fourier series expansion of a discrete time series $x(t)$ of period T can be given by Equation 3.6

$$X(f) = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-j2\pi ft}dt \tag{3.6}$$

where f represents equally spaced frequencies at multiples of period T. The power spectrum density estimates the portion power in each frequency bin obtained by the Fourier series and is given by Equation 3.7

$$P(f) = E[X(f)X^*(f)] \tag{3.7}$$

where * is the complex conjugate and $E[]$ the expected value.

As a rolling element passes over a small defect, a pulse of vibration is generated. However, due to the short duration of the pulse and the presence of vibrations from other machine elements, detecting the additional energy via traditional spectral analysis is difficult and almost impossible when analysing AE signals. Extraneous noise can sometimes be reduced through spectral averaging but additional techniques are usually required to extract the relevant frequencies. Furthermore, the defect signal is a high frequency wave modulated at a lower frequency dependent on the damaged component.

Characteristic Defect Frequencies

When a rolling element passes over discrete, rather than distributed, defects, such as those arising due to fatigue spalls or plastic deformation, the raceway is subjected to an impulse load which generates a broadband stress wave. The frequency at which these impulses occur is a function of operating conditions and bearing geometry. Each bearing component; the inner race, outer race and rolling elements, has a different defect frequency and the inner race defect frequency f_{ir}, the outer race defect frequency f_{or} and the rolling element defect frequency f_{re} are given by Equations 3.8-3.10 respectively [57].

$$f_{ir} = \frac{n}{2} f_s \left(1 + \frac{BD}{PD} cos\beta\right) \tag{3.8}$$

$$f_{or} = \frac{n}{2} f_s \left(1 - \frac{BD}{PD} cos\beta\right) \tag{3.9}$$

$$f_{re} = \frac{BD}{PD} f_s \left[1 + \left(\frac{PD}{BD}\right)^2 cos\beta\right] \tag{3.10}$$

where n is the number of rolling elements, f_s the rotational frequency of the bearing shaft, β the contact angle between the raceway and the rolling element, BD the roller diameter and PD the bearing pitch diameter. The dimensions are shown in Figure 3.5. Many bearing monitoring techniques are based upon extracting these frequencies from an enveloped signal thus indicating which component is damaged. These equations are approximations based on the assumptions of pure rolling contact and minimal elastic deformation. Defect frequencies of bearings operating under normal conditions are not expected to deviate significantly from those calculated given by Equations 3.8-3.10 however those subjected to light loads and high shaft speeds have been reported to experience slip at levels up to 50% [58].

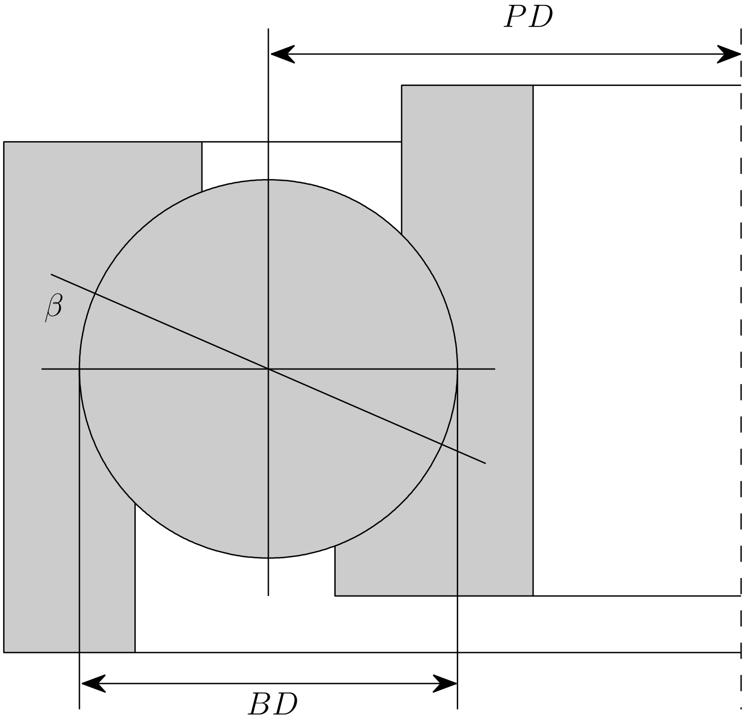

Figure 3.5: Bearing Dimensions

High Frequency Resonance Technique

The high frequency resonance technique (HFRT), sometimes referred to as envelope analysis, is one method to overcome the difficulties of noise and modulation presented in Section 3.2.2 and is based upon the principle that when an element passes over a defect the system generates a resonance at a much higher frequency than those generated by other machine elements. Therefore, the additional energy is concentrated in a narrowband which can be more readily detected than in the entire spectrum. The resonance is generally considered to be amplitude modulated at the characteristic defect frequency of the damaged component [59]. The aim of the HFRT is increase the signal to noise ratio and demodulate the signal in order to analyse the envelope signal.

The HFRT involves three stages. The process is shown graphically in Figure 3.6.

Firstly, the measured signal shown in (a) which contains unwanted broadband noise in addition to the frequencies generated by the defect, the spectrum of which is shown in (b), is bandpass filtered. The bandwidth contains the resonant frequency, boosting the presence of the defects in the time domain (c). The corresponding spectrum is shown in (d). Secondly, the filtered signal is enveloped to find the defect frequencies. This can be done efficiently via the Hilbert transform (e). Finally, the spectrum can be analysed since, in theory, any non zero values will occur at integer harmonics of the defect frequency (f). The enveloped signal can be downsampled at this point to reduce storage requirements as the resulting frequencies will be lower than those in the original measured signal.

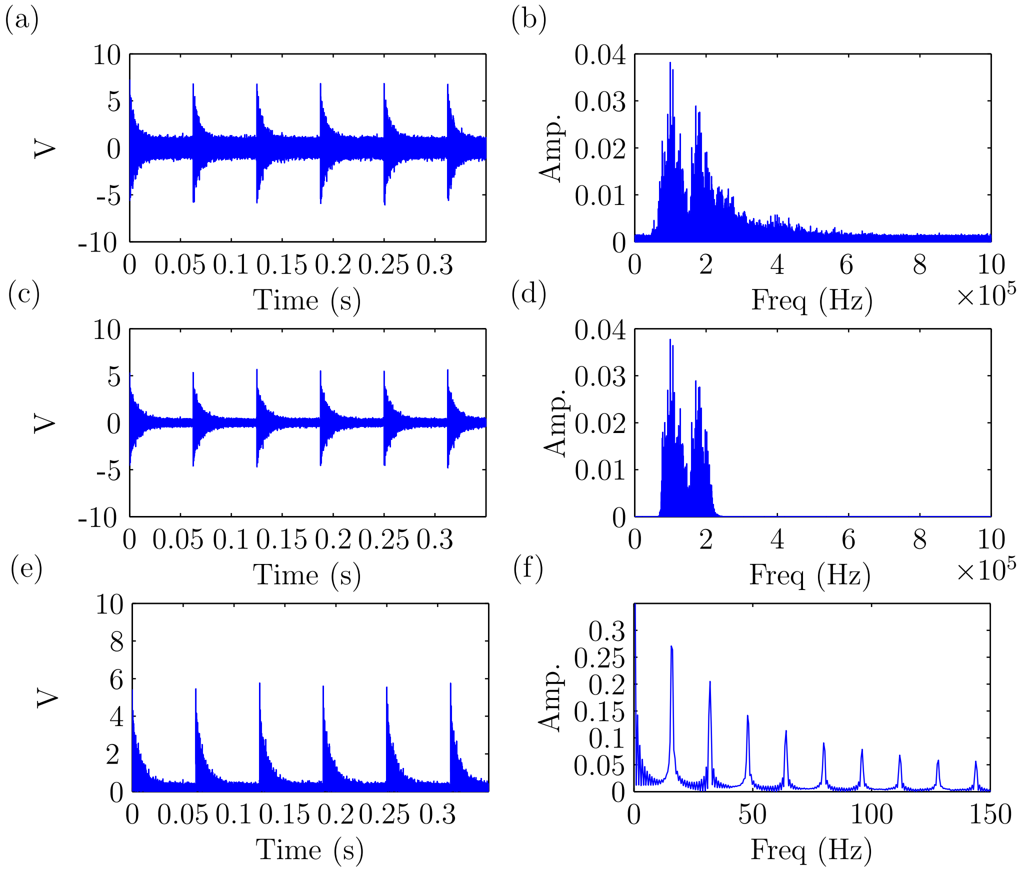

Figure 3.6: HFRT Procedure

Cepstrum

Cepstrum is the general term used to refer to the inverse Fourier transform of the logarithm of a signal's spectrum [60] and provides a measure of periodicity in a signal.

Such analysis can be used to indicate the presence of multiple faults which may not be clear in an FFT derived spectrum. A number of cepstrum exist including the real, complex and phase however the power cepstrum given in Equation 3.11 was first defined by Bogert et al [61] in 1963 and has found many applications in the condition monitoring of rotating machinery. An example of the cepstrum is shown in Figure 3.7. The independent variable here is called the quefrency which has the units of time. Any periodicity in the signal will manifest itself as a peak at the reciprocal of the quefrency. The peak in Figure 3.7 at 0.066s corresponds to the defect frequency of 15.1Hz. it should be noted that the plot has been zoomed in to show the peak clearly.

$$Power\ Cepstrum = |\mathcal{F}^{-1}\{log(|\mathcal{F}\{f(t)\}|^2)\}|^2 \tag{3.11}$$

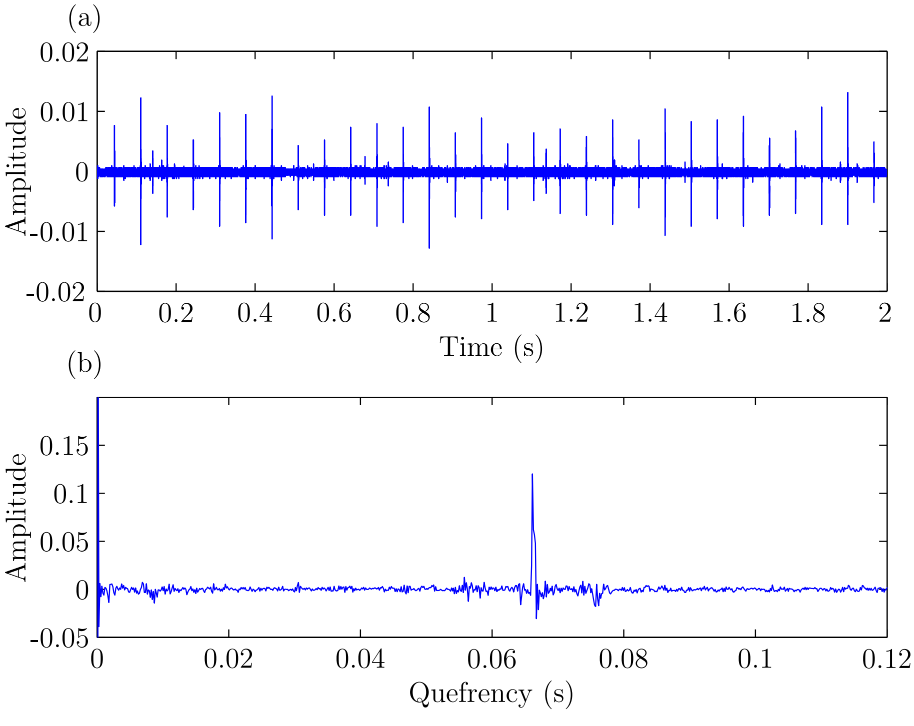

Figure 3.7: Measured data from (a) a damaged bearing and (b) the corresponding cepstrum

3.2.3 Time-Frequency Domain

The Fourier transform assumes that the signal being analysed is stationary. Since defect signals are transient in nature, a better approach may be to analyse signals in the time-frequency domain.

Short Time Fourier Transform

The short time Fourier transform (STFT) is the simplest method of analysing the frequency content of non-stationary signals. A window function, often a Hann or Gaussian window, is multiplied by the signal and the Fourier transform is calculated. The window is shifted in time to generate a 2D time-frequency representation of the signal which can be viewed graphically as a spectrogram, an example of which is shown in Figure 3.8 where the defect signals can be seen as an increase in frequency content between 0 and 400kHz at approximately 0.045 and 0.11 seconds where the rolling element impinges on the defect. The principal drawback of the STFT is one of resolution. A compromise between time and frequency resolution must be met by choosing a suitable window width for the application in question. A wide window gives good frequency resolution where similar frequencies may be separated in the spectrum however it is impossible to know at which point in the window each frequency occurs. The length of window can be reduced to improve the time resolution. Taking this to its extreme the window could be reduced to a Dirac impulse, the Fourier transform of which contains all possible frequencies meaning the frequency components will be smeared across the frequency axis. This compromise describes Heisenberg's uncertainty principle which states it is impossible to know the exact frequency of a signal at any given point in time.

Wavelet Analysis

Given the drawback of the STFT, the next logical step is to take an approach whereby the size of the window is no longer fixed allowing the data to be analysed at different scales. This idea forms the basis of wavelet analysis, a relatively recent mathematical development which has since found diverse applications in areas of signal processing such as noise reduction, data compression and the analysis of non-stationary signals.

In comparison to Fourier analysis whereby the basis function is a sine or cosine wave which by definition extends from minus to positive infinity and therefore is suitable for analysing stationary signals, the basis function for the wavelet transform is a wavelet which has been defined as “a ‘small wave’, which has its energy concentrated in time to

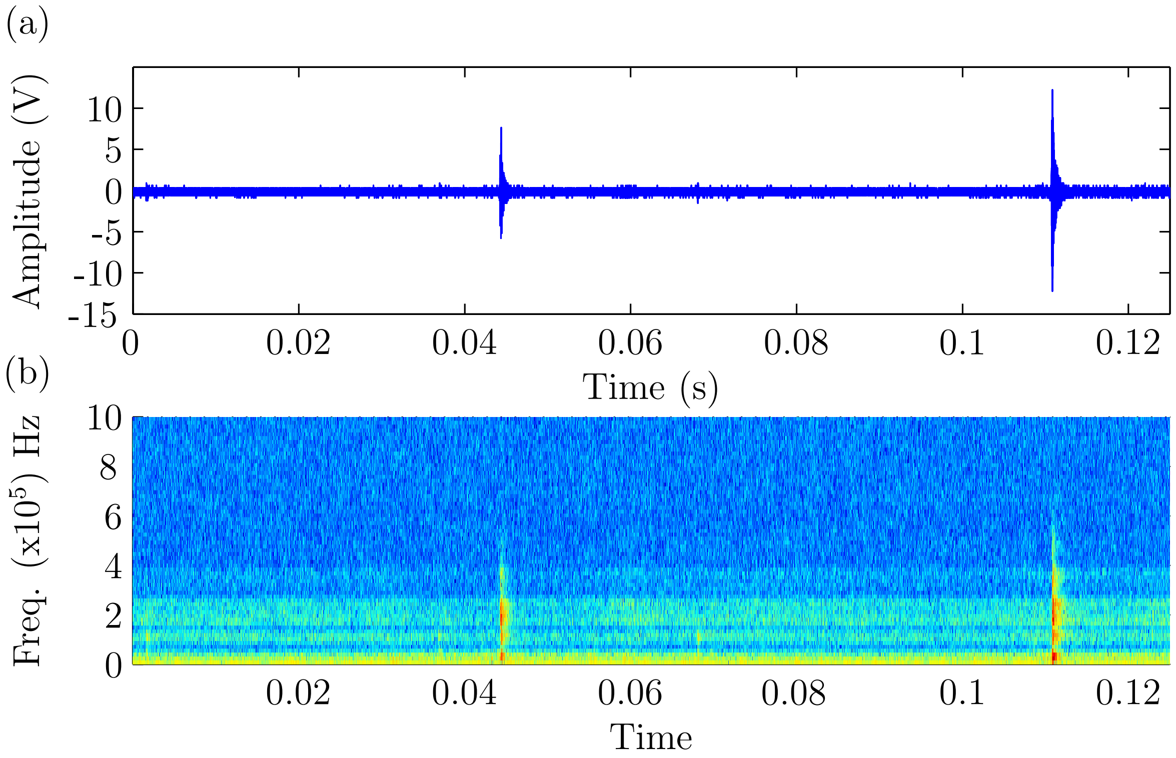

Figure 3.8: Measured data from (a) a damaged bearing and (b) corresponding spectrogram

give a tool for the analysis of transient, non-stationary, or time-varying phenomena" [62]. A number of common wavelets exist and the shape of the wavelet is largely determined by the application with researchers often attempting to match the wavelet to features in the signal of interest. Previous researchers, for example, have used the Morlet wavelet in detecting bearing damage due to its similarity to an impulse signal [63, 64]. Four commonly used wavelets are shown in Figure 3.9. In their unmodified form these wavelets will be referred to as mother wavelets denoted by $\mathit{\Psi}$. Although there is a potentially infinite choice of wavelets, they must fulfil certain criteria such as integrating to zero and having finite energy, properties which mean the wavelet must be oscillatory in nature. Wavelets of similar properties form a part of a family of wavelets, the name gives the family while the number denotes the how many vanishing moments.

The mother wavelet can undergo a number of translations and dilations. The resulting daughter wavelet is related to the mother wavelet by Equation 3.12 where τ and s govern the translation and dilation respectively. As will be shown, time resolution is governed by translation τ while s determines the frequency, or more

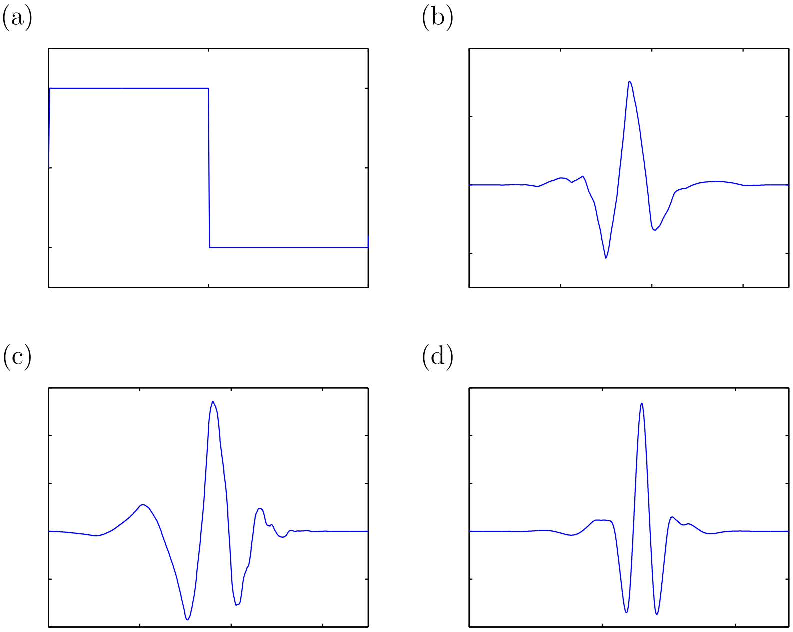

Figure 3.9: Examples of wavelets (a) Haar (b) Symlets 4 (c) Coiflets 4 (d) Daubechies 4

accurately scale, resolution of the analysis. The constant $\frac{1}{\sqrt{s}}$ is a normalising factor to preserve energy of the transform.

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\Psi(\frac{t-\tau}{s}) \tag{3.12}$$

The continuous wavelet transform (CWT) is generated by calculating the inner product of the daughter wavelet and the signal $x(t)$ for every possible combination of τ and s according to Equation 3.13

$$WT_{s,\tau} = \int_{\mathbb{R}} x(t)\Psi^{*}_{s,\tau}(t)dt \tag{3.13}$$

where * denotes the complex conjugate. The coefficients therefore provides a measure of similarity or correlation between the daughter wavelet and the corresponding section of the signal. A visual representation of these coefficients is called a scaleogram. Lin and Qu [63] explored the use of the CWT applied to the vibration signals of bearings and gears. Wavelet parameters were optimised by minimising the entropy of the transform and it was found that a superior denoising results could be achieved using the Morlet wavelet compared to Donoho's soft thresholding method [65].

Discrete Wavelet Transform

To calculate the wavelet coefficients at every scale and translation is computationally expensive. The discrete wavelet transform (DWT) has been developed as a practical means of overcoming this problem. Rather than compute wavelet coefficients for every possible value of τ and s, it is possible to interpret the transform as a subset of those which will accurately represent the contents of the data so that Equation 3.12 becomes

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi(\frac{t - k\tau_0 s_0^j}{s_0^j}) \tag{3.14}$$

for integers j and k. In order to maintain the possibility of reconstructing the signal this subset may be sampled on a dyadic grid whereby $s_0 = 2$ and $\tau_0 = 1$. A coarser sampling would prevent the signal to be accurately reconstructed.

In reality, the DWT is performed by convolution of the signal $x(t)$ with the wavelet coefficients thereby acting as a low pass finite impulse response (FIR) filter $g(z)$ where the bandwidth is a quarter of the sampling frequency. The signal is simultaneously filtered by a high pass quadrature mirror filter $h(z)$. Since a quadrature mirror filter is a filter whose frequency response is a mirror image of another filter around $\pi/2$, the resulting bandwidth covers the remainder of the spectrum. The initial signal $x(t)$ is represented by two new signals of equal lengths to $x(t)$ which introduces a level of redundancy. To overcome such redundancy, both signals are downsampled by two. The result is a series of approximation coefficients (cA) and a series of detail coefficients (cD). This process can be repeated any number of times whereby the approximation coefficients are decomposed at each level until the number of coefficients is reduced to two, each time the time resolution is halved and the frequency resolution is doubled. For n levels of decomposition, $n + 1$ sets of coefficients will be generated. A two level DWT is shown in Figure 3.10. An important property of the DWT is that each set of coefficients can be reconstructed to form a signal of equal length to the original $x(t)$ relating to its frequency band so for example cAA of $N/4$ samples can be reconstructed to AA of N samples. The original signal can then be restored through the addition of all reconstructed signals from the same level. For example, $x(t) = A + D$ after a single level decomposition. The use of the DWT has been shown to be effective in isolating the defect frequencies of bearings with single and multiples fault on both the inner and outer race [66].

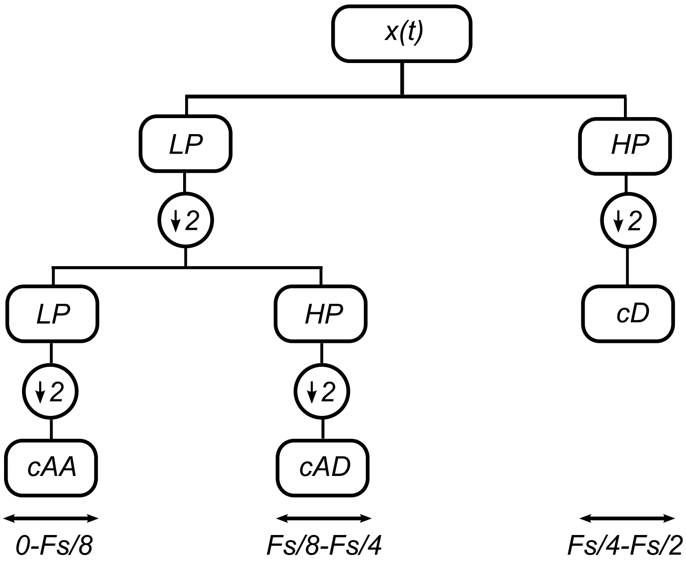

Figure 3.10: Two Level DWT

Wavelet Packet Decomposition

Wavelet packet decomposition (WPD) is a generalisation of the DWT which offers a richer degree of analysis. Rather than filtering the approximations at each level, the WPD high and low pass filters both the approximation and the detail resulting in 2^n sets of coefficients for an n level decomposition compared to $n + 1$ as would be the case for the DWT. Figure 3.11 shows a two level decomposition with the extra sets of coefficients compared to Figure 3.10. The bandwidth is equal for all sets of coefficients in each level. Nikolaou and Antoniadis take this approach employing an automated selection criterion to find the vector where the characteristic defect frequency is most prominent [67].

3.3 Review of AE Monitoring of Rolling Element Bearings

Conventional vibration analysis is a well established method for the diagnosis of rolling element bearings. Many parameters and signal processing techniques have been investigated, the majority of which are either directly applicable to AE data or initially

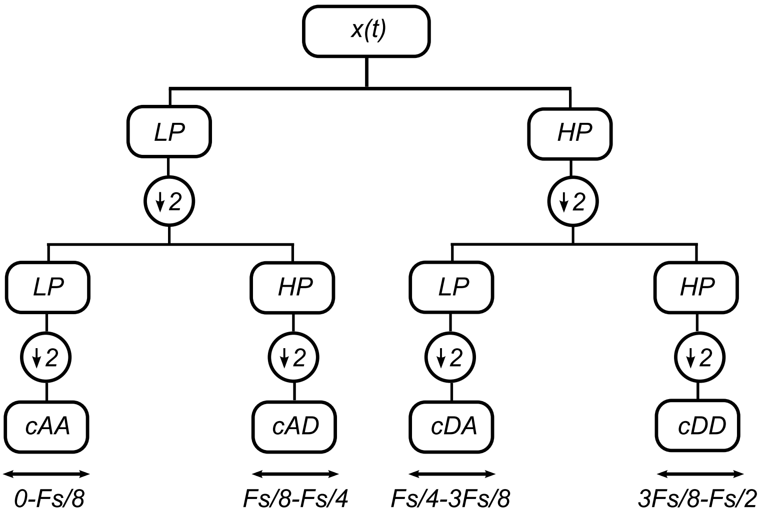

Figure 3.11: Wavelet Packet Decomposition

require a degree of preprocessing. Rather than providing a complete review of such techniques, the following section chronologically introduces methods explored by AE researchers for bearing monitoring, some of which being unique to AE technology.

Modern AE research began in the 1950s however, the first known study of acoustic emissions from rolling element bearings was performed by Rogers [68] in 1979 following concern regarding the safety and performance of slewing cranes on offshore gas production platforms. The use of AE was initially proposed as vibration analysis in the range of 1 to 20kHz and strain gauges were not deemed sensitive enough to detect incipient damage and fatigue cracks in slow speed bearings. Two diametrically opposed sensors were bonded to the slew ring and discrete AE events were recorded whilst the crane was in motion. A fixed threshold of one volt was used to count events. The amplitude and spatial distribution of events above the level of background noise could then provide a basis against which bearings of an unknown state could be compared. No comment was able to be made regarding the sensitivity of AE in detecting damage since only an undamaged bearing was monitored.

In 1982 Smith [69] carried out a comparison of AE monitoring with not only vibration analysis but also shock pulse metering (SPM) and jerk. The stationary outer race of a test rig was instrumented with a high frequency accelerometer, a shock pulse transducer and an acoustic emission transducer while the inner race was seeded with a 20μm scratch. Furthermore a back-to-back gear test rig was mounted on the same base in order to provide some typical background noise. In the first instance the bearing test rig was run without the gear test rig running and it was found that the signal to noise ratio of each method was similar. However, when the noise from the gears was present, it was not possible to visually detect the defect from the acceleration or acoustic emission data below 1000rpm. The SPM and jerk measurements showed it was capable of detecting defects at speeds above 500rpm. Additional experiments were performed at very low speeds and, even in the absence of the background noise, the defect signals of all methods, the exception of acoustics emissions, were not detectable in the electrical noise of the measurement system and the transducer was responding to every time a ball came into contact with the defect. It was suggested that this was due to the sensitivity of the transducer to base bending. Although no lower speed limit was established, the defect was detectable from 10rpm up to approximately 300rpm. This study suggests that the use of acoustic emissions should be considered in slow rotating bearings.

A direct comparison between AE and conventional vibration monitoring was made by Hawman and Galinaitis [70] to determine the minimum detectable defect size of each method. The digitised signal of a single AE sensor bonded to the stationary outer race of a ball bearing was analysed in both the time and frequency domain. Defects were scratched on the outer race, the size of which was increased after each test. The inner race was rotating at a fixed speed of 5700rpm, much faster than previous studies. Peaks in the spectrum of the demodulated signal at the outer race defect frequency indicated that AE was capable of detecting defects with a width of 0.001in (0.0254mm). The spectrum of the vibration data did not clearly indicate the presence of a defect until the width had reached 0.065in (1.65mm), an increase of over 6 times. At this point the RMS had increased by 38% indicating that AE is not only more suitable for detecting incipient damage in high noise environments but is also capable of indicating which component is damaged.

The early detection of defects was investigated through the use of advanced signal processing and pattern recognition techniques by James Li and Li [71]. Bearings with and without damaged components were monitored with AE sensors of bandwidth 20-90kHz, lower than is used in most AE monitoring, in order to reduce the amount of data after sampling. The ‘peaks and valleys’ derived from the auto-correlation of two features, namely the short-time energy function and the zero crossing rate, were used as inputs to a pattern classifier which plotted the two features and attempted to find a boundary separating the damaged from undamaged bearings. The classifier was trained by a gradient descent algorithm to label new data as damaged or not. A success rate of 94% indicated a significant improvement over vibration monitoring and the previous state of the art AE technique HFRT whose success rate was reported to be 86%. Furthermore, the study both demonstrated the success of monitoring damage on rolling elements which introduces an additional challenge due to its time varying position and transmission path and found that signal to noise ratios were stronger in AE transducers than accelerometers when placed remotely to bearing.

The aim of an investigation by Shiroishi et al [56] was to not only indicate the presence of a defect by AE monitoring but also attempt to establish a relationship between signal features and defect size. A cup and cone Timken bearing operating at 1200rpm was scratched with a diamond scribe to replicate natural damage. The HFRT and an adaptive line enhancer (ALE) were used to enhance the signal and increase the signal to noise ratio (SNR). Results were compared with both commonly used statistical parameters and those obtained from vibration signals processed in the same manner. Through the use of a peak ratio, defined as the ratio of the sum of the defect frequency and its harmonics to the average of the whole spectrum, the minimum detectable defect size was found to be 38μm; larger than the smallest defect size detectable by vibration analysis; 15μm. No meaningful correlation was found between any of the AE features investigated and the defect width whereas the peak value of the accelerometer signal demonstrated a strong relationship. Furthermore, it was recommended that time domain methods are employed when monitoring for distributed damage.

The monitoring of contaminants or starved lubrication introduces a further challenge since induced vibrations are not periodic and cannot, therefore, be monitored by traditional methods such as enveloping the measured signal and tracking the characteristic defect frequencies. Miettinen and Andersson [72] compared the use of AE with a number of vibration based methods to monitor bearings lubricated with contaminated grease. Under laboratory conditions a number of bearings were tested with and without contaminated grease. The pulse count method and RMS were used as AE measures. A clear increase in AE activity was demonstrated with the introduction of contaminants and amongst the methods investigated, AE was found to be the most sensitive. Further laboratory investigations by the same authors found that weight concentrations of contaminants as low as 0.02% could induce a clear increase in AE activity however such a relationship did not increase linearly with concentration as the pulse count measure became saturated [73]. The results also showed greases contaminated with the same proportion of ferrous material gave rise to higher levels of acoustic emission and that bearings replaced with clean grease exhibited higher levels of AE activity compared to a new bearing with clean lubricant. Miettinen and Leinonen [74] recognised the limitations of laboratory investigations and carried out a study aiming to determine how wear particles can be monitored in a field environment and to gain some experience in the type and level of disturbance that can be expected. In order to differentiate changes in AE activity arising from the bearing from external influences a reference sensor was used that was in the vicinity of but not bonded to the bearing. Measurements were taken from four scenarios; from a bearing which had been re-greased two years prior, the same bearing with clean grease, following the introduction of contaminated grease and finally, after two months of running in normal production with the contaminated grease. The results clearly show how environmental disturbances influence AE activity and the reference sensor, used in pulse count mode, could be used to correct the data from the bearing sensor. It was also concluded that from the measures investigated, kurtosis is the most sensitive to the introduction of contaminants into the grease.

The use of ring down counts, as described in Section 2.5.2, for defect detection was evaluated by Choudhury and Tandon [75]. The sensitivity to operating conditions was determined as bearings of different bore diameters were run under a range of loads and speeds. An automatic threshold was employed to eliminate the variance in background noise. The value of counts exhibited a sharp rise over the range of smallest tested defects but showed no further increase once the defect had progressed to 0.5-1mm indicating the parameter would be useful for defect detection but not size estimation. The results have been reproduced in 3.12. The number of ring down counts was shown to be insensitive to bearing size and applied load but increased with rotational speed as presented in Figure 3.13.

Further investigations into emissions from low speed machinery were presented by Jamaludin et al [76]. Two explanations for the difficulties in monitoring bearings at low speeds are presented: firstly, that acceleration is the most widely used parameter to measure vibration which reduces with a reduction in rotational speed. It is suggested that below 600rpm, displacement is a more suitable parameter. Secondly, at low speeds, the low frequency components of a defect signal are likely to be lower than the cut-off frequency of the high pass filters used in many instrumentation systems

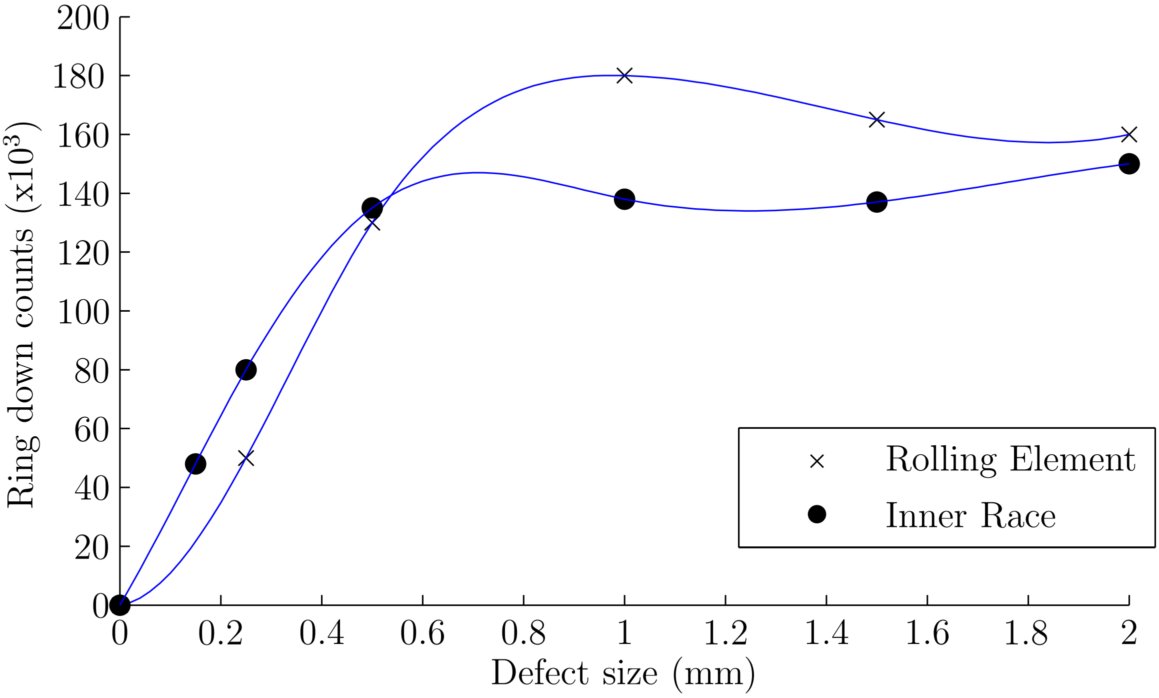

Figure 3.12: Ringdown counts at 75kg and 1500rpm [75]

to reduce inherent instrument noise. Previous studies had tested bearings operating at rotational speeds of no less than 10rpm [69]. Bearings were run at 1rpm with defects generated by electrodischarge machining (EDM) and spark erosion to imitate the progression of damage. However at this speed it was not possible to detect defects representative of early damage. The measured signal was represented by a number of auto-regressive coefficients which provided an efficient means of classification into damage states.

Mba validated the use of RMS to aid diagnosis and extended the work of Choudhury and Tandon [75] by determining the significance of the threshold on AE counts for the suitability of bearing monitoring [42]. Bearings, seeded with artificial defects, were run under a range of speeds (600-3000rpm) and loads (0-4.8kN). It was concluded that RMS provides a robust means of bearing diagnosis, generally demonstrating an increase with increasing load, speed and defect size. The relationship between the mechanical integrity of the bearing and the value of AE counts was found to be independent of the threshold which was calculated as a percentage of the peak amplitude for each loading case. A similar procedure was undertaken by Morhain and Mba [77] however the use of maximum event amplitude and energy was also investigated. In addition

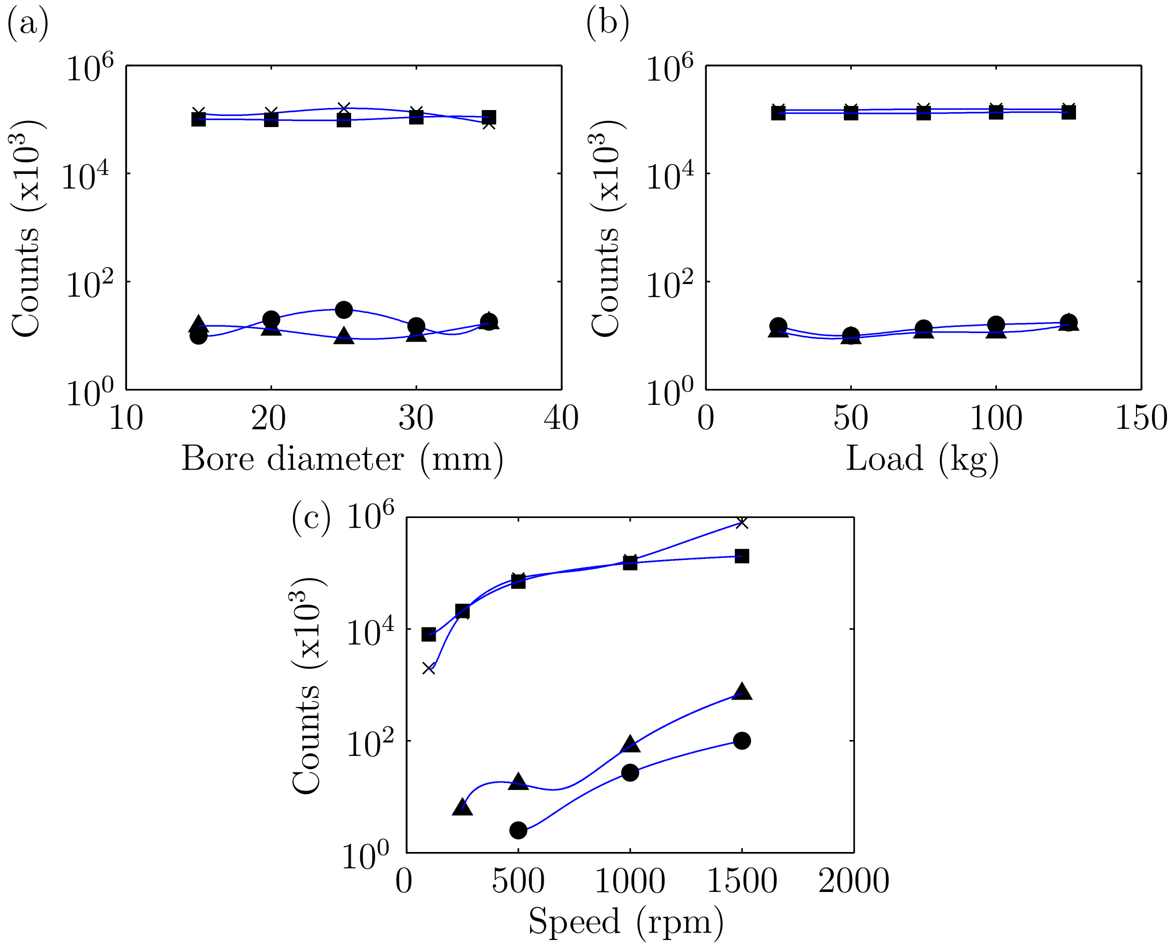

Figure 3.13: Ringdown counts under varying conditions ●Rolling element (no defect) ▲ Inner race (no defect) x Rolling element ■ Inner race [75]

to RMS and counts, the energy of AE events was found to correlate with load, speeds and defect size however such a relationship did not exist for the maximum amplitude.

Although Al-Ghamd and Mba [78] reinforced the use of commonly used AE parameters as a robust means of bearing diagnosis, their most significant findings related to the fundamental source of AE activity and an attempt to establish a relationship between defect size and signal features. Defects were seeded by a carbide tip engraving machine with and without protrusions above the mean surface roughness of the bearing raceway. Since the presence of smooth defects could not be distinguished from the bearings with no defect it was assumed the principal source of AE activity was a result of direct material contact. A direct relationship between defect length and burst duration was established while a comparison was made between AE and vibration analysis as a means of determining defect size. Figure 3.14 shows how both the

AE RMS and maximum amplitude increase with defect sizes while the bearing was rotating at 2000rpm, D1(0.85x1.35mm)-D7(13x10mm), whereas the equivalent vibration measures decrease at the largest defect sizes; a finding which may prove significant for bearing prognostics.

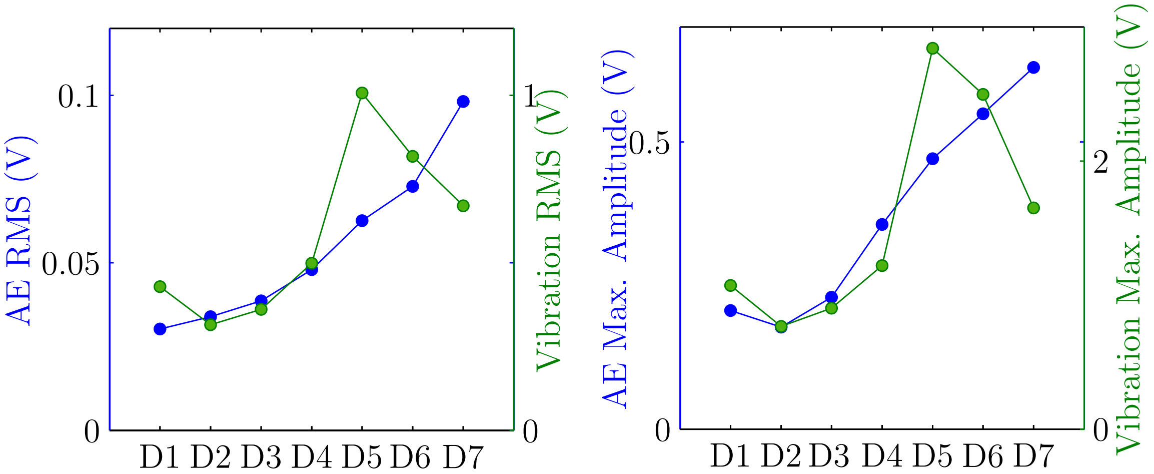

Figure 3.14: Comparison of AE and vibration RMS and maximum amplitudes [78]

In response to the high failure rates experienced by ball bearings in induction motors, Tandon et al [79] carried out a comparison of the commonly used condition monitoring methods; vibration, stator current, SPM and AE measurements at 1440rpm. AE measurements were shown to exhibit the greatest percentage increase in peak amplitude compared to the equivalent parameters measured for the other methods for both the minimum and maximum defect size tested.

Couturier and Mba [80] first investigated the importance of the specific film thickness λ in relation to AE levels under near iso-thermal conditions. The RMS of a grease lubricated Cooper split bearing was recorded whilst varying speed and load. As expected the RMS demonstrated an inverse relationship with λ; as the load increases, the film thickness decreases therefore increasing the number and severity of asperity contacts. However, a similar relationship did not exist whilst varying the speed, rather the opposite. Although higher speeds result in larger film thicknesses, it was postulated that the resultant increase in strain rates experienced by asperities in contact was the reason for the increase in AE RMS indicating that AE activity is not merely a function of specific film thickness. It is also suggested that as speed increases further and bearing moves towards hydrodynamic lubrication, AE levels will then decrease as no asperity contact is taking place.

The majority of previous studies have focused on validating the use of AE technology for condition monitoring by correlating the size of seeded defects or operating conditions with parameters characterising the defect signal. Elforjani and Mba [81] performed accelerated life tests on thrust bearings to monitor natural degradation. One raceway was replaced with a flat plate to increase contact stresses and induce fatigue. Traditional parameters were monitored but the authors recognised the limitations of these methods when monitoring signals exhibiting transient behaviour. To overcome this, the concept of the energy index (EI) and the K-S test were investigated. Both methods were shown to be more sensitive to the onset of fatigue when compared to the signal energy calculated over a predefined time constant.

Al-Dossary et al [82] aimed to further the work of [78] by investigating the relationship between defect size and various AE waveform parameters. A correlation between defect size and AE burst duration was demonstrated for outer race defects however such a relationship could not be observed for inner race defects due to, it is believed, variation in the transmission path as the bearing rotates. This difference is shown in Figure 3.15. The theoretical values, based upon the size of the defect and the relative speed of the passing elements, is consistently smaller than the measured burst duration. This is assumed to be a result of the signal decay which is not taken into account when calculating the theoretical values.

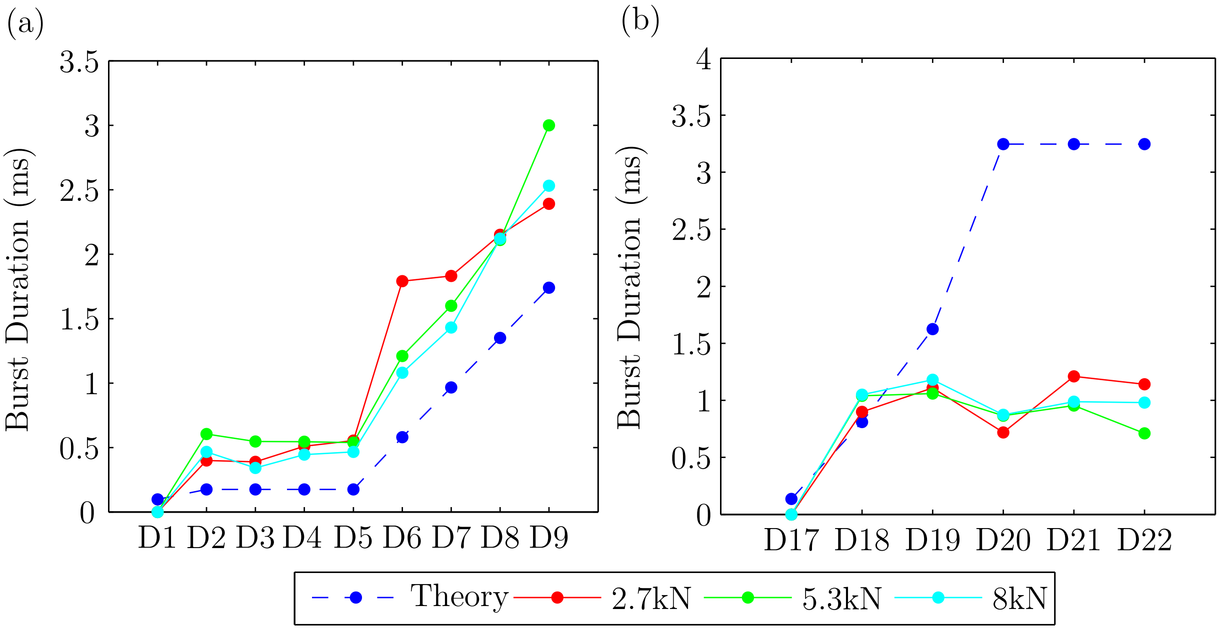

Figure 3.15: Burst duration for various defect sizes (a) outer race 3000rpm (b) inner race 3000rpm [82]

Defect frequencies were extracted by the short-time RMS and autocorrelation functions in a study by He at al [83]. A close match was noted between the theoretical and measured frequencies confirming the use of such a method to determine which component is damaged. The sensitivity of various AE parameters was investigated under a range of operating conditions. Radial load was shown to be insensitive to AE counts, energy, amplitude and kurtosis whereas rotational speed had a strong influence on all of the parameters reinforcing the findings of Choudhury and Tandon [75]. Increasing defect size was responsible for a less marked increase suggesting that such parameters may not be suitable in situations of variable speed and more complex signal processing techniques may be more suitable.

Widodo et al [84] developed a technique to successfully diagnose and classify bearing damage in low speed machinery rotating at a minimum speed of 10rpm. The bearings were seeded with defects representing both cracks and spalls on the inner race, outer race and rolling elements. The defects, which were created by a diamond cutting bit and an air-speed grinding tool, ranged in widths from 0.1-1.6mm. AE and vibration signals were enveloped and a number of time and frequency domain parameters were extracted. The dimensionality was reduced through component analysis to increase the performance of the classifier. The suitability of a multi-class relevance machine (RVM) and a support vector machine (SVM) was compared. The RVM was able to successfully differentiate between hairline cracks and spalls on the inner and outer race and rolling element defects. Furthermore, it was noted that AE signals were more sensitive to damage than vibration signals at rotational speeds less than 80rpm.

The concept of EI, initially investigated by [81], was examined in greater detail by Al-Balushi et al [85]. Defect signals were simulated by superposing Gaussian white noise upon a sequence of rapidly decaying high frequency sinusoidal waves; in a similar manner to the process followed in Section 5.1.1. By varying the amplitude of the bursts it was found that the EI could pick out transients when the SNR was 25%. The technique was also applied successfully to experimental signals contaminated with noise. It was suggested that by choosing a suitable window length the effect of electrical noise, which tends to be of shorter duration than defect signals, can be minimised.

Further parameters were investigated by Elforjani and Mba [86] to increase sensitivity to bearing degradation. In particular, information entropy (IE) was demonstrated to outperform both kurtosis and crest factor. Additionally, three non-linear power spectral methods; Prony energy, eigen-analysis and auto-regressive spectra, were investigated due to limitations of the FFT: notably leakage and resolution. Peaks appeared in all three spectrum before the FFT while the Prony energy spectrum exhibited the best performance. An attempt was made to provide a relationship between burst duration and defect length on naturally degrading bearings however no indication of the accuracy was provided. Maps of the coefficients derived from the continuous wavelet transform (CWT) were presented to provide a time-frequency representation of the signal. These maps provided an indication of periodicity at the defect frequency but no further use was made of this potentially useful tool.

3.4 Conclusion

The decision to employ one method of NDT must take into account, amongst other factors such as cost and ease of obtaining and analysing data, the sensitivity to damage.

A number of authors have attempted to determine whether AE is more sensitive to damage than conventional vibration analysis [56, 69, 70, 71, 78, 79, 84]. While it does not seem to be clear which method is more effective, there is a general consensus that AE is more suitable at low rotational speeds [69, 76, 84]. Kuboyama suggests that at high rotational speeds the defect vibrations have a large amount of energy concentrated over a short period of time whereas at lower speeds the vibrations have less energy spread over a longer period thus obscuring the defect signal in the background noise [87]. Jamaludin et al also state that conventional vibration systems tend to employ accelerometers whereas displacement sensors would be better suited to low-frequency analysis and the cut-off frequency of the high pass filters found in many systems is higher than the defect frequency of many slowly rotating bearings [76].

Another common thread running through many studies concerns the variation of AE parameters with operating conditions such as load, speed, defect size and bearing size. Although a number of common findings exist, many papers report seemingly contradictory results. For example, [75] and [83] found counts to be insensitive to load but strongly influenced by speed whereas [77] found that counts demonstrated a marked increase with load. It could be argued that film thickness or lubrication regime which is a function of all these conditions can explain AE activity better than any single condition. The only study to take this approach and focus on film thickness rather than individual conditions was performed Couturier and Mba [80] and could explain these apparent conditions. For example, an increase in speed may increase AE activity if the film thickness is too small to prevent full separation due to the increased strain rates. However a further increase may reduce emissions as the film thickness grows and no asperity contact takes place.

A handful of studies attempted to correlate defect size with various AE measures. The most successful of which were made by correlating the burst duration to the size of the defect [82, 86]. It must be noted however that such an approach will only work in those cases where the load on the defect is constant, therefore the damaged component must be stationary. No relationship could be established between defect size and AE parameters; RMS, kurtosis and crest factor [56].

Although not explicitly stated, it is assumed that previous studies have generally concentrated on cases where a full lubricant film is formed and no or little asperity contact takes place. The findings are likely to be significantly different under mixed or boundary lubrication where asperity contact could lead to increased emissions. Whilst similarities can be drawn between asperity contact and lubricant contamination which has been investigated, this appears to be a significant void in the field of AE bearing monitoring.

Chapter 4

Bearing Test Rig & Instrumentation

This chapter presents the experimental test rigs used throughout this thesis. Information regarding the design, the bearings and the seeded defects is provided followed by a description of the acoustic emission instrumentation and sensors.

4.1 MultliLife™ Bearing Rig

In order to both validate acoustic emissions as a means of monitoring bearing health and to test the concept of the MultiLife™ bearing, engineers at Ricardo have designed and built a full scale test rig to perform accelerated bearing life tests with loading conditions representative of those experienced by planetary support bearings. The rig has been designed to also allow the monitoring of the oil film thickness through ultrasound transducers bonded to inner surface of the inner race and is the focus of another doctoral project. The assembled rig and a sectioned schematic are shown in Figures 4.1 and 4.2 respectively.

The outer race of the test bearing is belt driven to reduce noise by a 7.5kW external motor and can reach a rotational velocity of 100rpm. The outer race is allowed to rotate by a larger self-aligning bearing which is shown in Figure 4.3 taken during disassembly of the rig. The inner race is fixed to the rig housing via a steel brace to prevent the inner race rotating during operation. A solid shaft fits through the inner race and

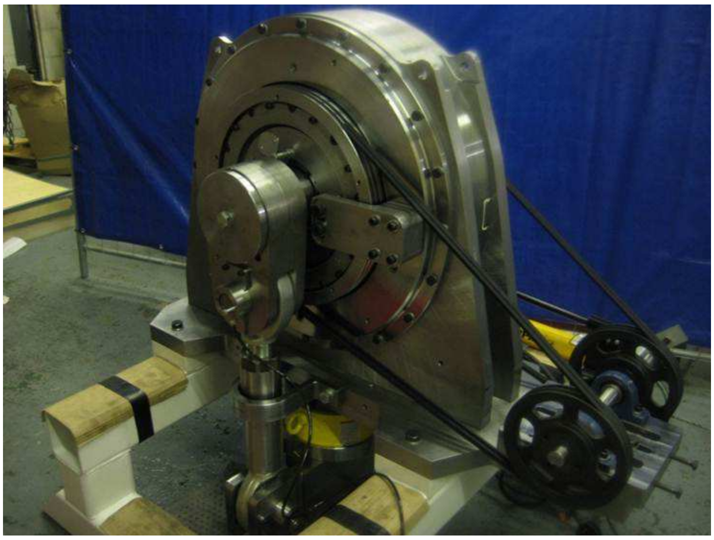

Figure 4.1: Photograph of Bearing Test Rig

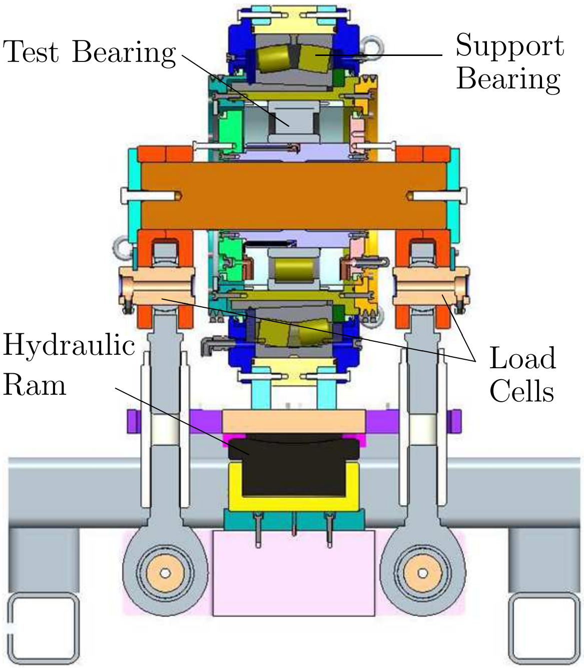

Figure 4.2: Schematic of Bearing Test Rig

is attached to two loading arms at either end designed to apply a purely radial load up to 1500kN. The inner race thus experiences uni-directional loading; representative of an epicyclic gear. Between the shaft and the inner race sits a sleeve which has been modified to allow sensors access to the inner race. Three equally spaced sensors, described in more detail in Section 4.3.1 are bonded to metal plugs machined to the correct radius and are spring loaded against the raceway. The rig was designed in this way as to allow the raceway to be rotated while the carrier and sensors remained stationary. A schematic and photo of the carrier is shown in Figures 4.4 and 4.5 respectively. The test bearing is pumped with a thin automotive oil of viscosity VG32 (see Table 4.1), whereas the support bearing is filled with a high viscosity VG320 oil in order to encourage the initiation of damage in the test bearing and prevent damage in the support bearing. The motor and oil pump is controlled by a PC running Labview which simultaneously records a low frequency log of bearing, inlet oil and ambient temperature in addition to oil pressure, applied load and rotation count of the outer race.

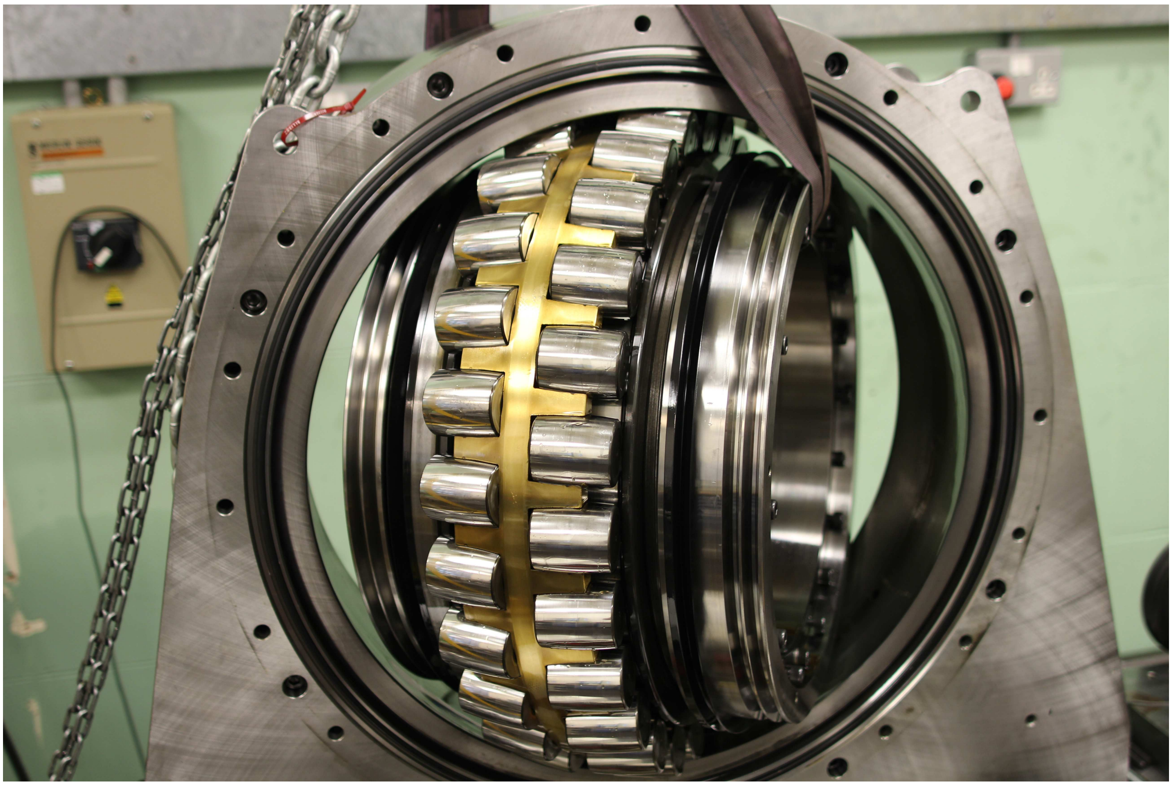

Figure 4.3: Support Self-Aligning Bearing

Table 4.1: Oil Viscosity

Temperature (° C)	Viscosity (cSt)
40	32
100	5.3

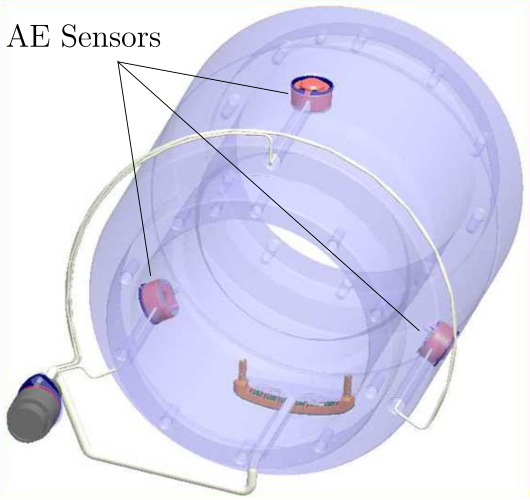

Figure 4.4: CAD Diagram of Sensor Carrier and AE Sensors

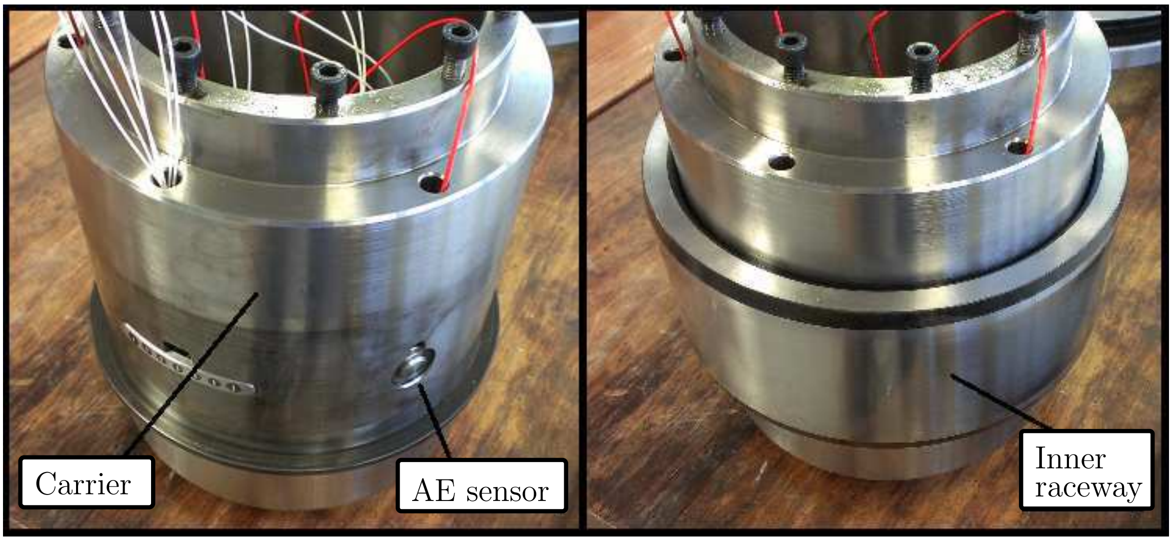

Figure 4.5: Sensor Carrier with and without Inner Raceway

4.1.1 Bearing

The specification of the cylindrical roller bearings (specification NU2244) manufactured by PSL in Slovakia and used for all testing on the MultiLife™ test rig is shown in Table 4.2.

Table 4.2: Bearing Specification

Dynamic load rating	1600kN
Static load rating	2300kN
Pitch diameter (PD)	313mm
Roller diameter (BD)	54mm
Inner race diameter	259mm
No. of rollers	15
Contact Angle	0°

Defect Frequencies

Using Equation 3.8 and the dimensions of the test bearing the inner race defect frequency can be calculated as a function of operating speed. Since the rotational speed experienced by a bearing in a wind turbine will not be constant, it is useful to express the defect frequency as a factor by which the shaft frequency f_s will be multiplied. This approach is generally referred to as order analysis rather than frequency analysis. In this case the inner race defect frequency is $8.79f_s$ or 14.65Hz when the f_s=100rpm. Likewise, the defect frequencies for the outer race and the rolling elements are $6.2f_s$ (10.33Hz) and $5.97f_s$ (9.95Hz) respectively.

Surface Roughness

In order to later calculate the specific film thickness of the bearing it is necessary to measure the surface roughness of the surfaces in contact. The profile of the inner raceway and a rolling element were measured with a 2D stylus based profilometer. Five measurements were taken at random points across both components only in the axial direction due to limitations of the profilometer. The value of R_q, the root mean squared roughness, given by Equation 4.1 was calculated for each measurement and averaged where n is the number of data points and y_i is the distance from the mean once the data has been corrected for waviness and tilt. The results are shown in Table 4.3. The composite surface roughness given by Equation 4.2 is therefore 0.195μm.

$$R_q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} \tag{4.1}$$

$$R_{q,T} = \sqrt{(R_{q,re}^2 + R_{q,ir}^2)} \tag{4.2}$$

Table 4.3: Bearing Surface Roughness (μm)

	Measurement No.						
Component	1	2	3	4	5	6	Average
Inner Race	0.131	0.172	0.181	0.193	0.175	0.128	0.173
Rolling Element	0.145	0.087	0.087	0.092	0.093	0.081	0.090

Seeded Defects

To minimise the length of each experiment and accelerate damage mechanisms the test rig will operate at a constant speed of 100rpm with a radial load of 1200kN; 75% of the bearing's dynamic load rating. The contact pressure between the inner race and the rolling element bearing can be calculated for a given load based upon the Hertzian theory of elastic contact. In the case of a cylindrical rolling bearing, the contact between the two components is assumed to be a line contact where the reduced radius R' can be derived from the radius of the rolling element R_1 and the radius of the surface of the inner race R_2 given by Equation 4.3.

$$\frac{1}{R'} = \frac{1}{R_1} + \frac{1}{R_2} \tag{4.3}$$

Using the dimensions provided in Table 4.2 the reduced radius is 22.3mm. For any applied load per unit length P', half the contact width b can then be estimated by Equation 4.4

$$b = \sqrt{\frac{4P'R'}{\pi E*}} \tag{4.4}$$

where $E*$ is the reduced modulus given by Equation 4.5.

$$\frac{1}{E*} = \frac{1}{2}(\frac{1-v_1^2}{E_1} + \frac{1-v_2^2}{E_2}) \tag{4.5}$$

Since the load applied to the bearing will be distributed over a number of rolling elements it is also necessary to calculate the load on the most heavily loaded element. The maximum load for rolling elements bearings with nominal clearance can be estimated by;

$$P_{max} = \frac{5F_r}{Zcos\alpha} \tag{4.6}$$

where F_r, Z and α are the radial load, number of rolling elements and contact angle respectively [88]. For an applied radial load of 1200kN, the most heavily loaded element is therefore subjected to 400kN. The width of the rolling elements is 108mm giving a load per unit width P' of 3.7MN. The full contact width $2b$ is therefore calculated as 1.35mm while the average p_{avg} and maximum contact pressures p_0 are 2.75GPa and 3.48GPa respectively according to Equations 4.7 and 4.8.

$$p_{avg} = \frac{P'}{2b} \tag{4.7}$$

$$p_0 = \frac{2P'}{\pi b} \tag{4.8}$$

Given such conditions and assuming the test bearing is well lubricated and free of contaminants, the L_{10} life for the NU2244 bearing, as calculated using Equation 1.1 on page 11, is approximately 8.7x10^6 revolutions. This corresponds to about 2 months when operating continuously at 100rpm. Clearly such a period is restrictive when attempting to perform a number of run to failure tests. It has, therefore, been deemed necessary to seed small defects on the raceway surface to act as a stress concentrator and thus an initiation point for further damage. An electrical-discharge machined (EDM) groove was eroded across the surface of the inner raceway on the two NU2244 bearings. The profile of the seeded defect measuring 100μm wide by 60μm deep taken from a scan using an optical interferometer is shown in Figure 4.6. It can be seen from Figure 4.6 that the erosion process has caused protrusions above the surface roughness on either side of the defect increasing the chance of metal-to-metal contact.

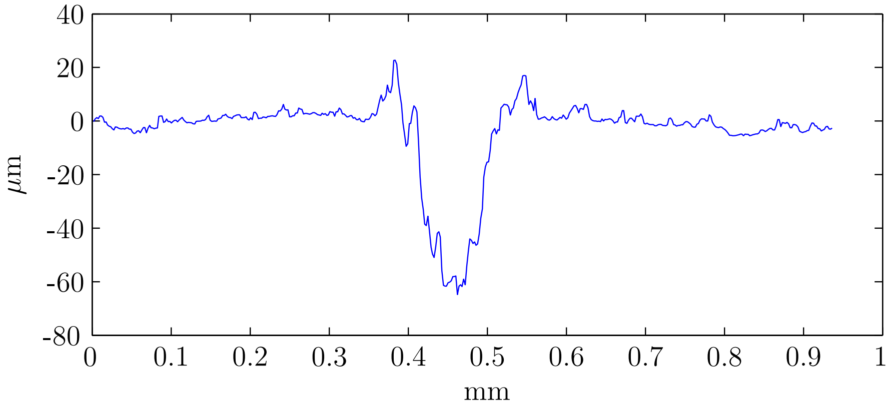

Figure 4.6: Profile of EDM Groove across Inner Raceway

4.2 Split Raceway Bearing Rig

Cooper Bearings manufacture split bearings whereby both raceways and the cage and roller assembly are each made of two semi-annular components to facilitate the installation and replacement of bearings since the shaft does not need to be removed during maintenance. The use of such bearings, but on a large scale, has been suggested to replace failed main shaft bearings on direct drive wind turbines to reduce downtime. One half of the cage and outer raceway are shown in Figure 4.7.

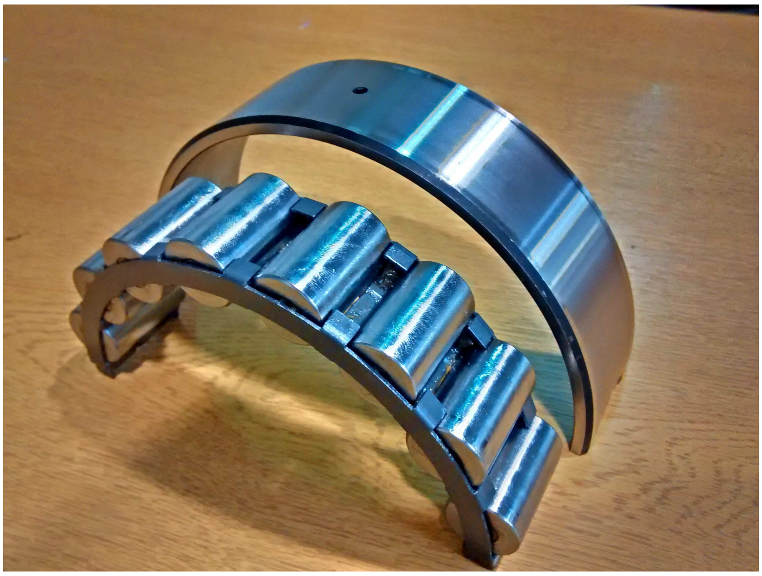

Figure 4.7: Photograph of Rolling Elements, Cage and Outer Race

Cooper Bearings have supplied the university with a test rig to test such bearings as shown in Figure 4.8. Two AE sensors were coupled to either end of the split outer raceway that is under load since damage is most likely to occur between these two points. Load is applied by a tightening a nut and measured by a load cell with a digital read-out. The bearing is grease lubricated. Results from this test rig are presented in Section 7.3. Two sensors were spring loaded and coupled against either end of the split outer race which was under load.

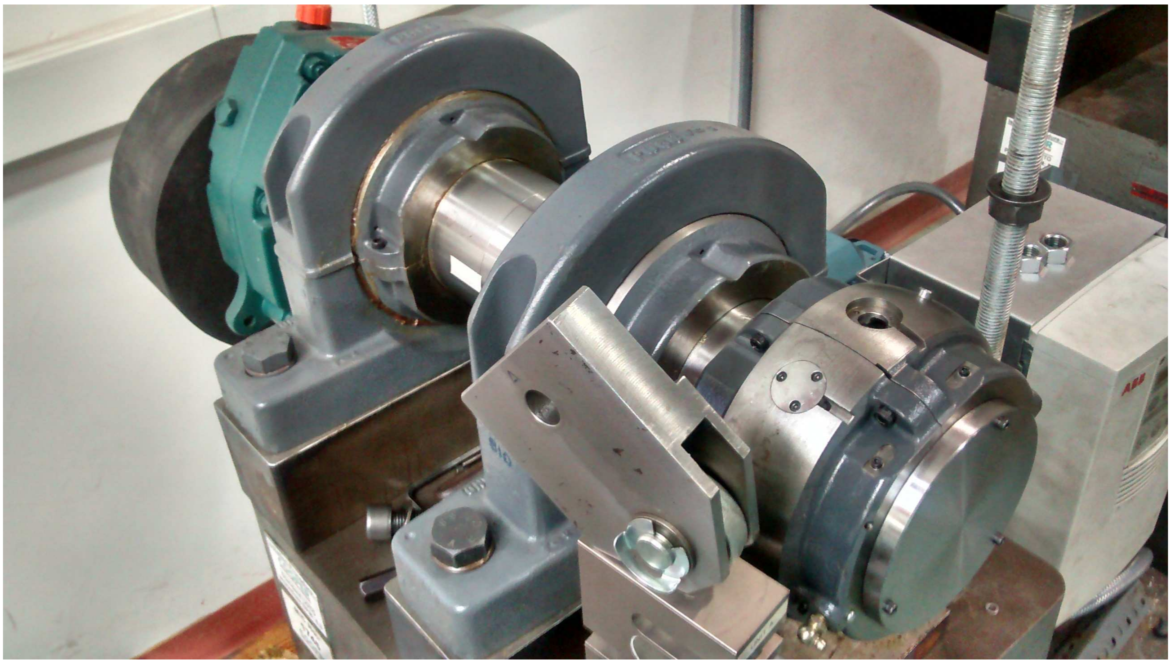

Figure 4.8: Split Race Bearing Test Rig

4.2.1 Bearing

The specification of the cylindrical roller split bearings (01E B 400EX) manufactured by Cooper and used for all the testing in this study is shown in Table 4.4.

Table 4.4: Cooper Split Bearing Specification

Dynamic load rating	320kN
Static load rating	421kN
Pitch diameter (PD)	138mm
Roller diameter (BD)	18mm
No. of rollers	16
Contact Angle	0°

The inner race, outer race and rolling element defect frequencies are therefore $9.05f_s$, $6.94f_s$ and $7.71f_s$ respectively.

4.3 Instrumentation

A schematic of the data acquisition employed for this work is shown in Figure 4.9. A Physical Acoustics Corporation PCI-2 based system was used to digitise and store the AE waveform data with a 16bit analogue/digital converter. The specialist software AEWin was used to record the waveform data and process the hit-based data. Each sensor was connected to a PAC 2/4/6 pre-amplifier with a selectable gain of 20, 40 or 60dB. An anti-aliasing filter was contained in each pre-amplifier. The system includes low and high pass filters which were set to the widest bandwidth of 20Hz-1MHz so the least information was lost during the acquisition stage. The sampling rate employed for all channels in this work was 2MHz. The same acquisition system was used on both test rigs. All subsequent processing was performed in Matlab.

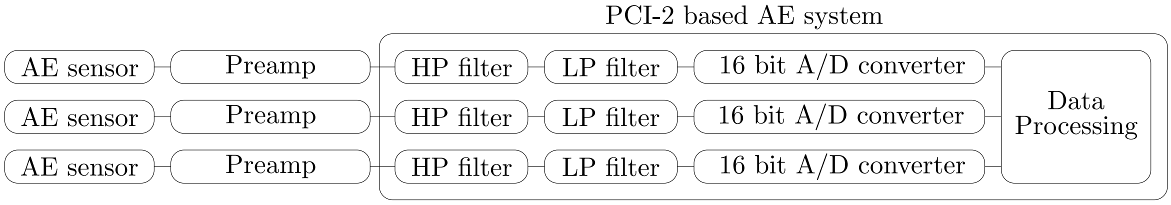

Figure 4.9: Schematic of DAQ system

4.3.1 Sensors

The sensors used on the Multilife™ test rig were three Physical Acoustics Nano 30 sensors measuring 8mm diameter by 8mm in height. The bandwidth of sensors is 125-750kHz with a resonant frequency of approximately 300kHz. A reproduction of the calibration certificate showing the sensor's frequency response curve is shown in Figure 4.10. The split bearing rig used two Physical Acoustic Pico sensors with a bandwidth of 200-750kHz and a resonant frequency of 250kHz. The Pico sensors measure 5mm in diameter by 4mm in height.

4.4 Conclusion

In this work, the majority of the testing is performed on a full scale bespoke test rig designed to replicate the loading scenario experienced by a planetary gear support

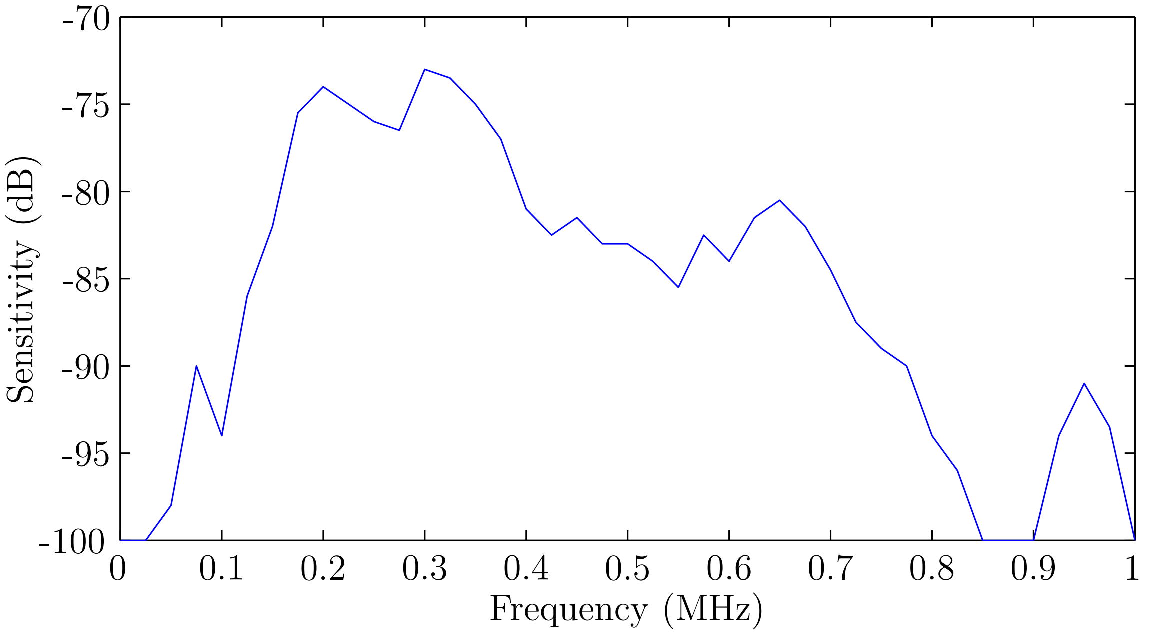

Figure 4.10: Sensitivity of PAC Nano 30 Sensor

bearing in a wind turbine gearbox whereby the inner race is stationary, experiencing uni-directional loads over a non-conformal contact. Operating conditions have been specified to reduce the expected life of the bearings as much as practically possible. Further tests are performed with a split bearing test rig. Both test rigs are instrumented with acoustic emission sensors and damage is seeded onto the raceways using electro-discharge machining.

Chapter 5

Feature Extraction and Damage Detection

This chapter presents a proposed method for determining the condition of a rolling element bearing using acoustic emission technology. The aim is to reduce the high dimensionality measured data to a vector of just a few dimensions whilst retaining the important characteristics of the original signal. Such an approach, reduces computational memory and prevents the overtraining of classification algorithms. The chapter initially describes the concepts at each stage of the proposed process and these are demonstrated using simulated defect signals. The creation of which is described in Section 5.1.1. A demonstration of outlier analysis, a technique used to highlight the degree of novelty of an unseen observation, again using simulated defect signals are included in this chapter. The techniques proposed in this chapter are applied to data from accelerated life tests later in Chapter 7.

5.1 Feature Extraction

Section 2.5 described a number of commonly recorded parameters when employing a hit-based monitoring system. This process relies on the amplitude of the defect signal being significantly higher than the background noise. In the case of incipient damage, the amplitude of such a signal is likely to be less or of a similar magnitude to the noise so such events will be missed. It is for this reason that in this work, raw waveform data, rather than hit based parameters, will be periodically analysed to determine the bearing condition. This section outlines the proposed process that will be followed when performing run-to-failure tests in Chapter 7. The aim of this process is to increase the signal to noise ratio, reduce the dimensionality of the signal for reduced storage requirements, demodulate the high frequency component of the defect signal and determine the presence of any dominate frequencies which relate to the characteristic defect frequencies presented in 3.2.2.

Initially, the use of the wavelet packet decomposition will act as a filter bank. The idea being that the defect frequencies will be concentrated in frequency bins of limited bandwidth and therefore will be more prominent in some reconstructed signals than others. The short time energy function will act as means of enveloping the signal and downsampling while the combination of the autocorrelation function and cepstrum will further reduce noise and indicate an periodicity in the signal. The proposed process is shown in a flow chart in Figure 5.1.

5.1.1 Creation of Simulated Data

The process outlined in Figure 5.1 will be demonstrated in the following sections using a simulated defect signal. This section describes the process for generating such a signal. A defect signal $x(t)$ can be assumed to be the sum of a measurement signal $y(t)$ and Gaussian noise $n(t)$ given by Equation 5.1.

$$x(t) = y(t) + n(t) \tag{5.1}$$

The measurement signal is transient series of short impulses representing each instance at which a rolling element strikes a defect whereas the additive background noise from other vibrations in the system is assumed stationary. Based upon the assumption that each impact generates the same waveform, a more general expression for a rolling element bearing defect signal can be given by Equation 5.2 [89].

$$x(t) = \sum_{i=1}^{M} A_i.s(t - iT - \tau_i) + n(t) \tag{5.2}$$

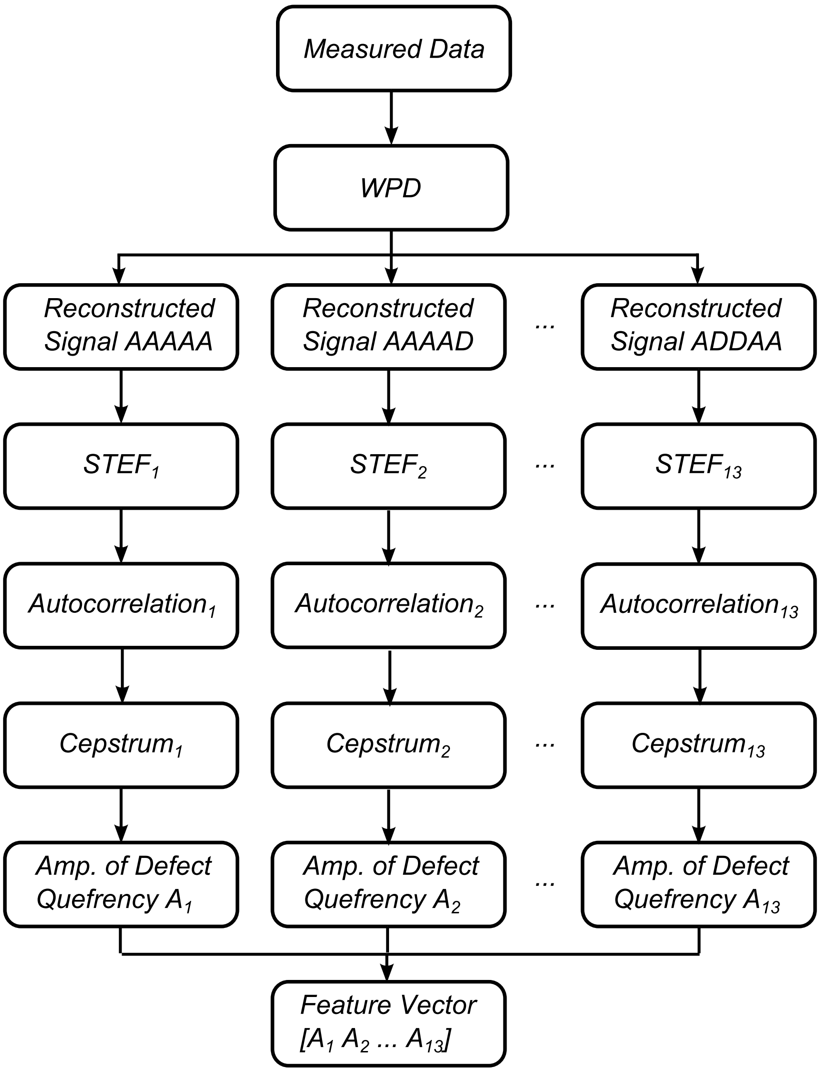

Figure 5.1: Flow Diagram of Signal Processing Procedure

This represents a series of M impulses occurring at an interval T, each subjected to a minor and random fluctuation τ_i due to slip or skid. A_i denotes the amplitude modulator and is periodic with shaft speed. This is either constant or approximately sinusoidal depending on whether the component is rotating or not and therefore experiencing a periodic load cycle. Some studies have modelled the waveform $s(t)$ as a heavily damped sinusoid at the resonant frequency f_r as given by Equation 5.3.

$$s(t) = e^{-Bt} cos(2\pi f_n t) \tag{5.3}$$

The damping coefficient B should be high enough to replicate attenuation in a real system ensuring the oscillations have disappeared prior to the following impulse.

Prior to obtaining the run to failure data of a number of bearings, a range of defect signals were simulated to provide an initial indication of the effectiveness of the diagnostic technique. The idealised case described above can sufficiently describe incipient faults in rolling element bearings [89]. However, in the following work a series of repeated artificial sources have been used to form $s(t)$. A Hsu-Nielsen source, was performed on the inner raceway of the bearing described in Section 4.1.1 and recorded with a Physical Acoustic Nano 30 sensor. The measured response is shown in Figure 5.2. This approached is preferred to using a synthesised waveform since finding the transients of a single frequency becomes a relatively trivial task.

A clean defect signal composed of a series of these sources is shown in Figure 5.3 at 16Hz, roughly typically of a defect frequency found in on the MultiLife™ bearing at rotational speeds slightly greater than 100rpm. Since the inner race is stationary and any defects would experience a constant load the signal has not been amplitude modulated by the shaft frequency.

Finally, Gaussian white noise is superposed upon the signal. Since the aim of this work is to determine the sensitivity of the proposed diagnostic methods to incipient damage, the scaling of the defect signal is adjusted to vary the signal to noise ratio (SNR). Conventionally, the SNR is the ratio of the power of the signal to power of the noise. However, given the transient nature of the defect signal the SNR has been defined as the ratio of the maximum amplitude A of the defect signal to three standard

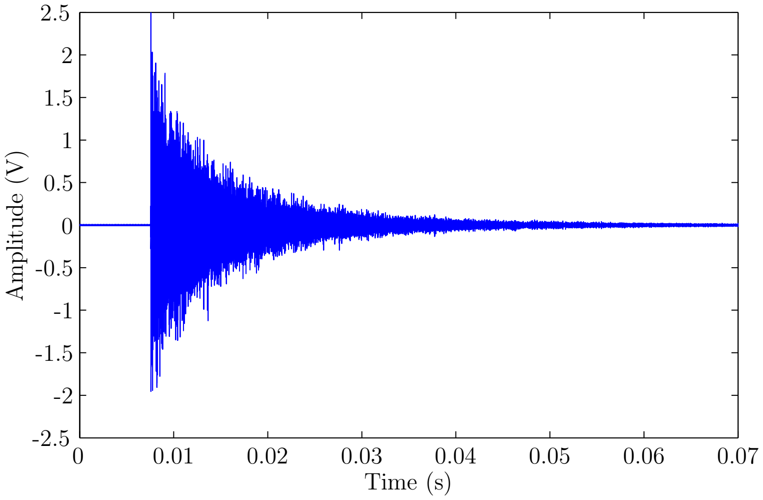

Figure 5.2: Hsu-Nielsen Source

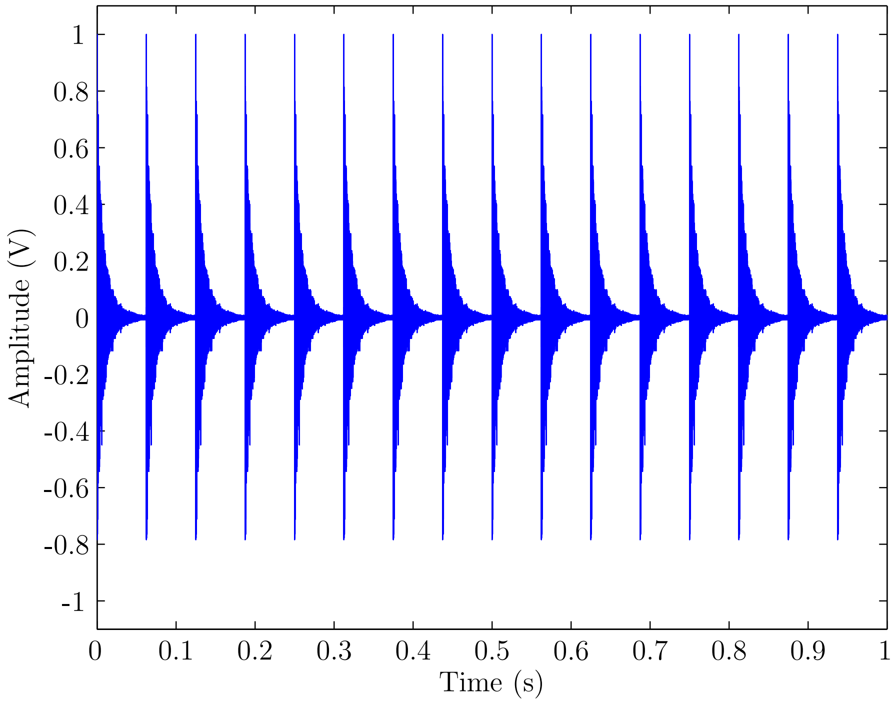

Figure 5.3: Pure Defect Signal

deviations σ of the noise (Equation 5.4). Previous work defined the SNR as the ratio of maximum amplitude of the signal to the maximum amplitude of the noise [85] however it was found that this produced too much variation between signals with the same SNR.

$$SNR = \frac{max(A_{signal})}{3\sigma_{noise}} \tag{5.4}$$

Two examples of the simulated signals with SNRs of 0.4 and 1.5 are shown in Figure 5.4. A random fluctuation up to 10% of the period T will also be introduced to verify the robustness of each diagnostic technique to slip and skid.

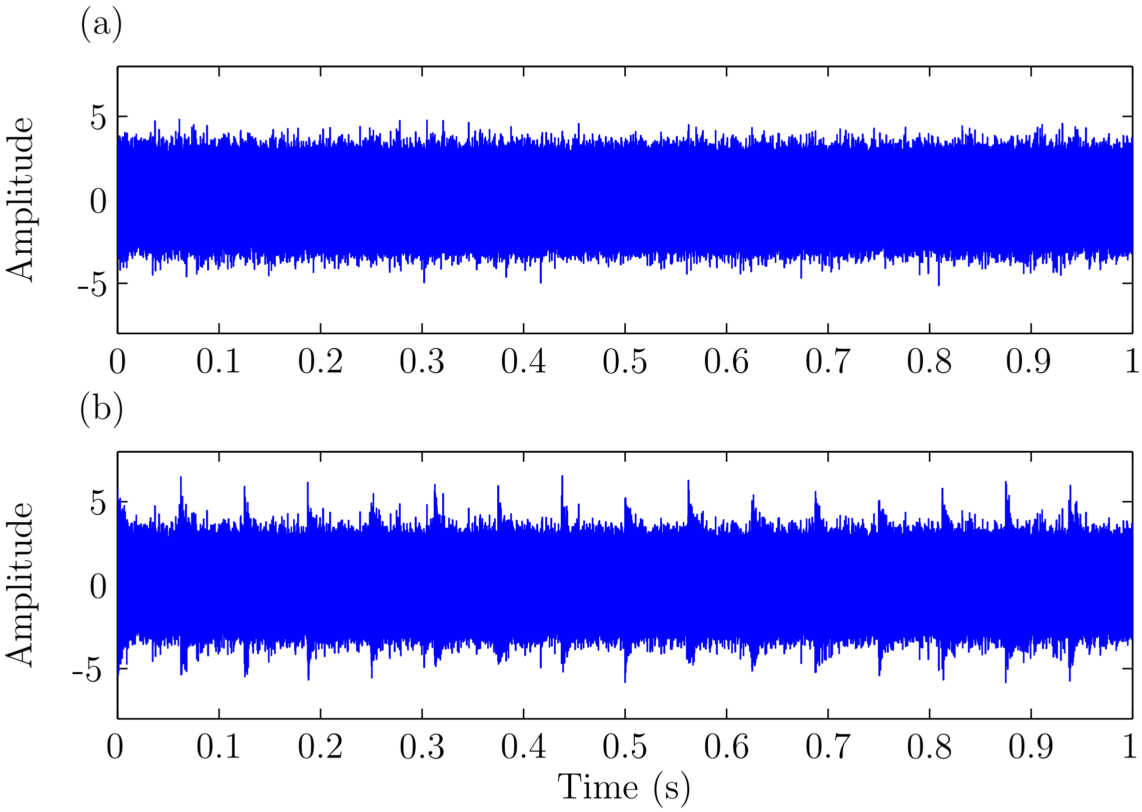

Figure 5.4: Simulated Defect Signals (a) SNR = 0.4, (b) SNR = 1.5

5.1.2 Wavelet Packet Decomposition

The theory behind the WPD is presented in Chapter 3. In this section, the decomposition will be demonstrated using simulated data described in Section 5.1.1. The data is one second in length and is sampled at 2MHz with a defect frequency of 16Hz. The SNR for the example shown is 0.8. In its unprocessed state no regular features can be seen in the signal and thus provides a good example for the subsequent demonstration. The frequency content of this signal is shown in Figure 5.6. The use of modelled white noise gives a uniform distribution across the frequency range whereas most of the additional content due to the modelled defect falls in the range of 75-300kHz. The inset plot in Figure 5.6 has zoomed in on the spectrum to the range relevant to the defect frequencies and the position of the defect frequency is shown by the dashed red line. Since the defect signal is buried within noise in addition to the fact that it manifests itself as an amplitude modulated high frequency signal, the characteristic defect frequency is not visible on the spectrum.

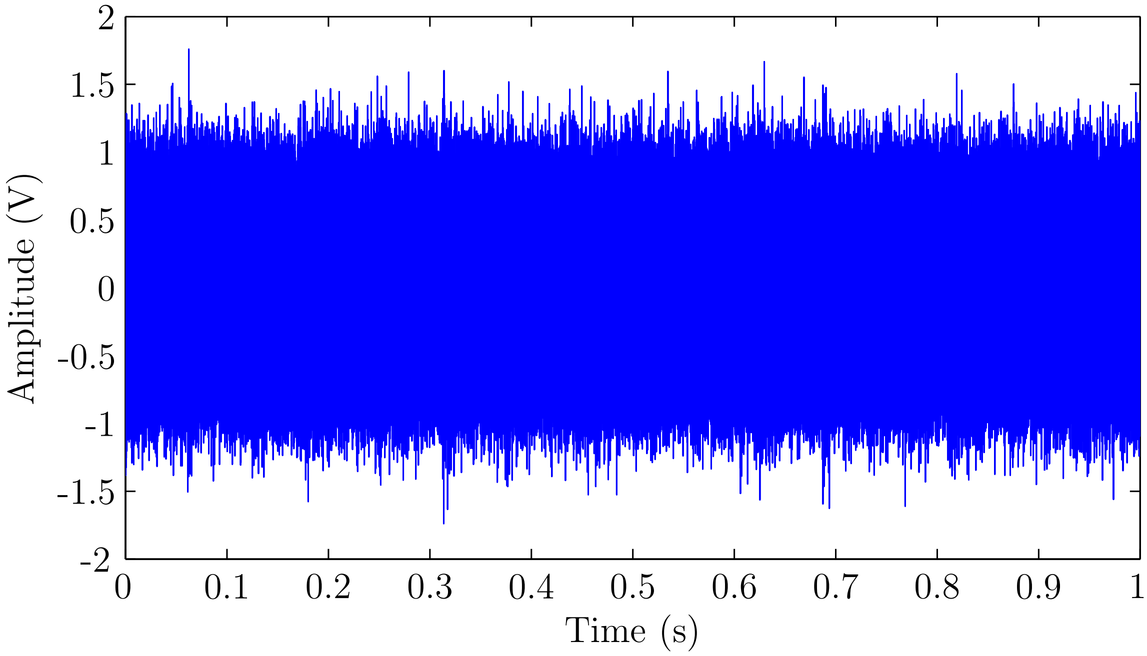

Figure 5.5: Simulated Defect Signal

When carrying out wavelet packet decomposition, two choices must be made; the wavelet used and the level of the decomposition. For all this work, the Debauchies 12 (db12) was used. The number relates to the number of vanishing moments. More vanishing moments tend to give more accurate results at the expense of computational power and time. In past studies, the db12 wavelet has been considered a reasonable compromise between accuracy and resources [67]. A level 5 decomposition gives $2^5 = 32$ sets of coefficients giving a filter bank where each set of coefficients has a bandwidth of approximately 31kHz. The signal shown in Figure 5.5 has been decomposed to 5 levels using the Wavelets Toolbox in Matlab and the signal reconstructed from the coefficients $cAADDD$ is shown in Figure 5.7. A regular series of events due to the

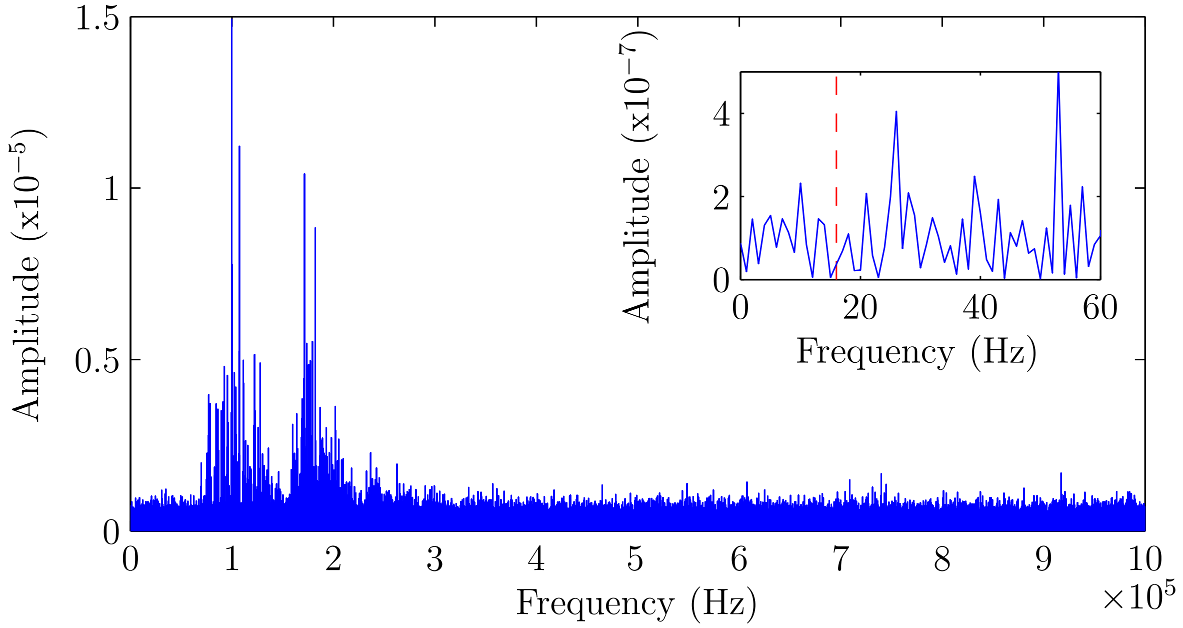

Figure 5.6: Frequency Content of Defected Bearing

simulated roller impacts is now visible at 16Hz.

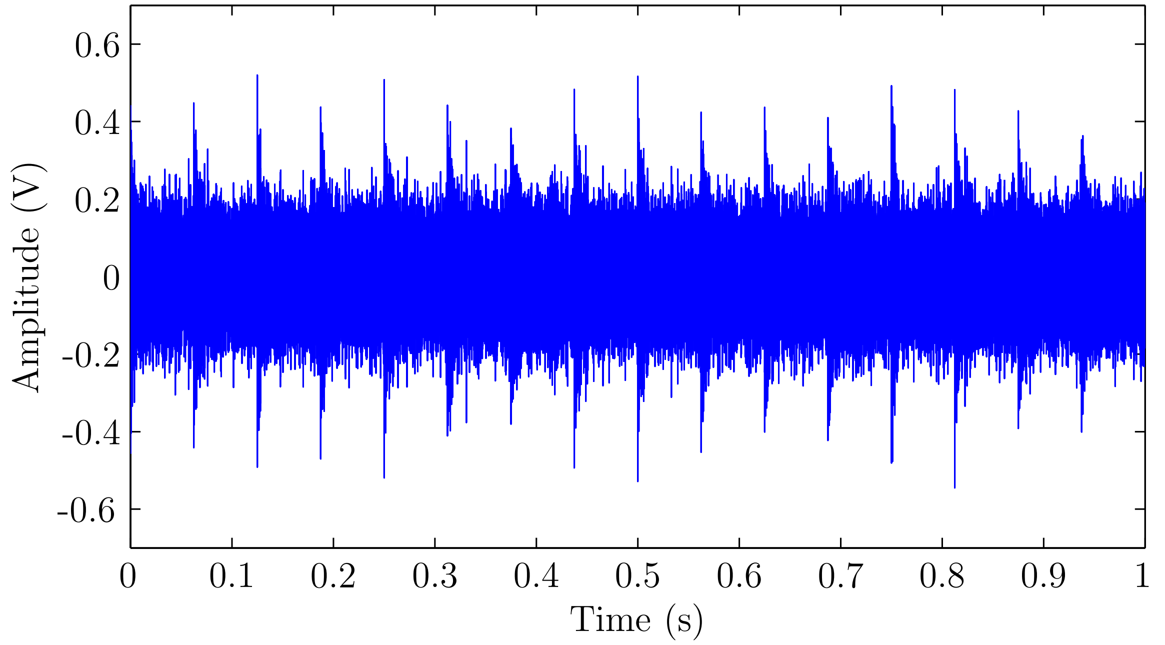

Figure 5.7: Reconstructed signal *ADDAD*

5.1.3 Short Time Energy Function

The purpose of the short time energy function (STEF) is to envelope the signal and act as a means of downsampling thus reducing the data storage requirements which is one of the major disadvantages of acoustic emission monitoring. The signal is divided into windows of equal length and the RMS is calculated for each window. This approach has the additional benefits of reducing the effect of spurious spikes of noise which tend to be of a shorter duration than that of an event due to roller impingement. A suitable window length must be chosen to achieve satisfactory demodulation and data reduction whilst ensuring that genuine events due to bearing damage are not smeared out in the averaging process. The window length was optimised by repeating the steps shown in Figure 5.1 using window lengths ranging from 10 to 10000 at intervals of 10 and summing the amplitude of the quefrencies in the cepstrum relating to the defect frequencies. The results are shown in Figure 5.8 with the red line indicating the trend following a median filter. It can be seen that the maximum occurs when the window length is about 750 samples. The STEF of the signal shown in Figure 5.7 is shown in Figure 5.9.

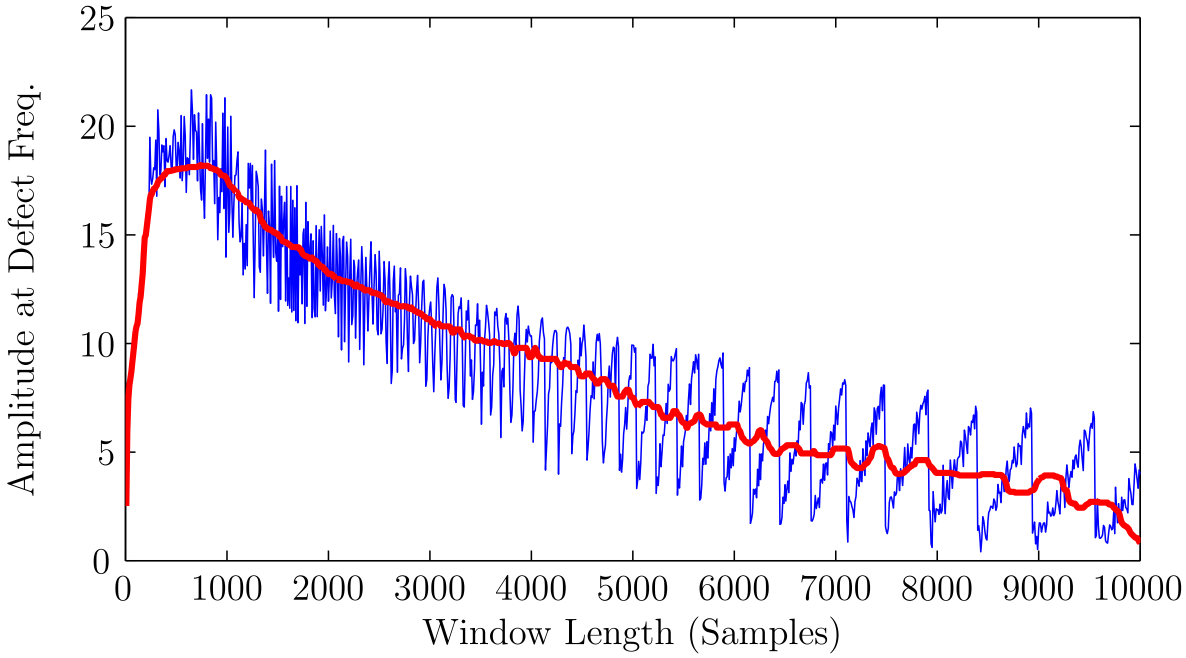

Figure 5.8: Window Length Optimisation for Short Time Energy Function

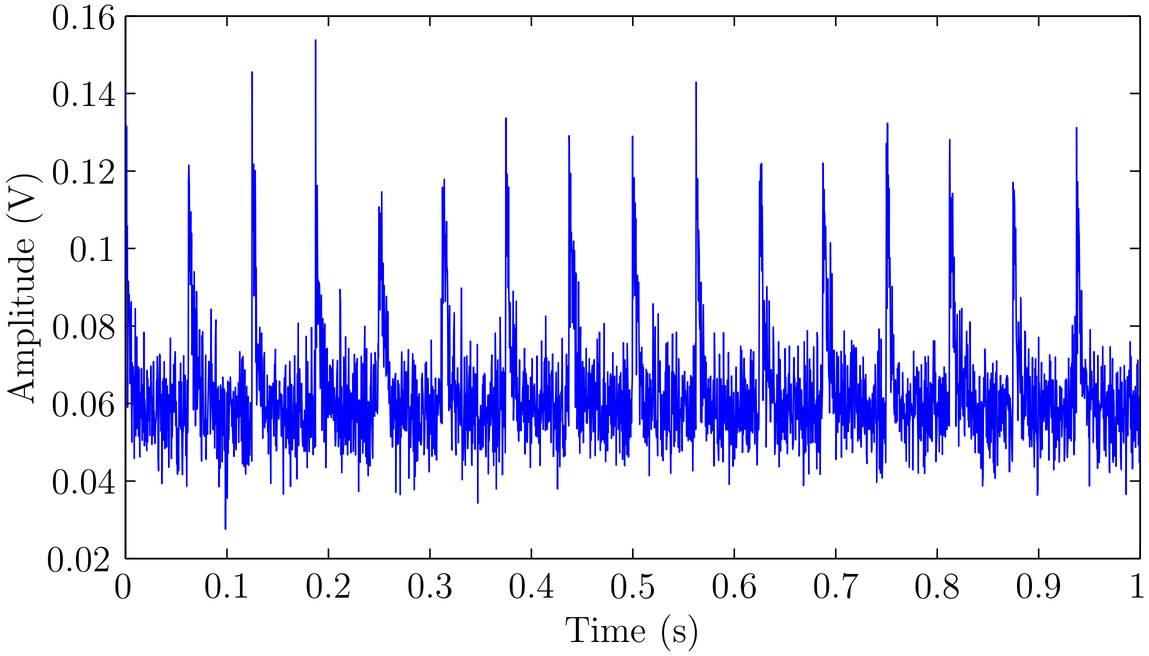

Figure 5.9: Short Time Energy Function of Reconstructed Signal

5.1.4 Autocorrelation

Cross-correlation is a mathematical function which gives a measure of similarity between two signals. Equation 5.5 describes cross correlation function between two continuous signals; $f(t)$ and $g(t)$

$$CC(\tau) = \int_{-\inf}^{\inf} f(t)g(t+\tau)dt \tag{5.5}$$

where τ is the time lag. Autocorrelation extends this idea and is defined as the cross correlation of a signal with itself and can be used to find periodicity within a signal. Equation 5.5 is therefore modified and the autocorrelation is defined by Equation 5.6.

$$AC(\tau) = \int_{-\inf}^{\inf} f(t)f(t+\tau)dt \tag{5.6}$$

The autocorrelation function will have a maximum at zero lag ($\tau = 0$) and subsequent peaks indicate a periodicity of 1/lag. As mentioned in Section 5.1.1, a defect signal can be assumed to be the sum of a regular pulse due to roller infringement on the defect and a random noise. If then, for example, the lag τ is equal to the time between the pulses, the correlation due to the defect signal will be large whereas the interval between the peaks should show no significant correlation thus increases the relative amplitude of the defect signal compared to the noise. Therefore the autocorrelation is used as a means of increasing the prominence of the defect frequencies in the subsequent stages. The autocorrelation of the STEF signal shown in Figure 5.9 is shown in Figure 5.10.

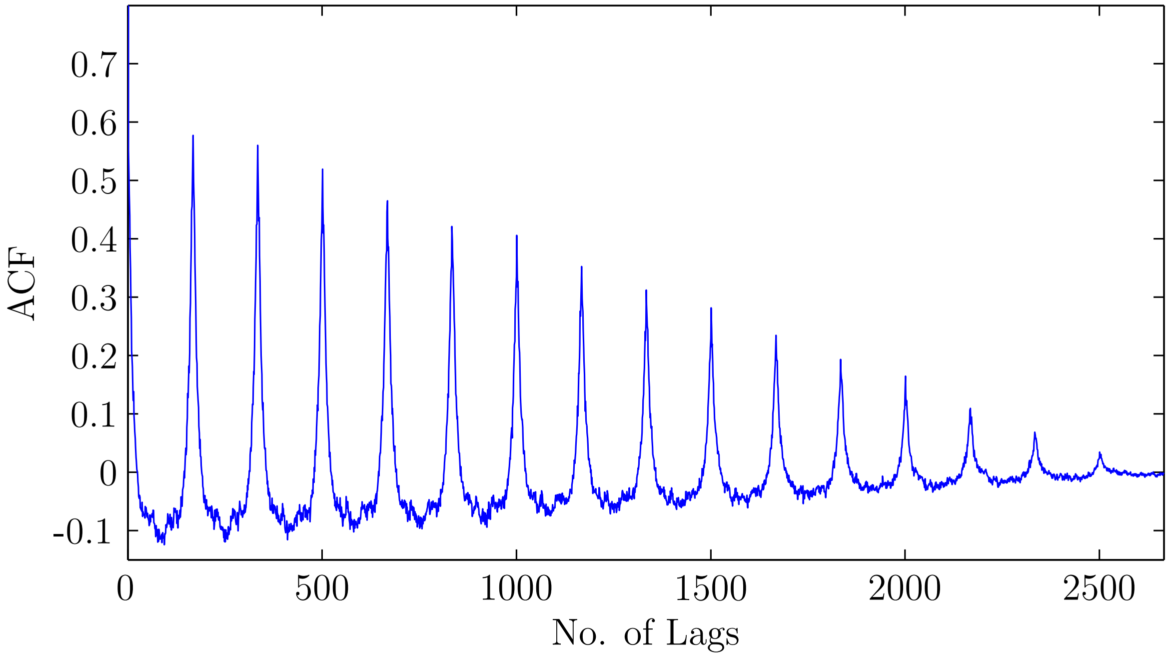

Figure 5.10: Autocorrelation of Enveloped Defect Signal

5.1.5 Cepstrum

The cepstrum, introduced in Section 3.2.2, gives a measure of periodicity in the signal. Figure 5.11 shows the cepstrum applied to the autocorrelation signal shown in Figure 5.10. The dominant peak at 0.062s relates to the defect frequency of 16Hz.

5.1.6 Feature Vector

The idea of generating a feature vector is to represent a large data set with a high level of redundancy by a much smaller dimension whilst retaining the relevant characteristics. In this work, three vectors, one relating to each bearing component, will

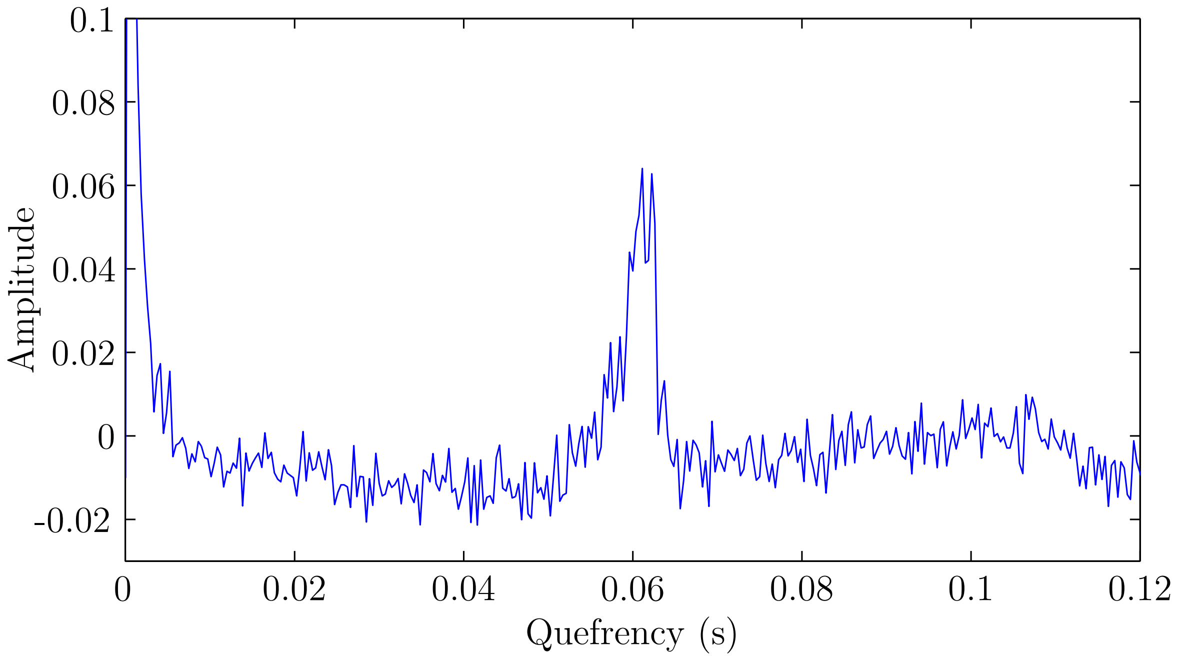

Figure 5.11: Cepstrum of Autocorrelation Signal

be retained for every sample. Each feature vector is formed by taking the amplitude of the quefrency corresponding to the defect frequency of either the inner race, outer race or rolling element for each of the reconstructed signals of the WPD. A fifth level decomposition, would therefore result in a vector of 32 dimensions. However, since the frequency content of the real signals is almost negligible above 400kHz, only the first thirteen reconstructed sets of coefficients will be used resulting in three thirteen dimension vectors for each sample.

5.2 Outlier Analysis

Outlier analysis is a multivariate statistical method that highlights previously unseen data which exhibits a degree of novelty. If a datum is deemed an outlier, it is assumed to be generated from a mechanism other than that which formed the training data. In the field of damage detection, data from the undamaged structure is used as the training data and outliers are therefore assumed to be indicative of damage. Section 5.2.1 describes the theory behind outlier analysis whilst Section 5.2.2 outlines a method for determining a suitable threshold above which a signal can be considered an outlier. Finally, the whole concept is demonstrated using simulated data in Section 5.2.3.

5.2.1 Outlier Identification

Rytter [90] established a natural hierarchy for problems of damage detection and identification as follows.

Level 1 - Damage Detection: Provides a qualitative indication that the component in question may be damaged.

Level 2 - Damage Location: Once the presence of damage has been confirmed the next level is to provide an estimation of the probable position of the damage.

Level 3 - Damage Classification: Provides an indication of the type of damage that has occurred.

Level 4 - Damage Assessment: Provides an estimation of the severity of the damage.

Level 5 - Damage Prognosis: Provides information regarding the safety of the component such as the remaining useful life that can be expected.

Identifying the presence of a fault is the most fundamental aspect of damage detection and forms the lowest level of the hierarchy. This must be fulfilled before other stages of the hierarchy are attempted. Outlier analysis is a multivariate statistical technique which can be used as a novelty detector thereby fulfilling this objective. The technique was first applied to the field of damage detection by Worden et al [91]. The technique does not necessarily provide information about the severity or location of the damage; concepts which occupy higher levels on Rytters hierarchy, however in some instances as will be shown later, a measure of distance from the normal state may be used for such purpose.

The technique of outlier analysis can be applied to univariate data. An observation will simply be deemed an outlier if it protrudes from either end of the data set. If a measure of discordancy of a given observation exceeds a pre-determined threshold, for example three standard deviations away from the mean, the observation will be classified as an outlier. The z-score z_ξ is one measure of discordancy which will later be extended to the multivariate equivalent and is given by Equation 5.7

$$z_\xi = \frac{|x_\xi - \bar{x}|}{\sigma} \tag{5.7}$$

where x_ξ is the data of the potential outlier and $\bar{x}$ and σ are the sample mean and standard deviation respectively.

The problem of novelty detection becomes more involved when applied to multivariate data since an outlier may not protrude from either end of any individual dimension yet can still exhibit discordancy. This is demonstrated graphically in Figure 5.12 with an artificially generated two variable example.

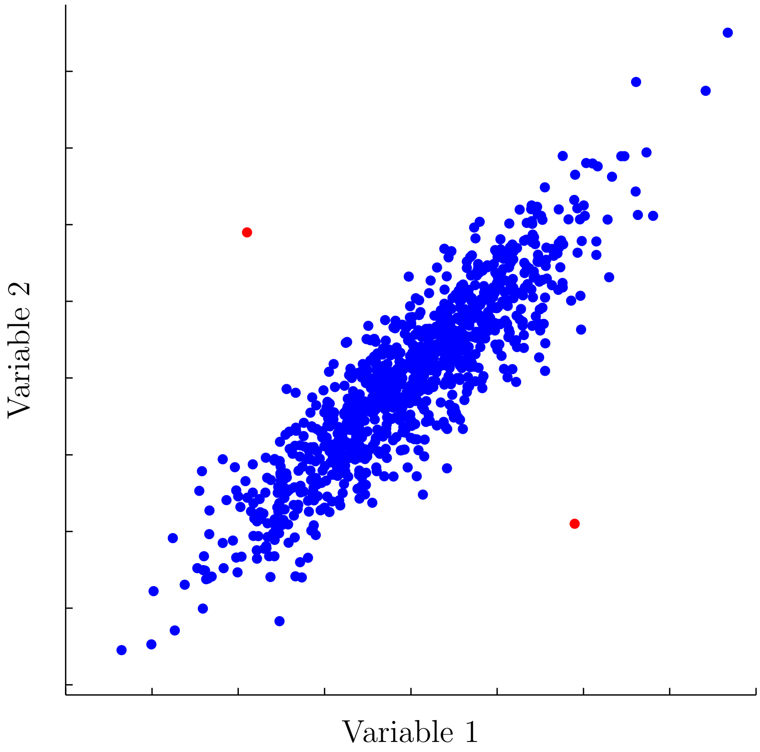

Figure 5.12: Two Variable Outlier Example

The data points in red are clearly outliers despite falling within the normal range for each individual variable. The Mahalanobis squared distance D_ξ is a multivariate extension of Equation 5.7 and is given by Equation 5.8

$$D_\xi = (\{x_\xi\} - \{\bar{x}\})^T [S]^{-1} (\{x_\xi\} - \{\bar{x}\}) \tag{5.8}$$

where $\{x_\xi\}$ is the feature vector of the potential outlier for which the discordancy will be calculated and $\{\bar{x}\}$ and $[S]$ are the sample mean vector and sample covariance matrix respectively. The Mahalanobis squared distance is often preferred to the Euclidean distance in multivariate analysis since it takes into account correlations of the data and is scale invariant. In other words, it compensates for differences in variance between dimensions. The suitability of using the Mahalanobis squared distance as a degradation indicator will be investigated.

5.2.2 Monte Carlo Threshold

As with the univariate equivalent, a multivariate observation is deemed an outlier if the Mahalanobis squared distance exceeds a given threshold. The Monte Carlo method is one method of statistically determining a suitable threshold. In general, Monte Carlo methods are a class of computational algorithms which rely on repeated random sampling in order to achieve numerical results. In this case, the principle is based on finding the extreme values which could be generated from the same mechanisms as the sample data. The calculation of the threshold is explained in detail in [91] and is summarised as follows:

Step 1 A matrix of p dimensions by n observations is generated whereby each element is randomly generated with a zero mean and unit standard deviation.

Step 2 The Mahalanobis squared distances are calculated for each observation where $\bar{x}$ and $[S]$ are calculated from the same matrix. The largest distance is stored.

Step 3 Steps 1 and 2 are repeated a large number of times. The stored distances are ordered by magnitude. The value above which 1 or 5% of the distances occur are commonly used thresholds. The 1% threshold will be used in the remainder of this work to reduce the likelihood of false alarms.

The mean vector and the covariance matrix may be calculated with or without the potential outliers. The threshold is said to be inclusive if calculated with the outliers and exclusive otherwise. The steps above calculate the inclusive threshold since each of the potential outliers have contributed to the calculation of the mean and the covariance matrix. Equation 5.9 allows the inclusive threshold T_{inc} to be converted to the exclusive threshold T_{exc} where n is the number of observations [91].

$$T_{exc} = \frac{((n-1)(n+1)^2)T_{inc}}{n(n^2-(n+1)T_{inc})} \tag{5.9}$$

The 1% exclusive threshold for an 13 dimension, 1000 observation case was found to be 48 after 10000 trials. The steps outlined in this section have been validated in the following section through the use of simulated defect signals before being applied to real data recorded from the test rig presented in Section 4.1 in Chapter 7.

5.2.3 Outlier Analysis with Simulated Data

The procedure outlined in Sections 5.1 and 5.2 was applied to a number of simulated signals with SNRs ranging from 0 to 0.5 in steps of 0.05 in order to determine the point at which defects may be detected. One thousand signals were simulated to represent the undamaged state and were used to form the mean vector $\bar{x}$ and the sample covariance matrix $[S]$ used for the outlier analysis. A further 100 signals for each damage condition were generated. The results of are shown in Figure 5.13. The 1% Monte Carlo threshold of 48 is marked in red.

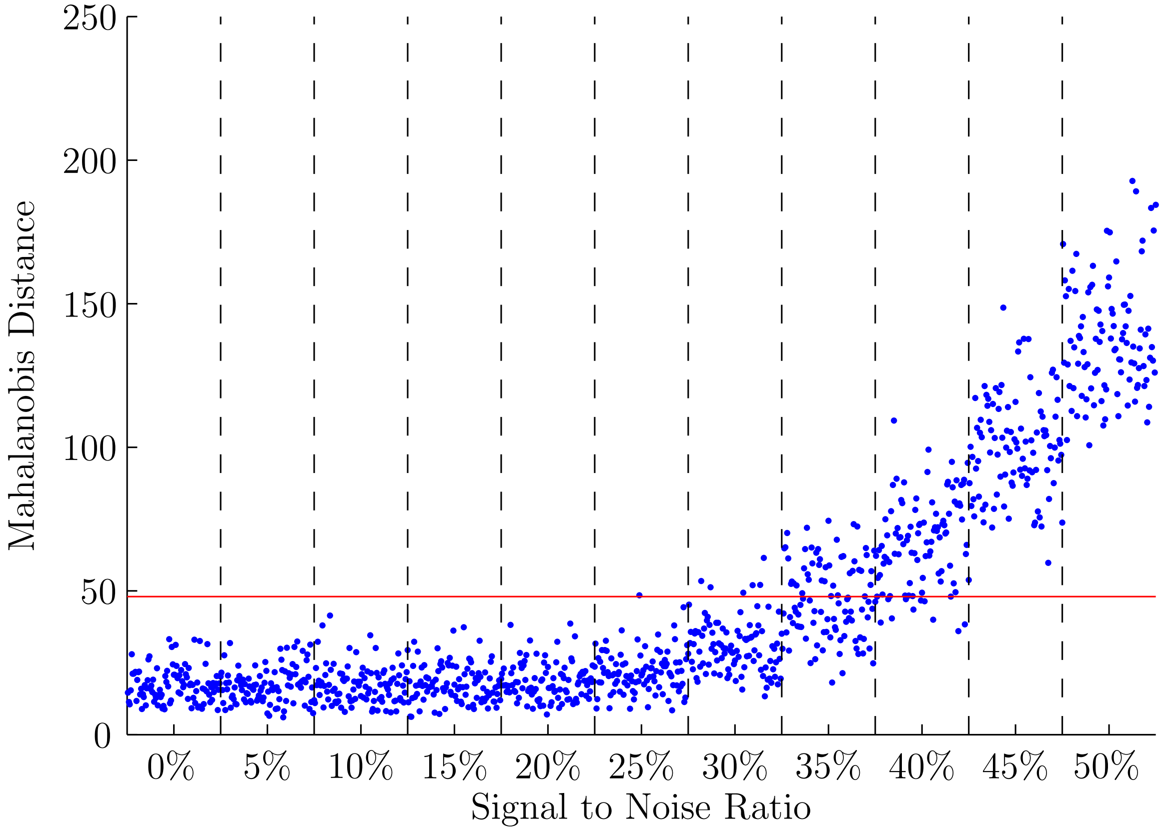

Figure 5.13: Outlier analysis results for range of simulated signals

Figure 5.13 shows outlier analysis can successfully differentiate between undamaged and damaged signals in cases where the signal to noise ratio is greater than approximately 0.4 or 40%. An example signal where the SNR is 0.4 is shown in Figure 5.4(a) at which point the transient signal representing the roller impingements are not visible to the eye. Figure 5.13 also demonstrates the potential to use discordancy, the Mahalanobis distance, as an indicator of bearing health once a defect has been detected since the increase in SNR shown in the plot is representative of a growing defect.

For sake of completeness, an outlier analysis has also been carried out by forming feature vectors comprising four of the traditional metrics presented in Section 3.2.1: RMS, peak, kurtosis and crest factor. The results are shown in Figure 5.14. This method demonstrates a clear difference in discordancy between the undamaged and all damaged signals suggesting this would be more successful than the method presented in Figure 5.13 however such a conclusion would be misleading. The simulated signals have been created by superimposing a defect signal of increasing amplitude on artificially generated Gaussian white noise of constant standard deviation. Therefore both the variance of the measures within each SNR group and the difference between the undamaged signals and those with a SNR of 50% are very small. For example, the average RMS when the SNR is 50% is just 0.45% greater than the non-damaged state while the increase in kurtosis is even smaller at 0.073%. In reality, the variance between signals in the same damage state would be much greater due to non-steady operating conditions, background noise and material imperfections. It is likely that there would be a overlap between signals of different SNRs in terms of RMS, peak and kurtosis values. Chapter 7 compares outlier analysis with the traditional measures and the proposed method presented in this chapter.

5.3 Conclusion

This chapter has presented a method of detecting bearing damage using acoustic emissions. It has been shown that with suitable processing techniques, it is possible to increase the signal to noise ratio of a signal from a defected bearing. In this instance this has been achieved through the use of wavelet packet decomposition acting as a filter bank, the short time energy function and autocorrelation while any periodicity can be measured using the cepstrum and has been used to form a suitable feature

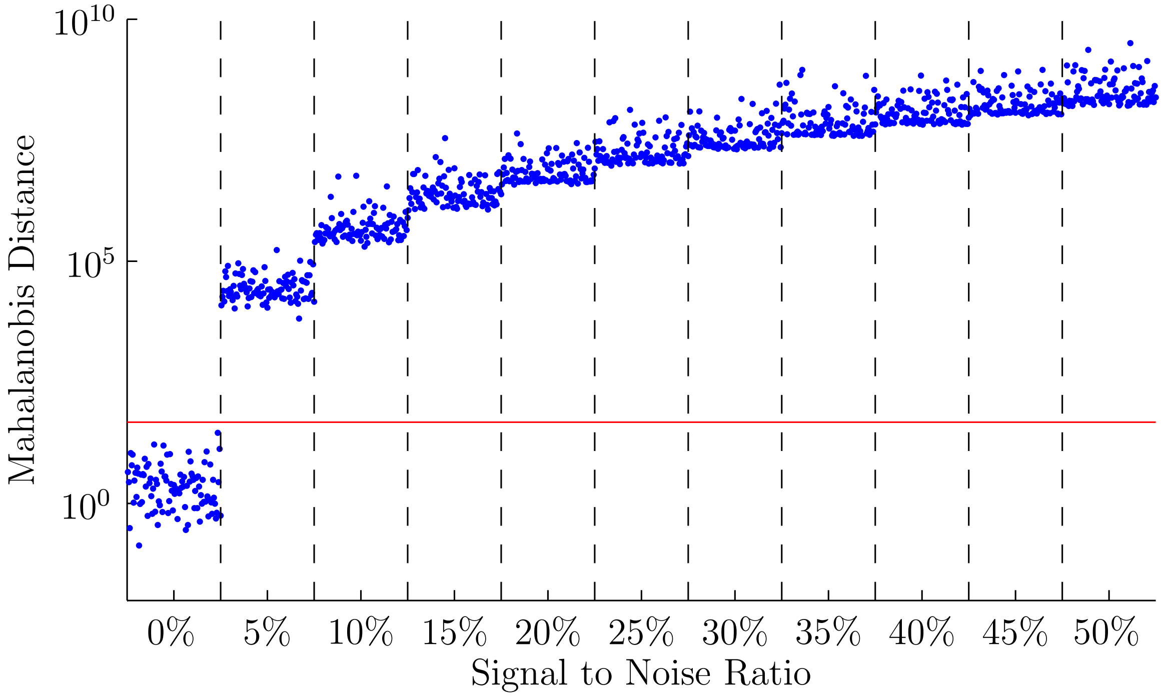

Figure 5.14: Outlier Analysis Traditional Metrics

vector.

Finally, outlier analysis has been presented with the aim of providing an indication of whether a bearing is damaged or not given the presence of undamaged training data. Furthermore the potential of the Mahalanobis distance as a measure of damage severity has been demonstrated with the use of a simulated defect signals.

Chapter 6

Source Location

This chapter investigates and compares three methods for estimating the position of a defect on a bearing using the arrival times of discrete events at AE transducers. An analytical approach is compared to two data driven techniques, namely Delta T mapping and neural networks. Errors are compared for artificial sources created with a pencil lead break and for a real, yet seeded, defect at a known position.

6.1 Introduction

One of the most commonly cited benefits of AE testing over other techniques lies in its ability to locate the sources of emissions of an entire structure during online monitoring. This chapter will introduce and compare a number of methods available for this purpose and make recommendations based on the findings. Other methods for source location using AE have been presented in the literature but are not relevant to this work. For example single sensor modal analysis is only suitable when monitoring plate like structures with low levels of background noise [92].

Techniques such as ultrasound which scan the structure's surface generally require complete access to the entire structure which, in the case of most machinery, will incur significant periods of downtime and therefore loss of production and reduced profits. Other methods, such as dye penetrant inspection can only locate surface damage and are subject to human error.

Since proximity to the source is not required, AE has the ability to locate damage on an entire structure with a sparse array of transducers. The attenuation of the wave when propagating through the structure should be taken into account when deciding where sensors are to be positioned. Since the transducers can be left in position during a complete life test, AE is particularly suited to online monitoring in hostile environments. Furthermore, source location is not restricted to material damage and techniques have been developed to locate continuous signals generated by mechanisms such as friction or leaks [31]. The majority of studies looking at source location have focused on low noise environments or simple geometries on applications such as thin plates, wires and pressure vessels.

Source location techniques have been developed for the two modes of AE introduced in Section 2.4; continuous and burst emission. The localisation of continuous signals is generally based on the cross-correlation function to find the delay between signals [93, 94] whereas a number of techniques have been developed for the source location of discrete events; the choice of which depends upon the application and accuracy required. Since a defected bearing is characterised by such events, this chapter will examine some commonly used techniques for the localisation of such signals. Many of these use arrival times of events at a given sensor based upon the time of the first threshold crossing or the arrival time of a given mode however in the cases where the transient signals do not exceed the noise the cross correlation of the signals may be used to find the difference in arrival times.

The majority of bearing maintenance decisions focus upon deciding at what point the bearing should be replaced so bearing life is maximised without risking catastrophic failure. Localising the point of damage has, until now, received little attention since bearings generally require a complete replacement once damage has been detected. However, further decisions must be made when bearings fitted with the Multilife™ actuator are considered. A suitable indexing strategy designed to maximise bearing life is dependent on how often and by how much the inner raceway is rotated. An accurate technique to locate the damage is required to optimise such a strategy. An intelligent indexing strategy will require knowledge of defect positions to ensure large or significant defects are rotated away from the loaded zone in order to prevent catastrophic failure. It will also allow the user to judge whether the defects progress once they are rotated away from the loaded region. Tracking damage in this manner will validate whether the assumptions made in the model described in Section 1.4 hold true and whether bearing life can indeed be extended by a factor of five. Finally, the positions of defects may provide an indication of the failure modes and causes. For example, a concentration of located emissions at one edge of the raceway may indicate side loading or misalignment, however given the number and position of sensors used in this work this is not possible and only the circumferential position will be estimated. Both analytical and data driven methods will be investigated and recommendations will be made based upon accuracy and computational requirements. This chapter will present a number of techniques currently employed for source location of other components and compare the performance of two techniques namely Delta T mapping and neural networks with an analytical solution where assumptions regarding the wave propagation are required.

6.2 Review of Source Location on Bearings

Comparatively few studies have been published regarding AE source location in rolling element bearings. Many bearings are small and relatively easy to replace when damage occurs. This means the location of defects is not important when making maintenance decisions. However in large bearings, such as those found in wind turbines or power plants, knowledge of where the failure occurs helps identify the root cause and may inform remedial operation.

The first investigation into source location of acoustic emissions from rolling element bearings was performed in the previously cited study by Rogers [68] whilst monitoring large slew ring bearings on platform cranes. Two sensors were bonded diametrically opposite each other on the fixed inside race of a bearing. The linear location of any bursts significantly above the background noise was calculated. This enabled the user to determine whether the bearing was damaged if a concentration of bursts developed in a given region. However, in this case the location is ambiguous since the emission may have originated on either side of the bearing and the use of just two sensors would not be able to differentiate between them. Since the bearing was only monitored in an undamaged state, the positions of the recorded events were evenly distributed; concentrations of hits would be indicative of damage. Furthermore, the lack of a damaged bearing meant there was no way of correlating the true position of the source with the estimations and therefore no indication of the error was provided.

Yoshioka and Fujiwara [95] developed a novel source location system using a single

AE transducer and magnetic detector to determine the position of a ball at the time of an emission. Such a method, whilst providing an accurate location of the emission, would also locate emissions at the position of each rolling element. The exact defect position would, therefore, not be clear until the raceway is visually examined. No indication of the errors introduced by this system was provided.

More recently, Elforjani and Mba [86] carried out source location whilst investigating the natural degradation of slow speed thrust bearings. The circumferential location was established in a similar manner to Rogers [68] however 4 sensors were employed allowing an exact circumferential estimation to be calculated. The estimated positions of a series of H-N sources were reported to be within 4% of the exact location. Eftekharnejad et al [96] extended this work to source location of high speed bearings and found similar errors. Both studies have made assumptions on wave propagation and wave speed.

6.3 Methods

The remainder of this chapter is concerned with validating which method of source location is most suitable in a large bearing application. The time of arrival (TOA) and Delta T methods are investigated. Furthermore, a method using neural networks, is explored.

Artificial sources have been used to train and validate each method. In this study the artificial sources were generated by a Hsu-Nielsen source [38, 39]; a standardised lead break of a retractable pencil is described with dimensions in Section 2.3.5.

The arrival time of each AE burst was taken as the time of first threshold crossing (FTC) as indicated in Figure 6.1. In most AE monitoring the threshold is defined by the user and usually dependent upon the level of background noise. In this work the threshold was set at 56dB.

As previously stated, the efficient operation of the MultiLife™ actuator is dependent upon being able to accurately estimate the position of damage and track this position during operation. Clearly only the circumferential position is required since the MultiLife™ actuators act in this direction. Furthermore, the positions of the sensors with respect to the raceway (Figure 4.4) restricts itself to estimating the circumferential position since there is no way of differentiating between sources emitted from the front and back of the bearing. Therefore, the following methods will only be concerned with the circumferential position of the damage.

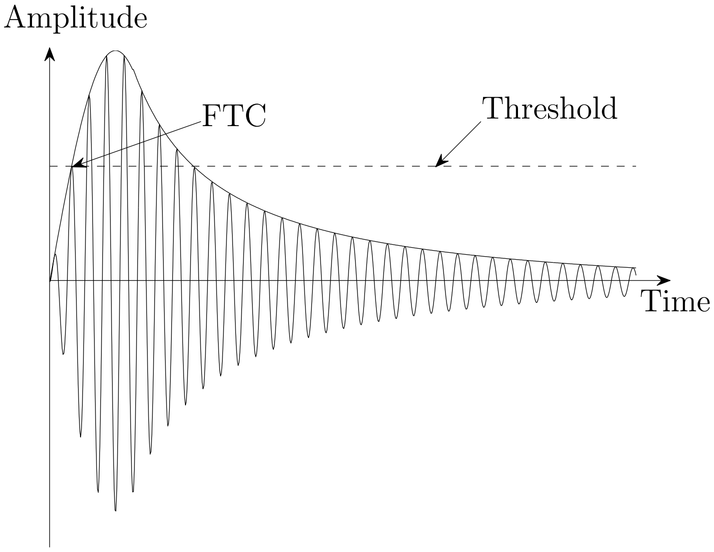

Figure 6.1: First threshold crossing of a simplified AE burst

6.3.1 Zone Location

For those cases where only coarse location is required, zone location may be employed. The arrival times of a given event at each sensor will depend upon its the distance from the source. The first arrival time will be experienced by the closest sensor. Therefore the source position can be reduced to an area around that sensor. The lightly shaded region in Figure 6.2 shows the position of a source if an event first arrives at sensor 1 (S1). The size of the area can be reduced by taking into account the order of subsequent arrival times, for example the darker region in Figure 6.2 if the event subsequently arrives at sensor 2 (S2). This darker region could be further subdivided again by taking into account the arrival times at sensors 3 and 7 (S3 & S7).

A similar principle may be employed for continuous signals. Rather than arrival times, the RMS at each sensor can be examined. The source can therefore be positioned within the zone around the sensor with the highest RMS assuming the sensitivity of the sensors and the coupling are similar.

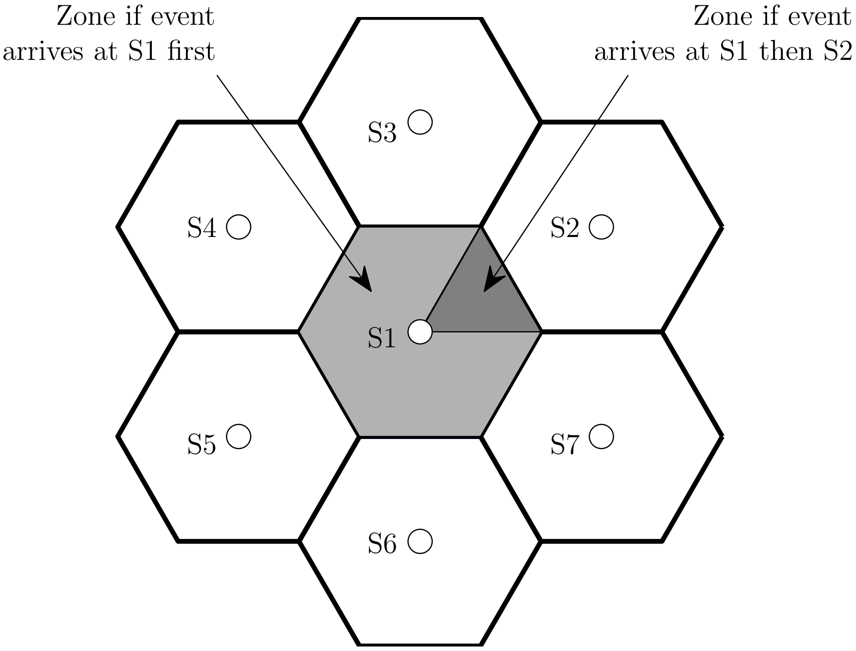

Figure 6.2: Zone Location

This technique can only estimate a zone within which the damage should be rather than provide an exact position. The accuracy depends upon the spacing of the sensors; the greater the distance between sensors, the larger the possible zones. Guard sensors may be also used to prevent the source location of events occurring outside the sensor array whereby those events which arrive at a guard sensor first are discarded.

If this technique were to be employed on the MultiLife™ test rig as described in Section 4.1 the maximum error that would be expected 271mm if using the arrival at the first sensor or 136mm if using the arrival times of the first two sensors given that the circumference of the rolling surface of the inner raceway is 814mm and three sensors are equally spaced around the raceway. Although this error is large, it can be used as a baseline by which other methods of source location may be compared.

6.3.2 Time of Arrival methods

Due to the limitations of zone location, time of arrival (TOA) methods are commonly used and employed by most commercial systems. TOA methods are analytical by nature and therefore require the wave propagation speed V and position of sensors to estimate the source position of a given event. The suitability of TOA methods depend upon the application; a number of applications are presented in the following sections.

Linear Location

The linear location of an AE event between two sensors is restricted to applications where one dimension is much greater than the other two; long struts, pipes and wires being typical examples. In general, the source location d between two sensors separated by a distance D is given by Equation 6.1.

$$d = \frac{1}{2}(D - \Delta t V) \tag{6.1}$$

where d is measured from the first hit sensor and Δt is the difference in arrival times between the sensors. The localisation of events is only meaningful when the source occurs between the sensors otherwise the difference in arrival times will be constant and the estimated source position will be located at either sensor.

2D Source Location

The principle of linear location can be extended to the location of sources in two dimensions assuming the speed of propagation is the same in all directions. If two sensors are bonded to an infinite plate, a hyperbola on which a source can be positioned, shown in Figure 6.3, can be described by Equation 6.2.

$$R = \frac{1}{2}\frac{D^2 - \Delta t^2 V^2}{\Delta t V + D cos\theta} \tag{6.2}$$

Since the use of two sensors is insufficient to pinpoint the source a third sensor must be used. This allows two hyperbolae to be calculated whereby the intersection determines the source location. Equations 6.3 and 6.4 can be solved simultaneously to find the intersection. A third hyperbole can also be calculated allowing for greater accuracy in the presence of noise. this principle is shown diagrammatically in Figure 6.4.

$$R = \frac{1}{2}\frac{D_1^2 - \Delta t_1^2 V^2}{\Delta t_1 V + D_1 cos(\theta - \theta_1)} \tag{6.3}$$

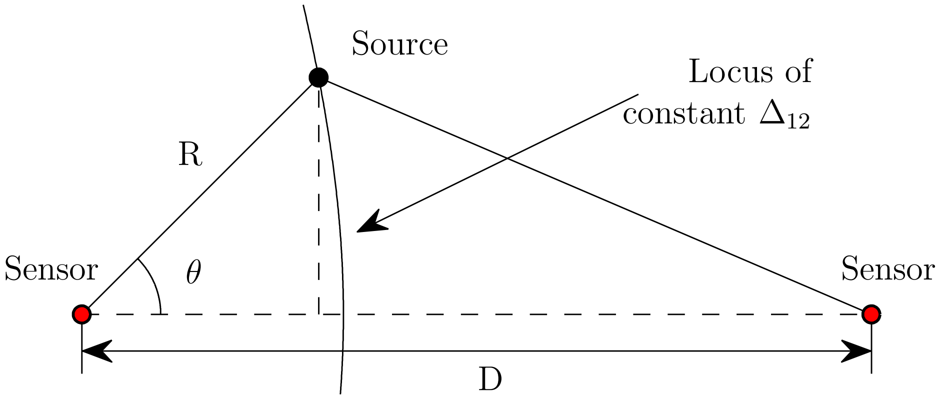

Figure 6.3: 2D Source Location with Two Sensors

$$R = \frac{1}{2}\frac{D_2^2 - \Delta t_2^2 V^2}{\Delta t_2 V + D_2 cos(\theta - \theta_2)} \tag{6.4}$$

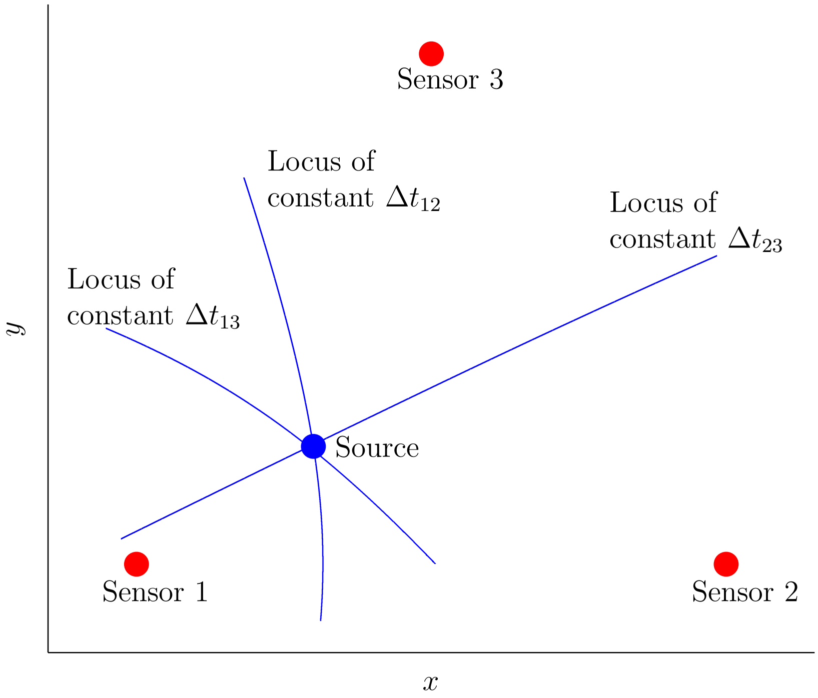

Figure 6.4: 2D Source Location by Triangulation

One of the first investigations to provide a solution for such a problem was undertaken by Tobias [97]. A general method providing an exact solution for a two dimensional problem is presented given the arbitrary placement of three sensors. It is shown however, that such a method can produce ambiguous results if the source is in the proximity of one of the sensors, in other words, two possible locations are calculated. The position of the sensors is shown to be important in the likelihood of producing an ambiguous result although the use of fourth sensor can overcome such a problem. Asty [98] extended the method outlined in [97] to develop a mathematical method to locate sources on a spherical surface and demonstrated how this method may be applied equally to planar surfaces if the radius of curvature is increased to infinity. In a similar manner, Barat et al [99] defined the method required for localization on a cylindrical surface based upon the principle of geodesics: the shortest route between two points. Such techniques provide a reasonable estimation of source position when simple geometries and homogeneous materials are used; however significant errors are introduced as the complexity of the structure increases. These principles can be extended to a third dimension by considering the intersection of hyperbolic surfaces however this introduces a number of problems [31].

Calculating the linear position of an AE hit between two sensors given a difference in arrival times Δt is a trivial matter (Equation 6.1). However the process becomes a little more involved when considering the location of a source on a bearing. Two potential positions can be calculated for any given difference in arrival times since an event will propagate clockwise and counter-clockwise around the raceway. Therefore potential positions may be calculated for double the number of sensor pairs. Thus three sensors, as in this case, will produce six potentials positions. A plot of all the positions will result in a cluster of three positions whereas the other will be spaced out.

Figure 6.5 shows this graphically. A hundred sources were created around the raceway using the Hsu-Nielsen pencil lead break as described in Section 2.3.5 and the time differences based on the FTC were calculated. The three sensors are were positioned at 0, 0.27 and 0.54m. The x-axis gives the actual position of these sources. For each position on the x-axis, the six potential positions have been calculated using Equation 6.1 and are each represented by a different colour. The concentration of estimated positions, three for each source, along the $x = y$ line shows the true estimations whereas the three other positions which deviate significantly from this line represented sources which have been incorrectly on the other side of the bearing. The correct three estimations must be identified and reduced to a single position.

k-means clustering is an algorithm which allows vectors of n dimensions to be quantised into k clusters. Further explanation of the algorithm is given in Appendix A.

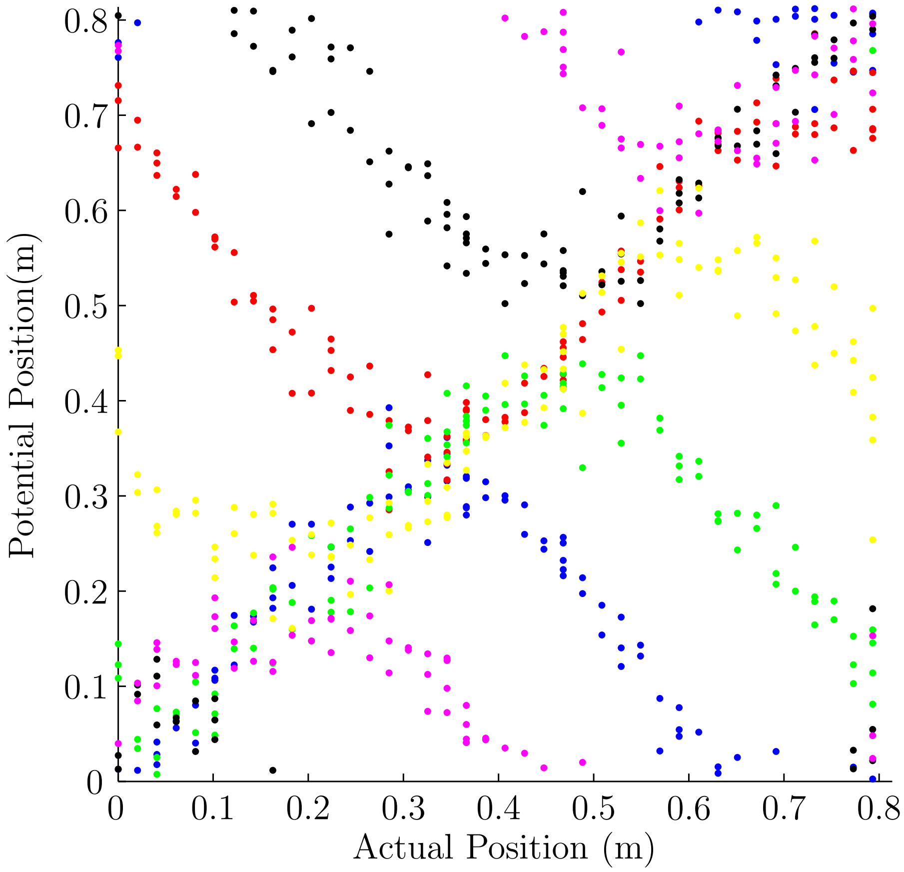

Figure 6.5: Six potential positions represented by each colour for each AE hit

Therefore, if the six dimension vector of potential positions are grouped into four clusters it would be expected that one cluster would contain three positions, the centroid of which can be taken as the estimate of source location.

This method requires a suitable wave propagation speed V. It would be possible to calculate V through simple time of flights of ultrasound pulses. However, rather than use a pre-determined value of V, the mean squared error (MSE) for a range of speeds was calculated and the minimum error taken to be the most suitable speed. This approach was taken since the FTC tends to underestimate the speed of sound since a finite time will elapse between the burst reaching the sensor and the threshold being exceeded; therefore a higher threshold equates to a lower effective wave speed.

To find a suitable wave propagation speed V for the TOA method, the average squared error between the predicted and target positions for a hundred artificial sources was calculated using the process of determining all potential positions and finding the centroid of those relating to those found in the cluster. This was repeated for a range of speeds from 10 to 10000ms^{-1}. The results are shown in Figure 6.6. The optimum propagation speed that minimises the error was found to be 3700ms^{-1}. This closely corresponds to a typical value for the speed of secondary waves in steel.

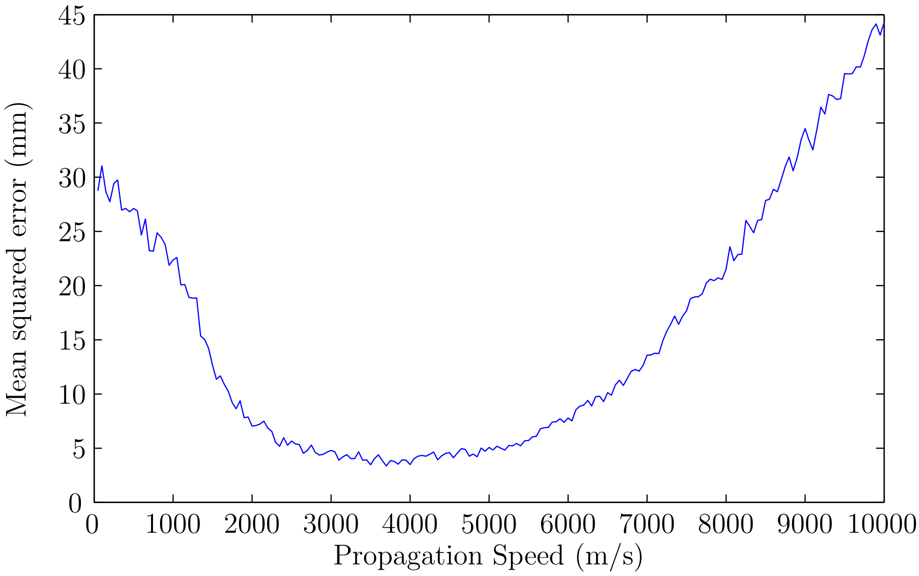

Figure 6.6: Mean squared error for range of propagation speeds

The estimated locations of the hundred artificial sources calculated when $V = 3700ms^{-1}$ are shown in Figure 6.7. Each datum in Figure 6.7 represents a source where the correct prediction of circumferential position would fall on the diagonal red line. A greater deviation from the line signifies a greater error. The error, defined as the average absolute difference between the expected and estimated position around the raceway, for this method is 42mm. Given the circumference of the inner race is 0.814m, this corresponds to a mean percentage circumferential error of 5.2%. Although most sources are located with reasonable accuracy there are three significant anomalies introduced by this method.

6.3.3 Delta T Mapping

The Delta T mapping method for source location was first proposed by Baxter et al [92] in order to overcome inaccuracies introduced by complex geometric structures and composite materials highlighted in previuosly in Section 6.3.2. Artificial AE sources performed on the area of interest enable maps of differences in arrival times (ΔT) between sensor pairs to be constructed; one for each sensor pair. The location of real

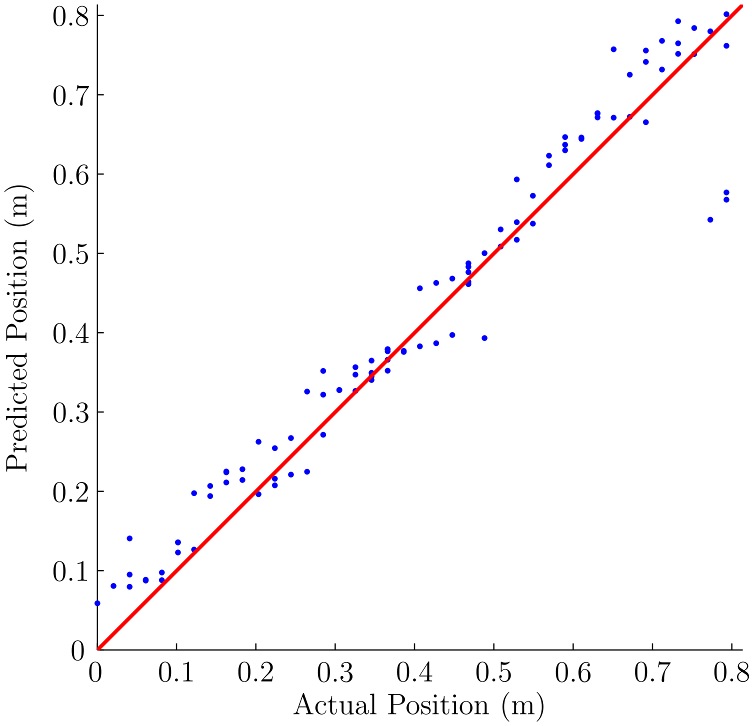

Figure 6.7: Predicted Positions using TOA Method

sources can then be estimated by the intersection of possible contours on the ΔT maps. This method eliminates the need for both a theoretical analysis of wave propagation and an accurate measurement of wave propagation speed. Furthermore, knowledge of the position of the sensors is not required assuming the sensors do not move during or between tests. A detailed methodology is provided in [92] and is summarised below.

Step 1 - Define area of interest: Restrict area to regions of high importance or high stress since conducting artificial sources can be time consuming. The rest of the structure may be monitored by analytical or zonal methods.

Step 2 - Determine grid density: Accuracy is determined by the density of artificial sources. This can be increased by bilinear interpolation.

Step 3 - Conduct artificial events at grid nodes: Previous work has suggested that between 5 and 10 sources are required at each node to reduce error to an acceptable level [92, 100].

Step 4 - Create ΔT maps: The differences in arrival times can be calculated and averaged for each sensor pair and plotted to form a contour map. By overlaying the maps from each of the sensor pairs, the location of an actual source can be estimated by identifying the intersection point of an actual AE event.

Baxter et al reported a reduction in source location error from 4.81% to 1.77 % when using the Delta T method compared to the analytical solutions to locate sources on a specific landing gear component [92].

These steps were carried out on the rolling surface of the inner raceway of the rolling element bearing. A grid was constructed with 40 nodes around the circumference and 5 across the raceway at approximately 20mm spacing in both directions. In order to verify how many sources were required at each node a number of Hsu-Nielsen sources were performed at an arbitrary point on the raceway, after approximately twenty five sources were generated, the percentage standard error of the mean difference in arrival time between two sensors was reduced to less than 1% (Figure 6.8). Therefore, twenty five sources were conducted at each node. The time differences, based on the FTC described in Section 6.3, for each sensor pair were calculated and averaged for each node.

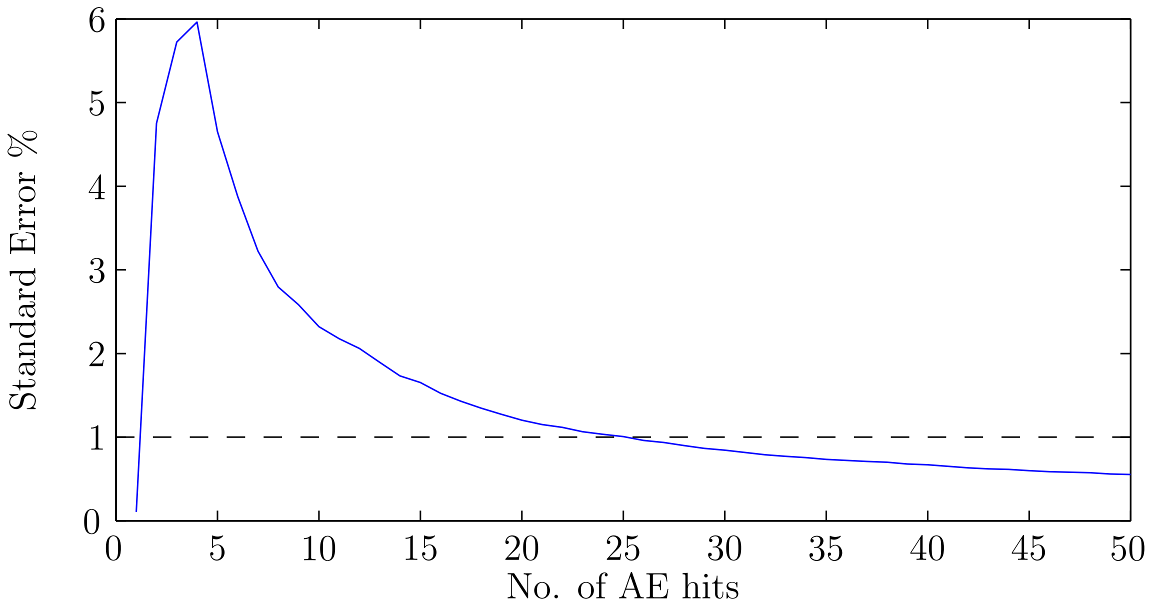

Figure 6.8: Percentage standard error of the mean time difference for a number of H-N sources

The three resulting ΔT maps are shown in Figure 6.9. Each map represents the inner raceway ‘unwrapped’ starting from sensor 1 at $x = 0$ following the race counter clockwise. Bilinear interpolation increased the resolution of the grid by a factor of 20 to approximately 1mm. Each source was located by finding the minimum of the sum of the differences between the arrival times of real data and the theoretical arrival times at each point on the ΔT maps.

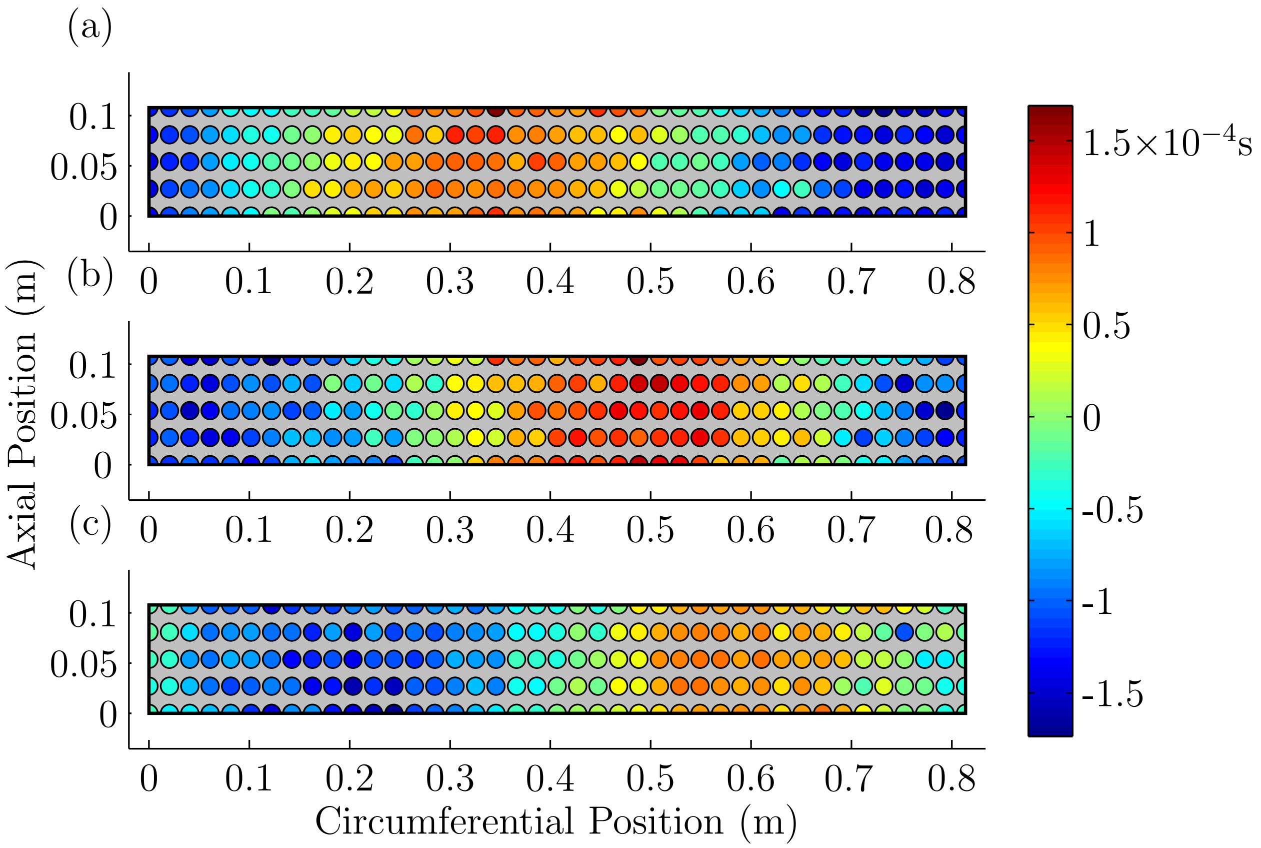

Figure 6.9: Delta T maps for sensor pairs (a) S1-S2, (b) S1-S3 and (c) S2-S3

The positions of the same hundred artificial sources used previously in Section 6.3.2 were estimated using the Delta T method to provide a direct comparison. The average error in this case is 29mm, a mean percentage error of 3.6%. The difference between the actual and estimated locations are shown graphically in Figure 6.10.

6.3.4 Neural Networks

Artificial neural networks (ANNs) are a series of connections mapping inputs to outputs designed to mimic the functionality of biological neural networks [101]. Each connection has an associated weight which is optimised during an iterative training process to minimise the MSE between the expected and actual output [102]. Neural networks are well suited to processing noisy data and given enough units are able to closely approximate any function. A typical multilayer feed forward network consists

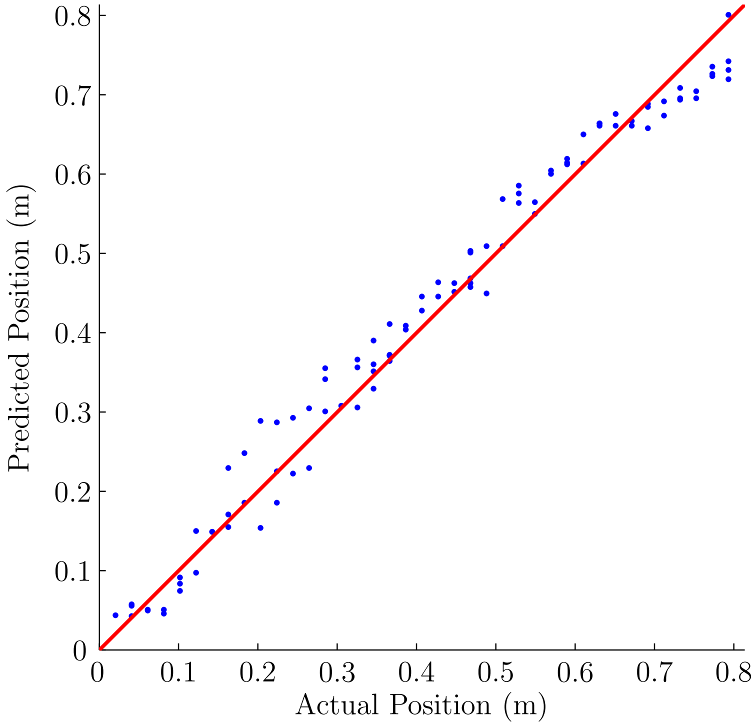

Figure 6.10: Predicted Positions using Delta T Maps

of an input layer, output layer and one or more hidden layers. The number of input and output units depends solely upon the number of inputs and outputs whereas each hidden layer in between can contain any of a number of units. Each unit takes the weighted sum of the previous layer and applies a non-linear activation function allowing a large range of inputs to be mapped onto a smaller domain often between -1 and 1. The function should be differentiable to allow backpropagation to train the network. The log-sigmoid and hyperbolic tangent are examples of such functions. Each layer may also have a bias which can be added to the weighted input of each unit. An example of a two layer feed forward network with 4 inputs (x_1-x_4), 3 hidden units and 2 outputs (y_1-y_2) is shown in Figure 6.11. A weight bias is added to the output layer. The network is said to be fully connected since every unit takes the weighted sum of every unit from the previous layer.

The most common method of training neural networks is through backpropagation [103]. This is a process whereby the error between target and output values for a training set are propagated backwards through the network (hence the need for a differentiable activation function) and the weights are updated to reduce the MSE. This process continues until the network convergences or reaches a satisfactory level

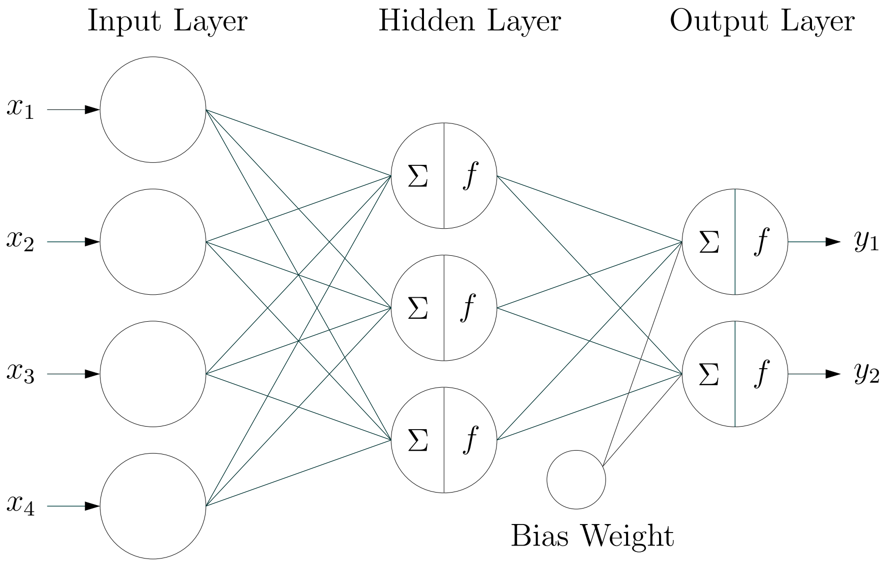

Figure 6.11: Example of Multilayer Feedforward Network

of performance. There are, however, a number of limitations associated with backpropagation. The most significant of which being the result may get caught in a local minima, training can be slow and that convergence is not guaranteed. Various modifications of the basic algorithm have been proposed to overcome these limitations. The Levenberg-Marquardt [104, 105] backpropagation algorithm, for example, has been designed to speed up the training process and other methods employ the use of momentum to prevent the algorithm getting caught in a local minima. Neural networks are proposed here for the purpose of source location by using arrival times as the inputs and training the network to output the coordinates of the source position.

In order to train and simulate a neural network, the architecture must first be defined. There are, however, no clear guidelines governing the optimum network architecture required for a given problem. A number of activation functions exist and any number of layers, hidden units or biases may be used to form a network. The most suitable architecture is often found through trial and error or from previous experience. This work has used a genetic algorithm to find the optimum topology in order to map differences in arrival times between sensors to source locations on the bearing raceway. A genetic algorithm is a problem solving technique designed to mimic natural evolution. This method follows an approach previously used to find the optimum network architecture for credit risk and credit analysis problems [106, 107]. A fully connected three layer network can be represented by a chromosome; in this case, a 26 digit binary string as shown in Figure 6.12.

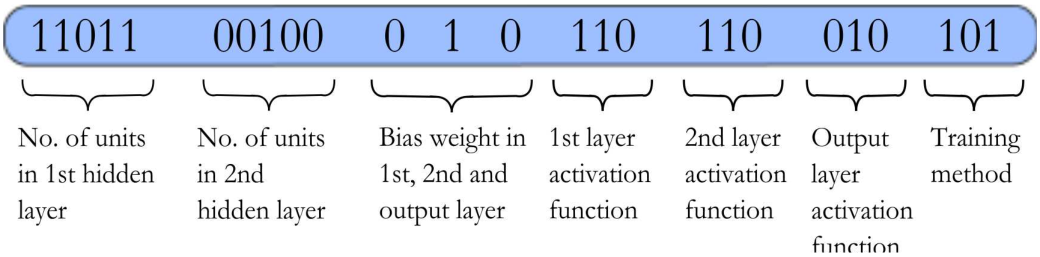

Figure 6.12: Binary Representation of a Neural Network

Each section of the binary string represents a feature of the network. Table 6.1 shows how these features can be decoded.

Table 6.1: Decoded Neural Network

Number of hidden units			
00000 00001	0 1	. . . 11111	. . . 32
Bias Weight			
0	No	1	Yes
Activation Function			
000	Hard Limit	100	Tan-Sigmoid
001	Log-Sigmoid	101	Softmax
010	Linear	110	Inverse
011	Positive Linear	111	Radial Basis
Training Algorithm			
0000	Levenberg-Marquardt back propagation	1000	One-step secant backpropagation
0001	Bayesian regulation back propagation	1001	One-step secant Gradient descent with momentum and adaptive learning rate backpropagation
0010	BFGS quasi-Newton back propagation	1010	Gradient descent with momentum backpropagation
0011	Resilient backpropagation	1011	Gradient descent with adaptive learning rate backpropagation
0100	Scaled conjugate gradient backpropagation	1100	One-step secant Gradient descent backpropagation
0101	CG backpropagation with Powell-Beale restarts	1101	One-step secant Sequential order incremental training with learning functions
0110	CG backpropagation with Fletcher-Reeves updates	1110	Cyclical order weight/bias training
0111	CG backpropagation with Polak-Ribiere updates	1111	One-step secant random order incremental training with learning functions

The algorithm randomly generates an initial population of 22 networks. Each network is trained using the Neural Networks Toolbox in Matlab and the cost function; the MSE of the predicted location, is calculated. The two networks resulting in the lowest cost function (the elite children) are retained to form a part of the following generation. The bulk of subsequent populations can be generated either by crossover or mutations of parents from the current generation.

In this work the remainder of the population is generated by three parent, eight point crossover plus the introduction of two random networks to ensure genetic diversity. The probability of a parent being chosen for crossover is inversely proportional to the ranking of the errors. This process is continued until convergence occurs and the cost function has reached a minimum. The network is trained with three inputs; the difference in arrival times between each sensor pair and a single output; the circumferential position of the source. Each network is trained 20 times and the cost function averaged since the same network architecture may produce different cost function due to the random initialisation of weights and random division of available data into training, validation and test sets. The cost function is calculated using an extra validation data set; the same as those used in the previous methods to provide a direct comparison. The same data used to create the ΔT maps has been used to train the networks; 5 artificial sources at 200 equally spaced nodes on the inner raceway. The data is randomly divided so that 70% of the data makes up the training set and 15% make up both the validation and testing sets. Figure 6.13 shows the convergence of the minimum mean-squared error of each population which occurs after approximately 40 generations. Table 6.2 shows the network architecture found by the algorithm.

Table 6.2: Optimum Neural Network Topology

No. of units in 1st layer	23
No. of units in 2nd layer	27
Bias weight in 1st layer	Yes
Bias weight in 2nd layer	No
Bias weight in output layer	No
Activation function in 1st layer	Hyperbolic Tangent Sigmoid
Activation function in 2nd layer	Softmax
Activation function in output layer	Hyperbolic Tangent Sigmoid
Training algorithm	Levenberg-Marquardt backpropagation

The optimised network, found by the genetic algorithm, was simulated, again with

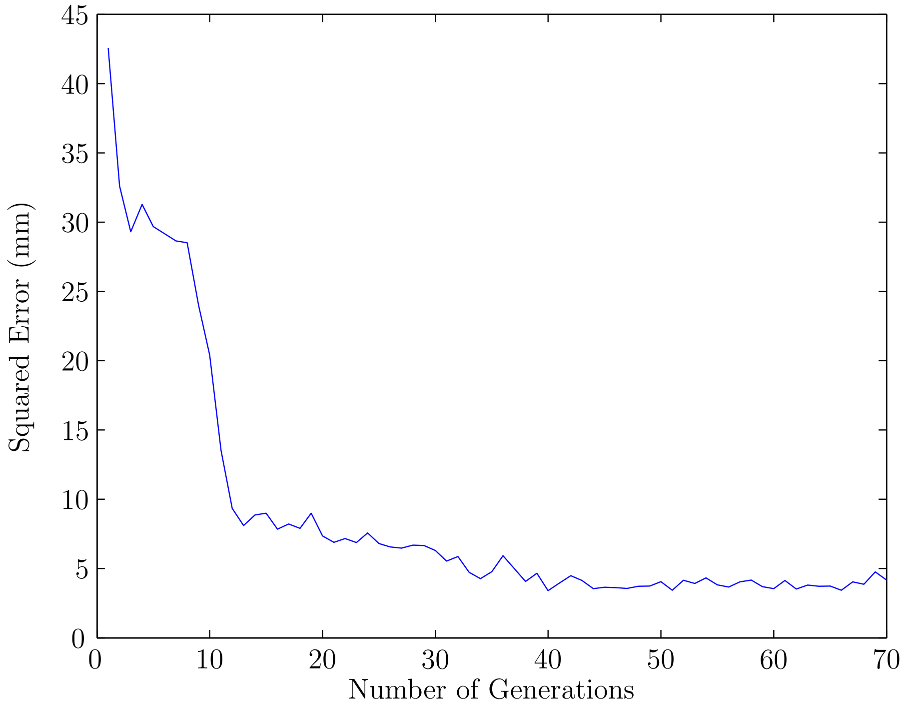

Figure 6.13: Convergence of Genetic Algorithm

the same hundred artificial sources. The average error is 17mm or 2.1%. The predicted positions are shown in Figure 6.14.

6.4 Sources Location of a Real Defect

The Multilife™ test rig described in Chapter 4 was run at 200kN and 100rpm with an artificially damaged bearing. A line defect, the details of which are given in Section 4.1.1, was created across the contact surface of the inner raceway and was positioned at top dead bottom (TDB) where the load is at a maximum. Under these conditions, a full lubricant film is expected (see Chapter 7), therefore any hits that arise are assumed to be generated by rollers impinging on the defect rather than due to asperity contact. The accuracy of the methods can be validated in a real situation by comparing the estimated positions to the position of the defect. A short sample of time series data is shown in Figure 6.15. A threshold of 0.015V (or 63dB) was chosen to capture the transient signals. This is significantly higher than background noise but chosen to ensure that the recorded hits were caused by the defect rather than noise for the purpose of validation. In reality a lower threshold would be used but in such case

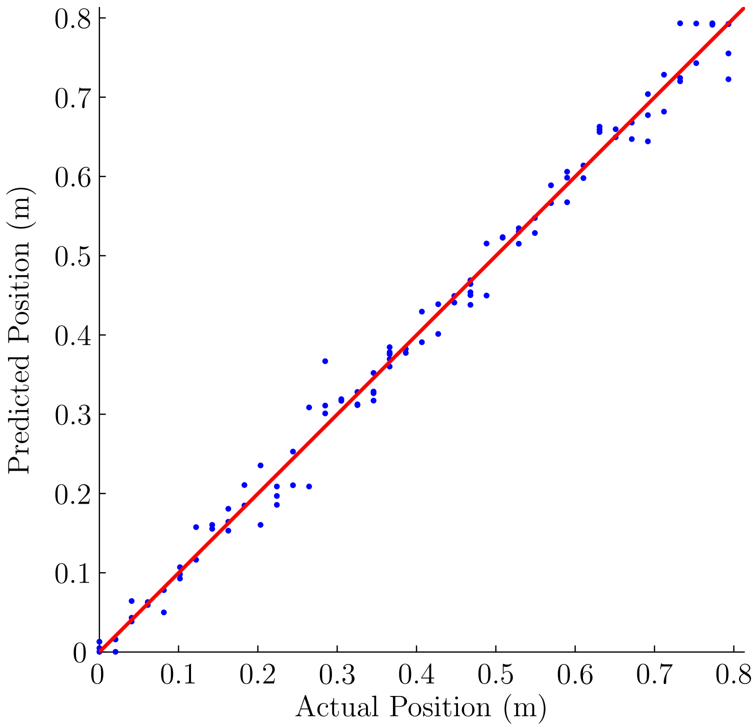

Figure 6.14: Predicted Positions using Neural Networks

concentrations of events in a single position would be used to determine the position of a defect rather than the estimated position of individual events.

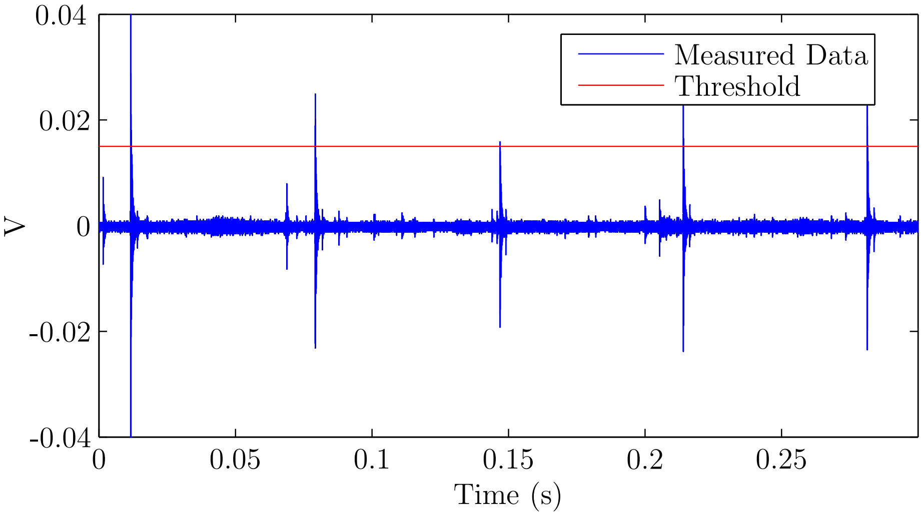

Figure 6.15: Signal from a Damaged Bearing with Threshold

The times of all threshold crossing were recorded until 30,000 individual hits had taken place across all channels. The data was cleaned up by only retaining data where the time difference between the earliest and latest threshold crossing was less than the time it would take a wave to propagate between the sensors. This removes any data where a threshold crossing occurred at only one or two sensors and increases the possibility that threshold crossings which occur at all three sensors within the this time are caused by the same source. As a result, just less than 2,000 events were analysed. A histogram of the estimated position of each approach, TOA, Delta T and neural networks, is shown in Figure 6.16. The average error, defined as the difference between the estimated and real position of the defect, associated with each of the methods is given in Table 6.3.

Table 6.3: Mean Error for each Source Location Method

Method	Average Error (mm)	% Error
Time of Arrival	40	4.9
Delta T Mapping	25	3.1
Neural Networks	22	2.7

6.5 Discussion

A comparison and summary of the accuracy of the three source location methods investigated in this study in addition to the maximum error expected by the zone location is shown in Figure 6.17.

Initially the methods were tested using artificial pencil lead breaks. It is clear that in this case both the Delta T mapping technique and neural networks offer an improved performance over the TOA method whose mean error is 42mm. Neural networks in particular have the capability of reducing circumferential error to 17mm. This corresponds to a percentage error in source location of just 2.1%. The Delta T technique has also been shown to reduce mean circumferential error to 29mm or 3.6%.

The artificial sources allowed many events to be generated quickly over the entire surface of the raceway and obtain an idea of the efficiency of each method before the test rig was commissioned. However, the introduction of background noise, different mechanisms for generating the transient bursts and potentially different propagation paths meant it was necessary to test each method using sources generated by a real defect. The corresponding errors were similar to those for the artificial sources, the

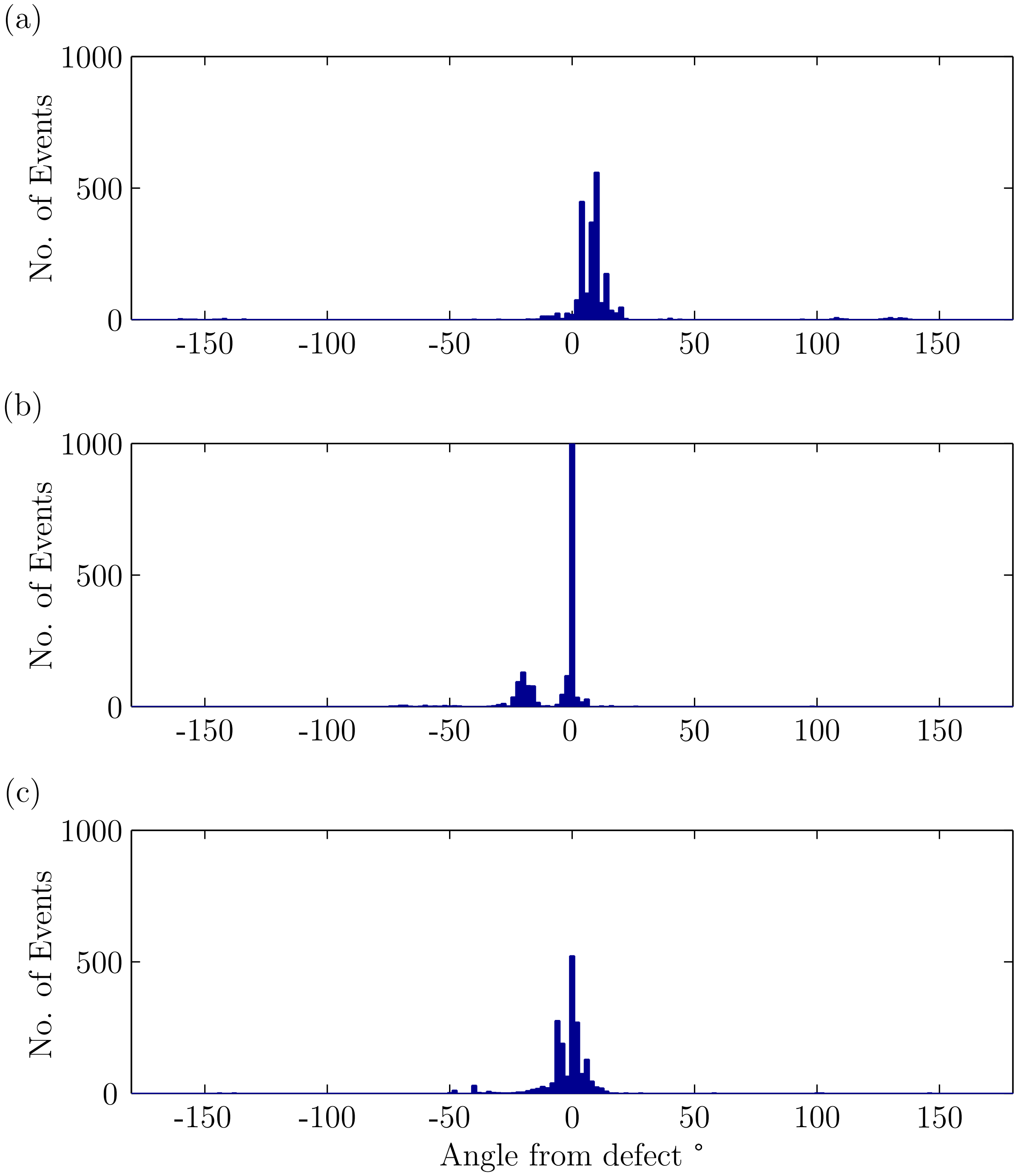

Figure 6.16: Histogram of Predicted Positions using (a) TOA (b) Delta T Mapping (c) Neural Network

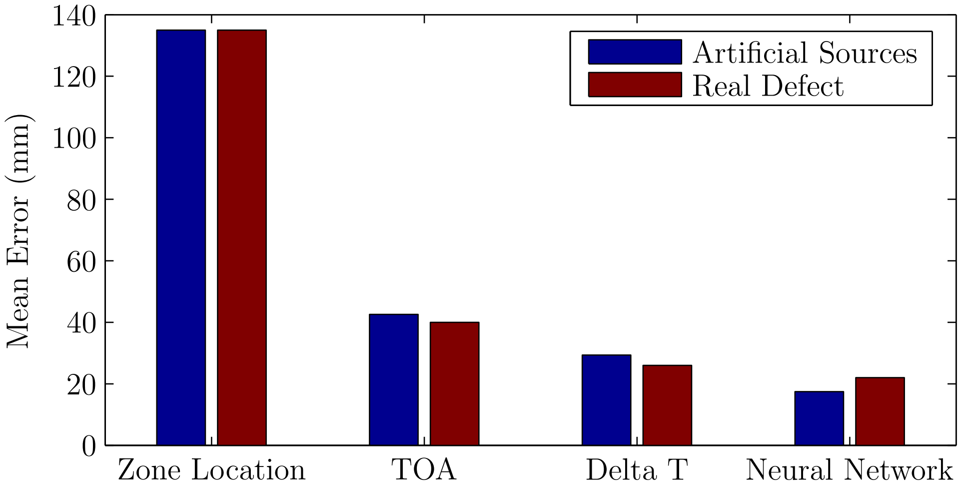

Figure 6.17: Summary of Location Methods

TOA and Delta T mapping errors were slightly lower at 40 and 25mm respectively whereas the performance of the neural network was less accurate at 22mm.

The TOA method provides a reasonable estimation of AE source location and would be suitable for large scale applications with simple geometries. Due to the limitations and time constraints of Delta T mapping and neural networks it would be better practice to restrict these techniques to areas of high importance or stress and monitor the rest of the structure by the TOA method.

The major drawback of both Delta T mapping and neural networks is in obtaining the data to create the ΔT maps and train the networks. This can be time consuming if the structure is large and there may be cases where access to the surface, for the purpose of performing artificial sources, is difficult or indeed impossible. This would be the case for ball bearings where the raceway cannot be removed. In these situations TOA methods may have to suffice. However, the training data only needs to be obtained once and can be used for any identical systems in the future.

During the early stages of damage when the amplitude of the AE events are small and background noise is greater due the rolling contact it is unsuitable to use the FTC. In this case, the use of cross-correlation of the enveloped signal or the use of a decomposed signal with a higher SNR as described in Section 5.1.2 may be required. However, such techniques may introduce slight errors due to the difference in time difference calculation.

The decision to use one method over another may be based on efficiency in addition to accuracy. Online application requires the computational time per event to be shorter than the hit definition time; the time the system waits after a hit is triggered to prevent multiple hits being recorded from a single event. The neural network method was most computationally expensive; the system used for this work took 0.0096 seconds to compute the position of the 100 events. The TOA method was faster, taking 67% of the time of the neural network while the Delta T method was a significantly faster still, only taking 15% of the time.

Since the indexing strategy of the MultiLife™ bearing will depend upon the position of the defect circumferentially the accuracy across the width of the raceway is of little importance. Any increase in axial accuracy would be impossible to achieve with a similar placement of sensors as used in this work since all three are positioned halfway across the raceway axially, 120° apart. Two emissions at the same angle but on opposing sides on the raceway axially would appear very similar to the sensors. By repositioning the sensors, greater accuracy may be possible potentially informing the user of instances of, for example, misalignment however this is beyond the scope of the project due to design restrictions.

6.6 Conclusion

The use of the Delta T method and neural networks for the purpose of estimating the source location of AE bursts has been investigated. Artificial sources generated on the removed inner race of a bearing taken from a planet gear of a wind turbine gearbox have been used to train and validate each method before being used to estimate the position of a real defect when run in a test rig. Both methods have been shown to offer an increase in accuracy when compared to TOA methods which have been used for bearing source location until present. The circumferential error for the real defect was shown to be 4.9% for the TOA method whereas this error was reduced to 3.1% and 2.7% for Delta T mapping and neural networks respectively. The use of a particular method should be dependent on the application. In those cases where many tests are being performed on identical bearings, such as in a laboratory, it may be preferable to invest some time in generating artificial sources and thus training data for an increase in accuracy whereas the time of arrival method would be more suited to those instances where time or access is restricted or if the bearing in question is only one of that size being investigated. Moreover, it may be considered that neural networks would be more suitable for online application due to a reduced computational time. Owing to the design of the test rig and configuration of the sensors it has not been possible to calculate the position of the defect axially.

Chapter 7

Test Cases

This chapter initially looks at the expected lubrication regime of the MultiLife™ bearing rig and how the AE signal responds to changes in oil film thickness. The findings suggest that the expected lubrication regime should influence how a bearing is monitored. The results of two run-to-failure tests with seeded defects are then presented using the proposed approach for damage detection described in Chapter 5. A brief remark is made concerning the instrumentation of the high speed shaft bearings of a 600kW wind turbine gearbox and some initial data under different loading conditions is presented. The chapter concludes by investigating data measured from an undamaged and damaged split race bearing test rig under various conditions.

7.1 MultiLife™ Rig Results

7.1.1 Oil Film Thickness

The expected lubrication regime can be estimated by the specific film thickness λ; a ratio of the minimum film thickness to the composite root mean squared roughness of the mating surfaces $R_{q,T}$ given by Equation 7.1. According to Khonsari [108], full surface separation and no asperity contact is expected when λ exceeds three while Williams [109] suggests asperity contact should be taken into account when $1 < \lambda <$

5.

$$\lambda = \frac{h_{min}}{R_{q,T}} \tag{7.1}$$

Based upon the simultaneous numerical solutions of Hertzian contact stresses, lubrication behaviour and the effect of pressure on increasing oil viscosity the minimum film thickness of a line contact can be estimated by;

$$\frac{h_{min}}{R} = \frac{2.65U^{0.7}G^{0.54}}{W^{0.13}} \tag{7.2}$$

where U, G and W are the dimensionless speed, material and load parameters respectively and R is the conjunction radius between the rolling element and the inner race [110]. Further details of which are given in Appendix B.

In order to calculate the minimum film thickness the load on the most heavily loaded rolling element in a bearing calculated according to Equation 4.6. Prior to performing full bearing life tests under steady conditions the bearing was run whilst subjected to a range of loads to vary the film thickness. The data required to calculate the film thickness is shown in Table 7.1.

Table 7.1: Experimental Constants for Calculating Film Thickness

Radius of rolling element (R_1)	27 mm
Radius of inner raceway surface (R_2)	129 mm
Width of contact (L)	72 mm
Dynamic viscosity of lubricant (μ)	1.13×10^{-8} Pa.s
Pressure viscosity coefficient (α)	0.02 Pa^{-1}
Elastic modulus (E)	210 GPa
Poisson's ratio (v)	0.3
Rotational speed (n)	100 rpm

Samples of AE data taken over a range of loads from 0-1000kN at 100rpm were recorded once the temperature had stabilised at 65±2° C. The RMS, maximum amplitude, kurtosis and crest factor, concepts introduced in Section 3.2 is calculated at each load while the specific film thickness λ is calculated using the composite surface roughness of 0.195μm given in Section 4.1.1. The results are shown in Figure 7.1.

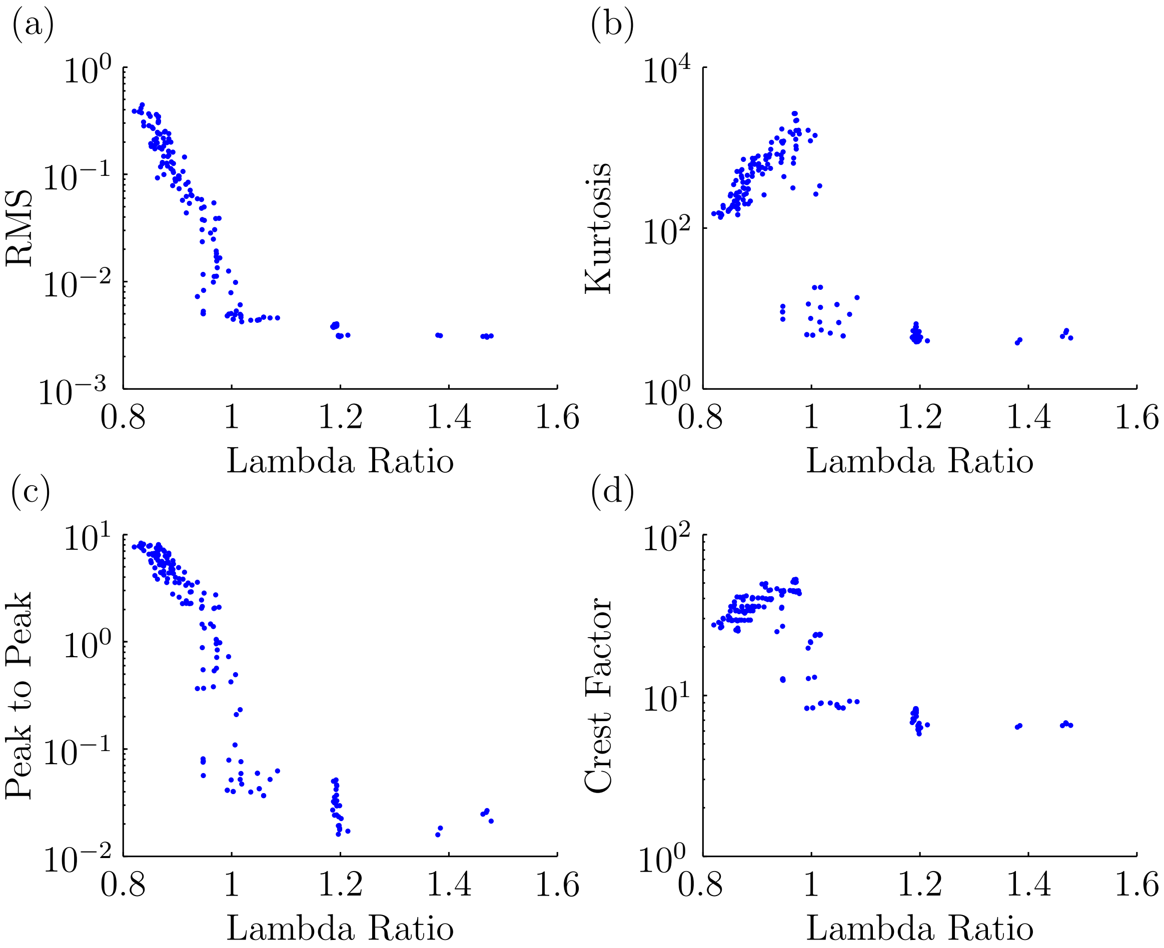

Figure 7.1: Statistical Parameters against Lambda Ratio

The operating conditions correspond to a range of specific film thicknesses from 0.8 to 1.5. Graph (a) shows that the RMS of the measured signal is relatively stable when $\lambda > 1$ suggesting the asperity contact is minimal and mechanisms through which the signal is generated is constant. As λ drops below 1, the film thickness is not large enough to ensure full separation between the components and asperity contact takes place. This manifests as an increase in RMS, the magnitude of which increases sharply with thinner film thickness as the rate and severity of the contacts increase. The kurtosis shows a similar pattern but suggests that some asperity contact may occur at slightly larger film thicknesses when $\lambda \leq 1.1$. At this point, the resulting transient signals, although sparse and of low amplitude, will increase the kurtosis but have a negligible effect on the RMS. As the film thickness decreases further, asperity contact becomes increasingly regular and the resulting transient signals start to blend together. As such, the signal approaches a Gaussian distribution and the kurtosis reduces. At larger film thickness the kurtosis is around three, typical of a fully lubricated undamaged bearing. The peak to peak value as shown in (c) also shows the maximum value decreasing as film thickness increases. At the smallest film thickness recorded, some of the samples were clipped due to the range of the ADC which otherwise would have resulted in larger amplitudes. The crest factor, the ratio of peak value to RMS, shown in (d) increases once metal to metal contact occurs but decreases at even smaller film thicknesses as the RMS increases at a faster rate than the peak value.

Measured AE waveforms are shown in Figure 7.2. Under a nominal load, the waveform shows an almost typically Gaussian distribution (a). There are some features in the signal but no events greatly exceed the background noise. At 300kN, the emergence of high amplitude events is clear (b), the frequency and amplitude of which become greater with increasing load (c). The presence of such events, indicative of boundary lubrication, means that many methods common to bearing monitoring such as peak value or kurtosis which relate to individual events of high amplitude are redundant since smaller events due to roller impingement will not affect such measures.

The data shown in Figure 7.1 suggests the transition from a fully lubricated contact to mixed lubrication takes place when $\lambda \approx 1.1$. This specific film thickness is thinner than that suggested by Khonsari [108] and Williams [109]. A closer look at the surface profile, an example of which is shown in Figure 7.3, of the rolling element shows that the largest deviations away from the mean are on the negative side. In other words, the troughs are of a greater magnitude than the peaks which means the calculated

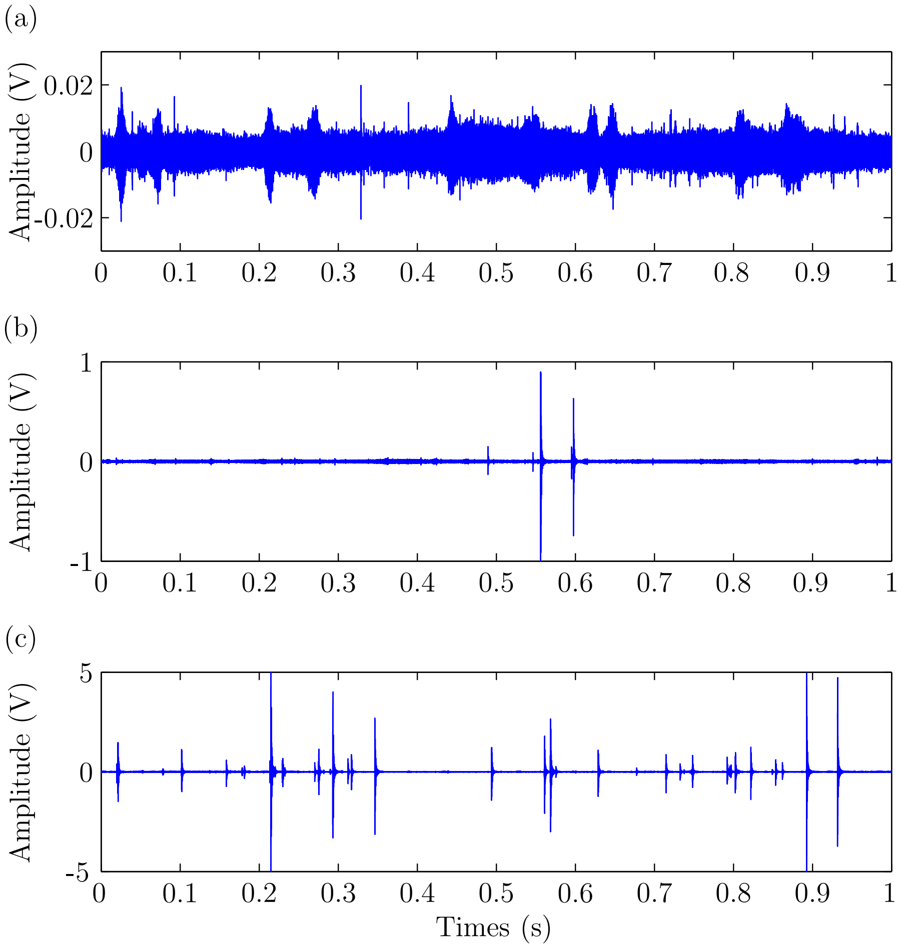

Figure 7.2: Measured AE waveforms at 0, 300 and 1000kN

R_q will suggest a higher level of asperity contact than actually takes place. Without these troughs the specific film thickness would be higher, closer to the values suggested by Khonsari and Williams. The oil temperature used in these calculations introduces further uncertainty. The samples were only taken once the temperature had stabilised but given the temperature was measured at the inlet it is likely to have cooled a small amount whilst passing through the pump meaning the temperature in the contact may be a slightly higher than the measured temperature giving a thinner lubricant film.

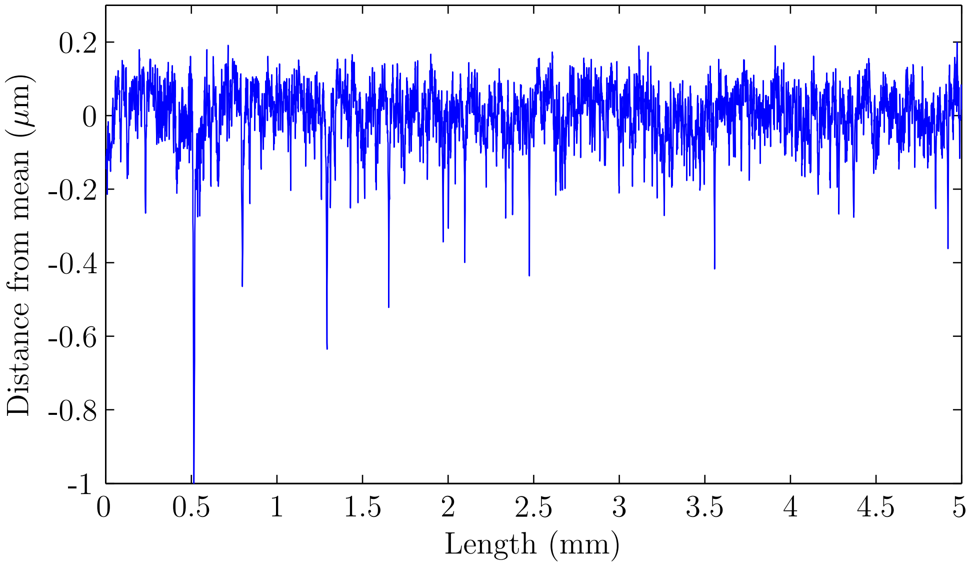

Figure 7.3: Surface Profile of Rolling Element

7.1.2 Run-to-Failure Tests

After some initial tests during commissioning of the rig, two run-to-failure tests were performed on the MultiLife™ rig under a constant radial load of 1200kN and constant speed of 100rpm. In each of the experiments the bearing was run initially in an undamaged state before being disassembled allowing a defect to be seeded on the inner raceway. The details of each experiment are given in the following sections.

Bearing 1

Test Sequence

After 700,000 revolutions of a new undamaged bearing, a defect measuring 100μm wide by 60μm deep machined on to the bearing surface, as shown in Figure 4.6, was positioned under maximum load. After a further 300,000 revolutions the defect had propagated significantly. The resulting damage is shown in Figure 7.4. It can be seen that the damage has propagated 'downstream' of the initial defect, potentially due to the rolling elements 'bouncing' off the defect and therefore dynamically loading the raceway causing higher contact stresses.

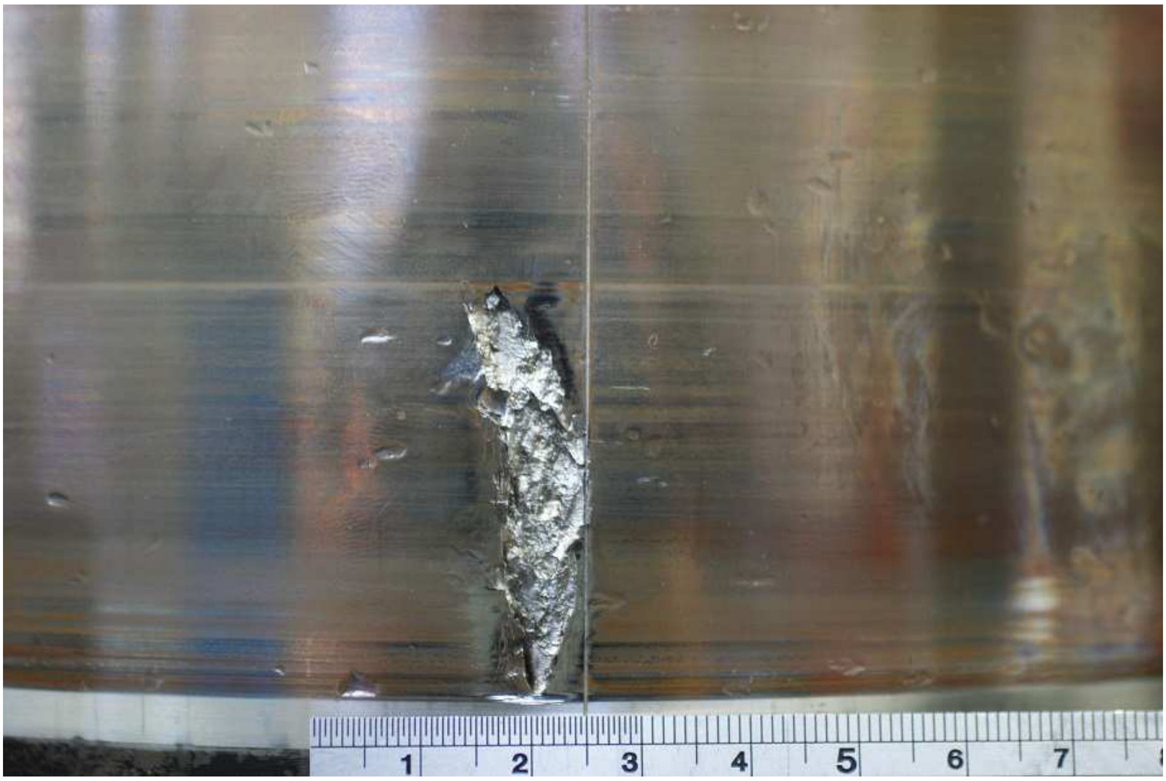

Figure 7.4: Inner Raceway Failure

The damaged inner race way was then manually indexed 180° whereby the defect would experience the minimal load while an undamaged part of the bearing was subjected to the maximum load. The test was continued for approximately 500,000 revolutions and on the subsequent inspection no further damage had taken place. This test sequence is summarised in Table 7.2.

Traditional Metrics

During the test, two seconds of acoustic emission data was recorded every 15 minutes across three channels. An average of the RMS, kurtosis, peak value and crest factor across the channels was calculated for each set of samples, plots of which are shown in Figure 7.5.

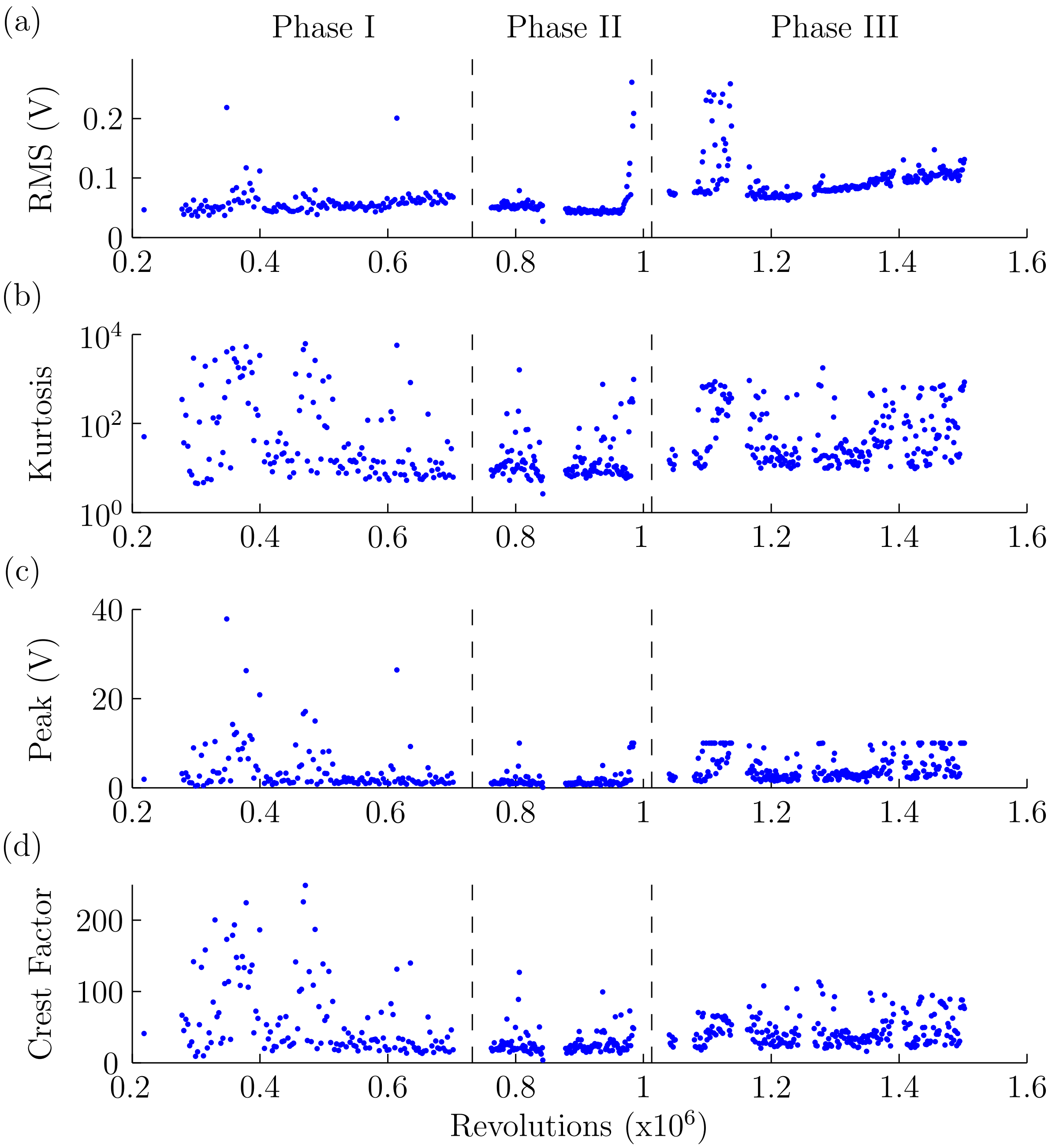

Figure 7.5: Trend of traditional metrics during Bearing 1 test.

Table 7.2: Bearing 1 Test Sequence

Phase	Bearing State	Revolutions
I	Bearing undamaged	0-700,000
II	Seeded defect positioned under maximum load	700,000-1,000,000
III	Raceway rotated 180°Ḋefect not under load	1,000,000-1,500,000

Figure 7.5 only shows the data when the test rig was running at or near to the stated conditions. After the inner raceway was rotated, the rig was disassembled at regular intervals to check the condition of the bearing and visually inspect for further damage. On each start-up it was necessary to gradually increase the load to allow the system to warm up, otherwise the drive belts would slip and melt (a mistake learnt the hard way). It is for this reason there are some gaps in the data. The dashed lines indicate the three sections of the experiment; the bearing in an undamaged state on the left hand side, the introduction and subsequent propagation of a seeded defect in the centre and the period following the indexing of the inner raceway on the right.

Figure 7.5 (a) shows that on the whole, the RMS is fairly constant, with the exception of a few high amplitude anomalies until the seeded defect propagates. At this point, the RMS rises sharply over a period of approximately 3 hours, or 14,000 revolutions. Once the damaged raceway has been indexed the RMS decreases but to levels marginally higher than prior to the failure. As the bearing condition continues to deteriorate, probably due to an increased level of debris denting the raceways and rolling elements, the RMS gradually increases. There is a temporary but significant increase in RMS at around 1.1×10^6 revolutions. The reason for such an increase is not clear but a possible explanation could relate to how the rig was assembled. It is noted that the increase in RMS falls between two disassemblies and could be a result of misalignment or looseness of a component in the rig. This shows that RMS can detect damage either when it is localised and significantly advanced or when the damage is distributed (as in the case of denting due to debris). This confirms the point made in Section 3.2.1 which states that when considering slow speed bearings, the additional energy due to localised defects is negligible compared to the background noise.

Figure 7.5 (b) shows the kurtosis is very susceptible to noise. Over the course of the test the value of kurtosis is spread over a range of three orders of magnitude due to the noise caused by asperity contact. Both the peak value (c) and the crest factor (d) are also unable to distinguish the difference between a damaged and undamaged bearing. The inability to detect incipient damage and determine the failed component in addition to the susceptibility to noise, confirms the need for a different approach. It should be noted here that after approximately 700,000 revolutions the gain setting of the pre-amplifier was changed from 40 to 60dB in an attempt to use the full dynamic range of the A/D converter. Unfortunately, this led to a number of signals clipping at 10V. This is addressed in the following section. The gain was increased during post-processing for those signals measured with a 40dB gain.

Outlier Analysis

The steps outlined in Chapter 5 were performed using the same raw data of a second sampled every 15 minutes used to generate the plot in Figure 7.5. The results are shown in Figure 7.6 with the three graphs (a)-(c) showing the results for the inner race, outer race and rolling elements respectively. The data representing the undamaged bearing is randomly divided into training and validation sets at a ratio of 85:15, the training data represented by green markers and the validation data represented by red markers. The blue markers represent the bearing in a damaged state. The horizontal green line represents the Monte Carlo threshold, the calculation of which is detailed in Section 5.2.2. It can be seen that across the three plots all but one datum in the validation set is correctly classified as undamaged, a rate of 95.2%. Once a defect is introduced, the measure of discordancy representing the inner race increases significantly above the threshold in contrast to the other plots representing the outer race and rolling element which generally remain below the threshold. This demonstrates that such a method is capable of detecting both early stage damage and determining which component is affected. It should be noted here that the threshold is a theoretical value based on a thirteen dimension vector under the assumption that the data will follow a Gaussian distribution. The threshold therefore is not necessarily fixed and another method of determining the threshold may be used depending on the relative importance of false positives and false negatives. For example, a second validation set could be used where the threshold is set as the maximum value of discordancy of this set.

The propagation of the defect can be seen prior to indexing the inner race however the logarithmic scale perhaps visually understates the degree of the damage. Between the introduction of damage and the point at which the defect propagates, the discordancy decreases slightly over time. It is likely that this is due to the raised lips on the artificial defect, which can be seen in Figure 4.6, being rounded off over time and thus reducing the impact energy. This agrees with the findings of [78] who found that smoothed off defects were not detectable compared to new defects of the same size with protrusions at the edge. Once the inner race is rotated, the defect is no longer subjected to a high load and the impact of the rolling elements is reduced. This is manifested as a reduction in discordancy to a level slightly higher than in the undamaged state. At this point it would be difficult to say which component is damaged.

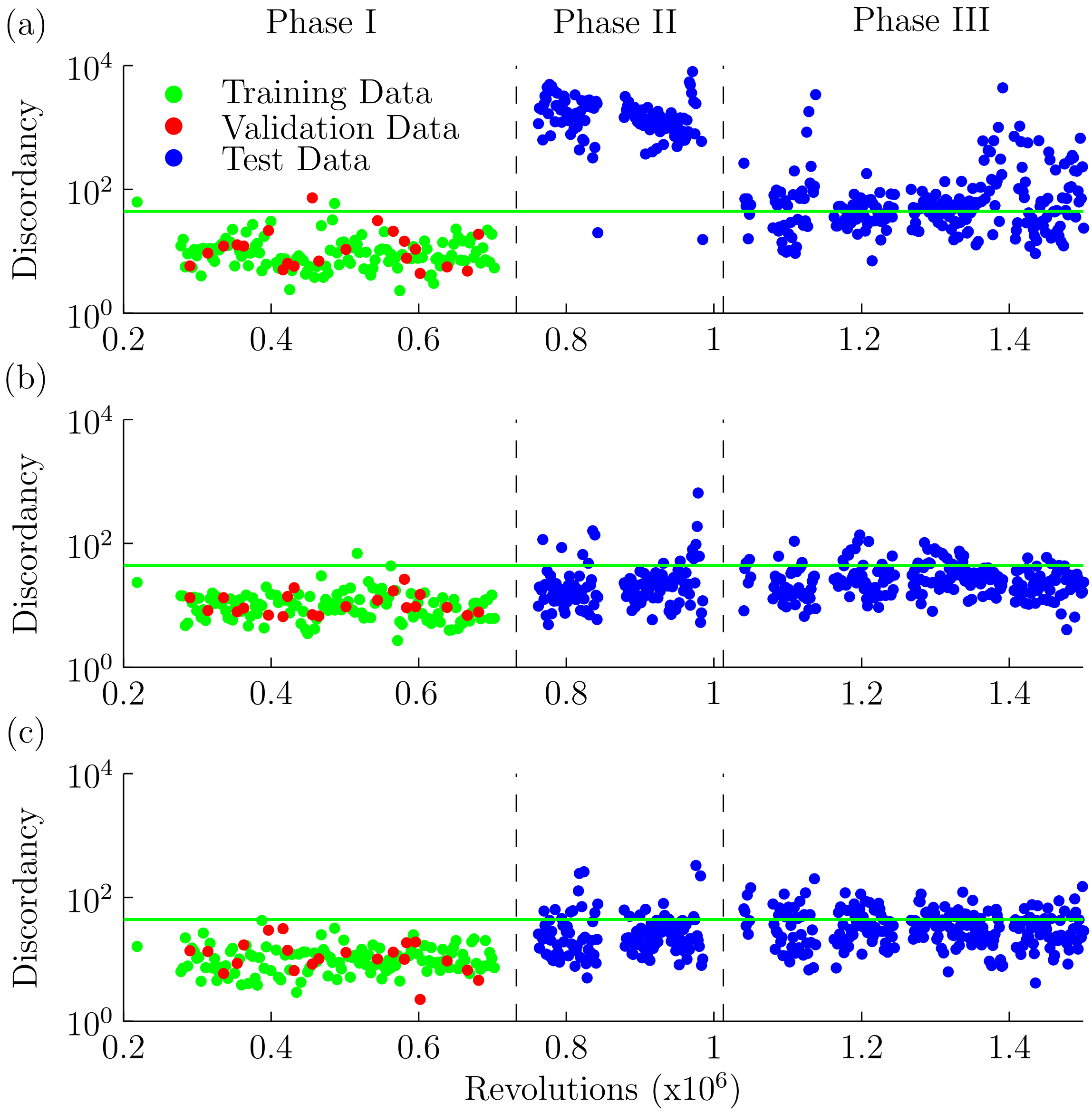

Figure 7.6: Outlier analysis of Bearing 1 for (a) inner race, (b) outer race and (c) rolling element

As noted in the previous section, a number of the signals clipped; limiting the peak values to 10V. To ensure this has not affected the results shown in Figure 7.6, the analysis was re-run excluding any signals in which clipping had taken place, a process which removed 10% of the signals. The results are shown in Figure 7.7. In this case, all but one of the validation data were correctly classified as undamaged, a rate of 94.7% while 99% of the damaged signals with the defect under load are correctly classified as damaged. Removing the clipped signals therefore had little effect on the performance.

As a means of comparison, outlier analysis has been performed with four dimension feature vectors comprising the traditional metrics shown in Figure 7.5. The same ratio of training and validation data has been used while the clipped signals were again excluded from the analysis. The results are shown in Figure 7.8. In this case, the metrics are not related to the characteristic defect frequencies and therefore are not component specific. It can be seen that in the undamaged state 100% of the validation set is correctly classified although only 2.5% of the signals are correctly classified once the seeded defect is introduced. In fact, the discordancy only increases above the Monte Carlo threshold once the damage has propagated significantly. As the damage is indexed away from the load, the discordancy falls below the threshold. The increase at 1.1x10^6 revolutions is driven by the increase in RMS at this point, the reason for which is postulated in Section 7.1.2.

Bearing 2

Test Sequence

The test sequence for the second bearing is summarised in Table 7.3. Again, a bearing was initially run in an undamaged state prior to being seeded with an EDM defect on the inner race after approximately 1.1×10^6 revolutions. With the defect acting as stress concentrator, coupled with the non-conformal contact and uni-directional loading, it was expected that the bearing would fail at the inner race with damage propagating from the seeded defect in a similar manner to the previous test. However, after 1.95×10^6 revolutions, the RMS of the bearing had increased sufficiently to warrant an inspection which revealed the outer race of the bearing had in fact failed (see photo in Figure 7.9) while no further damage had taken place on the inner race. One large spall can be seen clearly while smaller spalls were distributed around the entire

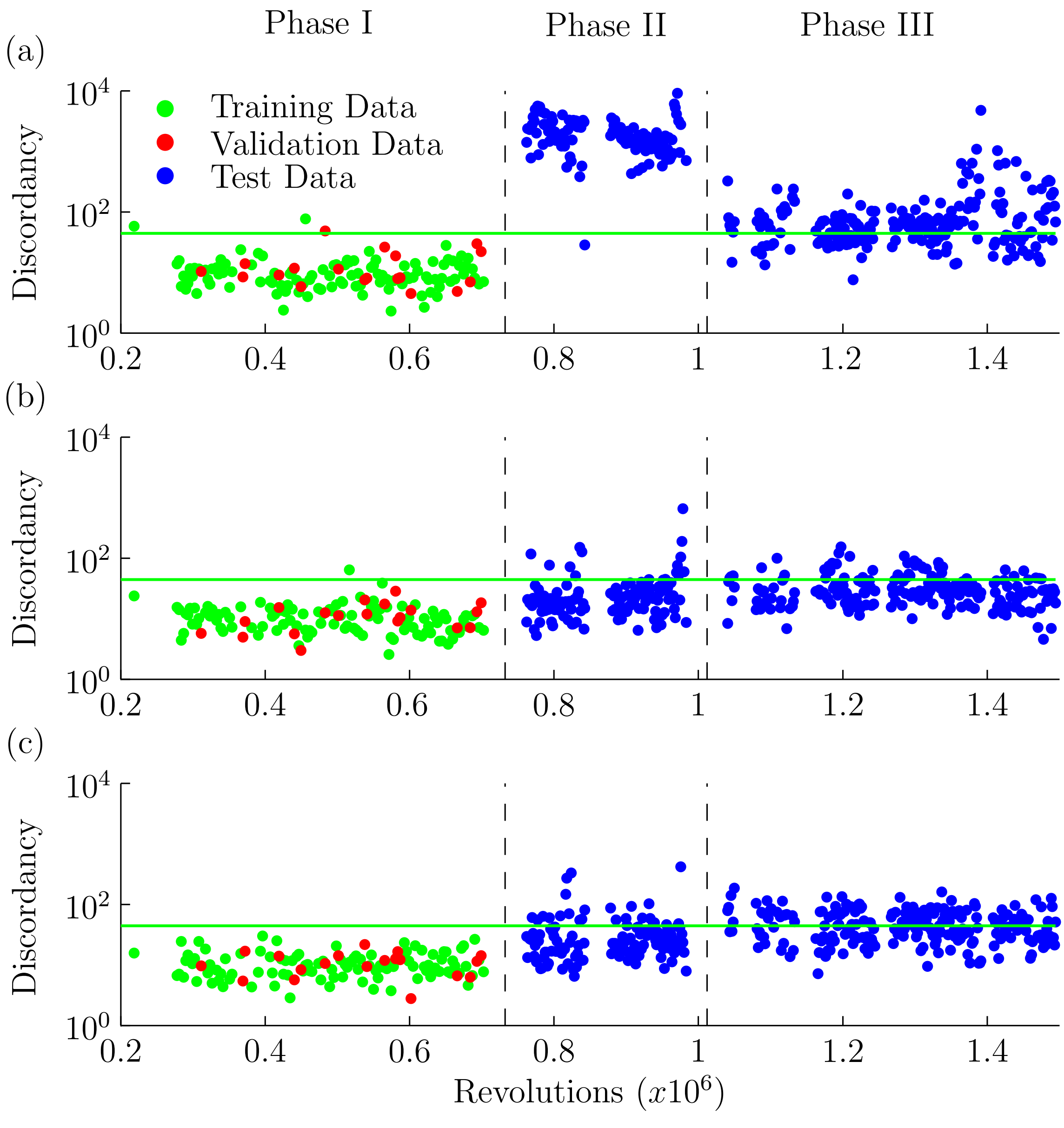

Figure 7.7: Outlier analysis of Bearing 1 excluding clipped signals for (a) inner race, (b) outer race and (c) rolling element

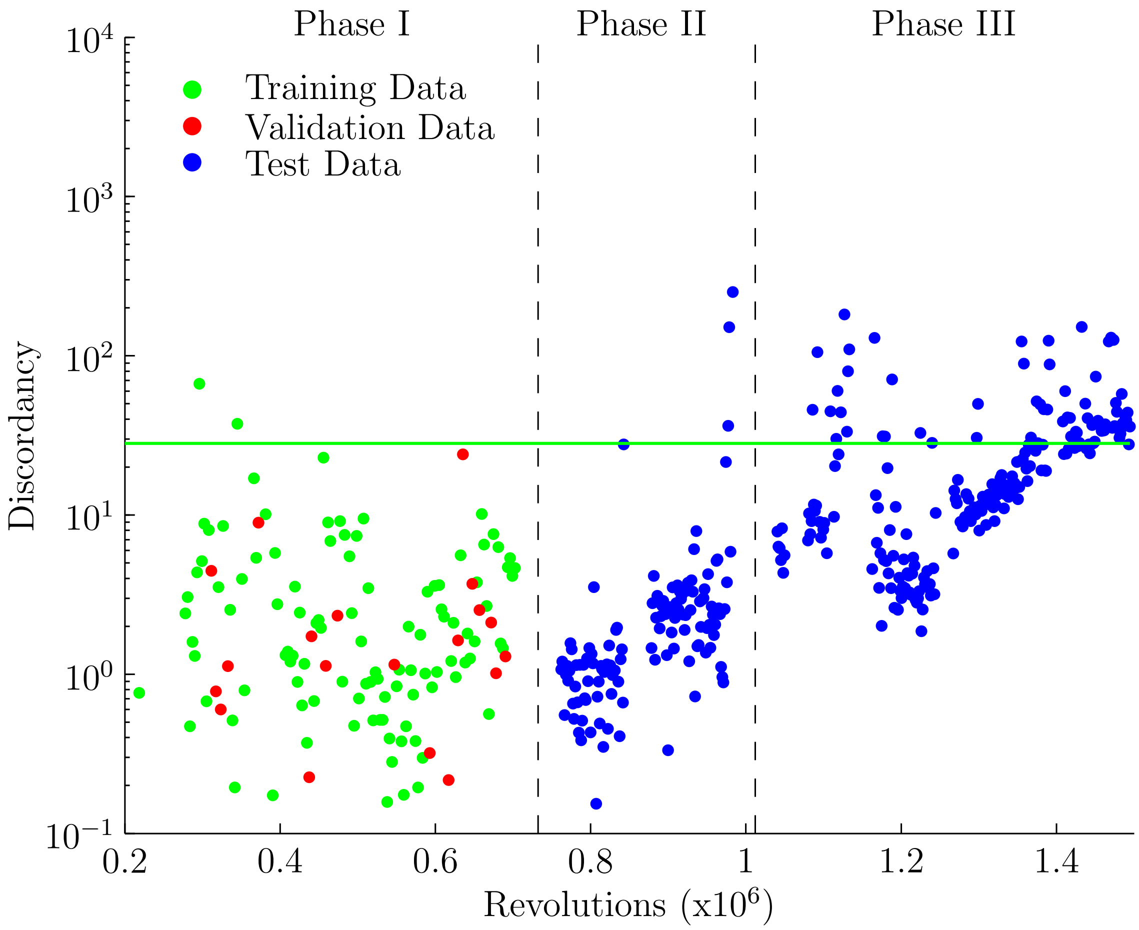

Figure 7.8: Outlier analysis with traditional metrics during Bearing 1 test.

circumference on one side of the bearing. The photo also shows a number of small dents on the raceway surface caused by rolling elements rolling over debris. Similar dents could also be seen on the inner race and rolling elements. The pitting that occurred at the edge of the raceway is likely to be due to misalignment caused by the self-aligning bearing shown previously in Figure 4.3. The misalignment meant the load was applied over a small contact area resulting in higher pressures ultimately leading to failure.

Table 7.3: Bearing 2 Test Sequence

Phase	Bearing State	Revolutions
I	Bearing undamaged	0-1,100,000
II	Seeded defect positioned under maximum load	1,100,000-1,950,000

Figure 7.9: Pitting to the Outer Raceway

Outlier Analysis

The results of the outlier analysis for this bearing are shown in Figure 7.10. On this occasion 99.2% of the validation set is correctly classified as undamaged. The introduction of the seeded defect on the inner race can clearly be seen in (a) with all but a few samples well exceeding the threshold. It should also be noted that the outlier analysis fails to classify the damage to the outer race towards the end of the test. The emissions resulting from damage to the outer race must propagate across two extra interfaces including the interface between the sensor face and metal plug to which it is bonded which will significantly reduce the energy reaching the AE transducers to such an extent that the signal becomes indistinguishable from the noise. There have been studies such as [75] that have shown it is possible to detect localised damage to a component other than that which the transducers is coupled to, however in that case the rotational speed was 1500rpm rather than 100rpm. Here, however, it is likely that the compound effect of the low rotational speed reducing the energy of the signal and the thin oil film increasing the background noise means the presence of a defect can not be detected and localised.

Figure 7.11 shows the specific film thickness falls between 0.52 and 0.7 with an average value of 0.6 over the duration of the experiment. Earlier in Section 7.1.1 it was shown that when $\lambda < 1$, noise, assumed due to asperity contact, increases significantly. Furthermore, the distributed damage around the circumference of the outer race means that a single frequency related to the damage will not be evident in the spectrum. Such damage lends itself to a different approach employing measures that are more suitable to continuous signals such as RMS, kurtosis and peak-to-peak value as shown in Figure 7.12. However, looking at the RMS in (a), it can be seen that highest values are found while the bearing is undamaged, at some point two orders of magnitude higher than when the defect is seeded. Such an increase was seen in Bearing 1 and it was postulated that another mechanism relating to the assembly of the rig could have been responsible. A similar level of disorder is seen with kurtosis (b), crest factor (c) and peak-to-peak (d) values since these are heavily affected by peaks and extreme values which can arise from means other than damage.

Again, these measures were used to form a four-dimension feature vector to directly compare with the method proposed in Chapter 5. The results are shown in Figure 7.13. In this case, 100% of the validation set is correctly identified as undamaged. There is a clear increase in discordancy once the defect has been seeded onto the raceway however not enough to satisfactorily distinguish the data from the undamaged case and only 13% of the samples are classified as damaged. Furthermore, no additional increase in discordancy is seen once the bearing condition starts to worsen at about $1.8\text{x}10^6$ revolutions.

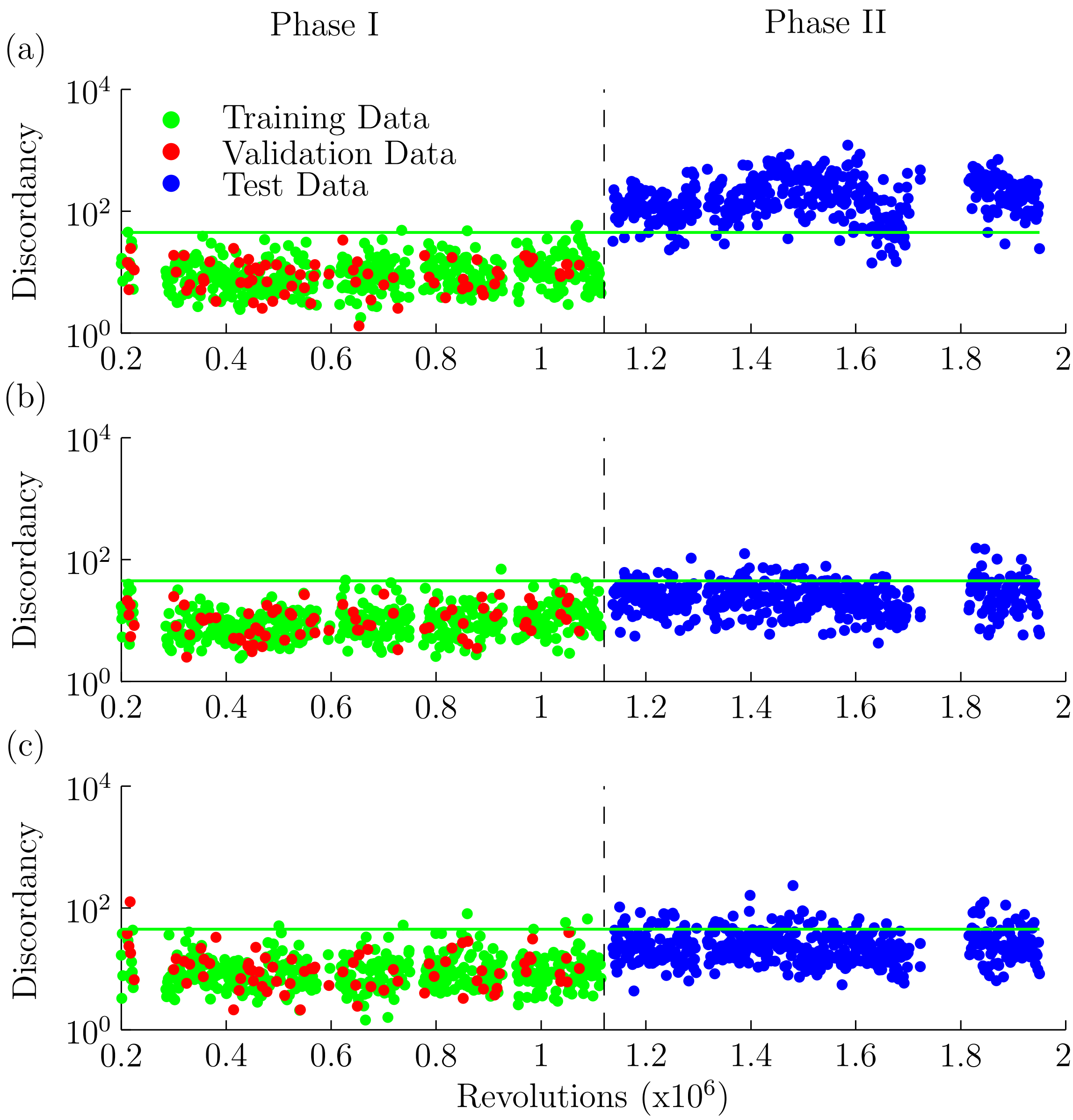

Figure 7.10: Outlier analysis of Bearing 2 for the (a) inner race, (b) outer race and (c) rolling element

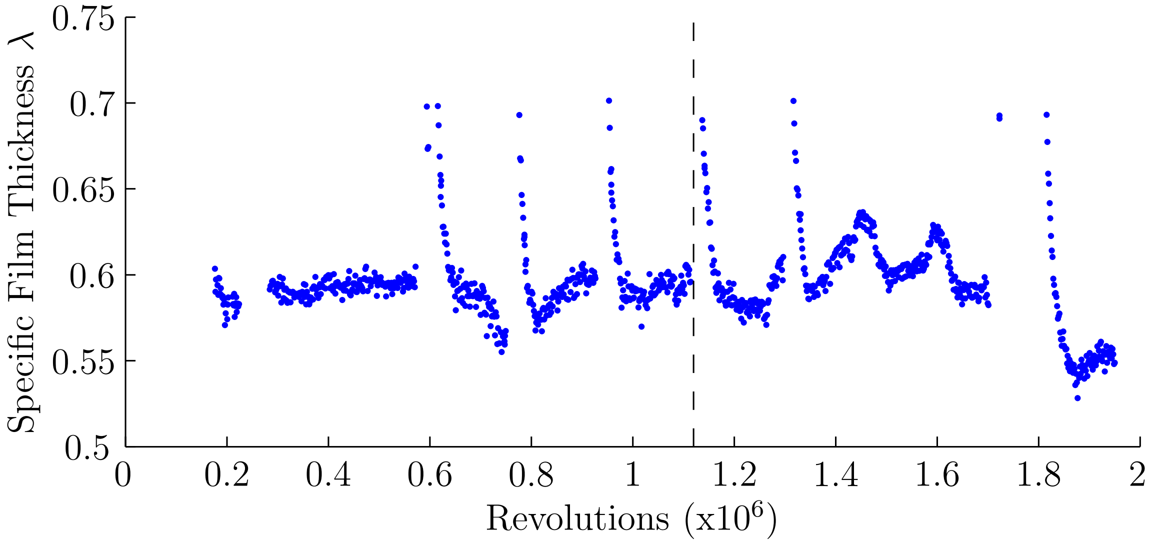

Figure 7.11: Specific Film Thickness during Bearing 2 Test

A more successful effort to monitor the degradation of the bearing towards failure is performed by using a method employed by environmental acousticians whereby the background noise is estimated by calculating the first percentile of the measured data. The first percentile value of each two second sample is shown in Figure 7.14. While no increase can be seen when the defect is introduced, which can be expected since the emissions due to the defect are of a much shorter duration than the time between them, there is a clear increase towards the end of the test as the outer race damage propagates towards failure.

7.2 Gearbox Commissioning

Wind turbines at the Barnesmore Wind Farm owned by Scottish Power situated on the border of the Republic of Ireland and Northern Ireland experience high failure rates due to the turbulent nature of the wind at the site. The increased chance of failure means that a turbine on this farm has been chosen to investigate the use of acoustic emissions and ultrasound as a means condition monitoring in an industrial environment. In the first instance, the outer race of a high speed shaft bearing on the gearbox of a Vestas V42 600kW wind turbine was instrumented with Physical Acoustics Micro-30D sensors due to relative ease of access, rather than having to deal with the problem of transferring the signal from a rotating component to the

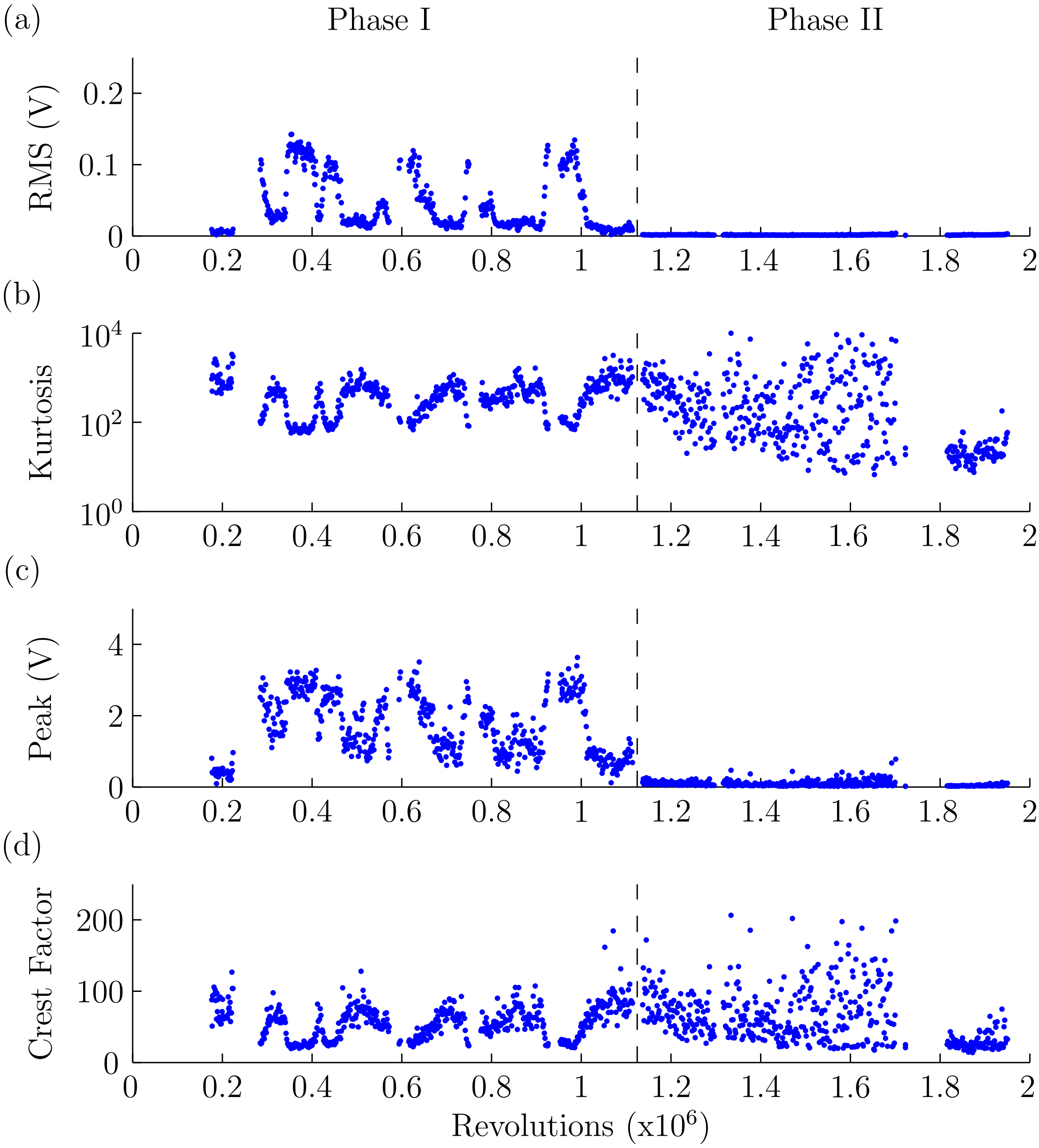

Figure 7.12: Trend of traditional metrics during Bearing 2 test

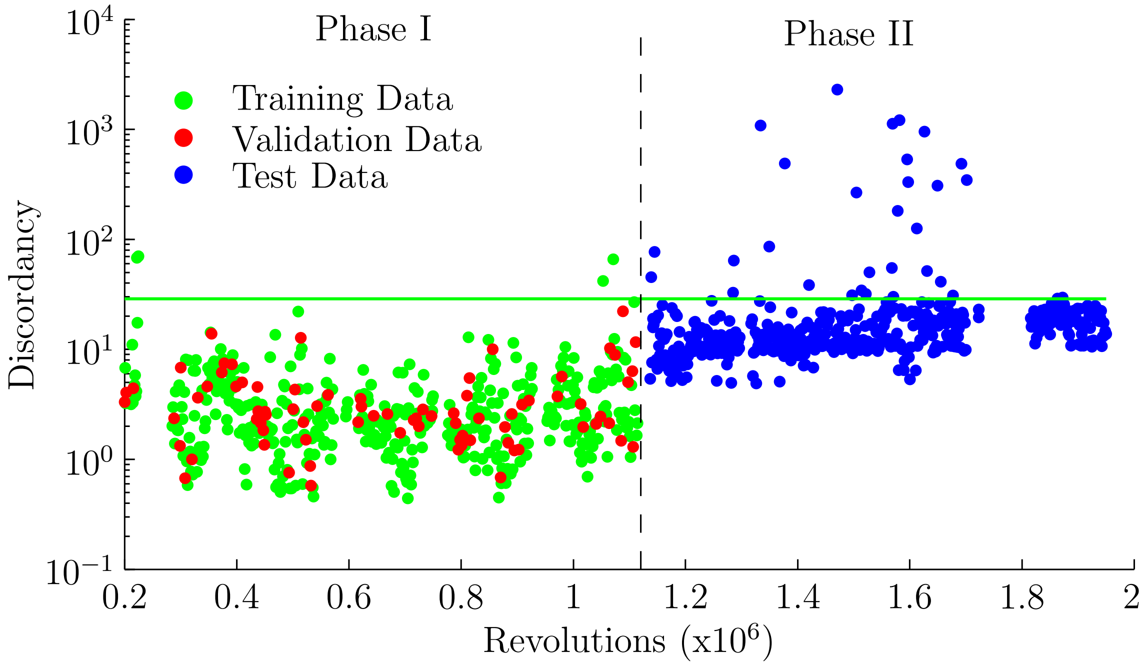

Figure 7.13: Outlier analysis with traditional metrics during Bearing 2 test

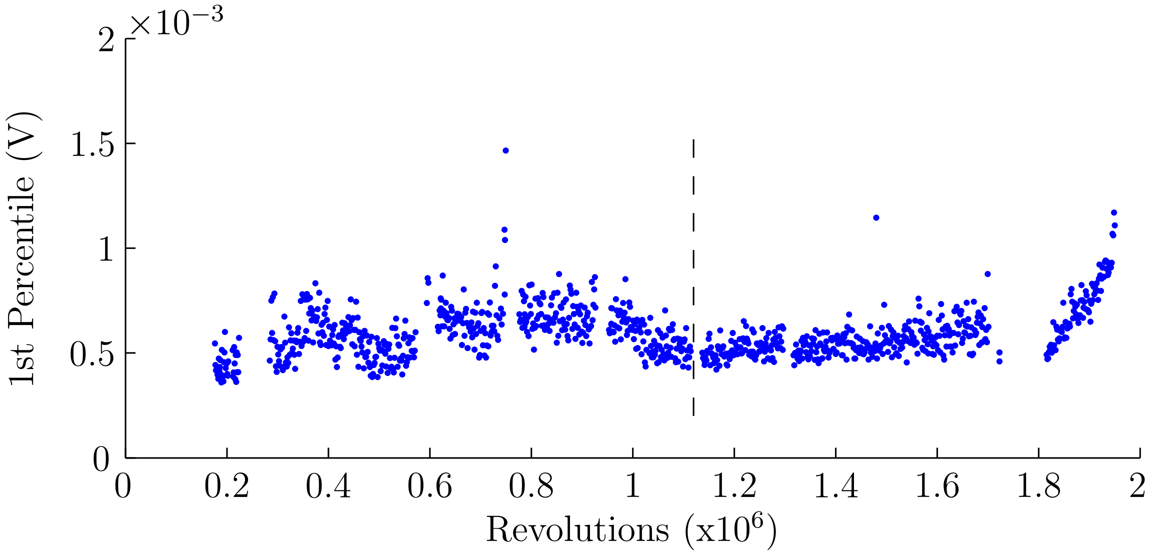

Figure 7.14: First Percentile Value during Bearing 2 Test

acquisition hardware. Prior to installation the gearbox was run on a test bed which allowed the integrity of the sensors to be checked and the noise assessed, providing a more realistic environment compared to the laboratory. The gearbox is shown in Figure 7.15, the low speed shaft driving the epicyclic gear can be seen on the right and the high speed shaft exiting on the left with an input to output ratio of 1:50.6. The gearbox was run under a number of test conditions, initially the speed at the high speed shaft was ramped up in three steps at 200, 1000 and 1500rpm followed by an increase in torque at 0, 50 and 100% of the maximum torque of the test bed which is approximately 30% of the rated torque of the gearbox. This process was repeated four times over a period of almost 5 hours. It is important to assess whether there is a marked increase in noise in the signals when the system is no longer in a laboratory environment.

Figure 7.15: Instrumented V42 Gearbox on Test Bed

The AE RMS for the test schedule as described above is shown in red in Figure 7.16. The RMS demonstrates a marked increase with speed from 200 to 1000rpm and again from 1000 to 1500rpm. This in direct contrast to the situation in which there is no discernible pattern with change in torque. While the bearing is rotating at 200rpm under no load, a number of short duration events are visible. These events correspond to actuations of the MultiLife™ device suggesting the system is susceptible to noise and that care must be taken to ensure signals are only analysed when the device is out of use.

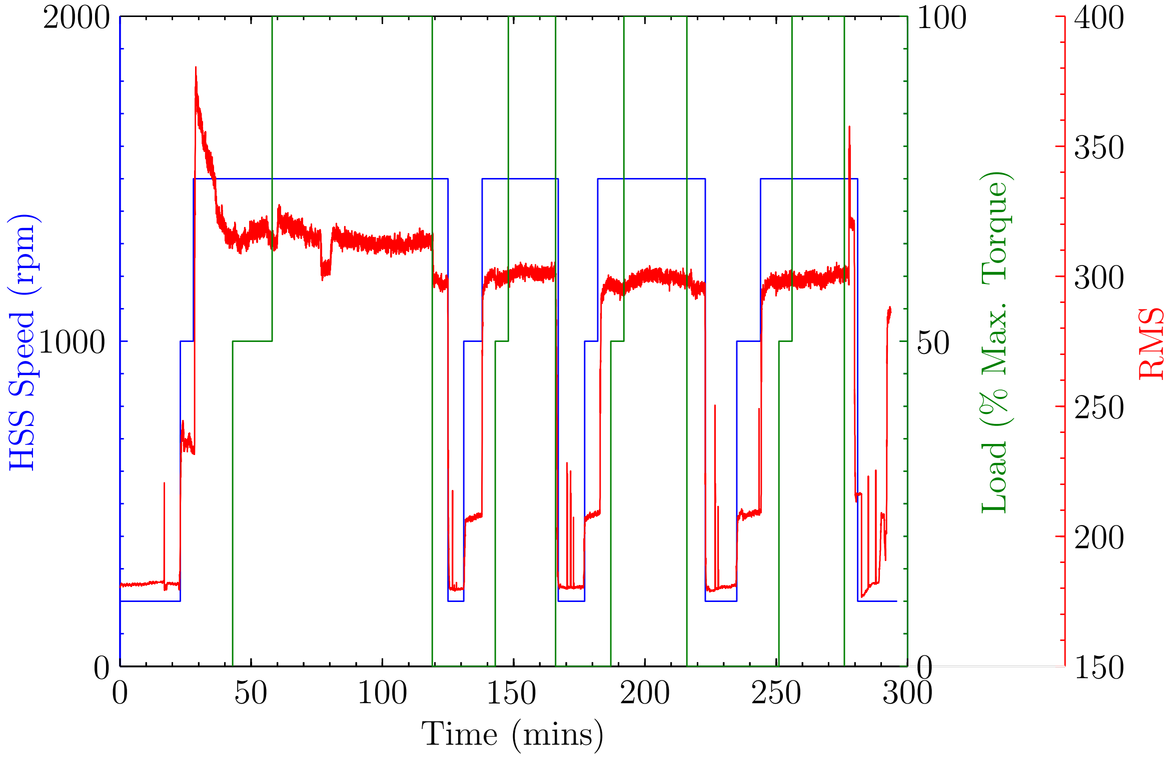

Figure 7.16: AE RMS of HSS at Different Conditions

7.3 Split Race Bearing Rig Results

The Cooper Bearing rig allows tests to be performed on a split bearing. Due to a lack of a control system and limited access to acquisition hardware, the tests carried out on this rig are limited to recording data over a range of operating conditions for the case of an undamaged and damaged bearing. The following section presents the results of such tests.

7.3.1 Typical Data

Data measured from a bearing with and without an artificially seeded defect on the outer race is shown in Figure 7.17. In both instances the bearings were running at 50rpm under a load of 18kN which is approaching the capability of the motor. It can be seen that under such operating conditions asperity contact does not play such an important role as is the case with the MultiLife™ rig due to the absence of random transient bursts. The undamaged case (a) shows a typically Gaussian distribution (kurtosis=2.9) with no distinct features in the data whereas peaks due to roller impingement are clearly visible in the damaged case (b). The measured defect frequency (8.81Hz) closely matches the theoretical defect frequency (8.87Hz) calculated by Equation 3.9.

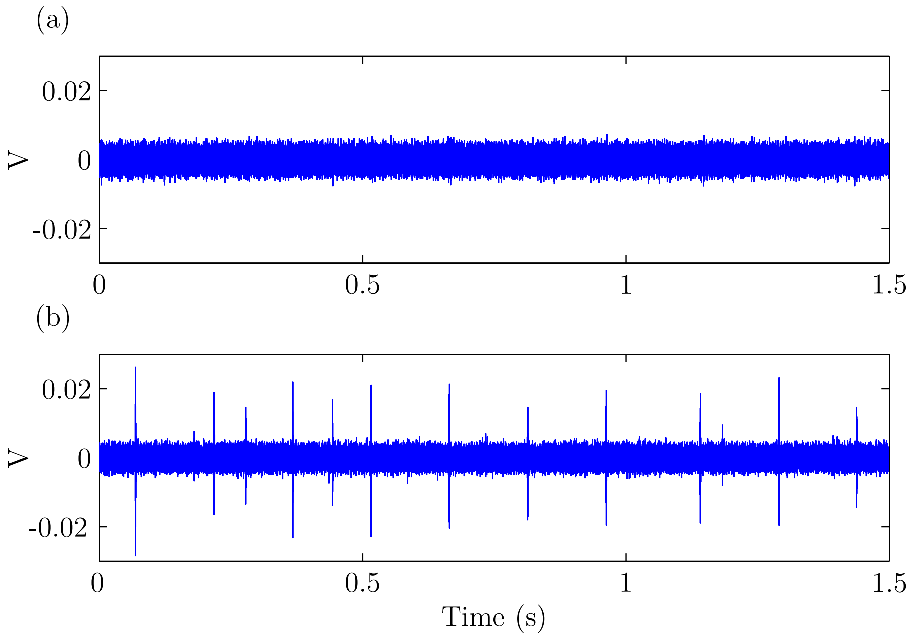

Figure 7.17: Measured data from (a) undamaged and (b) damaged bearing

7.3.2 Operating Conditions

The review of acoustic emission monitoring of rolling element bearings undertaken in Section 3.3 concludes that it has been difficult to establish a clear relationship between operating conditions and the measured signal with some studies reporting seemingly contradictory results. As suggested in previous sections this could be due to differences in lubrication regime where a fully lubricated contact may exhibit different behaviour in response to changes in operating conditions compared to a mixed contact. Therefore, it would be preferable to calculate the film thickness and infer the lubrication regime for the following work but limited access to test rig meant that a thermocouple was not bonded to the bearing and no temperature readings were taken. As a result, the film thickness was not calculated but a direct comparison could be made between an undamaged and a damaged bearing under like for like conditions. The lowest applied load was 3kN and was increased incrementally by 3kN to 18kN whilst for each load case, measurements were taken at 10-60rpm increasing by 10rpm at a time. For each operating condition, five data files of two seconds were recorded, the results of which were then averaged to reduce the effects of noise. Data was recorded from both an undamaged bearing and a bearing seeded with a localised artificial defect. Figures 7.18-7.22 show the effect of load and speed on the kurtosis, RMS, crest factor, K-S statistic and peak-to-peak values respectively.

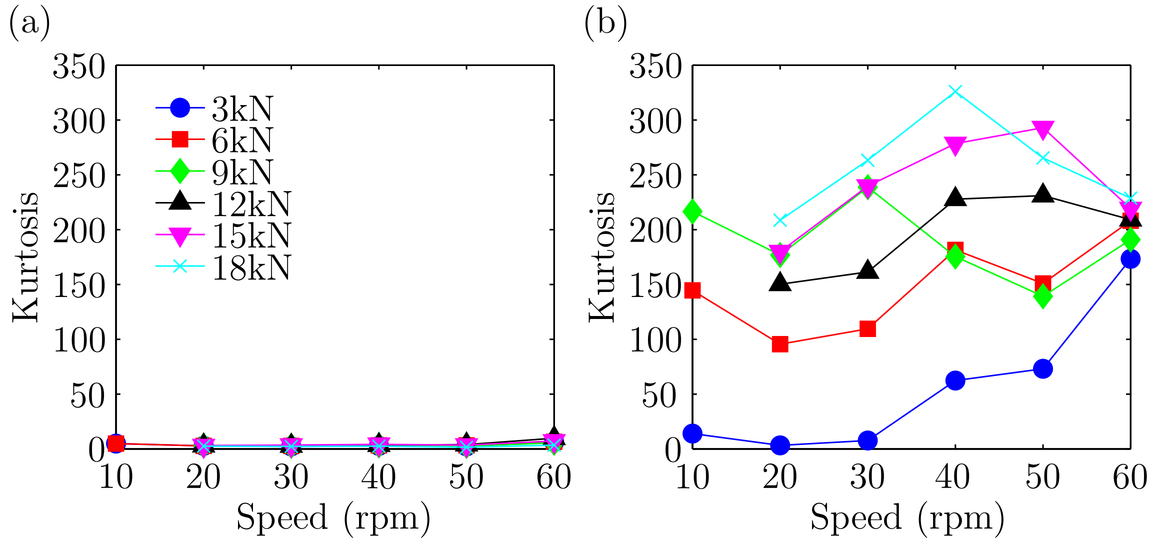

Figure 7.18: Kurtosis of (a) undamaged and (b) damaged bearing

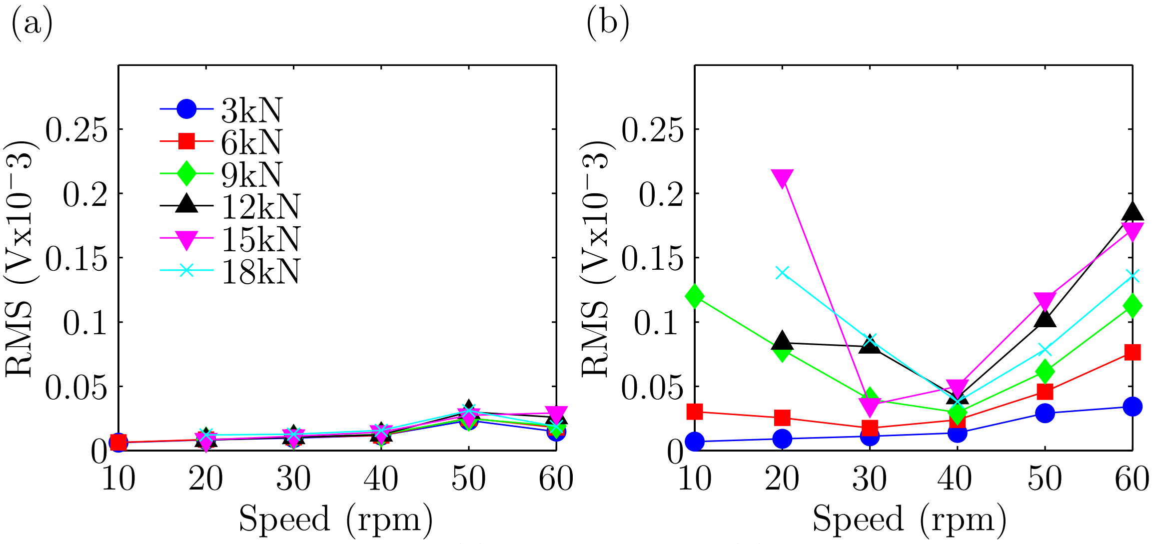

Figure 7.19: RMS of (a) undamaged and (b) damaged bearing

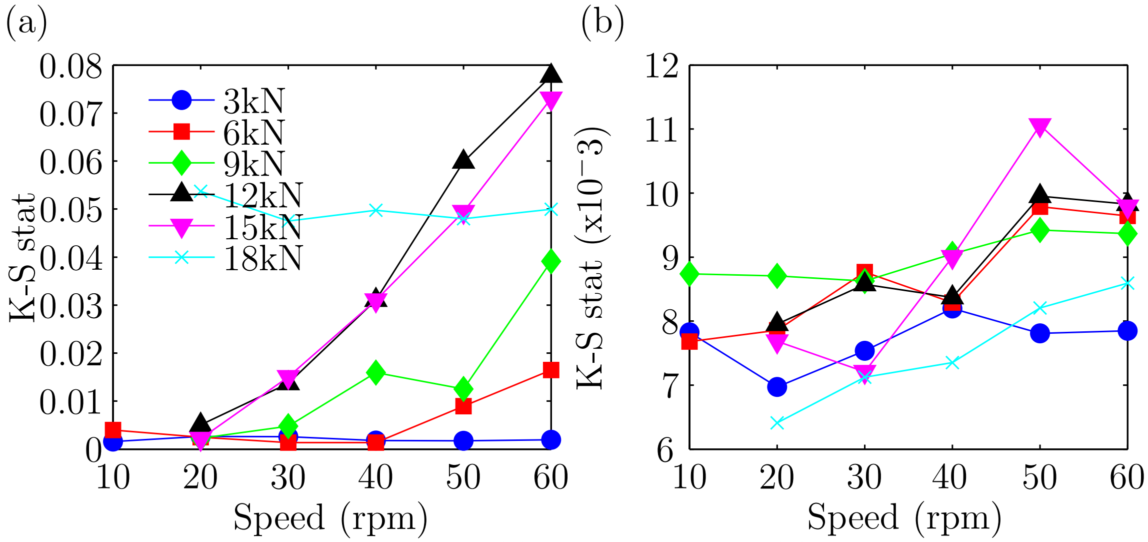

Figure 7.20: KS Statistic of (a) undamaged and (b) damaged bearing

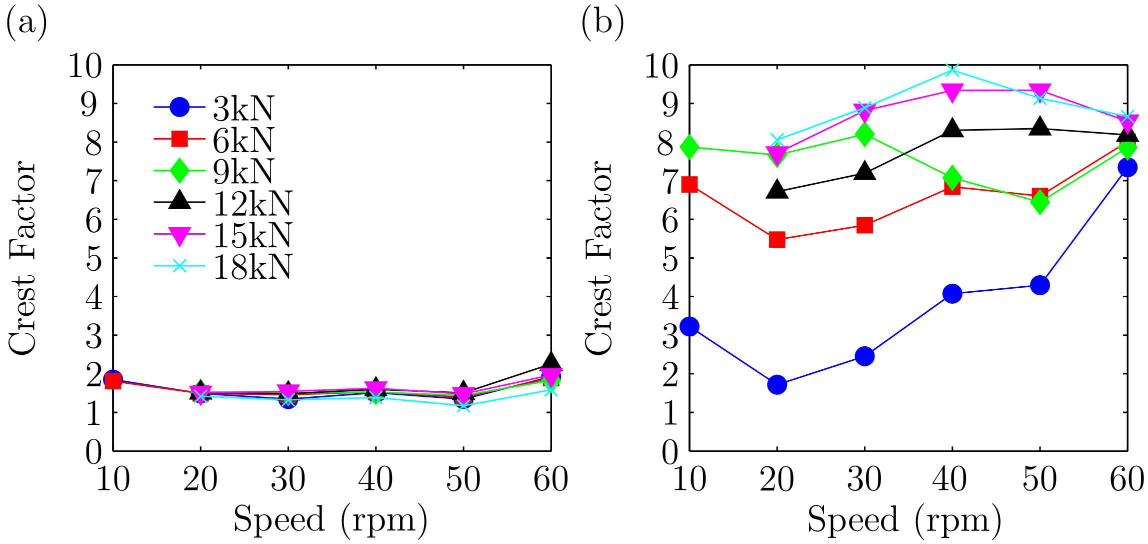

Figure 7.21: Crest Factor of (a) undamaged and (b) damaged bearing

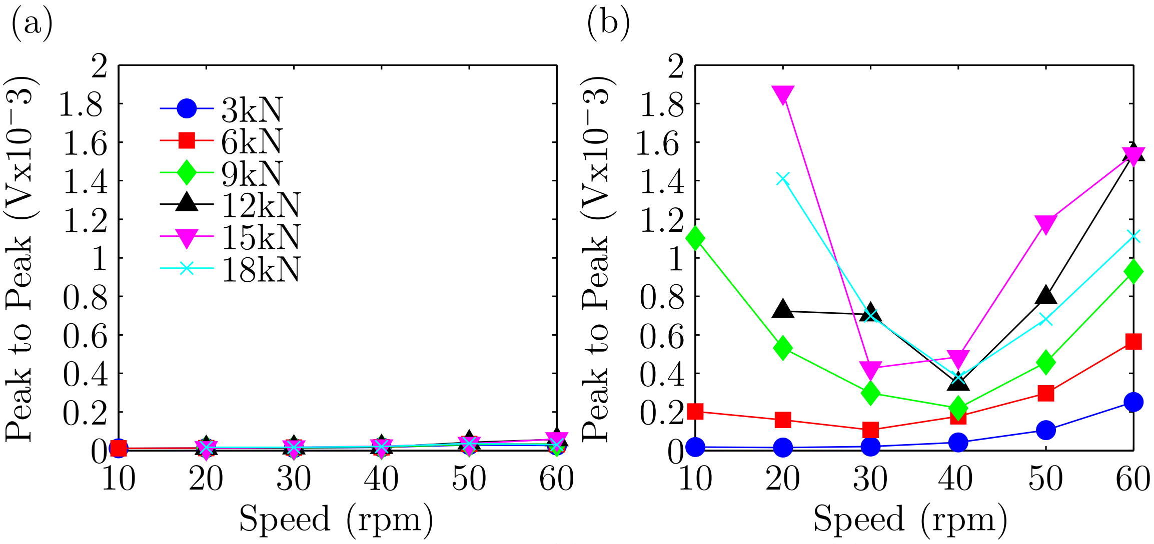

Figure 7.22: Peak to peak value of (a) undamaged and (b) damaged bearing

Since the value of kurtosis increases with sparse, high amplitude peaks, care must be taken when analysing signals from different bearing speeds. Two samples of a given length taken from a bearing with the same defect but run at different speeds will produce different kurtosis values since peaks due to roller impact will occur more often at higher speeds. To counter this effect, the theoretical outer race defect frequency was calculated and the signal was divided in such a way that each segment would contain one peak if present. The final value was then taken to be the median of all the segments. The same approach was applied to all the measures to reduce the effect of spurious noise which was present in some of the data files.

All parameters, with perhaps the exception of the KS statistic, demonstrate a clear increase with the introduction of damage. To quantify the sensitivity of each measure, the percentage increase for each operating condition was calculated and averaged. The results are shown in Table 7.4.

Table 7.4: AE Parameter Sensitivity

Parameter	Average % increase
Kurtosis	5480
RMS	384
KS Statistic	82
Crest Factor	359
Peak to Peak	2517

Clearly, kurtosis and the peak value show the greatest increase indicating such measures are the most suited to determining whether a bearing has suffered localised damage or not. However this experiment fails to highlight the subsequent decrease in kurtosis as damage increases. The sensitivity of the measures are clearly visible at the highest speeds and loads yet while the bearing is running at 10-30rpm at 3kN the presence of damage is much less distinct. In order to attempt to clearly distinguish the damaged bearings a multivariate approach may be taken again using outlier analysis.

Each sample used to generate Figures 7.18-7.22 (5x1 second for each condition) is again described by a thirteen dimension feature vector, each dimension representing the amplitude of the cepstrum at the defect frequency for each reconstructed signal as described in Chapter 5. The covariance matrix is formed from samples taken from the undamaged bearing with 15% of samples reserved for validation. The results are shown in Figure 7.23. The operating conditions for each sample number are shown in Tables 7.5 and 7.6. The samples from the undamaged bearing are on the left side of the plot in a red and green, the samples in red being the training data used to form the covariance matrix while those in green are used to validate the analysis. The samples from the damaged bearing are on the right side in blue. A percentage of correctly classified samples is calculated using the validation set of the undamaged bearings and all the samples from a damaged bearing. In this case only 47% of samples were correctly classified. All the validation observations were correctly classified as undamaged however the analysis failed to classify many of the observations from the damaged bearing with the only success occurring when the bearing speed was greater than 30rpm.

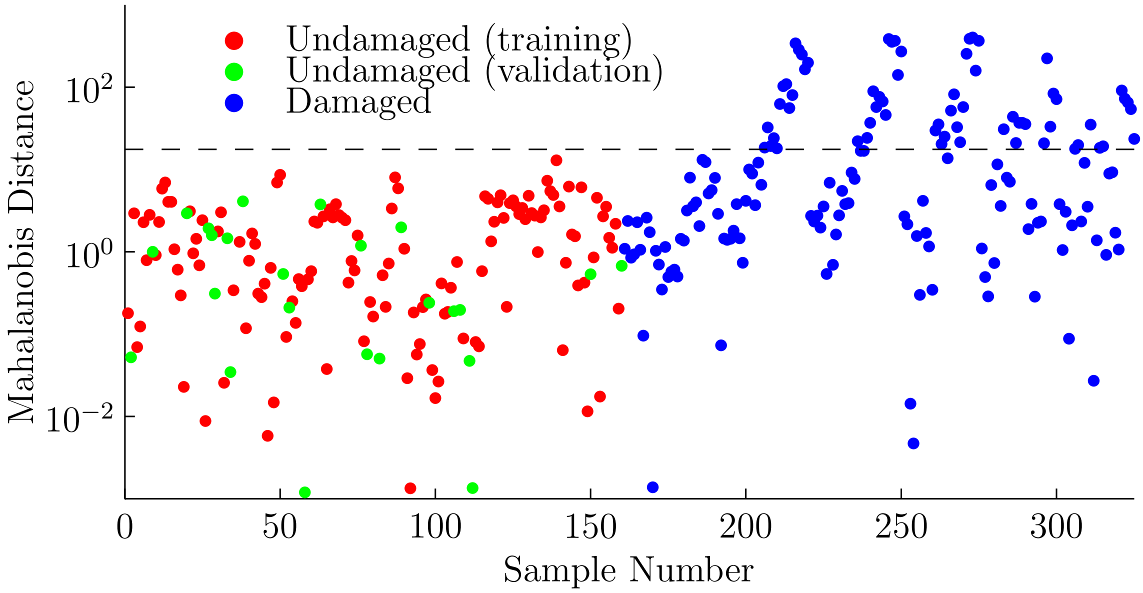

Figure 7.23: Outlier Analysis for Damage Detection of Cooper Bearing.

Due to the full lubricant film and lack of asperity contact, the same process is repeated but rather than use the feature vectors described in Chapter 5, this time they are formed using the measures presented earlier in this section in Figures 7.18-7.22. In this case 95% of samples were correctly classified, with all the incorrectly classified samples falling between 10 and 20rpm at 3kN. At these speeds and loads the impact energy generated during a roller impingement is insufficient to cause a burst of energy that can be distinguished from the background noise and the signal remains a Gaussian distribution. Interestingly once the bearing is operating at loads and speeds above 3kN and 30rpm, the measure of discordancy is relatively stable with respect to operating conditions indicating that within these bounds such technique may be suitable for severity assessment in transient conditions.

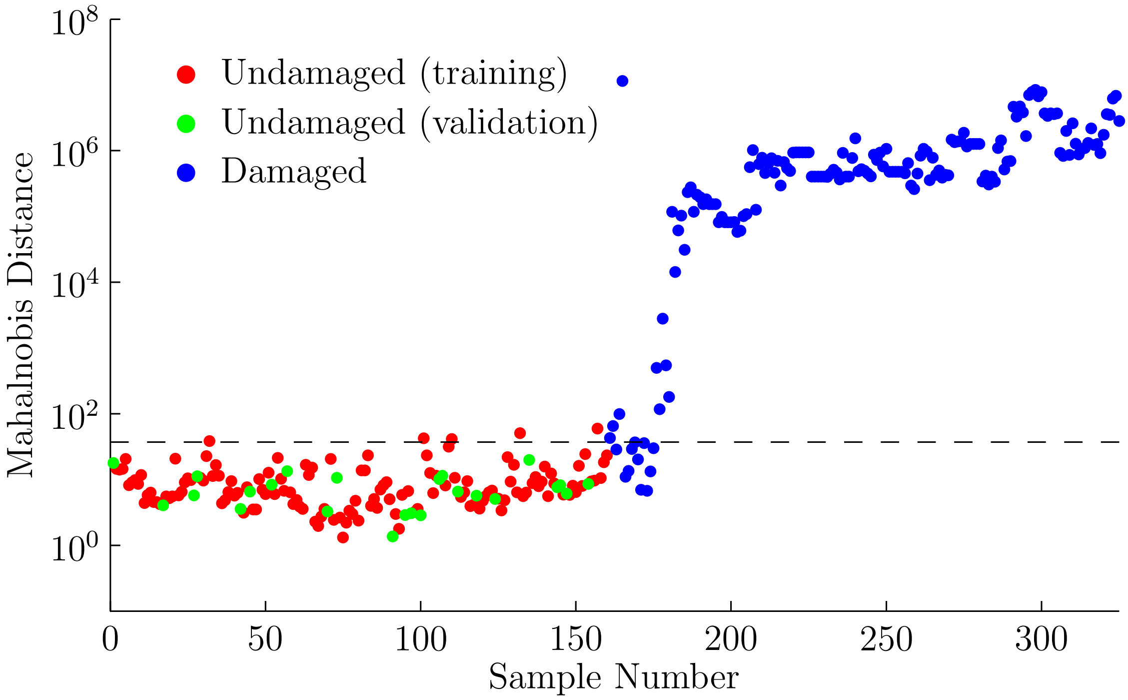

Figure 7.24: Outlier Analysis for Damage Detection of Cooper Bearing.

Table 7.5: Sample Numbers for Undamaged Bearing

	Speed (rpm)					
Load (kN)	10	20	30	40	50	60
3	1-5	6-10	11-15	16-20	21- 25	26-30
6	31-35	36-40	41-45	46-50	51-55	56-60
9	-	61-65	66-70	71-75	76-80	81-85
12	-	86-90	91-95	96-100	101-105	106-110
15	-	111-115	116-120	121-125	126-130	131-135
18	-	136-140	141-145	146-150	151-155	156-160

7.4 Conclusion

The results in this chapter demonstrate the importance of the lubrication regime on bearing monitoring. Section 7.1.1 shows that as film thickness decreases high amplitude events due to asperity contact become prominent reducing the effectiveness of more traditional measures of bearing health such as RMS, kurtosis and crest factor.

Table 7.6: Sample Numbers for Damaged Bearing

	Speed (rpm)					
Load (kN)	10	20	30	40	50	60
3	161-165	166-170	171-175	176-180	181- 185	186-190
6	191-195	196-200	201-205	206-210	211-215	216-220
9	221-225	226-230	231-235	236-240	241-245	246-250
12	-	251-255	256-260	261-265	266-270	271-275
15	-	276-280	281-285	286-290	291-295	296-300
18	-	301-305	306-310	311-315	316-320	321-325

The presence of transient signals due to incipient damage would have little or no effect on such measures as proved by the RMS trend in the run to failure tests presented in Section 7.1.2. The method proposed in Chapter 5, designed to reduce noise and boost the periodicity relating to localised defects, has been shown to be a more effective means of determining the bearing condition and can indicate which component is damaged in the case of a partially lubricated bearing.

This has been compared to a similar approach using traditional bearing measures for a fully lubricated split race bearing. In this case, these traditional measures are more capable of indicating the presence of a defect confirming the importance of understanding the lubrication regime.

Data has also been presented from an instrumented wind turbine gearbox on a factory test bed showing how the energy of the AE signal is highly dependent on the rotational speed but not on the torque the gearbox experiences while it has also been shown that care must be taken that other systems in the vicinity do not corrupt the AE data.

Chapter 8

Conclusion

The use of acoustic emissions as a means of monitoring the condition of the planetary support bearings of wind turbine gearboxes has been investigated. Despite best design practice being followed, high failure rates and the resulting loss of production during maintenance have lead to engineers seeking novel solutions to the problem. As a result, Ricardo have patented the Multilife™ bearing, based on the premise that bearing life can be increased by periodically rotating the inner raceway and thus distributing damage. The efficient use of which requires knowledge about the bearing condition and the position of any damage if present.

Previous studies have shown that in certain cases acoustic emission technology can provide an earlier warning to the onset of catastrophic failure compared to vibration analysis. A review of the literature suggests that this advantage is most distinct when monitoring low speed bearings, which in general refers to rotational speeds of 100rpm and less; speeds typically found in the planetary support bearings in question.

The work presented here has attempted to derive a suitable means of detecting incipient damage and if present, estimating its severity and location. The data has been recorded from two test rigs; the first, a bespoke full scale test rig designed to replicate the damage seen on planetary support bearings in wind turbine gearboxes, the second supplied by Cooper Bearings designed to run smaller split type bearings. The following sections highlights the main findings.

8.1 Defect Detection

In order to encourage the initiation and propagation of damage on the MultiLife™ bearings, conditions were chosen as to reduce the expected bearing life as much as practically possible. A high radial load, low speed and low viscosity lubricant meant the oil film was thinner than the composite roughness of the raceway and rolling element surfaces resulting in a mixed lubrication regime characterised by asperity contact. Until now, the monitoring of acoustic emissions in bearings under such a regime had not been investigated. Analysing data from the Multilife™ bearing at different conditions clearly shows a distinct change in the signature with the onset of asperity contact with RMS, kurtosis and peak values all rising sharply. This suggests such measures are not suitable if a mixed lubrication regime is expected and different approach is required.

8.1.1 Feature Extraction

To reduce the dimensionality of the acoustic emission data, a process, described in Chapter 5, was proposed to highlight any periodicity in the signal which occurs for localised defects. Wavelet packet decomposition breaks a signal down into a set of groups of coefficients which represent the original signal at different scales (or frequencies). Reconstructing individual groups of coefficients acts as a filter bank whereby the presence of a defect will be more prominent at some scales than others. At this point the signals need to be enveloped since the frequency content of the defect signal is modulated at the characteristic defect frequency of the damaged component, this was achieved using the short time energy function. Finally, the autocorrelation function, the cross correlation of a signal with itself, was employed to boost the periodicity and reduce noise between the events due to roller impact. A measure of the periodicity was provided by the cepstrum and a feature vector was formed of the amplitude of the quefrency relating to the defect frequency for each component.

8.1.2 Run-to-Failure Tests

Run-to-failure bearing tests on the Multilife™ rig showed that this process, in conjunction with outlier analysis, was possible to detect the presence of inner race defects measuring 100μm wide by 60μm deep by calculating the Mahalanobis distance from the normal condition. Smaller defects may have been detectable but were not investigated. The results were compared with the RMS trend which only increased once the seeded defect had propagated into a large spall or the general condition of the bearing had deteriorated due to debris denting the raceways and rollers. The first of the two tests showed that the Mahalanobis distance decreased once the defected raceway was rotated away from the loaded zone however the average value of discordancy was still greater than that of the undamaged bearing. The Monte Carlo method was used to determine a suitable damage threshold and provided a reasonable means of determining the condition of the bearing. However in this field, a false positive would prove more costly than failing to detect the smallest defects. It may therefore be preferable to use a higher threshold for future work.

The second run-to-failure test was stopped prior to the inner race failing due to significant damage to the outer raceway, presumably caused by edge loading, in the form of a single large localised spall in addition to a number of smaller distributed spalls around the raceway circumference. The outlier analysis failed to detect this damage. It is believed that this is due to a combination of high attenuation since the signal needed to propagate across three component interfaces and high levels of background noise due to the mixed lubrication regime. Previous studies have demonstrated the possibility of detecting damage to the non-stationary raceway [75] however the higher rotational speeds result in a greater energy loss and thicker film thickness reducing the occurrence of asperity contact. In this case, using the first percentile as a measure of background noise indicated the deteriorating condition of the bearing as damage occurred to the outer race.

The same processing techniques were employed on the split raceway bearing test rig to determine the presence of a defect. In this instance, the raw data suggested that a full lubricant film was formed and the undamaged bearing followed a Gaussian distribution. In this case, the technique was only successful at determining the state of the bearing at speeds greater than 30rpm. A more successful attempt was made when a feature vector was formed using the statistical parameters in the time domain. The lack of asperity contact during normal operation meant that any transient signals were the result of roller impingement on a defect. Therefore increases in these measures, which are heavily dependent on the peak values, reflected the bearing condition. The results from both test rigs highlights the importance of understanding the expected lubrication regime prior analysing acoustic emission data.

8.2 Source Location

Previously the position of the source of acoustic emissions from bearings had been investigated using time of flight methods. In this work, three methods of defect location on rolling element bearings employing the transient features of an AE signal have been investigated. A series of artificially generated sources were created on a bearing inner race and the errors were compared. Initially, the positions were estimated analytically based upon geometry and assumptions on the wave propagation which provided a reasonable degree of accuracy comparable to previous studies. Further improvements could be achieved using Delta T mapping or neural networks which could reduce the average error from 42mm to 29mm or 17mm respectively. The time and effort required to obtain the training data for these two methods will determine which method is likely to be used in practice.

The same methods were investigated whilst the bearing was in operation following the introduction of a seeded defect at a known location. The accuracy of the methods were similar to those when performed with artificial sources, with the average error for TOA method being slightly less at 40mm while the errors of Delta T mapping and neural networks were 25mm and 22mm respectively.

The choice of source location method will be dependent on the application in question. In those cases where accuracy is paramount neural networks may offer the best solution however the process of obtaining training data may be time consuming or indeed impossible if access is restricted. The geometry of a bearing is relatively simple however the difference in accuracy between the analytical and data driven methods is likely to increase with geometric complexity. This is demonstrated using a plate with a number of holes of different sizes by Hensman et al [43].

8.3 Future Work

Whilst seeded surface defects, such as those created by EDM, allow the duration of accelerated bearing life tests to be significantly reduced and provide a means for researchers to assess the use of defect detection algorithms, the resulting damage is not representative of the mechanisms by which the failure tends to occur. Field studies have suggested that a large proportion of planetary bearing damage is initiated via overload events which are likely to occur during emergency shutdowns and/or turbulent wind conditions. Further experiments will be conducted on bearings that have been subjected to loading conditions to replicate such damage. A modification to an impact rig situated in Cranfield University has been designed by engineers at Ricardo whereby a roller is dropped onto the raceway surface so that the subsurface shear stress exceeds the elastic limit of the material but the surface remains undamaged. A colleague is in the process of cutting up a bearing subjected to such loading to investigate whether any subsurface cracks have appeared. If this approach has not been successful, a different approach may be required whereby a roller is repeated pressed against the raceway to initiate fatigue damage

After the problems with misalignment during the second test causing the outer race to fail, the self-aligning bearing has been replaced by a parallel bearing to reduce chance of edge loading and therefore outer race fatigue damage. The new bearing is shown in Figure 8.1 before being reassembled with a new test bearing. It is hoped that the resulting damage will be more representative of the field conditions with the inner race failing in the loaded zone.

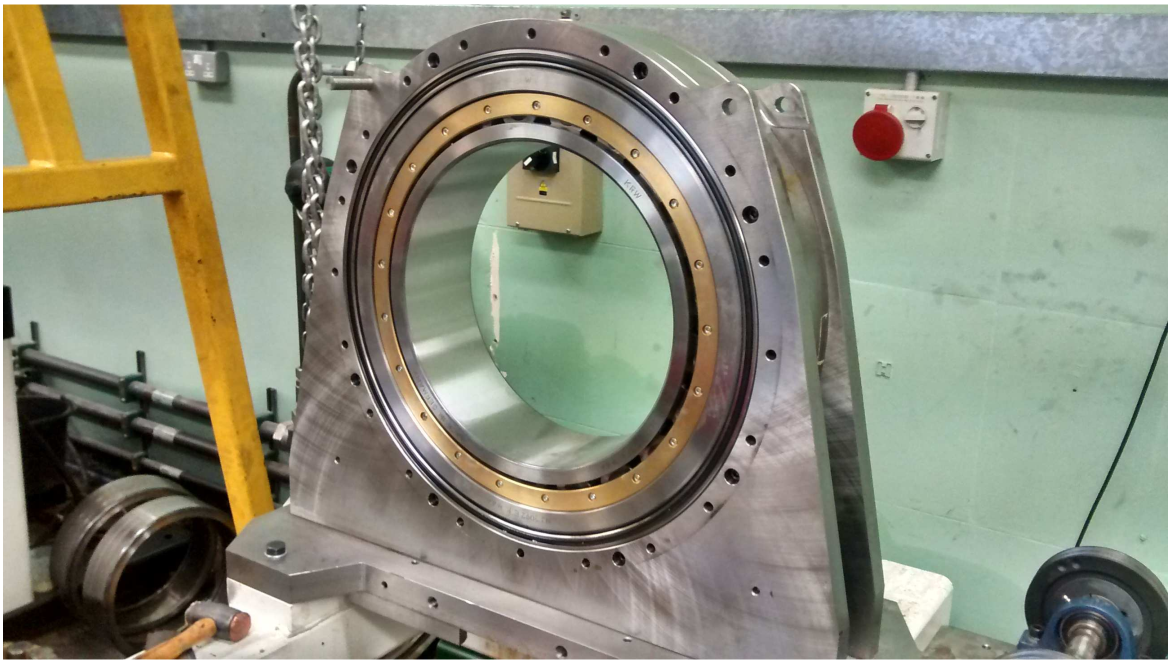

Figure 8.1: New Parallel Support Bearing for MultiLife™ Rig

The first run-to-failure bearing test presented in Section 7.1.2 proved that bearing life can be increased by rotating the damage 180° away from the most heavily loaded point and continuing to operate the bearing. This does not however provide any indication of the possible percentage increase in bearing life that can be expected with the introduction of the MultiLife™ actuators. Further experiments are required whereby an indexing strategy similar to those suggested by Figure 1.11 is employed. A larger data set will also allow data driven models to be designed that will enable the prediction of remaining useful life of the bearing. It would also be preferable to rotate the inner raceway using the MultiLife™ device to test the mechanical reliability in a laboratory environment and to reduce the labour required to regularly disassemble and reassemble the test rig, which could be the cause of noise in some of the experiments.

Finally, since the noise levels are at present unknown and potentially variable in an operating wind turbine, the use of a reference sensor as employed by Miettinen and Leinonen [74] means that the signal could be corrected for noise. Such a process would be relatively simple to perform with certain measures such as RMS or ring down counts and could be initially validated in the laboratory.

Appendix A

k-means clustering

k-means clustering is a simple, efficient algorithm designed to allocate n observations ($\mathbf{x}_1$, $\mathbf{x}_2$, ...,$\mathbf{x}_n$) to one of k clusters defined by its centroid where k is defined before the algorithm starts. The aim of the algorithm is to minimise the objective function J given:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^j - c_j\|^2 \tag{1}$$

The iterative process of the algorithm is described below.

- The number of k clusters is first chosen.
- Each cluster is assigned a centroid at random.
- Each observation is assigned to the cluster whose centroid is closest.
- The position of each centroid is recalculated as the average of the observations assigned to it.
- The previous two steps are repeated until no further re-assignments take place.

It has been shown that this algorithm will always converge yet the solution will not necessarily be the optimum.

Appendix B

Oil Film Thickness Calculations

Equations 2-4 give the dimensionless speed, material and load parameters required to calculate the minimum film thickness of a line contact [110]

$$U = \frac{\mu u}{E^* R} \tag{2}$$

$$G = \alpha E^* \tag{3}$$

$$W = \frac{P}{E^* R L} \tag{4}$$

where U, G and W are the dimensionless speed, material and load parameters respectively. The conjunction radius R formed between the surfaces in contact is given by Equation 5 where R_1 is the radius of the rolling element and R_2 is the radius of the rolling surface of the inner raceway.

$$R = \frac{R_1 R_2}{R_1 + R_2} \tag{5}$$

The entraining velocity u is the speed the element passes over the race. If pure rolling contact is assumed it is given by Equation 6.

$$u = \frac{\pi R_2 n}{30} \tag{6}$$

The equivalent elasticity modulus E^* for similar materials of elastic modulus E and Poisson's ratio v is given by Equation 7.

$$E^* = \frac{E}{1 - v^2} \tag{7}$$

References

[1] Global Wind Energy Council. Global wind report - annual market update. Technical report, GWEC, 2011.

[2] ExxonMobil. The outlook for energy: A view to 2040. Technical report, ExxonMobil, 2012.

[3] U.S. Energy Information Administration. Annual energy outlook 2012 with projections to 2035. Technical report, EIA, 2012.

[4] S. Sorrel, J. Speirs, R. Bentley, A. Brandt, and R. Miller. Global oil depletion: A review of the evidence. *Energy policy*, 38:5290–5295, 2010.

[5] M. A. Adelman. Comment on: R. W. Bentley,“Global oil & gas depletion”, Energy Policy 30 (2002) 189-205. *Energy Policy*, 31(4):389–390, 2004.

[6] P. R. Odell. *Why carbon fuels will dominate the 21st century's global energy economy.* Multi-Science Pub Co, Brentwood, England, 2004.

[7] S. H. Mohr and G. M. Evans. Long term forecasting of natural gas production. *Energy policy*, 39:5550–5560, 2011.

[8] EU directive 2009/28/EC, 2009.

[9] Department of Energy and Climate Change. Digest of United Kingdom Energy Statistics. Technical report, London, 2011.

[10] A. Kent and D. Mercer. Australia's mandatory renewable energy target (MRET): an assessment. *Energy Policy*, 34(9):1046–1062, 2006.

[11] K. Nishio and H. Asano. The amount of renewable energy and additional costs under the renewable portfolio standards in Japan. Technical report, Central Research Institute of Electric Power Industry, 2003.

[12] R. Wiser, K. Porter, and R. Grace. Evaluating experience with renewables portfolio standards in the United States. *Mitigation and Adaptation Strategies for Global Change*, 10(2):237–263, April 2005.

[13] R. Wiser and M. Bolinger. 2010 wind technologies market report. Technical report, U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, 2011.

[14] P. Nielsen, J. Lemming, P. Morthorst, N. Clausen, H. Lawetz, H. Lindboe, E. James-Smith, N. Bang, S. Strom, and J. Larsen. Vindmollers okonomi (the economy of wind power). Technical report, EMD International, 2011.

[15] Danish Energy Agency. Wind power in Denmark: Technologies, policies and results, 1999.

[16] M. Bolinger and R. Wiser. Understanding trends in wind turbine prices over the past decade. Technical report, Lawrence Berkeley National Laboratory, 2011.

[17] E. Lantz, M. Hand, and M. Wiser. The past and future cost of wind energy. In *2012 World Renewable Energy Forum*, 2012.

[18] Parsons Brinckerhoff. Powering the nation update, 2010.

[19] U.S. Energy Information Administration. Levelized cost of new generation resources, 2012.

[20] Scottish Enterprise. A guide to offshore wind, 2011.

[21] C. Kost, J. M. Mayer, J. Thomsen, N. Hartmann, C. Senkpiel, S. Philipps, S. Nold, S. Lude, N. Saad, and T. Schlegl. Levelized cost of electricity renewable energy technologies. Technical report, Fraunhofer Institute for Solar Energy Systems ISE, 2013.

[22] E. Echavarria, B. Hahn, G. J. W. Van Bussel, and T. Tomiyama. Reliability of wind turbine technology through time. *Journal of solar energy engineering*, 130(3):31005, 2008.

[23] F. Spinato, PJ Tavner, G. J. W. Van Bussel, and E. Koutoulakos. Reliability of wind turbine subassemblies. *Renewable Power Generation, IET*, 3(4):387–401, 2009.

[24] J. Ribrant and L. Bertling. Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005. In *Power Engineering Society General Meeting, 2007. IEEE*, pages 1–8. IEEE, 2007.

[25] GCube. GCube Top 5 US Wind Energy Insurance Claims Report. Technical report, 2013.

[26] W. Musial, S. Butterfield, and B. McNiff. Improving wind turbine gearbox reliability. In *Proceedings of the European Wind Energy Conference*, 2007.

[27] A. Ragheb and M. Ragheb. Wind turbine gearbox technologies. In *Nuclear & Renewable Energy Conference (INREC), 2010 1st International*, pages 1–8. IEEE, 2010.

[28] R. Poore and T. Lettenmaier. Alternative design study report: Windpact advanced wind turbine drive train designs study; November 1, 2000-February 28, 2002. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2003.

[29] J. C. Wheals, P. Guern, R. S. Dwyer-Joyce, M. B. Marshall, and T. Howard. Ricardo MultiLife™bearing programme for increased reliability of offshore wind turbines. In *European Wind Energy Conference*, 2011.

[30] ASTM E1067 / E1067M 11. Standard practice for acoustic emission examination of fiberglass reinforced plastic resin (FRP) tanks/vessels.

[31] R. K. Miller. *Acoustic Emission Testing*, volume 5 of *Nondestructive Testing Handbook*. American Society of Nondestructive Testing, 2nd edition, 1987.

[32] T. F. Drouillard. A history of acoustic emission. *Journal of acoustic emission*, 14(1):1–34, 1996.

[33] J. Kaiser. *Untersuchungen uber das auftreten Geraushen beim Zugversuch (An Investigation in to the Occurence of Noises in Tensile Tests or a Study of Acoustic Phenomena in Tensile Tests)*. PhD thesis, Technische Hochschule, Munich, 1950.

[34] H. L. Dunegan, C. A. Tatro, and D. O. Harris. Acoustic emission research. status report, December 1963–August 1964. Technical report, Lawrence Radiation Lab., Univ. of California, Livermore, 1964.

[35] A. T. Green, C. S. Lockman, and R. K. Steele. Acoustic verification of structural integrity of polaris chambers. *Modern Plastics*, 41(11):137–139, 1964.

[36] W. P. Mason, H. J. McSkimin, and W. Shockley. Ultrasonic observation of twinning in tin. *Physical Review*, 73(10):1213, 1948.

[37] Meysam Akbari and Mehdi Ahmadi. The application of acoustic emission technique to plastic deformation of low carbon steel. *Physics Procedia*, 3(1):795–801, 2010.

[38] N.N. Hsu. Acoustic emissions simulator, April 19 1977. US Patent 4,018,084.

[39] A. Nielsen. Acoustic emission source based on pencil lead breaking. *The Danish Welding Institute Publication*, 80:15, 1980.

[40] J. J. Hensman. *Novel Techniques for Acoustic Emission Monitoring of Fatigue Fractures in Landing Gear.* PhD thesis, University of Sheffield, 2009.

[41] J. F. Chretien and N. Chretien. A bibliographical survey of acoustic emission. *Non-Destructive Testing*, 5(4):220–224, 1972.

[42] D. Mba. Acoustic emissions and monitoring bearing health. *Tribology transactions*, 46(3):447–451, 2003.

[43] J. Hensman, R. Mills, S. G. Pierce, K. Worden, and M. Eaton. Locating acoustic emission sources in complex structures using Gaussian processes. *Mechanical Systems and Signal Processing*, 24(1):211–223, 2010.

[44] G. Drummond, J. F. Watson, and P. P. Acarnley. Acoustic emission from wire ropes during proof load and fatigue testing. *NDT & E International*, 40(1):94–101, 2007.

[45] M. Huang, L. Jiang, P.K. Liaw, C.R. Brooks, R. Seeley, and D.L. Klarstrom. Using acoustic emission in fatigue and fracture materials research. *JOM*, 50(11):1–14, 1998.

[46] NDT Resource Centre. Theory - acoustic waves. Last seen: 9 May 2015.

[47] A. K. Maji, D. Satpathi, and T. Kratochvil. Acoustic emission source location using lamb wave modes. *Journal of engineering mechanics*, 123(2):154–161, 1997.

[48] A. A. Pollock. Acoustic emission inspection. Technical report, Physical Acoustics Corporation, 2003.

[49] Y. Zhao. The importance of lubricant and fluid analysis in predictive maintenance. Spectro Scientific, 2014.

[50] Mohamed El Hachemi Benbouzid. A review of induction motors signature analysis as a medium for faults detection. *Industrial Electronics, IEEE Transactions on*, 47(5):984–993, 2000.

[51] S. Sheng. *Wind turbine micropitting workshop: a recap.* National Renewable Energy Laboratory, 2010.

[52] K. Stadler and A. Stubenrauch. Premature bearing failures in industrial gearboxes. In *Antriebstechnisches Kolloquium (ATK)*, pages 19–20, 2013.

[53] M. H. Evans. White structure flaking (wsf) in wind turbine gearbox bearings: effects of 'butterflies' and white etching cracks (wecs). *Materials Science and Technology*, 28(1):3–22, 2012.

[54] M. S. Patil, J. Mathew, and P. K. Rajendra Kumar. Bearing signature analysis as a medium for fault detection: A review. *Journal of Tribology*, 130(1), 2008.

[55] B. Tao, L. Zhu, H. Ding, and Y. Xiong. An alternative time-domain index for condition monitoring of rolling element bearingsa comparison study. *Reliability Engineering & System Safety*, 92(5):660–670, 2007.

[56] J. Shiroishi, Y. Li, S. Liang, T. Kurfess, and S. Danyluk. Bearing condition diagnostics via vibration and acoustic emission measurements. *Mechanical systems and signal processing*, 11(5):693–705, 1997.

[57] R. B. Randall and J. Antoni. Rolling element bearing diagnostics - a tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011.

[58] R. M. Stewart. Application of signal processing techniques to machinery health monitoring. Noise and Vibration, edited by White, RG, and Walker, JG, John Wiley & Sons, New York, 1982.

[59] P. D. McFadden and J. D. Smith. Vibration monitoring of rolling element bearings by the high-frequency resonance techniquea review. *Tribology international*, 17(1):3–10, 1984.

[60] Y.R. Hwang, K.K. Jen, and Y.T. Shen. Application of cepstrum and neural network to bearing fault detection. *Journal of mechanical science and technology*, 23(10):2730–2737, 2009.

[61] B. P. Bogert, M. J. R. Healy, and J. W. Tukey. The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the symposium on time series analysis*, volume 15, pages 209–243. chapter, 1963.

[62] C.S. Burrus, R.A. Gopinath, and H. Guo. *Introduction to wavelets and wavelet transforms: a primer*, volume 23. Prentice Hall Upper Saddle River NJ, 1998.

[63] J. Lin and L. Qu. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration*, 234(1):135–148, 2000.

[64] H. Qiu, J. Lee, J. Lin, and G. Yu. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, 289(4):1066–1090, 2006.

[65] D. L. Donoho. De-noising by soft-thresholding. *Information Theory, IEEE Transactions on*, 41(3):613–627, 1995.

[66] S. Prabhakar, A. R. Mohanty, and A. S. Sekhar. Application of discrete wavelet transform for detection of ball bearing race faults. *Tribology International*, 35(12):793–800, 2002.

[67] N. G. Nikolaou and I. A. Antoniadis. Rolling element bearing fault diagnosis using wavelet packets. *Ndt & E International*, 35(3):197–205, 2002.

[68] L. M. Rogers. The application of vibration signature analysis and acoustic emission source location to on-line condition monitoring of anti-friction bearings. *Tribology international*, 12(2):51–58, 1979.

[69] J. D. Smith. Vibration monitoring of bearings at low speeds. *Tribology International*, 15(3):139–144, 1982.

[70] M. W. Hawman and W. S. Galinaitis. Acoustic emission monitoring of rolling element bearings. In *Ultrasonics Symposium, 1988. Proceedings., IEEE 1988*, pages 885–889. IEEE, 1988.

[71] C. James Li and S. Y. Li. Acoustic emission analysis for bearing condition monitoring. *Wear*, 185(1):67–74, 1995.

[72] J. Miettinen and P. Andersson. Methods to monitor the running situation of grease lubricated rolling bearings. In *VTT Symposium*, volume 180, pages 92–101. Valtion Teknillinen Tutkimuskeskus, 1998.

[73] J. Miettinen and P. Andersson. Acoustic emission of rolling bearings lubricated with contaminated grease. *Tribology International*, 33(11):777–787, 2000.

[74] J. Miettinen and P. Leinonen. Monitoring of contaminants in a grease lubricated rolling bearing by acoustic emission in field environment. In *Proceedings of the 2nd COST*, volume 516, pages 243–252, 1999.

[75] A. Choudhury and N. Tandon. Application of acoustic emission technique for the detection of defects in rolling element bearings. *Tribology International*, 33(1):39–45, 2000.

[76] N. Jamaludin, D. Mba, and R. H. Bannister. Condition monitoring of slow-speed rolling element bearings using stress waves. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 215(4):245–271, 2001.

[77] A. Morhain and D. Mba. Bearing defect diagnosis and acoustic emission. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 217(4):257–272, 2003.

[78] A.M. Al-Ghamd and D. Mba. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing*, 20(7):1537–1571, 2006.

[79] N. Tandon, G. S. Yadava, and K. M. Ramakrishna. A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings. *Mechanical systems and signal processing*, 21(1):244–256, 2007.

[80] J. Couturier and D. Mba. Operational bearing parameters and acoustic emission generation. *Journal of Vibration and Acoustics*, 130(2):024502, 2008.

[81] M. Elforjani and D. Mba. Observations and location of acoustic emissions for a naturally degrading rolling element thrust bearing. *Journal of Failure Analysis and Prevention*, 8(4):370–385, 2008.

[82] S. Al-Dossary, R. I. Hamzah, and D. Mba. Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing. *Applied Acoustics*, 70(1):58–81, 2009.

[83] Y. He, X. Zhang, and M.I. Friswell. Defect diagnosis for rolling element bearings using acoustic emission. *Journal of vibration and acoustics*, 131(6):61012, 2009.

[84] A. Widodo, E. Y. Kim, J. D. Son, B. S. Yang, A. C. C. Tan, D. S. Gu, B. K. Choi, and J. Mathew. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Systems with Applications*, 36(3):7252–7261, 2009.

[85] K.R. Al-Balushi, A. Addali, B. Charnley, and D. Mba. Energy index technique for detection of acoustic emissions associated with incipient bearing failures. *Applied Acoustics*, 71(9):812–821, 2010.

[86] M. Elforjani and D. Mba. Accelerated natural fault diagnosis in slow speed bearings with acoustic emission. *Engineering Fracture Mechanics*, 77(1):112–127, 2010.

[87] K. Kuboyama. Development of low speed bearing diagnosis technique. *Detection, Diagnosis and Prognosis of Rotating Machinery to Improve Reliability, Maintainability, and Readiness Through the Application of New and Innovative Techniques*, pages 178–185, 1986.

[88] T. A. Harris. *Rolling bearing analysis.* J Wiley, New York, 2001.

[89] R. B. Randall, J. Antoni, and S. Chobsaard. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mechanical Systems and Signal Processing*, 15(5):945–962, 2001.

[90] A. Rytter. *Vibrational based inspection of civil engineering structures.* PhD thesis, unknown, 1993.

[91] K. Worden, G. Manson, and N. R. J. Fieller. Damage detection using outlier analysis. *Journal of Sound and Vibration*, 229(3):647–667, 2000.

[92] M. G. Baxter, R. Pullin, K. M. Holford, and S. L. Evans. Delta T source location for acoustic emission. *Mechanical systems and signal processing*, 21(3):1512–1520, 2007.

[93] I. Grabec. Application of correlation techniques for localization of acoustic emission sources. *Ultrasonics*, 16(3):111–115, 1978.

[94] I. Grabec, T. Kosel, and P. Mužič. Location of continuous AE sources by sensory neural networks. *Ultrasonics*, 36(1):525–530, 1998.

[95] T. Yoshioka and T. Fujiwara. A new acoustic emission source locating system for the study of rolling contact fatigue. *Wear*, 81:183–186, 1982.

[96] B. Eftekharnejad, A. Addali, and D. Mba. Defect source location of a natural defect on a high speed-rolling element bearing with acoustic emission. In *Annual Conference of the Prognostics and Health Management Society*, 2011.

[97] A. Tobias. Acoustic-emission source location in two dimensions by an array of three sensors. *Non-destructive testing*, 9(1):9–12, 1976.

[98] M. Asty. Acoustic emission source location on a spherical or plane surface. *NDT International*, 11(5):223–226, 1978.

[99] P. Barat, P. Kalyanasundaram, and B. Raj. Acoustic emission source location on a cylindrical surface. *NDT & E International*, 26(6):295–297, 1993.

[100] R. Pullin, M. Baxter, M. Eaton, K. Holford, and S. Evans. Novel acoustic emission source location. *Journal of Acoustic Emission*, 25:215–223, 2007.

[101] F. Rosenblatt. *Principles of neurodynamics*. Spartan Book, 1962.

[102] J. Han and M. Kamber. *Data mining: concepts and techniques*. Morgan Kaufmann, 2006.

[103] P. J. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[104] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.

[105] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

[106] E. F. M. Filho and A. de Carvalho. Evolutionary design of MLP neural network architectures. In *Neural Networks, 1997. Proceedings., IVth Brazilian Symposium on*, pages 58–65. IEEE, 1997.

[107] B. A. Correa and A. M. Gonzalez. Evolutionary algorithms for selecting the architecture of a MLP neural network: A credit scoring case. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 725–732. IEEE, 2011.

[108] M. M. Khonsari and E. R. Booser. *Applied tribology: bearing design and lubrication*, volume 12. John Wiley & Sons, 2008.

[109] J. Williams. *Engineering Tribology*. Engineering Tribology. Cambridge University Press, 2005.

[110] D. Dowson. Elastohydrodynamic lubrication. 1970.