DEVELOPMENT OF EFFECTIVE GEARBOX FAULT DIAGNOSIS METHODOLOGIES UTILISING VARIOUS LEVELS OF PRIOR KNOWLEDGE

by **Stephan Schmidt**

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Mechanical Engineering)

in the

Department of Mechanical and Aeronautical Engineering

Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

January 2019

ABSTRACT

DEVELOPMENT OF EFFECTIVE GEARBOX FAULT DIAGNOSIS METHODOLOGIES UTILISING VARIOUS LEVELS OF PRIOR KNOWLEDGE

by

Stephan Schmidt

Supervisor:	Prof. P. Stephan Heyns
Department:	Mechanical and Aeronautical Engineering
University:	University of Pretoria
Degree:	Philosophiae Doctor (Mechanical Engineering)
Keywords:	Gear diagnostics; Bearing diagnostics; Varying operating conditions; Prior knowledge; Historical data

Effective fault diagnosis techniques are important to ensure that expensive assets such as wind turbines can operate reliably. Vibration condition monitoring data are rich with information pertaining to the dynamics of the rotating machines and are therefore popular for rotating machine diagnostics. However, vibration data do not only contain diagnostic information, but operating condition information as well. The performance of many conventional fault diagnosis techniques is impeded by inherent varying operating conditions encountered in machines such as wind turbines and draglines. Hence, it is not only important to utilise fault diagnosis techniques that are sensitive to faults, but the techniques should also be robust to changes in operating conditions.

Much research has been conducted to address the many facets of gearbox fault diagnosis e.g. understanding the interactions of the components, the characteristics of the vibration signals and the development of good vibration analysis techniques. The aforementioned knowledge, as well as the availability of historical data, are regarded as prior knowledge (i.e. information that is available before inferring the condition of the machine) in this thesis.

The available prior knowledge can be utilised to ensure that effective gearbox fault diagnosis techniques are designed. Therefore, methodologies are proposed in this work which can utilise the available prior knowledge to effectively perform fault diagnosis, i.e. detection, localisation and trending, under varying operating conditions. It is necessary to design

different methodologies to accommodate the different kinds of historical data (e.g. healthy historical data or historical fault data) that can be encountered and the different signal analysis techniques that can be used.

More specifically, a methodology is developed to automatically detect localised gear damage under varying operating conditions without any historical data being available. The success of the methodology is attributed to the fact that the interaction between gear teeth in a similar condition results in data being generated which are statistically similar and this prior knowledge may be utilised. Therefore, a dissimilarity measure between the probability density functions of two teeth can be used to detect a gear tooth with localised gear damage.

Three methodologies are also developed to utilise the available historical data from a healthy machine for gearbox fault diagnosis. Firstly, discrepancy analysis, a powerful novelty detection technique which has been used for gear diagnostics under varying operating conditions, is extended for bearing diagnostics under varying operating conditions. The suitability of time-frequency analysis techniques and different models are compared for discrepancy analysis as well. Secondly, a methodology is developed where the spectral coherence, a powerful second-order cyclostationary technique, is supplemented with healthy historical data for fault detection, localisation and trending. Lastly, a methodology is proposed which utilises narrowband feature extraction methods such as the kurtogram to extract a signal rich with novel information from a vibration signal. This is performed by attenuating the historical information in the signal. Sophisticated signal analysis techniques such as the squared envelope spectrum and the spectral coherence are also used on the novel signal to highlight the benefits of utilising the novel signal as opposed to raw vibration signal for fault diagnosis.

Even though a healthy state is the desired operating condition of rotating machines, fault data will become available during the operational life of the machine. Therefore, a methodology, centred around discrepancy analysis, is developed to utilise the available historical fault data and to accommodate fault data becoming available during the operation of the machine. In this investigation, it is recognised that the machine condition monitoring problem is in fact an open set recognition problem with continuous transitions between the healthy machine condition and the failure conditions. This is explicitly incorporated into the methodology and used to infer the condition of the gearbox in an open set recognition framework. This methodology uses a different approach to the conventional supervised machine learning techniques found in the literature.

The methodologies are investigated on numerical and experimental datasets generated

under varying operating conditions. The results indicate the benefits of incorporating prior knowledge into the fault diagnosis process: the fault diagnosis techniques can be more robust to varying operating conditions, more sensitive to damage and easier to interpret by a non-expert. In summary, fault diagnosis techniques are more effective when prior knowledge is utilised.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank the following people and organisations:

- Prof. Stephan Heyns for providing me the opportunity to do my doctorate under his supervision, the freedom to perform research on topics of interest to me, for providing his support and guidance and the opportunity to extend my professional network in this field.
- Prof. Konstantinos Gryllias for providing me the opportunity to perform research on bearing diagnostics under his joint supervision with Prof. Stephan Heyns, for exposing me to new techniques in the condition monitoring field and for his hospitality during my visit to KU Leuven.
- The Eskom Power Plant Engineering Institute (EPPEI) for funding the redevelopment of the gearbox test facility used in this work.
- Mr. George Breitenbach and Mr. Herman Booysen for their assistance with the experimental setup and during the experiments.
- Ms. Bonolo Mokoka for her administrative support.
- My wife, Jeané, for her continuous support and for her words of encouragement throughout this period.

LIST OF PUBLICATIONS BASED ON CANDIDATE'S RESEARCH

Journal papers:

1. Schmidt, S., Heyns, P.S. and De Villiers, J.P., 2018. A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques. *Mechanical Systems and Signal Processing*, 100, pp.152-166.

2. Schmidt, S., Heyns, P.S. and De Villiers, J.P., 2018. A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process. *Mechanical Systems and Signal Processing*, 100, pp.630-646.

3. Schmidt, S., Heyns, P.S. and Gryllias, K.C., 2019. A discrepancy analysis methodology for rolling element bearing diagnostics under variable speed conditions. *Mechanical Systems and Signal Processing*, 116, pp.40-61.

4. Schmidt, S., Heyns, P.S., 2019, An open set recognition methodology utilising discrepancy analysis for gear diagnostics under varying operating conditions. *Mechanical Systems and Signal Processing*, 119, pp.1-22.

5. Schmidt, S., Heyns, P.S., 2019, Localised gear anomaly detection without historical data for reference density estimation. *Mechanical Systems and Signal Processing*, 121, pp.615-635.

6. Schmidt, S., Heyns, P.S. and Gryllias, K.C., A pre-processing methodology to enhance novel information for rotating machine diagnostics. Submitted to *Mechanical Systems and Signal Processing* and accepted for publication.

7. Schmidt, S., Heyns, P.S. and Gryllias, K.C., A methodology using the spectral coherence and healthy historical data to perform gearbox fault diagnosis under varying operating conditions. Submitted to *Applied Acoustics*.

   This paper is invited for a special issue.

Conference papers:

1. Schmidt, S., Heyns, P.S. and De Villiers, J.P., Discrepancy signal processing techniques for gearbox condition monitoring applications, First World Congress on Condition Monitoring, London, United Kingdom, 13-16 June, 2017.

2. Schmidt, S., Heyns, P.S. and Gryllias, K.C., A probabilistic novelty detection methodology based on the order-frequency spectral coherence, The sixth International Conference on Condition Monitoring of Machinery in Non-Stationary Operations, Santander, Spain, 20-22 June, 2018.

   Nominated for the best paper in the category: Young researcher.

3. Schmidt, S., Heyns, P.S. and Gryllias, K.C., Discrepancy analysis for gearbox condition monitoring: A comparison of different healthy data models, The 31st International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management, Sun City, South Africa, 2-5 July, 2018.

4. Schmidt, S., Heyns, P.S. and Gryllias, K.C., A comparison of different features for discrepancy analysis-based bearing diagnostics, The 28th Biennial ISMA conference on Noise and Vibration Engineering, Leuven, Belgium, 17-19 September, 2018.

LIST OF ABBREVIATIONS

ACP	Averaged Cyclic Periodogram
AF-IPS	Angle-Frequency Instantaneous Power Spectrum
BPFI	Ball Pass Frequency Inner Race
BPFO	Ball Pass Frequency Outer Race
BSF	Ball Spin Frequency
CCP	Conventional Classification Procedure
CIC	Change in Condition
CSR	Closed Set Recognition
CWT	Continuous Wavelet Transform
DR	Decision Rule
DWT	Discrete Wavelet Transform
EEMD	Ensemble Empirical Mode Decomposition
EM	Expectation Maximisation
EMD	Empirical Mode Decomposition
FS	Fault Severity
FTF	Fundamental Train Frequency
GM	Gaussian Model
GMF	Gear Mesh Frequency
GMM	Gaussian Mixture Model
GNB	Gaussian Naive Bayes
HMM	Hidden Markov Model

KL	Kullback-Leibler
LDA	Linear Discriminant Analysis
LI	Linear Interpolation
LR	Logistic Regression
MCMC	Markov Chain Monte Carlo
NLL	Negative Log-Likelihood
OFSC	Order-Frequency Spectral Correlation
OFSCoh	Order-Frequency Spectral Coherence
OSR	Open Set Recognition
OT	Order Tracking
PCA	Principal Component Analysis
PDF	Probability Density Function
PMMDM	Posterior Mode of the Marginal Distribution of the Mean
QDA	Quadratic Discriminant Analysis
QM	Quality Measure
RMS	Root-Mean-Square
SA	Synchronous Average
SE	Squared Envelope
SES	Squared Envelope Spectrum
STFT	Short-Time Fourier Transform
SV	Synchronous Variance
WC	Wavelet Coefficients
WPT	Wavelet Packet Transform

Contents

1 Introduction **1**

1.1 Background . . . 1

1.2 Literature survey . . . 2

1.2.1 Gearbox vibrations for fault diagnosis . . . 2

1.2.2 Signal analysis for fault diagnosis . . . 6

1.2.3 Data-driven fault diagnosis . . . 13

1.3 Scope of work . . . 16

1.3.1 Prior knowledge . . . 16

1.3.2 Methodologies for effective fault diagnosis . . . 18

1.3.3 Summary of proposed methodologies . . . 21

1.4 Layout of thesis . . . 21

2 Localised gear anomaly detection methodology **24**

2.1 Introduction . . . 24

2.2 Methodology . . . 25

2.2.1 Feature extraction and processing . . . 26

2.2.2 Kullback-Leibler divergence . . . 27

2.2.3 Data analysis for automatic condition inference . . . 29

2.3 Validation . . . 33

2.3.1 Experimental setup . . . 33

2.3.2 Investigation 1 . . . 34

2.3.3 Investigation 2 . . . 40

2.3.4 Diagnosis of the pinion . . . 47

2.4 Conclusion and recommendations . . . 48

3 Discrepancy analysis methodology for bearings **50**

3.1 Introduction . . . 50

3.2 Methodology . . . 51

3.2.1 Overview . . . 51

3.2.2 Feature extraction . . . 52

3.2.3 Feature modelling . . . 54

3.2.4 Discrepancy signal generation . . . 55

3.2.5 Processing the discrepancy signal . . . 55

3.3 Numerical validation . . . 58

3.3.1 Phenomenological gearbox model . . . 58

3.3.2 Vibration data . . . 61

3.3.3 Results . . . 62

3.4 Experimental validation . . . 64

3.4.1 Investigation 1 . . . 64

3.4.2 Investigation 2 . . . 67

3.5 Additional investigations . . . 72

3.6 Conclusion and recommendations . . . 74

4 Spectral coherence novelty detection methodology **75**

4.1 Introduction . . . 75

4.2 Methodology . . . 76

4.2.1 Estimating the residual signal . . . 76

4.2.2 Order-Frequency Spectral Coherence (OFSCoh) . . . 77

4.2.3 Feature extraction . . . 78

4.2.4 Healthy data model evaluation . . . 80

4.2.5 Novelty detection . . . 82

4.3 Numerical validation . . . 82

4.3.1 Phenomenological gearbox model . . . 82

4.3.2 Results . . . 84

4.4 Experimental validation . . . 87

4.4.1 Procedure . . . 87

4.4.2 Results . . . 89

4.5 Conclusion and recommendations . . . 90

5 Novel information enhancement methodology **92**

5.1 Introduction . . . 92

5.2 Methodology . . . 93

5.2.1 Signal decomposition and feature extraction . . . 93

5.2.2 Feature modelling . . . 95

5.2.3 Filter design to enhance novel information . . . 96

5.2.4 Novel signal estimation . . . 96

5.3 Phenomenological gearbox model . . . 99

5.3.1 Model . . . 99

5.3.2 Results with raw data . . . 101

5.3.3 Results with proposed methodology . . . 102

5.4 Experimental gearbox dataset . . . 106

5.4.1 Overview of dataset . . . 106

5.4.2 Results . . . 108

5.5 Conclusion and recommendations . . . 117

6 Open set recognition methodology **118**

6.1 Introduction . . . 118

6.2 Condition recognition for condition monitoring . . . 119

6.3 Methodology . . . 120

6.3.1 Discrepancy analysis . . . 121

6.3.2 Discrepancy feature extraction . . . 124

6.3.3 Condition recognition procedure . . . 125

6.3.4 Final remarks on the methodology . . . 128

6.4 Synthetic data investigation . . . 128

6.4.1 Synthetic dataset . . . 129

6.4.2 Discrete machine conditions for model optimisation . . . 131

6.4.3 Complete dataset for model optimisation . . . 133

6.5 Experimental validation . . . 135

6.5.1 Experimental dataset . . . 135

6.5.2 Application of proposed methodology . . . 136

6.5.3 Comparison to a Conventional Classification Procedure (CCP) . . . 144

6.6 Conclusion and recommendations . . . 147

7 Conclusion and recommendations **148**

7.1 Conclusion of thesis 148

7.2 Recommendations for future work 151

Bibliography **152**

A Additional results and investigations **A1**

A.1 Sensitivity analysis for prior parameters in Chapter 2 . . . . . A1

A.2 Bearing discrepancy analysis methodology information A4

A.2.1 Compensating for rotational speed variations A4

A.2.2 Comparing the suitability of different features A7

A.2.3 Comparing the suitability of different data models A17

A.3 Quality metrics used in Chapter 5 . . . . . A24

B Phenomenological gearbox model information **B1**

B.1 Model used in Section 3.3 . . . . . B1

B.2 Model used in Section 4.3.1 . . . . . B2

B.3 Model used in Section A.2.2 . . . . . B3

B.4 Model used in Section A.2.3 . . . . . B4

Chapter 1

Introduction

1.1 Background

Gearboxes are critical assets used in industries such as the power generation (e.g. wind turbines) and mining industries (e.g. draglines). It is crucial to use effective maintenance strategies for gearboxes due to their high downtime when performing maintenance, their high replacement cost and their susceptibility to damage [1, 2]. The most common maintenance strategies can be ineffective as corrective maintenance leads to unexpected failures, unexpected downtime and potential secondary damage, while time-based preventative maintenance is inefficient due to unnecessary downtime and components being replaced with unutilized remaining useful life [3–5].

Condition-based maintenance is an effective alternative to time-based preventative and corrective maintenance, because the inferred condition of the machine is for example used as a basis for maintenance decisions as opposed to the time of operation [3]. In condition-based maintenance, event and condition monitoring data are collected from the machine under consideration, whereafter the data are processed and subsequently used to perform maintenance decisions. Condition-based maintenance consists of diagnostics, where the collected data are used to detect, localise and trend damage, and prognostics, where the remaining useful life or time to failure is estimated from the collected data [3–5]. While prognostics can result in more effective maintenance planning, it remains imperative to be able to detect and characterise damage early in the degradation process, i.e. using an effective diagnostic system.

In wind turbines, the most prominent failure modes in gearboxes are for example teeth surface pitting, teeth bonding, gear fracture and bearing damage and are attributed to for example improper installation, inadequate lubrication, and excessive overloading [1, 6]. It

is therefore very important to be able to detect gear and bearing damage to ensure that the gearboxes operate as intended. Vibration-based condition monitoring techniques are most commonly used for rotating machine diagnostics (e.g. bearing and gear diagnostics), because the vibration data are rich with information pertaining to the current condition of the machine, it is possible to use sophisticated signal processing techniques to detect and identify the damaged component and the data are relatively easy to acquire [1, 7, 8].

However, rotating machines such as wind turbines [9], conveyor systems [10], bucket wheel excavators [11] and draglines [12] operate continuously under varying operating conditions e.g. varying load and speeds. The varying operating conditions result in changes in the vibration data, which make detecting changes in machine condition more difficult [11, 13–15]. This can result in the wrong condition to be inferred and subsequently inappropriate maintenance decisions to be made. Hence, it is critical to use gearbox fault diagnosis techniques that are sensitive to changes in machine conditions, while their performance is unaffected by varying operating conditions.

There is a need for effective gearbox fault diagnosis techniques, capable of detecting, localising and trending gear and bearing damage under varying operating conditions. A literature survey of vibration-based condition monitoring techniques for gearbox diagnostics is performed in the next section.

1.2 Literature survey

A literature survey is conducted on vibration-based condition monitoring techniques for gearbox diagnostics (i.e. gear and bearing diagnostics) and is partitioned as follows: In Section 1.2.1, the characteristics of the vibration data of gearboxes are investigated, whereafter signal analysis techniques for fault diagnosis are considered in Section 1.2.2 for the vibration data. Finally, data-driven fault diagnosis techniques are investigated in Section 1.2.3 to incorporate the available historical data into the fault diagnosis process.

1.2.1 Gearbox vibrations for fault diagnosis

It is important to understand the relationship between the interacting machine components in different conditions and the resulting vibration signal to ensure that the appropriate signal analysis techniques can be used to diagnose the condition of the machine. It is also important to generate numerical gearbox data to validate new fault diagnosis methods in a completely

controlled environment. Therefore, the characteristics of gearbox vibration signals are investigated in Sections 1.2.1.1 and 1.2.1.2, while methods to generate numerical gearbox data are presented in Section 1.2.1.3.

1.2.1.1 Gearbox vibration signals under constant operating conditions

Gearbox vibration signals are dominated by the vibrations resulting from the periodic meshing of the gear teeth. This is attributed to the variation of the gear mesh stiffness due to non-integer contact ratios, deviations from the involute gear profile and the deformation of the gear teeth under load which ultimately manifest as strong excitations at the gear mesh frequency [7, 16]. Gear damage results in changes in the gear mesh stiffness of the gear [17, 18] which manifest as changes in the amplitude of the vibration signal [16] and changes in the instantaneous angular speed of the gears [19]. Localised gear damage results in localised amplitude modulation i.e. amplitude modulation with a short duration relative to the gear rotation period, while distributed gear damage results in amplitude modulation with a longer period [16].

In rolling element bearings, damage on the inner or outer race of the bearing results in impacts as the load-carrying rolling elements move through the damaged portion, whereas a damaged rolling element results in impacts as the damaged region makes contact with the inner and outer races of the bearing [20]. The magnitude of the bearing impulses is dependent on the load zone. For example, inner race and rolling element damage signals are amplitude modulated as the fault moves through a constant load zone, while the outer race damage signal remains constant with respect to a constant load zone [20]. The magnitude of the bearing impulses is also non-linearly related to the condition of the bearing. El-Thalji and Jantunen [21] and Cerrada et al. [22] illustrated how changes in the condition of the bearing influence the magnitude of the impacts due to the damage and the measured vibrations.

The vibrations resulting from interacting machine components such as the impacts due to bearing damage in a gearbox are transmitted through the machine components (e.g. shafts and bearings) and casing of the machine to reach sensors such as accelerometers. The frequency response function of the transmission path acts as a filter on the vibration signal and therefore the measured vibration signal is dependent on the frequency response function of the machine, the location of the sensors and the rotational speed of the system [19, 23–25]. The interaction between the rotating components e.g. gear mesh interactions or bearing damage, also results in changes in the statistics of the vibration signal of the machine over time i.e.

the vibration signals are inherently non-stationary. However, due to the periodic nature of the interactions, the statistics of the vibration signals are periodic as well and therefore the signals are a special class of non-stationary, namely, cyclostationary [23, 26, 27]. To be more precise, the signals are angle-cyclostationary under stationary operating conditions due to the angle-locked periodic mechanisms of rotating machine components [23]. Different types of cyclostationary signals are formally defined and related to machine vibration by Antoni et al. [23]. Pure first-order cyclostationary signals, generated by misalignments and gear meshing for example have periodic means, while pure second-order cyclostationary signals generated by wear for example have periodic autocovariance functions. Hence, this distinction necessitates using appropriate techniques for first-order and second-order cyclostationary component detection [23].

1.2.1.2 Gearbox vibration signals under varying operating conditions

It is evident that the vibrations of the machines are inherently operating condition dependent; the gear mesh signal is load and speed dependent due to the load dependent deformation and the speed dependent interactions between the gear teeth during meshing [13, 25]. The frequency of the impacts in damaged bearings is for example dependent on the rotation of the shaft, while the magnitude of the impacts is dependent on the rotational speed and the applied load [28, 29]. The transmission path effect also plays a significant role under varying speed and varying load (which inherently result in varying speed [13, 19]) conditions. The magnitude and phase of the frequency response function between the interactions of the machine components and the sensor are frequency dependent, which means that changes in rotational speed result in changes in the amplitude and the phase of the measured vibration signals [25, 30]. Therefore, the statistics of the vibration signals are operating condition dependent, which result in the cyclostationarity of the signal being compromised i.e. the signal is cyclo-non-stationary under varying operating conditions [31].

The frequency modulation due to varying rotational speeds is usually attenuated by converting the signal from a time domain representation to an angle representation by using rotational speed information with order tracking [32]. The rotational speed can be estimated with measurement equipment such as a zebra tape shaft encoder and an optical probe [33] or from the vibration signal itself using tacholess speed estimation methods [34–39]. Instantaneous angular speed measurements also contain diagnostic information and have been used for fault diagnosis as opposed to vibration transducers e.g. accelerometers, because it min-

imises the number of sensors, the path between the vibration source and the measurement equipment is simpler and the noise contamination is less in the signals [19, 40]. Recently, much research are focused on tacholess order tracking methods i.e. performing order tracking without rotational speed measurements. Tacholess methods are attractive for condition monitoring, because the rotational speed measurements are not always possible due to limited access to the rotating components and harsh operating environments, the rotational speed measurement equipment can increase the cost of performing condition monitoring and phase distortion occurs when utilising the rotational speed measurements for order tracking signals from transducers with different transfer paths [30, 35–37].

The amplitude modulation due to varying load and speed conditions can be cyclic stationary (e.g. a sinusoidal load with a period equal to the rotational period of the gear) or non-cyclic stationary. Stander et al. [13] developed a load demodulation normalisation technique for normalising the amplitude modulation for cyclic stationary conditions and used the synchronous average, discussed further in Section 1.2.2.5, for non-cyclic stationary conditions. Abboud et al. [41] developed the generalised synchronous average to accommodate the amplitude modulation caused by varying speed conditions.

1.2.1.3 Physics-based modelling of gearboxes

It is further useful to develop gearbox models which can replicate and help to understand experimental phenomena, help to understand the dynamics of the rotating components and to provide a controlled environment for testing new fault diagnosis techniques. The modelling approaches are categorised as lumped mass models, finite element models and phenomenological models [28, 42, 43].

Lumped mass models aim to lump the continuous components that have mass, stiffness and damping, into discrete mass-spring-damper components. These models are relatively easy to implement and computationally fast to solve compared to finite element models for example. Some interesting lumped mass models have been developed by Bartelmus [44], Howard et al. [45], Chaari et al. [17, 46], Bartelmus et al. [10], and Mohammed et al. [18, 47]. The time-varying gear mesh stiffness, due to the interaction of the gear teeth during meshing, is usually modelled using analytical techniques or obtained from finite element analyses [42, 46, 48]. Finite element models are capable of modelling the interaction of the dynamics of the interacting components more accurately at the expense of increased modelling complexity and the high computational cost [42].

Phenomenological models aim to directly model the signal components generated by gearboxes based on experimental observations and by having a physical understanding of the interaction between the components e.g. understanding the statistical properties of the signals. Some important phenomenological models include a bearing damage model [29], a revised gearbox signal model [49] and a bearing and distributed gear damage model developed for varying speed conditions [28]. Even though the phenomenological models may not be as accurate as lumped mass or finite element models, it is computationally very efficient to generate data that can replicate complicated experimental observations e.g. amplitude modulation due to varying speeds and therefore it is advantageous to use in the early stages of developing fault diagnosis techniques.

The appropriate signal analysis techniques can therefore be selected by understanding the characteristics of vibration signals and by testing the techniques on data generated from a controlled environment. In the next section, different signal analysis techniques are considered by using the information contained in this section.

1.2.2 Signal analysis for fault diagnosis

The vibration data of rotating machines contain much information related to the machine and its operation. Fault diagnosis is impeded by non-diagnostic information, because the weak diagnostic information associated with bearings is masked by strong non-diagnostic information such as gear mesh interactions for example. It is therefore beneficial to process the vibration data by enhancing the diagnostic information and attenuating the non-diagnostic information. Various methods are investigated in this section for processing and analysing vibration data for gearbox fault diagnosis.

1.2.2.1 Condition indicators

Statistical time domain condition indicators are one of the simplest metrics for analysing vibration signals. The assumption is made that changes in the condition of the machine result in the average statistical properties of the signal such as the energy and the impulsiveness to change as well. Tandon [50] compared different metrics for bearing diagnostics and found that the root-mean-square and the peak value are good condition indicators. However, localised changes in the vibration signal can be masked by other dominating components such as gear mesh frequencies for example, which has resulted in more sophisticated metrics such as the FM4, i.e. the kurtosis of the residual signal to be used instead [7]. Hu et al. [51] used the

averaged logarithmic ratio, calculated from the side band ratios of the gear mesh frequencies, for diagnosing distributed gear damage. The degree-of-cyclostationarity and indicators-of-cyclostationarity can also be used as diagnostic metrics. McCormick and Nandi [52] used the degree-of-cyclostationarity for bearing diagnostics and Raad et al. [53] used indicators-of-cyclostationarity for gear diagnostics. Antoni and Borghesani [54] developed an interesting methodology to systematically design condition indicators for effective fault diagnosis and applied it for bearing diagnostics.

However, varying speed and load conditions result in changes in the statistical properties of the vibration signal as well. This makes it difficult to distinguish between changes in operating conditions and changes in the condition of the machine. Bartelmus and Zimroz [11] found that the statistical properties become more sensitive to changes in operating conditions as a gearbox deteriorates and used this information to distinguish between a healthy and a damaged bucket wheel excavator gearbox. Zimroz et al. [9] used the regression coefficients, determined for a statistical feature and the operating condition information, as a diagnostic metric to detect and trend wind turbine gearbox damage. Gryllias et al. [55] extended indicators of cyclostationarity for cyclo-non-stationary signals e.g. bearing signals acquired under varying operating conditions.

1.2.2.2 Time-frequency analysis

A frequency domain representation of the vibration data can provide useful insight into the characteristics of rotating components and can assist with localising the fault i.e. determine the component that is damaged. However, the Fourier spectrum is only useful to analyse signals that are first-order time-cyclostationary e.g. typically gear mesh signals [23, 27]. The time-cyclostationarity of the vibration signals is violated under varying load and speed conditions and therefore it is useful to use a multi-dimensional representation.

Time-frequency analysis techniques describe the flow of different frequency characteristics over time and are therefore useful for analysing vibration signals. The Short-Time Fourier Transform (STFT) assumes the signal is quasi-stationary in its calculation and can be used to represent the time-frequency properties of signals [3, 56]. For example, the spectrogram, i.e. the squared magnitude of the STFT, has been used for early gear fault detection [57]. Wang and McFadden [58] also used image processing techniques to quantify the characteristics of the damage in the spectrogram, where it was possible to effectively detect changes in the condition of the gear in a helicopter gearbox.

However, the STFT has a constant time and frequency resolution which can impede the characterisation of non-stationary signals [17]. Hence, it is beneficial to use more sophisticated techniques for analysing non-stationary signals. The Chirplet transform [35], the Wigner-Ville distribution [13, 19], the wavelet transform [59] and the empirical mode decomposition [60] are examples of techniques which are better suited for characterising non-stationary signals. The wavelet transform and the empirical mode decomposition are discussed further in Section 1.2.2.3 and Section 1.2.2.4 respectively.

1.2.2.3 Wavelet analysis

The wavelet transform uses a dilated and translated wavelet basis function i.e. a short, oscillatory, non-zero function, to decompose signals into time-scale representations. The benefits of using the wavelet transform is that it has good time and frequency localisation which makes it appropriate for analysing non-stationary signals and it is well-suited for detecting discontinuities in the signals. These aforementioned properties make the wavelet transform well-suited for fault diagnosis [59, 61–63].

Three variations of the wavelet transform that are mostly used for rotating machine fault diagnosis are the Continuous Wavelet Transform (CWT), the Discrete Wavelet Transform (DWT) and the Wavelet Packet Transform (WPT). The CWT decomposes the signal into continuous predetermined frequencies over time and has been extensively used for gear diagnostics [24, 61, 64–68] and bearing diagnostics [67, 69]. The CWT has for example been used for signal denoising [67], combined with residual signal analysis for more effective gear crack detection [24], used to develop a condition indicator for varying load conditions [70] and used with data-driven methodologies for automatic fault diagnosis [61, 64, 65]. However, much redundant information exists in the CWT and it can be difficult to select the specific frequencies when applying the transform.

The DWT uses quadrature mirror filters to decompose the signal into approximation and detail coefficients, whereafter the approximation coefficients are decomposed again with quadrature mirror filters [71]. This results in a set of wavelet coefficients to be obtained, where the high frequency bands have poor frequency resolution and good time resolution and lower frequency bands have good frequency resolution and poor time resolution [71]. Using the DWT is advantageous for fault diagnosis due to its computational efficiency and less data are being generated compared to the CWT. It has performed well for engine diagnostics [72] and bearing diagnostics [73].

However, in rotating machine diagnostics the important diagnostic information is usually located in the higher frequency bands, which means that good frequency resolution is required for those frequency bands as well. The WPT works on a similar principle as the DWT, but instead of only decomposing the approximation coefficients at each decomposition level, the approximation and the detail coefficients are decomposed at each level [71]. The consequence of this is that each set of approximation and detail coefficients at a specific decomposition level have the same bandwidth [71]. The aforementioned property and the computational efficient decomposition procedure make the WPT ideal for rotating machine fault diagnosis [74–77].

However, wavelet analysis has some drawbacks. Its performance is dependent on the choice of the wavelet basis function as well as the level of decomposition (for DWT and WPT) or the frequencies that are analysed (for the CWT) [59, 69]. It is also difficult to interrogate the wavelet transformed signal for the presence of damage [61] and therefore wavelet transform is frequently combined with data-driven techniques discussed in Section 1.2.3.

1.2.2.4 Empirical mode decomposition

The Empirical Mode Decomposition (EMD) is a time-frequency, non-parametric decomposition of a signal into intrinsic mode functions. The intrinsic mode functions are natural modes embedded in the vibration signal and can be used for diagnosing rotating machines [60, 78–81]. The EMD is performed using a recursive sifting process, where the higher frequency content is embedded in the first intrinsic mode functions and the lower frequency content manifests in the last intrinsic mode functions [60].

The benefits of the EMD are that it is self-adaptive i.e. the basis functions do not have to be specified before the signals can be analysed and it is suitable for analysing non-stationary and non-linear signals [60]. These properties have made it very attractive for bearing diagnostics [78] and gear diagnostics [80, 81] for example. However, the EMD has serious problems such as mode mixing which adversely influences its performance. This has resulted in many improvements being investigated for fault diagnosis applications [60]. For example the noise assisted EMD, usually referred to as the Ensemble EMD (EEMD), aims to reduce the mode mixing phenomenon by calculating the ensemble average of intrinsic mode functions [82, 83].

1.2.2.5 Cyclostationary analysis

Cyclostationary analysis techniques are well-suited for detecting damage which manifest as cyclostationary components in the vibration signals. First and second-order cyclostationary analysis techniques are considered in this section, because first and second-order cyclostationary signals are frequently encountered in rotating machine diagnostics [23, 27, 28, 84].

The synchronous average is a popular first-order cyclostationary technique; it is an estimate of the first-order cyclostationary components in a signal under the assumption of cyclo-ergodicity [23]. It can be used to retain mechanisms with periodic mean functions e.g. gear mesh signals and synchronous modulation due to varying loads [13, 23, 41] and it can be used to attenuate non-synchronous mechanisms e.g. non-synchronous modulation and second-order cyclostationary signals [19]. However, the efficiency of the synchronous average for gearbox diagnostics is poor when used in isolation [24, 43, 85]; the synchronous average of the filtered signal [85] and the residual signal e.g. by removing the gear mesh components have been much more successful for gearbox diagnostics [24, 43, 86].

Second order cyclostationary signals, which have zero mean functions and periodic autocovariance functions, manifest in the vibration signal due to wear and bearing damage for example [23, 87]. It is firstly necessary to separate the deterministic i.e. first-order cyclostationary components and random components e.g. second-order cyclostationary components in the vibration signal before the appropriate analysis techniques can be applied [41]. The second-order cyclostationary signal can be estimated by subtracting the periodic mean function i.e. the synchronous average from the vibration signal [20, 28, 41] or by extracting bandlimited second-order cyclostationary signals with bandpass filters [20, 28, 88]. The appropriate structural resonance frequency band can be calculated with healthy historical data [89, 90], historical fault data or without historical data, with techniques to address the latter presented in Section 1.2.2.6. The synchronous variance i.e. the synchronous average of the mean instantaneous power of the residual signal, is a second-order technique which has been successful for gear and bearing diagnostics [23, 68, 91].

Envelope analysis is another important second-order cyclostationary technique [28, 88, 92]. In a damaged bearing vibration signal, the time invariant structure is being modulated by periodic impulses, produced by bearing damage, and masked by deterministic components such as gear mesh signals [7, 20]. In envelope analysis, either the spectrum of the amplitude demodulated random part of the vibration signal is calculated [88, 93] or the spectrum of

the squared envelope (i.e. instantaneous power) of the random part is calculated [28, 93], whereafter the spectrum is interrogated for the presence of the modulating signal e.g. the cyclic frequency of the bearing impulses.

The spectral correlation of the signal is a bi-spectral distribution which indicates the power of a signal in a cyclic frequency, spectral frequency plane [84] and it is robust for detecting and characterising second-order cyclostationary signals [31]. It is often useful to enhance weak second-order cyclostationary components in the spectral correlation. This can be done by utilising the spectral coherence, a normalised form of the spectral correlation [27, 84, 87]. Antoni [84] formalised the spectral correlation and compared different estimators, whereas Antoni [87] investigated cyclic spectral analysis techniques for bearing diagnostics. Antoni et al. [94] developed an algorithm for the fast spectral correlation and Borghesani and Antoni [95] developed a faster spectral correlation, which significantly reduces the computational cost for estimating the spectral correlation. Capdessus et al. [96] used the spectral correlation and spectral coherence to detect the presence of a spall on a gear tooth.

There is actually a close connection between envelope analysis and cyclic spectral analysis; the squared envelope spectrum and the enhanced envelope spectrum can be obtained by marginalising over the frequency bands of the spectral coherence and the absolute spectral coherence respectively [94]. The enhanced envelope spectrum is able to detect weak components better than the squared envelope spectrum, and is therefore better suited for incipient damage detection [94].

However, the cyclostationarity of the vibration signals is destroyed under varying speed conditions; the bearing vibration signals are neither time- nor angle-cyclostationary, but rather cyclo-non-stationary [31]. The impacts and modulation induced by bearing impacts are angle-dependent, while the carrier frequency is time-invariant and for limited speed fluctuations [31] the vibration signals can be approximated as angle-time cyclostationary [31, 97, 98]. Abboud et al. [28] and Borghesani et al. [92] investigated the extension of envelope analysis under varying speed conditions, while Abboud et al. [31, 97] developed the order-frequency spectral correlation and the order-frequency spectral coherence for detecting second-order angle-time cyclostationary components. The instantaneous power spectrum is another important cyclostationary analysis technique that describes the instantaneous power of the different frequency bands of a signal over time or angle and has also been very useful for gearbox diagnostics under varying load conditions [99]. Urbanek et al. [15, 100] used the synchronous averaged angle-frequency instantaneous power spectrum for detecting gear and

bearing damage in wind turbine gearboxes under varying operating conditions.

1.2.2.6 Narrowband feature extraction

Bearing damage manifests in specific frequency bands which result in changes in the average statistical properties, e.g. impulsiveness, in those frequency bands. Hence, narrowband feature extraction techniques (i.e. techniques which extract features from a set of narrowband signals) are capable of identifying frequency bands with changing statistical properties due to damage.

The spectral kurtosis estimates the kurtosis of different frequency bands and can therefore indicate the frequency bands that contain much transient information [56]. The spectral kurtosis has to be estimated using time-frequency analysis approaches e.g. STFT and its performance depends on the parameters of the time-frequency distribution e.g. the window length [56]. Antoni [56] provided suggested parameters for the window overlap and the window type for the STFT. The optimal window length depends significantly on the characteristics of the investigated data and therefore Antoni [56, 101] suggested that the spectral kurtosis needs to be calculated for all potential window lengths. The spectral kurtosis over a range of candidate window lengths is referred to as the kurtogram. The kurtogram can be estimated by calculating many STFTs [101] or by using the wavelet coefficients, calculated at each level of the WPT [102], with the latter referred to as the fast kurtogram due to its computational efficiency. The kurtogram can ultimately be used to design a bandpass filter for envelope analysis, where the resulting envelope spectrum will contain the information which had the maximum impulsiveness in the spectral frequency bands [56, 101, 102]. Combet and Gelman [103] proposed using the spectral kurtosis to design a filter that enhances impulsive frequency bands. Barszcz and Randall [104] also successfully applied the spectral kurtosis for wind turbine gearbox diagnostics.

Even though the spectral kurtosis and the kurtogram have successfully been used for rotating machine diagnostics, they have inherent deficiencies: The spectral kurtosis and the kurtogram are adversely affected by the presence of spurious impulses and their performance is dependent on the repetition rate of the impulses [105]. Hence, other estimators have been proposed to improve the robustness of the important frequency band estimations, namely, the protrugram [106], the improved kurtogram [107], the enhanced kurtogram [108], the optimised spectral kurtosis [109], the sparsogram [108], the infogram [105, 110, 111] as well as various other estimators [112]. The envelope harmonic-to-noise ratio is another useful technique to

detect periodic impulses in frequency bands [113]. Hence, it is evident that many statistics can be estimated or features can be extracted from localised frequency bands to detect subtle changes due to damage.

1.2.3 Data-driven fault diagnosis

The aforementioned signal analysis tools can be used to perform fault diagnosis of rotating machines, however, it can be difficult to implement and to interpret for a non-expert. Hence, it is beneficial to use an intelligent system or decision making model to learn the important characteristics of the data, whereafter the model[1] is used for automatic decision making. This is performed by building a data-driven model of the available data, whereafter the model can for example return a diagnostic metric e.g. a health index, or the condition of the machine [22].

It is necessary to train the model on the available historical data i.e. finding the parameters of the model which minimise a loss function for the data under consideration. In this thesis, two different learning approaches are considered: Supervised learning using historical fault data and novelty detection when only healthy gearbox data or data of a gearbox in a reference condition are available.

1.2.3.1 Supervised learning with historical fault data

Classification models such as neural networks and support vector machines can be optimised by providing the model with the data and the loss function with a corresponding label [71, 114, 115]. This results in models that can be used to automatically assign labels to new data i.e. to automatically infer the condition of the machine from the processed condition monitoring data.

A popular approach for classification is to have a separate feature extraction and feature modelling step [65, 73, 116–121]. Firstly, features are extracted and selected based on the knowledge of an expert or by using a data-dependent criterion e.g. selecting features which maximise the distance between classes [71]. In most cases, signal analysis tools such as the average statistical indicators of vibration data [79], wavelet analysis [65, 77, 79, 117], EMD [119, 122, 123] and EEMD [83] are used as basis for feature extraction due to their ability to highlight or extract diagnostic information from a vibration signal. Thereafter, different

[1] In this section, the term *model* refers to a data-driven model (e.g. neural network, Gaussian mixture model) of the acquired vibration data and should not be confused with physics-based models discussed in Section 1.2.1.

models such as support vector machines [116, 124], neural networks [116, 125], hidden Markov models [73, 126], Gaussian mixture models [126], k-Nearest Neighbours [127], k-means clustering [128], and non-Naive Bayes classifiers [123], are used to learn the relationship between the extracted features and the labels with a supervised learning optimisation procedure. The optimised model can then be used to subsequently learn to automatically infer the condition of the machine from the extracted features.

Dimensionality reduction techniques are also important in the machine learning field, because the problems associated with high-dimensional data e.g. the curse of dimensionality can be circumvented by projecting the data onto a lower-dimensional subspace or manifold. Van der Maaten et al. [129] compared various dimensionality reduction approaches on artificial and real datasets, and their results indicated the importance of principal component analysis for dimensionality reduction of real datasets. Zimroz and Bartkowiak [130] used conical discriminant analysis and principal component analysis to project high dimensional features to a two-dimensional feature space which can be used to separate a healthy gearbox and a damaged gearbox under non-stationary operating conditions. Bartkowiak and Zimroz [131] performed dimensionality reduction by performing feature subset selection for gearbox fault diagnosis. Wan et al. [77] compared linear and different non-linear dimensionality reduction techniques for gear diagnostics, with the use of principal component analysis before different classification models having the best predictive performance.

The recent advances of deep learning approaches in the computer vision field [132] has resulted in deep learning approaches to be applied for condition monitoring [133, 134]. This is based on the assumption that the features are selected manually with expert knowledge of the domain and the standard neural networks are not sufficiently flexible to capture the important characteristics in the data of condition monitoring problems [133]. This has resulted in interesting methodologies based on deep neural networks [133], deep support vector machines [134], deep random forests [135], deep autoencoders [136], convolutional neural networks [137] and generative adversarial networks [138] to automatically extract features from the data, whereafter the condition of the machine is inferred from the extracted features. A review of deep learning approaches in system health monitoring is given by Khan and Yairi [139]. Their review indicated that even though there is much potential for deep learning approaches, there are also some shortcomings, e.g. the design of the network architecture is machine and application dependent and there is not a systematic method to design the architecture.

Even though powerful classification techniques exist, historical fault data can be difficult

or impractical to acquire for rotating machines. This has resulted in researchers developing physics-based approaches, where firstly physical models are used to generate numerical vibration data from a machine with different damage modes. Thereafter, a data model is trained on the numerical data and used to predict the condition of an actual machine. Gryllias et al. [140] and Sobie et al. [141] used physical models to generate bearing data, which are subsequently used to train a classification model for predicting the condition of measured data.

However, most supervised learning methodologies fail to recognise or acknowledge that for many practical problems only a sample of the population of class labels are available at the time of model optimisation. The supervised learning-based optimised models, only provide the label of the most probable class during inference; the models cannot distinguish between a novel class and a class that was present during model optimisation, which could lead to erroneous inferences. It is impractical and expensive to develop models based on the assumption that data from the population of all class labels are available at training time i.e. this means that the machine has to have failed in all of its damage modes before the models are optimised. Therefore, it is sensible to recognise that the training data do not represent the population of class labels, but only a sample of the class labels.

It is therefore important to be able to distinguish between data that are present in the training set and data from novel classes when making predictions. The notion that models should have a reject option is not a new concept [114, 127, 142], however Scheirer et al. [143, 144] formally recognised that many machine learning problems are in fact Open Set Recognition (OSR) problems and not Closed Set Recognition (CSR) problems. Both CSR and OSR frameworks aim to infer which class is associated with the data, but the OSR framework realises that the data can potentially be from a new class that was not present in the dataset used to optimise the model. Hence, it is necessary to be able to reject the inferred label by labelling the data as a novelty. The OSR problem has also been investigated by [145] for different datasets.

1.2.3.2 Novelty detection with only healthy historical data

It is desired to have rotating machines that operate in a healthy condition, which makes access to healthy condition monitoring data relatively easy. Hence, it is practical to use data-driven approaches in a novelty detection framework, where the objective is to determine whether the condition of the machine is healthy or not. This is a special case of OSR, where only one

class is available when optimising the model. Good overviews of different novelty detection approaches for a wide range of applications are found in Refs. [146–148].

Specifically in the condition monitoring field, many methods have been investigated for novelty detection. Timusk et al. [149] compared different novelty detection models for rotating machine diagnostics under varying operating conditions. Fernandez-Francos et al. [90] used a one-class support vector machine model for bearing novelty detection, with the model used to also determine the frequency band for envelope analysis. The Kullback-Leibler divergence is also popular for novelty detection in various applications; for example, changes in the probability density function of the data have been used to detect distillation process anomalies [150, 151]. Figueiredo et al. [152] used a Bayesian Gaussian mixture model approach to detect novelties caused by damage for structural health monitoring applications. Heyns et al. [153] proposed an interesting approach where autoregressive models, trained on the healthy gearbox data, are used to perform gear diagnostics under varying operating conditions.

Discrepancy analysis is another powerful rotating machine diagnostics technique when only healthy vibration data are available [154–156]. Localised novelty detection measures are generated for the angular data under investigation to form a discrepancy signal. This is performed by using a model of the angular data from a healthy machine. The discrepancy signal can then subsequently be analysed with synchronous averaging to perform fault diagnosis under varying operating conditions. Discrepancy analysis has been successful for gearbox diagnostics under varying load and speed conditions [154–157].

1.3 Scope of work

The scope of this work is divided into three sections: The term *prior knowledge* is defined in Section 1.3.1, whereafter the focus of the research (i.e. developing effective fault diagnosis methodologies) is discussed in Section 1.3.2. Lastly, a summary and the connectivity of the proposed methodologies are given in Section 1.3.3.

1.3.1 Prior knowledge

It is evident from Section 1.2.1 that much research has been conducted in the condition monitoring field to understand the physics of rotating machines and the implications of damaged components on the dynamics of the system. As seen in Sections 1.2.1 and 1.2.3,

much research has also been conducted to develop and use different signal analysis techniques that can provide representations of the data that are rich with diagnostic information, and to implement data models that can be used to assist with the decision making process by utilising historical data. The aforementioned information is categorised into engineering knowledge and knowledge extracted from machine learning algorithms and collectively referred to as prior knowledge, with the prior knowledge summarised in Figure 1.1.

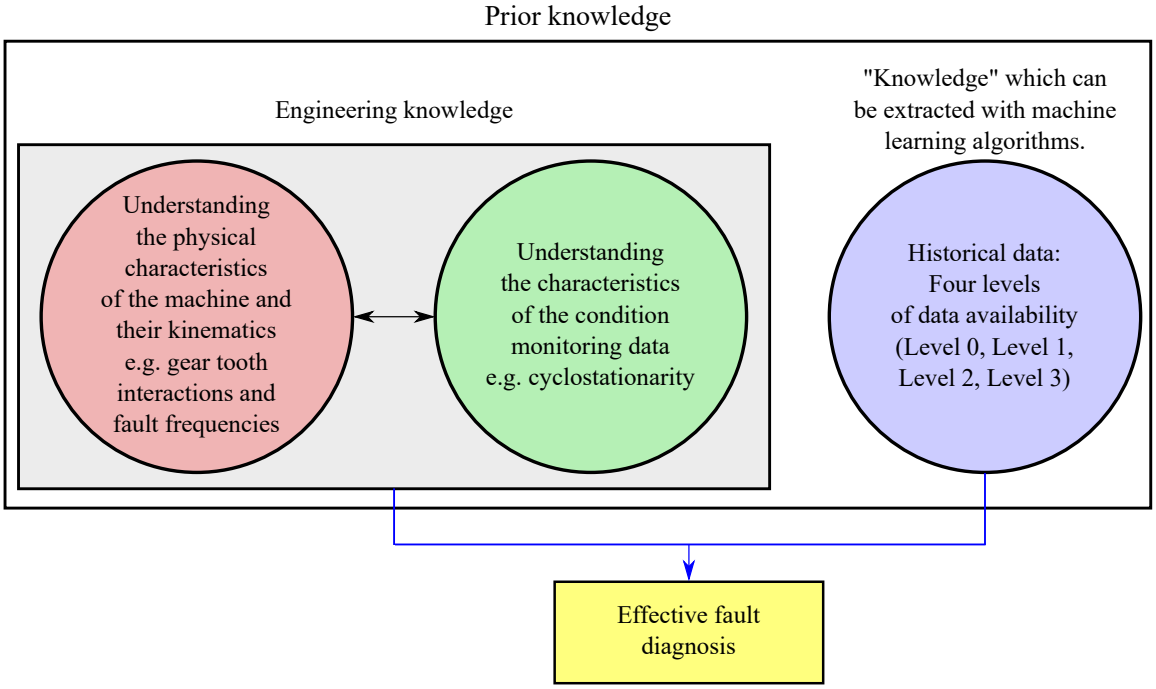


Figure 1.1: The prior knowledge that can be used for performing effective fault diagnosis. Two different types of categories are identified: Engineering knowledge obtained from understanding the dynamics of the machines and information learned by a machine learning model from the historical data. The engineering knowledge, for example, entails understanding that gear teeth interactions result in periodic phenomena, outer race bearing damage manifests at specific cyclic frequencies and gear mesh signals are first-order cyclostationary, while wear results in second-order cyclostationary vibration signals. The machine learning model can for example learn how a healthy system's response looks like from the historical data and use this information to automatically detect a deteriorating system. The different levels of historical data are discussed in Section 1.3.2 and presented in Figures 1.2 and 1.3.

The approach presented in this work, is in a sense philosophically different from deep learning, because it not only relies on models of the historical data, but also on engineering knowledge to perform fault diagnosis. Deep learning aims to overcome the need for expert diagnosticians to apply advanced signal processing techniques or to perform feature engineering in the fault diagnosis process. However, the engineering knowledge obtained from the condition monitoring literature (e.g. rotating machine signals are inherently cyclostationary) or from knowledge of the machines can be used to design methodologies that effectively

extract the diagnostic information from the data, it can result in models which require less data and less time to optimise, and it can result in predictions that are easier to interpret for a non-expert. Therefore, the focus of this thesis is developing effective fault diagnosis methodologies by using the available prior knowledge.

1.3.2 Methodologies for effective fault diagnosis

It is well-known that the gears and bearings found in gearboxes are prone to failure, with long downtimes associated with each failure [1, 6]. Therefore, in this thesis, different novel methodologies are proposed and investigated to systematically incorporate the available prior knowledge for performing effective gearbox diagnostics under varying operating conditions. This is done by firstly identifying different degrees of historical data availability and to develop methods which incorporate the available historical data for gearbox fault diagnosis. The identified degrees of historical data availability that can be encountered are summarised in Figure 1.2 and addressed separately in the subsequent sections.

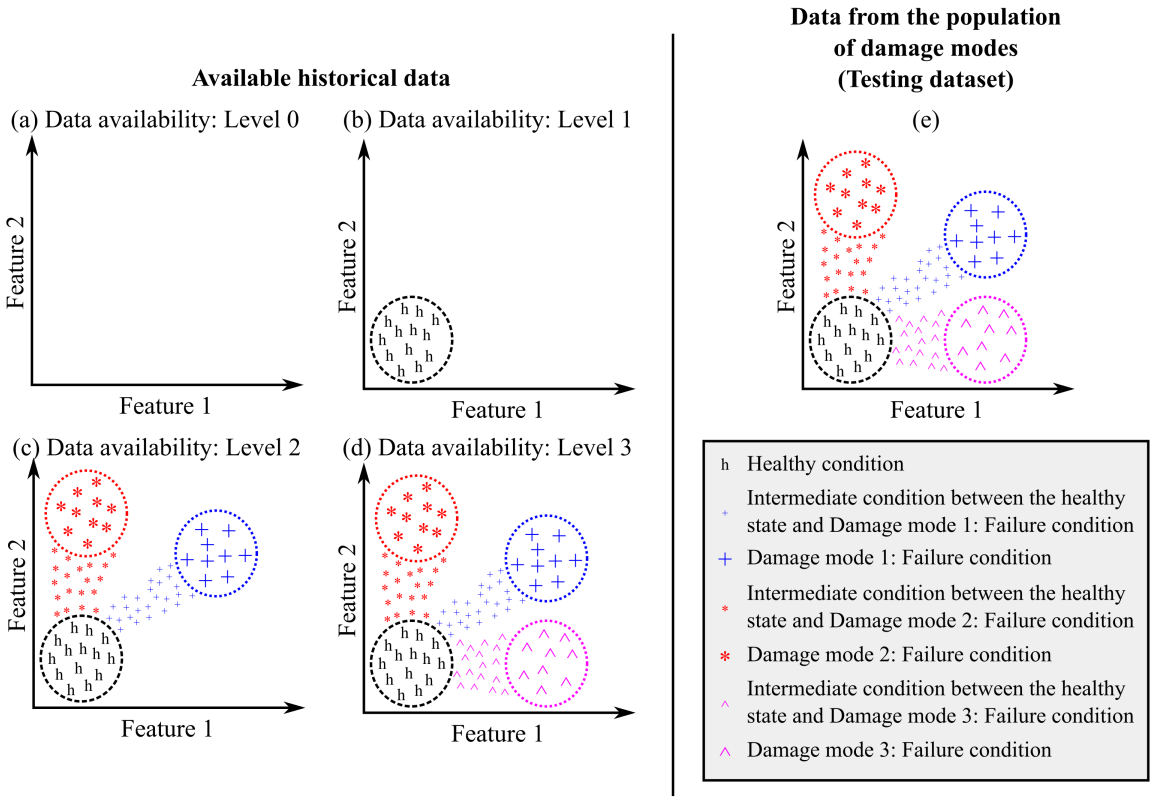


Figure 1.2: Different levels of historical data availability are illustrated in Figure 1.2(a)-(d) with respect to the population of conditions that could be encountered throughout the life of the machine in Figure 1.2(e). The data are shown in a general feature space. For example, each point in the power spectral density can be considered a dimension in the feature space.

1.3.2.1 Level 0: Historical data are not available

If a gearbox is new or if condition monitoring has not previously been performed on the gearbox, it is possible that there is no historical data available to use data-driven models for automatic condition inference. In this case, it is possible for an expert to use well developed signal analysis techniques such as cyclostationary analysis [27, 92, 94] to interrogate current data for potential damage. It is also possible to use condition indicators to detect changes in the properties of the data [53, 54]. However, by using prior engineering knowledge, it is for example possible to obtain a data-driven methodology for localised gear damage detection.

Localised gear damage is especially important to detect, because this could significantly shorten the life of the gearbox and it can be difficult to detect in industrial applications [67, 112]. It is also important to detect localised damage for gear discrepancy analysis. Discrepancy analysis, which requires healthy historical data, assumes that each portion of the gear is in the same condition i.e. no localised gear damage is present. Therefore, if localised damage is unwittingly present on the gear in the training data, it can significantly impede the performance of discrepancy analysis.

Hence, a methodology is proposed in this thesis to automatically detect the presence of localised gear anomalies due to localised damage. This is performed by exploiting the fact that interactions between healthy gear teeth result in vibration characteristics that are statistically similar. Therefore, this would not be possible without using prior knowledge of the physics of gears.

1.3.2.2 Level 1: Healthy historical data are available

Gearboxes are expected to operate for long periods in a healthy condition, while potentially operating for short periods in a damaged condition. The damaged component is replaced by applying the appropriate or available maintenance strategy e.g. time-preventative maintenance. The implication of this is that healthy condition monitoring data are easily acquired and therefore healthy historical data will have the highest availability. Hence, a large portion of this thesis is dedicated to developing gearbox fault diagnosis methodologies which utilise the available healthy historical data and to supplement the available signal analysis techniques. The developed methodologies aim to not only detect deviations from the normal condition like standard novelty detection approaches, but to also provide information on the characteristics of the damage, i.e. the damaged component as well as the extent of the

damage.

It is possible to utilise different categories of signal analysis techniques e.g. time-frequency or cyclic frequency-spectral frequency analysis and therefore separate methodologies are developed for each category. This is performed to illustrate how different signal analysis categories could be supplemented with prior knowledge such as historical data and also to ensure that this work is not constrained to a specific category of signal analysis techniques i.e. the spectral coherence and the instantaneous power spectrum are both very important for gearbox fault diagnosis. The following signal analysis categories are considered in this thesis[2]:

- Time-frequency analysis techniques such as empirical mode decomposition, wavelet analysis, and cyclostationary analysis.
- Cyclic frequency-spectral frequency representations obtained from second-order cyclostationary tools such as the spectral correlation.
- Narrowband feature extraction techniques resulting in the kurtogram and infogram for example. Time-domain analysis techniques such as the synchronous average and the synchronous variance and frequency domain analysis techniques such as envelope analysis are also considered in this work, but they are supplemented with historical data indirectly using the narrowband feature analysis techniques discussed here.

The focus is placed on the aforementioned signal analysis categories, because they are very relevant to gearbox fault diagnosis (i.e. fault detection, isolation and trending) and all can be used with historical data to obtain effective fault diagnosis techniques.

1.3.2.3 Level 2: Historical fault data of a sample of the damage modes

Even though the desired condition of the gearbox is healthy, specific damage modes from a population of possible damage modes will develop within a gearbox during its operation. Hence, it is not only important to develop a methodology that can utilise the available historical fault data, but the methodology should also acknowledge the fact that this is an open set recognition problem. Historical fault data could also become available during the operation of the machine and need to be accommodated. This is in contrast to most supervised learning techniques applied in literature for machine fault diagnosis; they do not

[2]The extensions of the categories to the angle and angle-frequency domains are implied. For example angle-frequency and angle-order analysis techniques are in the same category as time-frequency analysis techniques.

account for the fact that the problem is an open set recognition problem and the fact that historical fault data could become available during the operation of the machine.

Hence, a methodology is developed which utilises discrepancy analysis as a pre-processing technique for performing open set recognition on the available historical fault data. The open set recognition methodology is centred around discrepancy analysis, because discrepancy analysis is sensitive to the presence of damage under varying operation conditions [154–157]. However, if desired, it is possible to utilise the other techniques developed in Section 1.3.2.2 instead.

1.3.2.4 Level 3: Historical fault data of the population of the damage modes

The last level that could be encountered is if historical fault data of all potential damage modes are available. The implication of this is that this becomes a closed set recognition framework and conventional supervised learning classification techniques could be used for this task. However, this is not investigated in this thesis, because the open set recognition framework developed for the data availability of Level 2 can be used for this task and many techniques have already been developed specifically for this.

1.3.3 Summary of proposed methodologies

A number of methodologies are developed to address the different data availability levels as well as different potential signal analysis techniques. A summary of these methodologies and their relationships with the data availability are given in Figure 1.3.

1.4 Layout of thesis

The methodologies are presented in this thesis in the order of their current data availability, i.e. data availability: Level 0 precedes historical data availability: Level 1. Therefore, in Chapter 2 the localised gear anomaly detection without historical data methodology is presented and investigated on experimental gearbox data.

The methodologies for data availability: Level 1 are presented in Chapter 3, Chapter 4 and Chapter 5. In Chapter 3, a discrepancy analysis methodology for bearing diagnostics is presented and investigated on a phenomenological gearbox model, an experimental bearing dataset acquired under constant operating conditions and a experimental dataset acquired under varying speed conditions. A methodology for utilising the spectral coherence and prior

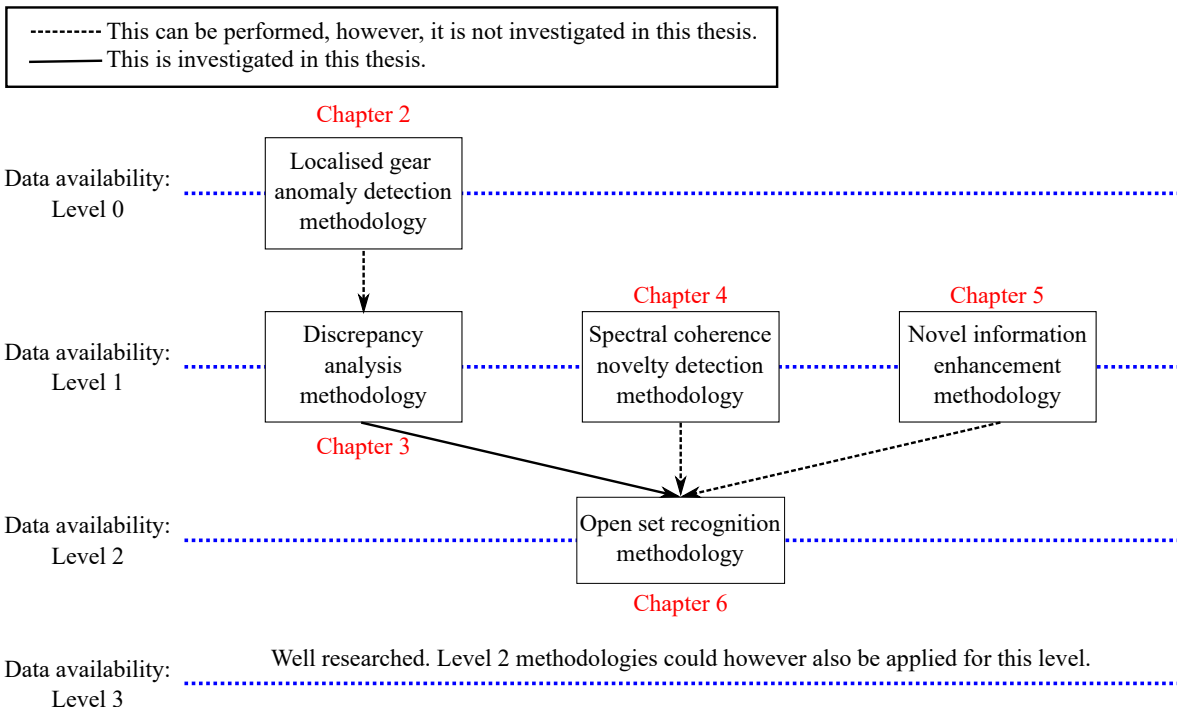


Figure 1.3: An overview is given of the methodologies presented in this thesis. The methodologies are presented with respect to the data requirement for applying the methodology, where more details on each data availability level are given in Figure 1.2. The dashed arrow between Level 0 and Level 1 reinforces that the localised gear damage methodology can be used to determine whether the training data actually corresponds to gears without localised gear damage. The arrows between Level 1 and Level 2 indicate that one of the methodologies can be used as input to the open set recognition methodology, however, only the discrepancy analysis methodology in Level 1 is used.

knowledge about the kinematics of the machines is presented in Chapter 4 and investigated on phenomenological gearbox data as well as experimental gearbox data. In Chapter 5, a methodology is presented to utilise narrowband feature extraction methods to enhance novel information in the vibration signal for more effective fault diagnosis. The methodology is investigated on phenomenological gearbox data as well as experimental gearbox data.

The open set recognition methodology for data availability: Level 2 is presented in Chapter 6. The methodology is able to incorporate historical fault data into the methodology for more effective fault diagnosis. Investigations are performed on synthetic as well as experimental gearbox datasets, where the results emphasise the importance of using an open set recognition framework for machine condition monitoring applications.

Finally, the work is concluded on in Chapter 7 of this thesis, where an overview of the methodologies is given, the contributions of the work are summarised and recommendations are made for future work.

Two appendices are included after the conclusion, where Appendix A contains additional information and results for the different chapters in this work. Appendix B contains addi-

tional information related to the phenomenological bearing model used throughout this work. References to the respective appendices are given in the relevant chapters.

The abbreviations used throughout this thesis are properly defined at its first usage and given on page viii as well. The definitions of the mathematical symbols, used in a specific chapter, are properly defined in each relevant chapter.

Chapter 2

Localised gear anomaly detection methodology

2.1 Introduction

It is important to monitor gearboxes for localised gear damage, because the localised gear damage results in increased stress concentrations which accelerate the deterioration of the gear tooth. This can also result in the condition of adjacent gear teeth and the condition of the gear teeth meshing with the damaged gear to deteriorate, because the contact during meshing is imperfect which results in portions of the gear teeth carrying larger loads. Historical data could potentially be used with a data-driven methodology such as discrepancy analysis to automatically detect the gear damage, however, historical data will not always be available. If discrepancy analysis is unintentionally optimised on data obtained from a gear with localised gear damage, it can obstruct its ability to diagnose damage.

Therefore, a novel diagnostic technique is proposed in this chapter to automatically detect localised anomalies i.e. localised segments that differ statistically from the other segments, using only the vibration data of a single measurement. This is performed by using prior knowledge that the data generated from the meshing of healthy gears should be statistically similar (i.e. generated from the same distribution) and localised damage result in data being statistically different. After the gears are identified as being healthy, it is possible to utilise more advanced approaches such as discrepancy analysis to detect localised and distributed changes in the condition of the gears. The purpose of this methodology is therefore not to replace techniques that utilise historical data such as discrepancy analysis or supervised learning approaches, but rather to address the special case where historical data are not

available.

The methodology is presented in Section 2.2, whereafter it is validated in Section 2.3 on experimental gearbox data acquired under different constant load and constant speed conditions and on experimental data acquired under time-varying operating conditions.

2.2 Methodology

The general methodology is presented in Figure 2.1(a), with the specific steps used in this chapter shown in Figure 2.1(b). For the implementation used in this chapter, it is assumed

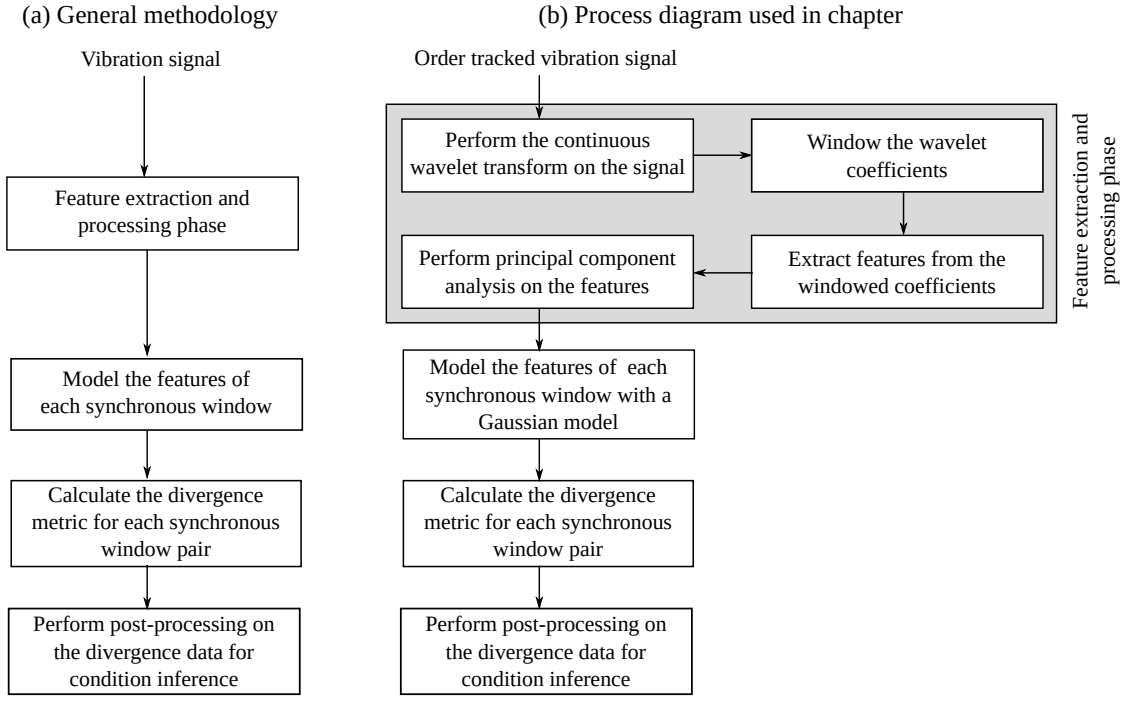


Figure 2.1: The process diagrams of the methodology and the specific implementation of the general methodology used in the chapter are shown in Figure 2.1(a) and 2.1(b).

that an order tracked vibration signal is available, which was measured over N_r shaft rotations. A gear with N_{teeth} teeth is connected to the aforementioned shaft and is investigated for potential localised damage. Machine condition features are extracted from the windowed, processed vibration signal, which allow localised changes to be detected within the signal. The processed data are windowed into N_w windows per gear revolution from which features are extracted and processed. The processed features are grouped together with other synchronous windows i.e. windows which correspond to the same angular position on the gear, whereafter the features of each synchronous window set are modelled with a probabilistic model. The windowing scheme's connection with the synchronous windows is illustrated in Figure 2.2. The windowing scheme and modelling approach result in N_w models, where each

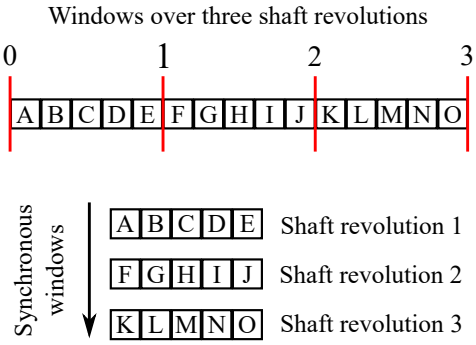


Figure 2.2: The synchronous (equal angular) windows are illustrated over three shaft revolutions ($N_r = 3$), with five windows being used per shaft rotation i.e. $N_w = 5$. The $N_r \times N_w$ windows are distinguished by different letters. The different synchronous windows are: A-F-K, B-G-L, C-H-M etc.

model captures the characteristics of the features within the respective synchronous window. The dissimilarity between the probability density functions of the models is used to generate a divergence matrix which is subsequently processed for fault detection.

The feature extraction and processing phase is considered in the next section, whereafter the modelling of the data, and the generation and processing of the dissimilarly measure for the models are discussed.

2.2.1 Feature extraction and processing

Time-frequency analysis methods are important for analysing non-stationary signals found in rotating machine condition monitoring applications. The temporal signal is transformed into a two-dimensional time-frequency plane, which makes it possible to detect time-localised changes in the frequency content of the signals e.g. bearing damage excites specific frequency bands at specific cyclic frequencies [3, 59]. The continuous wavelet transform [3]

$$\mathrm{CWT}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-b}{a} \right) dt, \tag{2.1}$$

is a time-frequency method which decomposes the signal $x(t)$ into a time-scale representation by translating and dilating a wavelet basis function $\psi(t)$ with the translation b and scale a parameters, respectively. The conjugate of the wavelet basis function $\psi(t)$ is denoted by $\psi^*(t)$. The continuous wavelet transform is well-suited for non-stationary signals, is appropriate for extracting features with diagnostic information and is sensitive to the presence of singularities [59]; this makes it a popular analysis tool for rotating machine diagnostics [3, 59, 65, 69, 70, 75]. However, the performance of the continuous wavelet transform is sensitive to the wavelet basis function that is used [69]. The Daubechies db1 wavelet is used in this chapter, due to

its good performance in the gear and bearing fault detection field [65, 69].

The gear mesh frequency contains diagnostic information and therefore the continuous wavelet transform is applied at the scales situated around the fundamental gear mesh frequency of the gearbox and its four harmonics. Twenty scales at a bandwidth of $3 \times k$ orders around the gear mesh frequency are selected, with $k = 1$ for the gear mesh frequency, $k = 2$ for its harmonic, etc. Each scale is windowed with rectangular windows into N_w segments per rotation, from which the Root-Mean-Square (RMS) of the windowed wavelet coefficients is calculated in a similar procedure as Heyns et al. [156]. The window length is set to $360/N_{teeth}$ degrees with a 0 degree overlap between consecutive windows, which results in a window for each gear tooth in each shaft rotation (i.e. $N_w = N_{teeth}$). It is expected that the vibration signal segments, associated with the portions of the gear that are in the same condition, are similar and therefore the extracted features will be statistically similar.

The RMS features, extracted from the 20 scales at the five gear mesh frequencies, result in a 100 dimensional feature space with a total of $N_r \times N_w$ observations and N_r synchronous observations at each window. The dimensionality of the feature space and the small number of synchronous observations make the model optimisation process susceptible to overfitting. Principal component analysis, a linear dimensionality reduction technique [114], is used to transform the original feature space to a lower dimensional feature space, while retaining most of the information content. Principal component analysis has performed very well compared to other non-linear dimensionality methods for gear fault diagnosis and is therefore well-suited for this application [77]. The accumulative contribution rate [83] is used to select the appropriate dimensionality of the new feature space to ensure that the information loss is minimal in the dimensionality reduction process.

2.2.2 Kullback-Leibler divergence

The Kullback-Leibler (KL) divergence, a subset of f-divergence [158], is very popular in the fault detection field [150, 151, 159–161] because of its sensitivity to changes in the density of the data [150]. The KL divergence [114]

$$KL(p_i||p_j) = -\int_{\boldsymbol{x}} p_i(\boldsymbol{x}) \log\left(\frac{p_j(\boldsymbol{x})}{p_i(\boldsymbol{x})}\right) d\boldsymbol{x}, \tag{2.2}$$

describes the dissimilarity between the probability density function p_i and p_j, associated with synchronous window i and j respectively, over the data space $\boldsymbol{x}$. If p_i and p_j are the

same, it means that the data in window i and j are from the same distribution and then $KL(p_i||p_j) = 0$, otherwise $KL(p_i||p_j) > 0$. The KL divergence has for example been used to detect gear damage [150, 159], distillation process anomalies [150, 151] and electrical motor damage [160]. Liu et al. [158] found that the symmetrised Pearson divergence outperforms the forward and the backward Pearson divergence for change point detection and therefore the symmetrised KL divergence

$$D_{ij} = KL(p_i||p_j) + KL(p_j||p_i), \tag{2.3}$$

over window pair i and j is used instead of the KL divergence given by Equation (2.2). Equation (2.3) is symmetric in the probability density functions p_i and p_j, which means that $D_{ij} = D_{ji}$ for any i and j. The symmetrised KL divergence is used to generate a divergence matrix over the gear, denoted by $\boldsymbol{D}$, which is a set of divergence measures over all synchronous window combinations. The divergence data of window i is represented by $\boldsymbol{d}_i \in \mathbb{R}^{N_w \times 1}$, which is used to construct the divergence matrix over all window combinations $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \dots \boldsymbol{d}_{N_w}]$, i.e. $\boldsymbol{D} \in \mathbb{R}^{N_w \times N_w}$.

The KL divergence only has a closed form solution for special probability density functions such as Gaussian distributions. If the datasets are non-Gaussian, the density ratio $p_j(\boldsymbol{x})/p_i(\boldsymbol{x})$ can be modelled using kernel density estimators to calculate the KL divergence [150, 158]. Ferracuti et al. [160] used kernel density estimators to model the probability density functions, whereafter the discrete form of the KL divergence was adopted. Equation (2.2) can also be evaluated using Monte Carlo integration, however it is very computationally expensive when performed over all window pair combinations i and j for each investigated dataset.

The objective is to find an efficient procedure to calculate the $N_w\left(N_w+1\right)/2$ unique divergence values for each matrix $\boldsymbol{D}$ of a specific measurement. It is assumed that the processed features are described sufficiently well by a multivariate Gaussian distribution. This results in the following closed form solution for Equation (2.3) [71]

$$D_{ij} = \frac{1}{2}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right)^T \left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right) + \frac{1}{2}\text{trace}\left(\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i^{-1}\boldsymbol{\Sigma}_j - 2\boldsymbol{I}\right), \tag{2.4}$$

where the mean and the covariance matrix of the processed features, associated with synchronous window i, are denoted by $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ respectively and the identity matrix is denoted by $\boldsymbol{I}$. The Jarque-Bera test, available in Reference [162], is used later in this chapter to deter-

mine whether the Gaussian assumption is correct. If the data are non-Gaussian distributed, more flexible models such as Gaussian mixture models [114] or density ratio estimation approaches [150, 158] can be used. However, due to the limited number of observations N_r in the dataset, it can be difficult to properly estimate and motivate the values of the hyperparameters of the more sophisticated modelling approaches.

In Figure 2.3, divergence matrices are presented for a healthy gear and a gear with localised damage; the matrix of each dataset is calculated by using Equation (2.4) over all window combinations. The divergence matrices are calculated from vibration data that was acquired during the gear fatigue experiments discussed in Section 2.3. The symmetry in the

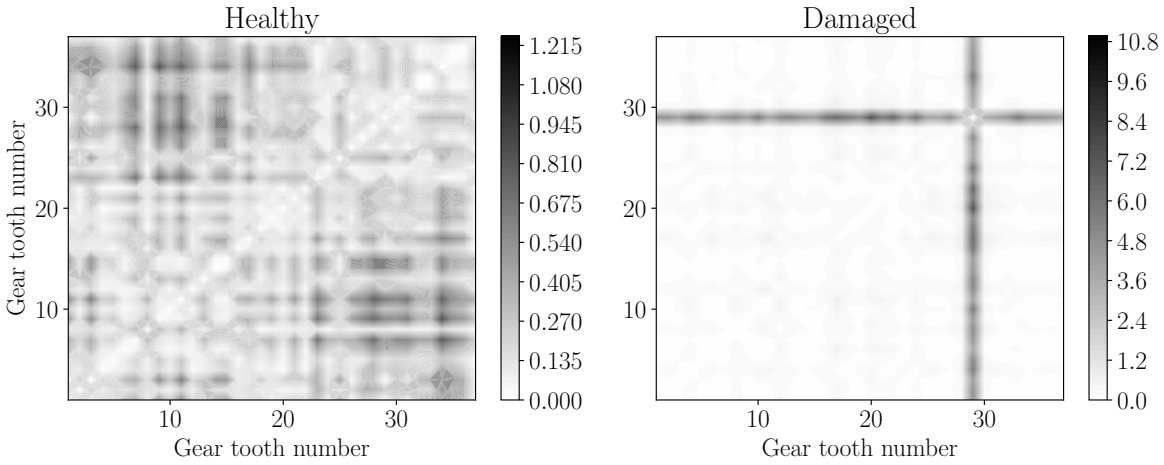


Figure 2.3: Divergence matrix data $\boldsymbol{D}$ for a healthy gear and a gear with localised damage obtained during experiments discussed in Section 2.3. Please note that different scales are used for the two plots and that a helical gear with 37 teeth is considered.

distribution can easily be seen, with the diagonal having a divergence of zero due to the properties of the KL divergence. The divergence data of the healthy gear differ slightly for each window, with small differences indicating that the teeth are approximately in the same condition. In the presence of localised damage, the data associated with the damaged tooth differ significantly from the data of the other healthy teeth, which results in dissimilar models that are easily identified in the image of the divergence matrix. The extent of the localised gear damage can be inferred from the techniques proposed in the next section.

2.2.3 Data analysis for automatic condition inference

The data in Figure 2.3 are shown in Figure 2.4(i) over a full gear rotation for a healthy and a damaged gear. The healthy gear divergence data are relatively close to one another when considering the same tooth, with some slight variations being present between different

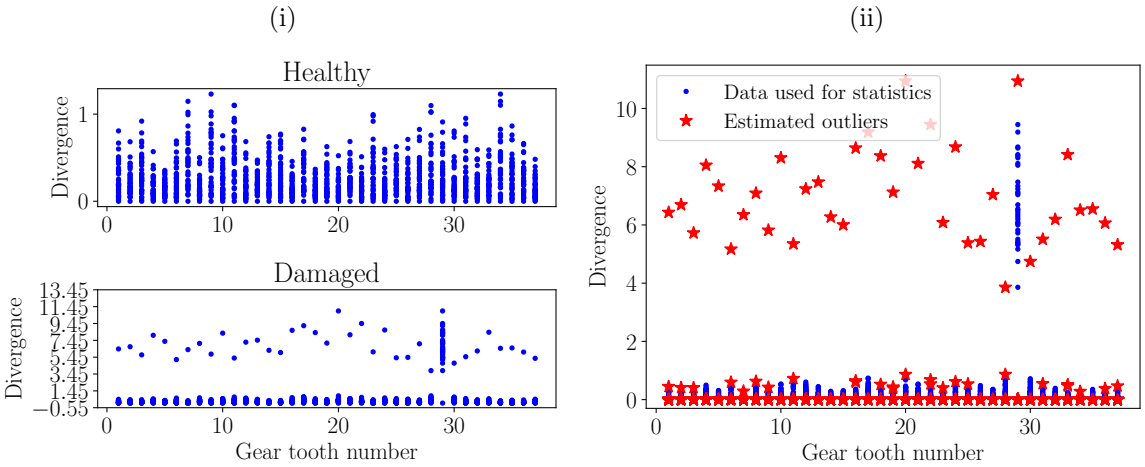


Figure 2.4: The synchronous divergence data are shown over a gear rotation for the healthy and the damaged gear in 2.4(i). The outliers are identified and shown in 2.4(ii) for the damaged gear using the proposed outlier removal process. The gear under consideration has 37 teeth.

teeth. The divergence data of the healthy portion of the damaged gear contains very similar characteristics to those of a healthy gear, with the exception that a large outlier is present. This outlier is due to the presence of localised damage on the gear, which manifests over all gear teeth as a result of the properties of $\boldsymbol{D}$ seen in Figure 2.3.

The outliers caused by localised damage and the zero divergence when $i = j$ in Equation (2.4) have to be removed from the dataset, given that they do not reflect the true condition of the portion of the gear associated with the window. The outliers in the divergence data, due to the presence of localised damage, are automatically estimated in a three step process; this is performed separately on the data of each window and is as follows for the data $\boldsymbol{d}_i$ of window i:

1. Calculate the $\upsilon/2$th and the $(100 - \upsilon/2)$th percentiles of $\boldsymbol{d}_i$, and momentarily only retain the data in between those percentiles. The retained divergence data in between the percentile bounds are denoted by $\tilde{\boldsymbol{d}}_i$.

2. Calculate the mean μ_i and the standard deviation σ_i of the retained samples of $\tilde{\boldsymbol{d}}_i$.

3. Subsequently use $r_j = \left(\frac{D_{ij}-\mu_i}{k_r\sigma_i}\right)^2$ for each $j \in [1,\ N_w]$, to determine whether the data point D_{ij} is an outlier $r_j > 1$ or not $r_j \leq 1$. The n data points, which are not removed in this process are denoted by $\boldsymbol{d}_j^{proc}$ and the data are subsequently used to perform condition inference.

The parameters υ and k_r, used in the outlier removal process, are equal to 5.0 and 3.0

respectively in this chapter. It should be noted that after the outlier removal process, each window may not have the same number of data points left and cannot strictly be written as a matrix. However, the processed divergence data are denoted by $\boldsymbol{D}^{proc}$ for the sake of notational simplicity where the superscript *proc* highlights that the processed data are used. It is important to emphasise that other outlier removal processes can be used if desired.

The outliers that are present in the damaged gear divergence data, shown in Figure 2.4(i), were identified with the proposed method and shown in Figure 2.4(ii). The benefits of the outlier removal process are shown in Figure 2.5 on the point estimates of the data. It is observed that the point estimates calculated from the processed data are significantly better than the raw data, because it has a smaller variance and a mean without a bias. Hence, the processed data can lead to more robust results.

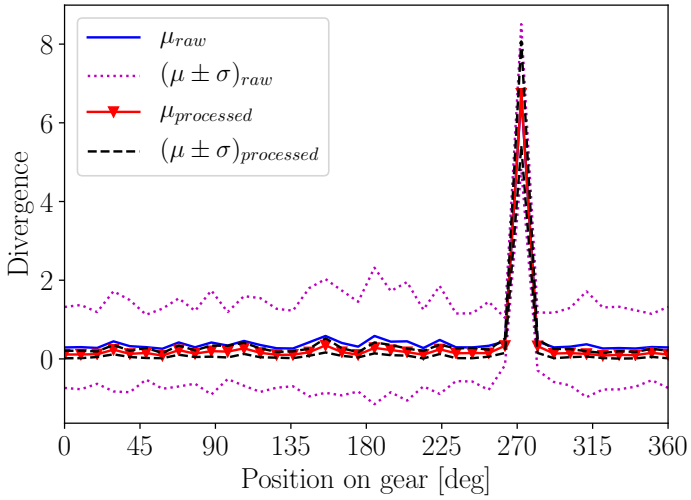


Figure 2.5: The point estimates of the processed data (using the outlier removal process) and the raw (unprocessed) data are compared for a gear with localised damage.

An automatic method is required to infer the condition of the gear i.e. to detect localised damage and evaluate changes in the extent of the damage. Bayesian analysis techniques [163, 164] are employed in this chapter to perform this task. Bayesian analysis techniques use probability densities to quantify the uncertainty in the random variables, which is subsequently used to make inferences on the random variables. It is assumed that the mean μ and the precision λ or the reciprocal of the variance of the data, of a specific window that is under consideration, are unknown and need to be inferred. The data are denoted by X in this section, because two distinct datasets are investigated with the approach outlined here. The prior distribution, incorporates prior knowledge of the parameters into the inference process

and is chosen to be a Normal-Gamma distribution in the form of

$$p(\mu, \lambda|\mu_0, \kappa_0, \alpha_0, \beta_0) = \text{Gaussian}\left(\mu|\mu_0, (\kappa_0\lambda)^{-1}\right) \text{Gamma}(\lambda|\alpha_0, \beta_0), \tag{2.5}$$

where μ_0, κ_0, α_0, β_0 are the unknown hyperparameters of the prior distribution. The data X are assumed to be independent identically distributed, where an individual data point is represented by a Gaussian probability density function with mean μ and precision λ. The choice of prior and likelihood function results in a Normal-Gamma posterior distribution in the form of

$$p(\mu, \lambda|X, \mu_0, \kappa_0, \alpha_0, \beta_0) = \text{Gaussian}\left(\mu|\mu_n, (\kappa_n\lambda)^{-1}\right) \text{Gamma}(\lambda|\alpha_n, \beta_n), \tag{2.6}$$

where [165]

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{x} \tag{2.7}$$

$$\kappa_n = \kappa_0 + n \tag{2.8}$$

$$\alpha_n = \alpha_0 + \frac{n}{2} \tag{2.9}$$

$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{1}{2}\frac{(\bar{x} - \mu_0)^2}{\frac{1}{n} + \frac{1}{\kappa_0}}, \tag{2.10}$$

are the parameters of the posterior distribution with n being the number of observations, and $\bar{x}$ being the sample mean of the data X of the considered window. The precision λ is a nuisance parameter and is marginalised out to obtain the posterior marginal distribution of the mean [114, 165]

$$p(\mu|X) = \int_{\lambda} p(\mu, \lambda|X)d\lambda \tag{2.11}$$

$$= \text{Student}_{2\alpha_n}\left(\mu \middle| \mu_n, \frac{\beta_n}{\alpha_n\kappa_n}\right), \tag{2.12}$$

which is in the form of the Student-t distribution, with $2\alpha_n$ degrees of freedom, a variance of $\beta_n/(\alpha_n\kappa_n)$ and a mean of μ_n. The dependence on the fixed hyperparameters are neglected for future notational simplicity. The mode and mean of the posterior marginal distribution of the mean, presented in Equation (2.11), are the same. The marginal distribution in Equation (2.11) is used to infer the condition of the gear, by calculating the probability that the

posterior mean exceeds an alarm threshold i.e.

$$P(\mu > \text{threshold}|X) = \int_{\text{threshold}}^{\infty} \int_{\lambda} p(\mu, \lambda|X) d\lambda d\mu, \tag{2.13}$$

where the preselected threshold indicates the region where the divergence is far from normal behaviour. The Bayesian data analysis approach is used on the divergence data of each window and it is used to calculate the sensitivity of the divergence matrix to machine condition changes as well. It should be noted that it is possible to use other analysis techniques to infer the condition of the gears, however some of the benefits of Bayesian data analysis techniques that are highlighted later in this chapter, will be lost.

2.3 Validation

The proposed methodology is validated in two experimental investigations in this chapter. The experimental setup, used to generate the data, is presented and discussed in the next section whereafter the results of the investigations are presented.

2.3.1 Experimental setup

An experimental setup, designed by Stander and Heyns [19], was refurbished to conduct gear fatigue tests under fluctuating operating conditions. The vibration data were acquired from the experimental setup in Figure 2.6, which consists of an alternator, three helical gearboxes and an electrical motor. The instantaneous load applied by the alternator and the instantaneous rotational speed applied by the electrical motor, were controlled with a personal computer. The axial acceleration from a 100mV/g tri-axial accelerometer, located

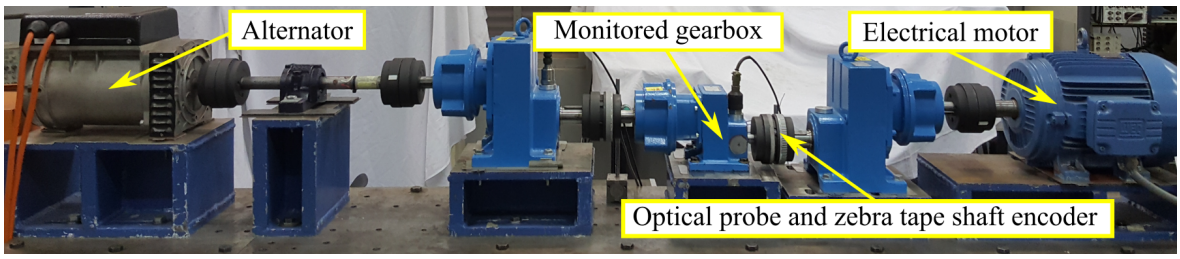


Figure 2.6: The experimental setup.

on the bearing housing of the monitored helical gearbox, is used in the condition monitoring process. The acceleration vibration signal was sampled at 25.6kHz and the optical probe signal was sampled at 51.2kHz with an OROS OR35 data acquisition device, where the

optical probe is used with an 88 pulse per revolution zebra tape shaft encoder to calculate the instantaneous angular speed of the input shaft of the monitored gearbox. Data were acquired from a healthy gearbox, whereafter the gearbox was disassembled so that localised damage could be seeded on the gear and thereafter the gearbox was reassembled. The gearbox, with the damaged gear shown in Figure 2.7(i), was subsequently operated for approximately twenty days under the operating conditions depicted in Figure 2.8, until the damaged tooth finally failed. During the fatigue experiments, approximately 1400 measurements were taken, with a smaller subset of the complete dataset used in this investigation. The gear with the

(i) Before fatigue experiment (ii) After fatigue experiment

Figure 2.7: The gear with seeded damage before and after the fatigue experiment was completed.

broken tooth (i.e. after the fatigue experiment was completed) is shown in Figure 2.7(ii).

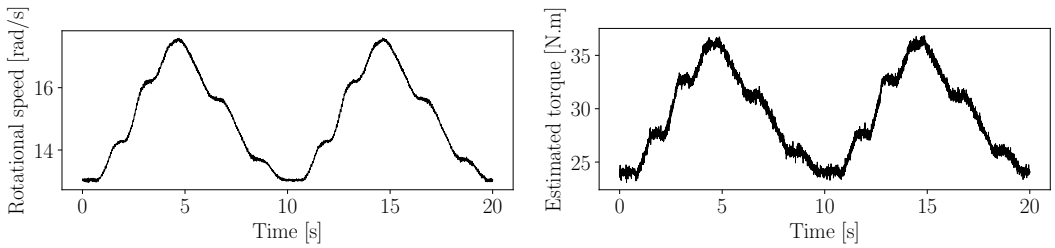


Figure 2.8: The operating conditions during the fatigue experiment at the input shaft of the monitored gearbox.

2.3.2 Investigation 1

In the first investigation, the effectiveness of the methodology is investigated and compared to conventional fault diagnosis techniques on gearbox data obtained from three cases with

different constant operating conditions. This investigation is performed to validate the robustness of the proposed methodology to detect localised anomalies in the data under different operating conditions.

The load and rotational speed properties of the three cases are presented in Table 2.1 and the operating conditions are present for measurements taken for a gearbox with a healthy gear and with a gear without a tooth (see Figure 2.7(ii)), respectively.

Table 2.1: Operating conditions at the input shaft of the monitored gearbox for investigation 1. These operating conditions were present for the healthy and the damaged experiments.

	Load [N.m]	Speed [rad/s]
Case 1	0	7.48
Case 2	32.48	14.81
Case 3	38.31	14.83

2.3.2.1 Conventional fault diagnosis techniques

The Synchronous Average (SA) of the vibration signal has been successfully used for gear fault diagnosis [19] and is investigated on the signals described in Table 2.1. The synchronous averages of the investigated computed order tracked vibration signals are presented in Figure 2.9(i). The amplitudes of the synchronous averaged signals in Figure 2.9(i) change signifi-

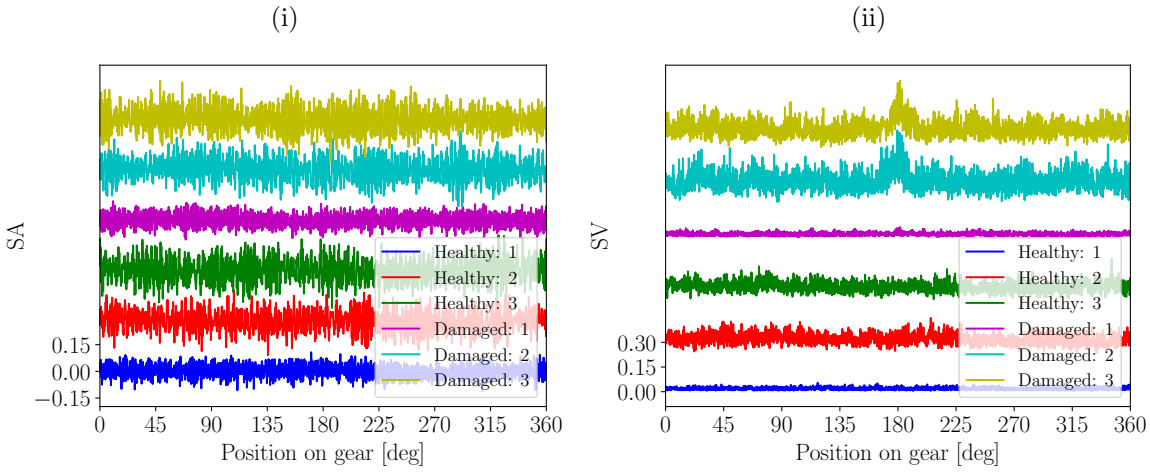


Figure 2.9: The Synchronous Average (SA) of the computed order tracked vibration signals of the gears and the Synchronous Variance (SV) of the gears in different conditions are compared for the operating conditions in Table 2.1. Each signal is given a unique offset to compare the results, with the damage being present at 180 degrees for the damaged gear cases. The legend indicates <Condition>: <Operating condition number>.

cantly with changes in operating conditions for gears in the same condition, with evidence of the broken tooth not being seen. Hence, it is difficult to distinguish between changes

in operating conditions and changes in machine condition utilising the synchronous average for fault diagnosis under different operating conditions. This also makes it difficult to set a threshold for automatic fault diagnosis; the gear damage is not evident and the data are significantly influenced by the different operating conditions.

The residual signal i.e. the departure of the vibration signal from the average deterministic vibration components can be used for more effective fault diagnosis [24, 86]; for example, it has performed significantly better than the conventional synchronous average for gear crack detection [24]. The Synchronous Variance (SV) i.e. the synchronous average of the squared residual signal is a second-order cyclostationary technique that has been used for gear and bearing fault diagnosis [23] and is displayed in Figure 2.9(ii). The SV is estimated with the procedure used in Ref. [23]. The gear damage is seen for operating condition Case 2 and Case 3, with the gear damage at Case 1 not being seen because no load is applied to the system. However, the results are also dependent on the operating conditions i.e. the variance of the synchronous variance is operating condition dependent and the damage is not very prominent for Case 2 and Case 3; the latter result is attributed to the fact that the helical gears have high contact ratios which mitigates the influences of the missing tooth on the gear mesh stiffness.

The Power Spectral Density (PSD) of the computed order tracked vibration signals is presented in Figure 2.10 similarly to the results in Figure 2.9. The variation of the gear

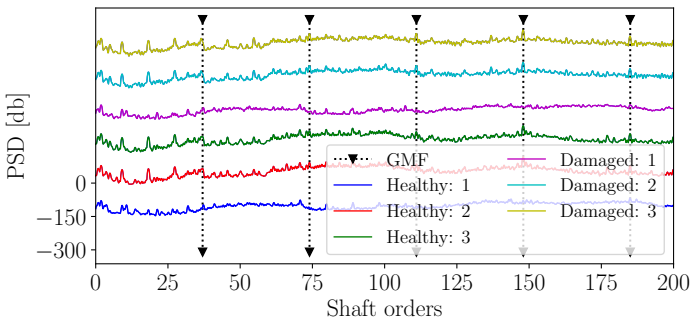


Figure 2.10: The Power Spectral Density (PSD) of the order tracked vibration signal for the different datasets described in Table 2.1.

mesh components with respect to operating condition changes is clearly observed, however, the differences between the gear conditions are not prominent. The results corroborate the observation that it is not easy to detect the damage using conventional techniques with the different operating conditions adversely influencing the fault diagnosis task. It is also difficult to assign a threshold to automatically detect the presence of localised damage with

the aforementioned techniques, because the data are operating condition dependent and the damage is not prominent in the processed signals.

2.3.2.2 Proposed methodology

Hence, a more sophisticated approach needs to be used to automatically infer the condition of the machine and therefore the proposed methodology is investigated on the same dataset. For each measurement considered in this investigation, the following procedure is followed: The vibration signal is order tracked using the optical probe and zebra tape shaft encoder whereafter the features are extracted and processed as described in Section 2.2.1. The new dimensionality of the principal component feature space is selected as two, because the corresponding accumulative contribution rate is approximately 95%. A multivariate Gaussian model is optimised on the principal component features of each set of N_w synchronous features, whereafter the divergence is calculated using Equation (2.4) for all synchronous window combinations. The outliers in $\boldsymbol{D}$ are removed with the procedure described in Section 2.2.3 to obtain the processed divergence matrix $\boldsymbol{D}^{proc}$, which is used in subsequent analyses.

Bayesian data analysis is used in the condition inference process as described in Section 2.2.3. The posterior distribution is a compromise between the prior information and the evidence obtained from the investigated dataset [114, 163]. The prior distribution conveys the prior beliefs in the data and are governed by a set of hyperparameters. The appropriate hyperparameters are unknown for the data under investigation and are set to $\mu_0 = 0.0$, $\kappa_0 = 0.01$, $\alpha = 10$, $\beta = 0.01$ so that the inferred posterior marginal distribution of the mean is dominated by the data. The consequences of the selected hyperparameter values are critically investigated in Appendix A.1.

It is necessary to specify an alarm threshold to infer the condition using Equation (2.13). If historical data of a healthy machine are available, it is possible to set an alarm threshold to perform automatic novelty detection and if historical fault data are available it is even possible to set an alarm threshold for the different stages of degradation. However, the latter is not possible in the absence of historical data and a different strategy is therefore required to perform this task.

The following procedure is used to estimate a threshold using only the data from a single measurement: It is assumed that a large portion of the gear is in the same condition and potentially a small portion of the gear is damaged i.e. localised damage is present. A threshold needs to be estimated from statistics that are robust to outliers due to localised gear damage,

therefore the median and the median absolute difference are used as opposed to the mean and the standard deviation to calculate the threshold

$$\text{threshold} = \text{median}\left(\boldsymbol{D}^{proc}\right) + k_{thres} \cdot \text{median}\left(\boldsymbol{D}^{proc} - \text{median}\left(\boldsymbol{D}^{proc}\right)\right), \tag{2.14}$$

where $\text{median}\left(\boldsymbol{X}\right) \in \mathbb{R}$ is the median of the data $\boldsymbol{X}$ and k_{thres} is a threshold factor and selected as five. Hence, it is possible to set a threshold with Equation (2.14) using only the current data under consideration i.e. historical data are not required.

The Posterior Mode of the Marginal Distribution of the Mean (PMMDM), for the healthy and the damaged gear divergence data of the three load cases, is superimposed in Figure 2.11 with the alarm threshold calculated with Equation (2.14). The alarm threshold for each dataset was approximately the same and therefore only their average is presented to keep the figures uncluttered. The results in Figure 2.11(i) indicate that in the absence of localised gear

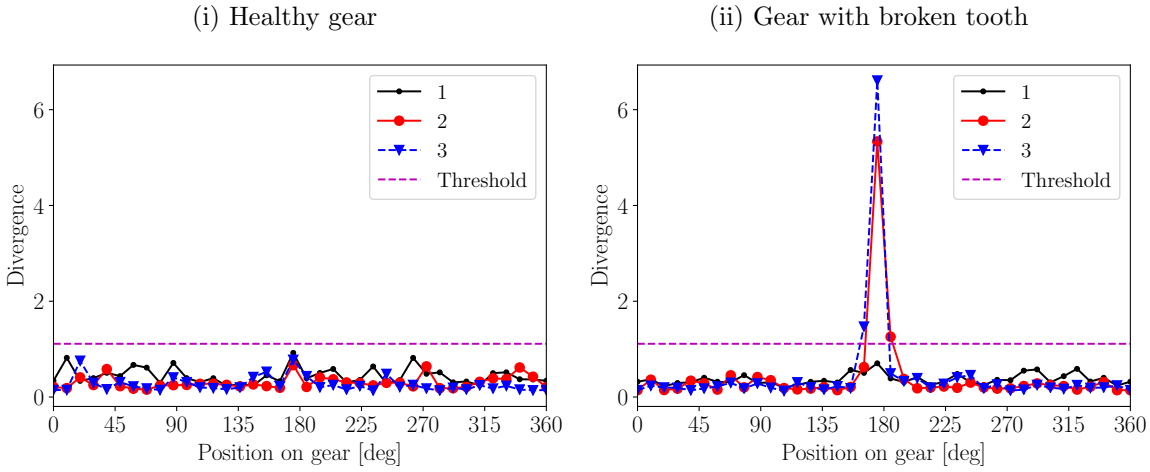


Figure 2.11: The Posterior Mode of the Marginal Distribution of the Mean (PMMDM) is shown for the healthy gear and the gear with a broken tooth for the operating condition cases described in Table 2.1. Only the average threshold of the different measurements is shown to ensure that the figure is not cluttered.

damage, the divergence does not exceed the threshold. It is clearly possible from the results in Figure 2.11(ii) to detect localised gear damage using the proposed technique when loads are applied to the system; in the absence of loads acting on the system, the damage is not detected. The magnitude of the divergence associated with the broken tooth varies slightly with changes in load and speed due to the fact that the physical impacts resulting from the missing tooth varies with operating conditions. The aforementioned results are reasonable, given the fact that helical gearboxes are investigated. Helical gears, that have large contact

ratios, will always have one tooth in contact, even if a tooth is missing. Hence, in the absence of loads, the impulses due to a missing tooth will be very small and may be undetected in many circumstances.

It is necessary to consistently compare the divergence data to the alarm threshold to perform automatic condition inference. This is performed by calculating the probability that the posterior mean is larger than the threshold using Equation (2.13) for each tooth. The results for the data in Figure 2.11 are shown in Figure 2.12. The results corroborate the

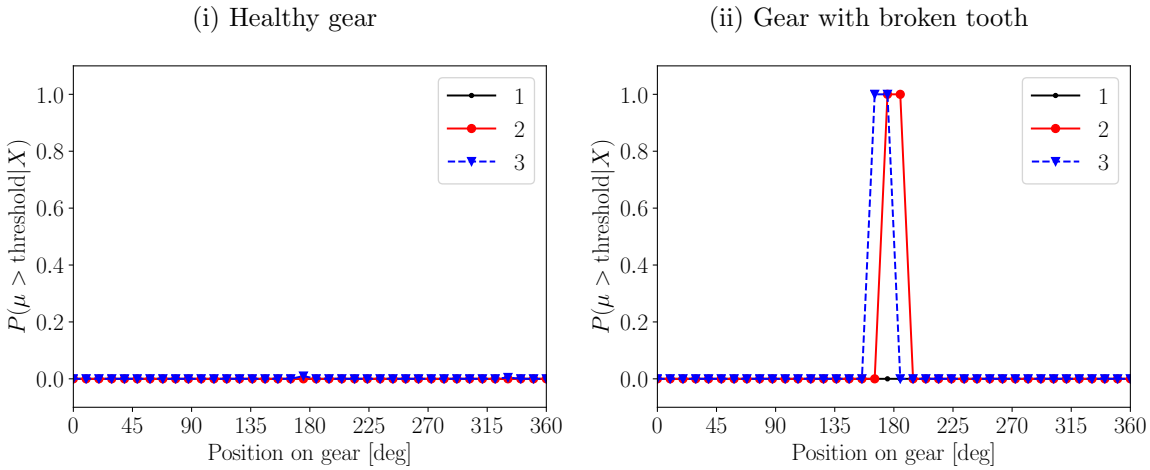


Figure 2.12: The probability that the posterior mean exceeds the threshold is calculated with Equation (2.13) for the data in Figure 2.11. The threshold, calculated for a specific measurement with Equation (2.14), is used with Equation (2.13) to infer the condition of the specific measurement i.e. the average threshold shown in Figure 2.11 is not used in the calculation procedure.

condition inferred from the results in Figure 2.11. The benefit of using the representation in Figure 2.12 is that it is easy to understand for a non-expert and the damage can clearly be seen.

The Jarque-Bera test [162] was performed on the features and it indicated that the features are not Gaussian distributed. Hence, even though the data are not Gaussian distributed, the developed technique detected the localised damage on the gear. Hence, the KL divergence detected a change in the mean and the covariance matrix of the data as the condition changed, even though the density itself is not well represented by a Gaussian distribution. More appropriate models and techniques could be used to possibly increase the sensitivity of the divergence matrix to damage, at the cost of an increase in computational time and it requires the hyperparameters to be estimated; which can be difficult if a single measurement is used.

It is evident from the results in Figures 2.11 and 2.12 that the proposed methodology

performs significantly better when compared to the results of the conventional fault diagnosis techniques in Figures 2.9 and 2.10; the damage is significantly more prominent, it is possible to automatically infer the condition of the gears and the healthy portions of the gear are unaffected by the different operating conditions.

2.3.3 Investigation 2

In the second investigation, the performance of the proposed method is investigated and compared to conventional fault diagnosis methods for the gearbox fatigue dataset described in Section 2.3.1. Conventional fault diagnosis methods are investigated in Section 2.3.3.1, the methodology's fault detection and localisation capabilities are considered in Section 2.3.3.2, whereafter the damage trending potential of the methodology is illustrated in Section 2.3.3.3. The data considered in this section were acquired under the operating conditions presented in Figure 2.8 and the same properties as in Section 2.3.2, such as the dimensionality of the feature space and the values of the hyperparameters are used.

2.3.3.1 Conventional fault diagnosis techniques

The synchronous averaged order tracked vibration signals and the synchronous variance of the residual signals are presented in Figure 2.13(i) and Figure 2.13(ii), respectively. The

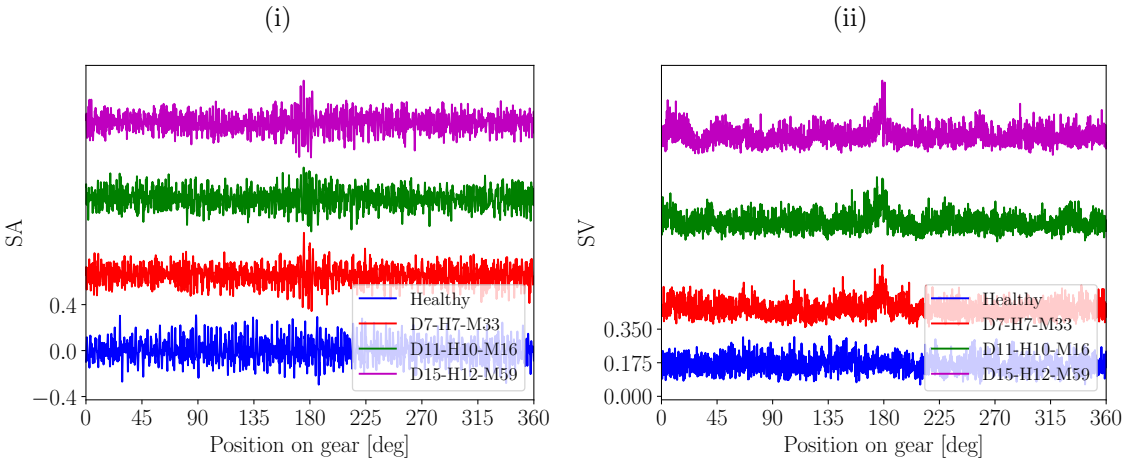


Figure 2.13: The Synchronous Average (SA) and the Synchronous Variance (SV) of the residual signal are compared for different datasets. The legend indicates the condition of the gearbox as well as the time from the start of the damaged experiment in Day-Hour-Minute format. Each signal is given a unique offset to make the comparison between the results easier, with the localised damage being present at 180 degrees.

residual signal, used to calculate the SV, is calculated by subtracting the generalised synchronous average from the synchronous average; the generalised synchronous average is a

better estimation of the periodic part under varying operating conditions than the conventional synchronous average [41]. A more detailed overview of its calculation is given in Section 4.2.1. The damage is seen in the synchronous average as well as the synchronous variance for the damaged vibration signals. Due to the fact that the operating conditions are the same for each dataset in this section, the influence of varying operating conditions on the data is not evident. The PSDs of the order tracked vibration signals are presented in Figure 2.14 for the measurements investigated in Figure 2.13 as well. It is difficult to detect the presence of

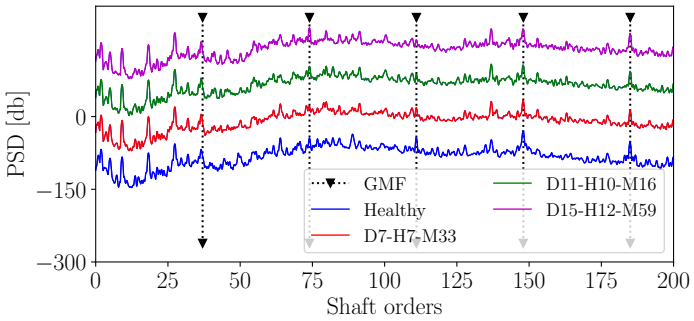


Figure 2.14: The Power Spectral Densities (PSDs) of the order tracked vibration signals, investigated also in Figure 2.13, are shown. The Gear Mesh Frequency (GMF) and its harmonics are also incorporated into the plot. Each signal is given a unique offset to make the comparison between the results easier.

localised gear damage in the PSD; it is even difficult when comparing the damaged gearbox data to the healthy gearbox data. It is only possible to observe changes in the harmonics of the gear mesh frequency for the damaged gearbox as the gearbox deteriorates.

It is also possible to compare the trending capabilities of conventional metrics and the metrics obtained from the proposed methodology i.e. the sensitivity of metrics to changes in condition. The kurtosis, which is an indication of the impulsiveness of the vibration signal, and the RMS, which is an indication of the average energy in the signal, can be used for fault trending and are compared in Figure 2.15 on the fatigue data i.e. as the gear deteriorated from Figure 2.7(i) to Figure 2.7(ii). The healthy gear data are also inserted at the start of

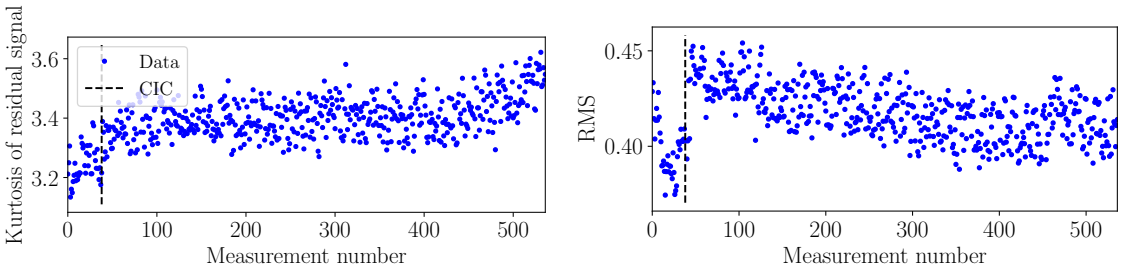


Figure 2.15: The kurtosis of the residual signal and the Root-Mean-Square (RMS) of the vibration signal are presented as function of measurement number.

the measurement number, with the healthy gearbox and the gearbox with the seeded damage separated with the Change In Condition (CIC) vertical line. A prominent change in the RMS is observed at the CIC line, which is attributed to the disassembling-reassembling procedure to damage the gear. Thereafter, the RMS decreases with measurement number. The kurtosis increased slightly at the CIC line, whereafter it remains constant between measurement number 100 and 450. A change in the kurtosis and the RMS is observed at the 480th measurement when the gear failed, however, the deterioration of the gear is not prominent from the data.

Hence, the conventional analysis techniques can be used for fault detection, fault localisation and fault trending by manually investigating the data. However, the results are not ideal e.g. the damage is not very prominent in the synchronous averages and the deterioration of the gear is not clearly seen in the trended results. Therefore, more robust metrics are required for automatic condition inference.

2.3.3.2 Automatic fault detection and localisation without historical data

The proposed methodology is implemented with exactly the same procedure as Section 2.3.2. The PMMDM is shown in Figure 2.16(i) for different time stamps, in day-hour-minute format, from the start of the fatigue experiment with the damaged gear. The PMMDM, calculated

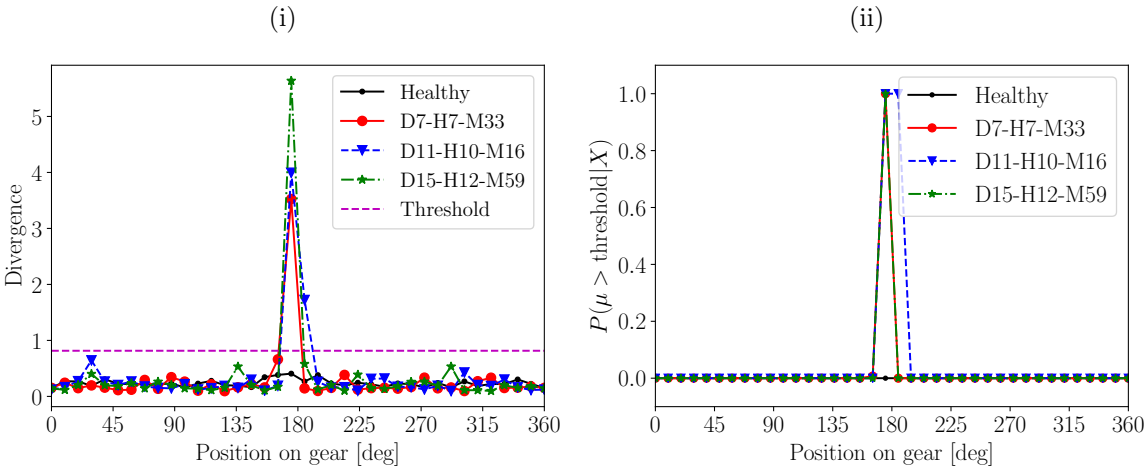


Figure 2.16: The Posterior Mode of the Marginal Distribution of the Mean (PMMDM) of the divergence for each window and the averaged threshold for the measurements are shown in Figure 2.16(i). The probability that the posterior mean of a specific measurement exceeds the threshold of the specific measurement is shown in Figure 2.16(ii). The legend indicates the day-hour-minute after the fatigue experiment was started with the damaged gear.

from Equation (2.11) for the processed divergence of the healthy data, does not exceed the

threshold; while the PMMDM of the damaged gear data exceed the threshold where the localised damage is situated (approximately 180 degrees).

The probability that the posterior mean exceeds the threshold, is presented in Figure 2.16(ii) and the results support the conclusions drawn from the PMMDM in Figure 2.16(i). Hence, this approach allows localised damage to be automatically detected without using historical data, with changes in damage severity observed as well.

2.3.3.3 Sensitivity of metrics to machine condition changes

Damage severity trending is a very important task in the diagnostics field, because it helps to infer the stability of the damage growth and provides support for subsequent maintenance decisions. Intrinsically, fault trending requires historical data to be available, whereafter changes in the data with respect to the reference data are investigated. In this section, fault trending is investigated to assess the sensitivity of the proposed methodology to changes in machine condition. More specifically, the following investigations are performed:

- The sensitivity fo the PMMDM to changes in machine conditions is investigated to determine whether it contains information on the severity of the damage.
- The performance of the automatic condition inference for different fault severities is investigated. This is performed by presenting the probability that the PMMDM exceeds the alarm threshold, calculated with Equation (2.14) by using the procedure described in Section 2.3.2, as a function of measurement number.
- The sensitivity of the statistics of the divergence data to changes in condition is investigated for all of the measurements. The PMMDM of the statistics of the divergence data are also calculated by utilising historical reference data to highlight the sensitivity of the statistics to changes in condition.
- Lastly, fault trending by utilising historical reference data with Bayesian data analysis techniques is illustrated. This is to further support the benefits of using Bayesian data analysis techniques and to illustrate that this method can be used even when historical data are available.

The PMMDM of the divergence is shown in Figure 2.17(i) for 38 healthy measurements and 498 damaged measurements, with the probability that the posterior mean exceeds the threshold shown in Figure 2.17(ii). The Change In Condition (CIC) line indicates the start

of the damage measurements. The healthy and the damaged measurements are presented

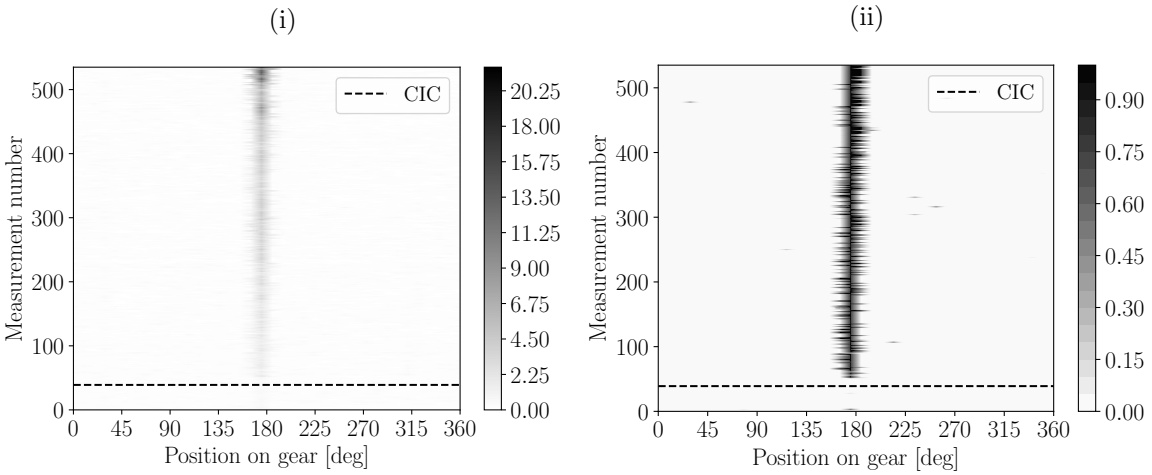


Figure 2.17: The surface plot of the PMMDM of the divergence data is shown in Figure 2.17(i) as function of measurement number and the probability that the posterior mean exceeds the threshold is shown in 2.17(ii). The threshold used to calculate the results in Figure 2.17(ii) is calculated separately for each measurement. The healthy and damaged data are separated by the Change In Condition (CIC) line.

together in Figure 2.17 to highlight the ability of the proposed approach to detect changes in machine condition. The change in the localised gear damage is evident from the results in Figure 2.17(i), especially in the last stages of the experiment.

In some instances in Figure 2.17(ii), it may seem that the angular arc of the damage is larger than in other cases. This occurs because the windows for feature extraction start arbitrarily between different datasets, but remain consistent for each dataset, because the order tracking is performed correctly. The consequence of this is that in some instances the damaged tooth is situated between two windows which result in the divergence data of two windows exceeding the threshold. This can potentially be alleviated by overlapping consecutive windows. Four statistics, calculated from the divergence metric are presented as a function of measurement number and compared to the metrics investigated in Figure 2.18. It is not only possible to observe the change in condition, but also the change from the seeded damaged gear to the gear with the broken tooth for the mean, standard deviation and maximum statistics in Figure 2.18. This is in contrast to the results in Figure 2.15, where only the change in condition is observed at the CIC line. Hence, the metrics are significantly more sensitive to changes in machine condition and therefore the proposed methodology is better suited for fault diagnosis.

In this part of the investigation, historical reference data are utilised for automatic fault

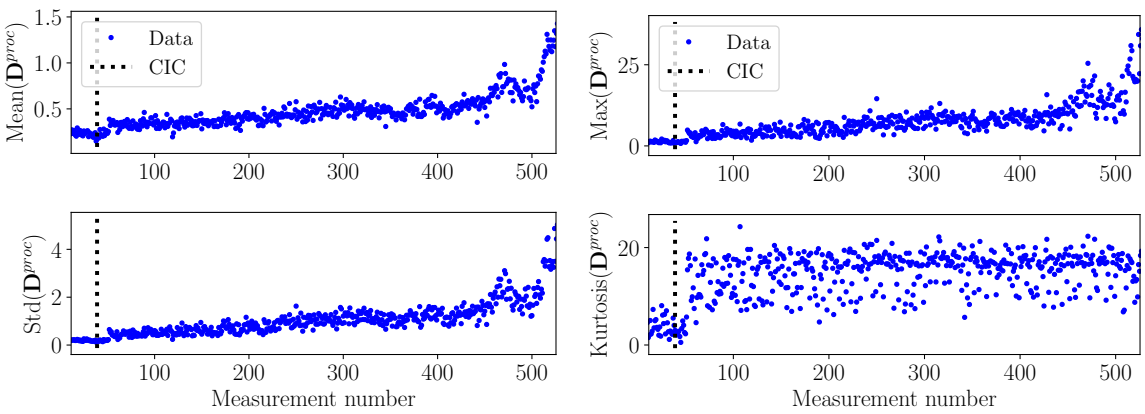


Figure 2.18: The mean, standard deviation (denoted by std), maximum (denoted by max) and the kurtosis of $\boldsymbol{D}^{proc}$ is shown as function of measurement number. The change in the gear condition is indicated by the Change In Condition (CIC) line.

trending to prove that the metrics do change with respect to measurement number as the condition of the gearbox changes. However, it is also investigated to emphasise that the metrics obtained from this methodology can be used for fault trending, if historical data do become available. It is assumed that twenty healthy measurements are available to calculate an alarm threshold, whereafter the trending procedure for all subsequent data is as follows:

- Calculate a statistic, such as the mean, of the whole processed divergence matrix $\boldsymbol{D}^{proc}$ for each measurement as done in Figure 2.18.
- Window the one-dimensional statistic into windows with length N_L and an overlap of N_O between consecutive windows.
- Model the windowed data with a Gaussian likelihood function and impose a Normal-Gamma prior distribution on the parameters. Use Equation (2.7) to (2.10), with the same hyperparameters as the previous sections, to obtain the posterior distribution parameters.
- Determine the probability that the mean of the windowed data exceeds the threshold with Equation (2.13). The threshold is calculated with Equation (2.14) for the first 20 measurements.

The four statistics of $\boldsymbol{D}^{proc}$ used in Figure 2.18 are independently investigated and compared to an independently calculated threshold, with the results shown in Figure 2.19. The windows have a length of $N_L = 20$ and an overlap of $N_O = 18$ measurements in all investigations. The results in Figure 2.19 indicate that it is essential to analyse the data with probabilistic techniques due to the variation thereof. It can be observed that the PMMDM of all statistics

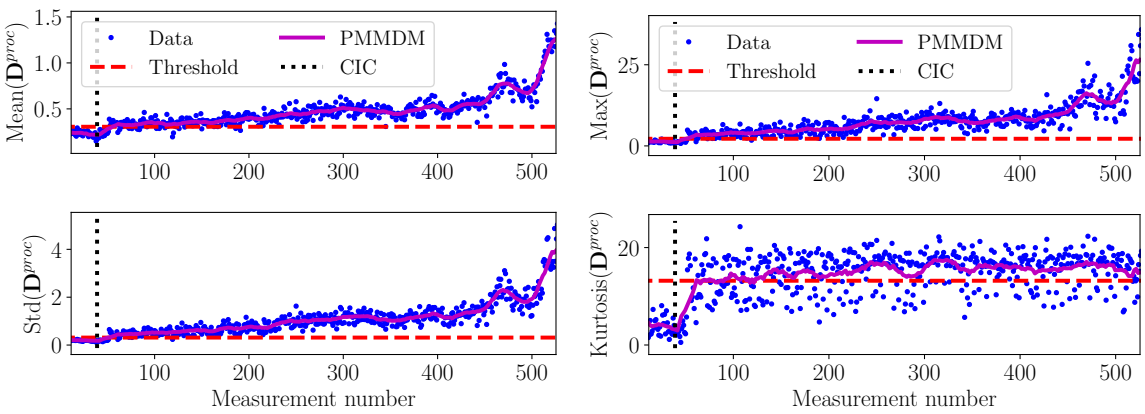


Figure 2.19: The PMMDM and the threshold of the data in Figure 2.18 are presented with the data in Figure 2.18.

changes at the change in condition line (i.e. where the healthy and the damaged measurements meet). Note that because the threshold is set far from the healthy data and windows are used, the threshold is not immediately exceeded for the first damaged measurements. A change in the metrics can however be observed.

The mean, standard deviation and the maximum value of $\boldsymbol{D}^{proc}$ contain less noise for the healthy measurements and therefore an alarm threshold is obtained which is close to the healthy data. The kurtosis, is very sensitive to the noise which is present in $\boldsymbol{D}^{proc}$. The implications of this is that the mean, standard deviation and the maximum of $\boldsymbol{D}^{proc}$ are more sensitive to changes in machine condition than the kurtosis, with the standard deviation and maximum values being the most sensitive.

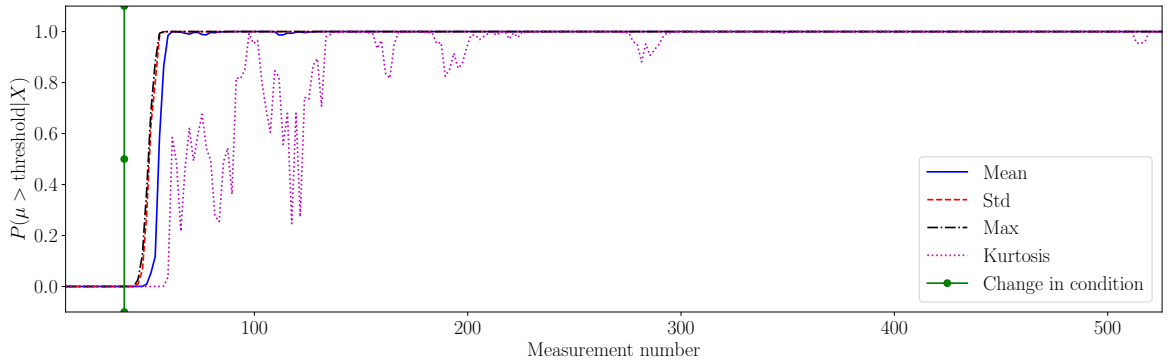


Figure 2.20: The probability that the mean of each statistic exceeds the threshold is shown in Figure 2.19. The change in condition line is shown as well.

If noise, for example from environmental effects or operating condition changes, is present in the trended statistics, it is uncertain whether the mean of the dataset exceeds the threshold. In Frequentist statistical approaches, the parameters are fixed [114] and therefore the mean will either be larger or smaller than the threshold. The Bayesian analysis approach allows

the uncertainty in the parameters due to the noise to be quantified, which helps with the maintenance decision process. Another observation that is made, is that using the posterior mean with Equation (2.13) is better suited for the condition inference task than using the individual values in Figure 2.19. This is because the posterior mean is less susceptible to noise and other outliers and it will therefore cause fewer false alarms if for example a new operating condition is observed for a specific measurement.

Hence, from this investigation it can be concluded that the PMMDM of the synchronous divergence data is sensitive to changes in machine condition, the condition can be automatically inferred throughout the life of the component, statistics such as the standard deviation of the divergence data are sensitive to changes in machine condition and that Bayesian data analysis techniques can be used for effective fault trending if historical reference data are available.

2.3.4 Diagnosis of the pinion

Up to this point, a gearbox with a damaged gear was investigated in this chapter, with the pinion remaining healthy for all the measurements that were made. It is important to distinguish between damage on the gear and damage on the pinion and therefore a brief investigation is now performed to determine whether the condition of the pinion is correctly inferred with the proposed technique.

The same process is performed as in Section 2.3.3.2 for the gear, with the exception that the number of windows (N_w) per revolution is 20 instead of 37, because the number of teeth on the gear and the pinion is different. The PMMDM and the probability that the posterior mean of a tooth exceeds the newly calculated threshold, is shown in Figure 2.21. Note that the exact same datasets are used as in Figure 2.16, with the same procedure performed as well. According to the results, the pinion is in a healthy condition and this proves that the condition of a healthy gear in a damaged gearbox can correctly be inferred by using the proposed technique.

The influence of the damaged gear on the statistics of the pinion's divergence data needs to be assessed, therefore, the statistics of the pinion are calculated for each measurement in the dataset. The same process is used as in Section 2.3.3.3 with the results shown in Figure 2.22 for the pinion. The results indicate that the statistics of the pinion divergence data contain much noise, however it remains fairly constant until the 480th measurement whereafter a slight increase is observed.

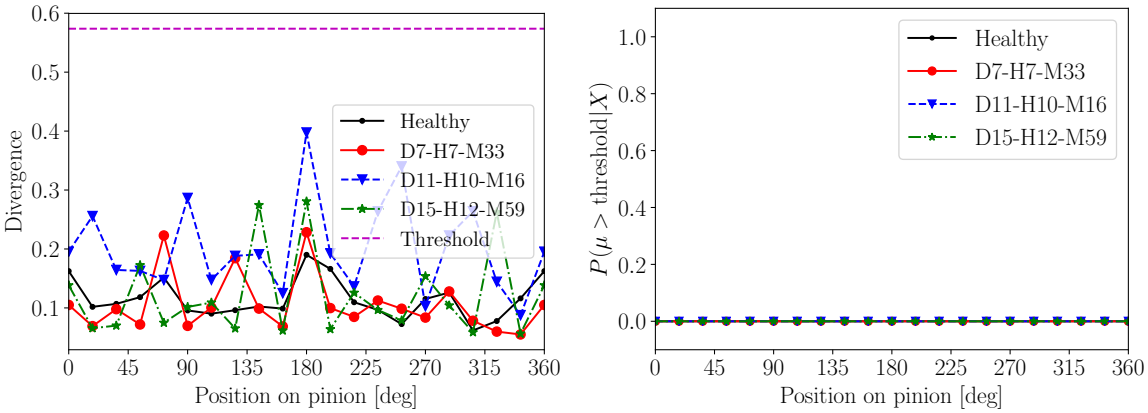


Figure 2.21: The PMMDM and the probability that the mean exceeds the threshold for the pinion.

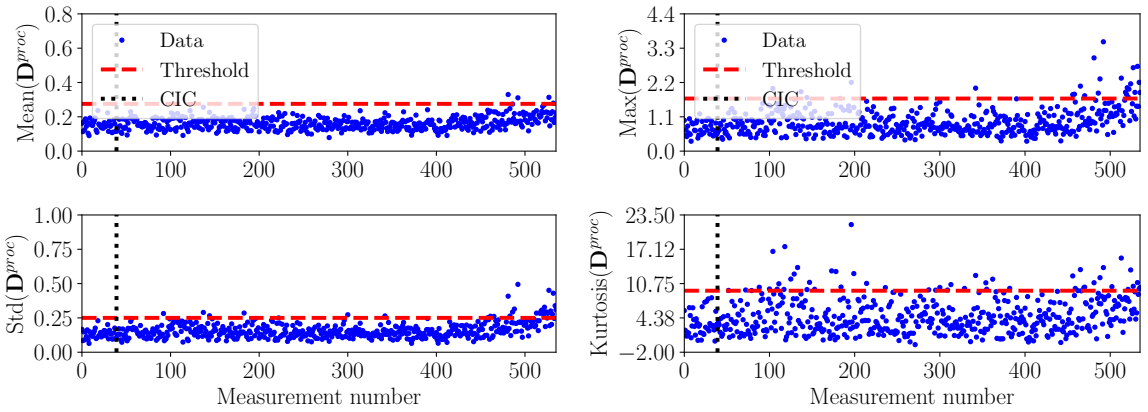


Figure 2.22: Statistics of $\boldsymbol{D}^{proc}$ as a function of measurement number for the divergence data, obtained by considering the pinion of the gearbox instead of the gear. The same measurements are used as in Figure 2.19.

The probability that the mean of the windowed statistics exceed the threshold, calculated with the procedure in Section 2.3.3.3, is shown in Figure 2.23. The results indicate that the data reflect the true condition of the pinion for most of the measurements. Only in the last stages of the experiment, the damaged portion of the gear dominates the vibration data, which is ultimately reflected in the pinion statistics as well. Hence, the increase in the pinion's statistics and the results from Section 2.3.3.3 indicate that failure of the gear is imminent at the 480th measurement; albeit actually because of damage on the gear, rather than on the pinion. Figure 2.23 however illustrates that the proposed approach is able to largely separate the gear and the pinion effects so that the correct gear condition can be inferred.

2.4 Conclusion and recommendations

In this chapter, a localised gear anomaly detection methodology is proposed to automatically detect localised gear damage with a data availability level 0, i.e. no historical data are

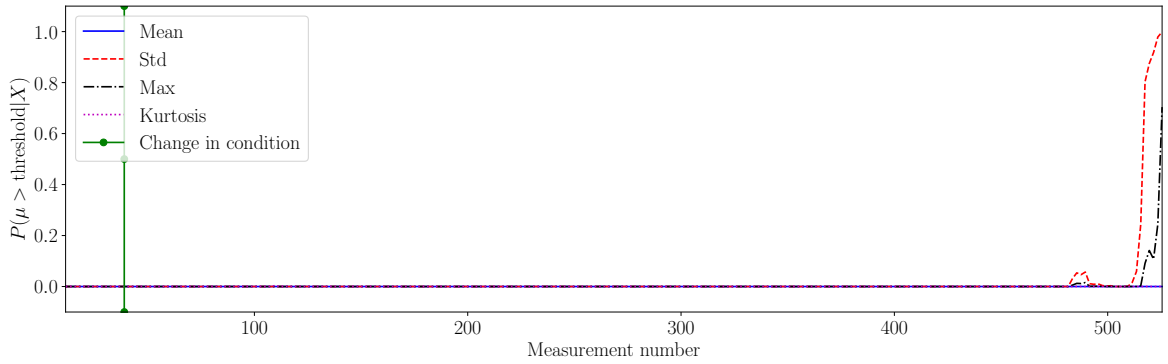


Figure 2.23: The probability that the mean of the statistics in Figure 2.22 exceeds the threshold is shown as a function of measurement number.

available. This is performed by utilising prior knowledge that gear teeth which are in the same condition should generate data that are statistically similar and localised gear anomalies generate data that are statistically different from other teeth. The Kullback-Leibler (KL) divergence is therefore calculated of the features of two gear teeth, with a large KL divergence indicating the gear teeth are in a different condition. Bayesian data analysis techniques are subsequently used to automatically detect the presence of localised gear damage from the processed divergence data.

The methodology is compared to the results obtained from conventional i.e. non-data-driven signal analysis tools on experimental gearbox datasets. The results emphasise the fact that data-driven techniques pose unique advantages to the condition monitoring domain e.g. damage can be detected automatically from diagnostic metrics which are sensitive to damage and robust to varying operating conditions. The results also emphasise the benefits of utilising engineering knowledge with data-driven techniques as opposed to purely data-driven techniques; it is possible to automatically detect the presence of localised gear damage without any historical data. Hence, an effective fault diagnosis methodology is obtained by using the available prior knowledge.

It is suggested that future work should investigate the benefits of using non-Gaussian models by comparing their results to the results of the Gaussian model used in this chapter.

Chapter 3

Discrepancy analysis methodology for bearings

3.1 Introduction

Discrepancy analysis is a novelty detection approach which has specifically been used for performing gear diagnostics under varying operating conditions. In discrepancy analysis, a statistical model is developed of short segments of the raw or processed vibration signal of a healthy gearbox. For example, the windowed raw data have been modelled by neural networks [154] and Gaussian mixture models [155], and the windowed wavelet transform of the vibration data have been modelled by a combination of hidden Markov models and Gaussian mixture models [156]. The models of the healthy data are subsequently used to assign a discrepancy measure or a localised novelty score to the windowed data of a new vibration measurement, whereafter the localised discrepancy measures are concatenated to form a discrepancy signal [154–157]. The synchronous averaged discrepancy signal [155, 156], the spectrum of the discrepancy signal [154] and the cepstrum of the discrepancy signal [154] have been used for gear diagnostics.

In this chapter, a discrepancy analysis methodology is proposed specifically for bearing diagnostics. Performing bearing diagnostics is important to ensure that gearboxes operate as intended [1, 6], and it is important to illustrate how time-frequency analysis techniques such as the wavelet packet transform can be supplemented with historical data for more effective diagnostics.

In Section 3.2, the methodology is presented, whereafter the methodology is validated on numerical gearbox data in Section 3.3 and on experimental bearing data in Section 3.4.

3.2 Methodology

3.2.1 Overview

The general process diagram for discrepancy analysis is presented in Figure 3.1, with the exact discrepancy analysis procedure implemented in this chapter presented in Figure 3.2. However, one of the benefits of the proposed methodology is that sub-processes such as the

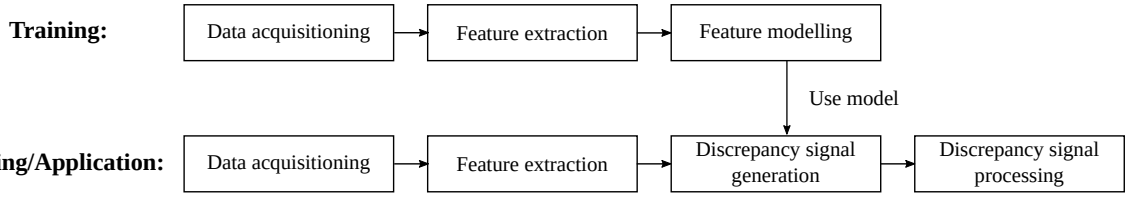


Figure 3.1: The key steps of the training and testing phase of the rolling element bearing diagnostic methodology are presented.

feature extraction procedure can be replaced with a more appropriate procedure if desired. In

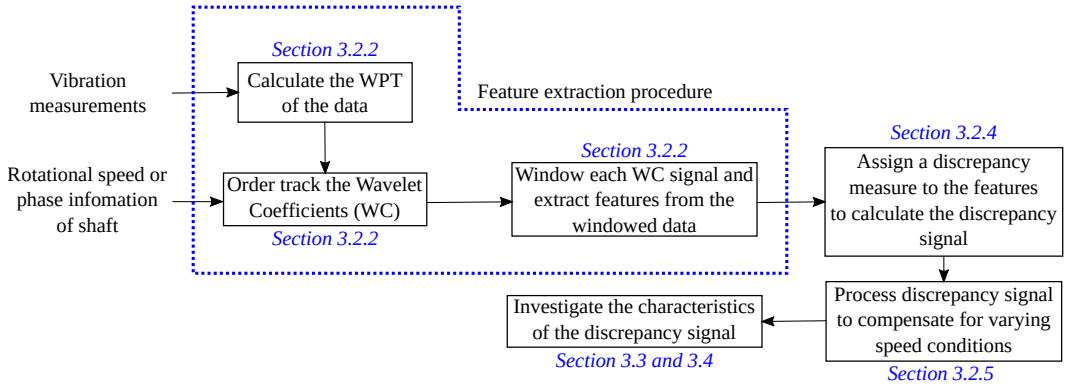


Figure 3.2: The steps in the methodology used in subsequent investigations are presented. It is assumed that the model parameters used to generate the discrepancy signal are already estimated with the method in Section 3.2.3 and therefore this step is excluded from the process diagram. Different characteristics of the discrepancy signal are investigated for the different datasets and therefore only Section 3.3 and Section 3.4.

the training phase, the features are extracted from the healthy vibration data, whereafter the features are modelled with a statistical model. In the testing phase, the features are extracted with the same approach as in the training phase, but the model of the healthy features are used to generate a discrepancy measure. The sequence of consecutive discrepancy measures is used to form a discrepancy signal which is processed to infer the presence of bearing damage.

3.2.2 Feature extraction

The feature extraction process that is used in this chapter, is presented in Figure 3.3. Wavelet

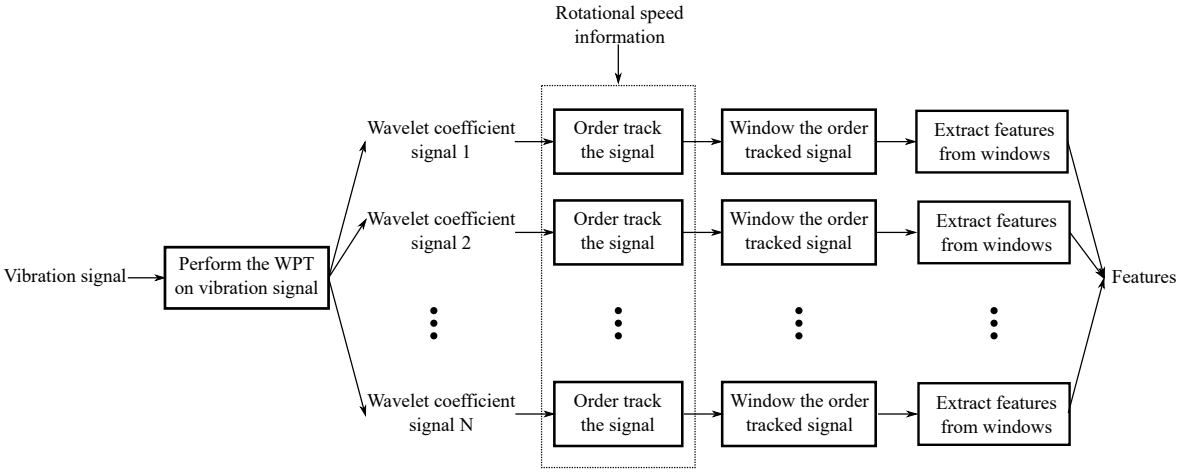


Figure 3.3: The feature generation process used in the proposed methodology. The features are in the angular domain due to the order tracking that is performed in the process. The abbreviations are, WPT: Wavelet Packet Transform; N: Total number of wavelet coefficient signals.

analysis is used in the feature extraction process, because it proved successful in detecting singularities and it is sensitive to impulses induced by bearing damage [59]. The WPT decomposes the signal into a set of wavelet coefficients with a filter bank of scaling and wavelet basis functions, where the wavelet coefficients are associated with specific sub-bands. The wavelet coefficients are dependent on the wavelet basis function, the appropriate choice depending on the characteristic that needs to be detected, and the level of decomposition, which depends on the frequency resolution that is required for the wavelet coefficients. The WPT is applied to the time domain signal, because the resonances which are excited by the bearing damage are independent of the rotational speed.

The features need to be sensitive to the presence of impulses, which manifest as time-localised bursts of energy and are quasi-periodic with respect to shaft angle due to rolling element slip. Hence, the extracted wavelet packet coefficients are subsequently order tracked with rotational speed information to transform it from the time to the angle domain. Hence, the order tracked wavelet coefficients preserve the angle-time cyclostationary properties of the bearing damage [31, 98]. Localised features are extracted from the wavelet coefficients, by windowing the wavelet coefficients with rectangular windows, whereafter the Root-Mean-Square (RMS) is extracted from each window. The RMS of the windowed wavelet coefficient is appropriate as a feature, because the localised energy and not the frequency characteristics

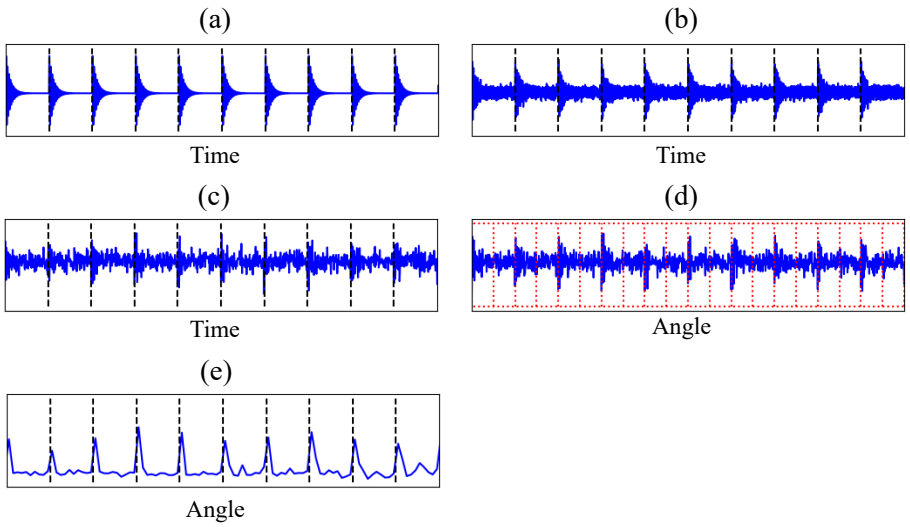


Figure 3.4: The windowing procedure is illustrated for a synthetic signal under constant operating conditions. In (a), a train of bearing impulses is shown. In (b), broadband Gaussian noise is added to the signal in (a) and is used for subsequent analyses. The WPT with a decomposition level of N_{dec} is performed on the signal in (b) which resulted in $2^{N_{dec}}$ wavelet coefficient signals. In (c), a single wavelet coefficient signal is shown and windowed with 20 non-overlapping windows in (d). The signal in (c) is order tracked to obtain the signal in (d). Finally, the RMS is extracted from the windows in (d) and a signal with 20 samples is shown in (e). The procedure in (d) and (e) is repeated for all $2^{N_{dec}}$ wavelet coefficients to obtain $2^{N_{dec}}$ signals similar to (e) which constitute the 2^{dec}-dimensional feature set which consist of 20 observations.

of the windowed data are important. The windowing procedure is illustrated in Figure 3.4 for a single wavelet coefficient signal extracted by applying the WPT on a synthetic bearing signal immersed in noise. The RMS of the windowed wavelet coefficients improves the sensitivity of the wavelet coefficients to damage and is therefore ideal for discrepancy analysis. The windows do not overlap in Figure 3.4 to make the illustration easier to understand. In the methodology, rectangular windows as opposed to Hamming windows for example are used to ensure that no information is lost and overlapping windows are used to ensure that no information on the edges of the non-overlapping windows is lost. The number of windows per shaft revolution should be more than twice the fault order that needs to be detected, but the window length should be sufficiently long to ensure that the statistics are properly estimated. It is suggested that the overlap should at least be 50% between coinciding windows to ensure that no information on the edges are lost.

Using additional statistics such as the kurtosis can potentially improve the diagnostic capabilities of discrepancy analysis, especially for incipient damage detection. However, optimising the discrepancy analysis procedure is beyond the scope of this work. The magnitude of the bearing impulses and of the vibration signal are dependent on the operating conditions,

which can cause problems when performing condition inference. This aspect is addressed in Section 3.2.5.

The window length is selected as $2\pi/20$ rad with a 50% overlap, to allow localised impulses to be detected. Wu and Liu [72] found that the Daubechies db20 wavelet basis function outperformed the Daubechies db4 and db10 wavelet basis functions for engine diagnostics and the Daubechies db20 has the shape of an impulse, which is why it is used in this chapter. The level of the decomposition, denoted by N_{dec}, is related to the number of wavelet coefficients by $2^{N_{dec}}$, with only the wavelet coefficients of the final decomposition level used for feature extraction. Hence, as only the RMS of the windowed wavelet coefficients is calculated and there are $2^{N_{dec}}$ wavelet coefficients, the dimensionality of the feature space (i.e. the number of features) is equal to $D = 2^{N_{dec}}$. The level of the decomposition is $N_{dec} = 4$ for all datasets investigated in this chapter which result in 16 wavelet coefficients. This means that the dimensionality of the feature space is 16, which provides an ideal compromise between having sufficient resolution to detect changes in localised spectral frequency bands but also to have a sufficiently low dimensionality to avoid the curse of dimensionality. Dimensionality reduction techniques, such as principal component analysis [114], can also be performed to alleviate the problems associated with the curse of dimensionality, but is not performed in this investigation due to the relatively low dimensionality of the feature space and the models that are used.

3.2.3 Feature modelling

The features, extracted in Section 3.2.2 from the healthy data, can be modelled with a Gaussian model, a Gaussian mixture model or a hidden Markov model to list a few examples. In this chapter, a Gaussian model is used to model the features extracted from the healthy data. The motivation for using a Gaussian model as opposed to other models is that it is very simple to implement and with the post-processing, which will be presented in Section 3.2.5, it is robust to varying speed conditions.

The D-dimensional features extracted from window i are denoted by $\boldsymbol{b}_i \in \mathbb{R}^{D\times 1}$, with the features of the healthy machine denoted by $\boldsymbol{b}_i^{train} \in \mathbb{R}^{D\times 1}$. The training feature set $\{\boldsymbol{b}_i^{train}\}$ comprises of N_{obs} observations of the D-dimensional features and is used to obtain the parameters of the model. The maximum likelihood estimates of the mean $\boldsymbol{\mu_b} \in \mathbb{R}^{D\times 1}$

and the covariance $\boldsymbol{\Sigma_b} \in \mathbb{R}^{D\times D}$ of the healthy feature set $\{\boldsymbol{b}_i^{train}\}$ are given by [114]

$$\boldsymbol{\mu_b} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \boldsymbol{b}_i^{train}, \tag{3.1}$$

$$\boldsymbol{\Sigma_b} = \frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} (\boldsymbol{b}_i^{train} - \boldsymbol{\mu_b})(\boldsymbol{b}_i^{train} - \boldsymbol{\mu_b})^T, \tag{3.2}$$

where the transpose is denoted by superscript T. If a single measurement is used to obtain the model parameters, N_{obs} is equal to the number of windows used for extracting features from the measurement. If multiple measurements are used to obtain the model parameters, then the extracted features of the respective measurements are concatenated together to form a single D-dimensional training set with a total of N_{obs} samples.

3.2.4 Discrepancy signal generation

The Mahalanobis distance is used as the discrepancy measure, which quantifies the deviation of the new data point from the model of the healthy data. The discrepancy measure is denoted by η_j for the features associated with the window j [114]

$$\eta_j = \sqrt{(\boldsymbol{b}_j - \boldsymbol{\mu_b})^T \boldsymbol{\Sigma_b}^{-1} (\boldsymbol{b}_j - \boldsymbol{\mu_b})}, \tag{3.3}$$

where the model parameters are obtained using Equations (3.1) and (3.2) from the healthy data, and $\boldsymbol{b}_j$ denotes the features, extracted from the wavelet coefficients in the jth window, of the testing feature set. The squared Mahalanobis distance is proportional to the negative logarithm of the probability density function of a Gaussian model, where the latter is in general available for all probabilistic models. The discrepancy measure, obtained from each time step, is used to form a discrepancy signal which can be processed to develop a diagnostic metric so that the condition of the bearing can be inferred. Therefore, the discrepancy signal $\boldsymbol{\eta} \in \mathbb{R}^{N_{obs}\times 1}$ corresponds to the features extracted from the N_{obs} windows applied to the vibration measurement under consideration.

3.2.5 Processing the discrepancy signal

The discrepancy signal is obtained in terms of shaft angle or shaft revolutions and need to be processed and analysed so that the condition of the bearing can be inferred. The simplest processing technique is to calculate the RMS of the discrepancy signal, but it will only be

possible to detect and trend damage and it will not be possible to perform fault localisation. Even though the discrepancy signal is relatively robust to varying operating conditions, it is not immune to it. Due to the sensitivity of the transmission path, between the excitation source and the measurement point, to the rotational speed and the presence of resonances, it is expected that the discrepancy measure η will be dependent on the rotational speed ω.

Bartelmus and Zimroz [11] and Zimroz et al. [9] showed that the parameters of a linear regression model, aiming to capture the relationship between the diagnostic metric and the operating conditions, are very good condition monitoring metrics. In this chapter, the conditional distribution $p(\eta|\omega)$ is modelled and used to mitigate the adverse influences of the varying speed conditions on the diagnostic metric. It is possible to model $p(\eta|\omega)$ by using Bayesian linear regression for example, however standard regression approaches assume that the noise has constant statistical properties [114], which is not always true for the data considered. In Figure 3.5, the sensitivity of the discrepancy measure with respect to rotational speed is presented for a phenomenological gearbox model, discussed in Section 3.3, that contains no bearing damage. The mean as well as the variance of the discrepancy

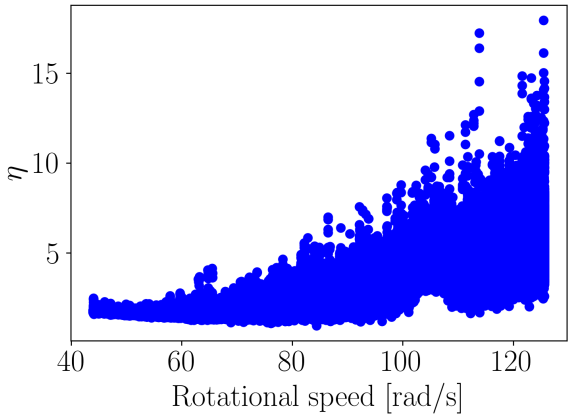


Figure 3.5: The discrepancy measure η vs. the rotational speed ω from the phenomenological gearbox model, considered in Section 3.3.

measure conditioned on the rotational speed are dependent on the rotational speed, with a resonance band being excited approximately at 105 rad/s. The data, presented in Figure 3.5, are introduced in this section to motivate and illustrate the processing method that is used in this chapter.

Heteroscedastic models are a class of models where the noise in the data is assumed to be a function of the inputs [114, 166, 167]. Mixture density networks [114], Gaussian processes [166] and Bayesian neural networks [167] have been used and extended for heteroscedastic models. The computational complexity of the aforementioned regression models increase

because Monte Carlo Markov Chain (MCMC) [166] or gradient-based optimisation methods are required [167] to estimate the noise parameters.

In this chapter, a very simplified non-parametric model is used to estimate the conditional distribution $p(\eta|\omega)$. If the assumption is made that the statistics of the discrepancy measure are independent of the input for small changes in the rotational speed, and that a sufficient amount of training data is available, then the conditional mean

$$\mu_{\eta|\omega}(\omega) = \mathrm{LI}\left(\{\Omega_i\}, \{\mu_{\Omega_i,\Delta\Omega}\}, \omega\right), \tag{3.4}$$

and variance

$$\sigma^2_{\eta|\omega}(\omega) = \mathrm{LI}\left(\{\Omega_i\}, \{\sigma^2_{\Omega_i,\Delta\Omega}\}, \omega\right), \tag{3.5}$$

can be good approximations for the actual conditional mean and variance for a specific rotational speed ω. The approach uses the discrepancy measures of the training dataset which correspond to a specific operating condition range $\omega \in \left[\Omega_i - \frac{1}{2}\Delta\Omega,\ \Omega_i + \frac{1}{2}\Delta\Omega\right]$ to calculate the mean $\mu_{\Omega_i,\Delta\Omega}$ and the variance $\sigma^2_{\Omega_i,\Delta\Omega}$. A set of window centres $\{\Omega_i\}$ are generated with a preselected window length $\Delta\Omega$ to generate a corresponding set of localised means $\{\mu_{\Omega_i,\Delta\Omega}\}$ and variances $\{\sigma^2_{\Omega_i,\Delta\Omega}\}$ of the discrepancy measure η; whereafter the linear interpolation function, denoted by LI, is used to predict the conditional mean $\mu_{\eta|\omega}$ given the rotational speed ω with Equation (3.4).

The benefits of this conditional modelling approach, as opposed to Bayesian regression for example, is that

- no a priori assumptions are made about the form of $\mu_{\eta|\omega}(\omega)$ nor $\sigma^2_{\eta|\omega}(\omega)$,
- it can be extended to other noise distributions,
- it is very computationally efficient as opposed to models requiring MCMC simulations, which is an advantage when performing condition monitoring.

The quality of the fit, using Equation (3.4) and Equation (3.5), depends on the distance between the window centres $\{\Omega_i\}$ and the window length $\Delta\Omega$. The unknown window length $\Delta\Omega$ is selected with a grid search optimisation process, discussed in Appendix A.2.1.1.

After, the parameters are estimated, equation (3.4) and Equation (3.5) are used to scale the discrepancy measure for the features associated with window j of the testing set, denoted

by η_j, with

$$\xi_j = \frac{\eta_j - \mu_{\eta|\omega}(\omega_j)}{\sigma_{\eta|\omega}(\omega_j)} \tag{3.6}$$

where ξ_j denotes the new, scaled discrepancy measure and ω_j denotes the average rotational speed for the features associated with window j. The above procedure can be extended to other distributions, for example if the actual distribution $p(\eta|\omega)$ is skew in the direction where the discrepancy measure increases, the logarithm of the discrepancy measure $\log \eta$ can be used instead of η to make the distribution more symmetric. The benefits of the proposed discrepancy processing procedure, used in this section, are only seen when applied under time varying speed conditions and will not be seen under constant operating conditions.

The proposed methodology as presented in Figure 3.2 is investigated on phenomenological gearbox model data in the next section, whereafter it is investigated on experimental data in Section 3.4.

3.3 Numerical validation

3.3.1 Phenomenological gearbox model

The phenomenological gearbox model that is used in this chapter, is obtained from the paper by Abboud et al. [28]. The model developed by Abboud et al. [28] approximates complicated physical phenomena such as other damaged machine components and the sensitivity of the signal's amplitude to varying rotational speeds, while it is simple to implement and computationally efficient to calculate, as opposed to lumped-mass and finite element gearbox models aiming to replicate the same phenomena. The casing or measured vibration signal for the gearbox is decomposed into four distinct parts [28]

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + x_b(t) + x_n(t), \tag{3.7}$$

where the deterministic gear component due to gear mesh interactions x_{gmc}, the random gear component due to distributed gear damage x_{dgd}, the outer race bearing damage component from a defective bearing x_b and the noise x_n components can be decomposed into the signal at the source, convoluted with the associated impulse response function to the point of

measurement [28]. The decomposition of the signals

$$x_{gmc}(t) = h_{gmc} \otimes q_{gmc}(t), \tag{3.8}$$

$$x_{dgd}(t) = h_{dgd} \otimes q_{dgd}(t), \tag{3.9}$$

$$x_b(t) = h_b \otimes q_b(t), \tag{3.10}$$

is written in terms of the impulse response function h_i, the signal at the source q_i and the convolution operator is denoted by $\otimes$. The impulse response function is written as a single degree of freedom system with viscous damping

$$h_i(t) = e^{-\zeta_i 2\pi f_{n,i} t} \sin\left(2\pi\sqrt{1-\zeta_i^2} f_{n,i} t\right), \tag{3.11}$$

where the damping ratio of component i is denoted by ζ_i and the natural frequency of component i is denoted by $f_{n,i}$. The deterministic gear component at the source, simulates the excitation due to the gear mesh interactions

$$q_{gmc}(t) = M_{gmc}(\omega_r(t)) \sum_{k=1}^{N_{gmc}} A_{gmc}^{(k)} \cos\left(kN_{t,g} \int_0^t \omega_r(\tau) d\tau + \varphi_{gmc}^{(k)}\right), \tag{3.12}$$

where N_{gmc} is the number of gear mesh components which are considered, $A_{gmc}^{(k)}$ is the amplitude of the kth component, $\varphi_{gmc}^{(k)}$ is the phase of the kth component, $\omega_r(t)$ is the rotational speed of the shaft in radians per second and $N_{t,g}$ is the number of teeth on the gear that is connected to the shaft under consideration. The monotonic function $M_{gmc}(\omega_r)$, simulates the dependence of the amplitude of the signal component on the rotational speed. The random gear component, simulates distributed gear damage, and is incorporated to complicate the bearing condition inference task

$$q_{dgd}(t) = M_{dgd}(\omega_r(t)) \cdot \varepsilon_{dgd}(t) \sum_{k=1}^{N_{dgd}} A_{dgd}^{(k)} \cos\left(k \int_0^t \omega_r(\tau) d\tau + \varphi_{dgd}^{(k)}\right), \tag{3.13}$$

and it incorporates similar components as Equation (3.12), except for a random component ε_{dgd}, obtained from

$$\varepsilon_{dgd}(t) \sim \mathcal{N}(0, \sigma_{dgd}^2), \tag{3.14}$$

where $\mathcal{N}(0, \sigma_{dgd}^2)$ is a zero-mean Gaussian distribution with variance σ_{dgd}^2 and $\sim$ denotes that a random sample is taken from the distribution. The noise from other components such as

ambient effects is written as

$$x_n(t) = \varepsilon_n(t) \cdot M_n(\omega_r(t)), \tag{3.15}$$

where the monotonic function of ω_r is denoted by M_n and the random component

$$\varepsilon_n(t) \sim \mathcal{N}(0, \sigma_n^2), \tag{3.16}$$

is a sample from a zero-mean Gaussian function with constant variance σ_n^2.

Lastly, the damaged bearing signal is addressed. The outer race damage is simulated as a train of Dirac functions

$$q_b(t) = M_b(t) \sum_{i=1}^{N_\mathcal{T}} A_b(i) \cdot \delta(t - \mathcal{T}_i), \tag{3.17}$$

centred at $\mathcal{T}_i$, where $\mathcal{T}_i$ depends on the characteristics of the bearing, slip in the bearing as well as the shaft speed. Hence, the angle of the impacts with respect to shaft orders, is calculated in the angle domain and transformed back to the time domain using the relationship between shaft angle and time [97]. The bearing slip is introduced by adding zero mean Gaussian noise with a standard deviation of 0.1 to the expected impact angles. The random factor $A_b(i) \sim \mathcal{N}(\bar{A}_b, \sigma_{A_b}^2)$ simulates the fact that the magnitudes of the bearing impulses, for a damaged bearing, are not constant over time. A change in the condition of the bearing, can be simulated by changing the magnitude of the bearing component in the casing vibration signal. This is performed by explicitly including a Fault Severity (FS) factor in the modified casing vibration signal

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + FS \cdot \tilde{x}_b(t) + x_n(t), \tag{3.18}$$

where $\tilde{x}_b(t)$ is a baseline bearing damage component and FS linearly scales the magnitude of the bearing impulse. The characteristic frequency of the outer race bearing component is 2.57 shaft orders and it manifest in a frequency band at 5 kHz. A healthy bearing is obtained by setting $FS = 0$.

3.3.2 Vibration data

The proposed methodology is investigated on the phenomenological gearbox model presented in Section 3.3.1, with the four rotational speed profiles investigated for this model given by

$$\omega_{r,1}(t) = 2\pi(1.3t + 7), \tag{3.19}$$

$$\omega_{r,2}(t) = 2\pi(5\sin(0.4\pi t) - 20\,(0.1t - 0.5)^2 + 15), \tag{3.20}$$

$$\omega_{r,3}(t) = 2\pi(6.5\cos(0.2\pi t) + 13.5), \tag{3.21}$$

$$\omega_{r,4}(t) = 2\pi(6.5\sin(0.2\pi t) + 13.5), \tag{3.22}$$

in rad/s. The rotational speed profiles are also presented in Figure 3.6, where it can be observed that the four operating conditions have the same maximum and minimum values and the same mean. Only the healthy data for the first three operating conditions are

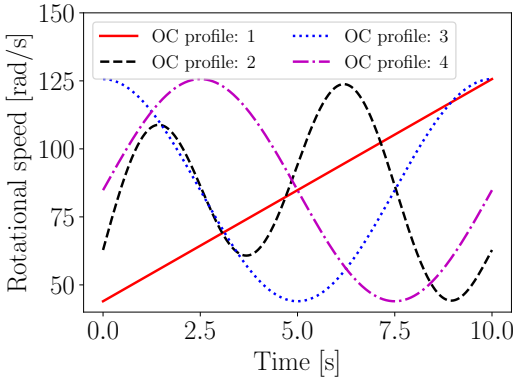
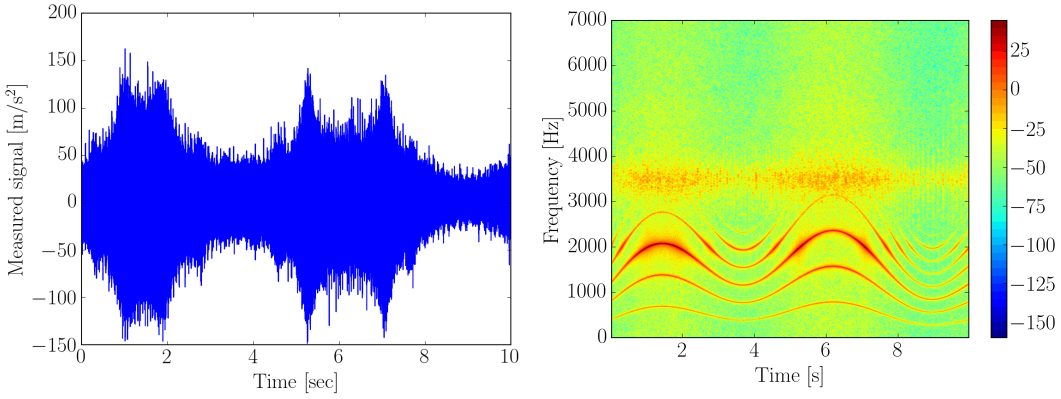


Figure 3.6: The four rotational speed profiles $\omega_{r,i}$ for the phenomenological model are presented, where the i in $\omega_{r,i}$ denotes the operating condition profile number indicated in the legend of the figure. The corresponding equations of rotational speed profiles are given in Equation (3.19) to Equation (3.22).

used for the model optimisation process. The casing vibration signal and the corresponding spectrogram for the second operating condition profile are presented in Figure 3.7 for a fault severity $FS = 1$ in Equation (3.18). The meshing components, the resonance band excited by the gear meshing components at 2.0 kHz as well as the resonance band excited by the random gear component at 3.5 kHz are clearly seen in the spectrogram presented in Figure 3.7(ii). However, the resonance band that is excited by the bearing impulses at 5.0 kHz is not seen in the spectrogram. The ability of the methodology to detect the presence of damage in the signal model is investigated in the next section.



(i) (ii)

Figure 3.7: The casing vibration signal is shown in its time and the time-frequency domain representations for the second operating condition profile in Figure 3.7(i) and Figure 3.7(ii), respectively.

3.3.3 Results

The feature extraction procedure in Section 3.2.2 and the feature modelling procedure in Section 3.2.3 were applied to the gearbox without bearing damage i.e. $x_b(t) = 0$ for $t \in \mathbb{R}$. This is not strictly a healthy gearbox, because distributed gear damage is present in the gearbox as well. However, this represents the gearbox in a reference condition, whereafter changes in the condition of the bearing is detected.

The model parameters are estimated from the vibration data of the first three operating condition profiles. The discrepancy measure, generated using Equation (3.3), is firstly obtained for the reference data and is presented with the corresponding rotational speed in Figure 3.5. Features are extracted from the order tracked and windowed wavelet coefficients after which a discrepancy measure is obtained for the 16 features extracted from each window. The discrepancy measure, denoted by η, is processed with Welch's power spectral density to calculate how its energy is distributed in the frequency domain with unit orders instead of Hertz. The spectrum in Figure 3.8 is presented for the four operating conditions that are under consideration as well as for three Fault Severities (FS) where FS, used in Equation (3.18), is equal to 1, 2 and 4, respectively.

The Ball Pass Frequency of the Outer race (BPFO) components in shaft orders are detected in all three cases, and as the damage increases the amplitude of the BPFO and its harmonics increase as well. However, the amplitude of the BPFO is different for the various operating conditions given the same fault severity. As a result, it is difficult to distinguish between changes in operating condition and changes in machine condition.

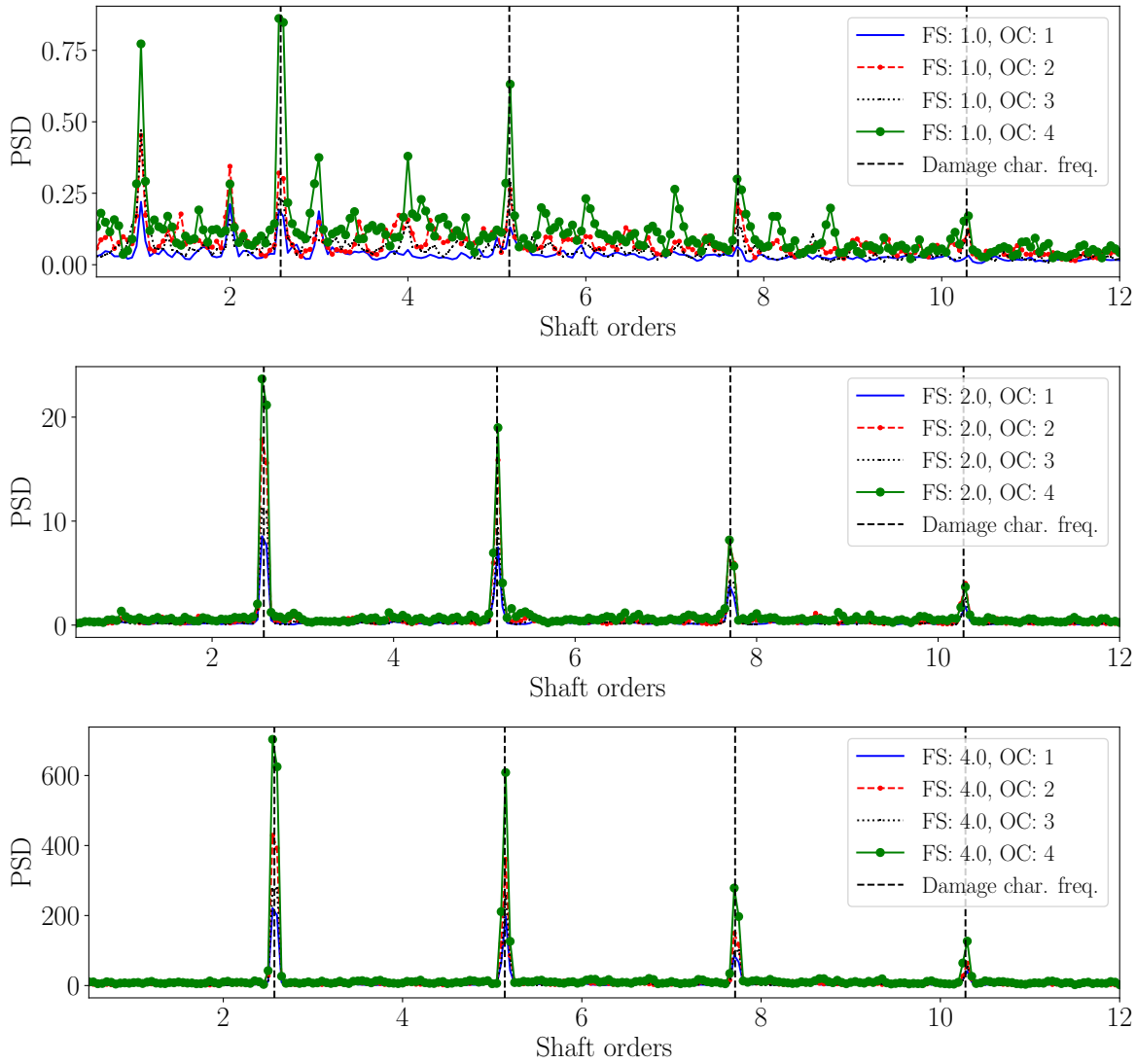


Figure 3.8: Comparing the spectrum of the discrepancy signals for the four operating conditions and the three fault severities (FS) which are investigated. Note that different scales are used for the three cases.

However, to make the results more robust, it is necessary to take into account the fact that the vibration signal and its components are dependent on the operating conditions. The conditional modelling procedure in Section 3.2.5 is used to obtain the conditional statistics. Equation (3.4) is used to obtain $\mu_{\eta|\omega}$ and Equation (3.5) is used to obtain $\sigma^2_{\eta|\omega}$, which are presented in Figure 3.9(i) as $\mu_{\eta|\omega} \pm 3 \cdot \sigma_{\eta|\omega}$ and superimposed on the reference data. A satisfactory fit is obtained with η in Figure 3.9(i) and therefore it is used further in this section. In Figure 3.9(ii), the scaled discrepancy measure that is obtained using Equation (3.6) is presented over the rotational speed. The scaled discrepancy signal, ξ_i, presents the desired property as its magnitude is relatively speed independent.

The spectrum of the processed discrepancy signal is presented in Figure 3.10 and it should be highlighted that the amplitude of the BPFO is independent of the rotational

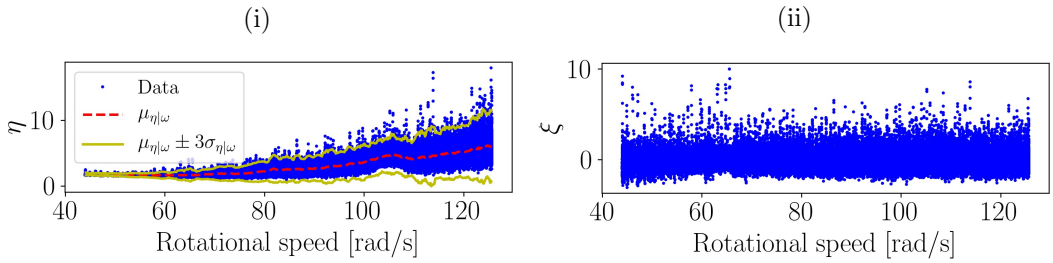


Figure 3.9: The discrepancy data over the rotational speed, originally presented in Figure 3.5, is presented with the fit $\mu_{\eta|\omega}(\omega)$ and $\sigma_{\eta|\omega}(\omega)$ in Figure 3.9(i). The processed discrepancy signal is presented in Figure 3.9(ii) for the reference data.

speed profile. The slight amplitude differences between the BPFO, estimated under the same operating condition, are attributed to a variation in noise, the random gear component and the model fit which is not perfect as seen in Figure 3.9. However, a significant improvement has been achieved by processing the discrepancy measure. This makes detecting, localising and trending the bearing damage easier under varying speed conditions.

3.4 Experimental validation

Two experimental investigations are performed in this section. In the first investigation, the ability of the methodology to detect damage, localise damage and to trend the diagnostic metric as a function of measurement time is investigated. The dataset was acquired under constant operating conditions and therefore the results do not reflect the ability of the methodology to be used under varying speed conditions. Hence, in the second investigation, the ability of the methodology to detect and localise damage is investigated on experimental data that were acquired under varying speed conditions.

3.4.1 Investigation 1

The IMS bearing dataset [168, 169] for the case where outer race damage occurred on a bearing, which ultimately resulted in the complete failure of the bearing, is investigated in this section. A schematic of the setup is presented in Figure 3.11. The run-to-failure test-rig contains four bearings supporting the same shaft, where the shaft is connected with a belt to an electric motor. A constant load of 6000 lbs and a constant shaft speed of 2000 rpm were applied during the experiments. Vibration measurements of 1 second long were sampled at 20 kHz, and ultimately 984 measurements were obtained during the experiment until the amount of debris in the oil, that were attached to a magnetic plug, exceeded a predefined

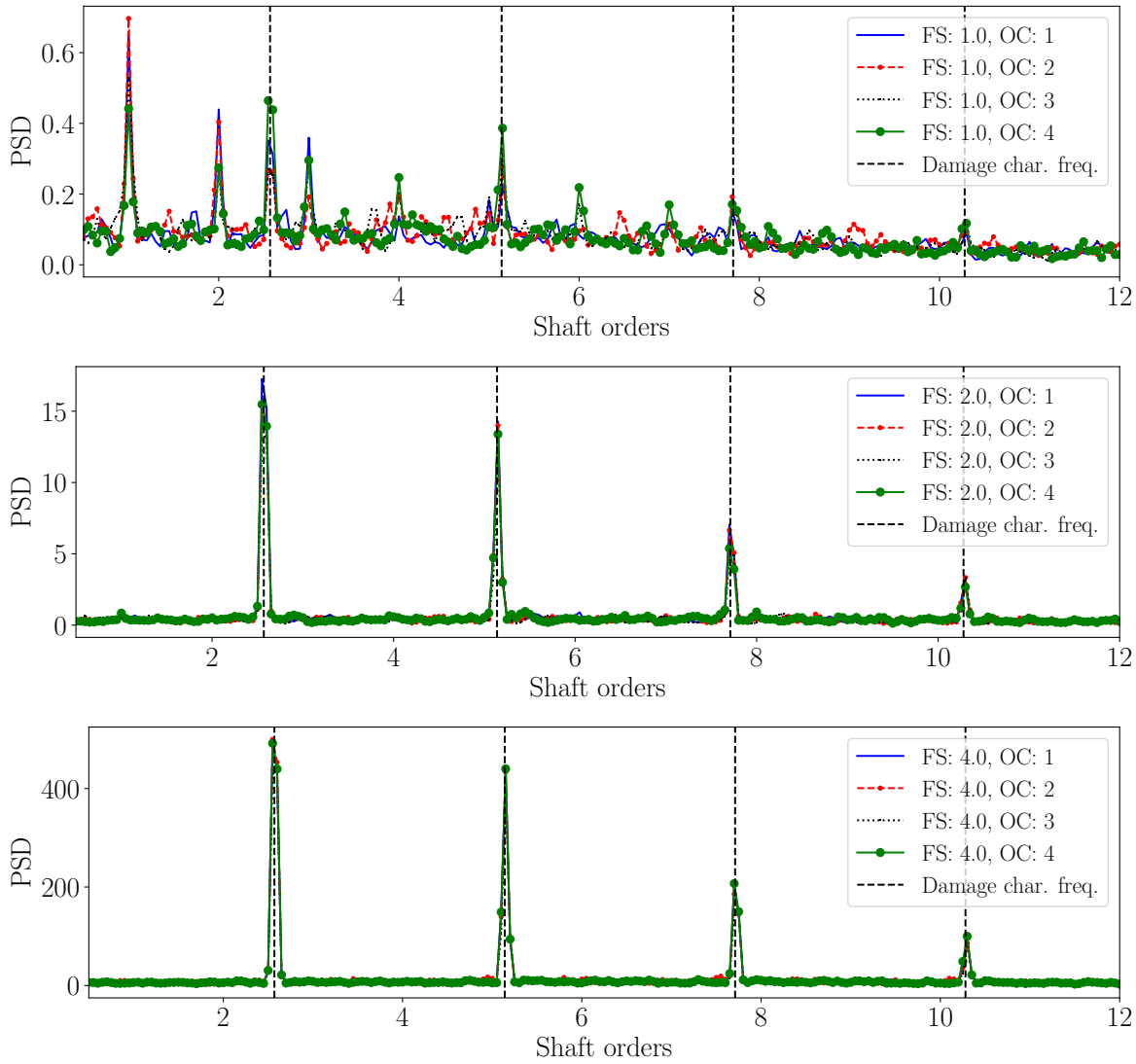


Figure 3.10: Comparing the spectrum of the processed discrepancy signals for the four operating conditions and the three fault severities that are investigated for the phenomenological gearbox model.

level [169].

The proposed methodology was applied to the dataset, where the first ten measurements of bearing B1 that ultimately failed, were used to optimise the Gaussian model. The processed discrepancy signal in the angle domain is used in this section, even though the data were acquired under constant operating conditions. There will not be any benefits from using the proposed discrepancy processing technique in Section 3.2.5 when the shaft speed is constant, however it is used for consistency. The power spectral density for each measurement was calculated and presented over the measurement number in Figure 3.12(i), while the spectrum for a single measurement is shown in Figure 3.12(ii). The theoretical BPFO component of 7.059 shaft orders and its harmonics are presented in Figure 3.12 as well. The damage components became visible at the 535th measurement in Figure 3.12(i).

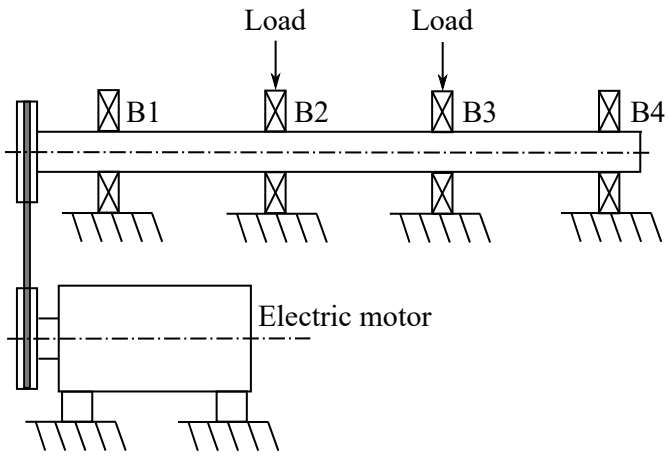
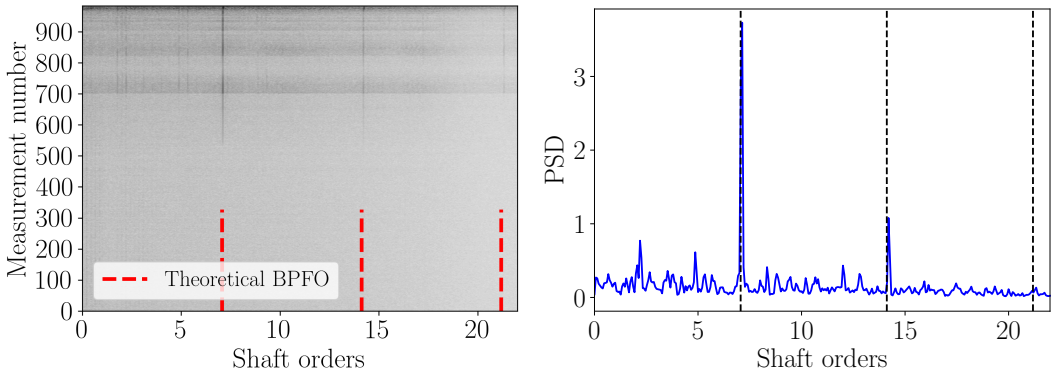


Figure 3.11: The IMS experimental setup [169]. The different bearings are indicated by B1, B2, B3 and B4. Bearing B1 is the bearing that failed.



(i) (ii)

Figure 3.12: The discrepancy as a function of measurement number and frequency (in shaft orders) is presented for all the measurement files in Figure 3.12(i), while the spectrum of a single measurement is presented in Figure 3.12(ii) for the IMS bearing dataset. The theoretical BPFO and its harmonics are presented as vertical, dashed lines in Figure 3.12(i) and Figure 3.12(ii).

Lastly, developing and trending a diagnostic metric over the measurement number are investigated. The first proposed metric, is to calculate the Root-Mean-Square (RMS) of the discrepancy signal for each measurement, which is presented over the measurement number in Figure 3.13. An alarm threshold is considered by calculating the mean μ and the standard deviation σ of the diagnostic metric for the first ten measurements, and then calculating $\mu + k\sigma$, where $k = 3$. If the number of measurements used to calculate the alarm threshold is too little, poor estimates of the mean and the standard deviation will be obtained. The second metric that is proposed is the amplitude of the fundamental component of the BPFO. An alarm threshold is calculated with the same method as the RMS metric. The change in condition, estimated at the time instant or measurement number where the mean of three consecutive diagnostic metrics exceed the threshold, is presented in Figure 3.13. The trending

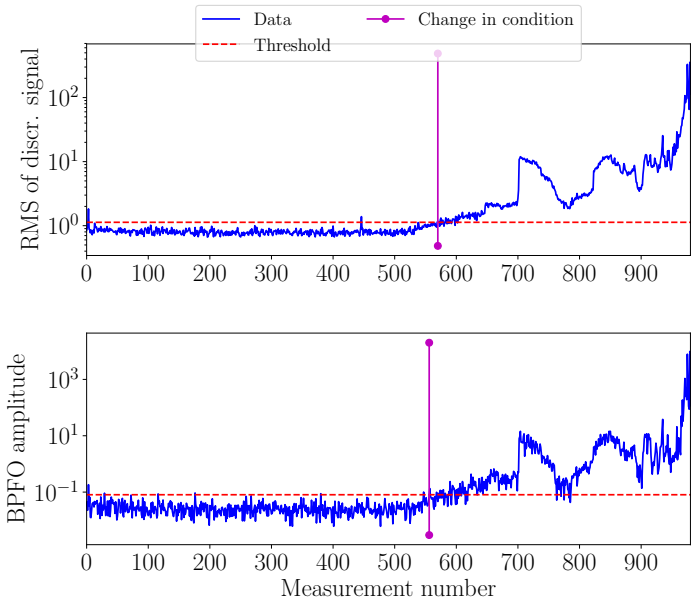


Figure 3.13: Two diagnostic metrics are shown on a logarithmic scale as function of the measurement number for the IMS dataset. For the first diagnostic metric, the RMS of the entire discrepancy signal is calculated and for the second diagnostic metric the fundamental outer race component in the spectrum for each measurement is calculated.

procedure is more robust to false alarms when using three consecutive measurements as opposed to a single measurement.

The diagnostic metrics presented in Figure 3.13 do not increase monotonically, which emphasises that the methodology's aim is to infer the presence of damage by detecting the presence of impulses in the signal. The progression of bearing damage, as indicated and discussed by El-Thalji and Jantunen [21], does not result in the magnitude of impulses to increase monotonically, but it can go through increasing and decreasing phases. This possibly reflects the characteristics that are observed in Figure 3.13.

Hence, from the results in Figure 3.12 and Figure 3.13 it is concluded that the developed methodology allows damage to be detected, located and trended prior to the failure of the bearing. The discrepancy signal is sensitive to changes in machine condition and the RMS of the discrepancy signal is a very good metric to be used for fault detection and trending. By using the amplitude of the BPFO component as a metric, the development of localised damage can be detected as well.

3.4.2 Investigation 2

In this section, four run-up datasets that correspond to four bearing conditions are analysed using the proposed diagnostic methodology. An overview of the setup as well as the data are presented in the next section, whereafter the methodology is applied on the data.

3.4.2.1 Overview of setup

Experimental data, acquired from a SpectraQuest, Inc. Machinery Fault Simulator, are processed in this section. The experimental setup, presented in Figure 3.14, consists of a 0.5 HP three-phase electrical motor that has a nominal speed of 3450 rpm and a shaft loader which applies a static load of 49.05 N to the system. The left bearing was in a healthy

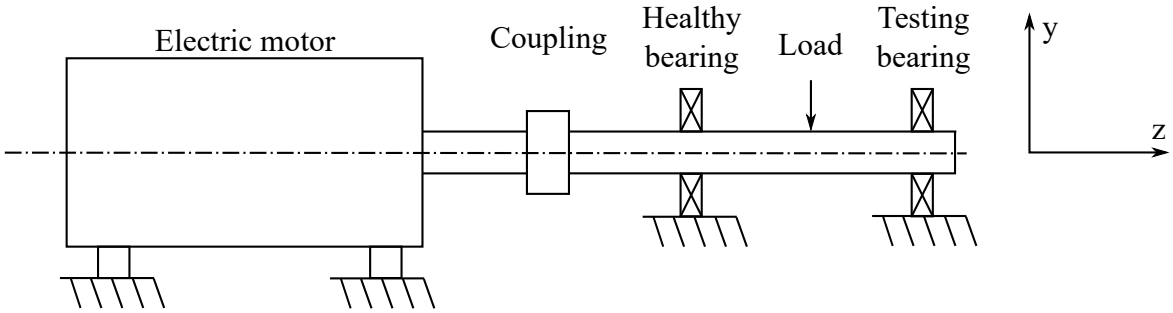


Figure 3.14: A schematic of the test bench that is investigated for the SpectraQuest, Inc. dataset.

condition, while the condition of the right bearing, i.e. the testing bearing, was changed between the different experiments. The four bearing conditions that were investigated for the right bearing are presented in Table 3.1, with the short name that is used in the figures and the analytical defect frequency, calculated from the properties of the bearing, included as well.

Table 3.1: Characteristics of the right bearing of the experimental setup in Figure 3.14.

Testing bearing condition	Short name for measurement	Defect frequency (shaft orders)
Healthy	HH	-
Outer race damage	HO	3.5913
Rolling element damage	HB	2.3751
Inner race damage	HI	5.4087

For each of the bearing conditions presented in Table 3.1, a speed run-up was performed while various signals such as acceleration measurements and a one-pulse per revolution tachometer signal were acquired. The rotational speed profiles for the four cases, using the tachometer signal information, are presented in Figure 3.15. The angular acceleration of the shaft was different for all four cases, with approximately the same maximum rotational speed obtained in each case. The vibration signals in the y-direction for the four cases listed in Table 3.1, sampled at 25.6 kHz, are presented in Figure 3.16. The sensitivity of the acceleration signal's amplitude to rotational speed is clearly observed for all four cases in Figure

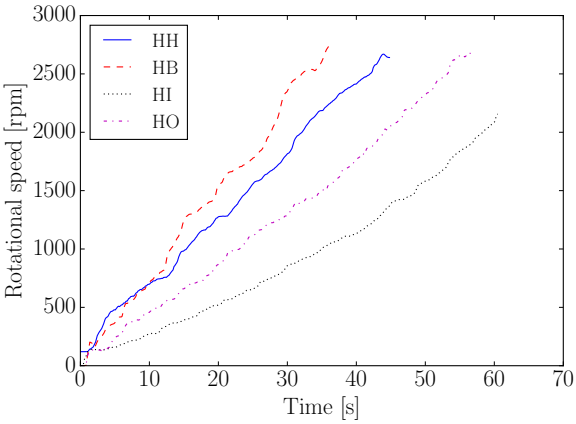


Figure 3.15: Rotational speeds as a function of measurement time of the four datasets from the SpectraQuest, Inc. setup. The data are labelled as follows: HH: Healthy bearing; HB: Rolling element damage; HI: Inner race damage; HO: Outer race damage.

3.16.

3.4.2.2 Results

The proposed feature extraction approach in Section 3.2.2 was applied to the healthy vibration data, whereafter a model of the features was created following the procedure described in Section 3.2.3. The discrepancy metric was generated using Equation (3.3), with the parameters of the aforementioned model, and used with the rotational speed in Figure 3.15 to estimate $\mu_{\eta|\omega}(\omega)$ and $\sigma_{\eta|\omega}(\omega)^2$ with Equation (3.4) and (3.5), respectively. The same feature extraction procedure was applied for the damaged cases, whereafter the discrepancy metric was generated with Equation (3.3) and processed using Equation (3.6). The processed discrepancy metric is analysed using Welch's power spectral density estimate and the results are separately presented for the four bearing conditions in Figure 3.17. More information on the bearings is given in Table 3.1.

The spectrum of the healthy bearing in Figure 3.17(i) contains the expected frequency components such as the shaft orders and its harmonics, but the magnitude of the amplitudes is very small. In the spectrum for the bearing with outer race damage in Figure 3.17(ii), the fundamental outer race component and its harmonics are seen within the spectrum with other frequency components such as the shaft order and its harmonics. The rolling element damage components dominates the spectrum in Figure 3.17(iii), with a slight difference between the theoretical defect frequency and the actual defect frequency presented. The inner race component and its harmonics are not clearly seen in the spectrum in Figure 3.17(iv), because the spectrum is contaminated by other components. The spectrum of all cases presented

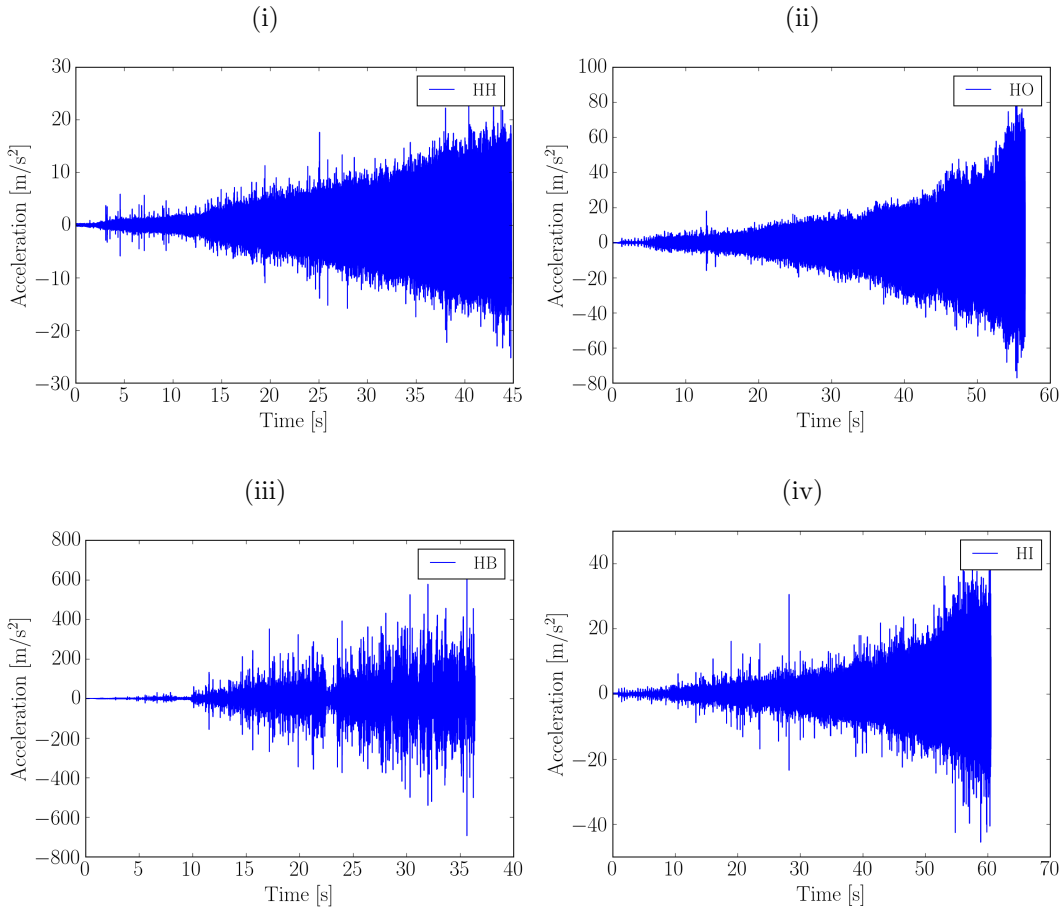


Figure 3.16: The vibration signals for the healthy testing bearing 3.16(i) and the testing bearing that has outer race damage 3.16(ii), rolling element damage 3.16(iii) and inner race damage 3.16(iv) are presented.

in Figure 3.17 is contaminated by harmonics of the Fundamental Train Frequency (FTF) which is equal to 0.4 shaft orders. The presence of these components can be attributed to an improper installation of the bearing.

The discrepancy data versus the rotational speed data are presented in Figure 3.18(i) with the fit $\mu_{\eta|\omega}(\omega)$ and $\sigma_{\eta|\omega}(\omega)$ presented as well. The fit in Figure 3.18(i) does not capture the characteristics of the data properly and this practically indicates that Equation (3.6) does not scale the data appropriately using the current fit. The discrepancy measure, for a fixed rotational speed, is skewed in the direction of increasing discrepancy which means that a Gaussian fit is not appropriate. It is possible to make the distribution more symmetric, by calculating the logarithm of the discrepancy measure i.e. $\log \eta$ instead. Hence, in Equation (3.4), Equation (3.5) and Equation (3.6) the discrepancy measure η is replaced by $\log \eta$ to obtain the result presented in Figure 3.18(ii). The fit is significantly better when using the logarithm of the data and it can be seen that only the lower bound of the fit (i.e. $\mu - 3\sigma$) deviates from the actual lower bound. The spectrum of the new processed logarithm of the

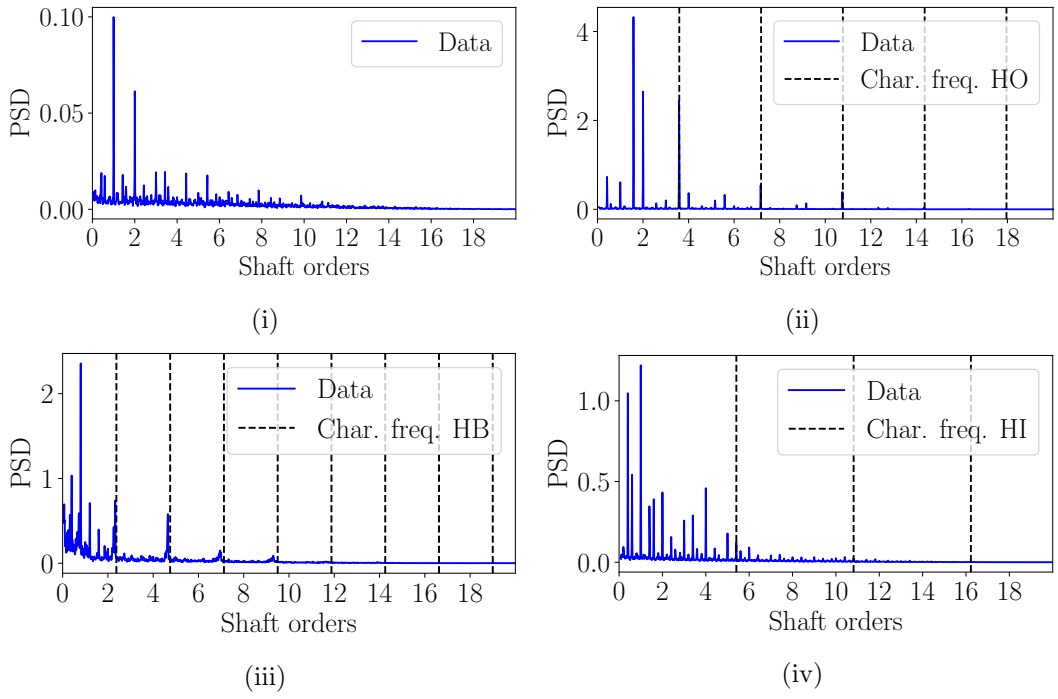


Figure 3.17: The spectrum of the processed discrepancy signals for the case where the testing bearing was healthy is presented in Figure 3.17(i) and where it had outer race damage is presented in Figure 3.17(ii), rolling element damage is presented in Figure 3.17(iii) and inner race damage is presented in Figure 3.17(iv).

discrepancy signal is presented in Figure 3.19.

If the spectrum of the healthy bearing in Figure 3.19(i) is compared to the spectrum of Figure 3.17(i), it can be observed that the shaft order components are seen more clearly. In Figure 3.19(ii), the outer race damage components are very prominent in the spectrum. The contaminating frequency components, surrounding the outer race damage components, are smaller with respect to the amplitude of the outer race damage components when compared to the result in Figure 3.17(ii). The spectrum in Figure 3.19(iii), contains the rolling element damage component and its harmonics. The rolling element bearing damage components seem slightly smaller with respect to the shaft order components when compared to the spectrum in Figure 3.17(iii). The improper scaling used in Figure 3.17(iii) amplified the defect frequencies of the rolling element damage. In the spectrum in Figure 3.19(iii), it is possible to detect the presence of rolling element damage and the scaling with $\log \eta$ can be consistently used to draw conclusions from future datasets.

The inner race damage components are slightly more prominent in the spectrum in Figure 3.19(iv) as compared to the spectrum in Figure 3.17(iv). A zoomed view of the inner race damage spectrum, presented in Figure 3.20, indicates that there is an inner race component present and it is surrounded by the shaft order and the other contaminating components.

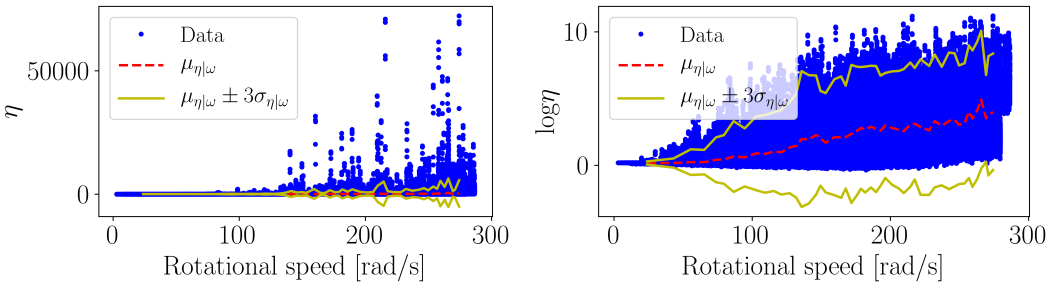


Figure 3.18: The fit of $\mu_{\eta|\omega}(\omega)$ and $\sigma_{\eta|\omega}(\omega)$ is compared in Figure 3.18(i) for the original discrepancy data η and for $\log\eta$ in Figure 3.18(ii).

Hence, from the results it is possible to detect the presence of outer race, rolling element and inner race damage components in varying speed conditions using the proposed methodology. Even though the initial results that were obtained by modelling η made it possible to detect the damage components, the results were more robust by using a more appropriate density i.e. when $\log\eta$ was modelled instead. The large speed changes in this dataset complicated the noise distribution of the discrepancy signal due to possible non-linear effects seen in Figure 3.18. It is therefore suggested that more elaborate noise models need to be investigated for equipment with significant speed changes throughout its operation to ensure that the processed discrepancy signal is speed independent. It is also prevalent from the results that are presented in Figure 3.19 and Figure 3.20 that the proposed procedure is not capable of removing the contaminating noise components generated by the 0.4 shaft order components nor the shaft order harmonics. The proposed method is designed to detect the presence of singularities and impulses and therefore the presence of impulses and singularities in the signal will be reflected in the discrepancy signal as well. The contaminating spectral components can be attenuated by methods proposed in literature for example see Refs. [28, 92].

3.5 Additional investigations

In the introduction of this chapter, it is stated that the methodology is presented without optimising each sub-process. This is because optimising each sub-process is machine dependent and it usually requires historical fault data to be available to maximise the detection capabilities of the model. Hence, the feature extraction and feature modelling steps were designed based on the information from literature and may not be optimal.

However, it is sensible to compare different time-frequency analysis procedures such as

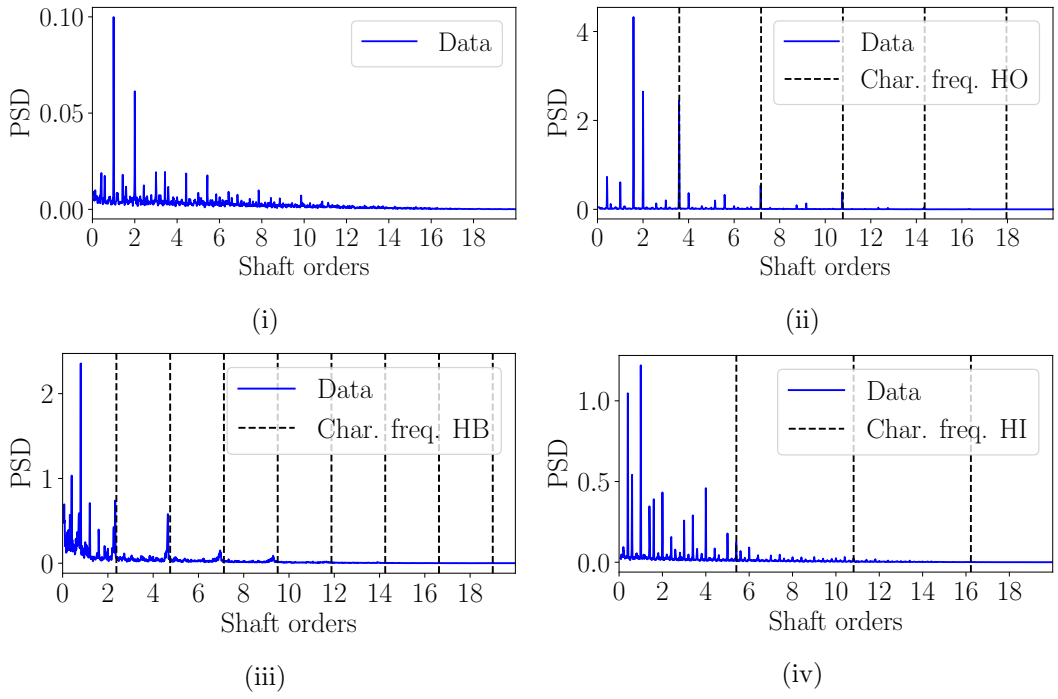


Figure 3.19: The spectrum of the processed discrepancy signals for the case where the testing bearing was healthy is presented in Figure 3.19(i) and where it had outer race damage is presented in Figure 3.19(ii), rolling element damage is presented in Figure 3.19(iii) and inner race damage is presented in Figure 3.19(iv).

the the angle-order spectrogram, the angle-frequency instantaneous power spectrum, the empirical mode decomposition and wavelet packet transform for bearing diagnostics. It is also sensible to compare the performance of Gaussian models to Gaussian mixture models and hidden Markov models i.e. other discrepancy models that have been used in literature as well as a principal component analysis model. Hence, separate investigations were performed to compare different time-frequency decomposition approaches and different models for discrepancy analysis. Only the results are discussed in this section, with the complete investigations and results found in Appendix A.2.2 and Appendix A.2.3.

In summary, the results in Appendix A.2.2 indicate that the angle-frequency instantaneous power spectrum and the wavelet packet transform using the db1, db4 and db44 have potential for discrepancy analysis under varying speed conditions. The benefit of the instantaneous power spectrum is that the basis function does not need to be determined beforehand and therefore it needs to be considered for future investigations using discrepancy analysis. Even though the investigation indicated that the angle-order spectrogram is sensitive damage, it is not suggested for bearing diagnostics, because the extracted features are inherently sensitive to varying speed conditions. The empirical mode decomposition performed poorly on the investigated datasets; this is attributed to the potential occurrence of mode mixing which

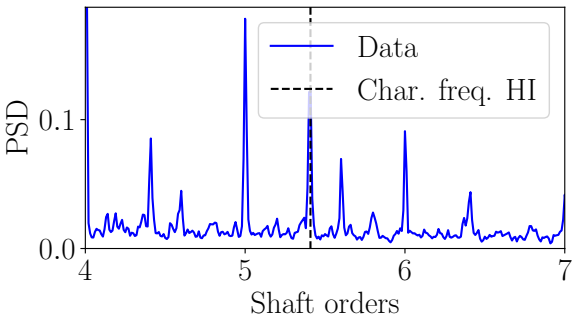


Figure 3.20: Zoomed view of the spectrum of the inner race damage bearing case (HI) originally presented in Figure 3.19(iv).

means that the features do not have a consistent meaning for discrepancy analysis.

The results indicate in Section A.2.3 that the Gaussian mixture model's discrepancy signal is the most sensitive to bearing damage and the most robust to varying speed conditions for the specific dataset under consideration. The Gaussian and hidden Markov models' discrepancy signals are also sensitive to bearing damage, but they are also sensitive to changes in operating conditions. The latter shortcoming can be circumvented by using the discrepancy processing technique in Section 3.2.5. Lastly, the principal component model, which uses the reconstruction error as a discrepancy measure, performed poorly on the investigated datasets.

3.6 Conclusion and recommendations

In this chapter, a novelty detection methodology is presented for bearing diagnostics under varying speed conditions. The methodology follows a similar process as gear discrepancy analysis methodologies [155, 156], however, care is taken to retain the angle-time cyclostationary properties of the bearing signal in the feature extraction process. After the features are extracted, a corresponding discrepancy signal is generated and processed to increase its robustness to varying speed conditions. The power spectral density of the processed discrepancy signal is used to diagnose the bearing.

The numerical gearbox and experimental bearing investigation indicate that it is possible to supplement time-frequency analysis approaches such as the wavelet packet transform with historical data, which can increase the effectiveness of the condition monitoring process under varying operating conditions. For a complex gearbox fault diagnosis methodology, it is suggested that a separate discrepancy methodology be used for the gears and the bearings, because the damage modes manifest differently in the signals. This is a potential avenue that can be investigated for future work.

Chapter 4

Spectral coherence novelty detection methodology

4.1 Introduction

Cyclic spectral analysis techniques such as the spectral correlation and the spectral coherence are some of the most powerful rotating machine fault diagnosis techniques [26, 84, 87]. This is attributed to their ability to effectively represent the cyclic and spectral frequencies of second order cyclostationary signal components in a two-dimensional frequency-frequency plane. The spectral coherence, for example, is able to amplify weak components in the frequency-frequency plane which makes it very effective for incipient fault detection [94]. It is even possible to localise the damaged component by utilising the available prior knowledge of the kinematics of the machine e.g. shaft frequency and bearing fault frequency. These properties make a novelty detection methodology, centred around the spectral coherence, sensible to ensure rotating machines can be effectively diagnosed.

Therefore, in this chapter, a methodology is proposed by combining the spectral coherence with historical data of a healthy machine and by incorporating prior knowledge about the kinematics of the machine. The methodology is capable of not only detecting changes in the condition of the machine, but it is also capable of determining which component is damaged. However, the conventional frequency-frequency spectral coherence is ill-suited for varying operating conditions and therefore the order-frequency spectral coherence, which generalises the spectral coherence to varying speed conditions [31, 97], is used throughout this chapter. The methodology is validated on the phenomenological gearbox model used in Section 3.3.1 and the experimental gearbox data presented in Section 2.3.3.

4.2 Methodology

An overview of the spectral coherence novelty detection methodology is presented in Figure 4.1. A temporal vibration signal $y(t)$ and a corresponding rotational speed signal $\omega(t)$ are

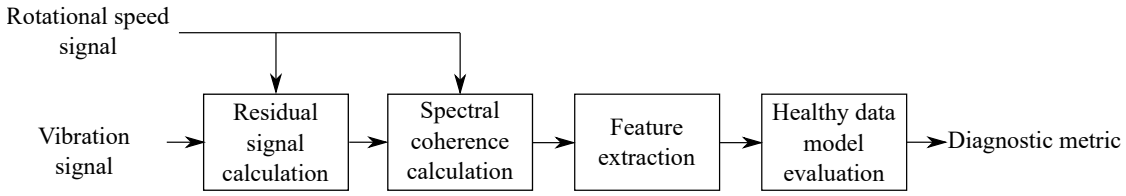


Figure 4.1: Overview of the proposed methodology.

acquired under non-stationary operating conditions, whereafter the residual signal $x(t)$ is calculated by subtracting the generalised synchronous average from the vibration signal. The Order-Frequency Spectral Coherence (OFSCoh) of the residual signal is calculated and features are extracted from the OFSCoh. The extracted features and a probabilistic model, optimised on the healthy data, are used with a discrepancy function to calculate a diagnostic metric. The diagnostic metric indicates the novelty of the measurement and is used to infer the condition of the machine.

A more detailed description of each step in the process is given in the subsections of this section.

4.2.1 Estimating the residual signal

The spectral coherence is a second order cyclostationary analysis technique, therefore it is good practice to remove the first order cyclostationary components from the signal to obtain a residual signal, whereafter the spectral coherence is calculated of the residual signal. It is possible to calculate the residual signal by using the synchronous average to estimate the first order cyclostationary components, whereafter the synchronous average is subtracted from the vibration signal. However, this procedure is only valid under constant speed operating conditions where the cyclo-ergodic assumption is valid. Varying speed conditions result in additional amplitude modulation due to the transmission path effect which violates the cyclo-ergodicity assumption made in the synchronous averaging procedure and therefore the residual signal cannot be properly estimated with the conventional synchronous average [13, 41].

Abboud et al. [28] compared different envelope enhancement techniques under varying speed conditions. According to their results, the generalised synchronous average performs

very well in attenuating deterministic components under varying speed conditions and therefore it is used in this methodology. In the generalised synchronous average, the measured vibration signal $y(t)$ and the rotational speed signal $\omega(t)$ are transformed from the time to the angle domain $y(\theta)$ and $\omega(\theta)$ using order tracking, whereafter the raw estimator of the generalised synchronous average of the signal $y(\theta)$ [41],

$$\hat{m}_y(\theta_\Theta, \omega_r) = \frac{1}{\text{card}\{K_r\,(\omega_r)\}} \sum_{k \in K_r} y(\theta_\Theta + k \cdot \Theta), \tag{4.1}$$

defined for $\theta_\Theta \in [0, \Theta)$, is calculated for R rotational speed regimes $\{\omega_r\}$. In Equation (4.1), the set of all $k \in \mathbb{N}^*$ for which the average rotational speed is in the bounds $[\omega_r - \Delta\omega/2, \omega_r + \Delta\omega/2]$ is given by [41]

$$K_r(\omega_r) = \left\{ k \in \mathbb{N}^* \middle| \omega_r - \frac{\Delta\omega}{2} \leq \frac{1}{\Theta} \int_{(k-1)\Theta}^{k\Theta} \omega(\theta) d\theta < \omega_r + \frac{\Delta\omega}{2} \right\}, \tag{4.2}$$

with the cardinality of K_r denoted $\text{card}\{K_r\}$ and $\mathbb{N}^* = \{1, 2, \ldots\}$.

A smoothed estimate of the generalised synchronous average using kernel density estimators $\mathcal{K}(\cdot)$ is used as opposed to the raw estimate [41]

$$\hat{\mathcal{M}}_y(\theta_\Theta, \omega) = \frac{1}{\sum_{r=1}^{R} \mathcal{K}\left(\frac{\omega - \omega_r}{\lambda}\right)} \sum_{r=1}^{R} \mathcal{K}\left(\frac{\omega - \omega_r}{\lambda}\right) \hat{m}_y(\theta_\Theta, \omega_r), \tag{4.3}$$

where λ is a smoothing parameter and $R \in \mathbb{N}^*$ denotes the number of regimes used in the estimation. Ultimately, the generalised synchronous average is used to estimate the angular residual signal

$$x(\theta) = y(\theta) - \hat{\mathcal{M}}_y\left(2\pi\frac{\theta}{\Theta}, \omega(\theta)\right), \tag{4.4}$$

whereafter the temporal residual signal $x(t)$ can be obtained from the relationship between the angle and the time domain i.e. θ and t.

4.2.2 Order-Frequency Spectral Coherence (OFSCoh)

The OFSCoh of the residual signal $x(t)$ is defined as [31]

$$\Gamma_{2x}(\alpha_\theta, f) = \frac{S_{2x}(\alpha_\theta, f)}{\left(S_{2x}(0, f) S_{2x_{\alpha_\theta}}(0, f)\right)^{1/2}}, \tag{4.5}$$

where $S_{2x}(\alpha_\theta, f)$ denotes the Order-Frequency Spectral Correlation (OFSC) of signal x at cyclic order α_θ and at spectral frequency f. The OFSC is calculated with [31]

$$S_{2x}(\alpha_\theta, f) = \lim_{T\to\infty} \frac{1}{\theta(T)} \mathbb{E}\left\{ \mathcal{F}_T\left(x(t)\right)^* \mathcal{F}_T\left(x(t)e^{-j\alpha_\theta\theta(t)}\omega(t)\right)\right\}, \tag{4.6}$$

where $\omega(t)$ is the rotational speed and $\theta(t) = \int_0^t \omega(u)du$ is the phase of the shaft, T is the time duration of the measurement, and $\mathcal{F}_T(x)$ is the Fourier transform of the signal x over the time interval T. The OFSC of the signal $x_{\alpha_\theta} = x(t)e^{-j\alpha_\theta\theta(t)}\omega(t)$ is denoted by $S_{2x_{\alpha_\theta}}(\alpha_\theta, f)$ in Equation (4.5).

The OFSC given by Equation (4.6) cannot be calculated analytically for measured vibration signals and therefore estimators need to be used. One of the most popular estimators for cyclic spectral analysis [84], namely, the Averaged Cyclic Periodogram (ACP) is used throughout this chapter. The ACP is used because it has good variance and bias properties [41, 98] and it is computationally efficient relative to other estimators [84]. The recent advances in more computationally efficient spectral correlation estimators [94, 95] make it even more beneficial to use the OFSCoh for condition monitoring applications with much data.

4.2.3 Feature extraction

It is possible to learn the complete spectral coherence of a healthy gearbox as an image with advanced machine learning models for example, whereafter the trained models can be used to determine whether the gearbox is healthy or whether a novelty is present from a spectral coherence. However, this learning task is complex, because the complete spectral coherence needs to be learned, many healthy datasets are required to ensure that the parameters can be properly estimated and the output of the model will probably only indicate the presence of anomalous behaviour and not its cause.

Hence, a feature extraction procedure is used in this methodology which reduces the dimensionality of the data and it allows not only novelties to be detected, but the condition of the different machine components in the machine can be inferred as well. This is performed by firstly utilising the available engineering prior knowledge of the machine i.e. the characteristic frequencies of the damage modes that need to be detected, whereafter the statistics are extracted from the spectral coherence at the identified frequencies.

The first step in the feature extraction procedure is to identify the characteristic frequencies of the set of machine components, denoted $\mathcal{S}_c$, that will be monitored. For example, if

$\mathcal{S}_c = \{\text{gear, pinion, BPFO, BPFI, BSF}\}$ is the set of characteristics which is being monitored, the corresponding set of fundamental cyclic orders $\{\alpha_c\} = \{\alpha_{\text{gear}}, \alpha_{\text{pinion}}, \ldots, \alpha_{\text{BSF}}\}$ can easily be determined from bearing catalogues and the gear ratios of the gearbox. Therefore, each characteristic $c \in \mathcal{S}_{cond}$ has a corresponding fundamental cyclic order denoted by α_c which is approximately known beforehand.

Secondly, it is necessary to extract features at the identified set of fundamental cyclic frequencies $\{\alpha_c\}$ and their harmonics. The Improved Envelope Spectrum (IES) has performed very well for enhancing specific frequency bands for bearing diagnostics [28] and therefore it is sensible to use as a starting point. The IES is calculated by integrating the squared spectral coherence over a specific frequency range $f \in [f_l, f_u]$ [98]

$$\text{IES}(k \cdot \alpha_c; f_l, f_u) = \frac{1}{f_u - f_l} \int_{f_l}^{f_u} |\Gamma(k \cdot \alpha_c, f)|^2 df, \tag{4.7}$$

where $k \cdot \alpha_c$ is the kth harmonic of the cyclic order of characteristic c. However, there are a few potential problems when using the IES as a feature in a data-driven fault diagnosis methodology:

- the exact cyclic order of characteristic c, denoted by α_c, needs to be known a priori. Usually, only an analytical estimate is available, which can differ slightly from the actual cyclic order.
- the integration bounds f_l and f_u also need to be known a priori i.e. the important spectral frequency bands are not known without historical fault data. It is possible to integrate over the entire frequency band i.e. $f_l = 0$ and $f_u = f_s/2$, however, this can potentially decrease the sensitivity of the methodology to subtle changes in a specific frequency band, because the signal-to-noise ratio of the spectrum will decrease.

The shortcomings of the IES in a data-driven fault diagnosis methodology are circumvented by using the modified IES

$$\text{IES}_{mod}(k \cdot \alpha_c, f_m; \Delta\alpha, \Delta f) = \frac{1}{\Delta\alpha \cdot \Delta f} \int_{k\cdot\alpha_c - \Delta\alpha/2}^{k\cdot\alpha_c + \Delta\alpha/2} \int_{f_m - \Delta f/2}^{f_m + \Delta f/2} |\Gamma(\alpha, f)|^2 df \; d\alpha, \tag{4.8}$$

as a feature, where $\text{IES}_{mod}(k \cdot \alpha_c, f_m; \Delta\alpha, \Delta f)$ denotes the average value of the spectral coherence for $\alpha \in [k \cdot \alpha_c - \Delta\alpha/2,\ k \cdot \alpha_c + \Delta\alpha/2]$ and $f \in [f_m - \Delta f/2,\ f_m + \Delta f/2]$. The modified IES is a multivariate function of both cyclic frequency $k \cdot \alpha_c$ and spectral frequency

f_m. The shortcomings of using the IES as a feature are circumvented with the modified IES as follows:

- The requirement of knowing the exact cyclic order of the characteristic is relaxed by assuming that the actual cyclic order is in the bounds $[k \cdot \alpha_c - \Delta\alpha/2,\ k \cdot \alpha_c + \Delta\alpha/2]$.
- The IES in Equation (4.8) is integrated for a specific integration bound $[f_l, f_u]$, but the modified IES is now a function of f_m as well, which means that a grid of spectral frequencies f_m can be used to obtain a two-dimensional IES.

The feature given by Equation (4.8) can be generalised by using the concept of a feature window shown in Figure 4.2. A feature window extracts features e.g. the mean or kurtosis

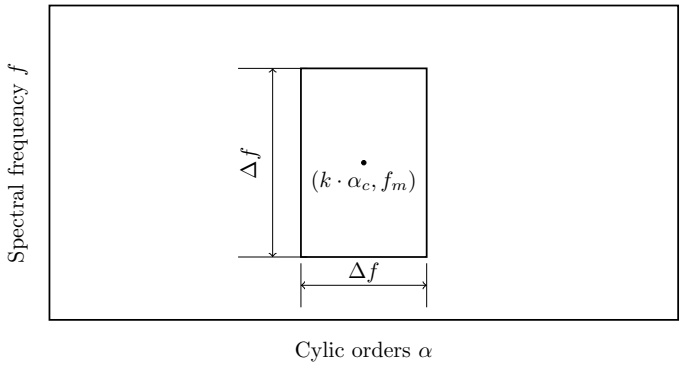


Figure 4.2: A feature window is presented on the $\alpha - f$ spectral coherence plane, where the centre of the rectangle is given by $(k \cdot \alpha_c, f_m)$ and the length and width of the rectangle are given by $\Delta\alpha$ and Δf respectively.

from a rectangular region in the spectral coherence. For example, if the mean of the spectral coherence in the rectangle in Figure 4.2 is calculated, it is equivalent to using Equation (4.8).

In Figure 4.3 an example of the application of feature windows is given, where the feature windows of four characteristics $\mathcal{S}_c = \{A, B, C, D\}$ are presented. It is illustrated that multiple harmonics can be used to extract features from a single characteristic and it is possible to make $\Delta\alpha$ and Δf characteristic dependent and $\Delta\alpha$ harmonic dependent. After the feature windows have been designed, it is possible to extract statistics from each feature window in Figure 4.3 as diagnostic features.

4.2.4 Healthy data model evaluation

A probabilistic model is optimised on the features extracted from the healthy datasets to estimate the probability density function of the healthy features. The estimated probability density function $p(\boldsymbol{b})$ of the features $\boldsymbol{b}$ is used throughout the process. It is subsequently

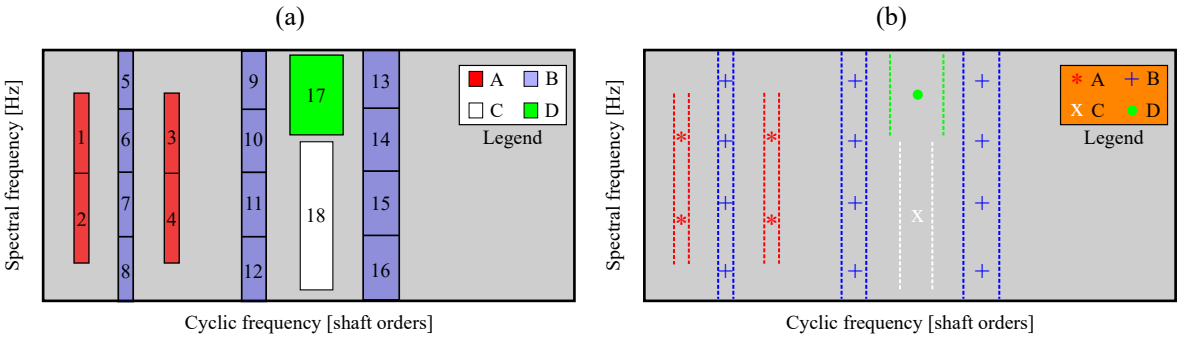


Figure 4.3: Two representations of the feature windows are shown. In (a) the feature windows and their numbering are given, while only the centre of the feature windows and their width are shown in (b). Different feature windows are used on four characteristics $\mathcal{S}_c = \{A, B, C, D\}$, where A can for example be the fundamental shaft order component and its harmonic.

possible to use a diagnostic metric such as the negative logarithm of the probability density function $-\log p(\boldsymbol{b})$ to infer whether the gearbox is healthy or whether anomalous behaviour is observed. However, it is not possible to determine which component is damaged with the diagnostic metric. Hence, a different approach needs to be used to infer the condition of the machine.

The features extracted from the squared-magnitude OFSCoh, denoted by $\boldsymbol{b} \in \mathbb{R}^{D\times 1}$, can be decomposed into N sub-features $\boldsymbol{b} = [\boldsymbol{b}_1, \ \dots \boldsymbol{b}_c, \ \dots \boldsymbol{b}_N]^T$, where $\boldsymbol{b}_c$ denotes the features associated with characteristic c and $N = \text{card}\{\mathcal{S}_c\}$. It is possible to evaluate how well the feature of characteristic c corresponds to the features of characteristic c in the healthy dataset with the marginal distribution

$$p(\boldsymbol{b}_c) = \int p(\boldsymbol{b}) d\boldsymbol{b}_{\acute{c}}, \tag{4.9}$$

where $\boldsymbol{b}_c$ denotes the features of the characteristic that is considered and $\boldsymbol{b}_{\acute{c}}$ denotes the complementary features.

Gaussian models are used in this investigation as opposed to more sophisticated models such as Gaussian mixture models. This is because Gaussian models have fewer parameters and therefore need less training data to ensure that the parameters of $p(\boldsymbol{b})$ are estimated properly. Gaussian models are also computationally efficient to implement with closed form solutions for the marginal and conditional distributions. The Gaussian model is denoted by $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ are calculated for the healthy features. The distribution over the complete feature space is therefore obtained with

$$p(\boldsymbol{b}) = \mathcal{N}(\boldsymbol{b}|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{4.10}$$

while the marginal distribution over the features of a specific characteristic, $p(\boldsymbol{b}_c)$, is obtained with [114]

$$p(\boldsymbol{b}_c) = \mathcal{N}(\boldsymbol{b}_c|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_{cc}), \tag{4.11}$$

where $\boldsymbol{\mu}_c$ is the mean and $\boldsymbol{\Sigma}_{cc}$ is the covariance of the healthy model that correspond to the features set c.

4.2.5 Novelty detection

The full or marginal distribution can easily be used to obtain a discrepancy measure in the form of the negative logarithm of the probability density function for novelty detection. For example, the corresponding discrepancy measure for Equation (4.11) is $-\log p(\boldsymbol{b}_c)$.

In the next section, the different modelling approaches given by Equation (4.10) and (4.11), are compared on numerical data for different fault cases.

4.3 Numerical validation

In this investigation, the phenomenological gearbox model used in Section 3.3 is used to simulate a healthy gearbox and a damaged gearbox that contains different severities of bearing and distributed gear damage. The purpose of this investigation is to validate that it is possible to detect and characterise damage with the proposed methodology and to illustrate the benefits of using the marginal distribution $p(\boldsymbol{b}_c)$ for fault diagnosis.

4.3.1 Phenomenological gearbox model

The casing vibration signal $x_c(t)$ of the gearbox can be decomposed in terms of the damaged outer race bearing component $x_b(t)$, the random gear component attributed to distributed gear damage $x_{dgd}(t)$, the deterministic gear component attributed to the gear mesh interactions $x_{gmc}(t)$ and the noise $x_n(t)$ as follows [28]

$$x_c(t) = x_b(t) + x_{dgd}(t) + x_{gmc}(t) + x_n(t). \tag{4.12}$$

The equations of the different casing vibration components are given by Equations (3.8), (3.9), (3.10) and (3.15). The bearing and distributed gear damage components in Equation (4.12) are decomposed into a fault severity factor FS_i and a baseline signal component $\tilde{x}_i$ to simulate different bearing and gear conditions, where the revised casing vibration signal is

given by

$$x_c(t) = FS_b \cdot \tilde{x}_b(t) + FS_{dgd} \cdot \tilde{x}_{dgd}(t) + x_{dg}(t) + x_n(t). \tag{4.13}$$

The investigated fault severity factors of the bearing and the gear are $FS_b \in \{0, 1, 2\}$ and $FS_{dgd} \in \{0, 1, 2\}$, respectively, with all combinations of the fault severity factors investigated. A healthy gearbox is defined as a gearbox with only the gear mesh and noise components i.e. $FS_b = FS_{dgd} = 0$ in Equation (4.13). The fault severity factors FS_i in Equation (4.13) are used to investigate the sensitivity of the proposed methodology to changes in the vibration amplitudes of the damaged components. The two damage modes have the following fixed properties: The bearing damage excites a natural frequency at 6.0 kHz with a cyclic order of $\alpha_b = 3.59$ shaft orders, while the distributed gear damage excites a resonance band at 1.5 kHz with a cyclic order of $\alpha_{dgd} = 1.0$ shaft orders. More information related to the model used in this investigation is given in Appendix B.2.

The gearbox operates under varying speed and constant load conditions. The rotational speed of the gearbox for a specific measurement is obtained by taking a random sample of a template operating condition profile. The rotational speed of each measurement used in this investigation is presented in Figure 4.4(ii) and obtained from sampling the minimum and maximum values of a template operating condition profile presented in Figure 4.4(i) from a uniform distribution. This is performed to increase the difficulty of the condition monitoring

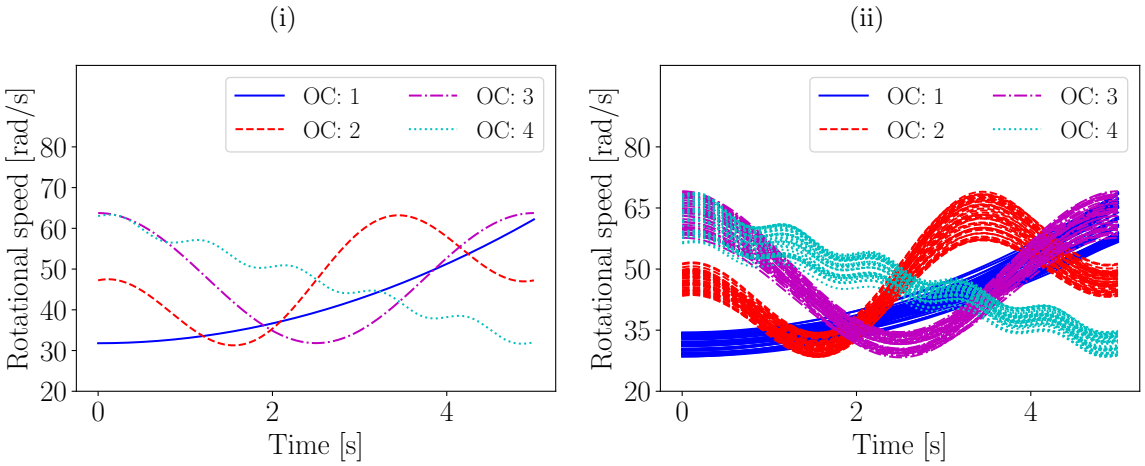


Figure 4.4: The template rotational speed profiles of the four operating conditions are presented in Figure 4.4(i) and the samples from the template rotational speed profile, used to generate the datasets, are presented 4.4(ii). The maximum and minimum values of the template profile are scaled separately with a factor sampled from $\mathcal{U}[0.9, 1.1]$ to obtain the maximum and minimum value of the actual rotational speed profile shown in Figure 4.4(ii). A uniform distribution with a domain of [0.9, 1.1] is denoted $\mathcal{U}[0.9, 1.1]$.

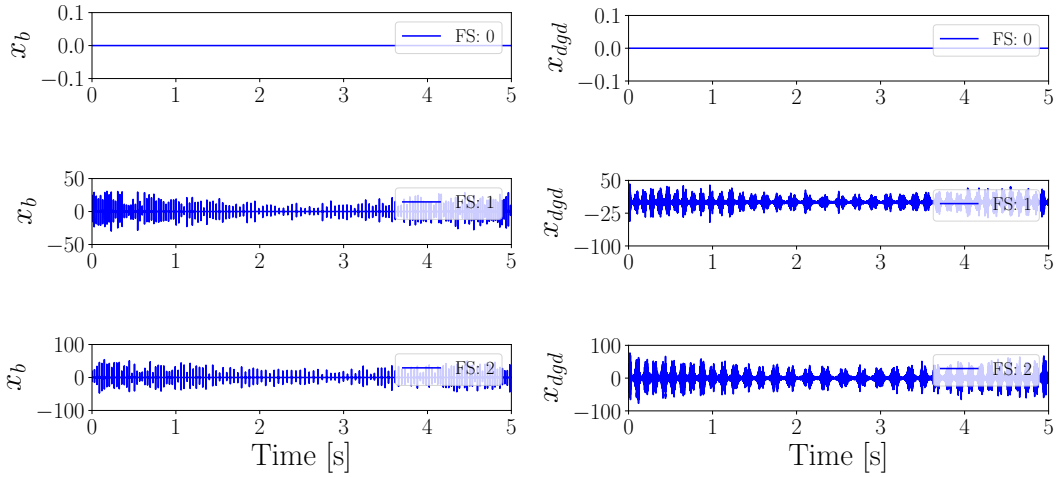


Figure 4.5: The bearing impulses and the distributed gear component magnitudes that correspond to the three investigated magnitudes for operating condition profile 3 (i.e. OC: 3).

task.

In Figure 4.5, the bearing and distributed gear signals are presented for the three fault severities that are investigated, i.e. $x_b(t) = FS_b \cdot \tilde{x}_b(t)$ and $x_{dgd}(t) = FS_{dgd} \cdot \tilde{x}_{dgd}(t)$ are presented.

4.3.2 Results

The methodology is applied on the phenomenological gearbox model discussed in Section 4.3.1, where the outer race component of a bearing and the gear are being monitored for changes in machine condition. Hence, the set of components being monitored consist of the gear and the outer race bearing component i.e. $\mathcal{S}_c = \{\text{dgd}, \text{ b}\}$, with the corresponding cyclic orders $c_{dgd} = 1.0$ and $c_b = 3.59$. These cyclic orders will be used to design the feature windows.

The OFSCoh of the residual casing signal for a gearbox with bearing and gear damage is shown in Figure 4.6(i). The bearing damage is clearly seen at $3.59 \cdot k,\ k \in \mathbb{N}^*$ shaft orders and a frequency band centred around 6 kHz, while the gear damage manifest at $1 \cdot k,\ k \in \mathbb{N}^*$ shaft orders and located at a frequency band centred around 1.5 kHz. The OFSCoh in Figure 4.6(i) confirms that the spectral coherence is capable of detecting second order components in the vibration signal.

The feature windows are designed based on the identified cyclic orders of the monitored characteristics and presented in Figure 4.6(ii). Four harmonics of each characteristic are used in this investigation which results in eight feature windows being used along the cyclic

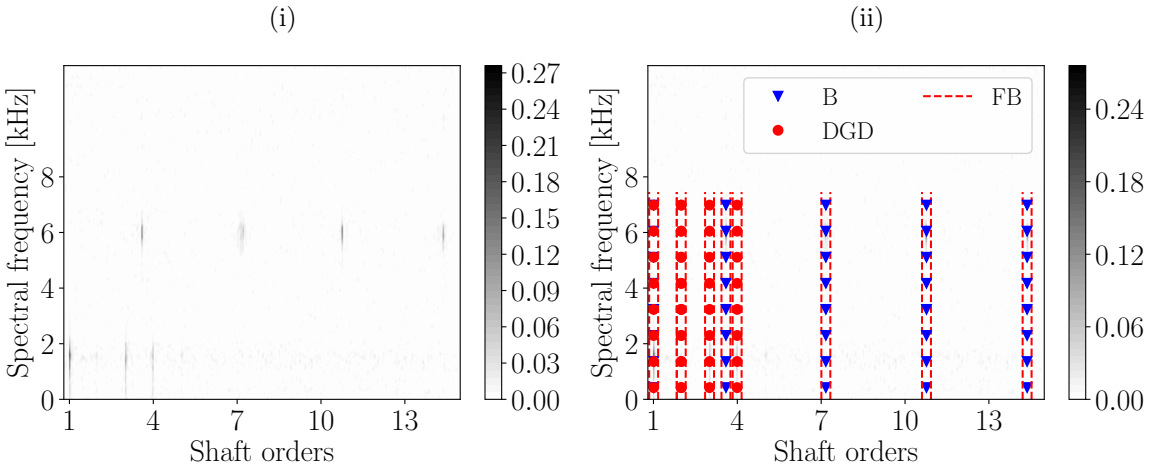


Figure 4.6: The OFSCoh for the gearbox that has bearing and distributed gear damage is shown in Figure 4.6(i) . The raw spectral coherence is shown as well as the 64 feature windows superimposed on the spectral coherence in Figure 4.6(ii). The following abbreviations are used: B - Outer race Bearing component; DGD - Distributed Gear Damage component; FB - Feature Bands.

order axis. However, there is a compromise between the number of feature windows that is used, the resulting number of data points available for estimating the features and the number of features that need to be learned by the model. Hence, eight feature windows along the spectral frequency axis are also used to ensure that the frequency resolution is sufficient to properly estimate the features and to detect bandlimited changes in the frequency bands. The frequency resolution and the cyclic resolution of the feature windows are 937.5 Hz and $\Delta\alpha = 0.16$ shaft orders. Note that if a significant deviation between the analytical characteristic frequency and the actual characteristic frequency are expected it is best to use a $\Delta\alpha$ which is harmonic dependent, e.g. $\Delta\alpha = 0.16 \times k,\ k \in \mathbb{N}^*$, to ensure that the components can be detected at the kth harmonic.

The mean of each feature window is extracted as a feature i.e. Equation (4.8) is used to extract the features, which means that a total of 64 features are extracted from the 64 windows for each measurement. The training features are extracted from 100 healthy measurements ($\text{FS}_{dgd} = \text{FS}_b = 0$ in Equation (4.13)) and subsequently used to obtain the mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ of the healthy model. The same feature extraction procedure is investigated for the different machine conditions and then the healthy model is subsequently evaluated with the procedure described in Section 4.2.4.

The diagnostic metric $-\log p(\boldsymbol{b})$, calculated for the healthy gearbox and the gearbox in the different damaged conditions, is shown in Figure 4.7. This diagnostic metric changes as the condition of the gearbox changes and therefore it can be used to detect novelties.

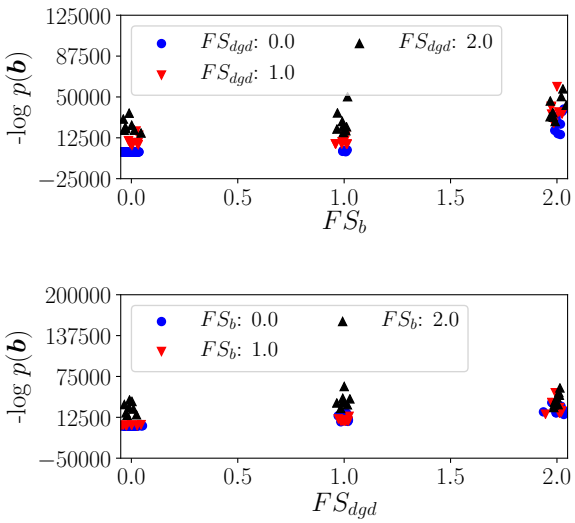


Figure 4.7: The diagnostic metric $-\log p(\boldsymbol{b})$ for all of the features is calculated for the combinations of gearbox conditions.

However, it is not possible to use the diagnostic metric to infer which component is damaged, because it changes as the condition of the bearing and the gear change.

The marginal distribution of each characteristic $p(\boldsymbol{b}_c)$ is also investigated on the same data, with the results presented in Figure 4.8(i). In Figure 4.8(i), the diagnostic metric of the

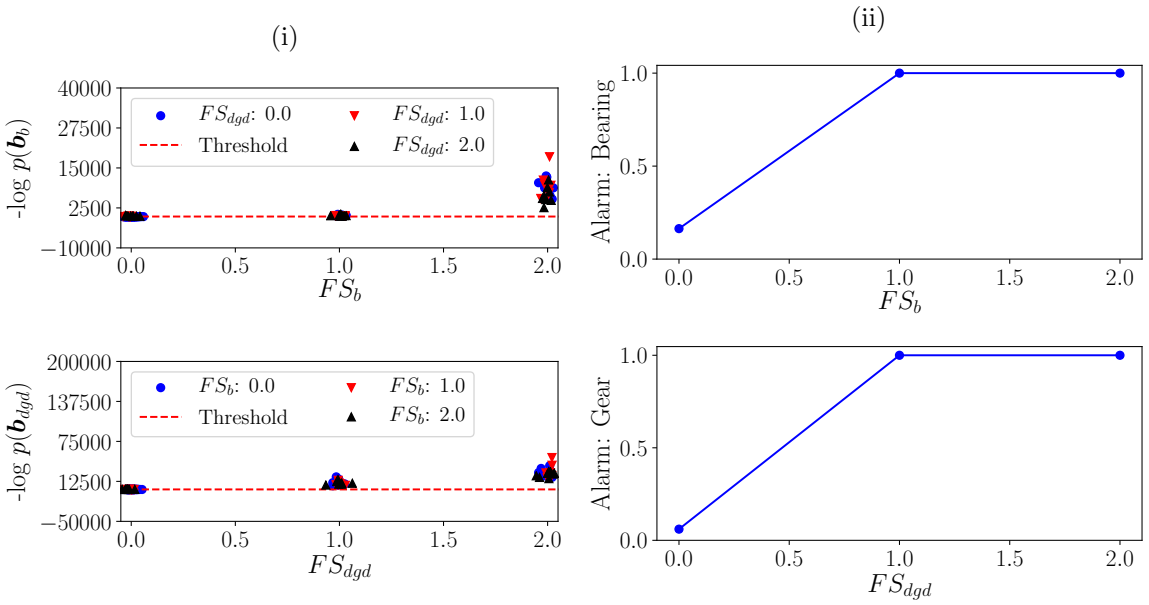


Figure 4.8: The diagnostic metric to detect distributed gear damage and the bearing damage, denoted by $-\log p(\boldsymbol{b}_{dgd})$ and $-\log p(\boldsymbol{b}_b)$, are presented in Figure 4.8(i) for different fault severity factors. The threshold in Figure 4.8(i) is used to calculate an alarm trigger, whereafter the average alarm trigger is presented in Figure 4.8(ii).

bearing $-\log p(\boldsymbol{b}_b)$ increases as the fault severity factor of the bearing FS_b increases, while being unaffected by changes in the condition of the gear. An equivalent result is obtained for

the distributed gear damage diagnostic metric $-\log p(\boldsymbol{b}_{dgd})$; it is sensitive to changes in the fault severity of the gear, while being robust to changes in the condition of the bearing.

A threshold is also calculated with the healthy data in Figure 4.8(ii) and used to determine whether the condition of the gearbox has changed or not. The threshold is compared to each data point in Figure 4.8 to calculate a corresponding alarm value i.e. if the data point exceeds the threshold the alarm value is unity, while being zero for healthy data. An average alarm value is presented in Figure 4.8(ii) for the different data points.

The benefits of using the proposed approach are highlighted by the results in Figure 4.8(ii). The alarm metric is easy to understand and to interpret, while being sensitive to damage and robust to varying operating conditions due to the advanced technology used in the process. Therefore, it is possible to automatically detect and characterise numerical bearing and distributed damage under varying speed conditions with the proposed methodology. In the next section, the methodology is applied to experimental gearbox data, where it is illustrated how to use the methodology for data acquired over the life of the machine.

4.4 Experimental validation

The methodology is investigated on the gearbox dataset presented in Section 2.3.1 to further highlight the potential of the proposed methodology. The ability of the methodology to detect gear damage and also to localise the damaged component are investigated in this section.

4.4.1 Procedure

A gearbox in a healthy condition was operated under varying operating conditions, whereafter the healthy gearbox was disassembled, damage was seeded on the gear, the gearbox was reassembled and operated until the gear tooth failed as discussed in Section 2.3.1. In this investigation, the gear and pinion are monitored i.e. $\mathcal{S}_c = \{gear,\ pinion\}$. From the information contained in Refs. [16], it is expected that the damage will manifest as modulation with a period equal to the shaft rotation and therefore the characteristic orders of the monitored components are $\alpha_{gear} = 1.0$ and $\alpha_{pinion} = 1.85$ shaft orders. The gear ratio of the monitored gearbox is therefore 1.85.

The localised gear damage manifests as localised higher order non-stationary behaviour in the vibration data and therefore attenuating the gear mesh component will enhance the diagnostic information in the signal [85, 86]. The deterministic components are removed

with the same procedure as the numerical data in the previous section i.e. the generalised synchronous average is calculated and subtracted from the vibration signal to obtain the residual signal. Therefore, the spectral coherence of the residual signal is calculated based on specification of [41]. The spectral coherence for two measurements are presented in Figure 4.9(i) and Figure 4.9(iii). The damaged gear is located on the reference shaft and therefore

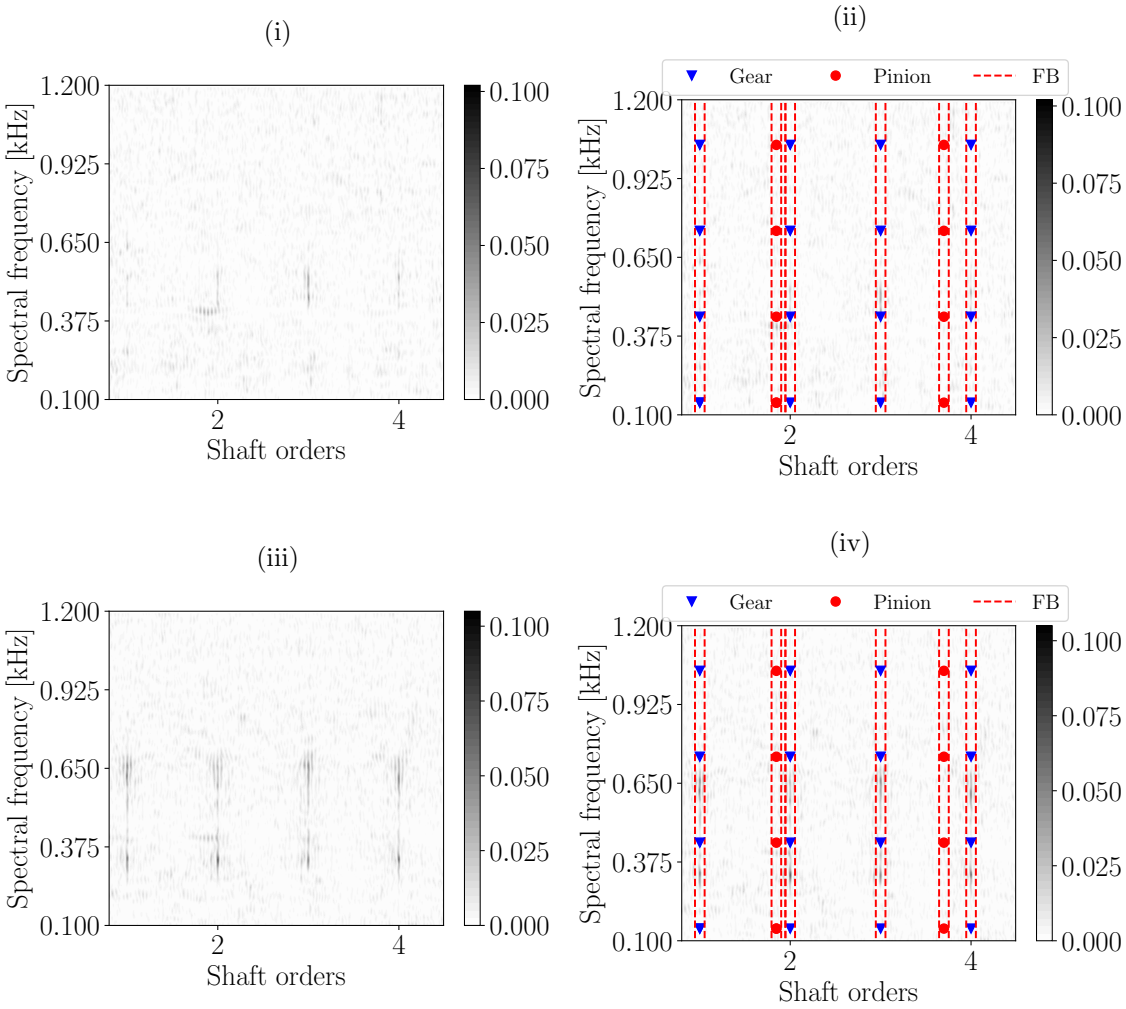


Figure 4.9: The OFSCoh of two damaged gearbox measurements are presented in Figure 4.9(i) and Figure 4.9(iii). The feature windows are superimposed on the two OFSCoh in Figure 4.9(ii) and 4.9(iv).

strong modulations are expected at $1.0 \cdot k,\ k \in \mathbb{N}^*$ shaft orders. The damaged gear component is seen in a frequency band between $[400, 500]$ Hz for the measurement in Figure 4.9(i) and as the damage progresses it manifests in a broader frequency range, with prominent peaks observed at approximately 350 Hz and 650 Hz, as seen in Figure 4.9(iii).

A frequency bandwidth of $[0, 1.2]$ kHz is used with four feature windows used along the spectral frequency axis. The cyclic order bandwidth is $[0, 4.5]$ shaft orders, which allows the

first four harmonics of the gear component and the first two harmonics of the pinion component to be monitored. The resulting 20 feature windows are presented for two measurements of the damaged gearbox in Figures 4.9(ii) and 4.9(iv). To ensure that it is possible to properly estimate the parameters of the model with the available data, it is important to not use too many feature windows relative to the available data.

In the training phase, 35 healthy measurements are used to estimate the parameters of the model, whereafter the model is used to infer the condition of the monitored components i.e. the gear and pinion are compared to a threshold with a similar procedure as Section 4.3.2. The damaged dataset consisted of 103 measurements, evenly spaced over the life of the gear.

4.4.2 Results

The diagnostic metric of the gear and the pinion, denoted $-\log p(\boldsymbol{b}_{gear})$ and $-\log p(\boldsymbol{b}_{pinion})$ respectively, are calculated for each measurement in the damaged dataset and presented in Figure 4.10(i). The diagnostic metric of the gear remains constant for the first few measure-

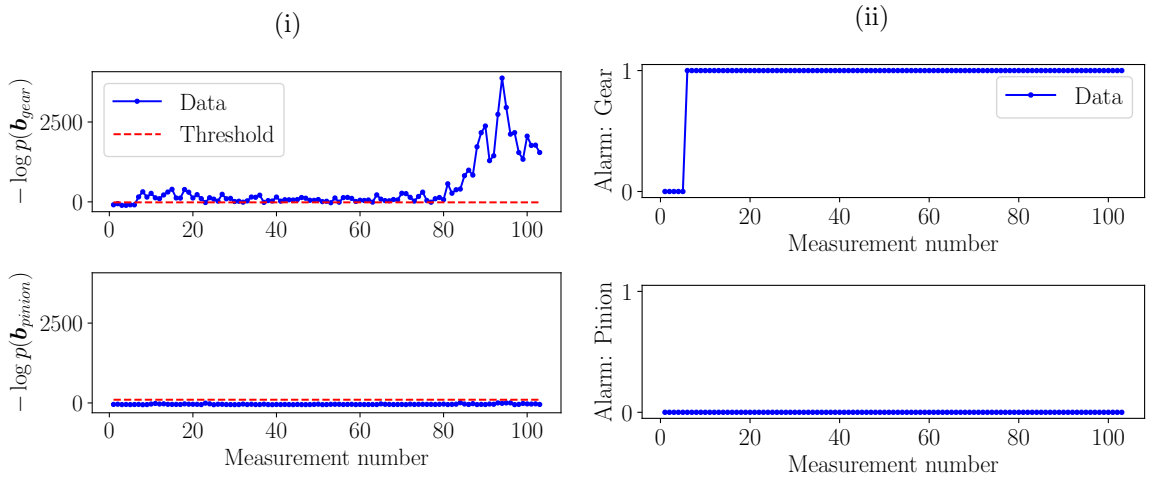


Figure 4.10: The diagnostic metric of the gear and the pinion are presented in Figure 4.10(i) for the damaged measurements. An alarm metric of each measurement is calculated and presented in Figure 4.10(ii). The alarm metric indicates whether the diagnostic metric of a specific measurement exceeds the threshold.

ments, whereafter the modulation in the vibration signal of the gear increases significantly. This resulted in the diagnostic metric to exceed the threshold and it indicates a change in the condition of the gear. After this occurrence, the diagnostic metric remains approximately constant, whereafter another significant change in the diagnostic metric is seen at measurement number 80. This significant change in the diagnostic indicator of the gear indicates

that failure of the gear is imminent and therefore the gearbox has to be stopped immediately. In contrast to the results of the gear diagnostic metric, the diagnostic metric of the pinion remains constant over measurement number and it does not exceed the threshold i.e. the condition of the pinion did not change.

The diagnostic metrics are compared to the threshold to infer the condition of the machine. The mean value of the previous N measurements is calculated and compared to the threshold to obtain the alarm metric shown in Figure 4.10(ii). The results in Figure 4.10(ii) corroborate the observations made in Figure 4.10(i). The alarm metric of the gear indicates that anomalous behaviour is present due to the presence of gear damage, while the alarm value of the pinion indicates that healthy behaviour is observed. It is sensible to use the previous N measurements to calculate an alarm metric of the current measurement, because if the average value of the diagnostic metrics of the previous N measurements is compared to the threshold to calculate the alarm value, it attenuates short-term (non-diagnostic) anomalous behaviour, attributed to anomalous operating conditions for example.

4.5 Conclusion and recommendations

A novelty detection methodology is presented which utilises the available historical data to detect anomalous behaviour in a gearbox. This methodology is not only capable of detecting gear and bearing damage under varying operating conditions, it is also possible to localise the damaged component under varying operating conditions. The methodology is applied for a numerical and experimental gearbox investigation.

In the numerical gearbox investigation, the gearbox is monitored for outer race bearing damage and distributed gear damage under varying speed conditions. In the investigation, it is illustrated that using the negative logarithm of the marginal distribution to calculate a diagnostic metric, damage can be detected and the damaged component can be inferred as well. In the experimental investigation, the gears of the gearbox are being monitored with vibration data acquired under varying load and speed conditions. The results further demonstrate the effectiveness of the methodology to detect, localise the damage component and detect changes in the severity of the damaged component under varying operating conditions.

The results from the two investigations illustrate the benefits of using prior knowledge i.e. knowledge about sophisticated signal processing techniques and the kinematics of the machine, in the fault diagnosis process. It is possible to design an effective fault diagno-

sis methodology that provides robust diagnostic metrics that can be processed to obtain alarm metrics. The alarm metrics are easy to interpret, which simplifies the fault diagnosis procedure under varying operating conditions.

It is suggested that future investigations should compare the suitability of different features (e.g. the energy of a feature window) and different models (e.g. kernel density estimators) for gearbox diagnostics using this methodology.

Chapter 5

Novel information enhancement methodology

5.1 Introduction

Narrowband feature extraction methods such as the kurtogram are very popular for rotating machine fault diagnosis applications [101, 102, 104, 106, 170]. In narrowband feature extraction methods, discussed in Section 1.2.2.6, the frequency band which maximises a specific statistic (e.g. kurtosis) is identified and can subsequently be used to automatically design the bandpass filter (i.e. it specifies the centre frequency and bandwidth of the filter) for envelope analysis. In the paper by Combet and Gelman [103], they illustrated that the spectral kurtosis can be used to design a filter, where frequency bands with much impulsive information are enhanced to improve their ability to detect gear faults.

Incorporating historical data with narrowband feature extraction methods could potentially lead to more effective fault diagnosis. It is therefore desired to develop a methodology where narrowband feature extraction methods and the available historical data of a machine in a healthy condition are utilised to obtain a novelty rich signal (i.e. a signal rich with information not present in the healthy historical data), by attenuating frequency bands with similar characteristics as a healthy machine. This will increase the effectiveness of conventional signal analysis techniques to perform fault diagnosis, because the diagnostic information in the signal under consideration is enhanced.

The methodology is presented in Section 5.2, whereafter the methodology is investigated on phenomenological gearbox data in Section 5.3 and experimental gearbox data in Section 5.4.

5.2 Methodology

The purpose of the methodology is to extract a novel signal $\boldsymbol{x}_{novel} \in \mathbb{R}^{N_x \times 1}$ from a measured vibration signal $\boldsymbol{x} = [x[0], x[1], \ldots x[N_x - 1]]^T$ with N_x samples i.e. $\boldsymbol{x} \in \mathbb{R}^{N_x \times 1}$ by using a dataset from the machine in a reference condition. The reference dataset contains N_{train} measurements and is denoted by $\left\{\boldsymbol{x}_{train}^{(i)}\right\}$, $i = 0, \ldots N_{train} - 1$.

A methodology to extract the novel vibration signal (i.e. it contains novel information) is presented in Figure 5.1. The methodology consists of two main phases, namely a training

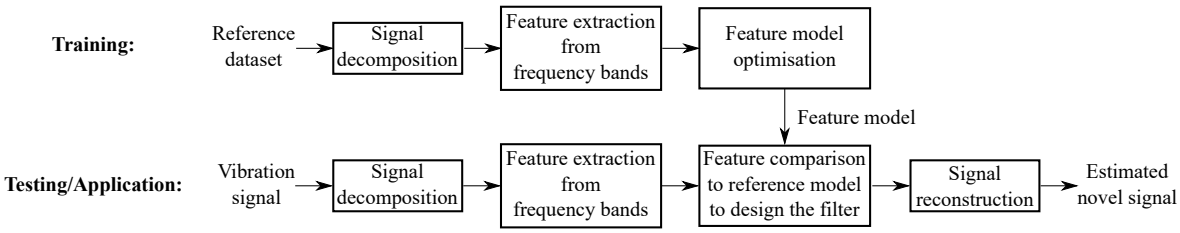


Figure 5.1: An overview of the methodology

phase and a testing phase. The training phase is used to obtain the model parameters through a model optimisation process on the reference dataset. The testing phase uses the model parameters to design a filter that is used to extract a novel signal that corresponds to the input vibration signal.

A detailed overview of each step in the methodology is given in subsequent sections.

5.2.1 Signal decomposition and feature extraction

The assumption is made that the machine component damage manifests in specific frequency bands and results in changes in the statistical properties of the associated feature bands e.g. the damage results in the impulsiveness to increase. Hence, features are extracted from narrowband signals to detect changes in the statistical properties of the frequency bands.

The first step of the feature extraction procedure is to decompose the signal into a time-frequency representation. The Short-Time Fourier Transform (STFT) of the signal $\boldsymbol{x} \in \mathbb{R}^{N_x \times 1}$,

$$S_x[n, m; \Delta f_l] = \sum_{k=0}^{N_w - 1} x[n \cdot (N_w - N_o) + k] \cdot \upsilon[k] \cdot e^{-j2\pi km \frac{\Delta f_l}{f_s}}, \tag{5.1}$$

is used for this task. The STFT is calculated for a pre-determined frequency resolution Δf_l and a corresponding time resolution Δt_l, with the time step and frequency that correspond to $S_x[n, m; \Delta f_l]$ given by $t = n\Delta t_l$ and $f = m\Delta f_l$, respectively. The windowing function

$\boldsymbol{v} \in \mathbb{R}^{N_w \times 1}$ can for example be a Hann function, where N_w is the length of the discrete window function and N_o denotes the number of samples overlapping between consecutive windows. The sampling frequency of the signal $\boldsymbol{x}$ is denoted by f_s.

The optimal window length for the STFT for machine diagnostics depends on the rotational speed of the component [103, 106] and therefore under varying speed conditions the optimal window length becomes difficult to estimate. Therefore, in this methodology N_l time-frequency decompositions are investigated to alleviate the influence of varying speed conditions. The novel information, detected in the different time-frequency distributions, is combined to obtain a single novel signal with the method discussed in Section 5.2.4. In the investigations performed in this chapter, a Hann window is used for all decomposition levels, with the following window lengths $\{N_w\} = \{8, 16, 32, 64, 128, 256, 512, 1024, 2048\}$ and a 75% overlap between consecutive windows being used.

A feature extraction function Ψ_{feat} is used to extract a diagnostic feature $\mathcal{U}(f, \Delta f_l)$ from the temporal signal associated with each centre frequency $f = m \cdot \Delta f_l$ of the STFT with a frequency resolution of Δf_l

$$\mathcal{U}(m \cdot \Delta f_l, \Delta f_l) = \Psi_{feat}(S_x[n, m; \Delta f_l]). \tag{5.2}$$

The feature extraction function Ψ_{feat} can be used to extract the kurtosis of the bandlimited signals, whereafter the kurtogram [56]

$$\mathcal{U}(m \cdot \Delta f_l, \Delta f_l) = \frac{\left\langle |S_x[n, m; \Delta f_l]|^4 \right\rangle_n}{\left\langle |S_x[n, m; \Delta f_l]|^2 \right\rangle_n^2} - 2, \tag{5.3}$$

is obtained. The squared envelope negentropy can also be extracted which results in the squared envelope infogram [105]

$$\mathcal{U}(m \cdot \Delta f_l, \Delta f_l) = \left\langle \frac{|S_x[n, m; \Delta f_l]|^2}{\left\langle |S_x[n, m; \Delta f_l]|^2 \right\rangle_n} \log \left(\frac{|S_x[n, m; \Delta f_l]|^2}{\left\langle |S_x[n, m; \Delta f_l]|^2 \right\rangle_n} \right) \right\rangle_n, \tag{5.4}$$

or the kurtosis and the squared envelope negentropy can for example simultaneously be used as features. In Equations (5.3) and (5.4), $\langle \cdot \rangle_n$ denotes the time averaging operator over N_t time steps e.g.

$$\langle |S_x[n, m; \Delta f_l]| \rangle_n = \frac{1}{N_t} \sum_{n=0}^{N_t - 1} |S_x[n, m; \Delta f_l]|. \tag{5.5}$$

The N_{feat} features extracted from the different frequency bands of the N_l STFT decompositions are written as a vector $\boldsymbol{u} \in \mathbb{R}^{N_{feat}}$ to simplify the notation and computations used in the subsequent sections. The features that correspond to the reference dataset are denoted by $\{\boldsymbol{u}_{train}^{(i)}\}, i = 0, \ldots, N_{train} - 1$ and is the training dataset that is modelled in the next section.

5.2.2 Feature modelling

In this step of the process, it is necessary to determine whether the features of specific frequency bands conform to the expected behaviour governed by the probability density function of the reference features or whether there is novelty in that region.

The features $\boldsymbol{u}$ are parametrised by the Gaussian Probability Density Function (PDF)

$$p(\boldsymbol{u}; \boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u) = \frac{1}{(2\pi)^{N_{feat}/2}} \frac{1}{|\boldsymbol{\Sigma}_u|^{1/2}} \exp\left(-\frac{1}{2} (\boldsymbol{u} - \boldsymbol{\mu}_u)^T \boldsymbol{\Sigma}_u^{-1} (\boldsymbol{u} - \boldsymbol{\mu}_u)\right), \tag{5.6}$$

where the mean and covariance of the features are denoted by $\boldsymbol{\mu}_u$ and $\boldsymbol{\Sigma}_u$ respectively. The Gaussian PDF is used because of its simplicity and effectiveness for detecting changes in the features. It is also sensible to use a Gaussian PDF for the features in this dataset as there can potentially be many features and only a few training datasets, which increases the difficulty of estimating the unknown parameters and hyperparameters of complicated models properly.

The naive assumption is made that the features are uncorrelated i.e. the full covariance matrix reduces to a diagonal matrix with the variances of the respective features being on the diagonal of the covariance matrix. This assumption is made to reduce the number of parameters that need to be estimated from the reference datasets. The mean

$$\mu_u[k] = \frac{1}{N_{train}} \sum_{i=0}^{N_{train}-1} u_{train}^{(i)}[k], \tag{5.7}$$

and diagonal covariance matrix

$$\Sigma_u[a, k] = \frac{1}{N_{train} - 1} \sum_{i=0}^{N_{train}-1} \left(u_{train}^{(i)}[k] - \mu_u[k]\right)^2 \cdot \delta(a - k), \tag{5.8}$$

of the Gaussian PDF in Equation (5.6) are obtained from the healthy features, where δ denotes the Kronecker delta function i.e. $\delta(0) = 1$ and 0 otherwise.

The parameters of the model are denoted $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u\}$ and a novelty detection scoring function $\mathcal{M}(u[k]; \boldsymbol{\Theta})$ is used with an associated threshold $\beta_{thres}[k]$ to detect novelties in

specific frequency bands i.e. a novelty is detected if $\mathcal{M}(u[k];\boldsymbol{\Theta}) > \beta_{thres}[k]$. A good novelty detecting scoring function for a Gaussian PDF is the squared Mahalanobis distance which can be simplified to

$$\mathcal{M}(u[k];\boldsymbol{\Theta}) = \left(\frac{\mu_u[k] - u[k]}{\sqrt{\Sigma_u[k,k]}}\right)^2, \tag{5.9}$$

for a diagonal covariance matrix.

5.2.3 Filter design to enhance novel information

It is desired to retain the frequency bands which contain novel information in the novel vibration signal. This is achieved by using the novelty detection score $\mathcal{M}(u[k];\boldsymbol{\Theta})$ and the associated threshold $\beta_{thres}[k]$ to find the filter coefficient $h[k]$ associated with feature $u[k]$

$$h[k] = \begin{cases} 1.0 & \text{if} \quad \mathcal{M}(u[k];\boldsymbol{\Theta}) > \beta_{thres}[k], \\ 0.0 & \text{if} \quad \mathcal{M}(u[k];\boldsymbol{\Theta}) \leq \beta_{thres}[k]. \end{cases} \tag{5.10}$$

The filter coefficient is equal to unity if a novelty is detected and zero if not. The filter has a similar form to the filter proposed by Combet and Gelman [103], where informative frequency bands are identified and amplified by the square root of the spectral kurtosis. However, in the aforementioned paper, the informative frequency bands were determined from a uniform statistical threshold and did not use historical data in the process. The threshold in Equation (5.10) is $\beta_{thres}[k] = 9$ for $k = 0, \ldots, N_{feat} - 1$. The motivation for the threshold is that it corresponds to $\mu \pm 3\sigma$ which covers 99.7% of the probability density function of a Gaussian distribution.

In Equation (5.10), it is possible to use a more sophisticated filter as well; for example, if a novelty is detected $h[k] = \mathcal{M}(u[k];\boldsymbol{\Theta})$, i.e. the novel information is amplified proportionally to its novelty score. However, the binary filter is investigated here to reduce the number of hyperparameters in the model.

5.2.4 Novel signal estimation

The novel signal is estimated from the novelty filter in two steps. Firstly, each frequency band m of the time-frequency representations of the different window lengths l are scaled by the novelty filter

$$\tilde{S}_x[n, m; \Delta f_l] = h[\kappa(m, l)] \cdot S_x[n, m; \Delta f_l], \tag{5.11}$$

where $\tilde{S}_x[n, m; \Delta f_l]$ denotes the scaled time-frequency representation $S_x[n, m; \Delta f_l]$. The filter coefficient associated with the centre frequency $m \cdot \Delta f_l$ and decomposition level l is given by $h[\kappa(m, l)]$, where $\kappa(m, l)$ is a function which relates the centre frequency and decomposition level to the index of a filter coefficient. In the second step, the inverse STFT of the time-scale distribution at level l, S_x^{-1}, is calculated by the scaled time-frequency representation to obtain an estimation of the novel signal

$$\hat{x}_{novel}[n; \Delta f_l] = S_x^{-1}(\tilde{S}_x[n, m; \Delta f_l]). \tag{5.12}$$

The $\hat{}$ makes it explicit that an estimate of the novel signal is obtained. The inverse STFT S_x^{-1} can be used to perfectly reconstruct the signal if the constant overlap-add constraint is satisfied.

The dependence of the estimation on the frequency resolution Δf_l of the time-frequency decomposition is emphasised in Equation (5.12). It is ideal to use an estimate that does not depend on the frequency resolution i.e. to use $\hat{x}_{novel}[n]$ instead of $\hat{x}_{novel}[n; \Delta f_l]$. This can be approximated with a weighted sum of the novel signals associated with each decomposition l

$$\hat{x}_{novel}[n] = \sum_{l=0}^{N_l-1} \hat{x}_{novel}[n; \Delta f_l] \cdot w[l], \tag{5.13}$$

where $\boldsymbol{w} \in \mathbb{R}^{N_l \times 1}$ denotes the weight vector, which has the following constraint $\sum_{k=0}^{N_l-1} w[k] = 1$. There are many potential methods to select the elements of the weight vector $\boldsymbol{w}$, with the simplest being to assign an equal weight to each frequency band, which results in

$$\hat{x}_{novel}[n] = \frac{1}{N_l} \sum_{l=0}^{N_l-1} \hat{x}_{novel}[n; \Delta f_l]. \tag{5.14}$$

Finally, the estimated novel signal can for example be analysed using cyclostationary techniques for the presence of damage in the machine components. Different techniques are investigated for the datasets and described in the following sections.

The methodology, discussed in Section 5.2.1 to Section 5.2.4, is summarised in Figure 5.2 to clarify the steps followed in the training and testing phases.

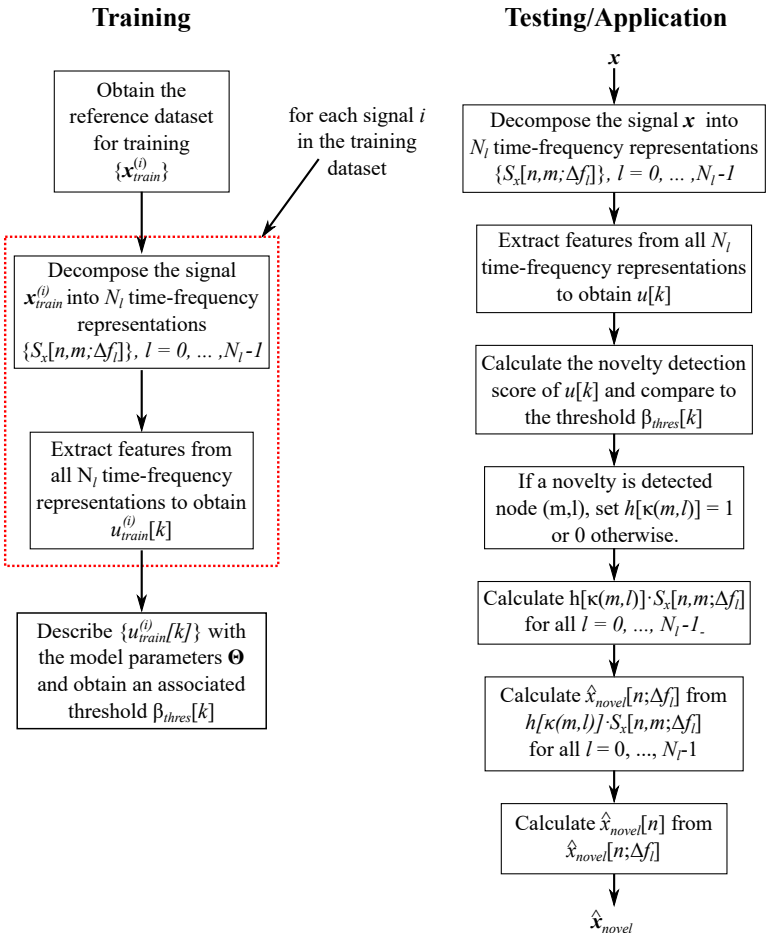


Figure 5.2: The proposed methodology is illustrated to obtain a novel vibration signal $\hat{\boldsymbol{x}}_{novel}$ from a vibration signal $\boldsymbol{x}$. In the training phase, the reference dataset is used to obtain the model parameters, whereafter a novel signal is extracted from the vibration signal in the testing phase.

5.3 Phenomenological gearbox model

The phenomenological gearbox model, used in Section 3.3.1, is used to generate data of a gearbox operating in various machine conditions under constant load and varying speed conditions. The model is slightly modified to include inner race damage as well.

5.3.1 Model

The phenomenological gearbox model is used to generate data in a reference condition, whereafter additional inner and outer race bearing damage is simulated in the gearbox. The gearbox signal in its reference condition consists of a gear mesh component x_{gmc}, a noise component x_n and a distributed gear damage component x_{dgd}, with the corresponding casing vibration signal given by

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + x_n(t). \tag{5.15}$$

The condition of the gearbox in its reference condition is changed by introducing localised outer race damage $x_{blo}(t)$ and localised inner race damage $x_{bli}(t)$ with the updated casing vibration signal given by

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + FS_{blo} \cdot x_{blo}(t) + FS_{bli} \cdot x_{bli}(t) + x_n(t). \tag{5.16}$$

The fault severity factors FS_i help to explicitly change the magnitude of the bearing damage components with respect to the other signal components e.g. $FS_i = 0$ indicates that the associated signal component is not present in the signal.

The gear mesh component, the distributed gear damage component and the noise component are given by Equations (3.8), (3.9) and (3.15) respectively and discussed in Section 3.3.1. The outer race bearing damage component

$$x_{blo}(t) = M(\omega(t)) \cdot h_{blo}(t) \otimes \left(\sum_{k=0}^{N_{\mathcal{T},blo}-1} A_{blo}(k) \cdot \delta\left(t - \mathcal{T}_{blo}^{(k)}\right) \right), \tag{5.17}$$

consists of a train of Dirac impulses δ convolved with the impulse response function $h_{blo}(t)$. The impulses are random in magnitude, captured by the random variable $A_{blo}(k)$, and are modulated by the rotational speed of the system with the function $M(\omega(t))$ as well. The time-of-arrival of bearing impulse k, denoted by $\mathcal{T}_{blo}^{(k)}$, is influenced by the rotational speed of the system as well as by slip [97]. Hence, the instantaneous phase of the shaft is used to

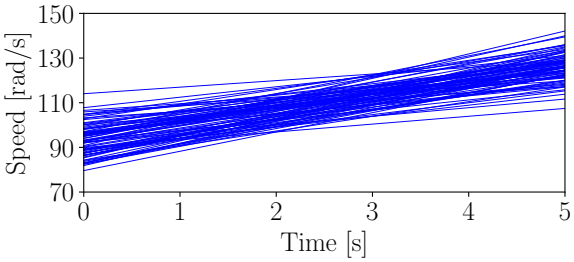


Figure 5.3: The rotational speeds used for the phenomenological gearbox model are presented, with the rotational speed associated with a specific measurement obtained from $\omega(t) = at+b$. The parameters a and b are sampled from a normal distribution with unit variance, with $\mathbb{E}\{\omega(t)\} = 10\pi \cdot t + 30\pi$ rad/s.

relate the angle and time of the impulses whereafter slippage effects are introduced by adding noise from a uniform distribution to the time of arrival of the impulses.

The inner race damage component

$$x_{bli}(t) = z_{stribeck}\left(\int_0^t \omega(\tau)d\tau\right) \cdot M(\omega(t)) \cdot h_{bli} \otimes \left(\sum_{k=0}^{N_{\mathcal{T},bli}-1} A_{bli}(k) \cdot \delta\left(t - \mathcal{T}_{bli}^{(k)}\right)\right), \quad (5.18)$$

has a similar form as the outer race component, except for the presence of the Stribeck equation [88]

$$z_{stribeck}(\varphi) = \begin{cases} z_0 \cdot \left(1 - \frac{1}{2\epsilon}\left(1 - \cos(\varphi)\right)\right)^{c_{str}} & \text{for } |\text{wrp}(\varphi)| < \varphi_{max} \\ 0 & \text{otherwise,} \end{cases} \quad (5.19)$$

which simulates the damaged portion of the inner ring moving in-and-out of the load zone. The wrap function, denoted by $\text{wrp}(\varphi)$, returns the phase of the shaft φ in the domain of $[-\pi/2,\ \pi/2]$. The constants associated with the Stribeck equation are selected as [88]: $c_{str} = 3/2$, $\epsilon = 0.49$, $\varphi_{max} = 0.99\pi/2$rad, and $z_0 = 1$.

However, it is important to highlight the model characteristics that assist with interpreting the results; the ball-pass outer race and inner race orders of the bearing are 4.12 and 5.88, while, the natural frequencies being excited by the outer race damage and inner race damage are 7.0 kHz and 5.5 kHz, respectively.

The rotational speeds for all casing vibration signals are linear ramp-ups (i.e. of the form $\omega(t) = at + b$), with the initial rotational speed and the gradient sampled from a normal distribution. The random rotational speed profiles are presented in Figure 5.3.

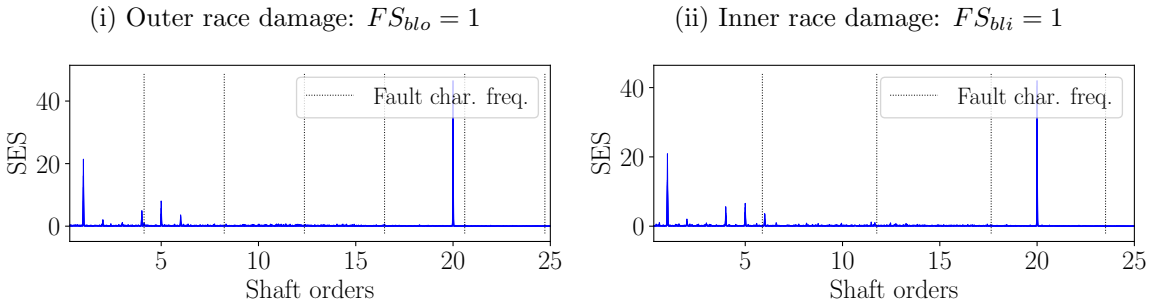


Figure 5.4: The Squared Envelope Spectrum (SES) of the original casing vibration signal (i.e. without any pre-processing performed) from the phenomenological gearbox model that contains outer race bearing damage and inner race damage, is presented. The analytical fault frequencies and their harmonics are superimposed on the spectra as well.

5.3.2 Results with raw data

The Squared Envelope Spectrum (SES) is one of the most powerful tools in bearing diagnostics [28] and therefore it is investigated on the casing vibration signals. In Figure 5.4(i), the SES of the raw vibration signal is presented, which corresponds to the bearing with outer race damage. In Figure 5.4(ii), the SES of the raw vibrations signal, which corresponds to the bearing with inner race damage is presented. The distributed gear damage manifests at the shaft orders and their harmonics are clearly seen, while the gear mesh component is prominent at 20 shaft orders. The bearing damage signal components' are not seen in Figure 5.4 for both cases. Usually, the SES of the random part is calculated i.e. where the deterministic components such as the gear mesh components are attenuated [41], but the distributed gear damage will not be removed due to the statistical nature of its characteristics i.e. it is second-order cyclo-non-stationary. The complete signal, as opposed to the residual signal, is used in all analyses to illustrate the ability of the proposed methodology to remove the historical information from the data.

It is possible to design a bandpass filter by selecting the centre frequency band and the frequency resolution that maximise the SE infogram shown in Figure 5.5(i) for the outer race damage and in Figure 5.5(ii) for the inner race damage. However, the distributed gear damage signal components dominate the feature plane which results in the wrong bandpass filter to be designed i.e. the resulting bandlimited signal only contains the distributed gear damage information. Hence, it is sensible to use the proposed methodology to incorporate the information of the historical data of the machine in a reference condition into the signal analysis procedure to assist with diagnosing the gearbox. The proposed methodology is investigated in the next section.

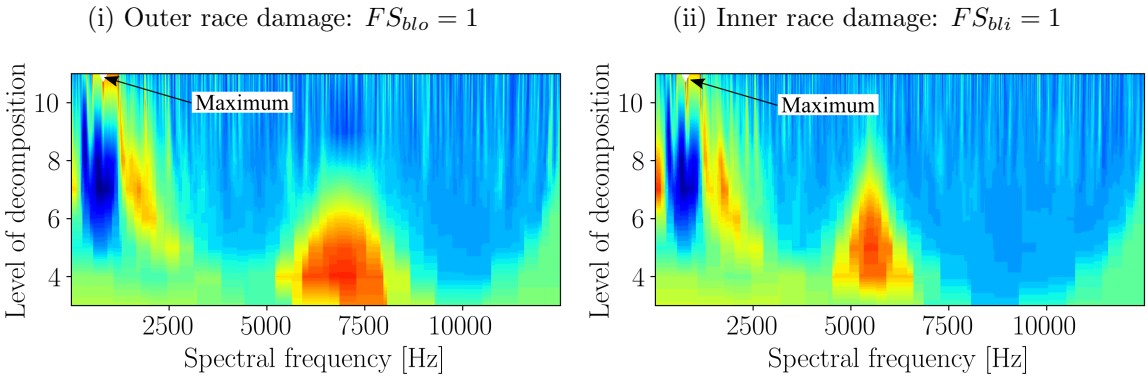


Figure 5.5: The squared envelope infogram of the raw casing vibration signal of the phenomenological gearbox model that contains outer race and inner race damage respectively, is presented. The infogram in Figure 5.5(i) corresponds to the SES in Figure 5.4(i) and the infogram in Figure 5.5(ii) corresponds to the SES in Figure 5.4(ii).

5.3.3 Results with proposed methodology

Twenty datasets of the gearbox in its reference condition are generated with Equation (5.15) and used in the training phase of the methodology. The squared envelope infogram given by Equation (5.4) is used as the features in this investigation, however other features could also be investigated as alternatives as discussed in Section 5.2.1. The model parameters are obtained from the features of the reference dataset using the procedure described in Section 5.2.2.

In the testing phase, the vibration data of the phenomenological gearbox model that contains either inner race or outer race damage are considered. Different magnitudes of the signal components associated with the damaged bearing are investigated by changing FS_i in Equation (5.16). This is performed to ensure that it is possible to not only detect and localise the bearing damage, but that it is also possible to infer changes in the condition of the bearing. The methodology is applied on the datasets by using the same feature extraction procedure, whereafter the model is used to assign a novelty score to each feature in the feature map with Equation (5.9). This is used with the threshold to design the filter given by Equation (5.10), whereafter the estimated novel signal is obtained with the procedure described in Section 5.2.4.

The squared envelope infogram is presented in Figure 5.6 for the novel signals corresponding to the squared envelope infogram in Figure 5.5. It is evident that the bearing damage located in the higher frequency bands are retained in the filtering process, while the distributed gear damage components are not dominating the feature plane any more. It is

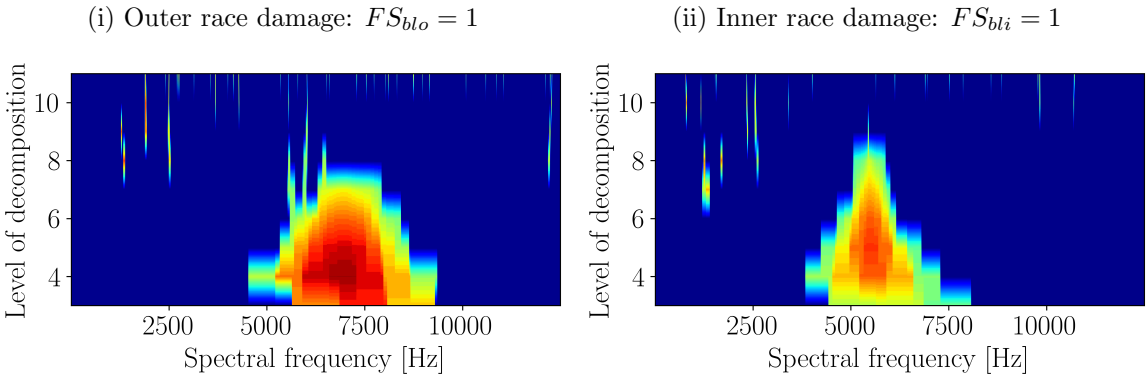


Figure 5.6: The squared envelope infogram of the estimated novel signal of the phenomenological gearbox model that contains outer race and inner race damage respectively, is presented. The raw signal counterpart of Figure 5.5(i) is presented in Figure 5.6(i), while the raw signal counterpart of Figure 5.5(ii) is presented in Figure 5.6(ii).

also observed that some spurious frequency bands are retained in the higher decomposition levels. This is possibly attributed to the statistics being improperly estimated due to the poor time resolution. However, with the averaging approach in Equation (5.14), the effect of the spurious components is attenuated.

The SES of the novel signals are presented in Figure 5.7. A significant improvement is observed in Figure 5.7 compared to the results in Figure 5.4 for a $FS_i = 1$. The bearing fault frequencies and the associated modulations are prominent in the spectra of the novel signals, which is in contrast to the SES of the raw signals where the distributed gear damage and the fundamental gear mesh frequency dominated the lower frequency range of the spectrum. The squared envelope spectra in Figure 5.7 also indicate that changes in the condition of the bearings can be detected i.e. an increase in FS_i results in an increase in the amplitudes of the fault components.

The vibration signals are cyclo-non-stationary under varying speed conditions, while the bearing vibration signal can be approximated as an angle-time cyclostationary signal. This makes the Order-Frequency Spectral Correlation (OFSC) [31] a powerful tool for analysing the vibration signals. The OFSC in Equation (4.6) is estimated with the Averaged Cyclic Periodogram (ACP) for the considered discrete signals [31].

The spectral correlation is presented for a bearing with outer race damage in Figure 5.8(i) and Figure 5.8(ii) for the original and estimated novel signals respectively. As opposed to the results in Figure 5.8(i) that are dominated by the gear components at 20 shaft orders, the bearing damage can be seen at the spectral frequency of 7 kHz, spaced at cyclic orders of 4.12 in Figure 5.8(ii).

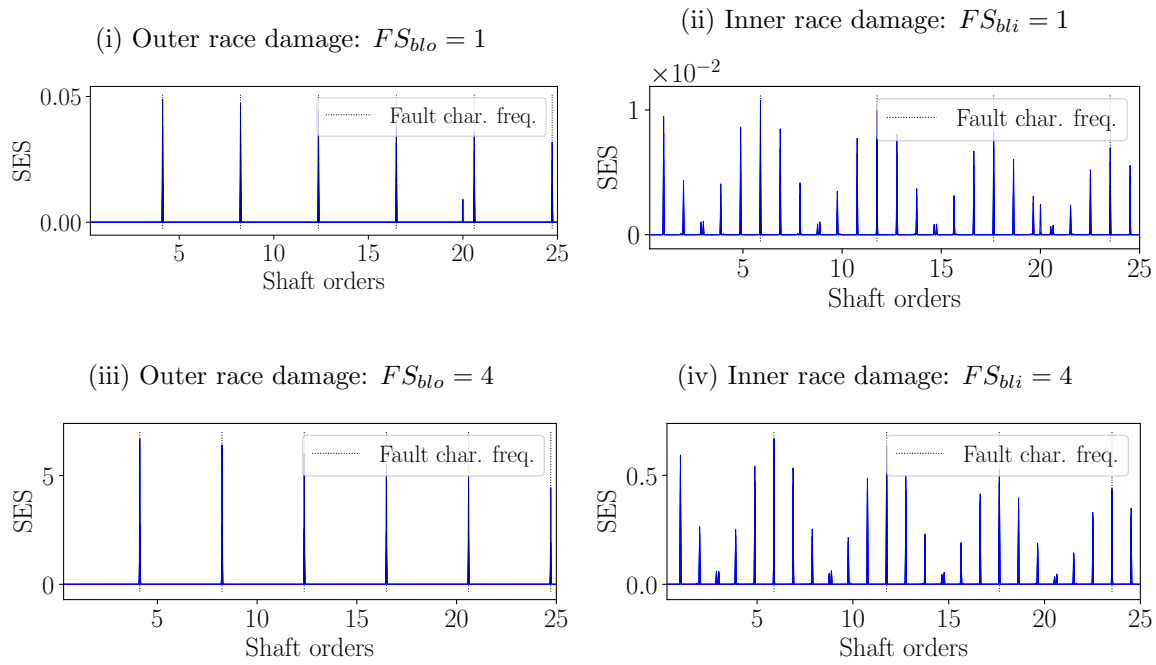


Figure 5.7: The Squared Envelope Spectrum (SES) of the novel vibration signals for outer race and inner race bearing damage are shown. The outer race damage spectra with FS_i of 1 and 4 are shown in Figure 5.7(i) and Figure 5.7(iii) respectively, while the inner race damage spectra with FS_i of 1 and 4 are shown in Figure 5.7(ii) and Figure 5.7(iv) respectively.

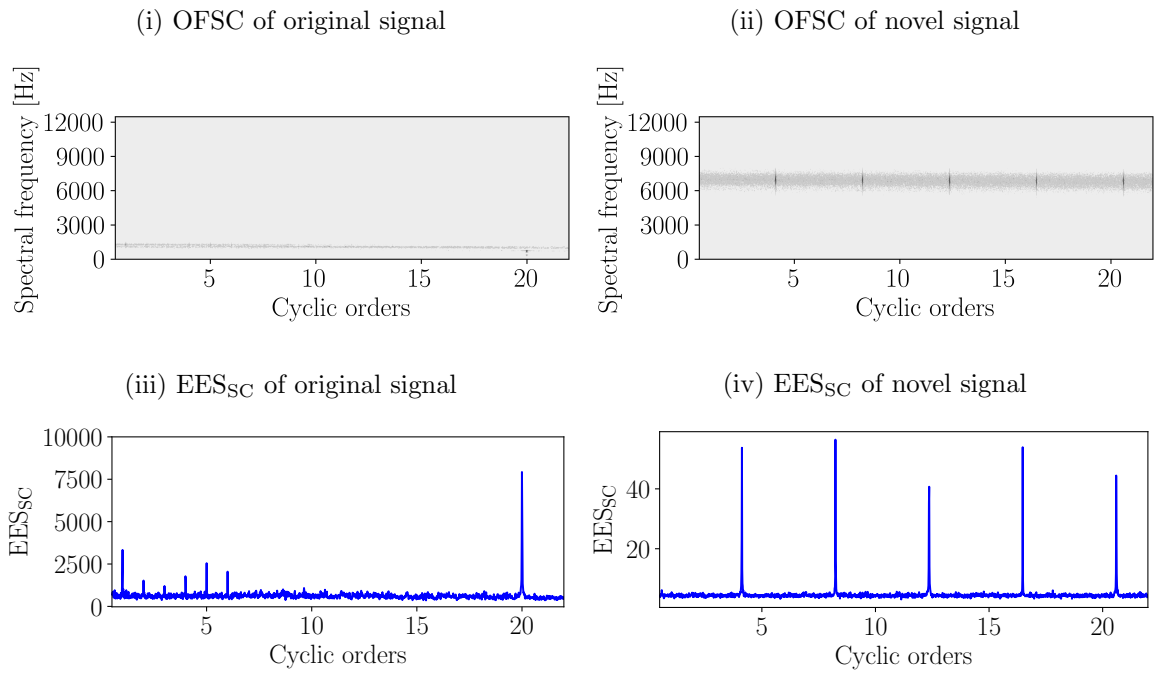


Figure 5.8: The magnitude of the spectral correlation of the gearbox with outer race damage is presented in Figure 5.8(i) for the original signal and 5.8(ii) for the estimated novel signal. The corresponding Enhanced Envelope Spectrum based on the Spectral Correlation EES_{SC} is shown in Figure 5.8(iii) for the original signal and Figure 5.8(iv) for the novel signals.

The Enhanced Envelope Spectrum (EES), estimated with the Spectral Correlation (SC)

$$\mathrm{EES_{SC}}(\alpha_\theta) = \int |\mathrm{S}_{2x}(\alpha_\theta, f)| df, \tag{5.20}$$

highlights the spectral frequency components that are present in the OFSC and is presented in Figure 5.8(iii) and Figure 5.8(iv) for the spectral correlation of the original and novel signals investigated in Figure 5.8(i) and Figure 5.8(ii). The $\mathrm{EES_{SC}}$ of the original vibration signal is dominated by the distributed damage and the gear mesh component, while the $\mathrm{EES_{SC}}$ of the novel vibration signal is similar to the results seen in the squared envelope spectrum in Figure 5.7(iii). Hence, by using the estimated novel vibration signal of the vibration signal, the novel information associated with the bearing damage is highlighted, while the historical information is attenuated. Similar conclusions were drawn from the inner race damage signal results and therefore the results are excluded for the sake of brevity.

The Order-Frequency Spectral Coherence (OFSCoh), given in Equation (4.5), is another useful technique for analysing second-order cyclo-non-stationary signals. The OFSCoh is a normalised representation of the SC, which helps to amplify small second-order components. The Enhanced Envelope Spectrum ($\mathrm{EES_{SCoh}}$) based on the Spectral Coherence (SCoh) [94]

$$\mathrm{EES_{SCoh}}(\alpha_t) = \int |\Gamma(\alpha_t, f)| df, \tag{5.21}$$

is calculated from the OFSCoh and is another useful quantity to detect small second-order cyclostationary components. The OFSCoh is presented in Figure 5.9(i) and Figure 5.9(ii) for the original and the estimated novel vibration signals also investigated in Figure 5.8. The attenuated gear components are more prominent in the OFSCoh as opposed to the OFSC. This is attributed to the fact that the OFSCoh enhances small signal components, which is also verified by the $\mathrm{EES_{SCoh}}$ of the signals, which shows that the attenuated gear components are enhanced. The bearing damage is more prominent with respect to the gear components in Figure 5.9(iv) than in Figure 5.9(iii). This illustrates two important points that while the OFSCoh and the $\mathrm{EES_{SCoh}}$ are very powerful and important for detecting small second-order cyclostationary components in vibration signals, these techniques do not accurately convey the novel information content of the vibration signal (i.e. they amplify the information that was previously attenuated).

The results in this investigation illustrate the potential of the proposed methodology. It

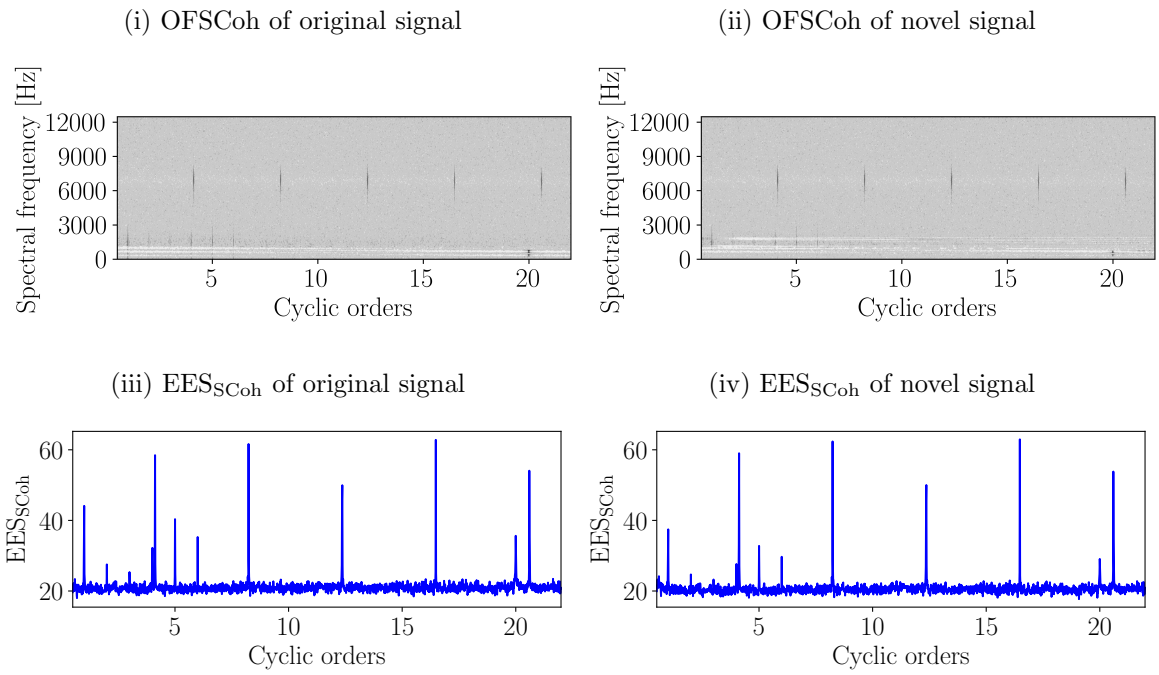


Figure 5.9: The magnitude of the Order-Frequency Spectral Coherence (OFSCoh) is shown for the original signal with outer race damage in Figure 5.9(i) and the estimated novel signal in Figure 5.9(ii). This corresponds to the Order-Frequency Spectral Correlation (OFSC) in Figure 5.8(i) and Figure 5.8(ii). The Enhanced Envelope Spectrum (EES$_{\text{SCoh}}$) is shown in Figure 5.9(iii) for the OFSCoh in Figure 5.9(i) and the EES is shown in Figure 5.9(iv) for the OFSCoh in Figure 5.9(ii).

is possible to attenuate the historical information from different interfering sources, which results in the novel damage information to be enhanced. The implication is that non-diagnostic frequency bands that dominate the signal and impede the fault diagnosis process, are automatically attenuated based on a statistical criterion.

5.4 Experimental gearbox dataset

5.4.1 Overview of dataset

The experimental gearbox dataset, discussed in detail in Section 2.3.1, is used to validate the proposed methodology in this section. In the training phase of the methodology, 30 measurements are used to calculate the parameters of the model in Section 5.2.2, while 100 measurements, equally spaced over the operational life of the gear, are used in the testing or application phase of the methodology.

The axial component of a tri-axial accelerometer located on the bearing housing of the test gearbox is used in this investigation. The spectrogram of the vibration signal of the gearbox with localised gear damage is presented in Figure 5.10(i) where the vertical lines indicate

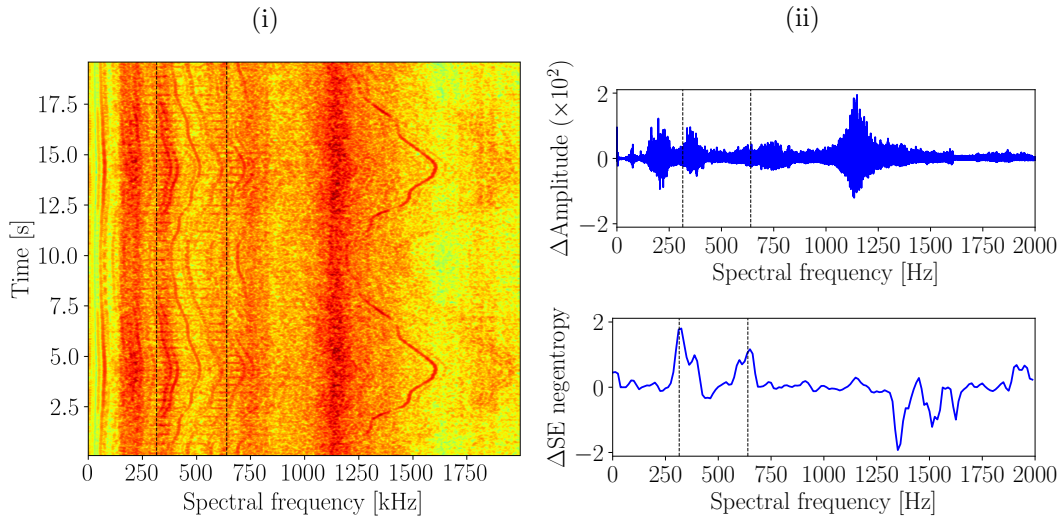


Figure 5.10: The spectrogram of a gearbox with localised damage under variable operating conditions is presented in Figure 5.10(i). The benefits for using the statistics of narrowband signals as opposed to the amplitudes of the frequency components for detecting frequency bands with novel information are illustrated in Figure 5.10(ii). The difference in the spectrum, between the damaged gearbox signal used in Figure 5.10(i) and the averaged spectrum of a healthy gearbox is presented in Figure 5.10(ii) and denoted as ΔAmplitude. The difference in the squared envelope infogram, between the damaged gearbox signal used in Figure 5.10(i) and the squared envelope infogram of a healthy gearbox is presented in Figure 5.10(ii) and denoted as ΔSE infogram. The squared envelope infogram is calculated for a specific window length to make the illustration easier.

the approximate spectral frequencies where the gear damage can be observed. The time varying frequency components, attributed to the gear mesh interactions, are proportional to the rotational speed of the input shaft presented in Figure 2.8, and clearly seen throughout the spectrogram of the signal. The time invariant frequency components are also clearly seen.

The benefit of using features of narrowband signals as opposed to spectral differencing is illustrated in Figure 5.10(ii), with the vertical lines being used to make the comparison easier between the two plots in Figure 5.10(ii) and the spectrogram in Figure 5.10(i). It is evident that the largest value for ΔSE infogram between the SE infogram of the damaged gearbox signal and the averaged SE infogram of the healthy gearbox signals occurs in the frequency bands where the damage is observed in Figure 5.10(i) i.e. they are highlighted with the vertical lines. This is in contrast to the results of Δ Amplitude; the amplitude difference is very sensitive to the operating conditions and the time-invariant frequency content seen at approximately 1100Hz. This supports using features in the methodology as opposed to differences in spectra for designing the novel filter.

The methodology is further motivated by:

- The damage results in significant changes in the SE infogram in two distinct frequency

bands in Figure 5.10(ii) as indicated by the vertical lines. This emphasises that more than one frequency band can potentially be used for demodulation for example. This is contrast to the methods in literature where only a single frequency band is selected for further analyses.

- The ΔSE infogram in Figure 5.10(ii) can be difficult to interpret manually, especially if many window lengths are investigated for the STFT decomposition, which is why an automatic statistical approach to incorporate the information of the different frequency bands into a single signal which can be used for further analyses, is required.

5.4.2 Results

The proposed methodology is implemented as in the numerical gearbox dataset. The SE infogram is used as features for consistency, whereafter a Gaussian model is used to describe the model parameters of the healthy dataset. The healthy dataset consists of 30 measurements. The novel information filter is designed using the proposed procedure and compared to the original signal using different signal processing techniques for different measurements.

5.4.2.1 Results with conventional methods

A popular signal analysis technique for gear diagnosis is the synchronous average of the vibration signal [13, 19]. The synchronous average is adversely affected by the modulation induced by varying operating conditions, but the load and speed variations are non-synchronous with respect to the shaft rotations and their influence are attenuated by the synchronous averaging process [19].

The synchronous averages for an original vibration signal and the corresponding novel vibration signal are presented for two measurements in Figure 5.11. Note that a total of 100 measurements are used between the two gear conditions shown in Figure 2.7, e.g. Measurement 75 means that 75% of the experimental time has elapsed. The gear damage was less developed for the measurement used in Figures 5.11(i) and 5.11(ii) compared to the measurement used in Figures 5.11(iii) and 5.11(iv) which resulted in vibration signals with larger amplitudes in the latter results for the novel signal. This is not evident in the original signal. Even though the synchronous averages of the raw vibration signals in Figure 5.11(i) and Figure 5.11(iii) do not clearly highlight the fact that only localised gear damage at approximately 135 degrees is present, the results are improved when calculating the synchronous average of

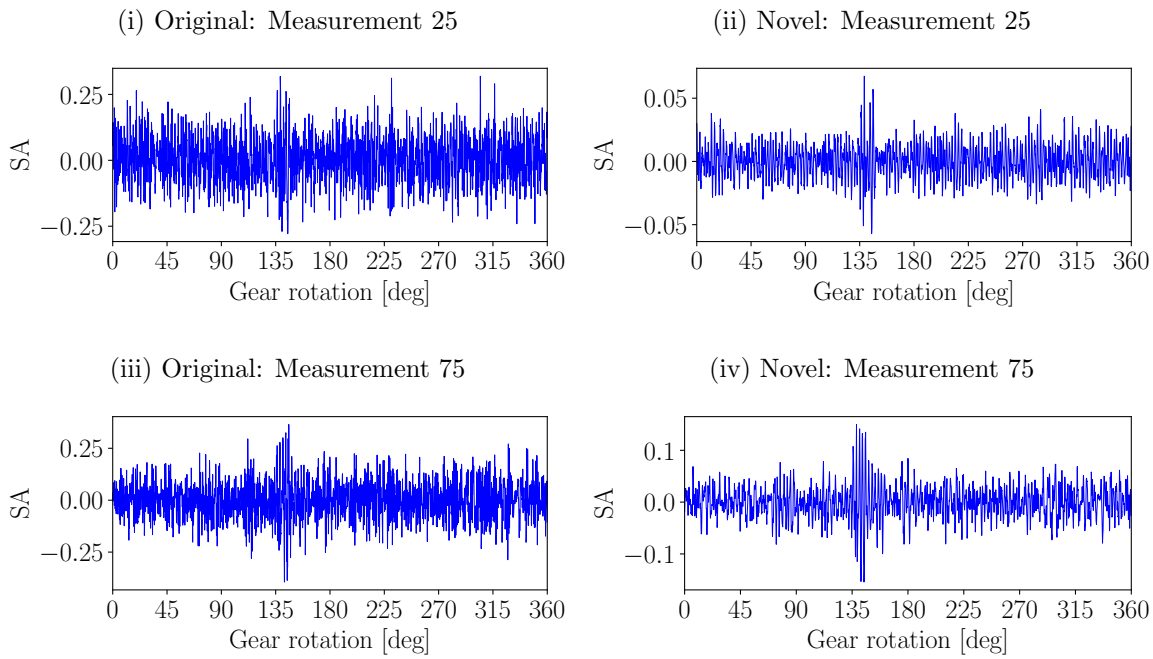


Figure 5.11: The synchronous averages of two measurements of the damaged gear dataset are presented.

the novel vibration signals shown in Figure 5.11(ii) and Figure 5.11(iv). In the novel vibration results, the damage component is more prominent with fewer spurious impulses that impede the condition inference process.

The squared envelope of the vibration signal indicates the instantaneous power of the vibration signal over time and can highlight potential impulses induced by damage. The localised gear damage results in periodical, localised changes in the power of the vibration signal and therefore the synchronous average of the squared envelope is investigated as well. The synchronous averages of the square envelope corresponding to the signals in Figure 5.11 are presented in Figure 5.12. It is evident from the results that the synchronous average of the instantaneous power highlights the localised damage better for the original and the novel cases as opposed to the synchronous average of the vibration signals. However, a significant improvement is seen in the novel vibration signals' results in Figure 5.12(ii) and Figure 5.12(iv) when compared to the corresponding original vibration signals' results in Figure 5.12(i) and Figure 5.12(iii). This highlights that the proposed pre-processing methodology is capable of highlighting the diagnostic information better and it also leads to results that are easier to analyse.

The SES is also investigated and presented in Figure 5.13 for the different signals being considered. Even though the SES is useful for characterising pure second-order cyclosta-

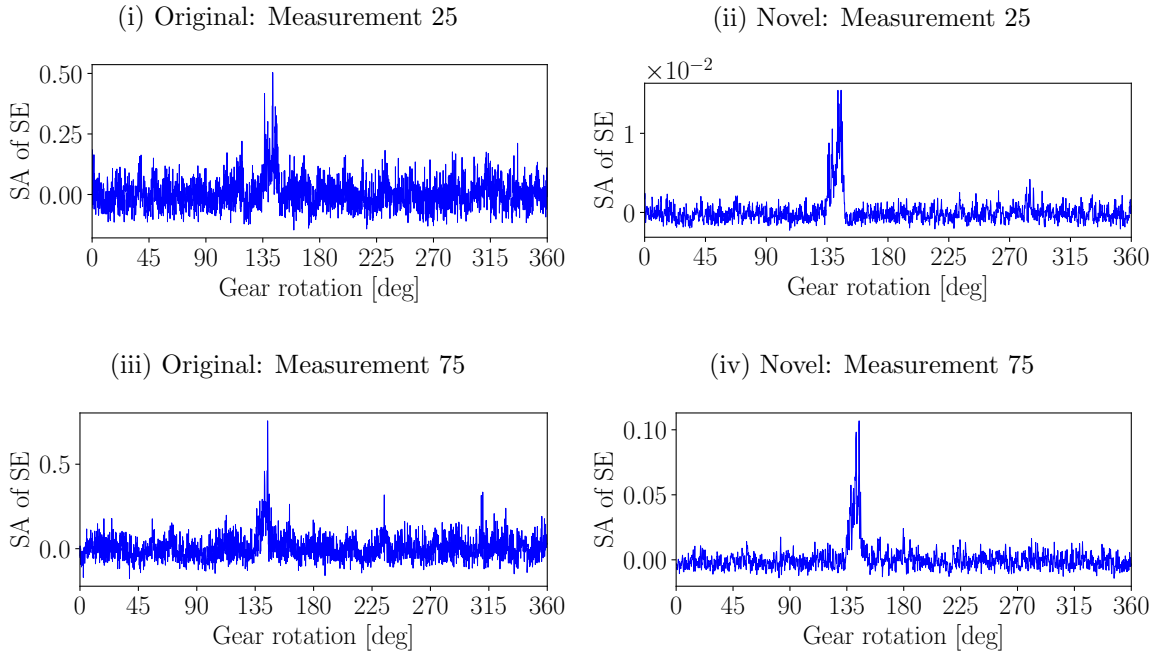


Figure 5.12: The synchronous averages of the squared envelope of two measurements of the damaged gear dataset are presented.

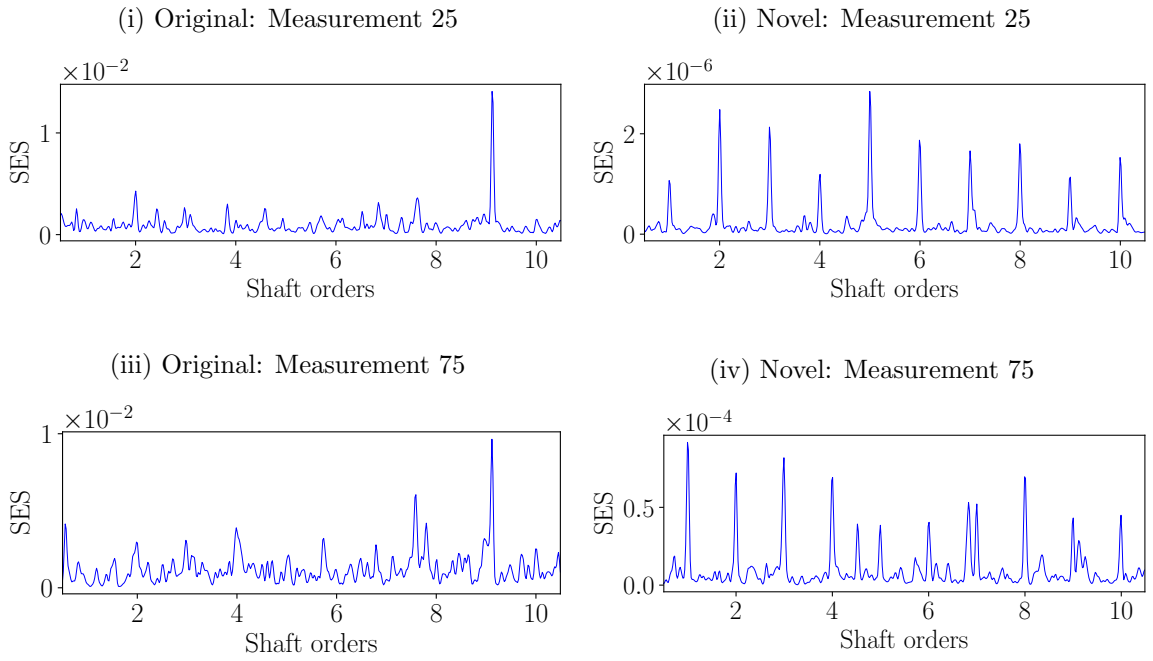


Figure 5.13: The Squared Envelope Spectrum (SES) is presented for the damaged gearbox data, with different y-axis scales being used. The SES presented in this figure corresponds to the results in Figure 5.12.

tionary signals, it is useful to quantify periodic changes in the instantaneous power of the vibration signal induced by the localised gear damage. The damaged gear is located on the shaft with the zebra tape shaft encoder and therefore the damage occurs at one shaft order and its harmonics. It is evident from the results in Figure 5.13(i) and Figure 5.13(iii) that the modulation induced by the localised gear damage is not prevalent in the original vibration signal. The signal component at 9.12 shaft orders is attributed to the shaft at the alternator being slightly unbalanced. This is in contrast to the SES of the novel vibration signals in Figure 5.13(ii) and Figure 5.13(iv). The signal components at the shaft harmonics due to the gear damage are clearly visible and dominate the SES. This further validates that the novel vibration signal highlights the diagnostic information in the vibration signal, with the periodicity of the damage clearly being detected.

It is reassuring to show that the results presented in Figures 5.11, 5.12 and 5.13 were not obtained by cherry picking and therefore the results are presented for all measurements used in this investigation. The results for the 100 damaged measurements are presented in Figure 5.14. The results in Figure 5.14 are obtained by adding a vertical offset to each measurement, with the offset increasing linearly with measurement number. It is clear from the results of the synchronous averages of the squared envelope i.e. SA of SE and the SES, that the novel signal is more sensitive to damage. The damage component and the change in of the damage component in the final stages of the gears' life are very prominent when compared to the original signal.

5.4.2.2 Quantification of performance

It is difficult to compare the performance of the methodology to the raw signal from Figure 5.14. Therefore, it is advantageous to use quantitative measures as basis of comparison and therefore three performance metrics are used on this dataset. The first metric is the kurtosis of the vibration signal $\boldsymbol{x} \in \mathbb{R}^{N_x}$

$$\text{kurtosis}(\boldsymbol{x}) = \frac{\frac{1}{N_x}\sum_{i=1}^{N_x}\left(x[i]-\bar{x}\right)^4}{\left(\frac{1}{N_x}\sum_{i=1}^{N_x}\left(x[i]-\bar{x}\right)^2\right)^2} - 3, \tag{5.22}$$

which is popular for incipient fault detection. In Equation (5.22), the mean of the vibration data $\boldsymbol{x}$ is denoted by $\bar{x}$. The kurtosis, as defined in Equation (5.22), is zero for Gaussian data and increases as the tails of the distribution increase e.g. as the impulsiveness of the dataset increase.

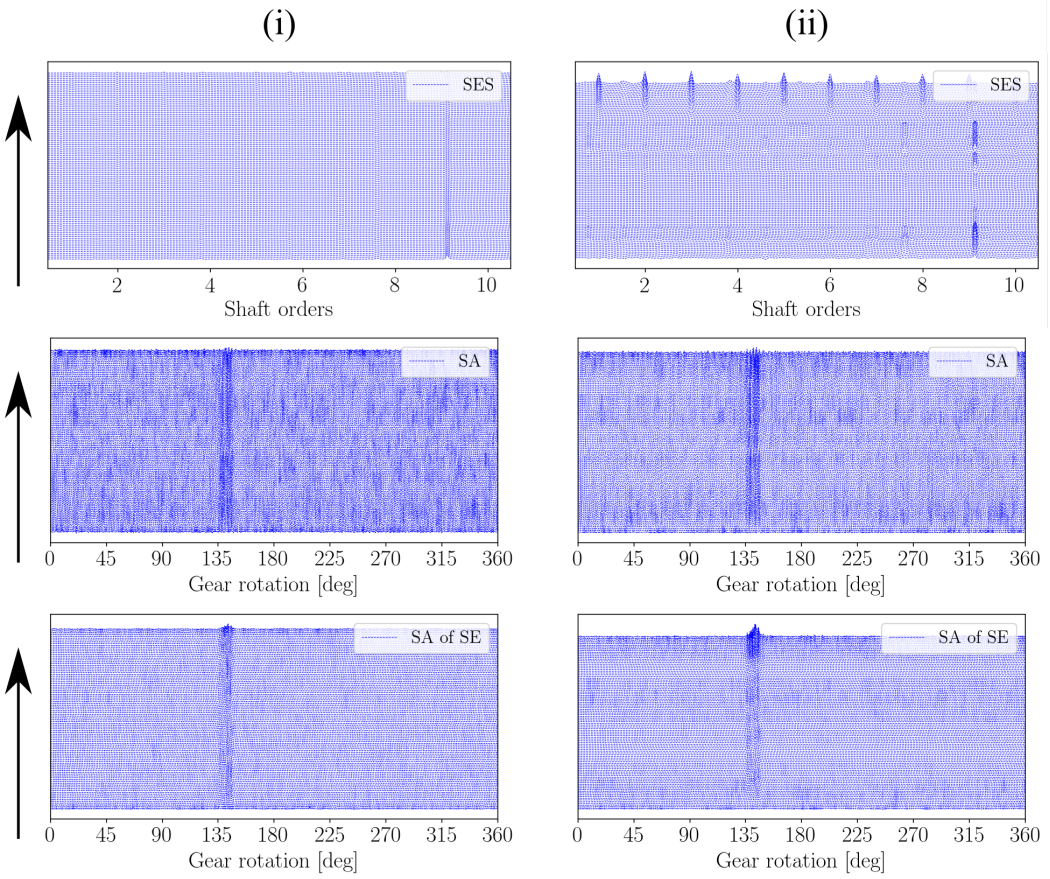


Figure 5.14: The waterfall plots of the smoothed Synchronous Average (SA) of the vibration signals, the smoothed Synchronous Average of the Squared Envelope (SA of SE) of the vibration signals and the smoothed Squared Envelope Spectrum (SES) are presented for the 100 damaged measurements. The 100 measurements were spaced throughout the life of the damaged gear i.e. from Figure 2.7(i) to Figure 2.7(ii). The smoothing, calculated from a second averaging process, is permissible for data acquired over a long time and is used to remove spurious noise which makes the results difficult to interrogate. The arrow indicates an increase in measurement number. A critical study of the statistics of the raw results is presented in Figure 5.15.

The second metric under consideration is the Quality Measure (QM) of the SES and given by

$$\mathrm{QM}_{\mathrm{SES}} = 10 \log \left(\frac{\mathrm{SES}_{damaged}(\boldsymbol{x})}{\mathrm{SES}_{noise}(\boldsymbol{x})} \right). \tag{5.23}$$

In Equation (5.23), $\mathrm{SES}_{damaged}(\boldsymbol{x})$ denotes the average of the damage components in the SES and $\mathrm{SES}_{noise}(\boldsymbol{x})$ is an estimate of the noise floor of the SES. Hence, it is desired to obtain a SES which maximises the QM for the SES i.e. $\mathrm{QM}_{\mathrm{SES}}$. The calculation procedure is illustrated in Figure A.16 in Appendix A.3.

The QM of the synchronous average

$$\mathrm{QM}_{\mathrm{SA}}(\boldsymbol{x}) = 10 \log \left(\frac{P^{\mathrm{SA}}_{damaged}(\boldsymbol{x})}{P^{\mathrm{SA}}_{healthy}(\boldsymbol{x})} \right), \tag{5.24}$$

is the last metric that is investigated. The QM of the synchronous average is a function of the ratio of the average power of the damaged portion $P^{\mathrm{SA}}_{damaged}(\boldsymbol{x})$ and the power of the healthy portion $P^{\mathrm{SA}}_{healthy}(\boldsymbol{x})$ of the synchronous average of the signal $\boldsymbol{x}$. A large $\mathrm{QM}_{\mathrm{SA}}(\boldsymbol{x})$ indicates that the synchronous averaged $\boldsymbol{x}$'s average power of the damaged portion is more significant than the average power of the healthy portion. The healthy portion and the damaged portions are calculated with the procedure given in A.3. The metric is used for the synchronous averaged vibration signals as well as the synchronously averaged square envelope.

The four metrics are presented in Figure 5.15 for the damaged measurements. It is evident from the various metrics that the novel signal is more sensitive to damage when compared to the original signal. However, the quality of the novel signal can vary over time, depending on the noise characteristics of the signal. If the long term data are effectively used i.e. by smoothing the results, it is evident from the results that the novel signal is significantly more sensitive to damage.

5.4.2.3 Additional analyses

The Instantaneous Power Spectrum (IPS) is also a powerful cyclostationary analysis tool which has been successful for gearbox diagnostics under varying operating conditions [15, 99,

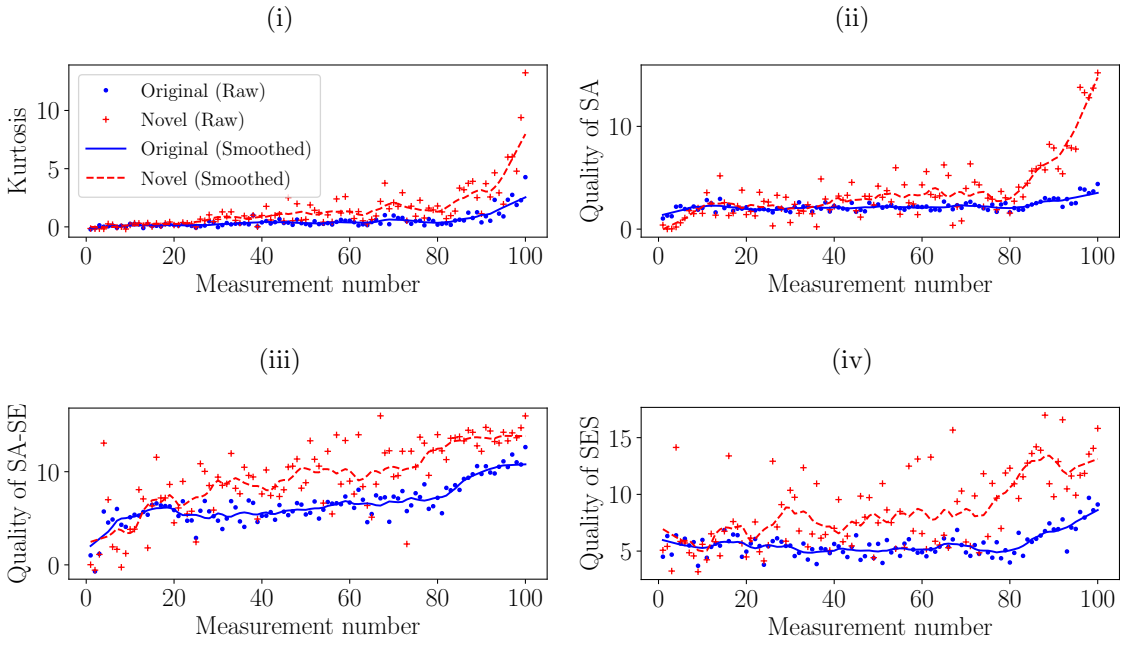


Figure 5.15: Different metrics of the healthy and the damaged vibration signals are calculated for all measurement numbers and with a smoothed estimate of the metrics being presented as well. The raw metrics, denoted by (*Raw*) in the legend, are calculated by using Equation (5.22), (5.23) and (5.24) on each measurement, where the smoothed estimate is calculated with a moving average with length eight and an overlap of 6. The kurtosis, calculated with Equation (5.22), of the synchronously averaged signal is calculated.

100]. The spectrogram estimator of the time-frequency IPS, given by [98]

$$I_{\text{TF-IPS}}[n,m] = \frac{1}{f_s \cdot N_x \cdot \sum_{a=0}^{N_w-1} \upsilon[a]} \left| \sum_{k=0}^{N_w-1} \upsilon[k] \cdot x[k + n \cdot (N_w - N_o)] \cdot e^{-j2\pi km/N_w} \right|^2, \tag{5.25}$$

is inappropriate for varying speed conditions and therefore the Angle-Frequency IPS (AF-IPS), obtained from order tracking each frequency band of the time-frequency IPS separately [98]

$$I_{\text{AF-IPS}}[a,m] = \text{OT}_{n\to a}\left(I_{\text{TF-IPS}}[n,m]\right), \tag{5.26}$$

where $\text{OT}_{n\to a}(\cdot)$ indicates that the time index n is converted to an angle index a through order tracking [32, 37]. The AF-IPS is presented in Figure 5.16 for the original vibration signal of the damaged gearbox and its corresponding novel vibration signal. Two frequency bands are prominent in Figure 5.16(i), whereas only one frequency band is prominent in Figure 5.16(ii). This indicates that the frequency band centred at 400 Hz contains significant novel information, while the frequency band centred at 1100 Hz was excited in the reference condition as well and is therefore attenuated in the methodology.

Urbanek et al. [15] illustrated the efficiency of the synchronous averaged IPS to perform

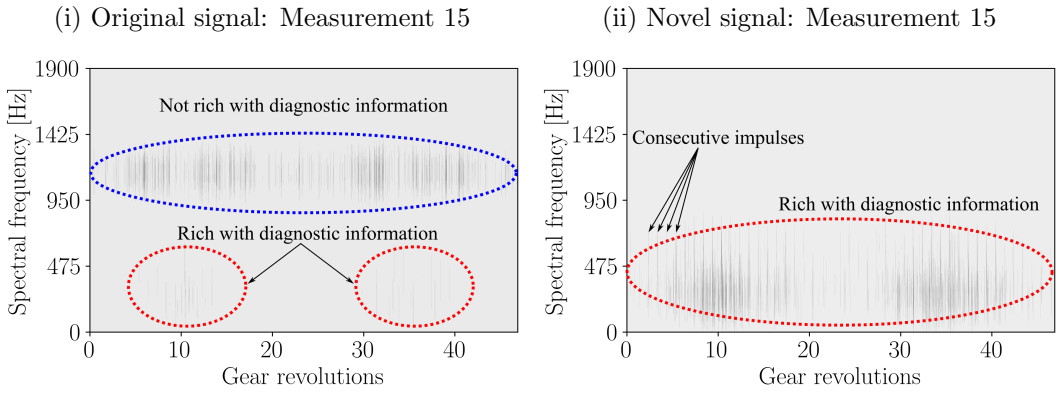


Figure 5.16: The Angle-Frequency Instantaneous Power Spectrum (AF-IPS) for the vibration signal of a damaged gearbox is presented. Figure 5.16(i) contains the AF-IAS of the raw vibration signal and Figure 5.16(ii) contains the AF-IAS for the corresponding novel signal. The two plots are not on the same scales to highlight the dominant components better.

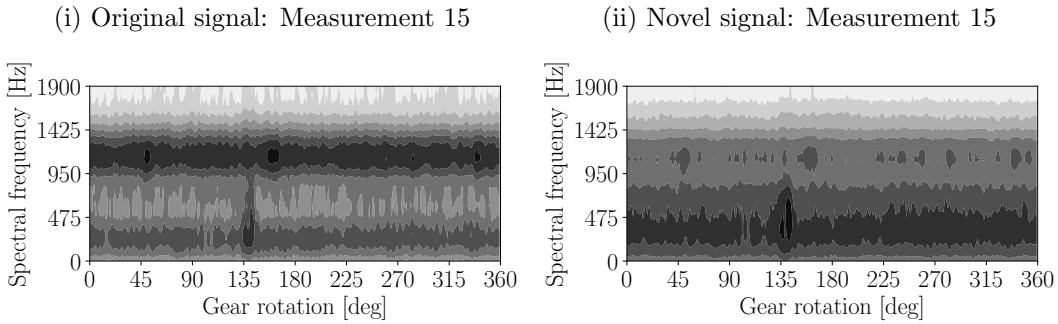


Figure 5.17: The synchronous average of the instantaneous power spectra shown in Figure 5.16(ii) are presented. The two plots are not on the same scale to highlight the dominant components better.

bearing diagnostics under varying operating conditions. This reduced the influence of varying operating conditions and non-synchronous components and therefore highlighted the repetitive diagnostic information. Hence, the synchronous average of the IPS, with a cycle period of one shaft order

$$\bar{I}_{\text{AF-IPS}}[a, m] = \frac{1}{N_{rot}} \sum_{k=0}^{N_{rot}-1} I_{\text{AF-IPS}}[a + N_s \cdot k, m], \tag{5.27}$$

is calculated for the signals, with N_{rot} indicating the number of revolutions completed by the shaft and N_s the number of samples per revolution. The synchronous averaged IPS presented in Figure 5.17 corresponds to the IPS presented in Figure 5.16. In Figure 5.17(i), the two excited frequency bands are clearly seen in the spectrum, with the localised gear damage at approximately 135 degrees seen in the lower frequency band. In the higher frequency band of Figure 5.17(i), spurious components are seen which can make it difficult to diagnose the

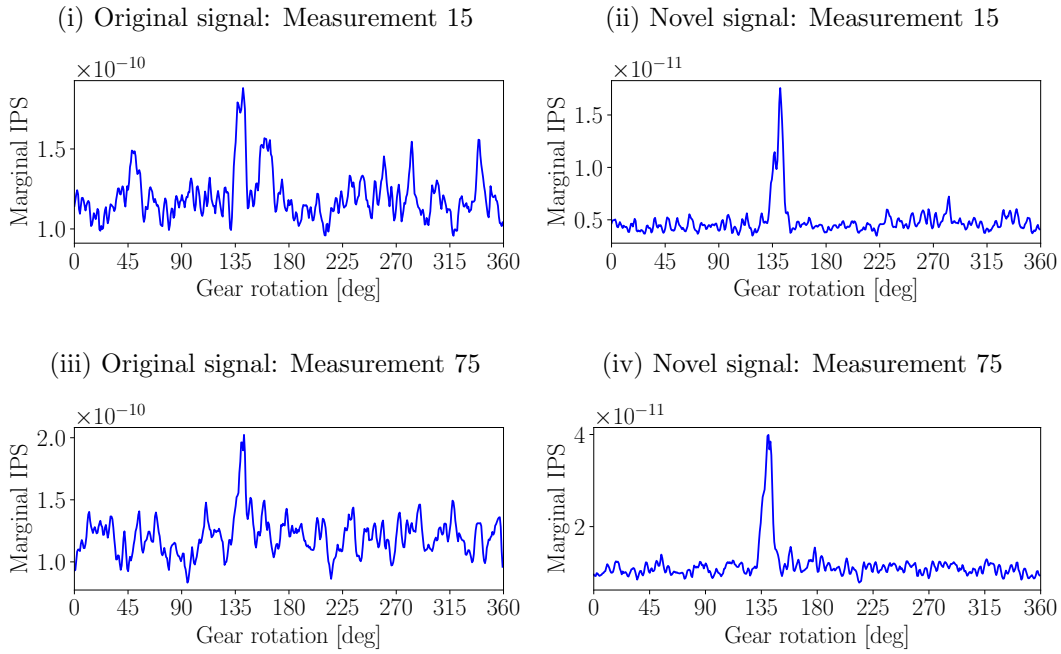


Figure 5.18: The marginal synchronous averages of the instantaneous power spectra shown in Figure 5.17 are presented.

machine. In Figure 5.17(ii), only one frequency band is prominent with the gear damage seen at approximately 135 degrees. The implications of the additional frequency band are not prevalent from the results in Figure 5.17 and therefore another representation is investigated.

The marginal synchronous average of the IPS i.e. by marginalising over the N_f spectral frequency bands

$$\bar{i}_{\text{AF-IPS}}[a] = \frac{1}{N_{rot} \cdot N_f} \sum_{m=0}^{N_f-1} \sum_{k=0}^{N_{rot}-1} I_{\text{AF-IPS}}[a + N_s \cdot k, m], \tag{5.28}$$

is used to highlight the localised gear damage and it makes it easier to illustrate the benefits of using the novel signal as opposed to the original signal. The marginal synchronous average of the IPS is presented in Figure 5.18 for the corresponding results presented in Figure 5.16(i) and Figure 5.17(i). Clearly, the presence of the second frequency band that does not contain diagnostic information in Figure 5.17(i) adversely influences the ability to detect the localised gear damage in Figures 5.18(i) and 5.18(iii). In the marginal IPS of the novel signal, presented in Figures 5.18(ii) and 5.18(iv), the damage is very prominent and clear, with the non-diagnostic information attenuated.

Hence, it is evident from the results that there are considerable benefits in incorporating historical information in the diagnosis procedure. Hence, it is possible to enhance the ability of existing signal processing techniques to detect damage, characterise damage and trend

damage for gear diagnostics under varying operating conditions.

5.5 Conclusion and recommendations

In this chapter, a methodology is presented to enhance the novel information in the vibration signal by attenuating frequency bands which do not have new information compared to the historical data. This is performed by using narrowband feature extraction techniques such as the squared envelope infogram or the kurtogram to extract statistics from narrowband signals, whereafter the statistics of a healthy machine are learned with a model. The model is subsequently used to identify frequency bands with anomalous behaviour, which are enhanced.

The methodology is applied to numerical gearbox data as well as experimental gearbox data where it is illustrated that the methodology can enhance the diagnostic information in the vibration data which subsequently leads to more effective fault diagnosis. It is specifically shown that by using the methodology as a pre-processing technique, it is possible to enhance the performance of signal analysis techniques e.g. cyclostationary analysis for fault diagnosis. Hence, it is sensible to combine narrowband feature extraction techniques with the available historical data to enhance the diagnostic information in the vibration data. This improves the effectiveness of conventional signal analysis techniques in the fault diagnosis field.

Lastly, the methodology is presented in its simplest form without optimising any of the sub-processes. However, there are many potential areas that can be investigated further, such as selecting different features, using non-Gaussian models, investigating different thresholds and different filter designs.

Chapter 6

Open set recognition methodology

6.1 Introduction

A central part of the condition-based maintenance is determining or inferring the condition of the machine from the condition monitoring data [1, 3]. However, performing this task reliably can be time consuming and difficult, especially when performed by a non-expert [3, 61]. Therefore, classification algorithms utilising machine learning [65, 77, 83] and even more recently deep learning techniques [133, 137, 139] are used to automatically infer the condition of the machine i.e. assigning a condition label to the machine from the condition monitoring data.

The machine learning models are trained with the available historical fault data to perform predictions in the application (or testing) phase of the condition monitoring process. However, the historical fault data can be difficult to acquire, because the machine operates mostly in a healthy condition (i.e. there is a class imbalance between the available healthy data and the historical fault data [171]), new machines do not have historical fault data available, historical fault data from all of the damage modes may not be available for the machine when optimising the model, and acquiring historical fault data from all of the damage modes is time consuming and expensive.

Hence, it is unrealistic to assume that historical data, representing a population of all possible damage modes, would be available when training the condition recognition model for most machines. Hence, the machine condition monitoring problem needs to be addressed as an open set recognition problem [143] and not as a closed set recognition problem, the latter being the conventional method for performing classification in condition monitoring applications. In Figure 1.2, different data availability levels are defined, with the OSR problem corresponding

to Data availability: Level 2 and the CSR problem corresponding to Data availability: Level 3.

In this chapter, an open set recognition methodology, using a discrepancy analysis procedure as a pre-processing step, is proposed. This is performed to not only illustrate the importance of open set recognition problems for the condition monitoring field, but also to illustrate that discrepancy analysis, conventionally used as a novelty detection approach, can be used to increase the predictive performance of the condition recognition system.

However, before presenting the methodology, it is firstly important to emphasise the differences between Closed Set Recognition (CSR) and Open Set Recognition (OSR) problems. This is performed in the next section.

6.2 Condition recognition for condition monitoring

In this section, the differences between CSR and OSR problems are highlighted for machine condition monitoring problems, with Figure 6.1(a) and Figure 6.1(b) containing the decision boundaries of the condition recognition problem for a machine with healthy historical data and historical fault data of two damage modes available. Discrete conditions are used in

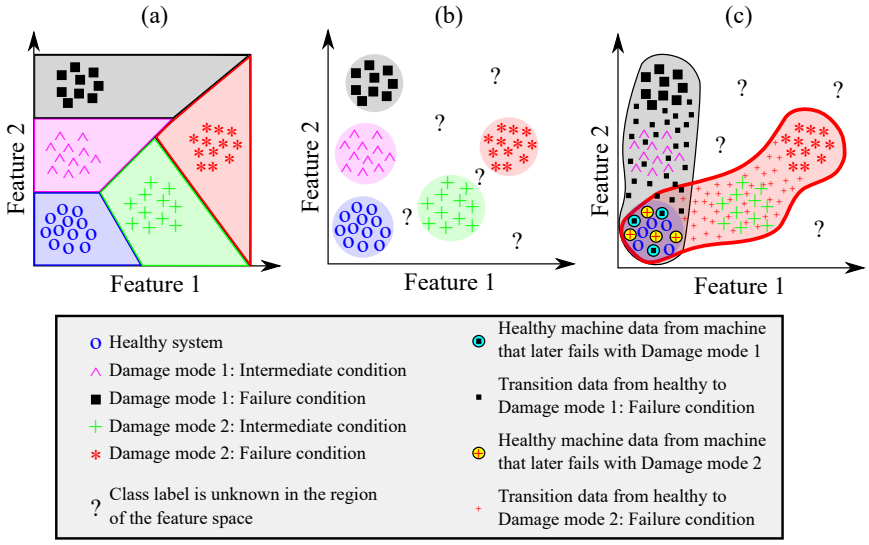


Figure 6.1: The differences between the decision boundaries of a Closed Set Recognition (CSR) framework and an Open Set Recognition (OSR) framework are shown in (a) and (b) respectively for artificial data. The proposed decision boundaries that are required for an ideal OSR framework for machine condition monitoring are shown in (c). In (c), the five classes used in (a) and (b) are reduced to three classes i.e. healthy, damage mode 1 and damage mode 2. The intermediate condition refers to a machine that is somewhere between a healthy and a failure state.

Figures 6.1(a) and 6.1(b), because this is the conventional method used in literature e.g.

[133, 172]. In the OSR framework (i.e. Figure 6.1(b)), predictions are only made in regions supported by historical data as opposed to the CSR (i.e. Figure 6.1(a)). In the CSR framework, class labels are assigned to the complete feature space which can lead to erroneous predictions when used for data from a new class or data not supported by the historical data, i.e. outliers.

However, a subtle but significant problem, arises when addressing the condition recognition procedure as open set recognition problem. Throughout the operational lifetime of a machine, the machine transitions from a healthy state to some damaged state e.g. a crack initialises and grows. However, this results in problems when using the OSR framework shown in Figure 6.1(b); only the discrete conditions can be labelled with a class label, with the transitions between conditions labelled as novelties when the conditions are well-separated in the feature space. Hence, it is important to learn the complete trajectory of the dataset, as the machine transitions from a healthy state to a damaged state. It is therefore necessary to exploit the complete dataset (i.e. as the machine transitions from a healthy state to a damaged state) to learn the transition path of the features so that the correct label can be assigned as the machine transitions between conditions as well. This approach results in the decision regions shown in Figure 6.1(c). It is also easier to assign class labels to the approach in Figure 6.1(c) as opposed to 6.1(b), because the damage initiation time and the intermediate states may not be known for the historical fault data and can be difficult to estimate without stopping and disassembling the machine.

Hence, it is not only important to use an open set recognition framework for condition monitoring applications, it is also necessary to use the full dataset (i.e. as the machine deteriorates) when training the models as opposed to labelling discrete states. Therefore, an open set recognition methodology is proposed for gear diagnostics under varying operating conditions in this chapter to obtain results similar to Figure 6.1(c). An overview of the proposed methodology is given in Section 6.3, whereafter an investigation is performed in Section 6.4 on synthetic data to illustrate the benefits of using an open set recognition methodology. Finally, the methodology is validated in Section 6.5 on experimental gearbox data.

6.3 Methodology

The proposed methodology is presented in Figure 6.2. Vibration data and rotational speed information of the machine component are converted to a processed discrepancy signal using

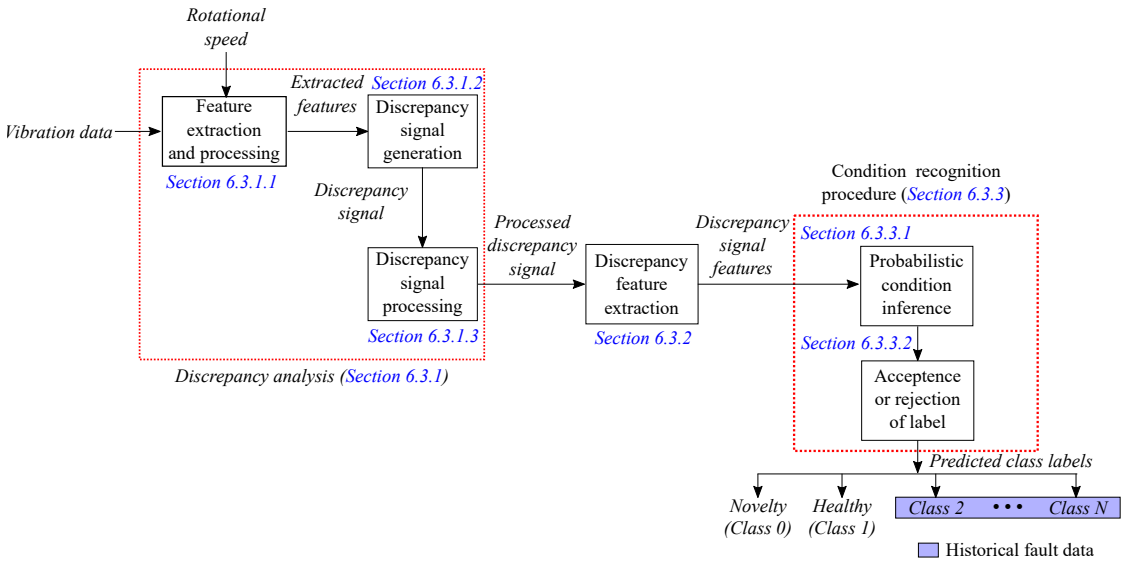


Figure 6.2: The proposed automatic novelty detection and condition inference methodology is presented, with the sections containing each part of the methodology given as well. In this figure, it is assumed that there are historical data of $N-1$ damage modes available, with the class labels used consistently throughout this chapter e.g. a class label of 0 will always indicate a novelty or a rejected class label.

the discrepancy analysis framework. The processed discrepancy signal is rich with diagnostic information and therefore discrepancy features are extracted from the processed signal. The extracted discrepancy signal features serve as an input to an OSR condition recognition system which is capable of performing novelty detection and inferring the condition of the machine under varying operating conditions. The rotational speed information can be acquired from tachometers or it can be estimated from the vibration data itself.

More information on each step in Figure 6.2 is given in the subsequent sections. The proposed methodology is a proof of concept and therefore different features and models can be used if desired.

6.3.1 Discrepancy analysis

Discrepancy analysis uses a model of the features, extracted from data that were acquired from a healthy machine, to assign discrepancy measures to the features of new data. This discrepancy measure is used to form a discrepancy signal which is analysed using techniques such as synchronous averaging to infer the condition of the gearbox [154–157]. A brief overview of the features and models used in this chapter are given in this section.

6.3.1.1 Feature extraction for discrepancy analysis

In discrepancy analysis, it is necessary for the features to be described in terms of time or shaft angle and therefore time-frequency analyses methods are employed [156]. Hence, the Continuous Wavelet Transform (CWT), which is popular for analysing non-stationary signals for rotating machine diagnostics [59, 65, 69, 70], is used with a Mexican hat basis function to obtain an angle-order distribution of the order tracked vibration signal, with the order tracking performed with the available rotational speed (or phase) information. The rotational speed can be estimated from measurement equipment [33] or estimated from the vibration signal [37] depending on the available equipment. The wavelet coefficients of the first three gear mesh components are extracted for different shaft angles, whereafter a windowing procedure is used to extract localised features from the wavelet coefficient signals. The windowing procedure is described in detail in Section 3.2.2. In this work, ninety rectangular windows per gear revolution with an overlap of 50% between coinciding windows are used on the wavelet coefficients of each gear mesh component. The features extracted from each window are given in Table 6.1. The ten statistical features in Table 6.1, applied to the wavelet coefficients of the three gear mesh components, result in a 30 dimensional feature space and are denoted by $\boldsymbol{b}_i \in \mathbb{R}^{30}$ for the features extracted from window i. Changes in the localised statistics of the wavelet coefficients, such as the localised impulsiveness for example, as opposed to changes in the values of the wavelet coefficients are detected by the healthy feature model. The benefit of this is illustrated in Figure 3.4, where the localised features can enhance the diagnostic information in the wavelet coefficients.

Table 6.1: The statistical features calculated for the windowed wavelet coefficients of each gear mesh frequency, with more details on the features found in Refs. [77, 122, 123].

1	Mean	2	Standard deviation
3	RMS	4	Maximum
5	Kurtosis	6	Skewness
7	Shape factor	8	Impulse factor
9	Crest factor	10	Clearance factor

6.3.1.2 Discrepancy signal generation

The discrepancy signal is generated by firstly modelling the features extracted from the healthy data, whereafter a discrepancy measure is assigned to the new features to construct a discrepancy signal. A Gaussian Mixture Model (GMM) is used to model the features

extracted with the process in Section 6.3.1.1, because it is able to represent any distribution to an arbitrary level of accuracy if a sufficient number of mixture components is used [114] and it has performed well for discrepancy analysis [155]. Heyns et al. [155] did not find an improvement by incorporating the operating conditions in the GMM for discrepancy analysis and therefore only the fault features are used to optimise the GMM in this section.

The GMM parameters are obtained from the Expectation Maximisation (EM) algorithm, which is very efficient in iteratively obtaining the maximum likelihood estimates of the model parameters [114]. The Negative Log-Likelihood (NLL) of a GMM with N_c mixture components [114]

$$\eta[l] = -\log\left(\sum_{j=1}^{N_c} \pi_{\boldsymbol{b},j}\mathcal{N}(\boldsymbol{b}_l|\boldsymbol{\mu}_{\boldsymbol{b},j}, \boldsymbol{\Sigma}_{\boldsymbol{b},j})\right), \tag{6.1}$$

is used as a discrepancy measure, where $\boldsymbol{b}_l$ denotes the features extracted from window l, $\pi_{\boldsymbol{b},j}$ is the mixture coefficient, $\boldsymbol{\mu}_{\boldsymbol{b},j}$ is the mean and $\boldsymbol{\Sigma}_{\boldsymbol{b},j}$ is the covariance of mixture component j of the GMM. The mixture component $\pi_{\boldsymbol{b},j}$ weights the contribution of the jth Gaussian model to the probability density function of the features and can be interpreted as the prior probability that component j is active in the GMM [114]. The mean $\boldsymbol{\mu}_{\boldsymbol{b},j}$, covariance $\boldsymbol{\Sigma}_{\boldsymbol{b},j}$ and mixture coefficient $\pi_{\boldsymbol{b},j}$ of each component j are the unknown parameters of the GMM and are estimated with the EM algorithm on the healthy features. The number of latent states or mixture components, denoted by N_c, is obtained by performing cross-validation. A multivariate Gaussian distribution over $\boldsymbol{b}$ is denoted by $\mathcal{N}(\boldsymbol{b}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in Equation (6.1), where its mean and covariance are denoted by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively.

6.3.1.3 Discrepancy signal post-processing

The synchronous average is a useful tool for analysing whether the discrepancy signal contains deterministic components attributed to gear fault impacts [155]. The synchronous average of the discrepancy signal of measurement j, denoted by $\boldsymbol{m}_\eta^{(j)} \in \mathbb{R}^{N_s}$, is calculated with

$$m_\eta^{(j)}[k] = \frac{1}{N_r}\sum_{i=1}^{N_r} \eta\left[k + (i-1)N_s\right], \tag{6.2}$$

where $1 \leq k \leq N_s$, the discrepancy signal is denoted by η, the number of rotations that was completed by the shaft is denoted by N_r and the number of samples per shaft revolution is an integer and denoted by N_s. The synchronous average can also be useful for attenuating the amplitude modulation due to non-cyclic stationary loads [19], which means that the

synchronous averaged discrepancy signal can be even more robust to non-cyclic stationary operating conditions.

In the next section, the features extracted from the processed discrepancy signal and subsequently used for the condition recognition procedure, are presented and discussed.

6.3.2 Discrepancy feature extraction

Two features are extracted from the synchronous average of the discrepancy signal, calculated with Equation (6.2), to infer the condition of the gear. The first feature

$$r_{j1} = \frac{\max(\boldsymbol{m}_\eta^{(j)}) - \min(\boldsymbol{m}_\eta^{(j)})}{\sqrt{\frac{1}{N_s}\left(\sum_{i=1}^{N_s} m_\eta^{(j)}[i]^2\right)}}, \tag{6.3}$$

makes it possible to discern between localised changes and uniform changes in the synchronous average of the discrepancy signal. This is an important generalisation of the previous work in discrepancy analysis which was focused on localised faults. The second feature

$$r_{j2} = \max(\boldsymbol{m}_\eta^{(j)}) - \min(\boldsymbol{m}_\eta^{(j)}), \tag{6.4}$$

is the range of the synchronous average of the discrepancy signal and is sensitive to localised changes in the discrepancy signal. The features extracted from the synchronous average of measurement j is denoted by a two-dimensional vector $\boldsymbol{r}_j = [r_{j1},\ r_{j2}]^T$. The two features, extracted from Equation (6.3) and Equation (6.4), have a low dimensionality which means that neither feature subset selection nor dimensionality reduction techniques are required to circumvent the curse of dimensionality or to visualise the features. The low-dimensionality of the feature space helps with understanding the performance of different classifiers for machine condition monitoring which subsequently supports the conclusions that are drawn from the results in this chapter. However, more optimised features can certainly be investigated to obtain classifiers that perform better.

The DC offset or mean value of the synchronous average of a healthy machine can be dependent on the discrepancy model which can adversely affect the features extracted from the discrepancy signal. Hence, the synchronous average of the first measurement of a dataset is subtracted from the entire dataset to obtain synchronous averages that are roughly model independent, with the new synchronous averages used in the feature extraction process. This is allowed, because it is assumed that the machine starts from an approximately healthy

condition and deteriorates over time.

6.3.3 Condition recognition procedure

In this section, the condition recognition procedure that uses the discrepancy signal features in Section 6.3.2 to infer the condition of the gear is presented. In the condition recognition procedure, the objective is to recognise that the data is part of a specific class and to classify or label the data accordingly. However, if the condition cannot be recognised, it means that either a new damage mode is present or an outlier is detected and the data is labelled as a novelty. This is implemented with a OSR model which consists of two steps; probabilistic condition inference is performed to predict the class label of the machine, whereafter the class label is either accepted or rejected.

6.3.3.1 Probabilistic condition inference

It is desired to cast the condition inference problem in a OSR framework as opposed to the conventional CSR framework. Scheirer et al. [143] formalised the OSR problem in their work and contrasted it to the CSR framework used in many applications. Casimir et al. [127] investigated a methodology using a nearest neighbour classifier with two rejection options for induction motor diagnostics, while Lazzaretti [173] investigated an OSR methodology for power distribution events. The work in Refs. [143–145, 173] used support vector machines to address the OSR problem, where Lazzaretti et al. [173] found a support vector data description model performs the best of the considered OSR models for power distribution event detection.

In the machine condition monitoring field, the machine starts from an approximately healthy condition and deteriorates to a specific damage mode until the machine is stopped or the component fails. Instead of searching for the crack initiation time in the historical data and discretizing the data into different conditions to create different damage classes which can result in classification problems as discussed previously, it is desired to learn the trajectory of the data by using the complete dataset for the OSR model optimisation problem. If the trajectory of the data is learned, it can be used to infer the condition of the machine at arbitrary damage states as seen in Figure 6.1(c). This is essential for applying OSR in the condition monitoring field.

The problem with learning the trajectory of the data is the resulting overlap between the data of the healthy class and the data from the damaged modes as seen in Figure 6.1(c).

This implicitly violates the objectives in the support vector optimisation task i.e. finding a hyperplane that maximises the distance between two classes, and due to the fact that the trajectory needs to be learned, it is concluded that support vector machines are not ideal for this task. GMMs have been successfully used for fault classification [126] and can potentially learn the trajectory of the data through the feature space. However, density-based methods such as GMMs suffer from the curse of dimensionality which adversely influence their performance in OSR problems [173]. The discrepancy features' dimensionality in condition monitoring is significantly lower than the dimensionality of most OSR applications and alleviates the problems of GMMs in OSR problems. Hence, GMMs are used with Bayes' rule for the OSR task in this work.

The condition is inferred from the discrepancy signal features by selecting the class with the highest posterior probability i.e. $\arg\max_i \left(P(C_i|\boldsymbol{r})\right)$, with the posterior probability of class C_i being obtained with Bayes' rule

$$P(C_i|\boldsymbol{r}) = \frac{p(\boldsymbol{r}|C_i)P(C_i)}{p(\boldsymbol{r})}. \tag{6.5}$$

The likelihood function and prior probability of class i are denoted by $p(\boldsymbol{r}|C_i)$ and $P(C_i)$ respectively, with the unconditional density over the features denoted by $p(\boldsymbol{r})$. A unique GMM likelihood function is used for each class, with the GMM associated with class i denoted by

$$p(\boldsymbol{r}|C_i) = \sum_{j=1}^{N_{c,r}} \pi_{\boldsymbol{r},j}^{(i)} \mathcal{N}\left(\boldsymbol{r}|\boldsymbol{\mu}_{\boldsymbol{r},j}^{(i)}, \boldsymbol{\Sigma}_{\boldsymbol{r},j}^{(i)}\right). \tag{6.6}$$

In Equation (6.6), the parameters of the GMM are estimated with the EM algorithm separately for each class i, with $\pi_{\boldsymbol{r},j}^{(i)}$, $\boldsymbol{\mu}_{\boldsymbol{r},j}^{(i)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{r},j}^{(i)}$ denoting the parameters of the jth component of class i. The appropriate number of mixture components $N_{c,r}$ in the GMM for the condition recognition procedure is estimated by cross-validation. The complete dataset, i.e. from a healthy machine to a damaged machine, is used in this methodology to optimise the GMM with each damage mode, with only the healthy data being used to optimise the GMM of the healthy machine.

In Refs. [117, 118], the class with the largest likelihood for the dataset is the inferred class i.e. $\arg\max_i p(\boldsymbol{r}|C_i)$, which is the same as $\arg\max_i P(C_i|\boldsymbol{r})$ when each class has the same prior probability. The same prior probability cannot be assigned to each class in this framework, because there is a large overlap in the healthy data space when using the complete datasets of

the damage modes for trajectory learning, which can lead to significant classification errors in the healthy data region. This problem can be resolved by giving the model that correspond to the healthy machine features preference by ensuring that its prior probability is larger than the other models i.e. $P(C_{\text{healthy}}) > P(C_i)$ for all $i \neq$ healthy. This comes from the expectation that the machine is healthy, unless there is strong evidence that the machine is in a damaged condition. In this work $P(C_i) = c$ for all $i \neq$ healthy and $P(C_{\text{healthy}}) = 2 \cdot c$, with the unknown constant c obtained by ensuring the prior is a valid probability distribution i.e. by solving $\sum_i P(C_i) = 1$.

The benefit of using Bayes' rule with the GMM is that it is efficient to incorporate historical data from a new damage mode into this framework. This is performed by optimising a GMM on the newly acquired data, whereafter the prior probabilities of all the models have to be recalculated with the prior probability of the new class i set to $P(C_i) = c$ as well. Thereafter, Equation (6.5) can be used to predict the class label.

6.3.3.2 Acceptance or rejection of inferred class label

The posterior prediction obtained with Bayes' rule in Equation (6.5) is dependent on the likelihood value of the other classes due to the unconditional density, and it does not account for other classes that can be present in the OSR problem [142]. Hence, a Decision Rule (DR) is used on the likelihood function of the model or class i to determine whether the model is capable of predicting the class label $s_i \geq 0$ or not $s_i < 0$, with

$$s_i = \log p(\boldsymbol{r}|C_i) - \alpha_i^{\text{DR}}. \tag{6.7}$$

The threshold and model fitness metric associated with the log-likelihood of class i is denoted by α_i^{DR} and s_i, respectively. If $s_i < 0$ for all i, it means that none of the models are capable of inferring the correct class and therefore the class label predicted in Section 6.3.3.1 is rejected and a novelty is detected. The posterior probabilities $P(C_i|\boldsymbol{r})$ corresponding to the detected novelties are invalid and are not used further in the analysis. If a novelty is detected by the condition recognition system, it does not necessarily indicate that a new damage mode is present, but the data should rather be investigated further by an expert in the field.

There are many criteria to select the threshold of class i, which will depend on the application e.g. a tighter threshold can be used for critical machinery. In this work, α_i^{DR} is selected so that 1% of the validation set is labelled as a novelty.

6.3.4 Final remarks on the methodology

If discrete conditions are used to define damage modes, performance metrics such as the class averaged accuracy and Youden's index can be used to quantify the performance of the OSR model [145]. However, in the methodology, the complete dataset is used for training the condition recognition system which poses some difficulties in quantifying the performance of the condition recognition system i.e. the damaged class contains healthy data and the condition varies significantly throughout the dataset. Hence, the following procedure is used for evaluating the performance of the proposed methodology:

- The measurement number at which the damage is detected and the characteristics of the posterior probabilities of the classes are investigated.
- The characteristics of the decision boundaries of different classifiers are investigated and compared, which is possible due to the two-dimensional discrepancy signal feature space. This makes it possible to investigate the implications of different decision boundaries for detecting a deteriorating gear condition which helps to evaluate the performance of the associated classification models.
- The OSR abilities of the methodology is investigated by visualising the two-dimensional discrepancy signal feature space. This makes it possible to evaluate the possibility of the OSR to generalise for other damage modes.

6.4 Synthetic data investigation

An investigation is performed in this section on the discrepancy feature extraction and condition recognition parts of the methodology presented in Figure 6.2, where it is assumed that the synchronous average of the different damage modes are already available. The purpose of this investigation is

- to investigate the characteristics of the discrepancy features,
- to highlight the importance of the decision boundaries for machine condition monitoring,
- to highlight that it is beneficial to use the complete dataset as the gear transitions from a healthy to a damaged state in OSR condition monitoring problems,

- and to compare the performance of the OSR condition recognition system to other conventional CSR classifiers on the same data.

6.4.1 Synthetic dataset

For the synthetic dataset, synchronous averages of six classes are generated with the following equations

$$m1_{\eta}^{(j)}[l] = 0.1 \cdot \epsilon[l], \tag{6.8}$$

$$m2_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \delta(l - \lfloor l_{med} \rfloor) + 0.1 \cdot \epsilon[l], \tag{6.9}$$

$$m3_{\eta}^{(j)}[l] = \kappa \cdot \beta + 0.1 \cdot \epsilon[l], \tag{6.10}$$

$$m4_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \left(\delta(l - \lfloor l_{med} \rfloor) + \sum_{j=1}^{3} \sum_{k=1}^{2} (0.8 - 0.2 \cdot j) \cdot \delta\left(l - \lfloor l_{med} \rfloor - j \cdot (-1)^k\right) \right) + 0.1 \cdot \epsilon[l], \tag{6.11}$$

$$m5_{\eta}^{(j)}[l] = \kappa \cdot \beta \cdot \sum_{j=0}^{\lfloor N_s/3 \rfloor} \delta(j - l) + 0.1 \cdot \epsilon[l], \tag{6.12}$$

$$m6_{\eta}^{(j)}[l] = \kappa \cdot \beta + 0.1 \cdot \sin\left(10\pi \cdot l/N_s\right) + 0.1 \cdot \epsilon[l], \tag{6.13}$$

where for example, $m1$ denotes the synchronous average of Class 1 in Equation (6.8). In Equations (6.8)-(6.13), j denotes the measurement number, l denotes the position on the gear with $1 \leq l \leq N_s$ and $l \in \mathbb{Z}$. The median of l is denoted by l_{med}. The condition of the component is indicated with a scalar factor $0 \leq \kappa \leq 1$, where 0 indicates a healthy condition and 1 indicates the component is in its failure state according to the respective damage mode or class. The magnitude of the damaged component in its failure state, denoted β, is sampled from a uniform distribution $\mathcal{U}$ with a domain of $[0.75,\ 0.90]$, i.e. $\beta \sim \mathcal{U}[0.75,\ 0.90]$. Zero mean Gaussian noise with a standard deviation of 1 is denoted by $\epsilon[l]$. The function $\delta(x) = 1$ for $x = 0$ and $\delta(x) = 0$ if $x \neq 0$ and the number of samples in the synchronous average is $N_s = 100$. The mean and standard deviation of the six synthetic synchronous average profiles are presented in Figure 6.3. Class 1 represents a healthy system, Class 2 and Class 4 represent localised damage such as a damaged gear tooth with slightly different characteristics [155], Class 3 and Class 6 represent damage over the whole profile of the gear and lastly, Class 5 contains a relatively small portion of the gear that is damaged. The similarity between Class 2 and 4 and Class 3 and 6 is purposefully used to make it a challenging problem for the classifiers which will result in interesting decision regions.

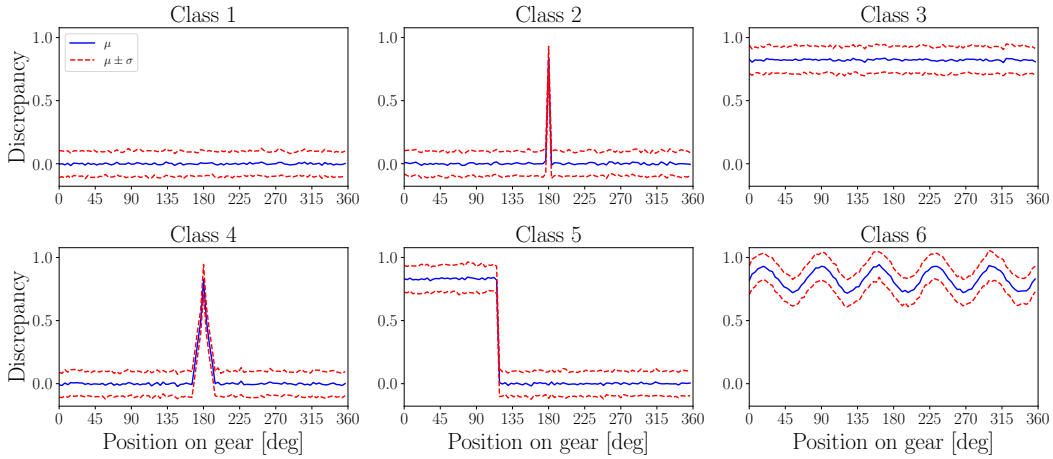


Figure 6.3: The mean μ and the standard deviation σ of 100 synchronous averages of the discrepancy signals of the six synthetic classes, investigated in Section 3. The data are generated with Equation (6.8) - (6.13).

The discrepancy features of the different classes, extracted with the procedure presented in Section 6.3.2 are presented in Figure 6.4(i) for the data in Figure 6.3. The assumption is

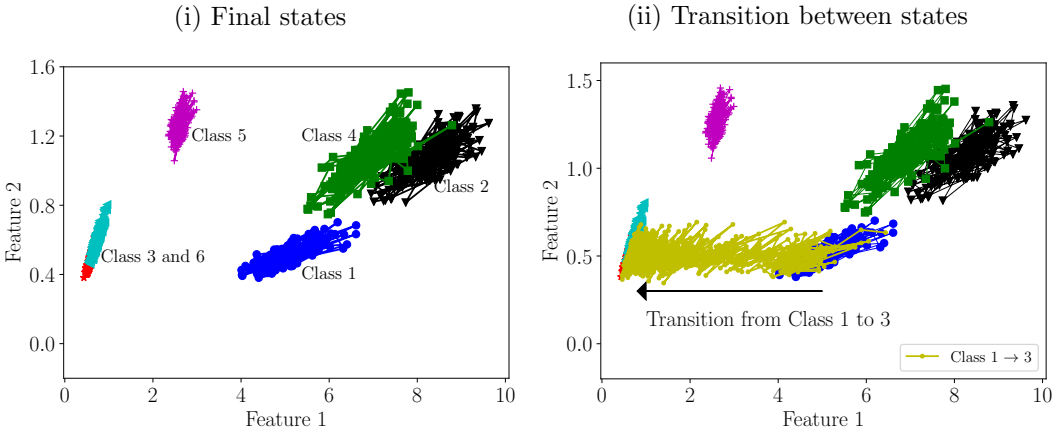


Figure 6.4: The features for the six classes in Figure 6.3 are presented in Figure 6.4(i). In Figure 6.4(ii), a stochastic transition from Class 1 to Class 3 is presented as an example.

made that the system starts from a healthy condition (i.e. Class 1 in Figure 6.3 and Figure 6.4(i)) and stochastically transitions to the other conditions with an example shown in Figure 6.4(ii). The transition between Class 1 and Class i is performed by linearly increasing the factor κ in the appropriate Equation e.g. the κ in Equation (6.13) is changed for a transition between Class 1 and Class 6. The stochastic transition seen in Figure 6.4(ii) is because β in Equation (6.10) is a random variable.

In the next section, CSR classification models are compared to the proposed OSR model with the data of the discrete classes in Figure 6.4(i) being used as training data.

6.4.2 Discrete machine conditions for model optimisation

In this section, a conventional classification process is used where discrete states (e.g. a gear with a broken tooth, a healthy gear etc.) are used as classes in the classification task to highlight the limitations of CSR models and the limitations of using discrete states for OSR models.

Due to the characteristics of the features in Figure 6.4(i), it is expected that Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) and Gaussian Naive Bayes (GNB) models will perform very well on the data and therefore the additional flexibility that neural networks and support vector machines provide are not required. The aforementioned models are implemented with scikit-learn [174] on the data in Figure 6.4(i) and compared to the OSR model discussed in Section 6.3.3.1 and Section 6.3.3.2. The training data and a query point are superimposed on the decision boundaries in Figure 6.5 for the LDA, QDA, LR and the OSR model. The query point highlights that the CSR classifiers predict a class label for data far from the training data and that the predictions are different due to the different characteristics of the decision boundaries. The classification error of the classifiers are given in Table 6.2 on the testing data, with the result of a GNB classifier presented as well. The LDA, QDA, LR and the OSR GMM classifiers perform equally well, with the GNB performing slightly worse. The classification error is attributed to the large overlap between Class 3 and Class 6. Even though the classification error is similar for the aforementioned models, their decision boundaries in Figure 6.5 have very different characteristics. The implication of this is that different predictions are made by the different classifiers as the data transition from Class 1 to another class. This emphasises that the classification errors are not the only important performance metric, but also the decision boundaries of the classifiers. The decision boundaries in a CSR framework are obtained from an optimisation process aiming to minimise the prediction error of the classifier for the historical data and may not necessarily have the desired properties for machine condition monitoring problems.

Table 6.2: Average classification error of the different models on the 200 testing measurements of each class, with the testing data generated similarly to the data in Figure 6.4(i).

LDA	QDA	GNB	LR	GMM
9.25%	9.92%	12.33%	9.25%	9.25%

The superiority of the OSR GMM with three mixture components is seen in Figure

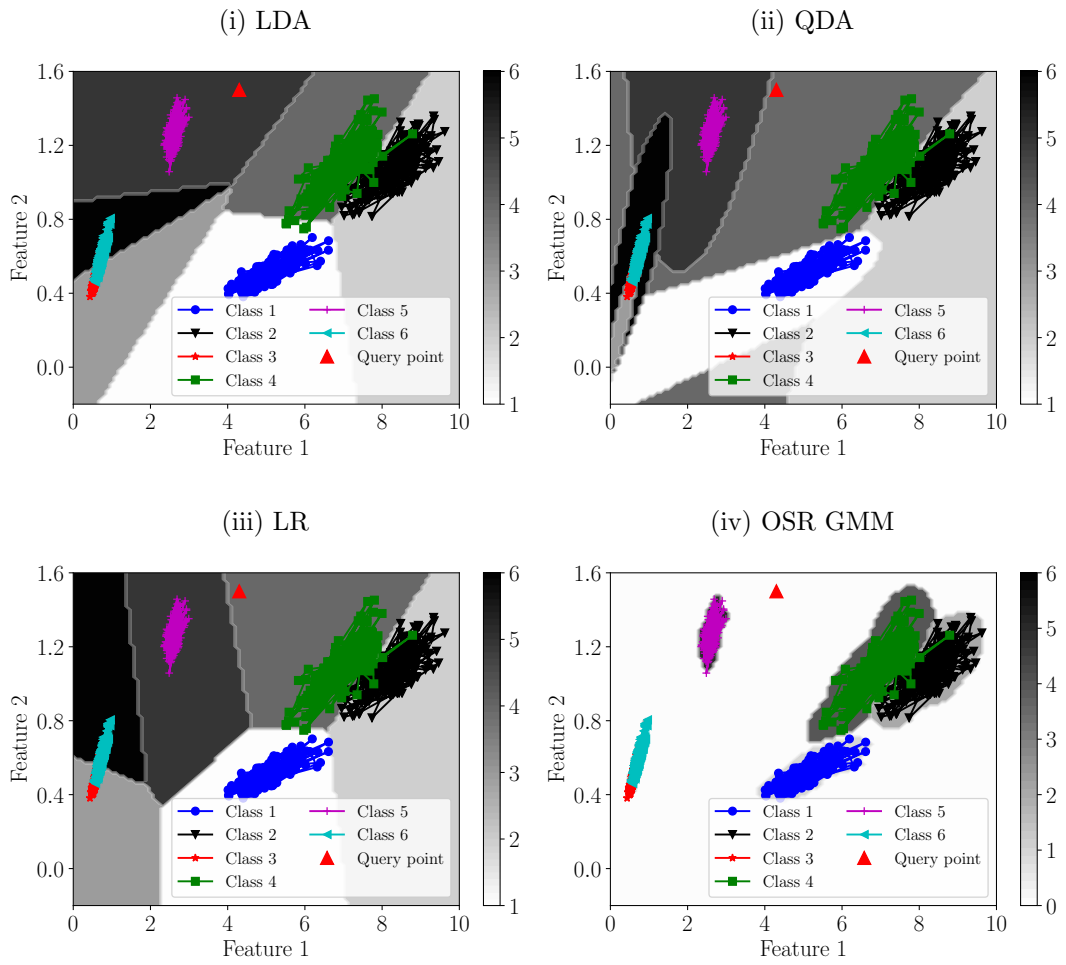


Figure 6.5: The decision regions for three classifiers using a CSR framework and the OSR model are superimposed on the training data in this figure. The colourbar indicates the class label, where a class label of 0 indicates that a novelty is detected in that region. Abbreviations: Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) and Open Set Recognition Gaussian Mixture Model (OSR GMM).

6.5(iv). The OSR GMM contains tightly fit boundaries around the regions supported by the historical data and a white region between the clusters which has a class label of 0. This indicates that the model is incapable of predicting the class label of the data in the regions that are not supported by the historical data and therefore novelties are detected in those regions. This is in contrast to the CSR classifiers, which can have misleading predictions in regions unsupported by the data.

The large novelty detection regions in Figure 6.5(iv) can however lead to problems when inferring the condition of a deteriorating machine, i.e. novelties will be detected in the transition process. A solution for this problem is therefore investigated in the next section.

6.4.3 Complete dataset for model optimisation

The OSR GMM decision boundaries in Figure 6.5(iv) are impractical for condition monitoring purposes because novelties will be detected almost continuously during the condition transition process. Therefore, it is proposed in this chapter that the data of the transition from Class 1 to the other classes are used for optimising the GMM classifier i.e. the whole dataset, as it transitions from Class 1 to Class 3 in Figure 6.4(ii), is labelled as Class 3 in this section. Therefore, no distinctions are made between the initial state and the final state when optimising the model.

The number of mixture components in the GMM of each class was determined with five-fold cross-validation, with the number of components determined to be four. The training data and the resulting decision regions of the OSR GMM are presented in Figure 6.6. It can be

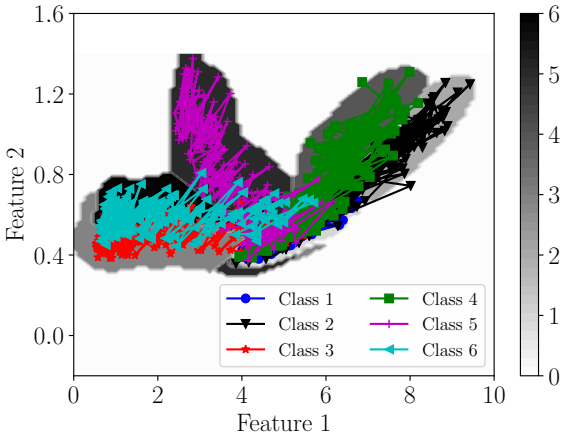


Figure 6.6: The decision regions for the OSR GMM that is optimised with the transition data is presented. For example, the data of the transition from Class 1 to Class 3 in Figure 6.4(ii) are the training data of Class 3 for the OSR GMM.

observed that the decision regions surround the training data, while the regions unsupported by the training data are classified as novelties i.e. a class label of 0 is obtained.

The posterior probability of the OSR GMM classifier using the complete dataset is compared to the posterior probability of the CSR LR classifier for the transition from Class 1 to Class 3 in Figure 6.7, with $\arg\max(P(C_i|\boldsymbol{r}))$ being used to infer the class label. A moving average with a length of 30 and an overlap of 25 measurements was used on the raw probability data to reduce the noise due to the stochastic transition process. The corresponding decision boundaries of the OSR GMM and CSR LR classifiers are shown in Figure 6.6 and Figure 6.5(iii) respectively. Initially, the data are in the healthy region and therefore the predicted class label is Class 1 for both classifiers. In Figure 6.7(i), the LR classifier incor-

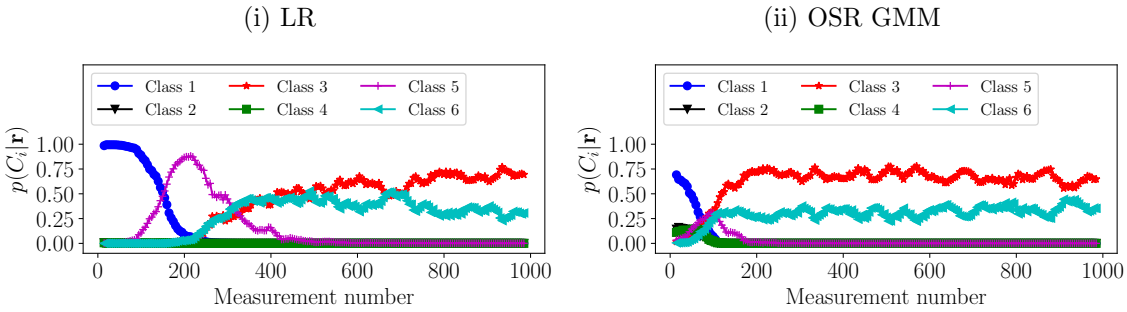


Figure 6.7: The posterior probability of the different classes for the LR classifier and the condition recognition system using the GMMs as the data stochastically transitions from Class 1 to Class 3. This is presented for the transition data in Figure 6.4(ii), using the decision boundaries in Figure 6.6 and Figure 6.5(iii).

rectly infers that Class 5 is present between measurement number 150 and 300 due to the characteristics of its decision boundary. In contrast to the LR classifier, the GMM condition recognition system in Figure 6.7(ii) correctly infers the class during the transition process. It is therefore clear from the results in Figure 6.6 and Figure 6.7(ii) that the decision regions can be controlled by learning the trajectory of the data with an appropriate classifier and the full training set (i.e. using continuous as opposed to discrete machine states), which is beneficial for the condition recognition process. This also emphasises that care should be taken when using non-linear classifiers that either transform the data to a high dimensional space e.g. support vector machines, or perform automatic feature selection e.g. neural networks. Using the aforementioned classifiers can result in the classes to be perfectly separated, but the decision regions may not preserve the important characteristics for machine condition monitoring applications.

The posterior probability in Figure 6.7(ii) does not indicate a change in condition between measurement number 200 and 1000 and can therefore not be used for fault trending. This is because the only aim of the condition recognition system is to determine whether the data belong to a specific class or not (when a novelty is detected). Therefore, it is suggested that the latent variables or hidden states of the GMM be used to detect changes in the features. The hidden states for the testing data, obtained from maximising the posterior distribution over the latent states as described in Ref. [114], are presented in Figure 6.8 for the model of Class 3, with three progression stages identified. The rapid alternation between two states indicate that the data reside in a region of the feature space well supported by two clusters i.e. there is a large overlap between the two clusters. The first hidden state is associated with the healthy data region, whereafter the active hidden state change as the data move

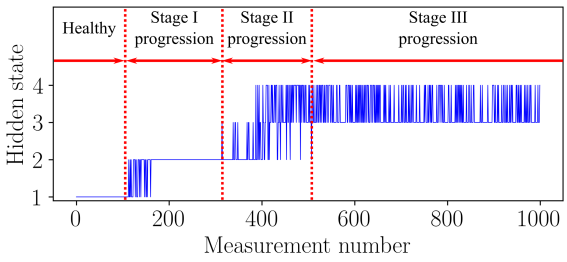


Figure 6.8: The hidden states of the GMM associated with the data of Class 3 as the data transition from Class 1 to Class 3.

through the different clusters, which can support maintenance decisions. For example, the first progression stage can signal a warning and if the latent state of the GMM reaches the final progression stage i.e. stage 3, the machine needs to be stopped immediately.

The results in this section validate the abilities of the proposed discrepancy features, they highlight the benefits of using an OSR as opposed to a CSR framework, and the importance of the decision boundaries in the machine condition monitoring field is emphasised as well. In the next section, the methodology is validated on experimental data.

6.5 Experimental validation

The proposed methodology is validated on experimental gearbox data, acquired under varying operating conditions, in this section.

An overview of the dataset is given in Section 6.5.1, whereafter the proposed methodology is investigated in Section 6.5.2 and compared to a conventional classification approach in Section 6.5.3.

6.5.1 Experimental dataset

Two experimental datasets are used in this investigation, with a detailed overview of the experimental setup given in Section 2.3.1. In the first experimental dataset, also discussed in Section 2.3.1, a healthy gearbox was monitored, whereafter the gearbox was disassembled and the gear was seeded with damage as shown in Figure 2.7(i). The gearbox was reassembled with the damaged gear and operated again with the operating conditions shown in Figure 2.8 until the gear tooth failed as shown in Figure 2.7(ii). In the second experiment, data were collected from a healthy gearbox whereafter the gearbox was disassembled and reassembled with a gear shown in Figure 6.9(i). The gearbox was operated again, with the operating conditions in Figure 2.8 imposed on the system, until the damaged tooth ultimately failed as

(i) Exp. 2: Before (ii) Exp. 2: After

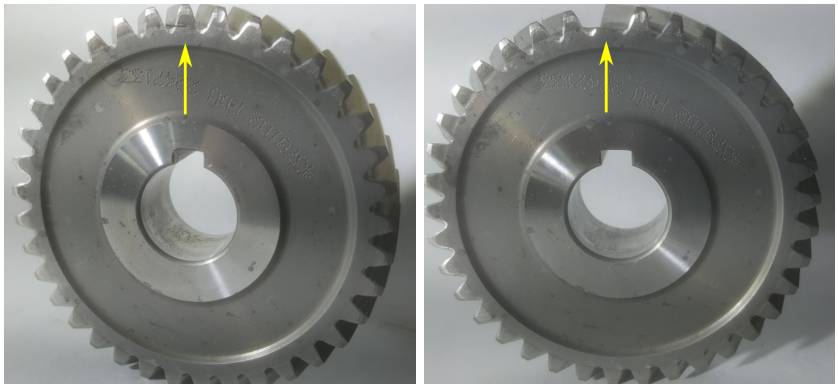


Figure 6.9: Damaged gears before and after the two fatigue experiments are presented with an arrow that indicates the location of the damage. The gear of the first experiment was tested for approximately 20 days before failure, in contrast to the second gear that ran approximately for 3 days before failing.

shown in Figure 6.9(ii). More severe damage was seeded in the gear in Figure 6.9(i) compared to the gear in Figure 2.7(i) by creating a slot that was 20% deeper in the gear tooth. This resulted in a significantly shorter time of operation (i.e. 3 days as opposed to 20 days) and a more abrupt failure for the damaged gear in the second experiment.

In all subsequent sections where the data are shown as a function of measurement number for the damaged gears, the first measurement was acquired for the gear in the initial condition (i.e. Figure 2.7(i) for experiment 1 or Figure 6.9(i) for experiment 2) and the last measurement was acquired for the gears in their final condition (i.e. Figure 2.7(ii) for experiment 1 or Figure 6.9(ii) for experiment 2). In contrast, the healthy gears were in the same condition for all healthy gear measurements of the two experiments. The healthy gear data will be referred to as healthy or a class label of 1 will be given, while the damaged gears will be referred to as damaged or a class label of 2 will be given where necessary. This is to be consistent with the class labels in Figure 6.2. If more damage modes were to be investigated, it would be necessary to use more descriptive labels for the damaged datasets.

6.5.2 Application of proposed methodology

In this section, the proposed methodology in Section 6.3 is investigated on the experimental data of Section 6.5.1. The first part of the investigation focuses on a novelty detection problem where only healthy historical data are available and the second part focuses on a novelty detection and classification problem where there are historical fault data available as well. A summary of the data used in the investigations of this section is presented in Table

6.3. Discrepancy analysis, the first step of the methodology in Figure 6.2, is investigated for

Table 6.3: The number of measurements used for the different datasets in the investigations, with the training and testing datasets being unique. The novelty detection training data refer to the case where the condition recognition system is only optimised with healthy data, where the classification training data refer to the case where the condition recognition system is optimised with historical fault data as well.

Dataset description	Training (Novelty detection)	Training (Classification)	Testing (Novelty detection & Classification)
Exp. 1: Healthy	44	44	50
Exp. 2: Healthy	0	0	50
Exp. 1: Damaged	0	166*	500*
Exp. 2: Damaged	0	0	215*

*The measurements are evenly spaced throughout the life of the gear.

the experimental datasets in the next section.

6.5.2.1 Discrepancy analysis

The discrepancy analysis step is implemented using the procedure described in Section 6.3.1 on the healthy historical experimental data. The number of latent states in the GMM is estimated by using five-fold cross validation which resulted in eight latent states i.e. $N_c = 8$ in Equation (6.1) to be used. The synchronous average of a healthy gearbox and a gearbox with localised gear damage, calculated with Equation (6.2), are presented in Figure 6.10. The

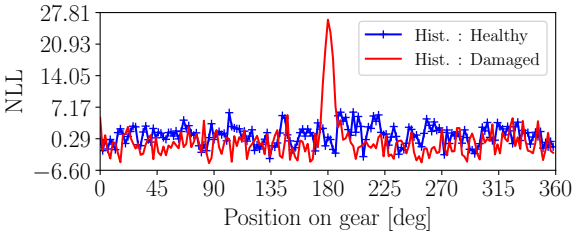


Figure 6.10: The synchronous average of the discrepancy signals of a healthy gearbox and of a gearbox with a damaged tooth at 180° is presented.

localised damage at 180° is easily identified when compared to the synchronous average of a healthy gearbox and looks similar to Class 2 and Class 4 in Figure 6.3.

Heyns et al. [155] illustrated that discrepancy analysis is better suited than using the synchronous average or power spectral density of the vibration signal for gearbox diagnostics under varying operating conditions. This is illustrated by comparing the synchronous averaged vibration signal and the synchronous averaged discrepancy signal in Figure 6.11 for two measurements of the damaged gearbox of the first experiment. It is evident that the synchronous averaged discrepancy signal is more sensitive to damage than the synchronous

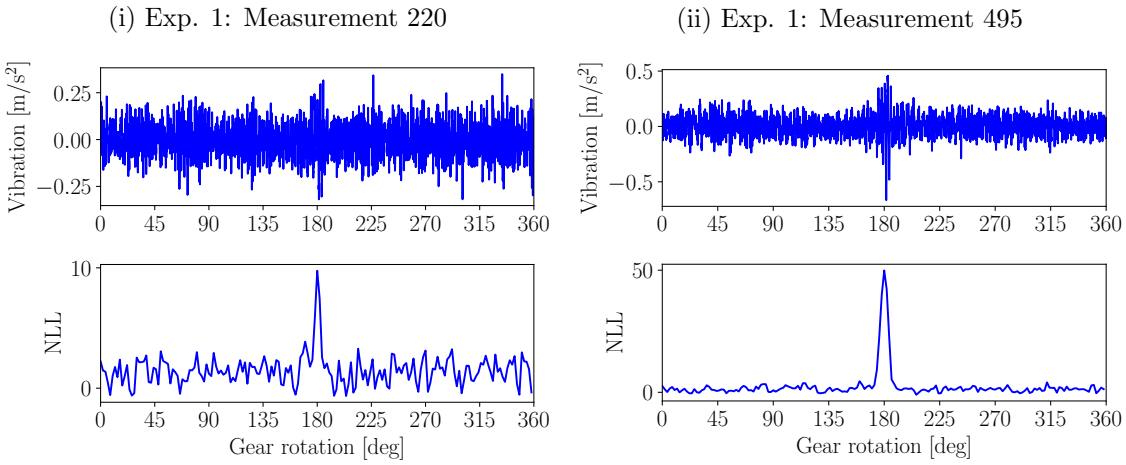


Figure 6.11: Synchronous averaged vibration signals are compared to synchronous averaged discrepancy signals of a gearbox with a damaged tooth at 180° for two measurements.

averaged vibration signal. Therefore, it is sensible to use the synchronous averaged discrepancy signal as opposed to the synchronous averaged vibration signal in the next step.

6.5.2.2 Discrepancy feature space of experimental dataset

The discrepancy features, discussed in Section 6.3.2, are presented in Figure 6.12 for the four gear datasets presented in Section 6.5.2 and Table 6.3. In Figure 6.12(i), all the datasets start at the zoomed portion box, also indicated by "Initial" and as the gear condition changes, the features move to "Final" with a trajectory in a similar direction as Class 2 and Class 4 in Figure 6.4(i). The large overlap in the healthy data region in Figure 6.12(ii) is attributed to

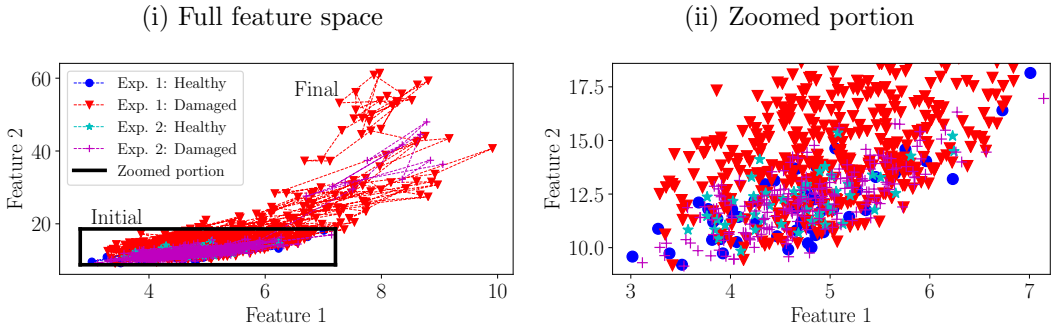


Figure 6.12: The full feature space of the four experiments is presented, with a zoomed view of the healthy feature region as well. In Figure 6.12(i), the starting position of the transition process, denoted "Initial", and the stopping position of the transition process, denoted "Final", are presented for clarity.

the fact that the damaged gears start approximately in a healthy condition, i.e. the seeded damage is small, whereafter the gears deteriorate during the experiments. The influence of the deterioration of the gear is clearly seen in Figure 6.12(i), with the two gears having

approximately the same trajectory. The slight differences in the trajectories are attributed to the fact that the seeded damage had different characteristics which resulted in a more abrupt failure for the second gear. The disassembling and reassembling process between the experiments also resulted in the features to have slightly different characteristics, however, the effect of this is negligible in comparison to the changes in the features as the condition of the gear changes.

In the next section, the condition recognition system is only optimised on healthy data, to test its ability to perform novelty detection.

6.5.2.3 Classification with only healthy historical data (Novelty detection)

It is necessary to test the ability of the condition recognition system to perform novelty detection when only historical data from a healthy machine are available. Therefore, it is assumed that only the healthy historical data in Table 6.3 are available for optimising the classifier presented in Figure 6.2. The discrepancy features are extracted from the aforementioned healthy historical data with the procedure discussed in Section 6.3.2, whereafter a single GMM is optimised on the healthy historical data with the procedure discussed in Section 6.3.3.1.

The objective of the condition recognition system is only to infer whether the features are from a healthy machine or not i.e. a novelty is detected and therefore it is unnecessary to use Bayes' rule given by Equation (6.5). Hence, the GMM of the healthy features and the decision rule given by Equation (6.7) are used in the condition recognition system to label the data as healthy (i.e. class label of 1) or as a novelty (i.e. class label of 0).

The next step is to evaluate whether the trained condition recognition system can infer the condition of the gearbox correctly for different gear conditions. The testing measurements from the four datasets in Table 6.3 are used to evaluate the performance of the condition recognition system, with the class labels presented in Figure 6.13. The healthy gear data in Figure 6.13(i) and 6.13(ii) are classified correctly with only three measurements being incorrectly labelled as novelties in Figure 6.13(i) i.e. they are labelled as class 0. These three measurements are outliers in the dataset.

Most of the initial damaged gear measurements of the first experiment are classified as healthy in Figure 6.13(iii) due to the small damage which resulted in an overlap in the healthy feature region as seen in Figure 6.12(ii). As the condition of the gear deteriorates, novelties are detected at an increasing rate until the gear tooth ultimately fails. In contrast

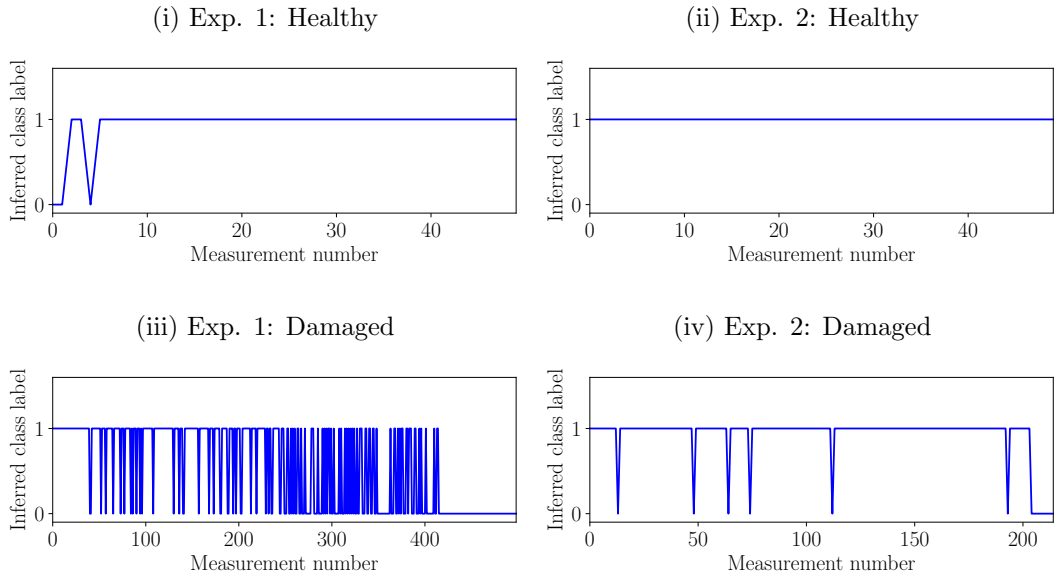


Figure 6.13: The output of the condition recognition system in Figure 6.2 for the different datasets when only healthy historical data are available. The inferred class label correspond to the class labels in Figure 6.2, with 0 indicating a novelty and 1 corresponding to the healthy class.

to the result in Figure 6.13(iii), the damaged gear of the second experiment does not contain evidence of a long transition from healthy to damage period in Figure 6.13(iv). A few of the measurements are classified as novelties which can indicate the presence of damage, with the failure occurring in the final stages of the experiment. The detection of the failure in the final stages of the experiment is not a deficiency of the method; the seeded damage was more severe for the second experiment than for the first experiment which resulted in a significantly shorter experimental time (i.e. 3 days as opposed to 20 days) due to the gear tooth failing abruptly.

The synchronous averaged discrepancy signals, corresponding to the measurements where novelties are detected in Figure 6.13(iii) and Figure 6.13(iv), are presented in Figure 6.14. The

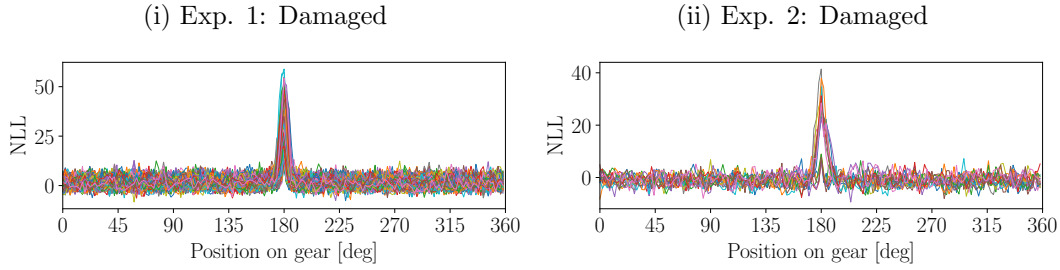


Figure 6.14: The synchronous averaged discrepancy signals which correspond to the novelties detected in Figure 6.13(iii) and 6.13(iv) are presented in Figure 6.14(i) and Figure 6.14(ii), respectively. Different colours are used to emphasise that the synchronous averages of different measurements are presented.

many synchronous averages in Figure 6.14(i) make it difficult to interpret the results in the figure and therefore a three-dimensional figure is presented in Figure 6.15 of the same results. There are not many novelties detected in Figure 6.13(iv) and not many synchronous averages presented in Figure 6.14(ii), which makes the results in Figure 6.14(ii) easy to interpret and therefore a three-dimensional surface is not shown for that data. The results in Figures 6.14 and 6.15 indicate that the novelties, detected in Figure 6.13(iii) and Figure 6.13(iv), are due to localised damage on a gear tooth. Hence, it is not only possible to automatically label the

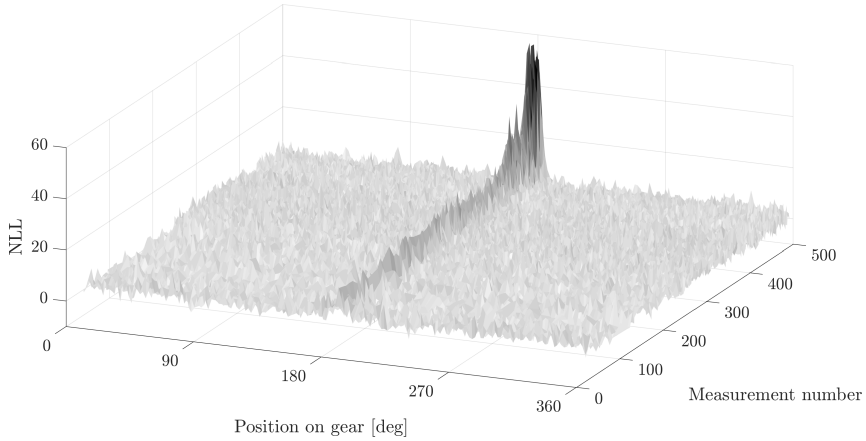


Figure 6.15: The results in Figure 6.14(i) are presented as a three-dimensional surface to make it easier to interpret the results.

measurements from a damaged gearbox under the varying operating conditions as novelties with the condition recognition system, but also to characterise the damage by manually investigating the corresponding synchronous averaged discrepancy signal.

In the next section, it is investigated whether it is possible to update the condition recognition system after historical fault data become available. The ability of the condition recognition system to correctly infer the condition of the gearbox is investigated as well.

6.5.2.4 Classification with healthy and damaged gear historical data

In this section, it is assumed that the historical damaged gear data, presented in Section 6.5.1 and Table 6.3, are available after the condition recognition system has already been optimised as a novelty detector in the previous section. Hence, the historical fault data, labelled as the training data in Table 6.3 need to be incorporated into the condition recognition system. The training data of the damaged class are spaced throughout the life of the gear from its initial damaged state and its failure state as described in Table 6.3.

The condition recognition system is updated by firstly optimising a new GMM on the

damaged class discrepancy signal features, whereafter the prior probabilities of the healthy and damaged gear classes are calculated as discussed in Section 6.3.3.1. The resulting prior probabilities of the classes are therefore $P(C_{healthy}) = 2/3$ and $P(C_{damaged}) = 1/3$. Hence, the condition inference procedure works in two steps: Firstly, the likelihood $p(\boldsymbol{r}|C_i)$ of the GMM of the damaged gear (optimised in this section) and the GMM of the healthy gear (optimised in the previous section) are calculated for the discrepancy signal features of a new measurement. Thereafter, the likelihoods and the prior probabilities are used with Bayes' rule to infer the condition of the gear as described in Section 6.3.3.1. The inferred class is then either accepted or rejected (i.e. labelled as a novelty) or not according to the procedure in Section 6.3.3.2.

The OSR GMM using Bayes' rule, with and without the decision rule given by Equation (6.7), are investigated on the four experimental datasets, with the resulting decision boundaries presented in Figure 6.16. Both classifiers distinguish the different classes well,

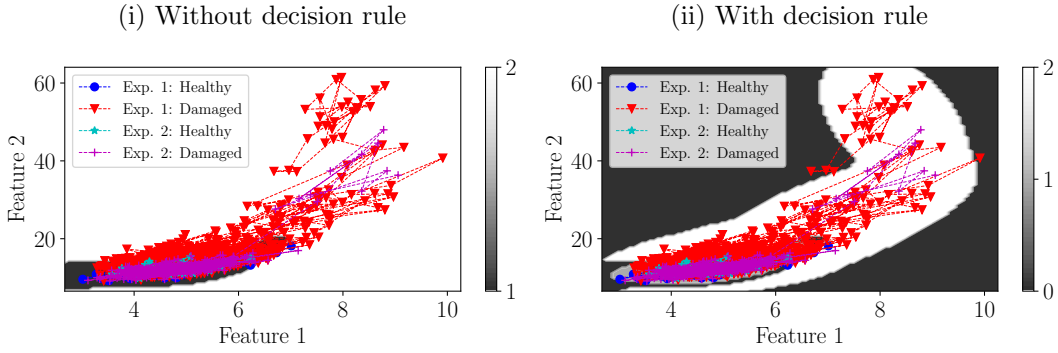


Figure 6.16: The decision region for the GMM condition recognition system in Section 6.3.3.1 with and without the decision rule is superimposed on the features of the experimental datasets. Note: Different colour bar ranges are used in Figure 6.16(i) and 6.16(ii)

but the classifier without the decision rule is incapable of detecting new classes and outliers as evidenced in Figure 6.16(i). The classifier with the decision rule learned the trajectory of the data as the gear's condition deteriorates and can indicate when there is a region in the feature space unsupported by the training data as seen in Figure 6.16(ii).

The posterior probability of the different classes calculated using Equation (6.5) for the four datasets, are smoothed using a moving average window with a length of 5% the number of measurements and a 90% overlap, and presented in Figure 6.17. The healthy gear data in Figure 6.17(i) and Figure 6.17(ii) are classified correctly. The damaged gear posterior probabilities in Figure 6.17(iii) and Figure 6.17(iv) indicate that the gears started from approximately healthy conditions, whereafter the damage progressed up to the failure of the

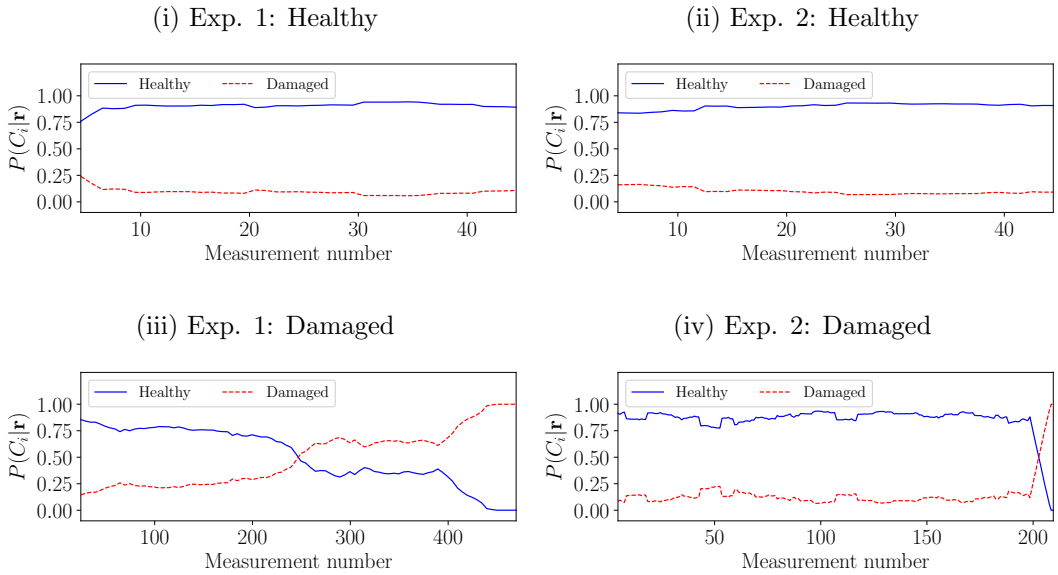


Figure 6.17: Posterior probability of the two classes for the four experimental datasets using the condition recognition system with the decision rule.

tooth. The damage in Figure 6.17(iii) can be detected at the 244th measurement, while the damage in Figure 6.17(iv) is detected at the 203rd measurement. The different characteristics in Figure 6.17(iii) and Figure 6.17(iv) are attributed to the different characteristics of the seeded damage and experimental time discussed in Section 6.5.1.

The change in probability for the damaged gears is evidence that the condition of the gears change with measurement number, however, it is shown in Section 6.4 that the hidden states contain useful information as well. The hidden states of the GMM for the associated class, determined by the procedure described in Bishop [114], are presented in Figure 6.18 for the damaged gears. The two damaged datasets exhibit very similar characteristics. The first

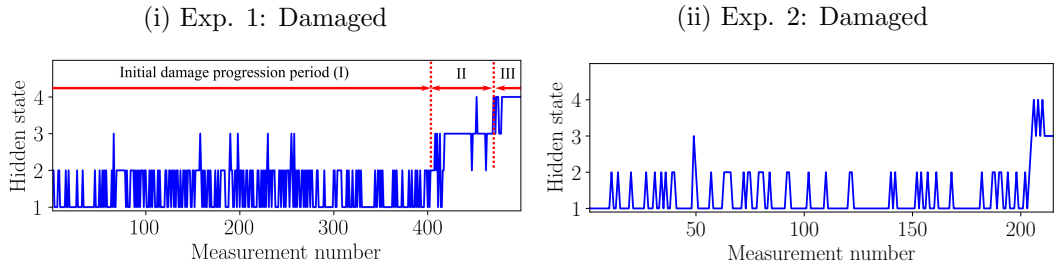


Figure 6.18: The hidden state in which the data reside as a function of measurement number for the damaged gears. Three stages are identified in Figure 6.18(i) and labelled with Roman numerals.

two states are visited for the initial measurements and as the gear deteriorates, state 3 and 4 become more active. The presence of Stage II and III in Figure 6.18(i) indicate that the

tooth failed or that failure is imminent. Due to the abrupt failure of the second gear tooth, the progression between Stage II and III is not seen. The hidden states can therefore be used in the maintenance decision making process, for example if the gear is classified as damaged and if state 2 is visited more frequently, a warning is issued and if state 3 and 4 (i.e. Stage II to III) is reached, the machine needs to be stopped as soon as possible.

It is concluded that the methodology is capable of detecting novelties and classifying the data correctly. The benefits of the open set recognition framework is clearly highlighted on the synthetic and experimental data. However, the benefits of using discrepancy analysis as a pre-processor for the classification problem has not yet been investigated. This investigation is performed in the next section.

6.5.3 Comparison to a Conventional Classification Procedure (CCP)

The aim of the investigation in this section is to highlight the benefits of using discrepancy analysis prior to extracting the features used for classification. The feature extraction procedure in Section 6.3.1.1 is extended for a Conventional Classification Procedure (CCP) to circumvent the use of discrepancy analysis in the condition inference procedure. The following process is used:

- Calculate the CWT of the order tracked vibration signal for the first three gear mesh components according to Section 6.3.1.1.
- Calculate the statistics given in Table 6.1 for the entire measurement. Hence, the windowing procedure is not used and this results in a single 30 dimensional feature set to be extracted for each measurement of the four datasets.
- The condition recognition system cannot be efficiently optimised due to the high dimensionality of the features, hence the dimensionality of the features is reduced with Principal Component Analysis (PCA). PCA is used, because it performed the best of all dimensionality reduction methods used for gear diagnostics in Ref. [77]. The PCA model is optimised on the historical data which is used to project the features on a lower-dimensional principal component feature space. The accumulative contribution rate is used to determine the appropriate dimensionality of the principal component feature space, which resulted in three dimensions to be used.

In Figure 6.19, the proposed discrepancy features and the CCP features are compared for 44 healthy historical measurements and 44 historical damage measurements, with the

latter spaced evenly over two days before the first gear failed. The 44 historical damage measurements used in Figure 6.19(ii) are not included in Table 6.3 and are used to compare the sensitivity of the features to faults. The two classes in Figure 6.19(i) are fairly well

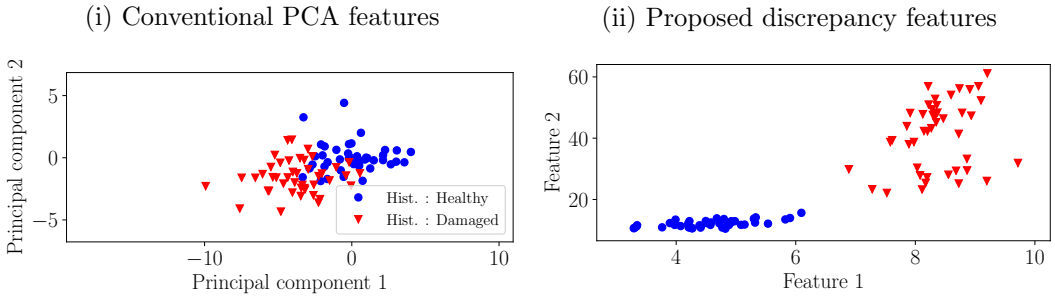


Figure 6.19: A comparison of the feature space using the proposed discrepancy features in Section 6.3.1.1 and the conventional features discussed in Section 6.5.3.

separated, with only a few samples overlapping. However, the proposed discrepancy analysis features in Figure 6.19(ii) separate the two classes significantly better than using the principal components of the raw features, which indicate that the proposed discrepancy features are more sensitive to damage. The large variance of the historical damaged data in Figure 6.19(ii) is attributed to features' sensitivity to the large change in machine condition as the gear tooth failed.

In Section 6.5.2, 44 healthy measurements are used for optimising the healthy GMM and 166 damaged gear measurements are used for optimising the model of the damaged GMM as outlined in Table 6.3. However, the 166 damaged measurements did not work very well with the PCA model due to the large overlap between the healthy historical data and the healthy portion of the historical fault data as shown in Figure 6.20(i). In contrast, the condition recognition system in Figure 6.20(ii) performs significantly better when using the 44 damaged measurements in Figure 6.19(ii). Performing dimensionality reduction with the complete dataset did not separate the classes well which resulted in poor decision boundaries. Hence, the historical features presented in Figure 6.19(i) are used in the CCP for all subsequent analyses.

The principal component feature space, superimposed on the decision boundaries of the condition recognition system, is presented in Figure 6.21 for the four datasets. It is difficult to see the trajectory of the features as the gear condition deteriorates for the two experiments in contrast to Figure 6.12. There a difference in the trajectory in the final stages of the condition transition process for the two experiments i.e. Exp. 1: Damaged and Exp. 2:

(i) With 166 historical damage measurements (ii) With 44 historical damage measurements

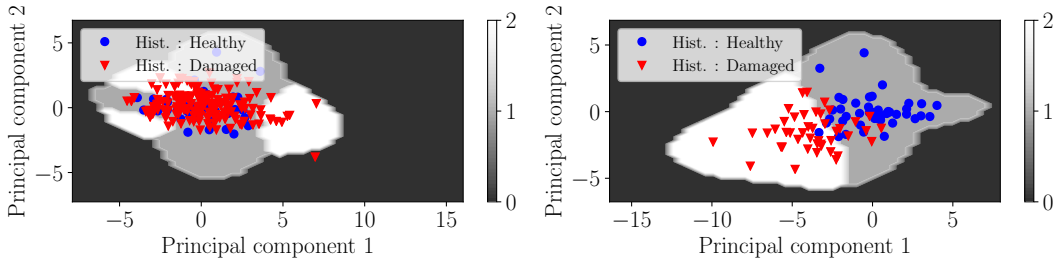


Figure 6.20: The decision region for the proposed condition recognition procedure with the 44 measurements in Figure 6.19(i) used for optimising the damaged model in contrast to using the damaged measurements discussed in Table 6.3. A class label of 0 corresponds to a novelty, a class label of 1 corresponds to a healthy gear and 2 corresponds to a gear with localised damage.

Damaged in Figure 6.12. This resulted in some measurements of the second experiment with

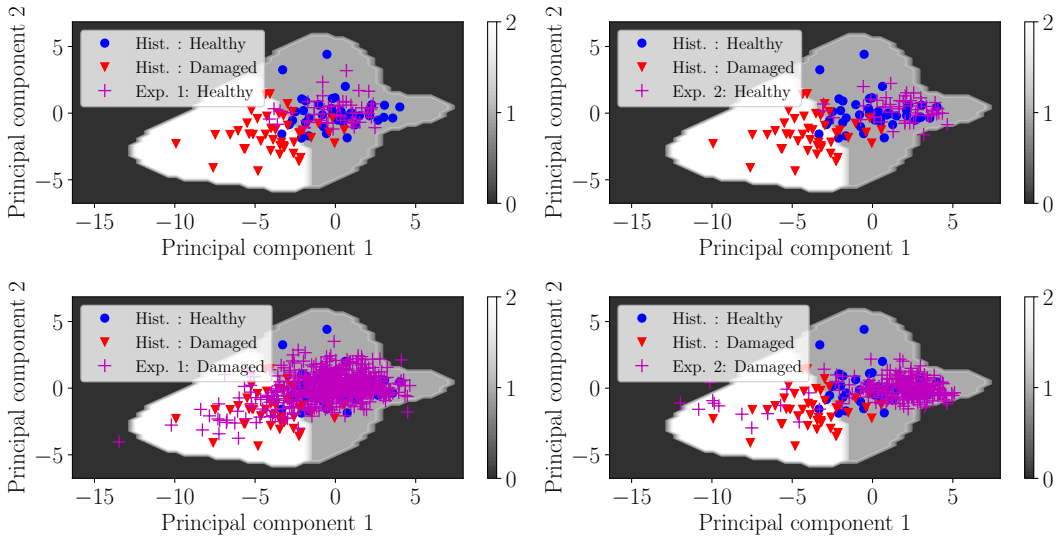


Figure 6.21: The principal component feature space of the features from the four experimental datasets using the procedure discussed in Section 6.5.3 is shown with the decision regions of the condition recognition system.

the damaged gear i.e. Exp. 2: Damaged to be incorrectly classified as outliers.

The posterior probability of the datasets, calculated with the same procedure as section 6.5.2, is presented in Figure 6.22. It is clear from the results that the classifier is capable of predicting the class label correctly for the healthy measurements and in the final measurements of the damaged gears, the damage is detected. The damage in Figure 6.22(iv) is detected at the same measurement number as in Figure 6.17(iv) due to the fact that the tooth failed abruptly. However, the damage in Figure 6.22(iii) is detected at the 460th measurement, which is significantly later than in Figure 6.17(iii), because the discrepancy features are more sensitive to changes in machine condition.

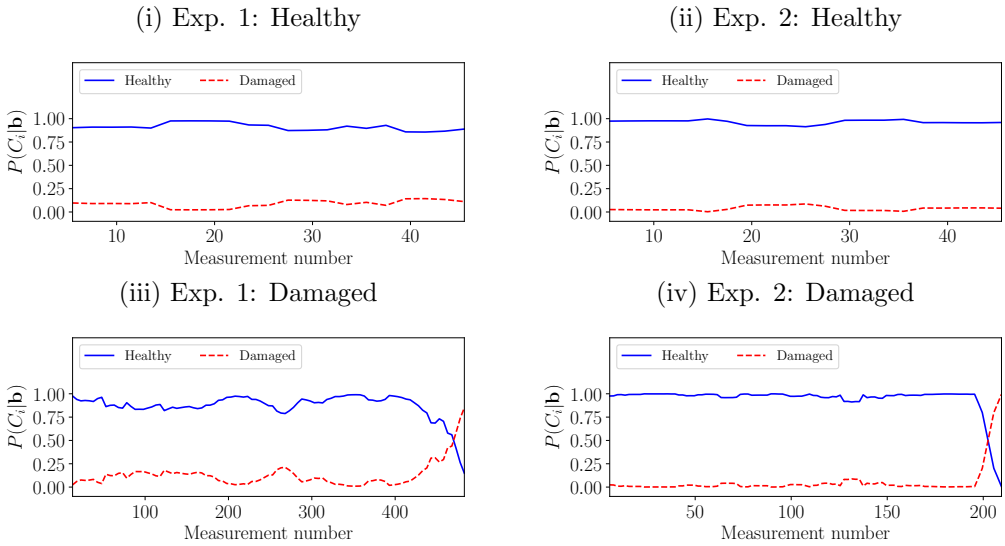


Figure 6.22: The posterior probabilities of the two classes for the four experiments using the principal component features with the condition recognition system, i.e. discrepancy analysis is not used.

Hence, processing the data with discrepancy analysis results in a better performance (e.g. it detects damage earlier and clearer decision boundaries are obtained) and therefore it is sensible to include discrepancy analysis in the OSR methodology.

6.6 Conclusion and recommendations

In this chapter, it is emphasised that machine condition monitoring problems should be addressed as open set recognition and not as closed set recognition problems. An open set condition recognition methodology is proposed which can be used to automatically infer the condition of gears using an open set recognition framework. The proposed methodology is effective in the fault diagnosis task, because the historical fault data are utilised in the condition recognition process and condition recognition predictions are only made in regions well supported by the historical data, with data from new damage modes or outliers being labelled as novelties. The methodology is validated on synthetic as well as experimental datasets, where the importance of using an open set recognition framework with the full training dataset is illustrated.

In future investigations, it is suggested that the methodology be extended to include the different damage modes that can be encountered in gearboxes. Deep learning methods, which have much potential in the condition monitoring field, also need to be extended to accommodate the openness of the condition monitoring problem.

Chapter 7

Conclusion and recommendations

7.1 Conclusion of thesis

In this thesis, effective fault diagnosis methodologies are proposed for gearboxes operating under varying operating conditions by utilising the available prior knowledge about the machine. Prior knowledge can be used to obtain effective fault diagnosis methods as follows:

- Prior knowledge related to the kinematics of the gearbox (e.g. gear ratios, bearing fault frequencies) can be used to extract diagnostic rich information, which can ultimately be used to not only detect damage, but the damaged component can be inferred as well.
- Prior knowledge related to the characteristics of the vibration data can be utilised to select the appropriate signal analysis techniques to extract or enhance the diagnostic information in the data and it can be used to discern between the healthy parts and damaged parts of rotating machine components e.g. to discern between healthy gear teeth and damaged gear teeth.
- Prior knowledge in the form of historical data can be used to supplement the available signal analysis techniques, to ensure that damage can be automatically detected, localised and that changes in the condition of the component can also be detected.

In this thesis, five methodologies are proposed to illustrate how prior knowledge can be used for effective fault diagnosis. More specifically, different data availability levels, summarised in Figure 1.2, and different signal analysis techniques, are investigated in the methodologies. In Chapter 2, a novel methodology is developed to automatically detect localised gear anomalies without any historical data being available. This methodology is especially

important, because some gearboxes may not have any historical data available (e.g. machines for which condition monitoring has not been performed in the past) and localised gear damage significantly decreases the life of the gearbox. It is also illustrated in this methodology that Bayesian techniques can be utilised to incorporate uncertainty in the condition inference process. The methodology was investigated on experimental gearbox data, acquired under constant operating conditions, and experimental gearbox data, acquired under varying operating conditions.

Designing fault diagnosis techniques for a gearbox with only healthy historical data is very important, because gearboxes operate for relatively long periods in a healthy condition, which makes healthy historical data easy to obtain. Therefore, three methodologies are developed for gearboxes where only healthy historical data are available, with each methodology specifically designed to highlight how historical data can be used to supplement a different class of signal analysis technique.

In the first methodology presented in Chapter 3, discrepancy analysis is extended for performing bearing diagnostics under varying speed conditions using time-frequency analysis techniques such as the wavelet packet transform, the empirical mode decomposition and the instantaneous power spectrum to extract diagnostic rich features, whereafter a healthy data model is used to assign a localised novelty detection score to the data, which is used to obtain a discrepancy signal which can be processed to detect, localise and trend damage. In this methodology, a discrepancy signal processing method is also developed to compensate for varying speed conditions, with the processed discrepancy signal being a reliable estimate of the condition of the machine under varying speed conditions. The methodology was investigated on numerical gearbox data, experimental bearing data acquired under constant operating conditions and experimental bearing data acquired under varying speed conditions.

In Chapter 4, a methodology is developed to supplement the spectral coherence, a very powerful second-order cyclostationary analysis technique, with healthy historical data. The spectral coherence is utilised to decompose the signal into a cyclic order-spectral frequency plane, whereafter features can be extracted from the spectral coherence and the healthy features are modelled with a probabilistic model. It is shown that the marginal distribution of the features of a specific characteristic can be used to not only detect localised damage, but also to determine which component is damaged. The methodology was investigated on numerical gearbox data and experimental gearbox data acquired under varying operating conditions, where the modified improved envelope spectrum is used as basis for feature

extraction.

The last methodology for this data availability class, presented in Chapter 5, utilises narrowband feature extraction techniques to extract the statistics of different frequency bands in a signal, whereafter the statistics are modelled and used to enhance the novel information in the vibration signal. The novelty rich signal can be analysed using conventional signal analysis techniques such as the synchronous average, synchronous squared envelope and the spectral coherence to infer the condition of the machine. After extracting a novelty rich signal with the proposed methodology, it is illustrated that the novelty rich signal is significantly more sensitive to damage, which improves the ability of conventional signal analysis techniques to detect, localise and trend damage. The methodology was investigated on numerical gearbox data as well as experimental gearbox data, acquired under varying operating conditions, with the squared envelope infogram used as a feature extraction technique.

In Chapter 6, the last data availability level is considered by developing a methodology to automatically infer the condition of the machine by using the available historical data. It is illustrated that the condition monitoring problem is actually an open set recognition problem, where only historical data of a sample of a population of damage modes are available and therefore the inferred condition should be treated with care. Conventional supervised learning approaches do not recognise the openness of the condition monitoring problem, which means that erroneous results can be obtained when a new damage mode, not present in the historical data, is encountered. The open set recognition framework makes it necessary to use the full dataset (i.e. as the healthy machine degrades in its failure state) as the training dataset, as opposed to using discrete states. This is performed to ensure that predictions can be made throughout the life of the machine, with the added benefit that it requires less human intervention to label the dataset. In this methodology, it is also illustrated that by using discrepancy analysis as a pre-processing step, the performance of the methodology can be significantly improved. The methodology was investigated on a synthetic dataset, where the benefits of using an open set recognition methodology are illustrated, and on two experimental gearbox datasets.

Hence, the proposed methodologies, which rely on the availability of prior knowledge, can be used to perform effective fault diagnosis.

7.2 Recommendations for future work

The following recommendations are made for future work in the gearbox condition monitoring field:

- The purpose of this thesis was not to find the optimal methodologies for gearbox fault diagnosis, and therefore standard signal analysis and feature modelling techniques, which have been widely established and proven to be successful in literature, were used. However, there is much potential for investigating and optimising the methodologies to improve their performance.

- The experimental gearbox data were only acquired for healthy gears and gears with localised damage. There are many other damage modes that can be encountered in gearboxes, for example, distributed gear damage is widely encountered in wind turbine gearboxes. It is therefore necessary to apply the proposed methods on other typical damage modes such as wear, pitting and scuffing to validate their capabilities further.

- The datasets considered in this thesis were generated in a controlled laboratory environment. Even though industrial gearboxes were used in the experimental gearbox datasets and the experiments were performed under time-varying operating conditions, the operating conditions do not have the same non-stationary characteristics that could be encountered in industrial machines such as wind turbines. Therefore, it is suggested that the methodologies need to be investigated on industrial gearboxes in real application environments, acquired under non-stationary operating conditions, to fully validate them.

- Instantaneous angular speed measurements have much potential for gearbox condition monitoring applications. Therefore, the benefits of using the proposed methodologies, with instantaneous angular speed measurements used as source of condition monitoring data, need to be investigated.

Bibliography

[1] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, and L. Rambault. Gearbox condition monitoring in wind turbines: A review. *Mechanical Systems and Signal Processing*, 111:251–264, 2018.

[2] I. El-Thalji and J. P. Liyanage. On the operation and maintenance practices of wind power asset. *Journal of Quality in Maintenance Engineering*, 18(3):232–266, 2012.

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

[4] A. Prajapati, J. Bechtel, and S. Ganesan. Condition based maintenance: A survey. *Journal of Quality in Maintenance Engineering*, 18(4):384–400, 2012.

[5] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel. Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications. *Mechanical Systems and Signal Processing*, 42(1-2):314–334, 2014.

[6] Y. Lin, L. Tu, H. Liu, and W. Li. Fault analysis of wind turbines in China. *Renewable and Sustainable Energy Reviews*, 55:482–490, 2016.

[7] R. B. Randall. *Vibration-based condition monitoring.* Wiley, 2011.

[8] F. Pedro, G. Márquez, A. Mark, J. María, P. Pérez, and M. Papaelias. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy*, 46:169–178, 2012.

[9] R. Zimroz, W. Bartelmus, T. Barszcz, and J. Urbanek. Diagnostics of bearings in presence of strong operating conditions non-stationarity - A procedure of load-dependent features processing with application to wind turbine bearings. *Mechanical Systems and Signal Processing*, 46(1):16–27, 2014.

[10] W. Bartelmus, F. Chaari, R. Zimroz, and M. Haddar. Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis. *European Journal of Mechanics, A/Solids*, 29(4):637–646, 2010.

[11] W. Bartelmus and R. Zimroz. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing*, 23 (5):1528–1534, 2009.

[12] B.L. Eggers, P.S. Heyns, and C.J Stander. Using computed order tracking to detect gear condition aboard a dragline. *The Journal of The Southern African Institute of Mining and Metallurgy*, 107, 2007.

[13] C. J. Stander, P. S. Heyns, and W. Schoombie. Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions. *Mechanical Systems and Signal Processing*, 16(6):1005–1024, 2002.

[14] W. Bartelmus and R. Zimroz. Vibration condition monitoring of planetary gearbox under varying external load. *Mechanical Systems and Signal Processing*, 23(1):246–257, 2009.

[15] J. Urbanek, T. Barszcz, R. Zimroz, and J. Antoni. Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions. *Measurement*, 45(7):1782–1791, 2012.

[16] R. B. Randall. A new method of modeling gear faults. *Journal of Mechanical Design*, 104(2):259–267, 1982.

[17] F. Chaari, W. Bartelmus, R. Zimroz, T. Fakhfakh, and M. Haddar. Gearbox vibration signal amplitude and frequency modulation. *Shock and Vibration*, 19(4):635–652, 2012.

[18] O. D. Mohammed, M. Rantatalo, J. O. Aidanpää, and U. Kumar. Vibration signal analysis for gear fault diagnosis with various crack progression scenarios. *Mechanical Systems and Signal Processing*, 41(1-2):176–195, 2013.

[19] C. J. Stander and P. S. Heyns. Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions. *Mechanical Systems and Signal Processing*, 19(4):817–835, 2005.

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

[21] I. El-Thalji and E. Jantunen. A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, 60-61:252–272, 2015.

[22] M. Cerrada, R. V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. Valente de Oliveira, and R. E. Vásquez. A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99:169–196, 2018.

[23] J. Antoni, F. Bonnardot, A. Raad, and M. El Badaoui. Cyclostationary modelling of rotating machine vibration signals. *Mechanical Systems and Signal Processing*, 18(6): 1285–1314, 2004.

[24] G. Dalpiaz, A. Rivola, and R. Rubini. Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears. *Mechanical Systems and Signal Processing*, 14(3):387–412, 2000.

[25] J. Lin. A Review and strategy for the diagnosis of speed-varying machinery. In *Prognostics and Health Management (PHM), 2014 IEEE Conference on. IEEE*, 2014.

[26] 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.

[27] J. Antoni. Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009.

[28] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mechanical Systems and Signal Processing*, 84:200–226, 2017.

[29] P. D. McFadden and J. D. Smith. Model for the vibration produced by a single point-defect in a rolling element bearing. *Journal of Sound and Vibration*, 96(1):69–82, 1984.

[30] C. J. Stander and P. S. Heyns. Transmission path phase compensation for gear monitoring under fluctuating load conditions. *Mechanical Systems and Signal Processing*, 20(7):1511–1522, 2006.

[31] D. Abboud, S. Baudin, J. Antoni, D. Rémond, M. Eltabach, and O. Sauvage. The spectral analysis of cyclo-non-stationary signals. *Mechanical Systems and Signal Processing*, 75:280–300, 2016.

[32] K.R. Fyfe and E.D.S. Munck. Analysis of computed order tracking. *Mechanical Systems and Signal Processing*, 11(2):187–205, 1997.

[33] D. H. Diamond, P. S. Heyns, and A. J. Oberholster. Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression. *Mechanical Systems and Signal Processing*, 81:402–418, 2016.

[34] F. Bonnardot, M. El Badaoui, R. B. Randall, J. Danière, and F. Guillet. Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). *Mechanical Systems and Signal Processing*, 19(4):766–785, 2005.

[35] M. Zhao, J. Lin, X. Wang, Y. Lei, and J. Cao. A tacho-less order tracking technique for large speed variations. *Mechanical Systems and Signal Processing*, 40(1):76–90, 2013.

[36] J. Urbanek, T. Barszcz, and J. Antoni. A two-step procedure for estimation of instantaneous rotational speed with large fluctuations. *Mechanical Systems and Signal Processing*, 38(1):96–102, 2013.

[37] Q. Leclère, H. André, and J. Antoni. A multi-order probabilistic approach for Instantaneous Angular Speed tracking debrie fi ng of the CMMNO'14 diagnosis contest. *Mechanical Systems and Signal Processing*, 81:375–386, 2016.

[38] S. Schmidt, P.S. Heyns, and J.P. de Villiers. A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process. *Mechanical Systems and Signal Processing*, 100:630–646, 2018.

[39] K. C. Gryllias and I. A. Antoniadis. Estimation of the instantaneous rotation speed using complex shifted Morlet wavelets. *Mechanical Systems and Signal Processing*, 38 (1):78–95, 2013.

[40] B. Li and X. Zhang. A new strategy of instantaneous angular speed extraction and its application to multistage gearbox fault diagnosis. *Journal of Sound and Vibration*, 396:340–355, 2017.

[41] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach. Deterministic-random separation in nonstationary regime. *Journal of Sound and Vibration*, 362:305–326, 2016.

[42] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen. Review on dynamics of cracked gear systems. *Engineering failure analysis*, 55:224–245, 2015.

[43] P. D. McFadden. Examination of a technique for the early detection of failure in gears by signal processing of the time domain average of the meshing vibration. *Mechanical Systems and Signal Processing*, 1(2):173–183, 1987.

[44] W. Bartelmus. Mathematical modelling and computer simulations as an aid to gearbox diagnostics. *Mechanical Systems and Signal Processing*, 15(5):855–871, 2001.

[45] I. Howard, S. Jia, and J. Wang. The dynamic modelling of a spur gear in mesh including friction and a crack. *Mechanical Systems and Signal Processing*, 15(5):831–853, 2001.

[46] F. Chaari, T. Fakhfakh, and M. Haddar. Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness. *European Journal of Mechanics, A/Solids*, 28(3): 461–468, 2009.

[47] O. D. Mohammed, M. Rantatalo, and J.-O. Aidanpää. Dynamic modelling of a one-stage spur gear system and vibration-based tooth crack detection analysis. *Mechanical Systems and Signal Processing*, 54-55:293–305, 2015.

[48] F. Chaari, W. Baccar, M. S. Abbes, and M. Haddar. Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. *European Journal of Mechanics, A/Solids*, 27(4):691–705, 2008.

[49] P. D. McFadden. A revised model for the extraction of periodic waveforms by time domain averaging. *Mechanical Systems and Signal Processing*, 1(1):83–95, 1987.

[50] N. Tandon. A comparison of some vibration parameters for the condition monitoring of rolling element bearings. *Measurement*, 12(3):285–289, 1994.

[51] C. Hu, W. A. Smith, R. B. Randall, and Z. Peng. Development of a gear vibration indicator and its application in gear wear monitoring. *Mechanical Systems and Signal Processing*, 76-77:319–336, 2016.

[52] A.C. McCormick and A.K. Nandi. Cyclostationarity in rotating machine vibrations. *Mechanical Systems and Signal Processing*, 12(2):225–242, 1998.

[53] A. Raad, J. Antoni, and M. Sidahmed. Indicators of cyclostationarity: Theory and application to gear fault monitoring. *Mechanical Systems and Signal Processing*, 22(3): 574–587, 2008.

[54] J. Antoni and P. Borghesani. A statistical methodology for the design of condition indicators. *Mechanical Systems and Signal Processing*, 114:290–327, 2019.

[55] K. C. Gryllias, H. Andre, Q. Leclere, and J. Antoni. Condition monitoring of rotating machinery under varying operating conditions based on cyclo-non-stationary indicators and a multi-order probabilistic approach for instantaneous angular speed tracking. *IFAC-PapersOnLine*, 50(1):4708–4713, 2017.

[56] J. Antoni. The spectral kurtosis: A useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006.

[57] W. J. Wang and P. D. McFadden. Early detection of gear failure by vibration analysis I. Calculation of the timefrequency distribution, 1993.

[58] W. J. Wang and P. D. McFadden. Early detection of gear failure by vibration analysis–ii. interpretation of the time-frequency distribution using image processing techniques, 1993.

[59] Z. K. Peng and F. L. Chu. Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mechanical Systems and Signal Processing*, 18(2):199–221, 2004.

[60] Y. Lei, J. Lin, Z. He, and M. J. Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing*, 35 (1-2):108–126, 2013.

[61] W. J. Staszewski and G. R. Tomlinson. Application of the wavelet transform to fault detection in a spur gear, 1994.

[62] M. J. Dowling. Application of non-stationary analysis to machinery monitoring. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 1, pages 59–62. IEEE, 1993.

[63] W. J. Wang and P. D. McFadden. Application of orthogonal wavelets to early gear damage detection. *Mechanical Systems and Signal Processing*, 9:497–507, 1995.

[64] J.-D. Wu and J. J. Chan. Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network. *Expert Systems with Applications*, 36 (5):8862–8875, 2009.

[65] Ł. Jedliński and J. Jonak. Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. *Applied Soft Computing*, 30:636–641, 2015.

[66] W. J. Wang and P. D. McFadden. Application of wavelets to gearbox vibration signals for fault detection. *Journal of Sound and Vibration*, 192(5):927–939, 1996.

[67] 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.

[68] W. Wang. Early detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15(5):887–903, 2001.

[69] J. Rafiee, M. A. Rafiee, and P. W. Tse. Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, 37(6): 4568–4579, 2010.

[70] X. Y. Wang, V. Makis, and M. Yang. A wavelet approach to fault diagnosis of a gearbox under varying load conditions. *Journal of Sound and Vibration*, 329(9):1570–1585, 2010.

[71] S. Theodoridis and K. Koutroumbas. *Pattern recognition*. Elsevier/Academic Press, 2009.

[72] J.-D. Wu and C.-H. Liu. Investigation of engine fault diagnosis using discrete wavelet transform and neural network. *Expert Systems with Applications*, 35(3):1200–1213, 2008.

[73] V. Purushotham, S. Narayanan, and S. A. N. Prasad. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT & E International*, 38(8):654–664, 2005.

[74] G. G. Yen and K.-C. Lin. Wavelet packet feature extraction for vibration monitoring. *IEEE Transactions on Industrial Electronics*, 47(3):650–667, 2000.

[75] J.-D. Wu and C.-H. Liu. An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. *Expert Systems with Applications*, 36(3):4278–4286, 2009.

[76] Y. Shao, D. Su, A. Al-habaibeh, and W. Yu. A new fault diagnosis algorithm for helical gears rotating at low speed using an optical encoder. *Measurement*, 93:449–459, 2016.

[77] X. Wan, D. Wang, P. W. Tse, G. Xu, and Q. Zhang. A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions. *Measurement*, 78:138–150, 2016.

[78] J. Dybała and R. Zimroz. Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal. *Applied Acoustics*, 77:195–203, 2014.

[79] Y. Lei, Z. He, and Y. Zi. Application of an intelligent classification method to mechanical fault diagnosis. *Expert Systems with Applications*, 36(6):9941–9948, 2009.

[80] S. J. Loutridis. Damage detection in gear systems using empirical mode decomposition. *Engineering Structures*, 26(12):1833–1841, 2004.

[81] R. Ricci and P. Pennacchi. Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions. *Mechanical Systems and Signal Processing*, 25(3): 821–838, 2011.

[82] J. Singh, A. K. Darpe, and S. P. Singh. Bearing damage assessment using Jensen-Rényi Divergence based on EEMD. *Mechanical Systems and Signal Processing*, 87:307–339, 2017.

[83] C. Y. Yang and T. Y. Wu. Diagnostics of gear deterioration using EEMD approach and PCA process. *Measurement*, 61:75–87, 2015.

[84] J. Antoni. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21:597–630, 2007.

[85] P. D. McFadden. Low frequency vibration generated by gear tooth impacts. *NDT International*, 18(5):279–282, 1985.

[86] W. Wang and A. K. Wong. Autoregressive model-based gear fault diagnosis. *Journal of Vibration and Acoustics*, 124(2):172, 2002.

[87] J. Antoni. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, 304(3-5):497–529, 2007.

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

[89] R. B. Randall. Detection and diagnosis of incipient bearing failure in helicopter gearboxes. *Engineering Failure Analysis*, 11(2):177–190, 2004.

[90] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, and A. Alonso-Betanzos. Automatic bearing fault diagnosis based on one-class ν-SVM. *Computers & Industrial Engineering*, 64(1):357–365, 2013.

[91] P. D. McFadden. Detection of gear faults by decomposition of matched differences of vibration signals. *Mechanical Systems and Signal Processing*, 14(5):805–817, 2000.

[92] P. Borghesani, R. Ricci, S. Chatterton, and P. Pennacchi. A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mechanical Systems and Signal Processing*, 38(1):23–35, 2013.

[93] D. Ho and R. B. Randall. Optimization of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, 14(5):763–788, 2000.

[94] J. Antoni, G. Xin, and N. Hamzaoui. Fast computation of the spectral correlation. *Mechanical Systems and Signal Processing*, 92:248–277, 2017.

[95] P. Borghesani and J. Antoni. A faster algorithm for the calculation of the fast spectral correlation. *Mechanical Systems and Signal Processing*, 111:113–118, 2018.

[96] C. Capdessus, M. Sidahmed, and J.L. Lacoume. Cyclostationary processes: Application in gear faults early diagnosis. *Mechanical Systems and Signal Processing*, 14(3):371–385, 2000.

[97] D. Abboud, J. Antoni, M. Eltabach, and S. Sieg-zieba. Angle\time cyclostationarity for the analysis of rolling element bearing vibrations. *Measurement*, 75:29–39, 2015.

[98] D. Abboud and J. Antoni. Order-frequency analysis of machine signals. *Mechanical Systems and Signal Processing*, 87(October 2016):229–258, 2017.

[99] N. Baydar and A. Ball. Detection of gear deterioration under varying load conditions By using the instantaneous power spectrum. *Mechanical Systems and Signal Processing*, 14(6):907–921, 2000.

[100] J. Urbanek, T. Barszcz, and J. Antoni. Time-frequency approach to extraction of selected second-order cyclostationary vibration components for varying operational conditions. *Measurement*, 46(4):1454–1463, 2013.

[101] J. Antoni and R. B. Randall. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2):308–331, 2006.

[102] J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007.

[103] F. Combet and L. Gelman. Optimal filtering of gear signals for early damage detection based on the spectral kurtosis. *Mechanical Systems and Signal Processing*, 23(3):652–668, 2009.

[104] T. Barszcz and R. B. Randall. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. *Mechanical Systems and Signal Processing*, 23(4):1352–1365, 2009.

[105] J. Antoni. The infogram: Entropic evidence of the signature of repetitive transients. *Mechanical Systems and Signal Processing*, 74:73–94, 2016.

[106] T. Barszcz and A. Jabłoński. A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. *Mechanical Systems and Signal Processing*, 25(1):431–451, 2011.

[107] Y. Lei, J. Lin, Z. He, and Y. Zi. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 25(5):1738–1749, 2011.

[108] D. Wang, P. W. Tse, and K. L. Tsui. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 35 (1-2):176–199, 2013.

[109] W. A. Smith, Z. Fan, Z. Peng, H. Li, and R. B. Randall. Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference. *Mechanical Systems and Signal Processing*, 75:371–394, 2015.

[110] X. Xu, Z. Qiao, and Y. Lei. Repetitive transient extraction for machinery fault diagnosis using multiscale fractional order entropy infogram. *Mechanical Systems and Signal Processing*, 103:312–326, 2018.

[111] D. Wang. An extension of the infograms to novel Bayesian inference for bearing fault feature identification. *Mechanical Systems and Signal Processing*, 80:19–30, 2016.

[112] J. Obuchowski, A. Wylomanska, and R. Zimroz. Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems and Signal Processing*, 48(1-2):138–152, 2014.

[113] X. Xu, M. Zhao, J. Lin, and Y. Lei. Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis. *Measurement*, 91:385–397, 2016.

[114] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

[115] C. M. Bishop. *Neural networks for pattern recognition*. Clarendon Press, 1995.

[116] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence*, 16:657–665, 2003.

[117] Q. Miao and V. Makis. Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models. *Mechanical Systems and Signal Processing*, 21(2):840–855, 2007.

[118] H. Ocak and K. A. Loparo. HMM-based fault detection and diagnosis scheme for rolling element bearings. *Journal of Vibration and Acoustics*, 127(4):299, 2005.

[119] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, 89:16–27, 2015.

[120] X. Qi, Z. Yuan, and X. Han. Diagnosis of misalignment faults by tacholess order tracking analysis and RBF networks. *Neurocomputing*, 169:439–448, 2015.

[121] W. Yang, P.J. Tavner, C.J. Crabtree, and M. Wilkinson. Cost-effective condition monitoring for wind turbines. *IEEE Transactions on Industrial Electronics*, 57(1):263–271, 2010.

[122] Y. Lei and M. J. Zuo. Gear crack level identification based on weighted K nearest neighbor classification algorithm. *Mechanical Systems and Signal Processing*, 23(5): 1535–1547, 2009.

[123] M. Y. Asr, M. M. Ettefagh, R. Hassannejad, and S. N. Razavi. Diagnosis of combined faults in rotary machinery by non-naive Bayesian approach. *Mechanical Systems and Signal Processing*, 85:56–70, 2017.

[124] L. M. R. Baccarini, V. V. Rocha E Silva, B. R. De Menezes, and W. M. Caminhas. SVM practical industrial application for mechanical faults diagnostic. *Expert Systems with Applications*, 38(6):6980–6984, 2011.

[125] M. Unal, M. Onat, M. Demetgul, and H. Kucuk. Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network. *Measurement*, 58:187–196, 2014.

[126] T. Marwala, U. Mahola, and F. V. Nelwamondo. Hidden Markov models and Gaussian mixture models for bearing fault detection using fractals. *International Joint Conference on Neural Networks, Vancouver, Canada*, pages 3237–3242, 2006.

[127] R. Casimir, E. Boutleux, G. Clerc, and A. Yahoui. The use of features selection and nearest neighbors rule for faults diagnostic in induction motors. *Engineering Applications of Artificial Intelligence*, 19(2):169–177, 2006.

[128] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 38(3):2888–2911, 2011.

[129] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1–41, 2009.

[130] R. Zimroz and A. Bartkowiak. Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions. *Mechanical Systems and Signal Processing*, 38(1):237–247, 2013.

[131] A. Bartkowiak and R. Zimroz. Dimensionality reduction via variables selection - Linear and nonlinear approaches with application to vibration-based condition monitoring of planetary gearbox. *Applied Acoustics*, 77:169–177, 2014.

[132] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[133] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72-73:303–315, 2016.

[134] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing*, 168:119–127, 2015.

[135] C. Li, R. V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing*, 76-77:283–293, 2016.

[136] S. Haidong, J. Hongkai, Z. Huiwei, and W. Fuan. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 95:187–204, 2017.

[137] L. Jing, M. Zhao, P. Li, and X. Xu. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, 111:1–10, 2017.

[138] Z. Wang, J. Wang, and Y. Wang. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing*, 310:213–222, 2018.

[139] S. Khan and T. Yairi. A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing*, 107:241–265, 2018.

[140] K. C. Gryllias and I. A. Antoniadis. A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Engineering Applications of Artificial Intelligence*, 25(2):326–344, 2012.

[141] C. Sobie, C. Freitas, and M. Nicolai. Simulation-driven machine learning: Bearing fault classification. *Mechanical Systems and Signal Processing*, 99:403–419, 2018.

[142] D. M. J. Tax and R. P. W. Duin. Growing a multi-class classifier with a reject option. *Pattern Recognition Letters*, 29(10):1565–1570, 2008.

[143] W. J. Scheirer, A. De Rezende Rocha, A. Sapkota, and T. E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7): 1757–1772, 2013.

[144] W. J. Scheirer, L. P. Jain, and T. E. Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2317–2324, 2014.

[145] M. D. Scherreik and B. D. Rigling. Open set recognition for automatic target classification with rejection. *IEEE Transactions on Aerospace and Electronic Systems*, 52(2): 632–642, 2016.

[146] M. F. Pimentel, D. Clifton, L. Clifton, and L. Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.

[147] M. Markou and S. Singh. Novelty detection: A review - Part 1: Statistical approaches. *Signal Processing*, 83(12):2481–2497, 2003.

[148] M. Markou and S. Singh. Novelty detection: A review - Part 2: Neural network based approaches. *Signal Processing*, 83(12):2499–2521, 2003.

[149] M. Timusk, M. Lipsett, and C. K. Mechefske. Fault detection using transient machine signals. *Mechanical Systems and Signal Processing*, 22(7):1724–1749, 2008.

[150] L. Xie, J. Zeng, U. Kruger, X. Wang, and J. Geluk. Fault detection in dynamic systems using the Kullback Leibler divergence. *Control Engineering Practice*, 43:39–48, 2015.

[151] F. Harrou, Y. Sun, and M. Madakyaru. Kullback-Leibler distance-based enhanced detection of incipient anomalies. *Journal of Loss Prevention in the Process Industries*, 44:73–87, 2016.

[152] E. Figueiredo, L. Radu, K. Worden, and C. R. Farrar. A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability. *Engineering Structures*, 80:1–10, 2014.

[153] T. Heyns, S. J. Godsill, J. P. De Villiers, and P. S. Heyns. Statistical gear health analysis which is robust to fluctuating loads and operating speeds. *Mechanical Systems and Signal Processing*, 27(1):651–666, 2012.

[154] T. Heyns, P. S. Heyns, and R. Zimroz. Combining discrepancy analysis with sensorless signal resampling for condition monitoring of rotating machines under fluctuating operations. *Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, 2(2):52–58, 2012.

[155] T. Heyns, P. S. Heyns, and J. P. De Villiers. Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. *Mechanical Systems and Signal Processing*, 32:200–215, 2012.

[156] P. S. Heyns, R. Vinson, and T. Heyns. Rotating machine diagnosis using smart feature selection under non-stationary operating conditions. *Insight-Non-Destructive Testing and Condition Monitoring*, 58(8):1–8, 2016.

[157] S. Schmidt, P.S. Heyns, and J.P. de Villiers. A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques. *Mechanical Systems and Signal Processing*, 100:152–166, 2018.

[158] S. Liu, M. Yamada, N. Collier, and M. Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, 2013.

[159] A. C. Bittencourt, K. Saarinen, S. Sander-Tavallaey, S. Gunnarsson, and M. Norrlöf. A data-driven approach to diagnostics of repetitive processes in the distribution domain - Applications to gearbox diagnostics in industrial robots and rotating machines. In *Mechatronics*, volume 24, pages 1032–1041, 2014.

[160] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti, and S. Longhi. Electric motor defects diagnosis based on kernel density estimation and Kullback Leibler divergence in quality control scenario. *Engineering Applications of Artificial Intelligence*, 44:25–32, 2015.

[161] J. Zeng, U. Kruger, J. Geluk, X. Wang, and L. Xie. Detecting abnormal situations using the Kullback Leibler divergence . *Automatica*, 50(11):2777–2786, 2014.

[162] Eric J., Travis O., Pearu P., et al. SciPy: Open source scientific tools for Python, 2001–. URL http://www.scipy.org/. [Online; accessed 2017-04-20].

[163] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC Texts in Statistical Science, 3th edition, 2013.

[164] B. P. Carlin and T. A. Louis. *Bayesian methods for data analysis*. CRC Press, 2008.

[165] K. P. Murphy. *Conjugate Bayesian analysis of the Gaussian distribution*, 2007.

[166] P.W. Goldberg, C.K.I. Williams, and C.M. Bishop. Regression with input-dependent noise: A gaussian process treatment. *Advances in Neural Information Processing Systems*, 10:493–499, 1997.

[167] C. M. Bishop and C. S. Qazaz. Regression with input-dependent noise: A Bayesian treatment. *Advances in Neural Information Processing Systems 9*, 9(1991):347–353, 1997.

[168] J. Lee, H. Qiu, J. Yu, and Rexnord Technical Services. IMS, University of Cincinnati. "Bearing Data Set", NASA Ames Prognostics Data. URL https://ti.arc.nasa.gov/c/3/.

[169] 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*, 289:1066–1090, 2006.

[170] P. W. Tse and D. Wang. The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "two automatic vibration-based fault diagnostic methods using the novel sparsity measurement - Parts 1 and 2". *Mechanical Systems and Signal Processing*, 40(2):499–519, 2013.

[171] F. Jia, Y. Lei, N. Lu, and S. Xing. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mechanical Systems and Signal Processing*, 110:349–367, 2018.

[172] R. Shao, W. Hu, Y. Wang, and X. Qi. The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform. *Measurement*, 54:118–132, 2014.

[173] A. E. Lazzaretti, D. M. J. Tax, H. Vieira Neto, and V. H. Ferreira. Novelty detection

and multi-class classification in power distribution voltage waveforms. *Expert Systems with Applications*, 45:322–330, 2016.

[174] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Appendix A

Additional results and investigations

In this appendix, supporting information is given for the main content of the thesis. In Section A.1, a sensitivity analysis is performed to determine the effect the prior hyperparameters in Chapter 2 have on the inferred results. In Section A.2, additional information is given and additional investigations are performed for the bearing discrepancy analysis methodology presented in Chapter 3. Lastly, additional information is also provided in Section A.3 for the novel information enhancement methodology presented in Chapter 5.

A.1 Sensitivity analysis for prior parameters in Chapter 2

In Chapter 2, the hyperparameters of the prior distribution are selected a priori to the analysis and subsequently used to perform Bayesian inference. However, the hyperparameters of the prior distribution can have a significant impact on the posterior distribution and therefore the choice of the hyperparameters, that were used in the previous analyses, needs to be motivated. The hyperparameters were chosen so that the posterior marginal distribution on the mean is dominated by the data and not by the prior information. However, the choice may not be appropriate and therefore the sensitivity of the hyperparameter choice on the resulting posterior distribution is investigated in this section.

Hierarchical or multilevel Bayesian analysis can be used to incorporate uncertainty in the hyperparameters into the inference process, by using hyperpriors [164]. Hence, the full joint distribution over the random variables is as follows

$$p(\mu, \lambda, \mu_0, \kappa_0, \alpha_0, \beta_0 | X, \boldsymbol{\phi}) = p(\mu, \lambda | \mu_0, \kappa_0, \alpha_0, \beta_0, X) p(\mu_0, \kappa_0, \alpha_0, \beta_0 | \boldsymbol{\phi}), \tag{A.1}$$

where $p(\mu, \lambda | \mu_0, \kappa_0, \alpha_0, \beta_0, X)$ is the posterior distribution if the hyperparameters are fixed,

$p(\mu_0, \kappa_0, \alpha_0, \beta_0|\boldsymbol{\phi})$ is the distribution over the prior hyperparameters, known as the hyperprior, and $\boldsymbol{\phi}$ is the hyperprior parameters. The joint distribution over the posterior mean and precision is subsequently calculated as

$$p(\mu, \lambda|X, \boldsymbol{\phi}) = \int_{\alpha_0} \int_{\beta_0} \int_{\mu_0} \int_{\kappa_0} p(\mu, \lambda, \alpha_0, \beta_0, \mu_0, \kappa_0|X, \boldsymbol{\phi}) \; d\alpha_0 \; d\beta_0 \; d\mu_0 \; d\kappa_0, \tag{A.2}$$

which allows the uncertainty in the hyperparameters to be reflected in the posterior distribution parameters which are subsequently used in the condition inference process. The model with the hyperpriors in Equation (A.1) is called a three-level model and the approach in Section 2.2.3, with the fixed hyperparameters, is known as a two-level model [164]. Sampling methods are investigated to allow arbitrary hyperpriors to be investigated. A sample s from the hyperprior distribution

$$\mu_0^s, \kappa_0^s, \alpha_0^s, \beta_0^s \sim p(\mu_0, \kappa_0, \alpha_0, \beta_0|\boldsymbol{\phi}), \tag{A.3}$$

is used to obtain the posterior Gaussian-Gamma distribution parameters using the process described in Section 2.2.3. Note that the superscript s denotes that a sample of the variable is obtained or considered. The posterior precision sample is obtained from

$$\lambda^s \sim \text{Gamma}(\lambda|\alpha_n^s, \beta_n^s), \tag{A.4}$$

and the posterior mean sample is obtained from

$$\mu^s \sim \text{Gaussian}(\mu|\mu_n^s, (\kappa_n^s \lambda^s)^{-1}), \tag{A.5}$$

which is used for subsequent inference procedures. The sampling process is repeated multiple times to obtain a sufficient number of samples, which is used to represent the posterior distribution over the parameters of interest.

In this chapter, the hyperparameters are assumed to be uncorrelated and uniformly distributed between a and b, which is denoted by $U[a, b]$. The true hyperparameter values are unknown and therefore all values within a sensible range are assumed to be equally probable. The hyperparameters for the processed divergence data $\boldsymbol{d}_i^{proc}$ are sampled from, $\kappa_0 \sim U[10^{-4}, n/5]$, $\mu_0 \sim U[\min(\boldsymbol{d}_i^{proc}), \max(\boldsymbol{d}_i^{proc})]$, $\alpha_0 \sim U[10^{-4}, 20]$, $\beta_0 \sim U[10^{-4}, 20]$, which is subsequently used to obtain the posterior parameters. The number of elements in

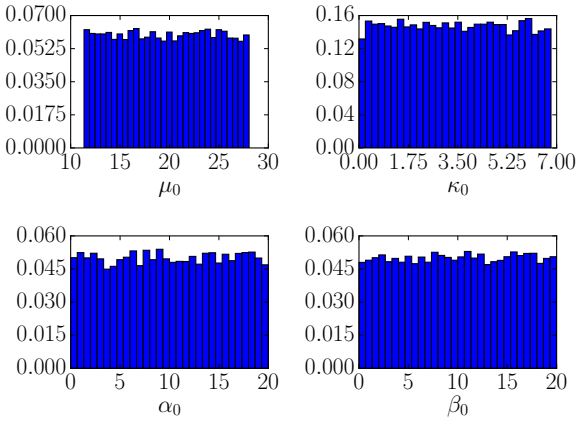


Figure A.1: Probability density function constructed from 20,000 samples taken from the hyperprior.

$\boldsymbol{d}_i^{proc}$ is n and $\min(\boldsymbol{d}_i^{proc})$ and $\max(\boldsymbol{d}_i^{proc})$ denotes the minimum and maximum value of $\boldsymbol{d}_i^{proc}$ respectively.

The samples from the hyperprior is shown in Figure A.1. The posterior distribution from this three-level approach is compared to the two-level approach in Figure A.2. The hyperparameters of the two-level approach are the same as in Section 2.3.2 for the divergence data of a specific window. The posterior marginal distribution on the mean is very similar

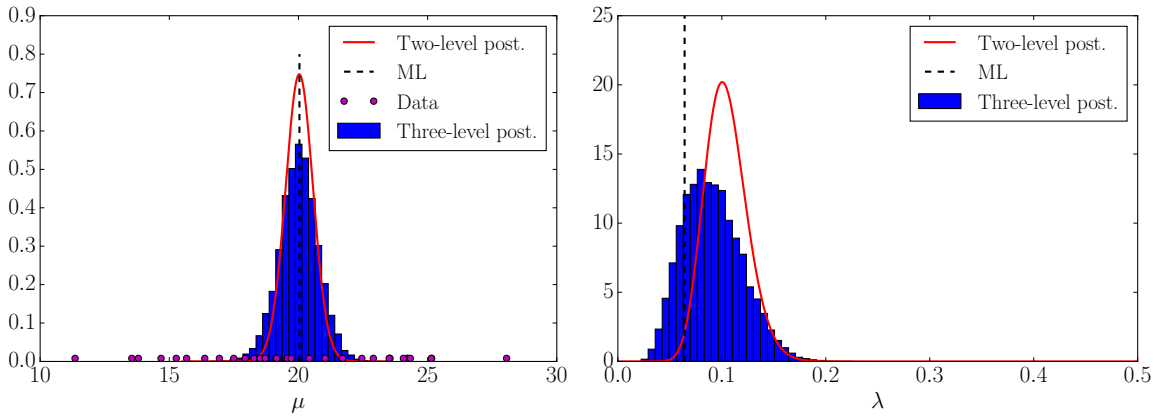


Figure A.2: The probability density function constructed from 20,000 samples, taken from the hyperprior of the three-level approach, is compared to the two-level approach and the maximum likelihood (ML) estimate.

for the two approaches, with the probability density function of the two-level approach being slightly more localised. This means that the uncertainty in the posterior mean is larger for the three-level approach due to the additional uncertainty in the hyperparameters. The precision is more sensitive to the choice of hyperparameters, and therefore a broader probability density function is seen for the three-level approach. The mode of the three-level precision posterior is closer to the maximum likelihood data than the two-level approach. The posterior marginal distribution of the mean is used for condition inference and is very similar for the two-level

and the three-level approaches, which validates that the choice of hyperparameters in the previous analyses is sensible.

As a result, the three-level approach was not used for the other investigations in Chapter 2, because similar inferences will be made if Equation (2.13) is used, but with a significant increase in computational time.

A.2 Bearing discrepancy analysis methodology information

Additional information is given in this section for the bearing discrepancy analysis methodology in Chapter 3. In Section A.2.1, supporting information is given for the derivation of the rotational speed compensation algorithm presented in Section 3.2.5 and in Section A.2.1.1 the window length optimisation procedure, used in the rotational speed compensation algorithm, is presented and discussed. It is also possible to use different features in the methodology as stated throughout Chapter 3. Therefore, in Section A.2.2 the suitability of different features and in Section A.2.3 the suitability of the different models are compared for discrepancy analysis.

A.2.1 Compensating for rotational speed variations

In this section, the procedure that is used in Section 3.2.5 to estimate the mean and the variance of the conditional distribution $p(\eta|\omega)$ is presented. It is assumed that the actual conditional density of the discrepancy measure η given the rotational speed ω, $p(\eta|\omega)$, is unknown and samples cannot be easily obtained from it. The process is started by considering the analytical form of the expected value $\mathbb{E}[\eta|\omega]$

$$\mathbb{E}[\eta|\omega] = \int \eta \cdot p(\eta|\omega) d\eta. \tag{A.6}$$

If an infinite number of samples is taken from the conditional distribution $p(\eta|\omega)$, then it can be used to calculate the expected value of η given ω

$$\mathbb{E}[\eta|\omega] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \eta_i|\omega, \tag{A.7}$$

where $\eta_i|\omega$ is a sample from $p(\eta|\omega)$. A different form of Equation (A.7) is analysed because it is assumed that it is not possible to sample from $p(\eta|\omega)$ directly

$$\mathbb{E}\left[\eta|\omega\right] = \lim_{d\Omega\to 0}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\eta_i|\omega \;\in\; [\Omega - d\Omega/2, \Omega + d\Omega/2]\,, \tag{A.8}$$

where $\eta_i|\omega$ is a sample from the conditional distribution $p(\eta_i|\omega)$, where the rotational speed $\omega \in [\Omega - d\Omega/2, \Omega + d\Omega/2]$ and defining the variable Ω in the equation is useful for later. If it is assumed that the actual set of discrepancy measures for a healthy dataset $\{\eta_i\}$ and the corresponding set of rotational speeds $\{\omega_i\}$ are samples from the actual distribution $p(\eta, \omega)$, then it can be used to approximate $\mathbb{E}[\eta|\omega]$. Equation (A.8) is extended to accommodate the experimental data

$$\mathbb{E}\left[\eta|\omega_j\right] \approx \frac{1}{N_{\eta|\omega}}\sum_{i=1}^{N_{\eta|\omega}}\eta_i|\omega_i \;\in\; [\Omega - \Delta\Omega/2, \Omega + \Delta\Omega/2]\,, \tag{A.9}$$

with $\omega_j \in [\Omega - \Delta\Omega/2, \Omega + \Delta\Omega/2]$ and $\eta_i|\omega_i$ denoting the sample η_i for which the corresponding rotational speed ω_i falls within the range $\omega_i \in [\Omega - \Delta\Omega/2, \Omega + \Delta\Omega/2]$. There are $N_{\eta|\omega}$ samples in the set $\{\omega_i\}$ for which $\omega_i \in [\Omega - \Delta\Omega/2, \Omega + \Delta\Omega/2]$. Hence, it can be seen from Equation (A.8) that the smaller the window length $\Delta\Omega$ is and the more samples $N_{\eta|\omega}$ are used, the better the approximation in Equation (A.9) becomes. However, because the actual data are considered, there is a compromise between the selected window length and the number of samples that can possibly be used.

Ultimately, this procedure is implemented by generating a grid $\{\Omega_i\}$ of window centres and preselecting the window width $\Delta\Omega$. For each window centre j a single value of $\mathbb{E}[\eta|\Omega_j]$ is obtained. Instead of assuming that $\mathbb{E}[\eta|\omega]$ is constant for a given window centre and within the bounds, it is assumed that $\mathbb{E}[\eta|\omega]$ varies linearly between window centres. Hence, the value of $\mathbb{E}[\eta|\omega]$ for a specific ω is obtained by linear interpolation by using the window centres $\{\Omega_i\}$ and the corresponding set $\{\mathbb{E}[\eta|\Omega_i]\}$. The notation $\mu_{\eta|\omega}(\omega) = \mathbb{E}[\eta|\omega]$ is used in Chapter 3. The approach is easily extended to calculate the variance of the discrepancy measure given the operating conditions.

A.2.1.1 Determining the window length of the speed compensation procedure

The following procedure is followed to calculate the window length used in the rotational speed compensation procedure in Section 3.2.5. A grid of candidate window lengths is generated

and then the following process is followed for each candidate window length on the grid and a preselected window overlap:

1. Calculate the set of window centres $\{\Omega_i\}$, the conditional mean $\{\mu_{\Omega_i,\Delta\Omega}\}$ and the variance $\{\sigma^2_{\Omega_i,\Delta\Omega}\}$ based on the current window length and the window overlap. The position of the window centre k is determined from

$$\Omega_k = \frac{1}{2}\Delta\Omega + k\left(\Delta\Omega - \Omega_{overlap}\right), \tag{A.10}$$

   for $k = 0, 1, \ldots$, where the window overlap is denoted by $\Omega_{overlap}$ and the window overlap fraction is given by $\Omega_{overlap}/\Delta\Omega$.

2. Discretize the rotational speed axis with a fixed resolution $\Delta\omega$, by starting at the minimum rotational speed ω_{min}. This discretization process is used in the next step to evaluate the suitability of the current window length. A relatively fine resolution of $\Delta\omega = 0.1$ rad/s is used for all future investigations in this chapter.

3. Evaluate the following cost function

$$C = \sum_{k=1}^{N-1}\left(\mu_{\eta|\omega}(\omega_{min} + k\Delta\omega) - \mu_{\eta|\omega}(\omega_{min} + (k-1)\Delta\omega)\right)^2, \tag{A.11}$$

   where N is the number of grid-points that were generated in the previous step. The function in Equation (A.11) gives an indication on how smooth $\mu_{\eta|\omega}(\omega)$ is.

4. Change the window length $\Delta\Omega$ and then repeat the process until all grid points are evaluated.

The cost function in Equation (A.11) was evaluated for the data presented in Figure 3.5 for a grid of window lengths and three window overlap fractions i.e. $\Omega_{overlap}/\Delta\Omega$. The window length which is selected from the process, is indicated by the ideal label in Figure A.3 and is not the window length that corresponds to the smallest value of the cost function. If the window length is too long, $\mu_{\eta|\omega}(\omega)$ will be too smooth to capture the localised trend within the data. This is presented in Figure A.4, where the results of different window lengths (indicated by the candidate lines in Figure A.3) are presented for the data shown in Figure 3.5. All of the curves $\mu_{\eta|\omega}(\omega)$ for the different window lengths $\Delta\Omega$ are shown over one another in Figure A.4(i). In Figure A.4(ii), each curve is given a different offset which helps to compare the different results that are obtained. It is concluded from the results that all

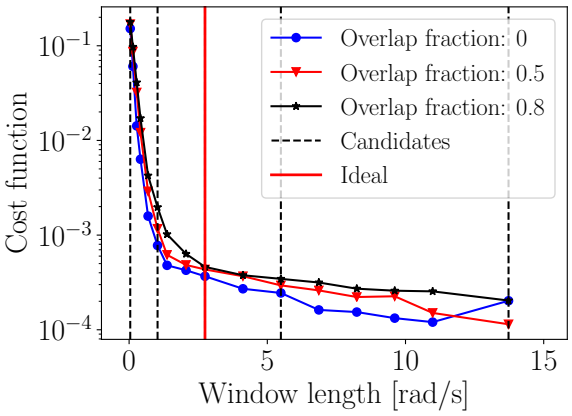
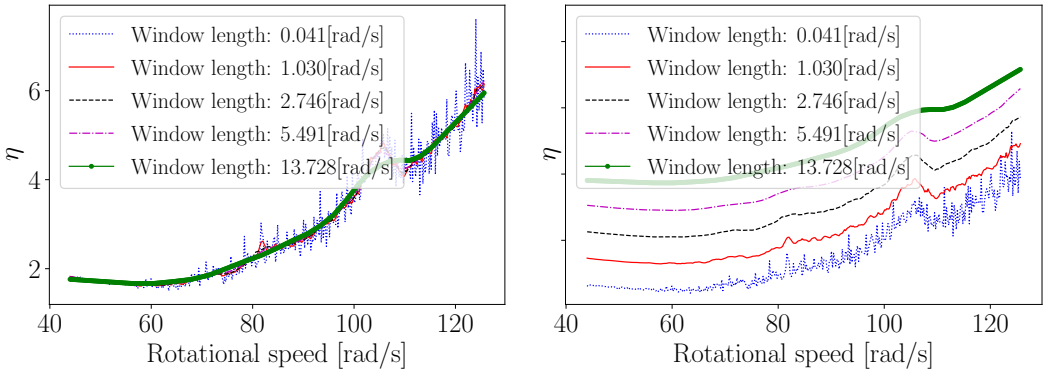


Figure A.3: The cost function in Equation (A.11) is evaluated for several window lengths, with the vertical candidate lines indicating the window lengths which are investigated in subsequent figures. The overlap fraction is obtained by $\Omega_{overlap}/\Delta\Omega$.



(i) (ii)

Figure A.4: Two plots of the function $\mu_{\eta|\omega}$ are shown for the candidate lines and the ideal line in Figure A.3. In Figure A.4(i) the fitted function is shown on the actual ordinate, and in Figure A.4(ii) a unique value is added to η for each dataset to separate the fitted functions for a simpler visual interpretation. An overlap fraction of 0.8 is used.

curves have the same trend, however if the window length is too small a very noisy function is obtained while a too large window length results in an overly smooth curve which does not fully capture the local characteristics of the data. A window length between 1.5 rad/s and 5.5 rad/s gives sufficiently good results for the data under consideration. The window fraction overlap of 0.8 was used for all investigations, because it provides smoother results without losing local information.

A.2.2 Comparing the suitability of different features

The purpose of this investigation is to compare the suitability of different features for discrepancy analysis. The sensitivity of the discrepancy signal for the presence of bearing damage and the robustness of the discrepancy signal to changes in operating conditions are compared for different features in this section. The same procedure for discrepancy analysis is

followed as Figure 3.1, except, different features are investigated and compared and instead of using the Mahalanobis distance as a discrepancy measure, the negative log-likelihood of the Gaussian model is used.

A.2.2.1 Features for discrepancy analysis

It is necessary to extract diagnostic information from the measured vibration signal to improve the condition inference capabilities of discrepancy analysis. In discrepancy analysis, the multi-dimensional features should represent information pertaining to the condition of the machine as a function of angle. Another important consideration of potential discrepancy features is that even though the modulation due to bearing damage is angle-dependent, the carrier component is time-invariant [31, 98]. The implication of this is that the signal is approximately angle-time cyclostationary as opposed to time- or angle-cyclostationary under varying speed conditions [31, 98] and therefore the feature extraction methods should be carefully designed. The Wavelet Packet Transform (WPT), angle-order Spectrogram, Angle-Frequency Instantaneous Power Spectrum (AF-IPS) and Empirical Mode Decomposition (EMD) are considered to perform feature extraction in this investigation.

Wavelet Packet Transform (WPT)

The Wavelet Packet Transform (WPT) forms part of wavelet analysis where a signal is recursively decomposed into approximation (i.e. low-frequency content) and detail (i.e. high-frequency content) coefficients by using a filter bank of quadrature mirror filters [74]. A two-level wavelet packet decomposition of the temporal signal $x(m)$ is shown in Figure A.5 with the Wavelet Coefficients (WCs) of each level presented as well. The impulse response

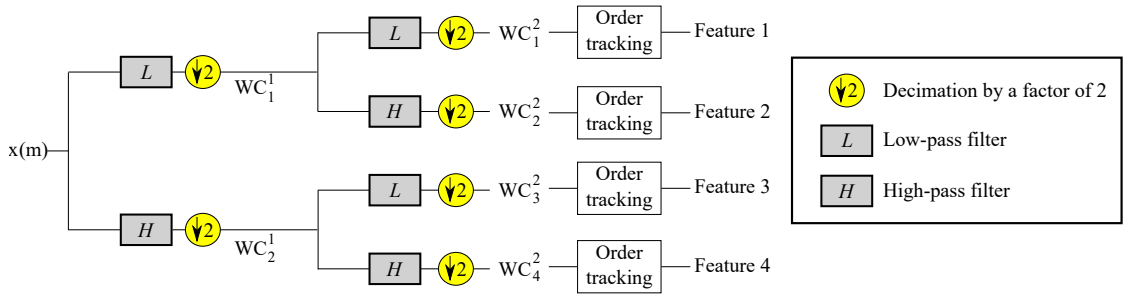


Figure A.5: Two-level wavelet packet decomposition of a discrete temporal signal $x(m)$ with a filterbank of low-pass (L) and high-pass (H) quadrature mirror filters for the WPT features is presented. $WC_i^j(n)$ indicates the ith coefficients of the jth level.

function of the low-pass filter h_L and of the high-pass filter h_H are determined by the specific wavelet basis function that is used [74]. For example, the filters are used to obtain filtered

versions of the original signal $x(m)$

$$x_L(m) = x(m) \otimes h_L, \tag{A.12}$$

$$x_H(m) = x(m) \otimes h_H, \tag{A.13}$$

where x_L and x_H denote the low-pass and high-pass filtered signals. The low-pass and high-pass filtered signals are decimated with a factor of two to obtain the wavelet coefficients at the first level. This is performed repeatedly on the wavelet coefficients as shown in Figure A.5 to obtain a set of wavelet coefficients. The effectiveness of the WPT for diagnostics depends on the level of the decomposition as well as the wavelet function that is used and therefore a few common wavelet basis functions are investigated.

As outlined in Figure 3.3 and Figure A.5, the WPT is performed on the temporal vibration signal to decompose the signal into a set of narrowband WCs. The WCs of the final level are order tracked to obtain an angle domain representation of the specific WCs which are subsequently used as features in the discrepancy analysis process. The resulting features are given by

$$\boldsymbol{b}_n = \left[WC_1^N(n), WC_2^N(n), \ldots, WC_{2^N}^N(n)\right]^T, \tag{A.14}$$

where $WC_i^N(n)$ denotes the n-sample of the ith order tracked wavelet coefficients at level N and the superscript T indicates the vector is transposed.

In the feature extraction procedure used in Chapter 3, the WCs are windowed to extract localised statistics as features to enhance the discrepancy signal. However, this is not performed on the wavelet packet features, because this can be applied to all investigated features and does not convey their actual ability to extract diagnostic information.

Angle-order Spectrogram

The spectrogram is the squared magnitude of the Short-Time Fourier Transform (STFT) and can show how the time-frequency components of the vibration signal evolve over time. However, there is a trade-off between the time and frequency resolution of the STFT and therefore the STFT is ill-suited for representing the time-frequency information of non-stationary signals in general. Despite its limitations, some diagnostic information may still be contained within the spectrogram and it has the benefit that it is relatively computationally efficient to calculate due to the Fast Fourier Transform (FFT) being an essential part of the calculation process. The fact that the signal has a relatively poor time (or angle) resolution is not

problematic for discrepancy analysis where the sampling frequency of the signal is usually reduced by windowing the signal [155, 156]. The feature extraction procedure is as follows:

- Order track the vibration signal.
- Calculate the STFT of the order tracked vibration signal.
- Calculate the spectrogram, as the squared magnitude of the STFT, and use as features.

The number of frequency bins is equal to the number of features that is extracted. The discrete spectrogram of the order-tracked vibration signal $x(m)$ is given by

$$S(l,k) = \left| \sum_{m=0}^{L_w-1} x(m + R_w l) w(m) e^{-2j\pi nk/L_w} \right|^2 , \tag{A.15}$$

and calculated for a windowing function $w(m)$ of length L_w and an overlap between consecutive windows O_w, where $R_w = L_w - O_w$. The D-dimensional features extracted from the spectrogram are denoted by

$$\boldsymbol{b}_n = [S(n,0), S(n,1), \ldots, S(n,L_w-1)]^T , \tag{A.16}$$

where n relates to the shaft angle with $n\Delta\varphi/2$ rad, where $\Delta\varphi$ is the angular resolution of the spectrogram in radians. Due to the fact that the angle-time cyclostationary properties of the bearing damage are not preserved in the calculation of the features, it is expected that the features will be sensitive to varying speeds, which is non-ideal under varying speed conditions.

Instantaneous Power Spectrum (IPS)

The interactions of rotating components (e.g. gear meshing, impacts due to bearing damage) of machines operating under stationary conditions result in the generation of time-cyclostationary signals i.e. stochastic vibration signals with periodic statistical properties in terms of time. This makes cyclostationary analysis very powerful for performing rotating machine diagnostics [27]. However, for bearings operating under varying speed conditions, the time-cyclostationary properties of the signal are lost and the signals are rather cyclo-non-stationary. Abboud et al. [31, 98] showed that the signals from damaged bearings can be approximated as angle-time second-order cyclostationary under varying speed conditions and gave angle-time cyclostationary estimators for classical second-order cyclostationary tools

such as the spectral coherence and the IPS. The IPS is very powerful for rotating machine diagnostics [15, 99] and is used as discrepancy analysis features in this work. The time-frequency IPS (TF-IPS)

$$I_{TF}(t,f) = \int_{-\infty}^{\infty} \mathcal{R}_{2x}(t,\tau) \cdot e^{-2\pi\tau f} d\tau, \tag{A.17}$$

describes how the power of the signal manifest in its frequency bands over time where $\mathcal{R}_{2x}(t,\tau)$ denotes the time-autocorrelation function. The Welch-based estimator of the TF-IPS [98]

$$\hat{I}_{TF}(l,k) = \frac{1}{f_s \cdot L \sum_{a=0}^{N_w-1} w(a)^2} \left| \sum_{m=0}^{L_w-1} x(m + R_w \cdot l) w(m) e^{-2j\pi mk/L_w} \right|^2, \tag{A.18}$$

has a very similar form to Equation (A.15) and is used in this work with the same notation. Under varying speed conditions, the Angle-Frequency IPS (AF-IPS) is more appropriate and obtained by order-tracking each frequency band of the TF-IPS separately [98]

$$\hat{I}_{AF}(n,k) = \mathrm{OT}_{t\to\varphi}\left(\hat{I}_{TF}(l,k)\right), \tag{A.19}$$

where $\mathrm{OT}_{t\to\varphi}$ indicates that the time index l is converted to an angle n with Order Tracking (OT). The features at angle index n is obtained by concatenating the values of the AF-IPS at all frequency bands for angle increment n and given by

$$\boldsymbol{b}_n = \left[\hat{I}_{AF}(n,0), \hat{I}_{AF}(n,1), \ldots, \hat{I}_{AF}(n,L_w-1)\right]^T. \tag{A.20}$$

It is evident from Equation (A.16) and Equation (A.18) that the dimensionality of the feature space is governed by the length of the window that is used, while Equation (A.15) and Equation (A.17) indicate the time resolution is dependent on the length of the window as well as the overlap between consecutive windows. In all subsequent investigations the AF-IPS as opposed to the TF-IPS is used for feature extraction and is shortened to IPS in the figures and when it is being referred to.

Empirical Mode Decomposition (EMD)

The Empirical Mode Decomposition (EMD) is a powerful self-adaptive time-frequency analysis technique that decomposes a signal into a set of orthogonal Intrinsic Mode Functions (IMFs). The EMD is capable of analysing non-stationary signals in contrast to the STFT and

it does not rely on a priori assumptions on the form of the signal components compared to the WPT [60]. However, the EMD has a few problems such as the IMFs not being orthogonal and the occurrence of mode mixing, a phenomenon where a signal component manifest within a few IMFs. The ensemble EMD and other techniques have been proposed in literature aiming to circumvent the aforementioned problems of the EMD [60]. In this investigation, the standard EMD without any improvements is used in spite of the aforementioned problems, because this is an initial investigation of the EMD. The EMD is performed on the temporal vibration signal to extract N temporal IMFs. The temporal IMFs are order tracked to obtain the angular IMFs which are subsequently used as features in the discrepancy analysis process, with the features given by

$$\boldsymbol{b}_n = [\mathrm{IMF}_1(n), \ldots, \mathrm{IMF}_N(n)]^T . \tag{A.21}$$

In discrepancy analysis, localised features are converted to a discrepancy measure. This means that phenomena which result in spurious changes in features such as mode mixing will result in corresponding spurious discrepancies. These spurious discrepancies may dominate the discrepancy signal which will cause erroneous inferences to be made.

A.2.2.2 Dataset

The phenomenological gearbox model, proposed by Abboud et al. [28], and presented in Section 3.3.1, is used in this section as well. The impulse response functions of the deterministic gear component, the random gear component and the bearing damage component have the same damping ratio of 0.05 and natural frequencies of 1.5 kHz, 2.5 kHz and 5.0 kHz respectively. The ball-pass outer race order of the bearing is 8.14 shaft orders. Six bearing

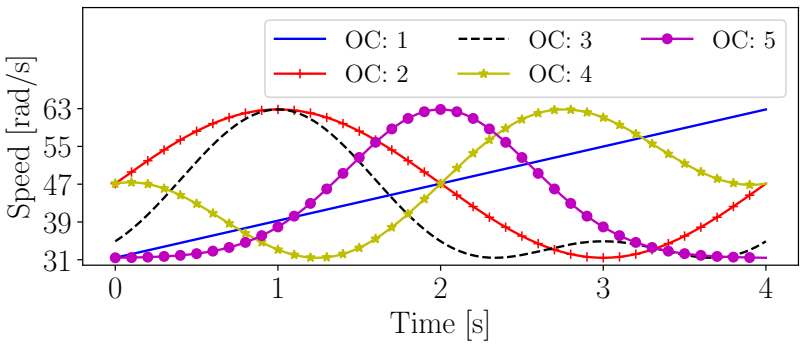


Figure A.6: Rotational speed profiles for the phenomenological gearbox model.

conditions are investigated with the model namely $FS = [0.0, 0.02, 0.04, 0.08, 0.12, 0.16]$.

A.2.2.3 Overview of investigation

The first step of the discrepancy analysis procedure used in this work is to extract diagnostic features from the vibration signal. The dimensionality of the features should be approximately the same to ensure that a fair comparison between the different features can be given. The dimensionality of the features in this investigation is set to 16 unless it is not practically possible. The WPT decomposition level is set to four, which results in 16 wavelet coefficients to be extracted from the vibration signal. Many wavelet basis functions can be used to perform the decomposition, with the focus placed on Daubechies wavelets which is very popular for bearing diagnostics. The following Daubechies wavelet basis functions are investigated in this section: db1, db2, db4, db8, db10, db16, db20, and db44.

The spectrogram features are calculated by using a Hann windowing function with a window length of 16 and an overlap of 75% between consecutive windows. The TF-IPS is implemented with the same characteristics as the spectrogram features, whereafter the TF-IPS is order tracked to obtain the AF-IPS. The window length is selected to have the same dimensionality as the WPT features.

The EMD is a self-adaptive signal processing technique to extract multiple IMFs which are used to obtain features as discussed in Section A.2.2.1. It was found that a different number of IMFs is extracted for different signals and to ensure that all measurements have the same dimensionality, the number of IMFs that is extracted is set to 12. If more IMFs are used in the training phase, there is a risk that some datasets in the testing phase may be unusable due to an insufficient number of IMFs being extracted.

The features of a system with a healthy bearing (i.e. FS = 0.0) for the first four operating conditions are used to obtain the healthy model as discussed in Chapter 3. The negative log-likelihood

$$\eta(n) = -\log \mathcal{N}(\boldsymbol{b}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{A.22}$$

is used to generate the discrepancy signals for the different datasets, whereafter the spectra of the discrepancy signals are calculated. The spectra are presented in the next section.

A.2.2.4 Sensitivity to damage

The spectra of the discrepancy signals are presented in Figure A.7, where the bearing damage component at 8.14 shaft orders and its harmonics are clearly seen for the AF-IPS, spectrogram and most WPT features. The spectra associated with the spectrogram features contain a

relatively large gear mesh component at 20 shaft orders when compared to the other features, while only the fundamental bearing component can clearly be seen in the spectra of the discrepancy signal associated with the WPT db44. In contrast to the other results, the EMD features do not contain any clear diagnostic information in its spectra. It is difficult to

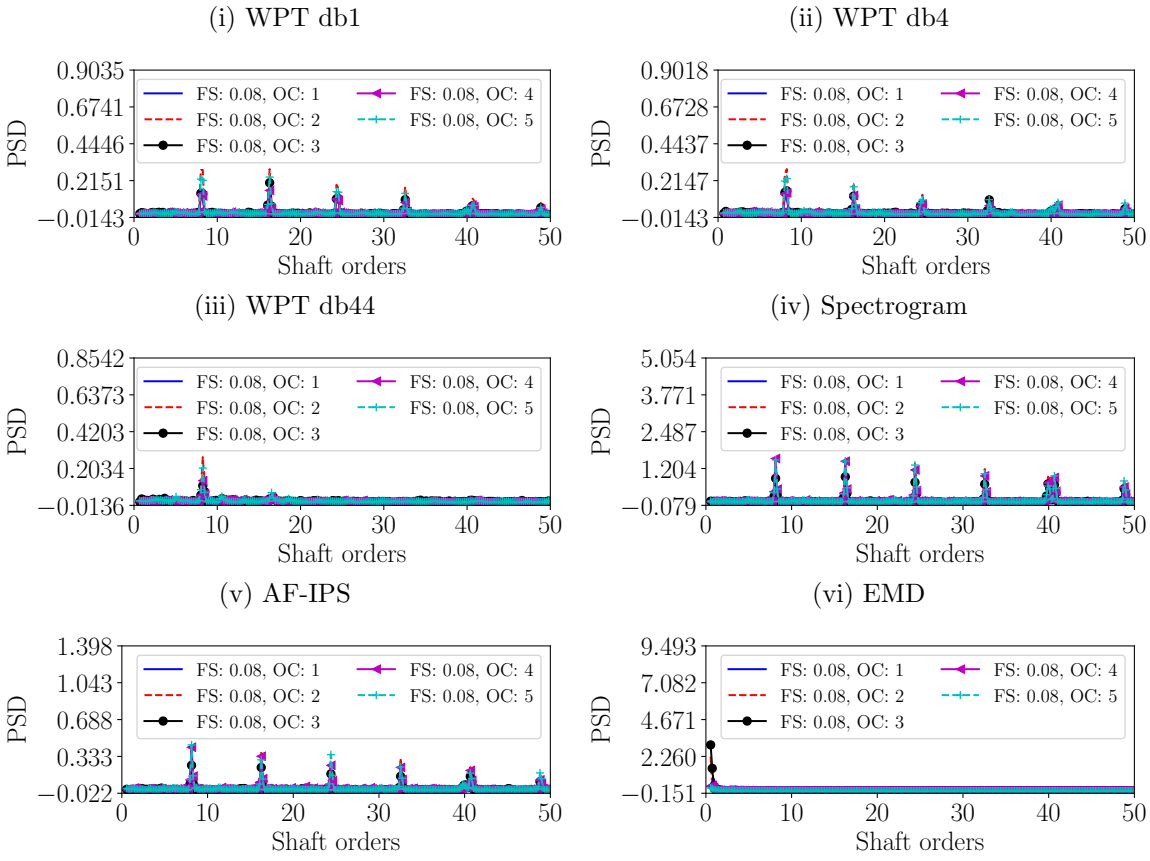


Figure A.7: Discrepancy spectra for the different features used for the phenomenological gearbox model with a fault severity of 0.08.

compare the performance of the features from the spectra in Figure A.7 and therefore their performance is quantified and compared next.

The amplitude of the fundamental ball-pass outer race component contains diagnostic information as seen in Figure A.7 and is investigated over fault severity to quantify the sensitivity of the features to damage. The raw amplitude is presented in Figure A.8(i) and the normalised amplitude is presented in Figure A.8(ii). The normalised amplitude, calculated by dividing the amplitude of the different fault severities by the amplitude of the healthy bearing component (FS = 0.0), is used to quantify the relative change of the amplitude over fault severity. Surprisingly, from the results in Figure A.8, the spectrogram features are the most sensitive to the damage, with the AF-IPS being second most sensitive. The discrepancy signal associated with the EMD features is the least sensitive to damage. This is attributed

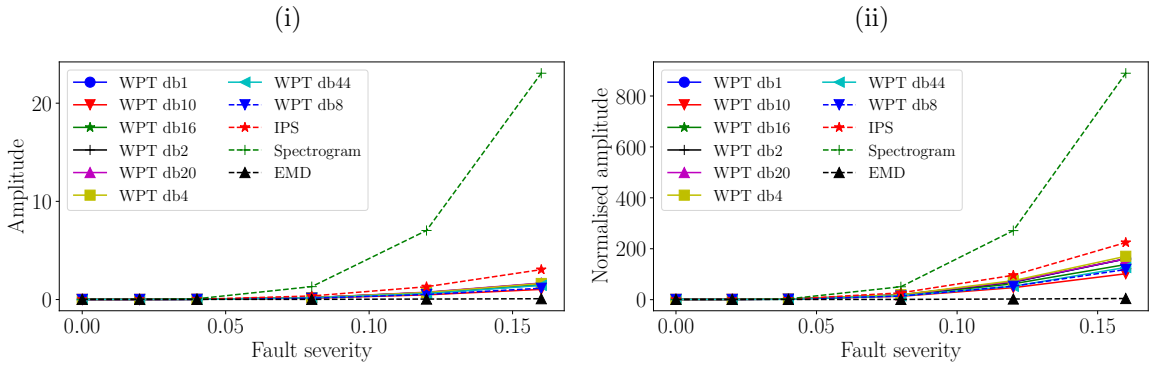


Figure A.8: The sensitivity of the amplitude of the fundamental bearing component versus fault severity.

to mode mixing, which results in spurious discrepancies to be detected.

It is recommended to use other performance metrics as well, to ensure that the performance of the features can be completely quantified.

A.2.2.5 Robustness to varying speed conditions

It is not only important to quantify the sensitivity of the features to damage, but also their robustness to varying speed conditions. The amplitude of the fundamental outer race component is presented over rotational speed in Figure A.9(i) and the normalised amplitude is presented in Figure A.9(ii). The normalised amplitude is obtained by dividing the mean am-

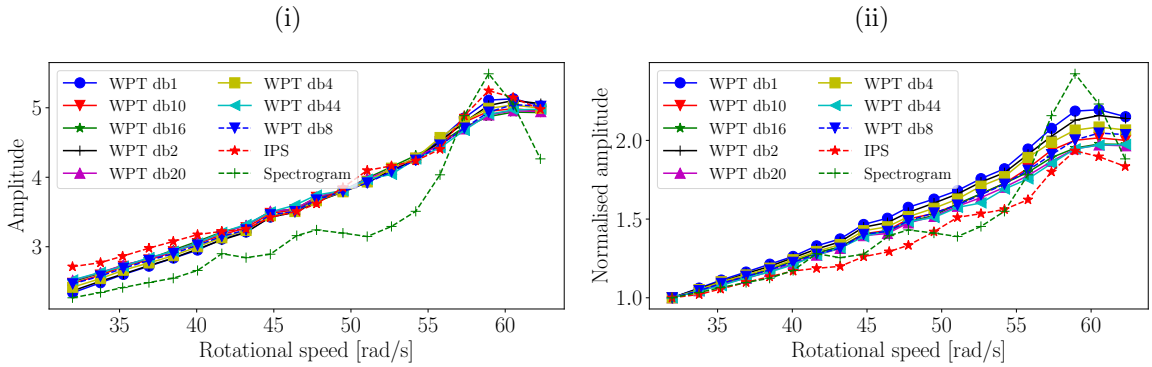
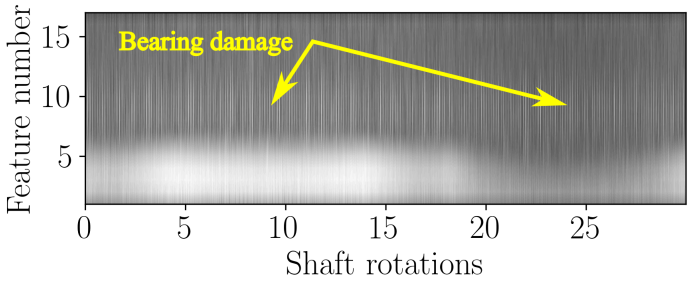


Figure A.9: The sensitivity of the amplitude of the fundamental bearing component versus fault severity.

plitude of the fundamental component at different speeds by the amplitude of the component corresponding to the minimum rotational speed. The results of the EMD features are not included in Figure A.9, because the EMD features are relatively insensitive to changes in operating conditions and distorted the results in the figure. The EMD features were the least sensitive to changes in speed, while the AF-IPS performed second best. Even though the

(i) AF-IPS



(ii) Spectrogram

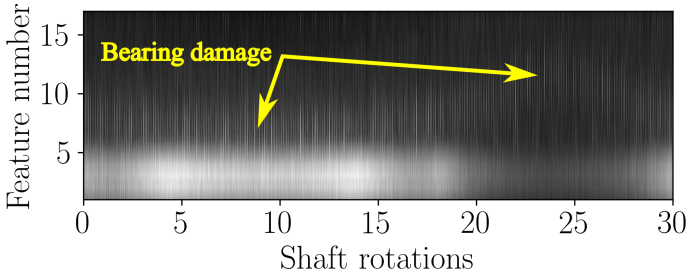


Figure A.10: A comparison of the spectrogram features and the AF-IPS features over shaft rotations for the phenomenological gearbox model with bearing damage. Operating condition profile three was used to generate the results.

Spectrogram features average normalised amplitude was not the most, the variation of the features in Figure A.9 was the most. This indicates that it is relatively sensitive to varying speed conditions with fairly large fluctuations with respect to rotational speed.

A.2.2.6 Discussion

The spectrogram features performed the best of the features in terms of its sensitivity to damage, however, the features were also sensitive to speed changes as seen in Figure A.9(ii). The spectrogram features and the AF-IPS features are presented for the phenomenological gearbox model with operating condition profile three being present in Figure A.10. It is clear that the spectrogram features are very sensitive to rotational speed changes because the bearing damage resides in different features as the speed varies which is in contrast to the AF-IPS features. The characteristics of the features depend on the rotational speed which makes feature selection approaches difficult to apply. This also means that a healthy Gaussian model, which is likely non-isotropic, may be more sensitive to bearing damage for specific features than others which is an undesirable characteristic.

The WPT features also performed very well compared to the AF-IPS. The db4 features performed the best, while the db10 features performed the worst as can be seen in Figure A.8(ii). The db44 wavelet basis function, which has performed very well in the literature [120],

does not perform well in this numerical investigation. Hence, it is clear that the performance of the WPT features is sensitive to the choice of wavelet basis functions and therefore care should be taken when using it in a novelty detection framework. This also highlights the benefits of using the AF-IPS; the basis function is fixed and does not have to be altered for different applications. In contrast to the other feature extraction approaches, the EMD was time-consuming to apply and also did not perform very well for discrepancy analysis due to mode mixing. However, there are many improvements to the EMD such as the ensemble EMD that is not investigated in this work but can potentially result in better features for discrepancy analysis.

A.2.3 Comparing the suitability of different data models

In discrepancy analysis, many different data models can be used to describe the characteristics of the data e.g. Gaussian mixture models [155]. The purpose of this investigation is to compare the suitability of different models for the discrepancy analysis methodology presented in Chapter 3. A similar feature extraction procedure is used as Chapter 3 and summarised in the next section, whereafter an overview of the different models are given and investigated on a numerical gearbox dataset.

A.2.3.1 Features for discrepancy analysis

The bearing features used in this investigation are based on the features used in Chapter 3. The wavelet packet transform is used with a Daubechies db1 wavelet function to decompose the time domain vibration signal into a set of wavelet coefficients. The Daubechies db1 basis function is used because it is able to detect singularities induced by bearing damage. A four level wavelet packet decomposition is performed which results in 16 wavelet coefficients, associated with different frequency bands, to be obtained. This allows the frequency bands to have a relatively fine resolution to detect damage manifesting in specific frequency bands and the features to have an acceptable dimensionality to avoid potential problems associated with the curse of dimensionality.

The wavelet coefficients of each frequency band are separately order tracked, whereafter the localised RMS is extracted from a moving average window with a length of $2\pi/50$ radians and a 50% overlap between adjacent windows. The short window length makes it possible to detect localised changes in the vibration signal, however, a method to select the optimal window length for bearing diagnostics still needs to be determined. Hence, instead of obtain-

ing a single 16 dimensional feature for a single vibration measurement i.e. by calculating the RMS of the wavelet coefficients for all shaft angles, many localised 16 dimensional features are extracted from a single vibration measurement by extracting features from the moving windows. A detailed description of the windowing and feature extraction procedures can be found in Chapter 3.

A.2.3.2 Feature modelling approaches for discrepancy analysis

The features of a healthy gearbox, extracted from the previous section, is subsequently modelled by the models investigated in this section, where new features extracted from a single window, denoted by $\boldsymbol{b}$, are evaluated with the discrepancy function to obtain a discrepancy measure. The discrepancy measures that correspond to the features extracted from each window are combined to form a discrepancy signal which is subsequently analysed for damage. The dimensionality of the feature space is 16, which means that the dimensionality of the feature vector is $\boldsymbol{b} \in \mathbb{R}^{16}$. A brief overview of each model and its associated discrepancy measure $\mathcal{M}(\boldsymbol{b})$ are given in this section.

Gaussian Model (GM)

A multivariate Gaussian Model (GM) is one of the simplest models that can be used for discrepancy analysis and has been successful for bearing diagnostics under varying speed conditions as seen in Chapter 3. The Mahalanobis distance for the features

$$\mathcal{M}(\boldsymbol{b}) = \sqrt{(\boldsymbol{b}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{b}-\boldsymbol{\mu})}, \tag{A.23}$$

uses the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, estimated from the healthy data, to calculate the discrepancy between the features $\boldsymbol{b}$ and the healthy features. If the features have a dimensionality of D, then the mean is defined $\boldsymbol{\mu} \in \mathbb{R}^{D\times 1}$ while the covariance is defined $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$. The features investigated in this work result in $D = 16$ and the mean and covariances have the same size throughout this work, unless stated otherwise. The GM has the benefit that it is easy (i.e. it has closed form maximum likelihood solutions) and computationally efficient to implement, however, it is incapable of describing multi-modal and non-Gaussian distributions.

Gaussian Mixture Model (GMM)

The Gaussian Mixture Model (GMM) is a generalisation of the GM, where a weighted sum of N-Gaussian densities are used to approximate the density of the features under investigation.

The GMM has the benefit that it can approximate any density to an arbitrary accuracy if a sufficient number of mixture components is used [114]. The GMM has been successfully used for discrepancy analysis-based gear diagnostics [155, 156], with the NLL of the discrepancy signal

$$\mathcal{M}(\boldsymbol{b}) = -\log\left(\sum_{i=1}^{N_c} \pi_i \mathcal{N}(\boldsymbol{b}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\right), \tag{A.24}$$

used as a discrepancy measure for the GMM. In Equation (A.24), π_i is the weight of the ith mixture component and denotes the Gaussian probability density function associated with the ith mixture component. Even though the GMM is very flexible, it requires the estimation of many parameters (i.e. N-means, N-covariances and N-weight parameters) with an optimisation procedure, and the appropriate number of mixture components needs to be estimated from the data as well. GMMs are also susceptible to overfitting and care needs to be taken when optimising the models [114]. In this investigation, the Expectation-Maximisation (EM) algorithm is used to obtain the model parameters using a maximum likelihood framework.

Hidden Markov Model (HMM)

The aforementioned models are optimised based on the assumption that each data point is independent and identically distributed (i.i.d.), which may be limiting. HMMs overcome the i.i.d. limitation by assuming that the features are generated by a latent state that follows a Markov process and the features are related to the latent states by an observation density. A first-order Markov process is used with a Gaussian observation distribution for each latent state in this investigation. The discrepancy measure for the HMM is calculated with

$$\mathcal{M}(\boldsymbol{b}) = -\log p(\boldsymbol{b}_i|\boldsymbol{b}_{i-1}, \ldots, \boldsymbol{b}_1), \tag{A.25}$$

for the feature at window i. The HMM is even more computationally expensive than the GMM, because the transition probability matrix of the Markov process needs to be estimated as well. The additional temporal dimension gives the HMM additional discriminatory power for detecting novelties over the GMM and GM. The mean and covariance of the Gaussian observation density of each state, the initial state, and the transition matrix are obtained using the Baum-Welch algorithm using a maximum likelihood framework in this investigation.

Principal Component Analysis-based Reconstruction (PCA-R)

Principal Component Analysis (PCA) is a popular dimensionality reduction technique, where the eigenvectors of the covariance matrix of the features are used to transform the features to a lower subspace by removing redundant information. The dimensionality reduction is performed with

$$\boldsymbol{z} = \boldsymbol{V}_{1:d}^{T}\left(\boldsymbol{b} - \boldsymbol{\mu}\right), \tag{A.26}$$

where $\boldsymbol{V}_{1:d} \in \mathbb{R}^{D\times d}$ represents the eigenvectors associated with the d-largest eigenvalues, $\boldsymbol{\mu}$ denotes the mean of the healthy features and $\boldsymbol{z} \in \mathbb{R}^{d\times 1}$ denotes the principal components. It is possible to obtain a one dimensional signal from the original signal with PCA (i.e. $d = 1$), however, this is not a discrepancy signal and very poor results are obtained when using it in a discrepancy analysis context.

In this investigation, a new discrepancy analysis approach based on PCA is investigated. The mean and the eigenvectors of the healthy features are used to map the lower dimensional principal component features with a dimensionality of $\mathbb{R}^{d\times 1}$ back to the original feature space with

$$\boldsymbol{b}_{rec} = \boldsymbol{V}_{1:d}\boldsymbol{V}_{1:d}^{T}\left(\boldsymbol{b} - \boldsymbol{\mu}\right) + \boldsymbol{\mu}, \tag{A.27}$$

where the reconstructed features are denoted by $\boldsymbol{b}_{rec}$ which is a vector. The root-mean-square of the reconstruction error $\boldsymbol{b} - \boldsymbol{b}_{rec}$ is used as a discrepancy measure

$$\mathcal{M}(\boldsymbol{b}) = \sqrt{\frac{1}{D}\left(\boldsymbol{b} - \boldsymbol{V}_{1:d}\boldsymbol{V}_{1:d}^{T}\left(\boldsymbol{b} - \boldsymbol{\mu}\right) - \boldsymbol{\mu}\right)^{T}\left(\boldsymbol{b} - \boldsymbol{V}_{1:d}\boldsymbol{V}_{1:d}^{T}\left(\boldsymbol{b} - \boldsymbol{\mu}\right) - \boldsymbol{\mu}\right)}. \tag{A.28}$$

If $d < D$, then the reconstruction error will be non-zero and if the machine condition changes, the reconstruction error will increase. The dimensionality d is estimated by selecting the minimum PCA dimension that has an accumulative contribution rate of 95%.

A.2.3.3 Numerical Validation

Five different FS, namely 0.0, 0.02, 0.04, 0.06 and 0.08 are investigated here as well. The FS merely quantifies the relative magnitude of the bearing impulses of the different cases e.g. a FS of 0.02 contains impulses that are on average four times smaller than a FS of 0.08, while FS of 0.0 simulates a healthy bearing.

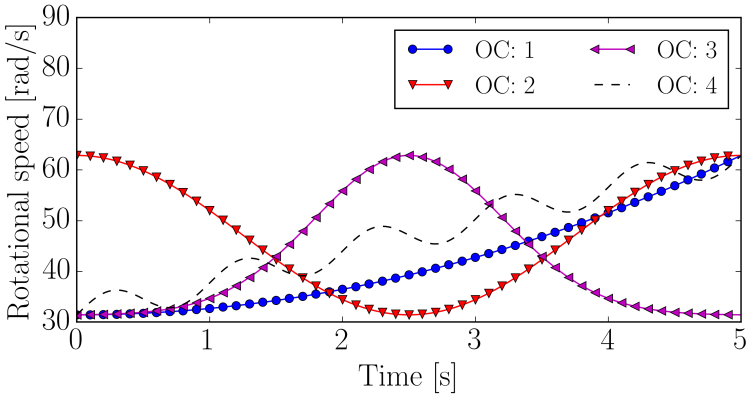


Figure A.11: The Operating Condition (OC) profiles investigated for the phenomenological gearbox model.

Spectrum of the discrepancy signals

The frequency content of the discrepancy signals, where the bearing has a FS of 0.02, is shown in Figure A.12 for the four investigated models. This is obtained by calculating the spectrum of the discrepancy signal of each measurement in the dataset. The damage has

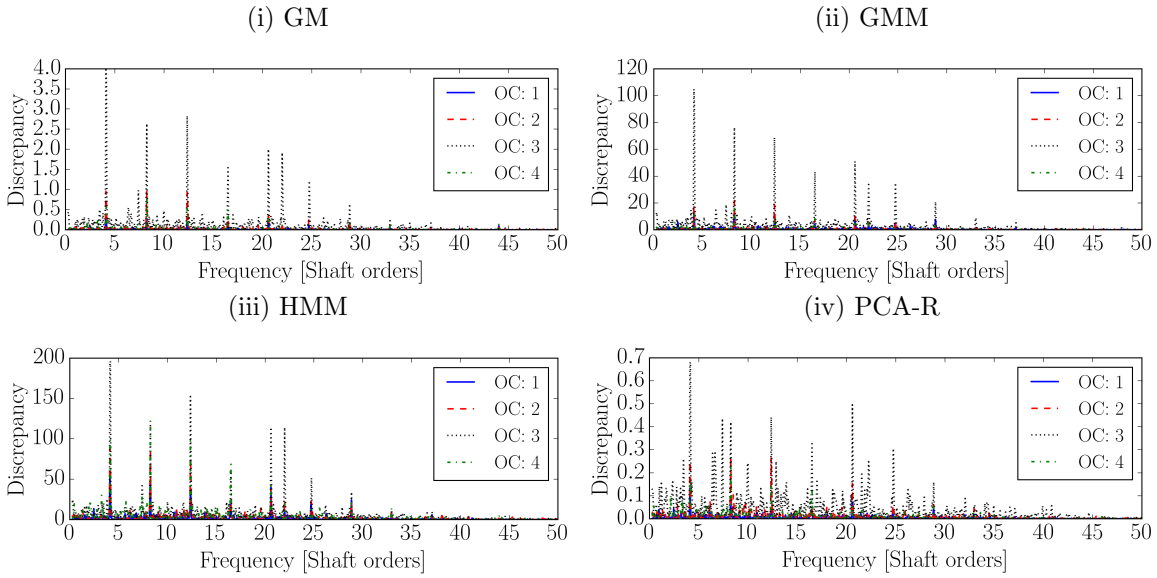


Figure A.12: The spectra of the discrepancy signals obtained from the different models for a bearing with a FS of 0.02.

a fundamental component at 4.12 shaft orders, with the fundamental component and a few harmonics being seen in the spectra of all the models. However, the spectra are definitely sensitive to changes in operating conditions as seen by the variation of the amplitudes between different operating condition profiles. The spectra of the discrepancy signals generated with the GM, GMM and the HMM contain mostly diagnostic information, whereas the PCA-R spectrum contains much noise which impedes the condition inference process.

It is difficult to quantify the robustness of the different models to varying operating

conditions and the sensitivity of the models to damage from the spectra and therefore two separate investigations are performed in the next sections to quantify the performance of the models.

Sensitivity to damage

The sensitivity of the discrepancy signals obtained from the different models to damage is investigated in this section. The amplitude of the fundamental component of the characteristic frequency (i.e. 4.12 shaft orders) in the spectrum is used to quantify the magnitude of the damage for the different FS values and the different models in this section. The ampli-

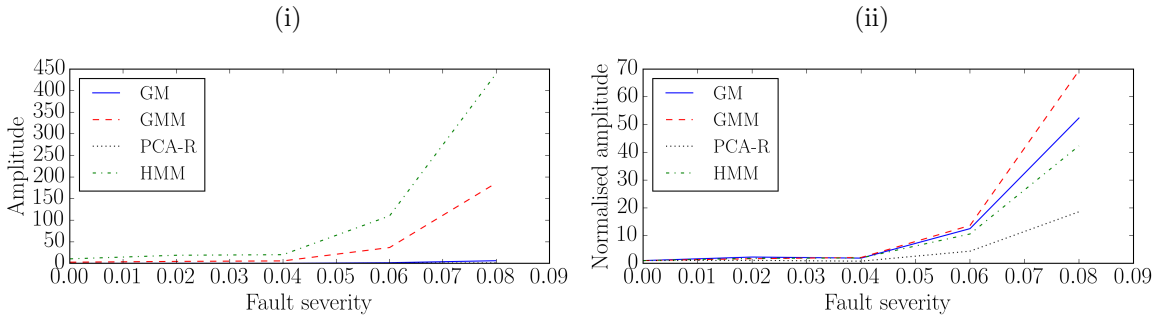


Figure A.13: The sensitivity of the models to bearing damage are compared. The amplitude (a) and normalised amplitude (b) of the fundamental ball-pass outer race component are shown over fault severity.

tude of the fundamental components in Figure A.13(i) are averaged over the four operating condition profiles, and presented over bearing fault severity. The HMM has the largest discrepancy values with respect to damage, where in contrast the PCA-R and the GM have the smallest discrepancy values. However, it is not desired to quantify the absolute values, but instead to quantify the relative increase in the values as the damage progresses. Therefore, the amplitudes in Figure A.13(i) are normalised by the amplitude of the component from a healthy machine, with the results presented in Figure A.13(ii).

It can be observed that the GMM is the most sensitive to damage, while the GM is second most sensitive to damage, performing even better than the HMM. The PCA-R model does not perform very well relative to the other models, however, it successfully detects changes in the bearings condition. The GM is very simple relative to the GMM and HMM and may not represent the exact density of the healthy features, however, it is clearly capable of detecting small changes in the spectrum of the machine and is therefore well suited for discrepancy analysis.

Sensitivity to operating conditions

The discrepancy signal of a healthy gearbox is presented over rotational speed in Figure A.14 for the different models. The conditional mean and variance of the discrepancy signal, given the rotational speed, varies over rotational speed. The GM, GMM and the PCA-R models contain very similar characteristics, however, the HMMs statistics vary significantly over rotational speed. It is desired to quantify the sensitivity of the discrepancy signal

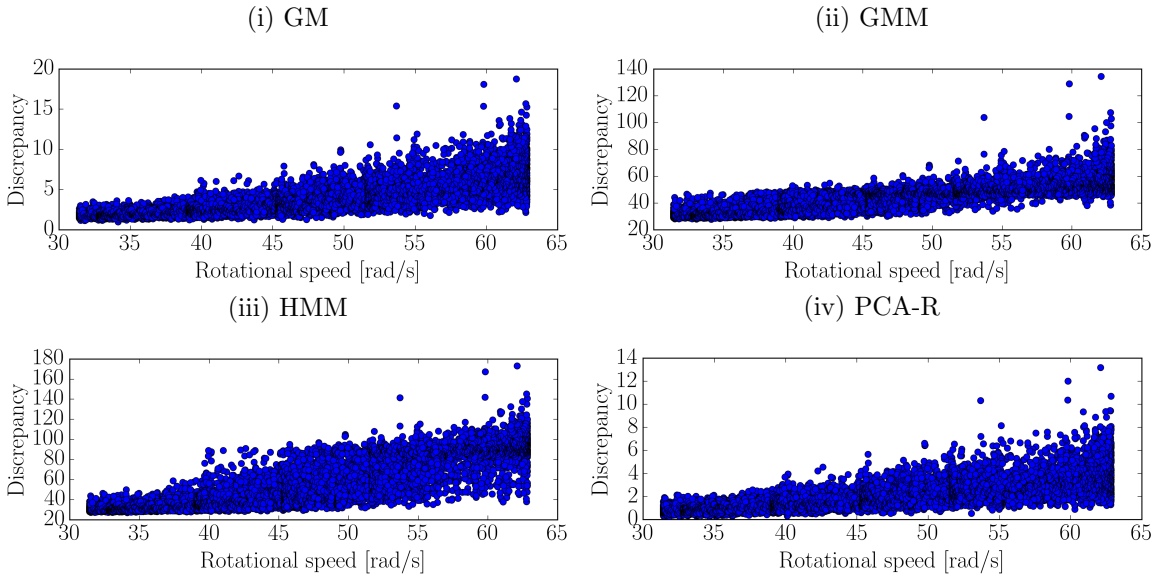


Figure A.14: The discrepancy signal of a healthy gearbox versus the rotational speed is presented for the different models.

to rotational speed and therefore the conditional mean of the discrepancy signal given the rotational speed is calculated with the procedure used in Chapter 3. The raw conditional mean results are presented in Figure A.15(i). It can be observed that the HMM contains the largest changes in discrepancy value, while the GM and the PCA-R models contain the smallest changes.

Similar to the investigation in Section 3.2.2., the normalised discrepancy, obtained from dividing the discrepancy value by the discrepancy value that correspond to the lowest rotational speed, is investigated as well. The results in Figure A.15(ii) indicate the PCA-R and the GM are the most sensitive to changes in rotational speed, while the GMM is the most robust. Even though the GM results are sensitive to rotational speed variations, it can easily be compensated for by the method proposed in Section 3.2.5.

Discussion

In this investigation, different models for discrepancy analysis were investigated and com-

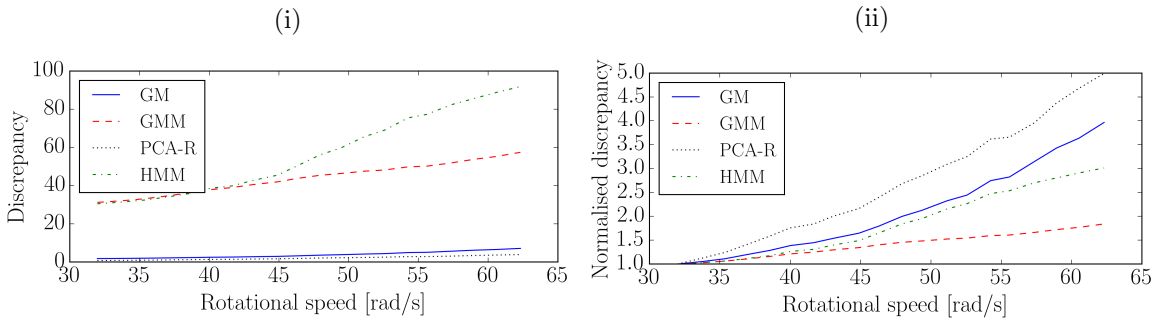


Figure A.15: The sensitivity of the discrepancy signal to rotational speed is compared for the different models by comparing the conditional mean A.15(i) and the normalised conditional mean A.15(ii) of the results in Figure A.14.

pared. The Gaussian mixture model is the most sensitive to damage and the most robust to varying speed conditions and is therefore best suited for discrepancy analysis according to this investigation. The hidden Markov model is quite sensitive to operating conditions which may be ascribed to the Markov process of the model which struggles to distinguish between changes in machine conditions and operating conditions. The Gaussian model performed very well with the bearing diagnostics task which has the advantage that it can be easily implemented without considering convergence and hyperparameter optimisation issues. Lastly, even though the principal component reconstruction-based model did not perform as well as the other models, the investigation indicated how the reconstruction model framework can be used for discrepancy analysis.

A.3 Quality metrics used in Chapter 5

In Chapter 5, different squared envelope spectra and synchronous averages are presented. It is difficult to compare the performance of the novel and the original signals from figures and therefore performance metrics are used. In this section, an illustration is given of the calculation of the performance metrics introduced in Section 5.4.2.2.

The calculation of the quality metrics used in Section 5.4 are presented in Figure A.16. The Quality Measure (QM) of the Squared Envelope Spectrum (SES) is calculated with Equation (5.23) where the average component of the fault frequency components shown in Figure A.16 is denoted by $\text{SES}_{damaged}$. The SES of the noise floor is calculated by calculating the median of the SES in the frequency range shown in Figure A.16, with the identified fault frequency components ignored.

The QM of the synchronous average is calculated with Equation (5.24). The healthy and

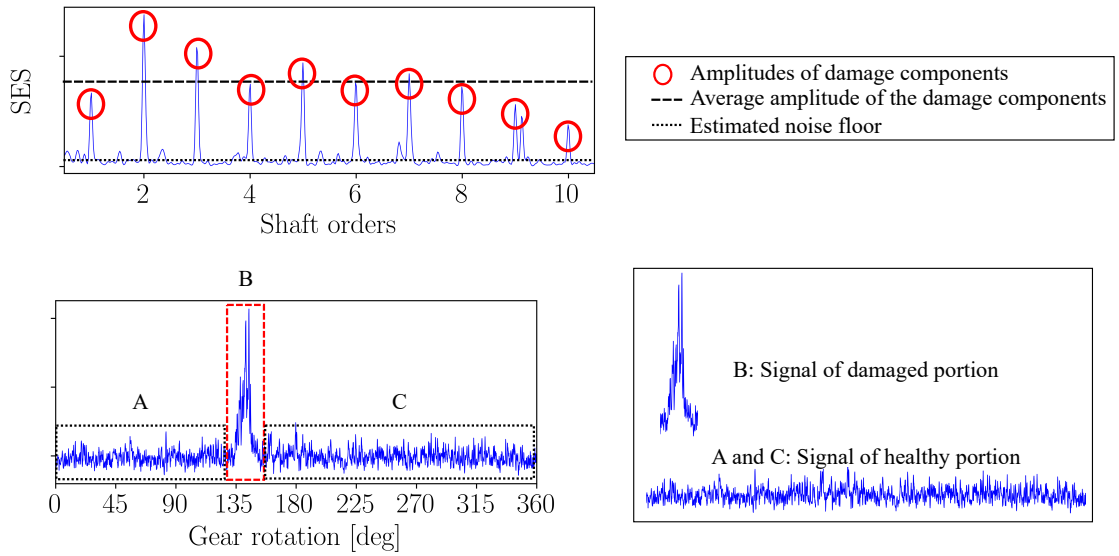


Figure A.16: A supplementary figure to illustrate how the quality of the synchronous average and of the squared envelope spectrum is calculated.

damaged portions are known beforehand as shown in Figure A.16. Therefore, the power of the healthy and damaged portions shown in Figure A.16 can easily be calculated and is used in Equation (5.24).

Appendix B

Phenomenological gearbox model information

The casing vibration signal of the phenomenological gearbox model comprises the following components

$$x_c(t) = x_b(t) + x_{dgd}(t) + x_{gmc}(t) + x_n(t), \tag{B.1}$$

where $x_c(t)$ is the casing or measured vibration signal, $x_b(t)$ is the bearing vibration signal induced by damage, $x_{dgd}(t)$ is the distributed (or random) gear component, $x_{gmc}(t)$ is the gear mesh component or deterministic gear component, and x_n is the broadband noise component. The phenomenological gearbox model is used in a few investigations, where each investigation used different parameters. In this appendix, the different parameters are summarised separately for each investigation.

B.1 Model used in Section 3.3

The single degree-of-freedom impulse response function, used to relate the raw vibration component at the source and the measured response, is described by the natural frequency of the system and the damping ratio. The natural frequency and the damping ratio are given in Table B.1 for this investigation.

The relative power of the signal component i with respect to the noise component is calculated with

$$\mathrm{SNR}_i = 20 \log \left(\frac{\sqrt{(1/T) \int_0^T x_i(t) \cdot x_i(t) dt}}{\sqrt{(1/T) \int_0^T x_n(t) \cdot x_n(t) dt}} \right), \tag{B.2}$$

and used to quantify the magnitude of the different components with respect to the noise.

Table B.1: The parameters of the impulse response function, defined in Equation (3.11), are given for the different signal components used in Section 3.3.1.

Component	f_n Hz	ζ
Gear mesh component	2000	0.05
Distributed gear damage component	3500	0.05
Outer race bearing damage	5000	0.05

The results are presented in Table B.2 for the different fault severities.

Table B.2: The ratio of the power of the signal components to the power of the noise signals, used in Section 3.3.1, is calculated with Equation (B.2) for the different Fault Severities (FS). Abbreviations: Gear Mesh Component (GMC), Distributed Gear Damage (DGD), Bearing (B).

FS	Casing	GMC	DGD	B
0.0	10.2777	9.47	-1.0045	-inf
1.0	10.2903	9.4707	-1.0052	-15.749
2.0	10.3245	9.4732	-0.9806	-9.7349
4.0	10.4478	9.4668	-1.0055	-3.7025

B.2 Model used in Section 4.3.1

The impulse response function parameters and the ratio of the power of the signal components and the noise components are given in Table B.3 and Table B.4, respectively, for the implementation in Section 4.3.

Table B.3: The parameters of the impulse response function are given for the different signal components used in Section 4.3.

Component	f_n Hz	ζ
Gear mesh component	2500	0.05
Distributed gear damage component	1500	0.05
Outer race bearing damage	6000	0.05

Table B.4: The ratio of the power of the signal components to the power of the noise signals, used in Section 4.3, is calculated with Equation (B.2) for the different Fault Severities (FS).

FS B	FS G	Casing	GMC	DGD	B
0.0	0.0	1.9436	-2.4888	-inf	-inf
1.0	0.0	2.0307	-2.4503	-inf	-15.964
2.0	0.0	2.221	-2.4715	-inf	-9.8811
0.0	1.0	3.309	-2.4011	-2.436	-inf
1.0	1.0	3.3216	-2.5261	-2.5057	-15.9222
2.0	1.0	3.4974	-2.4689	-2.4428	-9.8773
0.0	2.0	5.8157	-2.4573	3.5268	-inf
1.0	2.0	5.8417	-2.4496	3.5042	-15.8276
2.0	2.0	5.8488	-2.5884	3.4262	-10.118

B.3 Model used in Section A.2.2

The impulse response function parameters are given in Table B.5 and the ratio of the power of the signal components to the power of the noise component is given in Table B.6 for the investigation performed in Section A.2.2.

Table B.5: The parameters of the impulse response function are given for the different signal components used in Section A.2.2.

Component	f_n Hz	ζ
Gear mesh component	5000	0.05
Distributed gear damage component	2500	0.05
Outer race bearing damage	1500	0.05

Table B.6: The ratio of the power of the signal components to the power of the noise signals, used in Section A.2.2, is calculated with Equation (B.2) for the different Fault Severities (FS).

FS	Casing	GMC	DGD	B
0	11.5775	11.2462	-12.3859	-inf
2	11.5662	11.2274	-12.4057	-21.2925
4	11.5611	11.2176	-12.4039	-15.3211
8	11.5985	11.225	-12.3522	-9.1889
12	11.6495	11.2368	-12.3244	-5.6542
16	11.7226	11.2379	-12.3859	-3.2184

B.4 Model used in Section A.2.3

The impulse response function parameters and the power of the signal components and the noise components are given in Table B.7 and Table B.8 for the investigation performed in Section A.2.3.

Table B.7: The parameters of the impulse response function are given for the different signal components used in Section A.2.3.

Component	f_n Hz	ζ
Gear mesh component	4000	0.05
Distributed gear damage component	1500	0.05
Outer race bearing damage	2200	0.05

Table B.8: The ratio of the power of the signal components to the power of the noise signals, used in Section A.2.3, is calculated with Equation (B.2) for the different Fault Severities (FS).

FS	Casing	GMC	DGD	B
0	4.6341	2.8064	-inf	-inf
2	4.627	2.7968	-inf	-21.9809
4	4.6545	2.79	-inf	-15.9372
6	4.6982	2.779	-inf	-12.4778
8	4.7943	2.8086	-inf	-9.9174