UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

A COST-EFFECTIVE DIAGNOSTIC METHODOLOGY USING PROBABILISTIC APPROACHES FOR GEARBOXES OPERATING UNDER NON-STATIONARY CONDITIONS

by **Stephan Schmidt**

Submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Mechanical Engineering)

in the

Department of Mechanical and Aeronautical Engineering

Faculty of Engineering, Built Environment and Information Technology

University of Pretoria

November 2016

SUMMARY

A COST-EFFECTIVE DIAGNOSTIC METHODOLOGY USING PROBABILISTIC APPROACHES FOR GEARBOXES OPERATING UNDER NON-STATIONARY CONDITIONS

by

Stephan Schmidt

Supervisors:	Prof. PS Heyns and Dr. JP de Villiers
Department:	Mechanical and Aeronautical Engineering
University:	University of Pretoria
Degree:	Master of Engineering (Mechanical Engineering)
Keywords:	Diagnostics; Fluctuating operating conditions; Novelty detection; Tacholess order tracking; Hidden Markov models; Probabilistic; Model combination

Condition monitoring is very important for critical assets such as gearboxes used in the power and mining industries. Fluctuating operating conditions are inevitable for wind turbines and mining machines such as bucket wheel excavators and draglines due to the continuous fluctuating wind speeds and variations in ground properties, respectively. Many of the classical condition monitoring techniques have proven to be ineffective under fluctuating operating conditions and therefore more sophisticated techniques have to be developed. However, many of the signal processing tools that are appropriate for fluctuating operating conditions can be difficult to interpret, with the presence of incipient damage easily being overlooked.

In this study, a cost-effective diagnostic methodology is developed, using machine learning techniques, to diagnose the condition of the machine in the presence of fluctuating operating conditions when only an acceleration signal, generated from a gearbox during normal operation, is available. The measured vibration signal is order tracked to preserve the angle-cyclostationary properties of the data. A robust tacholess order tracking methodology is proposed in this study using probabilistic approaches. The measured vibration signal is order tracked with the tacholess order tracking method (as opposed to computed order tracking), since this reduces the implementation and the running cost of the diagnostic methodology.

Machine condition features, which are sensitive to changes in machine condition, are extracted from the order tracked vibration signal and processed. The machine condition features can be sensitive to operating condition changes as well. This makes it difficult

to ascertain whether the changes in the machine condition features are due to changes in machine condition (i.e. a developing fault) or changes in operating conditions. This necessitates incorporating operating condition information into the diagnostic methodology to ensure that the inferred condition of the machine is not adversely affected by the fluctuating operating conditions. The operating conditions are not measured and therefore representative features are extracted and modelled with a hidden Markov model. The operating condition machine learning model aims to infer the operating condition state that was present during data acquisition from the operating condition features at each angle increment. The operating condition state information is used to optimise robust machine condition machine learning models, in the form of hidden Markov models.

The information from the operating condition and machine condition models are combined using a probabilistic approach to generate a discrepancy signal. This discrepancy signal represents the deviation of the current features from the expected behaviour of the features of a gearbox in a healthy condition. A second synchronous averaging process, an automatic alarm threshold for fault detection, a gear-pinion discrepancy distribution and a healthy-damaged decomposition of the discrepancy signal are proposed to provide an intuitive and robust representation of the condition of the gearbox under fluctuating operating conditions. This allows fault detection, localisation as well as trending to be performed on a gearbox during fluctuating operation conditions.

The proposed tacholess order tracking method is validated on seven datasets and the fault diagnostic methodology is validated on experimental as well as numerical data. Very promising results are obtained by the proposed tacholess order tracking method and by the diagnostic methodology.

Acknowledgements

I would like to acknowledge and thank the following people for their contribution to this work:

- Prof. Stephan Heyns
- Dr. Pieter de Villiers
- Prof. Fakher Chaari
- Mr. George Breitenbach
- Mr. Herman Booysen
- Dr. Abrie Oberholster
- Mr. Dawie Diamond
- Mr. Renier Coetzee

My family and friends for their continuous support throughout my study.

The Eskom power plant engineering institute (EPPEI) for funding the redevelopment of the gearbox test facility. The experimental work of this dissertation was conducted on the gearbox test facility.

Contents

List of abbreviations

ACR	Accumulative contribution rate
ANN	Artificial neural network
BIC	Bayesian information criterion
CB	Confidence bound
CBM	Condition-based maintenance
COT	Computed order tracking
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
GMM	Gaussian mixture model
HMM	Hidden Markov model
IAS	Instantaneous angular speed
IF	Instantaneous frequency
IMF	Intrinsic mode function
LDT	Laser displacement transducer
MC	Machine conditions
MCF	Machine condition features
NLL	Negative log-likelihood
OC	Operating conditions
OCS	Operating condition state
PCA	Principal component analysis
PPR	Pulses per revolution
RMS	Root mean square
RMSE	Root mean square error
STFT	Short-time Fourier transform
SVM	Support vector machine
VKF	Vold-Kalman filter
WPT	Wavelet packet transform
WVD	Wigner-Ville distribution

List of symbols

General

Roman symbols:

a_1	Dilation or scale parameter of CWT
a_2	Translation parameter of CWT
D	Dimensionality of original feature space
d	Dimensionality of reduced feature space
f	Frequency
f_s	Sampling frequency
$\boldsymbol{J}$	Gear-pinion discrepancy distribution
n	Sample or increment
p	Probability density function
T	Transpose
t	Continuous time
$\boldsymbol{v}$	Eigenvector
W	Wavelet coefficients
$\boldsymbol{X}$	Multidimensional dataset
x	Signal
$\boldsymbol{x}_n$	Multidimensional data vector at increment n
$\boldsymbol{Z}$	Multidimensional latent variable at all time increments
$\boldsymbol{z}_n$	Multidimensional latent variable at increment n

Greek symbols:

μ	Mean
$\boldsymbol{\mu}$	Multidimensional mean vector
ξ	Eigenvalue
$\boldsymbol{\Sigma}$	Covariance matrix
σ	Standard deviation
υ	Time-limited window
ψ	Wavelet basis function

Chapter 2

Roman symbols:

$\boldsymbol{C}$	Viscous damping matrix

$c_{a\,i}$	Electric motor constants
c_{CR}	Contact ratio
c_k	Stiffness damping constant
c_m	Mass damping constant
$\boldsymbol{F}$	External load vector
g_n	Proportional drop in electric motor speed
g_b	Slip
I_{11}	Inertia of the motor
I_{12}	Inertia of the pinion
I_{21}	Inertia of the gear
I_{22}	Inertia of the machine
$\boldsymbol{K}$	Total stiffness matrix of the gearbox model
$\bar{\boldsymbol{K}}$	Mean stiffness matrix of the gearbox model
$\boldsymbol{K}_s$	Static contribution to total the stiffness matrix
$\boldsymbol{K}_t$	Dynamic contribution to the total stiffness matrix
k_{gm}	Gear mesh stiffness
$\bar{k}_{gm}$	Mean gear mesh stiffness
k_{xi}	Translational stiffness component in the x-direction at block i
k_{yi}	Translational stiffness component in the y-direction at block i
$k_{\theta\,i}$	Torsional stiffness of the shaft associated with block i
$\boldsymbol{M}$	Mass matrix of the gearbox model
m_i	Mass of block i, or pinion (m_p) or gear (m_g)
N_i	Number of teeth on the pinion (N_p) or the gear (N_g)
$r_{b\,i}$	Base circle radius of the pinion (r_{b12}) or the gear (r_{b21})
$\boldsymbol{S}$	Dynamic stiffness coefficient matrix
s_i	Dynamic stiffness coefficient i
T_b	Torque at breakdown
T_M	Electric motor torque
T_L	Torque applied by the machine
x_i	Translational degree of freedom in the x-direction of block i
$\boldsymbol{Y}$	Displacement degrees of freedom of the gearbox model
y_i	Translational degree of freedom in the y-direction of block i

Greek symbols:

α	Pressure angle
δ	Displacement in the line of action
θ_{11}	Torsional degree of freedom of the driving motor
θ_{12}	Torsional degree of freedom of the pinion
θ_{21}	Torsional degree of freedom of the gear

θ_{22}	Torsional degree of freedom of the machine
ω_r	Rotational speed of the electric motor
ω_s	Synchronous speed of the electric motor

Chapter 3

Roman symbols:

$\boldsymbol{A}$	The Vold-Kalman filter square diagonal matrix
B_w	Bandwidth factor
c_s	Relationship between fundamental frequency of IF and f_r
e	Phase error
e_R	Relative phase error
f_c	Centre frequency of maxima tracking constraint
f_F	Polynomial function
f_{IF}	Instantaneous frequency
$\boldsymbol{f}_{\max}$	Maxima tracked instantaneous frequency vector
$f_{\max}$	Maxima tracked instantaneous frequency
f_r	Rotational frequency of shaft of interest
H	Hilbert transform
j	Imaginary unit
k	Harmonic number
$\mathcal{L}_n$	Lagrangian equation to be minimised at time increment n
N_m	Number of points used for linear regression
N_p	Order of the polynomial used for linear regression
$\boldsymbol{Q}$	Design matrix
s_i	Slack variable associated with constraint i
$\boldsymbol{w}$	Weight vector
w_i	Weight associated with component i
x_{IF}	Mono-component sinusoidal signal with frequency f_{IF}

Greek symbols:

β	Envelope of a mono-component signal of interest
Γ	Short-time Fourier transform
Δt	Time step or sampling period, reciprocal of f_s
$\Delta f_{c\,i}$	Frequency bandwidth of feasible region of constraint i
Δt_ϕ	Denotes the phase difference between two signals in time increments
$\boldsymbol{\varepsilon}$	Non-homogeneity term

$\boldsymbol{\zeta}$	Noise and other sinusoidal components vector
κ	Cost function
λ_i	Lagrange multiplier associated with constraint i
ρ_i	Penalty factor associated with constraint i
ϕ_{IF}	Instantaneous phase of the instantaneous frequency component
ϕ_r	Instantaneous phase of the shaft

Chapter 4

Roman symbols:

$\boldsymbol{b}_n$	Multidimensional machine condition features at increment n
$\boldsymbol{J}$	Gear-pinion discrepancy distribution
$\mathcal{M}_i$	Model i
N_m	Number of measurements
$N_{r,k}$	Number of rotations associated with measurement k
N_r^i	Number of averages completed by the gear N_r^g and the pinion N_r^p
N_s	Number of samples per revolution
N_s^i	Number of samples per revolution for the gear N_s^g and the pinion N_s^p
N_T	The number of samples in a signal
$N_{\boldsymbol{X}}$	Number of data in dataset $\boldsymbol{X}$
N_θ	Number of parameters
$\boldsymbol{o}_n$	Multidimensional operating condition features at increment n
s	Scale index
$\mathcal{T}_{\alpha,N_h-1}$	Student-t indicator with $100 \times (1-\alpha)\%$ confidence interval generated from N_h healthy measurements
$\boldsymbol{V}_{1:d}$	Eigenvectors associated with the d-largest eigenvalues
$\boldsymbol{y}_n$	Multidimensional data with a reduced dimensionality
$\mathbf{z}_n^o$	Latent variable of the operating condition model

Greek symbols:

$\boldsymbol{\mu}_{\boldsymbol{X}}$	The multidimensional mean of dataset $\boldsymbol{X}$
η	Discrepancy signal
η_i	Discrepancy signal of measurement i
η^i	Discrepancy signal of gear η^g or pinion η^p
η_{ref}	Reference discrepancy signal
$\boldsymbol{\theta}$	Model parameters

$\hat{\boldsymbol{\theta}}$	Maximum likelihood estimate of the parameters
$\{\boldsymbol{\theta}_b\}$	Set of machine condition model parameters
$\boldsymbol{\theta}_b^j$	Parameters of machine condition model j
$\boldsymbol{\theta}_o$	Operating condition model parameters
μ_W	Average of the non-zero, windowed wavelet coefficients at a scale
$\mu_\eta^{(1)}$	Synchronous average of the discrepancy signal
$\mu_{\eta,k}^{(1)}$	Synchronous average of the discrepancy signal generated from measurement k
$\mu_\eta^{(2)}$	Synchronous average of $\mu_\eta^{(1)}$ between measurements or the second synchronous average

Chapter 5

$\mathrm{CB_v}$	Confidence bound used as an alarm threshold
μ_d	Mean discrepancy of the damaged portion
μ_h	Mean discrepancy of the healthy portion
μ_t	Synchronous or second synchronous average of the discrepancy signal (depends on section)
μ_v	Synchronous average of the discrepancy signal
σ_t	Standard deviation of the second synchronous average of the discrepancy signal

Chapter 1 Introduction

1.1 Background

Maintenance on machines and equipment is an unavoidable and an essential task in industrial environments where expensive equipment is operated. Unexpected breakdowns due to a lack of maintenance and replacing equipment based on the time it was used are inefficient and expensive for asset management. A more efficient alternative, namely condition-based maintenance (CBM), is required when performing maintenance on critical equipment that requires a high level of reliability (Jardine et al., 2006). CBM is a maintenance strategy that uses the condition of the machine to determine which components require maintenance and replacement and which are still in an acceptable condition for operation. CBM consists of diagnostics which aims to detect, isolate and characterise faults present in the system as well as prognostics, which predicts when a fault or failure will occur (Jardine et al., 2006, Lee et al., 2014).

Gearboxes are used in the power (wind turbines), mining (draglines, continuous miners, bucket wheel excavators, conveyor belts), aeronautical and automotive industries. The gearbox within a wind turbine has the highest relative cost over a 20-year period of all components and it contributes to 10.9% of the total wind turbine cost (Busby, 2012). The gearboxes are subjected to harsh operating conditions and inherent fatigue cycles due to the repetitive loading and unloading cycles of the gear teeth during meshing. This can result in bending fatigue, surface fatigue etc. which can result in ultimate failure of the gearbox and lead to production, financial etc. losses.

CBM of gearboxes remains a challenging task due to the inherent and unavoidable fluctuating conditions in which the machines regularly operate. The fluctuating operating conditions are induced by the continuously fluctuating wind speed for wind turbines (Zimroz et al., 2014), non-homogeneous ground properties for bucket wheel excavators (Bartelmus and Zimroz, 2009*b*), changes in transported loads for draglines (Eggers et al., 2007) and conveyor systems (Bartelmus et al., 2010). The fluctuating operating conditions manifest very similar to faults in a measured vibration signal (Randall,

1981) and it results in additional complex phenomena such as varying signal-to-noise ratios (Chaari et al., 2012). The assumption that any change in the measured vibration signal is only attributed to changes in the condition of the machine, is violated in fluctuating operating conditions (Stander et al., 2002). Varying rotational speeds also influence the order tracking process, since the transmission path from the impact, produced by local damage, is different for a tachometer located at a shaft and a transducer located on a bearing housing. This results in phase distortion when performing computed order tracking on the signal (Stander and Heyns, 2006). There is a need for a cost-effective gearbox diagnostic methodology which is able to detect, locate and characterise faults in fluctuating operating conditions.

1.2 Literature survey

A literature survey is performed on vibration-based condition monitoring approaches for rotary machines with the focus placed on gearbox diagnostics. The first section deals with vibration-based condition monitoring techniques. The second section deals with techniques which learn the characteristics within the data and use this information to infer the condition of the machine.

1.2.1 Vibration-based condition monitoring

Vibration measurements are regularly used in the rotating machine diagnostics field, since changes in the condition of the machine are instantaneously reflected in the vibration signal (Randall, 2011). However, changes in the machine condition (i.e. developing faults) are usually concealed by dominant components within the vibration signal, such as the gear mesh frequencies, changes in operating conditions and noise. Sophisticated signal processing techniques are required to extract the characteristics or machine condition features, that reflect the condition of the machine, from the vibration signal.

An overview of the most common signal processing techniques applied to gearbox condition monitoring is discussed in the succeeding sections.

1.2.1.1 Statistical time domain features

Statistical time domain metrics such as the root mean square (RMS), kurtosis and crest factor can be used to infer changes in the condition of the machine due to a de-

veloping fault. Tandon (1994) investigated statistical time domain metrics for bearing fault detection and found that the RMS, vibration peak and power are sensitive to a deteriorating bearing, while the crest factor is not in all cases. Even though the aforementioned features are sensitive to damage in the rotary machine, they are sensitive to changes in operating conditions as well (Lin and Zhao, 2014). This makes it difficult to ascertain the cause of the changes within the features which make the time domain features inappropriate for diagnostics in fluctuating operating conditions (Dowling et al., 1993, Lin and Zhao, 2014).

However, it is possible to alleviate the problems associated with the aforementioned statistical features by incorporating operating condition information into the fault detection process. Zimroz et al. (2014) extracted the peak-to-peak and RMS of the vibration signal from a wind turbine's gearbox as diagnostic features. The features have bimodal probability density functions due to the fluctuating operating conditions during data acquisition. The relationship between the machine condition features and the power generated by the wind turbine (i.e. operating condition information) is obtained by linear regression analysis. The parameters or weights of the linear regression model are sensitive to changes in the bearing condition in the presence of fluctuating operating conditions.

1.2.1.2 Frequency domain analysis

The frequency domain representation of a vibration signal is obtained by means of a discrete Fourier transform which makes the implicit assumption that the signal under consideration is stationary and periodic. Sidebands are common in the spectrum of the vibration signal due to modulation effects such as impacts produced by faults (McFadden, 1985), but it is also produced by operating condition changes (Randall, 1981, 2011). The fluctuating rotational speeds result in spectral smearing to occur as well, since the discrete Fourier transform is unable to handle fluctuating characteristic frequencies. Hence, frequency domain analysis of a temporal signal is inappropriate for fault detection in fluctuating operating conditions (Lin and Zhao, 2014). Order tracking, discussed in section 1.2.1.3, and time-frequency analysis, discussed in section 1.2.1.8, overcome the requirement for constant rotational speed operation during data acquisition.

1.2.1.3 Order tracking

The spectral smearing of rotary frequency components (i.e. shaft frequencies, gear mesh frequencies etc.) is alleviated by representing the time domain signal in the angle

domain instead. This is usually performed by using computed order tracking (COT) which requires a transducer, such as an optical sensor or a proximity probe, and a shaft encoder or a shaft key to generate a tachometer signal. This tachometer signal contains the instantaneous phase of the shaft over time and is used to convert the time domain vibration signal to the angle domain. However, the quality of the results from COT depends on the accuracy of the measured arrival times of the shaft encoder or key, the resampling or interpolation procedure, the number of pulses per revolution (Fyfe and Munck, 1997) as well as the geometrical characteristics of the shaft encoder itself (Resor et al., 2005). Resor et al. (2005) developed a geometry compensation technique for incremental shaft encoders to reduce the errors in the tachometer signal in slightly varying rotational speeds. Diamond et al. (2016) developed a geometric compensation technique which can be used in arbitrary varying shaft speeds by incorporating Bayesian linear regression theory in the geometry estimation process.

Tacholess order tracking is actively investigated by the engineering community because, (Gryllias and Antoniadis, 2013, Qi et al., 2015, Urbanek et al., 2013, Zhao et al., 2013):

- Tachometers and shaft encoders can be impractical or even impossible to install.
- It is critical to the success of COT to accurately measure the arrival times of the increments on a shaft encoder and therefore high sampling frequencies are required. The high sampling frequencies lead to additional hardware requirements.
- The equipment required to perform order tracking raises the financial costs of performing CBM.
- COT under varying rotational speeds are sensitive to transmission path effects.

In most tacholess order tracking approaches, the instantaneous phase (or frequency) of a shaft is estimated from the measured vibration signal over time. This is subsequently used to resample the signal in constant angular intervals. No additional equipment is required to perform order tracking and the phase distortion, due to the transmission path effects, is removed. Bonnardot et al. (2005) was the first to introduce a tacholess order tracking procedure and Combet and Gelman (2007) improved the technique by automating the harmonic selection process used in the methodology. Both the aforementioned methods are limited to small speed fluctuations which are impractical for gearboxes undergoing large rotational speed changes such as the gearboxes used in draglines and wind turbines. Heyns, Heyns and Zimroz (2012) used a clustering maxima tracking procedure to extract the instantaneous frequency (IF) from the time-frequency spectrum. The low-pass filtered estimate of the IF is used to resample the vibration signal into constant angular intervals.

Urbanek et al. (2013) and Zhao et al. (2013) introduced tacholess order tracking procedures that are appropriate in applications where there are large rotational speed variations. Urbanek et al. (2013) proposed performing maxima tracking on the short-time Fourier transform (STFT) to extract an IF. The signal is sampled to the angle domain so that conventional bandpass filtering can be performed on the signal. The signal is reverse resampled to obtain a mono-component signal in the time domain. Zhao et al. (2013) used the Chirplet transform, due to its superior time-frequency resolution, with maxima tracking. The Vold-Kalman filter, which is a bandpass filter with a time varying centre frequency, is utilised and overcomes the need to use the conventional bandpass filtering with reverse resampling as used by Urbanek et al. (2013). Urbanek et al. (2013) and Zhao et al. (2013) extracted the instantaneous phase from the IF by using the Hilbert transform on the mono-component signal. The instantaneous phase is used to resample the signal at constant angular intervals. Vinson (2014) applied a combination of the techniques proposed by Urbanek et al. (2013) and Zhao et al. (2013) to perform tacholess order tracking.

Gryllias and Antoniadis (2013) used complex shifted Morlet wavelets to extract the complex-valued wavelet coefficients from a vibration signal and used this to estimate the instantaneous angular speed of a rotary machine. This technique performed well and is not sensitive to the choice of wavelet hyperparameters. He et al. (2016) used the discrete spectrum correction technique to order track the vibration signal. The energy centrobaric correction method is used to estimate the instantaneous meshing frequency of a wind turbine's gearbox from the STFT. A three-point convolution was applied to the spectrogram to improve the resolution of the peaks (i.e. meshing frequency components etc.) and attenuate the background noise. The proposed method performed well in estimating the rotational speed of the system as well as order tracking the vibration signal.

1.2.1.4 Synchronous averaging

Time synchronous averaging, applied to a temporal signal with a constant rotational speed, and synchronous or rotation domain averaging, applied to an order tracked signal, are used to filter out components which are non-synchronous with the reference shaft under consideration. The components which are periodic and a constant multiple with respect to a reference shaft are retained in the averaging process (Randall, 2011). McFadden (1985) applied synchronous averaging to a raw and a low-passed vibration signal of a gearbox with a localised gear crack. The synchronous average of the low-passed vibration signal clearly indicates the gear fault, since all frequencies exceeding the gear mesh frequency and its sidebands, which dominate the vibration signal, are

filtered out. This is similar to residual analysis, discussed in section 1.2.1.5, where the dominant components which have little diagnostic information are removed from the signal to highlight localised impulses.

Stander et al. (2002) proposed using a load demodulation normalisation procedure to remove the amplitude modulation caused by fluctuating loads to ensure that the synchronous average is more sensitive to localised faults. The filtered modulation signal, due to the varying loads, is estimated and used to normalise the amplitude modulated vibration signal. The synchronous average of the load normalised signal highlights the presence of a fault which present itself with periodic impulses within the signal. Stander and Heyns (2005) investigated the instantaneous angular speed (IAS) for gear diagnostics in fluctuating operating conditions and found that cyclic stationary loads (i.e. remains the same for each rotation) should be removed by using the load demodulation normalisation technique developed by Stander et al. (2002). Modulation due to non-cyclic stationary (i.e. changes between rotations) loads is attenuated by using synchronous averaging. Stander and Heyns (2006) performed phase domain averaging on the vibration signal to overcome the phase distortion in the order tracking process due to the varying rotational speeds. Phase domain averaging is more effective than synchronous averaging, since it requires significantly fewer averages to attenuate the non-synchronous components.

1.2.1.5 Residual signal analysis

A residual signal represents the departure of the gearbox synchronous average signal from the tooth meshing vibration signal (McFadden, 1987). Dominant periodic non-fault related frequency components, that dominate the synchronous average, are removed so that concealed impulses due to localised faults are easily observed. The residual signal is obtained by only removing the gear mesh characteristics, by removing the gear mesh sidebands as well or by using linear prediction techniques such as autoregressive modelling (Randall, 2011). Heyns, Heyns and Zimroz (2012) investigated artificial neural networks (ANN), discussed in section 1.2.2.3, to generate a residual signal. McFadden (1987) found that the residual signal is significantly more sensitive to incipient localised faults than the synchronous average, since this removes the non-diagnostic information from the signal. The kurtosis of the residual signal is a very powerful diagnostic tool as well (Randall, 2011).

1.2.1.6 Cyclostationary analysis

Cyclostationary analysis was first introduced to analyse communication signals, but it emerged as a powerful diagnostic technique for rotating machines (Antoni, 2009). Cyclostationary signals have statistical properties that vary periodically in the time or in the angle domain, where the latter describes signals generated by rotating machines well (Antoni et al., 2004). Many of the cyclostationary analysis techniques have been developed under specific operating condition requirements and these techniques have to be modified to be applicable for data measured in arbitrarily varying operating conditions (Abboud et al., 2017). Vibration signals that are measured during fluctuating rotational speed conditions can only be cyclostationary if the random speed fluctuation is periodic, stationary or cyclostationary (Antoni et al., 2004).

Fourier analysis is an appropriate analysis tool for first order cyclostationary signals that contain a periodic mean. A second order cyclostationary signal has a periodic autocorrelation function which indicates periodic amplitude and frequency modulation are present within the signal (Antoni et al., 2004). Antoni et al. (2004) successfully applied cyclostationary approaches to signals obtained from an internal combustion engine, a gearbox with a damaged gear and a gearbox with a damaged bearing. Impulses induced by bearing faults have random components due to slip etc. and by using cyclostationary approaches the fault characteristics can be extracted (Antoni et al., 2004). Antoni (2009) wrote an extensive tutorial on cyclostationary tools applied to rotating machines and illustrated many of the concepts of cyclostationary analysis with examples.

1.2.1.7 Envelope analysis

Mcfadden and Smith (1984) investigated the high-frequency resonance technique for bearing fault detection. It is expected that the impulses caused by local damage excite the resonance bands, which does not interfere with the low-frequency content of the signal (i.e. shaft rotation frequencies etc.). The process requires the estimation of a resonance band and the application of a bandpass filter, with a preselected bandwidth, to retain the resonance frequency and its sidebands. The filtered signal is demodulated to obtain an envelope signal or subsequently an envelope spectrum which contains the frequencies at which the resonance band are excited. The deterministic components, such as tooth meshing in gearboxes, also excite the resonance bands and therefore it is difficult to discern between the many spurious components and the component of interest (i.e. the fault frequency) in the envelope spectrum (Abboud et al., 2017). Ocak and Loparo (2005) noted that the shortcomings for this technique are that the

resonance band, the rotational speed and the specifications of the bearing should be known beforehand. Bartelmus and Zimroz (2009*b*) used an envelope signal, obtained from demodulation analysis, as an estimate of the external load applied to the gearbox of a bucket wheel excavator. The vibration signal is more sensitive to varying loads in a damaged condition as opposed to a healthy condition (Bartelmus and Zimroz, 2009*b*).

Wang et al. (2001) proposed a resonance demodulation technique where envelope analysis is performed on the residual signal to remove the strong influence of the deterministic components in the envelope spectrum. The squared envelope signal is sensitive to incipient faults as well as their location.

Abboud et al. (2017) noted that the squared envelope spectrum is one of the most powerful techniques for analysing cyclostationary signals, but it is only applicable for constant operating conditions. An envelope enhancing procedure, which attenuates the source of spurious components in the squared envelope spectrum (the deterministic and cyclo non-stationary components), is proposed. The generalised synchronous average is able to reduce the deterministic components, while an additional filtering step is required to remove the cyclo non-stationary components from the envelope spectrum. This results in a robust fault sensitive squared envelope spectrum.

1.2.1.8 Time-frequency analysis

Time and frequency domain analyses are ill-suited for analysing non-stationary signals when performed separately, but a simultaneous time-frequency analysis provides invaluable insight into the characteristics within a non-stationary signal (Jardine et al., 2006). The short-time Fourier transform (STFT) aims to overcome the fault localisation inabilities of frequency domain analysis by windowing the vibration signal and performing the discrete Fourier transform on each of the resulting vibration segments. The STFT is developed with the assumption that the signal is quasi-stationary within each windowed segment (Jardine et al., 2006). The STFT has a constant frequency and time resolution which are inversely related to one another. This limited resolution is problematic, especially when detecting characteristics at different frequencies and with different lengths in the time domain. Hence, the use of the STFT for fault detection in non-stationary operating conditions is limited (Chaari et al., 2012, Lin and Zhao, 2014, Peng and Chu, 2004).

Bartelmus and Zimroz (2009*a*) extracted the amplitudes of the gear mesh frequency components from the short-time acceleration spectrum as machine condition features. However, the extracted features were unable to separate a gearbox in a good condition from a gearbox in a poor condition, because the features contain operating condition

information as well. The gearboxes in the good and the poor condition were successfully separated by incorporating operating condition information into the fault detection process. The results of the investigation indicate that a damaged gearbox is more susceptible to loads than a healthy gearbox, similarly to Bartelmus and Zimroz (2009*b*).

The Wigner-Ville distribution (WVD), is an alternative to the STFT, which obtains a better time-frequency resolution, but it suffers from severe interference within the spectrum (Jardine et al., 2006, Peng and Chu, 2004). This interference impedes the fault detection process. Stander et al. (2002) used the pseudo-WVD, which overcomes the interference terms of the WVD by applying a frequency smoothing window, to perform gear fault detection. Stander and Heyns (2005) used the smooth pseudo-WVD for gear fault localisation in conjunction with an amplitude ratio to trend the gear fault severity. Baydar and Ball (2000) noted that the pseudo-WVD does not remove all of the interference effects and these interference effects impede the fault detection process, especially of incipient faults. Baydar and Ball (2000) developed the instantaneous power spectrum to overcome the limitations of the bilinear transforms (i.e. WVD). It is possible to detect localised gear faults under stationary and non-stationary operating conditions with the proposed instantaneous power spectrum.

1.2.1.9 Wavelet analysis

Wavelet basis functions are short non-zero oscillations in the time domain that are stretched (known as dilation) and translated. The wavelet transform is a time-scale distribution obtained by convolving the stretched and translated wavelet functions with the vibration signal of interest (Addison, 2002). The time-scale distribution is well suited for non-stationary signal analysis, denoising and singularity detection (Peng and Chu, 2004). Some of the popular basis functions for fault detection in rotary machines are presented in figure 1.1.

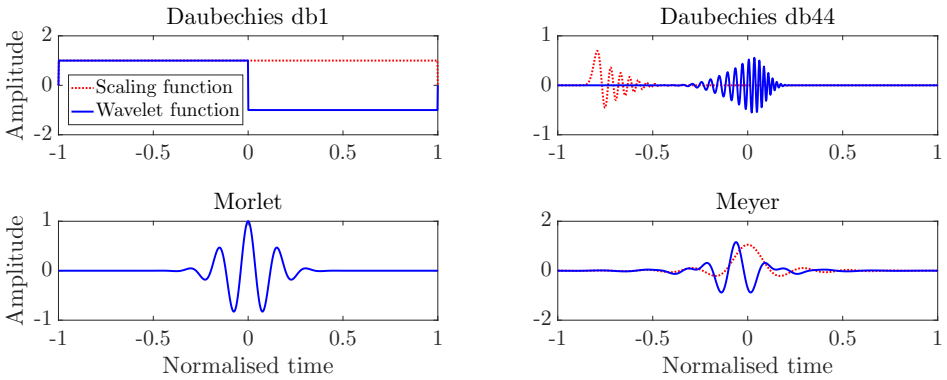


Figure 1.1: Some wavelet basis functions, where the scaling and wavelet function is the low-pass and high-pass filter, respectively.

Three different variations of the wavelet transform are discussed in this section, namely the continuous wavelet transform (CWT), the discrete wavelet transform (DWT) and the wavelet packet transform (WPT). The CWT makes no a priori assumption about the resolution of the scales, which are inversely proportional to frequency, as opposed to the DWT and the WPT. The continuous wavelet coefficients, denoted by W, are obtained from the CWT, calculated from,

$$W(a_1, a_2) = \frac{1}{\sqrt{a_1}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t - a_2}{a_1} \right) dt \tag{1.1}$$

where a_1 and a_2 represent the dilation (or scale) and translation parameters, respectively. The complex conjugate of the mother wavelet basis function and the investigated signal are denoted by ψ^* and $x(t)$, respectively.

Staszewski and Tomlinson (1994) used the CWT for detecting gear damage, since the damage presents itself as non-stationary, abrupt occurrences within the vibration signal. It is difficult to detect small abrupt changes within the resulting wavelet distribution by visual inspection and therefore a similarity analysis with the Mahalanobis distance was conducted which enabled fault detection and severity estimation.

Wang and McFadden (1996) found that the CWT is capable of detecting different gearbox faults simultaneously as opposed to the STFT due to STFT's limited time-frequency resolution. Lin and Qu (2000) used wavelet denoising to identify impulses caused by local bearing and gear damage within noisy vibration signals. Dalpiaz et al. (2000) detected localised gear faults with the CWT. Miao and Makis (2007) extracted the wavelet modulus maxima as features, since it is sensitive to singularities generated by damage on a gear. Wang et al. (2010) investigated incipient fault detection of gearboxes under varying loads. The authors proposed a fault growth parameter, operating on the amplitude of the CWT, which is sensitive to incipient faults and insensitive to varying loads.

The performance of the wavelet transform depends on the correlation between the wavelet basis function and the characteristic of interest such as impulses due to faults etc. and should therefore be selected carefully (Wang and McFadden, 1996). Rafiee et al. (2010) investigated the optimal wavelet basis functions for gear and bearing fault detection with the CWT. The central moment of the wavelet coefficients performs the best for fault detection purposes and the Daubechies db44 is the most similar to bearing and gear vibration signals from the 324 wavelet basis functions that are investigated. However, the Haar wavelet basis function performed the best on bearing and gear fault classification, while the Daubechies db44 performed the worst on the investigated signals. Jedliński and Jonak (2015) investigated the CWT for early fault detection of gearboxes utilising various wavelet basis functions. The Haar and the Meyer basis

functions obtained the best fault classification accuracy, while the Daubechies db4 basis function performed the worst in the investigation by Jedliński and Jonak (2015).

The DWT uses a Dyadic discretization scheme for the scaling and translation parameters (Addison, 2002, Lin and Qu, 2000). A signal is decomposed into approximation (low-pass filtered signal) and detail (high-pass filtered signal) coefficients by using two quadrature mirror filters. The low-pass and the high-pass filters are known as the scaling and wavelet functions, with their impulse responses given in figure 1.1. The orthogonal filters ensure that no redundant information (i.e. only unique information) is contained within the coefficients and that the DWT can be efficiently implemented. In the next step of the DWT, only the approximation coefficients are decomposed into another set of detail and approximation coefficients, which increases the resolution associated with the coefficients containing low frequency information. Each set of detail and approximation coefficients are associated with a specific band in the spectrum of the signal. Sung et al. (2000) investigated the use of the DWT and STFT for localised gearbox fault detection, where the DWT had superior fault localisation capabilities as opposed to the STFT on the investigated data. The DWT has also been successful in internal combustion fault detection (Wu and Liu, 2008) and detecting bearing faults (Purushotham et al., 2005).

The WPT is very similar to the DWT, except that a finer frequency resolution is allocated to the higher frequencies by filtering the detail coefficients as well (Addison, 2002, Theodoridis and Koutroumbas, 2009). This is more appropriate than the DWT for cases where the wavelet coefficients need to be sensitive to changes in high-frequency content (Theodoridis and Koutroumbas, 2009). The WPT has been successful in detecting localised (Shao et al., 2014, Wan et al., 2016), distributed gear faults (Shao et al., 2014), internal combustion engine faults (Wu and Liu, 2009) and gear, bearing and shaft faults (Yen et al., 2000). See Addison (2002) for more details on wavelet analysis.

1.2.2 Machine learning-based fault detection

The machine learning field emerged from computer science and aims to develop models which recognise the underlying patterns within data and then uses this knowledge to make intelligent predictions. Features are extracted from the measured data to separate the characteristics of interest from the irrelevant characteristics. The models, optimised on the extracted features located in high-dimensional spaces, are prone to problems (i.e. performance, memory etc.) due to the curse of dimensionality (Bishop, 2006). The features may contain redundant information which results in more hardware and

software requirements (i.e. data storage, processing capabilities, computational time etc.). The problems associated with large dimensional feature spaces are circumvented with feature selection and dimensionality reduction techniques. The optimised model is subsequently used to make predictions from features extracted from the new data.

1.2.2.1 Feature extraction

The measured data contain characteristics pertaining to the operating conditions, the effects of damage and other phenomena that may not be of interest to the fault detection process. It is important to extract features that capture the characteristics of interest so that the optimised model is only sensitive to changes in those characteristics. Common features for fault detection is obtained from time domain statistics like RMS (Zimroz et al., 2014), spectral analysis (Gryllias and Antoniadis, 2012, Yiakopoulos et al., 2011), short-time analysis techniques (Bartelmus and Zimroz, 2009*a*, Zimroz and Bartkowiak, 2013), processing the wavelet transform coefficients (Qi et al., 2015, Wan et al., 2016) and the intrinsic mode functions (IMF) obtained from the empirical mode decomposition (EMD) or the ensemble EMD (EEMD) (Yang and Wu, 2015).

Operating condition information such as the power generated by wind turbines (Zimroz et al., 2014), the envelope of the vibration signal (Bartelmus and Zimroz, 2009*b*), the rotational speed of a gearbox shaft (Bartelmus and Zimroz, 2009*a*, Heyns, Heyns and De Villiers, 2012) and smart feature selection techniques (Vinson, 2014) have been incorporated into the fault detection process to make it more robust in fluctuating operating conditions.

Linear scaling methods such as standardisation ensure that the features are of the same order of magnitude and can be applied to data to alleviate potential model optimisation problems (Bishop, 1995). Feature selection approaches to reduce the dimensionality of the feature space by discarding features that do not improve the performance of the model. This is done by selecting features that enhance the separation between the classes (or machine conditions in the feature space) and by removing features which contain redundant information. Class separability measures such as divergence, scatter matrices and feature correlation measures are used as criteria for selecting or discarding features in supervised learning problems (Theodoridis and Koutroumbas, 2009). It is only possible to select features based on class separability measures if historical fault data is available. In the next section, data-driven dimensionality reduction techniques are investigated.

1.2.2.2 Dimensionality reduction techniques

Dimensionality reduction techniques transform the current feature space to a lower dimensional feature space by retaining most of the information content of the signal with respect to some criterion. The intrinsic dimension of the feature space can be significantly lower than the apparent or raw dimension of the features. This has a significant impact on the optimisation of the models used to describe the feature space (Bishop, 1995). Principal component analysis is a popular dimensionality reduction technique which transforms the features to its principal axes so that each dimension is independent. This allows the removal of features that are highly correlated with little loss in information content (Bishop, 2006).

Van Der Maaten et al. (2009) completed an extensive study on dimensionality reduction techniques on artificial as well as natural datasets. The non-linear dimensionality reduction techniques such as the auto-encoder and the local tangent space analysis performed well on highly non-linear artificial datasets distributed along complex manifolds. In contrast, natural datasets rarely have these highly non-linear manifolds and principal component analysis (PCA), a linear reduction procedure, outperforms all the non-linear dimensionality reduction methods on the natural datasets.

Zimroz and Bartkowiak (2013) investigated multivariate statistics for gearbox diagnostics under non-stationary operating conditions. PCA and conical discriminant analysis are investigated on a bucket wheel excavator operating under non-stationary conditions. The amplitudes of the gear mesh components are used as machine condition features which resulted in a fifteen dimensional feature space with an intrinsic dimensionality of two. The healthy and damaged gearbox are linearly separated in the two dimensional principal component feature space.

Shao et al. (2014) investigated PCA and kernel-based PCA as dimensionality reduction procedures for WPT machine condition features to detect different gear faults in constant operating conditions. The new dimension is chosen so that the retained features consist of 95% of the original information content. The PCA does not have linear separation in the feature space between the different machine conditions, but kernel-based PCA does at the cost of selecting an optimal parameter.

Wan et al. (2016) performed an investigation on different dimensionality reduction techniques (PCA, kernel-based PCA, Isomap, local linear embedding and Laplacian eigenmaps) on gear crack sensitive wavelet features, using the Daubechies db44 mother wavelet, with different statistical models. PCA, a linear dimensionality reduction technique, does not require any hyperparameter optimisation, while an exhaustive grid search is completed to obtain the optimal hyperparameters for the non-linear dimen-

sionality reduction techniques. Statistical models such as the Naive-Bayes classifier, linear and quadratic discriminant have the highest classification accuracy on the data transformed with PCA as opposed to the non-linear methods.

If all of the historical fault data of all of the possible machine conditions are available, then the parameters of the non-linear methods can be optimised to perform well on the fault detection task. For novelty detection problems, where only data of one condition is available, an optimal parameter for all possible conditions cannot be selected.

1.2.2.3 Machine learning methods

Machine learning methods are used in diagnostics to obtain a model that describes the relationship between the extracted features and the condition of the machine etc. The machine learning model is optimised in the training phase, tested with labelled data not used during training in the validation phase and tested with new (unlabelled) data (Theodoridis and Koutroumbas, 2009).

Three model training or optimisation philosophies are usually used namely, supervised, unsupervised and novelty detection (Theodoridis and Koutroumbas, 2009). Supervised learning makes the assumption that correctly labelled historical fault data of all the possible conditions, that need to be classified, are available. This is an ineffective method for practical diagnostic strategies, since it is expensive to acquire historical fault data, especially for new machines. Unsupervised learning does not label any of the data and hence it identifies clusters of data that have similar characteristics. The labels are assigned manually by an expert if they are required. Finally, novelty detection approaches only require labelled data of one class (or condition) and the objective of the model is to identify whether new data is part of the labelled dataset or not (Theodoridis and Koutroumbas, 2009).

Pimentel et al. (2014) performs a review of novelty detection approaches for various applications and the following categories are identified: probabilistic-, distance-, domain-, reconstruction- and information-based approaches. Some of these categories are discussed in the subsequent sections.

Reconstruction-based approaches

Reconstruction-based approaches map the multidimensional feature space to an output space which can be a set of class labels (classification problem) or an output function (regression problem). Artificial neural networks (ANN) and radial-basis functions are some examples of reconstruction-based approaches.

An ANN is a mathematical model inspired by biological neural networks. The most common implementation form is a feed-forward neural network that is optimised by minimising the least squares error function (Bishop, 1995, 2006). An ANN can even be used to fit likelihood functions if it is viewed from a probabilistic perspective (Bishop, 2006). ANNs are successfully implemented for internal combustion engine diagnostics (Wu and Liu, 2008, 2009) and for bearing diagnostics (Samanta et al., 2003, Unal et al., 2014). Unal et al. (2014) optimised the ANN with genetic algorithms (a global optimisation algorithm), due to the non-convex optimisation space being prone to the existence of many local minima. The ANN is able to accurately predict the condition of the bearings from the features extracted from the result of envelope analysis.

Yang and Wu (2015) proposed a gear diagnostics procedure involving features extracted from the IMFs obtained from the EEMD and performing the Hilbert-Huang transform. The dimensionality of the extracted features is reduced by utilising PCA and subsequently optimising a neural network on the lower dimensional features.

Qi et al. (2015) compared the abilities of the ANN and radial-basis function to classify misalignment faults using WPT features, and even though the features are non-linearly separated, both obtained satisfactory results.

Heyns, Heyns and Zimroz (2012) investigated the use of residual and discrepancy analysis for gearbox fault detection under non-stationary operating conditions. A residual signal is generated from autoregressive modelling and an ANN, respectively. A discrepancy signal, which represents the deviation in the expected response of a healthy system, is generated from the envelope of the residual signal and is treated further in section 1.2.2.3. It was found that the autoregressive model is incapable of detecting faults in non-stationary operating conditions, while the ANN is able to detect the presence of localised faults.

Even though the neural network approaches provide sufficient flexibility to fit arbitrary functions, neural networks have some drawbacks (Bishop, 1995):

- ANNs are computationally expensive to optimise and can easily be trapped in local minima by utilising gradient-based optimisation approaches.
- The flexibility of the ANNs can easily result in overfitting.
- Multi-modal target functions lead to suboptimal results if the least squares error function is used. Bishop (2006) proposed using mixture density networks instead for modelling multi-modal target functions.

Domain-based approaches

Domain-based approaches aim to model the boundary separating the classes (or conditions) and then classify new data according to its position with respect to the boundaries. Support vector machines (SVM) are the most popular example of domain-based approaches. Classes that are non-linearly separated in the original feature space, are transformed with kernel functions into a high-dimensional feature space where the classes are linearly separated. The kernel function and its parameters influence the performance of the SVM (Bishop, 2006). Unlike the ANN, the SVM cost function is convex.

Samanta et al. (2003) compared the capabilities of ANNs and SVMs for bearing diagnostics utilising time domain features such as RMS and central moments. The SVM outperformed the ANN in terms of classification results.

Jedliński and Jonak (2015) investigated the use of ANN and SVM for gearbox fault detection with CWT coefficients. The ANN and SVM trained on the CWT features outperformed the ANN and SVM trained on raw data, while the SVM outperformed the ANN in each case.

Gryllias and Antoniadis (2012) noted that fault specific data is scarce and expensive to acquire for industrial environments. A physics-based bearing diagnostic methodology is developed where the measured data is tested on a model trained on simulated data. A bearing model is used to generate the simulated bearing fault data. A two-stage SVM approach is used, where the first stage determines whether a fault is present or not and the second stage determines the characteristic of the fault. This performed well on experimental as well as industrial data.

Fernández-Francos et al. (2013) investigated SVMs for bearing diagnostics with only healthy data available (i.e. novelty detection). A one-class SVM is applied to features extracted from the envelope spectrum of the vibration data during a bearing fatigue test. It is possible to detect a novelty (or a deterioration in machine condition) early with this approach.

Probabilistic approaches

Probabilistic approaches aim to create a parametric (a Gaussian distribution etc.) or non-parametric (kernel density estimators etc.) model of the density that generated the data. The flexibility of the parametric approaches is limited to the choice of the model, while the non-parametric methods can provide arbitrary flexibility. However, a

non-parametric approach requires the training data to be stored, it can be expensive to evaluate the function, it requires extensive hyperparameter optimisation and it scales poorly with an increase in feature space dimensionality in many cases (Bishop, 2006).

Parametric distributions are investigated in this study. A Gaussian distribution,

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{1.2}$$

is motivated from the central limit theorem and its convenient mathematical properties, but it is limited to data with uni-modal distributions (Bishop, 2006). Features which exhibit multi-modal behaviour are poorly modelled with a Gaussian distribution. The mean $\boldsymbol{\mu}$ and the covariance $\boldsymbol{\Sigma}$ of the data $\boldsymbol{X}$, with dimensionality D, control the form of the distribution. A single observation of the dataset $\mathbf{X}$ is denoted by $\mathbf{x}$. A mixture of weighted Gaussian distributions provide sufficient flexibility to model data with multi-modal densities. Gaussian mixture models (GMM) have the capability to model any distribution to arbitrary accuracy if a sufficient number of Gaussian mixture components are used (Bishop, 2006).

Both the aforementioned distributions make the assumption that the data are identically and independently distributed (i.i.d) which ignores sequential patterns within the data when optimising and evaluating the model. In the speech recognition field strong sequential patterns exist within the data and Hidden Markov models (HMM) are regularly used to model the features extracted from audio signals (Bishop, 2006, Gales and Young, 2008, Rabiner, 1989). A stochastic process follows a Markov process if the current state only depends on a finite number of previous states. The only evidence of the unobserved Markov process between the discrete hidden states in a HMM is the noisy observations (or features). The graphical model in figure 1.2 represents a HMM over three time steps.

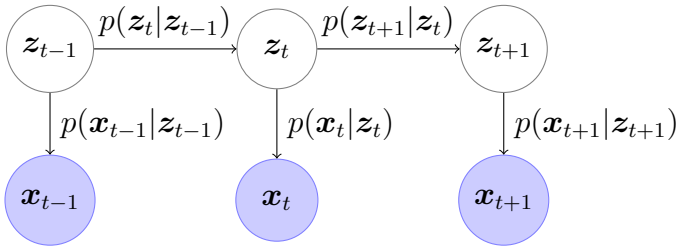


Figure 1.2: A graphical model of a HMM with the observed data and latent states denoted by $\boldsymbol{x}$ and $\boldsymbol{z}$ respectively. The circles indicate a random variable, with a highlighted circle indicating that the outcome of the variable is observed.

The latent variable $\boldsymbol{z}_t$ denotes which hidden state generated the observed data or features $\boldsymbol{x}_t$ at time increment t. The observation density for a specific hidden state can

be a Gaussian distribution, a GMM etc. The likelihood function of a HMM does not only take the coordinates of the features in the feature space into account, but also the transition between the hidden states. Hence, the HMM is more powerful than the GMM and the Gaussian distribution in the case of multi-modal data with sequential characteristics.

In the case of novelty detection, the likelihood function of only one condition is available and therefore it is impossible to use Bayes' theorem to evaluate the posterior probability of the condition or class. Bishop (1994) proposed a procedure for utilising neural networks for novelty detection procedures. The ANN models the likelihood of the known class, and then the probability density function of the novelty class is chosen broad or uninformative. This information is used with Bayes theorem to determine whether the new data are a novelty or not.

Marwala et al. (2006) investigated HMMs and GMMs for bearing fault diagnosis. The time domain data are divided into small segments and then a box counting multi-scale fractal dimension technique is applied from which a GMM and a HMM is trained separately. Principal component analysis (PCA) is applied to the extracted features. The HMM performs better than the GMM on the investigated data, but the GMM requires less computational resources during the optimisation and evaluation phase.

Ocak and Loparo (2005) proposed a two-stage HMM bearing diagnostic methodology. Quasi-stationary windows are extracted from an amplitude demodulated vibration signal in the first stage. The linear prediction coefficients are extracted from the aforementioned signal. The linear prediction coefficients of a healthy gearbox are modelled with a HMM and then new data are classified as healthy or damaged according to its likelihood and a predetermined threshold. Wavelet packet coefficients are extracted in the second stage and modelled with a separate HMM for each fault case using supervised learning. If the bearing is damaged according to the first stage, the second stage determines which damage mode is present.

Purushotham et al. (2005) investigated a discrete observation HMM optimised on mel-frequency cepstral coefficients for bearing diagnostics. The features are extracted from the windowed time domain data and used to train a HMM for each bearing condition.

Miao and Makis (2007) proposed a two-stage fault detection process for gear diagnostics. The first stage aims to determine whether the gearbox is damaged or not and the second stage aims to determine the characteristic of the fault (i.e. cracked gear tooth etc.). The likelihood of the parameters of the HMM can be compared to a threshold which determines whether the gearbox is healthy or not. In the second stage, the condition of the gearbox is inferred by selecting the condition which has the

highest likelihood when evaluating its HMM. A parameter re-estimation procedure is proposed to update the parameters of the HMM as the gearbox deteriorates.

Timusk et al. (2008) investigated gearbox diagnostics in the presence of fluctuating operating conditions with different features and novelty detection algorithms. Timusk et al. (2008) combined many classifiers for decision making which performed better than using each classifier separately.

Heyns, Godsill, De Villiers and Heyns (2012) investigated an adaptive autoregressive model to generate a residual signal for gear fault detection which is robust to fluctuating operating conditions. An autoregressive model is implemented for an operating condition state and the final prediction of the vibration signal is a weighted average of all of the autoregressive models. The weights are determined by using Bayesian statistics where the weights reflect the suitability of a specific autoregressive model to predict the signal value given the current operating condition. Fault localisation and trending are possible with numerical and experimental data.

Figueiredo et al. (2014) used a GMM to model the data acquired from a healthy bridge during different weather conditions. The minimum Mahalanobis distance is trended over time, where a sequence of outliers indicates that damage is present in the bridge in the presence of various weather conditions.

The aim of discrepancy analysis is to detect a deviation in the expected behaviour of a dataset according to an optimised machine condition model. The deviation is attributed to a change in machine condition and by investigating the characteristics of the discrepancy signal, the characteristic of a fault is inferred. The advantage of discrepancy analysis is that only data of a gearbox in a healthy condition is required to detect faults. Heyns, Heyns and Zimroz (2012) used an ANN for modelling the vibration signal of a healthy gearbox. The discrepancy signal is generated from the envelope of the error between the signal and the prediction of the ANN. Localised faults are successfully detected by this approach.

Heyns, Heyns and De Villiers (2012) investigated the use of discrepancy analysis for gearbox fault detection in fluctuating operating conditions. The authors implemented a windowing scheme on an order tracked vibration signal to extract signal segments which are used with operating features (the average speed) to train GMMs. A discrepancy signal is generated from the data in the form of a negative log-likelihood (NLL) signal. The authors incorporated operating condition information into the fault detection process to increase the robustness of the discrepancy signal in non-stationary operating conditions. The discrepancy signal, generated by only using the windowed segments, is able to locate the faults and distinguish between different fault severities.

Vinson (2014) proposed a gear fault diagnostic methodology using smart features and machine learning technology. A HMM is optimised on operating condition sensitive features from which a GMM is trained on machine condition features for each hidden state of the HMM. Hence for a specific window in time, the operating condition state is classified first and then the appropriate GMM is selected to generate a discrepancy measure. The purpose of this is to make the discrepancy signal more robust to fluctuating operating conditions.

1.3 Scope of the research

Rotary machine diagnostics in the presence of fluctuating operating conditions remain an important and challenging task. Bartelmus and Zimroz (2009*a*,*b*) and Zimroz et al. (2014) developed fault detection methodologies for large machines, such as bucket wheel excavators as well as wind turbines, in the presence of fluctuating operating conditions. Even though the developed approaches perform well with the fault detection task and are simple to implement, it is difficult to ascertain the location and the nature of the damage (i.e. localised gear fault etc.). This information is invaluable for asset integrity management since the remaining useful life of the machine depends on the component that is damaged and the nature of the damage. Ultimately, this information provides support for maintenance decision-making tasks such as planning downtime, component procurement etc.

The outcomes of this study is to develop a fault diagnostic methodology that,

- Is able to detect, locate and trend (i.e. detecting progressing damage) localised gear faults from vibration data measured in fluctuating operating conditions.
- Is cost-effective by not relying on equipment which increases the cost and restricts its applicability (i.e. tachometers and shaft encoders).
- Only depends on data that are measured during normal operation. Historical fault data are expensive to acquire, especially for new machines.
- Does not require experts to interpret the results.

Vibration transducers such as accelerometers are widely available, easily mounted on rotary machines such as gearboxes and the measured vibration signal is rich with diagnostic information. This motivates extracting diagnostic information from a vibration signal, measured from an accelerometer.

Vibration signals measured from rotary machines such as gearboxes are inherently

angle-cyclostationary which motivates processing the signals in the angle domain (Antoni et al., 2004). However, converting from the time to the angle domain requires knowledge about the instantaneous phase of a reference shaft over time, which is usually obtained from a tachometer signal. Single pulse per revolution signals (from a shaft key and proximity probe) do not have sufficient resolution to measure large angular acceleration. In contrast, shaft encoders provide sufficient resolution for measuring large angular accelerations. However, digital shaft encoders are not easy to install and are not durable in harsh conditions. Zebra tape shaft encoders are relatively easy to install in most cases and its geometrical errors can be compensated for, but its performance may deteriorate in harsh conditions. The shaft encoders require large sampling frequencies as well (Fyfe and Munck, 1997), which results in more hardware and storage requirements. This motivates using tacholess order tracking methods to obtain a cost-effective fault diagnostic methodology.

A computationally efficient tacholess order tracking method is proposed in this study which is robust to noise and other unimportant frequency components. This procedure is based on the methods by Urbanek et al. (2013), Zhao et al. (2013) and Vinson (2014). The maxima tracking process is used to isolate an instantaneous frequency of interest and is vital to the success of the tacholess order tracking method. The maxima tracking process, that is used by Urbanek et al. (2013) and Vinson (2014), is improved by using a probabilistic approach to incorporate acceleration information in the process. This increases the robustness of the tacholess order tracking method in the presence of noise and large angular accelerations. The two generations of the Vold-Kalman filter are investigated and compared in this study as well. In the final results, the second generation Vold-Kalman filter is used, since it seems slightly more robust and also less prone to errors according to Blough (2003) and Pan and Lin (2006). The first generation Vold-Kalman filter was investigated by Zhao et al. (2013) and Vinson (2014). The proposed tacholess order tracking method is validated on seven datasets. The results from the proposed maxima tracking method is compared to the maxima tracking method proposed by Urbanek et al. (2013).

Bartelmus and Zimroz (2009*a*,*b*), Chaari et al. (2012) and Zimroz et al. (2014) have emphasised that operating condition information must be incorporated into the fault detection process to ensure that the cause of changes within the machine condition features can be ascribed to a change in machine condition with confidence. However, Timusk et al. (2008) and Heyns, Heyns and De Villiers (2012) do not find an improvement in their results by incorporating operating condition information in the fault detection process. Timusk et al. (2008) attributes this to similarities between the investigated operating condition modes. Stander and Heyns (2005) proves that synchronous averaging attenuates the amplitude modulation generated from non-cyclic

stationary loads. Heyns, Heyns and De Villiers (2012) performs synchronous averaging on the discrepancy signal generated in non-cyclic stationary load conditions, which may explain why the operation condition information did not improve the robustness of the fault diagnostic methodology. Vinson (2014) is able to detect faults in the synchronous average of the discrepancy signal, with only a correct operating condition classification of 53%. It is concluded from this that the synchronous averaging process makes the fault diagnostic methodology more robust to fluctuating operating conditions. However, operating condition information needs to be incorporated into the fault diagnostic methodology, since:

- Operating condition information helps to determine how the machine condition information or features are expected to react under those operating conditions. It is expected that this becomes more important when machines like draglines, which have large varying operating states, are monitored.

- The operating condition information helps to determine whether predictions can be made from the current operating condition so that erroneous machine condition predictions are avoided.

In this study, a fault diagnostic methodology is investigated by using machine learning techniques. The data are modelled so that the expected response of the system in a specific machine condition can be determined. Incorporating prior knowledge (i.e. fault historic data) to diagnose a machine condition arises naturally in machine learning models, which the classical methods such as residual, time-frequency, cyclostationary etc. analysis cannot provide. Hence, it is possible to obtain a robust fault diagnostic methodology by using expert knowledge, advanced processing techniques, large amounts of data and machine learning models to infer the condition of the machine.

A novelty detection approach is considered in this study, since it does not constrain the fault diagnostic methodology to only detect specific faults which are available in the historical fault data. The physics of the system do not have to be modelled with novelty detection approaches, which makes a novelty detection approach simpler to implement than physics-based approaches.

Machine condition features, which are sensitive to machine condition changes within the system, are extracted from the vibration signal. However, the machine condition features can be sensitive to operating condition changes as well. This necessitates incorporating operating condition information into the fault diagnostic methodology. Since the operating conditions cannot be measured, it is extracted from representative operating condition features. The operating condition features are modelled with a HMM and the operating condition state is predicted from the aforementioned HMM

on new data. The machine condition features, extracted at the same time step as the operating condition features, receive the same operating condition state or label as the respective operating condition features. The machine condition features labelled with a specific operating condition state are modelled with a HMM machine condition model. If the operating condition classification process is correct, the output of the machine condition model is approximately operating condition independent. This emphasises that the operating condition features must be insensitive to machine condition changes to ensure that the operating condition classification is correct.

A probabilistic approach is used to combine the information from the different machine condition models and the operating condition model to generate a discrepancy signal. The discrepancy signal is post-processed with a synchronous averaging process, similarly to Heyns, Heyns and De Villiers (2012) and Vinson (2014), but additional post-processing techniques are proposed in this study:

- A second averaging process is proposed between measurements which renders a more robust diagnostic indicator.
- The condition of the machine and the location of the fault is inferred by using statistical theory to generate an alarm threshold.
- A gear-pinion discrepancy distribution is proposed to evaluate the condition of a two shaft gearbox.
- A healthy-damaged decomposition of the discrepancy signal is proposed for easy fault trending.

The fault diagnostic methodology is validated on data from a numerical gearbox model and data from experiments conducted in a laboratory.

1.4 Layout of document

The experimental setup and the numerical gearbox model, used to generate data in this study, are presented and all of their characteristics are discussed in chapter 2. The investigated operating conditions and damage modes, with the resulting vibration signals, are presented in the chapter as well.

In chapter 3, the proposed tacholess order tracking method is introduced and investigated on the vibration data presented in chapter 2 as well as additional datasets. The proposed maxima tracking algorithm is critically compared to the benchmark maxima tracking algorithm, proposed by Urbanek et al. (2013).

In chapter 4, the fault diagnostic methodology is introduced and discussed in detail. Chapter 4 starts with an overview of the fault diagnostic methodology whereafter each aspect is discussed in detail and motivated if necessary. The experimental data in chapter 2 is used to illustrate many of the concepts introduced in chapter 4.

The proposed fault diagnostic methodology, introduced in chapter 4, and the tacholess order tracking method, introduced in chapter 3, are used to obtain a cost-effective diagnostic methodology. The ability of the fault diagnostic methodology to perform fault detection, localisation and fault trending is evaluated in chapter 5 on the numerical and experimental datasets introduced in chapter 2.

The work is concluded in chapter 6 and recommendations are made for future work in this field of study. Two appendices are included at the end of this dissertation. Some auxiliary information are provided in chapter A for the experimental setup and some auxiliary results, which support the statements in the dissertation, are included in chapter B.

Chapter 2 Gearbox vibration data

The performance of a diagnostic methodology needs to be tested on data generated in an environment where the condition of the system is known and the required equipment are available. Numerical and experimental data are used in this study to evaluate the abilities of the proposed diagnostic methodology to detect, locate and trend damage on the gear by only using vibration data.

A numerical gearbox model is simple and feasible to construct, computationally feasible to solve and is created in a controlled environment. The aforementioned properties make the numerical gearbox model ideal to use in the early stages of developing a diagnostic methodology. There is no uncertainty pertaining to the instantaneous operating conditions, machine condition (i.e. the degree of damage) and the properties of the system (resonances etc.). The similarities of the numerical gearbox model and an actual gearbox depend on the modelling assumptions. The gearbox model, presented in section 2.1, is only used to test the feasibility of the diagnostic methodology and therefore many simplifications such as the gear mesh stiffness, damage progression etc. are permissible. The characteristics of the model such as the gearbox type, the dimensions, operating conditions etc. differ from the experiments, because it is only used as another source of data.

The experimental setup and the measured data are presented in section 2.2. Two accelerated life tests are performed experimentally in this study, with the focus placed on gear failure due to localised faults. The experimental data provides insight on the performance of the diagnostic methodology on data from a real machine.

2.1 Numerical gearbox model

The vibration data from a numerical gearbox is generated from an eight degree of freedom lumped mass gearbox model used by Chaari et al. (2008, 2012). The model of the gearbox in figure 2.1 contains a pinion and a gear that are connected through shafts

to an electric motor and an external load or machine, respectively. The original articles

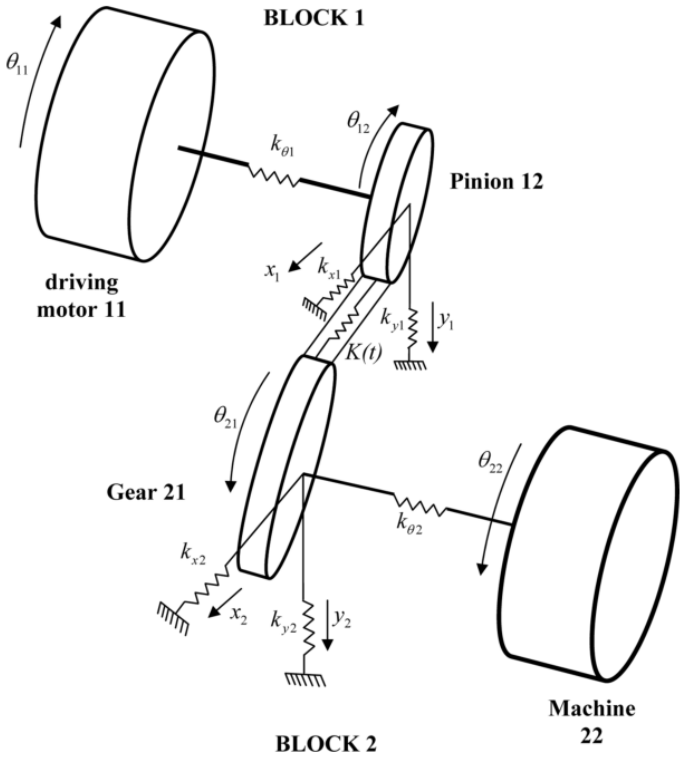


Figure 2.1: Eight degree of freedom lumped mass gearbox model (Chaari et al., 2012).

by Chaari et al. (2008, 2012) contains printing errors within the equations describing the gearbox model. The printing errors are corrected in this study.

The vibration data from the model in figure 2.1 is obtained from solving the equation of motion,

$$\boldsymbol{M}\ddot{\boldsymbol{Y}}(t) + \boldsymbol{C}\dot{\boldsymbol{Y}}(t) + \boldsymbol{K}(t)\boldsymbol{Y}(t) = \boldsymbol{F}(t) \tag{2.1}$$

describing the dynamic response of the discrete system for the applied external loads $\boldsymbol{F}$. The mass, damping and stiffness characteristics of the system are described by the matrices denoted by $\boldsymbol{M}$, $\boldsymbol{C}$ and $\boldsymbol{K}(t)$, respectively. The stiffness matrix is a function of time due to the varying gear mesh stiffness discussed later in the text. The displacement degree of freedom of the system is,

$$\boldsymbol{Y}(t) = \{x_1,\ y_1,\ \theta_{11}, \theta_{12}\ x_2,\ y_2,\ \theta_{22},\ \theta_{21}\} \tag{2.2}$$

where x_i, y_i and θ are the horizontal, vertical and angular displacements of the different blocks indicated in figure 2.1. The last two columns in equation (2.2) are interchanged from the original article to make the subsequent corrections easier. The characteristics of the spur gearbox with the notation in table 2.1 are the same values used by Chaari et al. (2012). The mass matrix associated with the gearbox system only contains

Table 2.1: Characteristics of the gearbox in the model, where i is p or g which denotes whether the characteristic applies to the pinion or the gear.

Characteristic	Symbol	Pinion (p)	Gear (g)	Unit
Number of teeth	N_i	20	40	–
Base circle radius	r_{bi}	28.19	56.38	mm
Pressure angle	α	20^o		degrees
Contact ratio	c_{CR}	1.6		–

non-zero terms on the diagonal, where the diagonal elements are given by,

$$\texttt{diag}(\boldsymbol{M}) = \begin{bmatrix} m_1 & m_1 & I_{11} & I_{12} & m_2 & m_2 & I_{22} & I_{21} \end{bmatrix} \tag{2.3}$$

with each component described in table 2.2. The mass of block 1 (m_1) and block 2 (m_2) refer to the mass of the pinion-bearing and gear-bearing blocks, respectively.

Table 2.2: Mass characteristics of the gearbox model

Characteristic	Symbol	Value	Unit
Mass of block 1	m_1	1.8	kg
Mass of block 2	m_2	2.5	kg
Moment of inertia of the motor	I_{11}	0.0043	kg.m^2
Moment of inertia of the pinion	I_{12}	2.7×10^{-4}	kg.m^2
Moment of inertia of the machine	I_{22}	0.0045	kg.m^2
Moment of inertia of the gear	I_{21}	0.0027	kg.m^2
Mass of the pinion	m_p	0.6	kg
Mass of the gear	m_g	1.5	kg

The bearings, the shafts and the gear teeth in contact during meshing are modelled as springs. The stiffness matrix of the system, denoted by $\boldsymbol{K}(t)$, describes the relationship between the displacements and the forces applied to the different degrees of freedom of the system. The stiffness matrix contains a static and a dynamic contribution,

$$\boldsymbol{K}(t) = \boldsymbol{K}_s + \boldsymbol{K}_t(t) \tag{2.4}$$

where the static and the dynamic contribution are denoted by $\boldsymbol{K}_s$ and $\boldsymbol{K}_t(t)$ respectively. The static contribution describes the time-invariant stiffness characteristics the

system, like the bearings and the shafts, and is given by,

$$\boldsymbol{K}_s = \begin{bmatrix} k_{x1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{y1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{\theta 1} & -k_{\theta 1} & 0 & 0 & 0 & 0 \\ 0 & 0 & -k_{\theta 1} & k_{\theta 1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{x2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_{y2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & k_{\theta 2} & -k_{\theta 2} \\ 0 & 0 & 0 & 0 & 0 & 0 & -k_{\theta 2} & k_{\theta 2} \end{bmatrix} \tag{2.5}$$

where the components of the static stiffness matrix are given in table 2.3. Equation (2.5) is corrected from the article which contains a negative definite stiffness matrix. The dynamic contribution is attributed to the periodical varying gear mesh stiffness

Table 2.3: Stiffness characteristics of the gearbox model (Chaari et al., 2012)

Characteristic	Symbol	Value	Units
Bearing translational stiffness	k_{xi}, k_{yi} for $i = 1, 2$	10^8	N/m
Torsional stiffness of shafts	$k_{\theta i}$ for $i = 1, 2$	10^5	N×m/rad
Gear-mesh stiffness			
- Maximum	$\max(k_{gm}(t))$	2.1×10^8	N.m
- Minimum	$\min(k_{gm}(t))$	0.81×10^8	N.m
- Mean	$\bar{k}_{gm}$	1.65×10^8	N.m

as the teeth move in and out of mesh. The contact ratio in table 2.1 indicates that for 60% of the meshing cycle two teeth are in contact, while only one tooth is in contact during the remaining 40% of the cycle. The kinematic relationship between the different degrees of freedom in equation (2.2), due to the meshing of the teeth, is derived by considering the line of action,

$$\delta(t) = (x_1 - x_2)\sin(\alpha) + (y_1 - y_2)\cos(\alpha) + \theta_{12} r_{b12} + \theta_{21} r_{b21} \tag{2.6}$$

where α, x_i, y_i and θ_i denotes the pressure angle and the displacement in the horizontal, vertical and angular directions, respectively. The displacement parallel to the contact force (or line of action) is denoted by $\delta(t)$ and is decomposed into translational and angular displacements. This is used to obtain the stiffness contribution of the dynamic component, due to varying gear mesh stiffness, to the total stiffness matrix. The dynamic component of the stiffness matrix is decomposed into two parts,

$$\boldsymbol{K}_t(t) = k_{gm}(t)\boldsymbol{S} \tag{2.7}$$

where $k_{gm}(t)$ is the gear mesh stiffness at time t and $\boldsymbol{S}$ is a non-dimensional, time-invariant matrix in the form of,

$$\boldsymbol{S} = \begin{bmatrix} s_3 & s_5 & 0 & s_7 & -s_3 & -s_5 & 0 & s_9 \\ s_5 & s_4 & 0 & s_6 & -s_5 & -s_4 & 0 & s_8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_7 & s_6 & 0 & s_{10} & -s_7 & -s_6 & 0 & s_{12} \\ -s_3 & -s_5 & 0 & -s_7 & s_3 & s_5 & 0 & -s_9 \\ -s_5 & -s_4 & 0 & -s_6 & s_5 & s_4 & 0 & -s_8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_9 & s_8 & 0 & s_{12} & -s_9 & -s_8 & 0 & s_{11} \end{bmatrix} \tag{2.8}$$

which couples the force degrees of freedom to the displacement degrees of freedom due to meshing. The components of equation (2.8) are given in table 2.4. The pressure

Table 2.4: Coefficients of matrix $\boldsymbol{S}$ in equation (2.8) which is associated with the dynamic component of the total stiffness of the gearbox model.

s_1	$\sin(\alpha)$	s_2	$\cos(\alpha)$	s_3	$\sin(\alpha)^2$	s_4	$\cos(\alpha)^2$
s_5	$\sin(\alpha)\cos(\alpha)$	s_6	$r_{b12}\cos(\alpha)$	s_7	$r_{b12}\sin(\alpha)$	s_8	$r_{b21}\cos(\alpha)$
s_9	$r_{b21}\sin(\alpha)$	s_{10}	r_{b12}^2	s_{11}	r_{b21}^2	s_{12}	$r_{b12}r_{b21}$

angle of the two meshing gears are denoted by α, the radius of the base circle for the gear and the pinion are denoted by r_{b21} and r_{b12} respectively. The values of these characteristics are given in table 2.1. Even though equation (2.8) is in essence the same as the original articles, it is different due to the new degrees of freedom in equation (2.2). It was found that some of the components are incorrect in the original article after deriving equation (2.8) and these mistakes are corrected in this study.

The gear mesh stiffness variations of a spur gear with damage are investigated by Chaari et al. (2008, 2009). Chaari et al. (2009) modelled the gear mesh stiffness fluctuations during meshing for an undamaged (healthy) gear and for damaged gears over time. Good agreement is found between the analytical and the finite element approaches to model tooth crack damage. The gear mesh stiffness has a similar shape in comparison to a square wave, with the reduction in stiffness due to the investigated cracks affecting more than one meshing period. The effect of the damage on the second meshing period seems relatively small compared to the effect on the meshing period where the broken tooth is primarily in contact. Chaari et al. (2009) found a linear relationship between the reduction in gear mesh stiffness and crack depth. Chaari et al. (2012) uses a square wave stiffness function to approximate the gear mesh stiffness with its properties presented in table 2.3. This is a good approximation to the results found in sophisticated analyses conducted by Chaari et al. (2009).

In this study the gear mesh stiffness of the healthy gearbox is approximated the same as done by Chaari et al. (2012) with the result in figure 2.2. The equation of motion

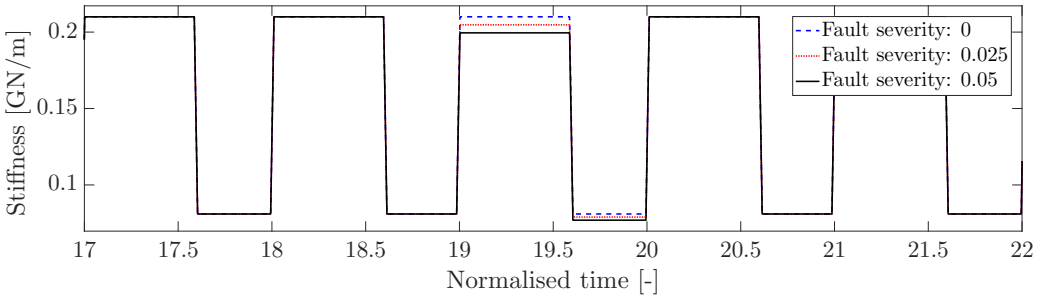


Figure 2.2: Gear mesh stiffness approximation in the presence of root crack faults with different fault severities. The time is normalised with the gear mesh period.

(equation (2.1)) is solved in the time domain and therefore the gear mesh stiffness period varies with a fluctuating rotational speed. The damage in figure 2.2 is induced by approximating the results obtained by Chaari et al. (2009). Damage is simulated by multiplying the gear mesh stiffness associated with the damaged tooth with a factor between 0.0 and 1.0. A fault severity of 0.2 indicates that the new gear mesh stiffness of the tooth is $0.8 \times k_{gm}(t)$, where $k_{gm}(t)$ is the original gear mesh stiffness.

The damping characteristics in the system are approximated by proportional damping and is calculated by,

$$\boldsymbol{C} = c_m \boldsymbol{M} + c_k \bar{\boldsymbol{K}} \tag{2.9}$$

where the proportionally constants that account for the mass and stiffness contributions are $c_m = 0.05$ and $c_k = 10^{-6}$ (Chaari et al., 2012). The mean stiffness matrix, denoted by $\bar{\boldsymbol{K}}$ in equation (2.9), is calculated by,

$$\bar{\boldsymbol{K}} = \boldsymbol{K}_s + \bar{k}_{gm} \boldsymbol{S} \tag{2.10}$$

where $\bar{k}_{gm}$ denotes the mean gear mesh stiffness given in table 2.3.

A four pole, three-phase, 50[Hz] ABB - MT 90 L electric motor is attached to the pinion through a shaft in the model with the properties given in table 2.5. It is the same electric motor used by Chaari et al. (2012). The relationship between the torque and the speed of the motor is (Chaari et al., 2012),

$$T_M(g_n) = \frac{T_b}{\left[1 + (g_b - g_n)^2 \left(\frac{c_{a1}}{g_n} - c_{a2} \times g_n^2\right)\right]} \tag{2.11}$$

where g_b and T_b are the slip in the motor and the torque at break down, respectively. The proportional drop in rotational speed by the motor, denoted by g_n, is calculated

Table 2.5: Characteristics of the electrical motor (Chaari et al., 2012)

Characteristics of motor	Symbol	Value	Units
Rated power	P	1.5	kW
Synchronous speed	ω_s	1500	rpm
Full load	T_f	10	N.m
Starting torque	T_s	27	N.m
Torque at break down	T_b	32	N.m
Slip	g_b	0.315	–
Motor constant	c_{a1}	1.711	–
Motor constant	c_{a2}	1.316	–

from,

$$g_n = 1 - \frac{\omega_r}{\omega_s} \tag{2.12}$$

where the rotational and synchronous speed of the motor is ω_r and ω_s, respectively. The relationship between the torque and the rotational speed in figure 2.3 is calculated from equation (2.11) and the properties in table 2.5.

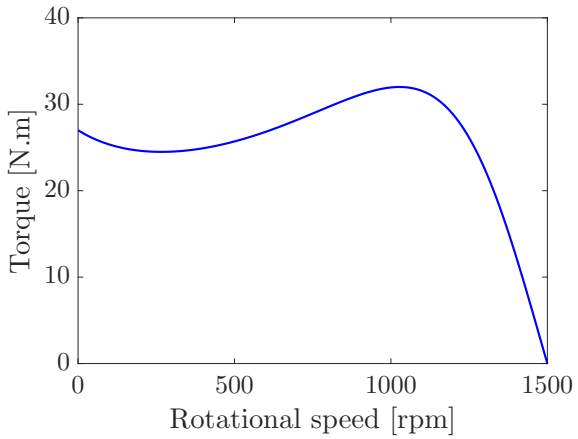


Figure 2.3: The relationship between the torque and the speed of the electrical motor.

The external load vector,

$$\boldsymbol{F} = \begin{bmatrix} 0 & 0 & T_M & 0 & 0 & 0 & T_L & 0 \end{bmatrix} \tag{2.13}$$

is used in equation (2.1). The load that is applied by the motor T_M is calculated from the rotational speed in the system and the motor model in equation (2.11). The external load applied by the machine T_L, in figure 2.1, is obtained from the relationship corrected from the article by Chaari et al. (2012),

$$T_M = -\frac{N_p}{N_g} T_L \tag{2.14}$$

where N_p and N_g are the number of teeth on the pinion and the gear, respectively.

All of the components that are required to solve the equation of motion (2.1) are available. The natural frequencies of the system is used to evaluate the plausibility of the model. The undamped natural frequencies are obtained by solving an eigenvalue problem,

$$\left(\bar{\boldsymbol{K}} - \xi \boldsymbol{M}\right) \boldsymbol{v} = \boldsymbol{0} \tag{2.15}$$

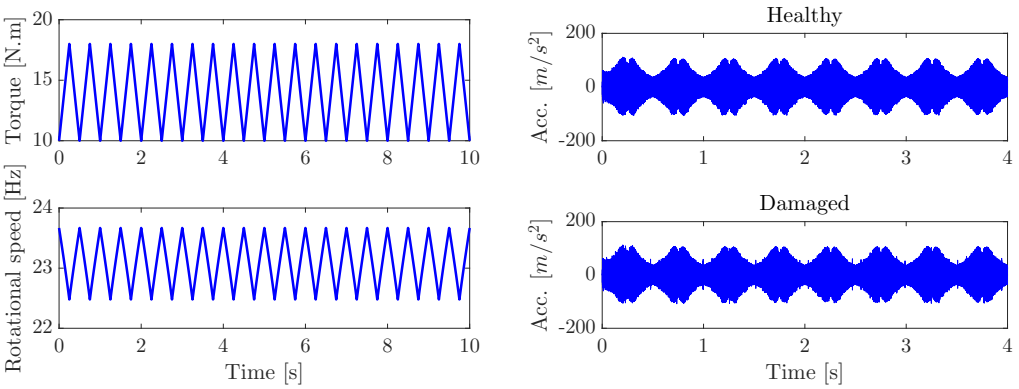
where $\sqrt{\xi}$ is an undamped natural frequency of the system with associated eigenvector $\boldsymbol{v}$. The mean stiffness matrix calculated from equation (2.10) is used in equation (2.15) as opposed to the original article. The solution of the eigenvalue problem in table 2.6 indicates that only one rigid body mode is present, which is expected. This means that the only free vibration mode in the system is associated with the shafts turning proportionally to one another, which is the expected result.

Table 2.6: Undamped natural frequency number and the associated value in Hertz of the gearbox model.

1	2	3	4	5	6	7	8
0.000	532.309	1006.584	1057.102	1186.271	1203.706	2067.666	5289.872

Now that all of the characteristics associated with the gearbox model are determined, the equation of motion can be solved. Equation (2.1) is solved with an implicit Newmark integration scheme due to the highly non-linear behaviour of the gear mesh stiffness. More information on this integration procedure is available in the textbook by Zienkiewicz and Taylor (2005). The simulated operating conditions in figure 2.4a has abrupt (non-smooth) changes during each duty cycle and are used in this study. The presence of the abrupt operating condition changes are not physical, but it tests the ability of the diagnostic methodology to detect faults in complicated operating conditions. The resulting vibration signal is compared in figure 2.4b for the gearbox in a good condition and the gearbox with local damage on the gear with a fault severity of 0.1. The acceleration of the fifth degree of freedom, denoted by $\ddot{x}_2$ in figure 2.1, is used in this study.

Severe amplitude modulation is present in the vibration signals presented in figure 2.4b due to the applied operating conditions. This severe amplitude modulation emphasises that the diagnostic technique should be robust to operating condition changes. There are slight differences between the healthy (undamaged gear) and the damaged gearbox vibration signals in figure 2.4b.



(a) Operating conditions (b) Vibration signal

Figure 2.4: The operating conditions at the motor-pinion shaft and the corresponding vibration signal for the numerical gearbox model.

2.2 Experimental setup

An experimental setup was constructed based on the design of Dr. CJ Stander (Stander and Heyns, 2005) to test the ability of the fault diagnostic methodology on real data. The experimental setup is shown in figure 2.5 with the corresponding legend in table 2.7. The objective of the experiments was to evaluate whether it is possible to detect,

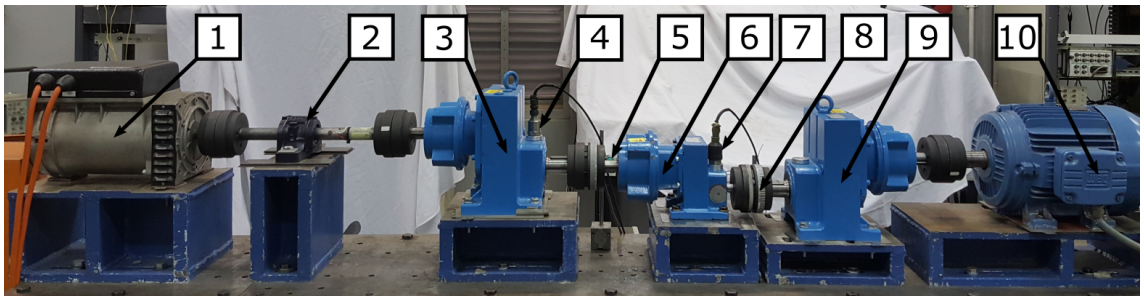


Figure 2.5: The experimental setup with its legend included in table 2.7.

localise and trend fault growth. This was performed in the form of an accelerated life test by seeding small defects in the teeth of the gears. The defects progressed and resulted in complete failure of the tooth as the gearbox was overloaded.

The electric motor supplies rotational energy to the system and is connected to a speed-reduction gearbox which is connected to the monitored gearbox (indicated by (6)). The monitored gearbox and the gearbox closest to the alternator increases the rotational speed of the input shaft. The output shaft of the gearbox closest the alternator is supported by a bearing and it provides the ability to attach a telemetry system for strain measurements etc. The alternator is connected to a resistor bank and it removes energy from the system by providing torsional resistance. The highest torque in the

Table 2.7: The legend used for the experimental setup in figure 2.5.

Item number	Description
1	5.5 kVA Mecc alte spa three-phase alternator
2	Support bearing
3	Siemens E68-A-100 helical gearbox
4	Accelerometer (500 [mV/g] sensitivity)
5	Proximity probe applied to key of shaft (1 pulse per revolution)
6	Siemens E38-A-100 helical gearbox (Monitored gearbox)
7	Accelerometer (500 [mV/g] sensitivity)
7*	Tri-axial accelerometer (100[mV/g])
8	Zebra tape shaft encoder and optical probe (88 pulses per revolution)
9	Siemens E68-A-100 helical gearbox
10	Weg 5,5 kW three-phase four-pole squirrel cage electrical motor

7* is mounted behind the accelerometer indicted by 7.

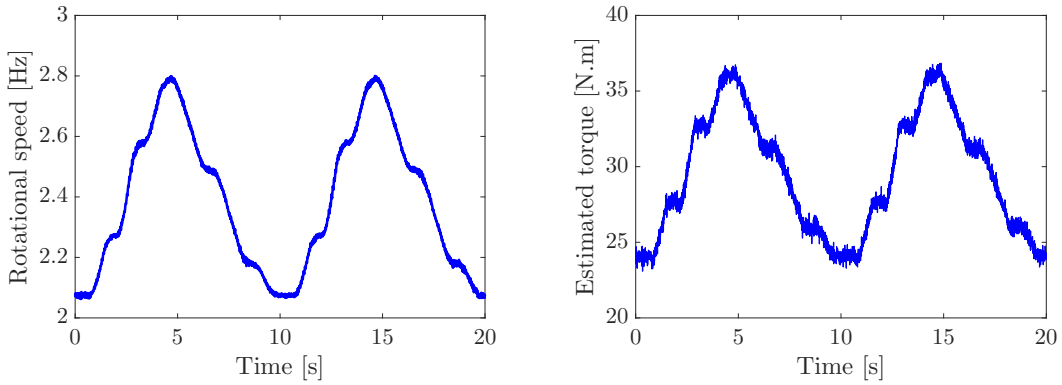
system is located at the input shaft of the monitored gearbox, which is indicated by 8 in figure 2.5. The rotational speed of the motor and the load applied by the alternator are controlled through a personal computer with the i3 package, developed by Dr. JJA Eksteen. The alternating voltage and current generated by the alternator is converted to direct voltage and current before being measured at a rate of 1kHz with a National instruments data acquisition card in the personal computer.

Two uni-axial accelerometers, a tri-axial accelerometer and two tachometers were used during the experiments which resulted in seven data streams being measured simultaneously by an Oros OR35 data acquisition system. A lot of data were measured to ensure that sufficient data are available for potential future work. The helical gearboxes have strong axial excitations which contain a lot of diagnostic information, hence the axial component of the tri-axial accelerometer is used in subsequent chapters of this study.

The accelerometer signals were sampled at 25.6kHz and the tachometers (indicated by 5 and 8 in figure 2.5) were sampled at 51.2kHz. The tachometer sampling frequencies affect the zero crossing time errors, which need to be minimal for performing order tracking successfully (Fyfe and Munck, 1997). Hence, the sampling frequencies of the tachometers were the highest sampling frequencies that were possible without any measurement problems. The zebra tape shaft encoder requires geometrical compensation (see (Resor et al., 2005)) was performed with the technique developed by Diamond et al. (2016) in this study. The instantaneous phase from the zebra tape shaft

encoder is used to evaluate the performance of the tacholess order tracking methods in chapter 3. A high sampling frequency was used for accelerometer signals to ensure that all the relevant information can be extracted to perform fault diagnosis, tacholess order tracking and future work on the data. The nature of tests only provide a single opportunity for measurement and therefore high sampling frequencies were motivated as well. Appendix A contains additional information pertaining to the experimental setup and the data.

The operating conditions (i.e. load or torque and speed) in figure 2.6 are investigated in this study and is present during data acquisition. The load and the rotational speed of the input shaft of the monitored gearbox are presented, but it can easily be converted to the operating conditions on other shafts by using the gear ratios included in Appendix A. The operating conditions that are investigated in this study differ from



(a) (b)

Figure 2.6: The rotational speed in figure 2.6a and the estimated torque in figure 2.6b at the input shaft of the monitored gearbox.

those found in wind turbines (He et al., 2016) and draglines (Eggers et al., 2007), but it provides sufficient complexity to critically evaluate the performance of the tacholess order tracking method as well as the diagnostic methodology. The duty cycle of the operating conditions was 10 seconds with a period of 10 - 20 minutes between consecutive measurements.

Experiments were firstly performed with a healthy gear and after sufficient data were measured, the experiments were stopped. The monitored gearbox was removed from the setup and the monitored gearbox was disassembled so that its gear could be damaged. After the gear was damaged, the monitored gearbox was reassembled and connected to the system. In this study, it is assumed that the diagnostic methodology is applied to a gearbox where the damaged gear resulted due to normal operation. However, the disassembling and reassembling process alters the system and changes the integrity of the data. Therefore, the following provisions were made to reduce the

effects of the disassembling and reassembling processes:

- The healthy data are generated from a gearbox with a gear in a healthy condition. This gear is removed, damaged and then inserted back into the system. The damaged dataset is measured from the gearbox with the aforementioned damaged gear.
- The system is aligned similarly in both cases with the same measurement equipment being used to ensure that there are minimal changes between the healthy and damaged experiments.

After sufficient data was obtained from a healthy gearbox during the first experiment, the damage was seeded into one of the teeth of the gear with the result presented in figure 2.7. The slot was 0.3mm high, 50% of the tooth width deep and is through the entire breadth of the tooth (i.e. in the axial direction) for the first experiment.

Figure 2.7: The damaged gear of the first experiment.

The measured vibration signal from the axial component of the tri-axial accelerometer is presented in figure 2.8 for the gearbox with a healthy and a damaged gear. The operating conditions in figure 2.6 were present during data acquisition. The damaged gearbox vibration signal was acquired during the early stages of the experiment. The vibration signals for both conditions contain clear impulses in the time domain signal which leads to broadband excitation. These type of impulses are normally indicative of damage. The source of the impulses was isolated to the monitored gearbox by mounting an accelerometer on various components in the setup during constant torque and speed conditions. The impulses were non-present in the vibration signals measured from equipment located with long transmission paths from the monitored gearbox. The period between successive impulses in figure 2.8 varies over time and it does not clearly

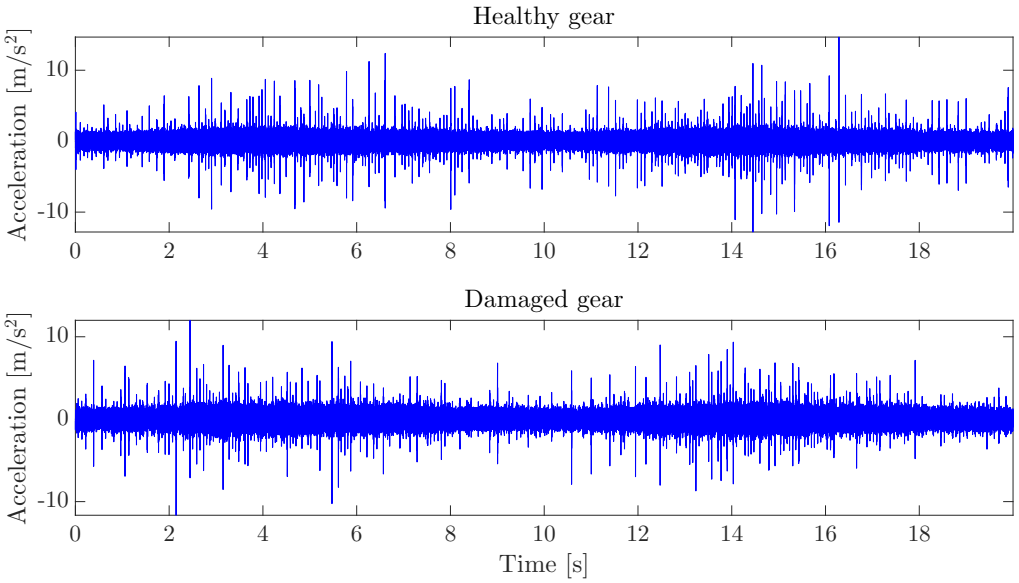


Figure 2.8: Vibration signal measured on the gearbox with a healthy (undamaged) a damaged gear.

correspond to a characteristic fault frequency, such as a ball-pass frequency etc., of the system. The impulses disappear in some cases for short time frames.

It was noticed that one of the supporting bearings of the monitored gearbox's input shaft is able to slide freely within its housing. Circlips prevent large axial movements, but it seems that some axial and radial movement with respect to the housing occur during experiments. It is suspected that the movement of the bearing results in the observed impulses in the vibration signal. The investigation that was conducted on this subject is included in appendix A.2.

The first experiment continued for more than 45 days with the damaged gear, since there was not a definitive indication that the damaged tooth and other adjacent teeth had failed. After the experiment was stopped, it was observed that the damaged tooth had failed during the experiment, as shown in figure 2.9. The time that the tooth failed is unknown, since the tooth failure was not clearly seen in the data and the gearbox cannot be disassembled during testing without affecting the data of subsequent experiments with the same gear. Usually, a few teeth fail, or the shaft fails etc. during the tooth failure which is easily seen within the data. The time of failure is investigated in chapter 5.

Another accelerated life test was completed to ensure that the exact time of failure is known when evaluating the fault diagnostic methodology. The height of the seeded damage was set to 0.2mm with a depth of 60% of the tooth width, which resulted in a significantly faster testing time of approximately three days. The second gear failed

Figure 2.9: The damaged gear in figure 2.7 after more than 45 days of experiments.

similarly to the first experiment, where only the damaged tooth failed.

2.3 Conclusion

Two gearbox datasets were presented in this chapter. The numerical gearbox model is corrected from the model obtained from the original articles by Chaari et al. (2008, 2012). The purpose of the numerical gearbox model is to test the diagnostic methodology in a completely controlled environment. An experimental setup was constructed so that the methodology can be tested on data from an actual gearbox. Two experiments were conducted in the laboratory. The first experiment was conducted to investigate the ability of the gearbox model to detect the presence of faults, locate them on the gear and to trend it over time. It was expected that more than one tooth will break after the failure of the damaged tooth. A gear with a few broken teeth are easily seen within the vibration signal and therefore that was some ambiguity whether the gearbox should be opened or not. It was decided to continue with the experiments instead of compromising the integrity of the experiment. A second test was conducted and stopped after it was suspected that the tooth of the gear had failed. These characteristics are thoroughly investigated in chapter 5. In the next chapter, the proposed tacholess order tracking is evaluated on the data described in this chapter with additional data as well.

Chapter 3 Tacholess order tracking methodology

The purpose of this dissertation is to develop a robust and cost-effective fault diagnostic methodology. This necessitates using tacholess order tracking methods to circumvent the practical and financial implications of installing and using tachometers. In this chapter, a maxima tracking algorithm, developed by Urbanek et al. (2013), is modified, improved and investigated on seven datasets.

3.1 Proposed tacholess order tracking method

The proposed tacholess order tracking process diagram in figure 3.1 is similar to the methods used by Urbanek et al. (2013), Zhao et al. (2013) and Vinson (2014). A vibration signal, measured from a rotary machine, is decomposed into a time-frequency distribution so that an instantaneous frequency (IF), which varies with the rotational speed of the reference shaft, can be tracked. A Vold-Kalman filter (VKF) is centred on the estimated IF so that a mono-component signal can be extracted from the multicomponent vibration signal. The first (angular-velocity) and the second generation (angular-displacement) VKFs are investigated in this study and it is found that the angular-displacement VKF performs slightly better than the angular-velocity VKF. The instantaneous phase is calculated from the mono-component vibration signal by using the Hilbert transform and is scaled to reflect the phase of the shaft of interest. The instantaneous phase is subsequently used to resample the vibration signal at a constant angular frequency to order track the signal.

3.1.1 Time-frequency distribution

The spectrum of a signal measured from a rotary machine contains rotational speed dependent and -independent components. The time and frequency domain information

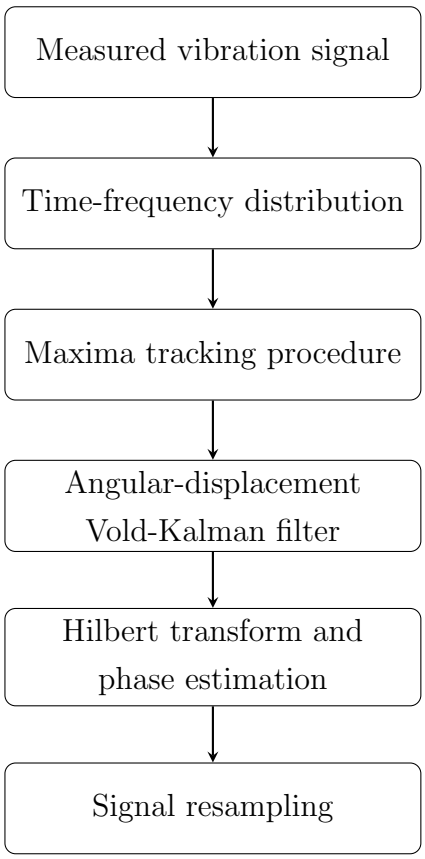


Figure 3.1: Overview of the proposed tacholess order tracking method.

in the vibration signal is investigated simultaneously in a time-frequency distribution so that the IF can be estimated at each time increment. Urbanek et al. (2013) used the spectrogram for maxima tracking and Zhao et al. (2013) used the spectrogram to estimate the chirp rate so that the Chirplet transform could be used to estimate the component of interest. In this study the spectrogram is used for maxima tracking and motivated by:

- It does not contain any interference terms such as the Bilinear transforms (Wigner-Ville distribution etc.) (Peng and Chu, 2004).
- It does not require any prior knowledge of the rotating frequency to estimate a characteristic such as a Chirp rate for the Chirplet transform or which scales to evaluate for the CWT.
- It is fairly computationally efficient compared to similar methods such as the Chirplet transform.

The short-time Fourier transform (STFT), performed on a continuous signal $x(t)$,

$$\Gamma(t, f) = \int_{-\infty}^{\infty} x(\tau)\upsilon(\tau - t)\exp(-2\pi f j\tau)d\tau \tag{3.1}$$

is used to represent the time and frequency information of the signal simultaneously.

The time-limited window, such as a Hamming window, is denoted by $\upsilon(\tau - t)$ and $j = \sqrt{-1}$. Equation (3.1) is easily implemented for discrete signals, since the fast Fourier transform is recursively applied on each windowed vibration signal segment to construct the time-frequency distribution. The spectrogram is related to the STFT in equation (3.1) with $|\Gamma[t, f]|^2$.

3.1.2 Proposed maxima tracking procedure

Maxima tracking is performed to find, isolate and track an IF, denoted by f_{IF}, in the form of,

$$f_{\text{IF}}(t) = k \times c_s \times f_r(t) \tag{3.2}$$

in a time-frequency distribution. The IF is related to the rotating frequency or speed of the shaft, denoted by $f_r(t)$, with the time-invariant factor $k \times c_s$. This indicates that it is the kth harmonic of a component with a frequency of $c_s \times f_r(t)$, such as a gear mesh frequency etc. This IF presents itself as a ridge or a local maximum in the discrete time-frequency distribution, where the IF is estimated by solving a constrained optimisation problem in the form of,

$$\begin{aligned} &\underset{f}{\text{minimise}} && -|\Gamma[t, f]|^2 \\ &\text{subject to} && (f - f_c[n])^2 \leq \Delta f_c^2 \end{aligned} \tag{3.3}$$

where Δf_c denotes the feasible bandwidth and f_c denotes the centre of the constraint, which is the predicted value of the IF at time increment n. A discrete signal evaluated at time increment n is denoted by $x[n]$, where a continuous signal or function is denoted by $x(t)$. The relationship between the time increment n and the continuous time t is,

$$t = n\Delta t = \frac{n}{f_s} \tag{3.4}$$

where f_s and Δt are the sampling frequency and sampling period, respectively. The solution of equation (3.3) at time increment n, denoted by $f_{\max}[n]$, is assumed to be related to the actual IF by,

$$f_{\text{IF}}(n\Delta t) \approx f_{\max}[n] \tag{3.5}$$

where Δt is the time difference between adjacent window centres in the spectrogram. If maxima tracking is performed as an unconstrained optimisation problem, then c_s will vary with time which will result in erroneous phase estimation and order tracking results. Urbanek et al. (2013) centred the constraint in equation (3.3) about the maxima tracked frequency estimated at the previous time increment,

$$f_c[n] = f_{\max}[n - 1] \tag{3.6}$$

where the implicit assumption is made that there is no acceleration between time increments. Note that if $f_c[n]$ is incorrect, a larger Δf_c has to be used to accommodate the errors in the maxima tracking process. However, if a large Δf_c is used in the maxima tracking process, it is more susceptible to tracking incorrect components which are present in the large feasible region.

The assumption that is made to obtain equation (3.6) is problematic when performing maxima tracking in the presence of large angular accelerations, especially in the presence of strong background noise. Higher order information needs to be incorporated into the maxima tracking process to make it more robust for tracking the IF in the presence of large accelerations. The Taylor series expansion of the IF in the signal about the previous time increment,

$$f_{\text{IF}}(t) = f_{\text{IF}}(t - \Delta t) + \Delta t \frac{d}{dt} f_{\text{IF}}(t - \Delta t) + \frac{1}{2}\Delta t^2 \frac{d^2}{dt^2} f_{\text{IF}}(t - \Delta t) + \dots \tag{3.7}$$

can be used to predict the value of the IF at the next time increment $t = n\Delta t$. This prediction is used to centre the constraint,

$$f_c[n] = f_{\text{IF}}(n\Delta t) \tag{3.8}$$

used in equation (3.3). Equation (3.6) is obtained from equation (3.7) by assuming that all the gradients are zero (i.e. no acceleration is present). Incorporating higher order information into the maxima tracking process can increase its robustness, however the higher order information is unknown, since the function $f_{\text{IF}}(t)$ is unknown. The spectrogram contains discrete time steps and therefore the gradients in equation (3.7) can only be estimated by using finite difference schemes on the previous IF estimates (i.e. $f_{\max}[n-1]$, $f_{\max}[n-2]$ etc.), which is not desired. The finite difference estimates of the higher order information result in erroneous maxima tracking estimates, since there are ridge smearing and background noise present in the spectrogram.

A probabilistic approach is used to infer the actual IF so that higher order information can be incorporated into the maxima tracking process. It is assumed that the true IF, f_{IF}, is related to its representation in the spectrogram or its estimate from the spectrogram, $f_{\max}$, with a Gaussian distribution,

$$p\left(f_{\max}[n] \,|f_{\text{IF}}[n\Delta t]\,, \sigma^2\right) \propto \exp\left(-\frac{\left(f_{\max}[n] - f_{\text{IF}}(n\Delta t)\right)^2}{2\sigma^2}\right) \tag{3.9}$$

where σ represents the standard deviation of the noise and is attributed to ridge smearing and other noise components in the spectrogram. The estimate of the true IF, f_{F}, is represented by a N_p order polynomial in the form of,

$$f_F(t) = w_0 + w_1 t^1 \dots w_{N_p} t^{N_p} \tag{3.10}$$

where $f_F(t)$ in equation (3.10) is used instead of $f_{\text{IF}}(t)$ in equation (3.9) and then the weights $\boldsymbol{w} = \{w_0, w_1, \dots w_{N_p}\}$ are obtained using maximum likelihood. The true IF, f_{IF}, is unknown and therefore an estimate, f_F, needs to be used at each time step. The weights (Bishop, 2006),

$$\boldsymbol{w} = (\boldsymbol{Q}^T\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\boldsymbol{f}_{max} \tag{3.11}$$

are obtained from the previous N_m estimates of the IF, denoted by $\boldsymbol{f}_{max}$, using maximum likelihood. The design matrix of the N_p order polynomial, in the form of,

$$\boldsymbol{Q} = \begin{bmatrix} 1 & t[n-1] & \dots & t[n-1]^{N_p} \\ 1 & t[n-2] & \dots & t[n-2]^{N_p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t[n-N_m] & \dots & t[n-N_m]^{N_p} \end{bmatrix} \tag{3.12}$$

and the vector of the previous N_m estimates of the IF,

$$\boldsymbol{f}_{max} = \begin{bmatrix} f_{\max}[n-1] \\ f_{\max}[n-2] \\ \vdots \\ f_{\max}[n-N_m] \end{bmatrix} \tag{3.13}$$

are used in equation (3.11) to obtain the weights in equation (3.10). The polynomial function in equation (3.10) satisfies the Taylor series expansion of the IF in equation (3.7), which means that if a first order polynomial function is used, then the assumption is made that the angular acceleration is constant etc. Note that even though frequentist statistics are used to infer the weights in this study, Bayesian linear regression can be investigated with a conjugate prior by slightly modifying equation (3.11). However, Bayesian linear regression requires additional hyperparameter optimisation, which is not investigated in this study.

It is strongly suggested to use a first order polynomial to ensure that only a few number of training points (N_m) are required, solving equation (3.11) remains computationally efficient and that errors in the extrapolation process are minimised. By using equation (3.6), the IF estimate at the first time increment in the maxima tracking process needs to be provided. If the constraint in equation (3.3) is centred by the prediction made by equation (3.10) from a first order polynomial, then the initial gradient needs to be provided as well. To make this process easier, it is assumed that the initial gradient is zero, which results in equation (3.6) to be used for the initial steps. In this study the first three estimates of the IF are obtained by using equation (3.3) with equation (3.6) and thereafter equation (3.10) can be used. The initial estimate of the IF is obtained by inspecting the spectrogram and $c_s \times k$, in equation (3.2), is estimated by

roughly knowing the operating range of the machine or using a simple device such as a hand-held tachometer.

Since it is a constrained minimisation problem, the Lagrangian in the form of,

$$\begin{aligned}\mathcal{L}_n(f, \boldsymbol{\lambda}, \boldsymbol{s}) = -|\Gamma(t[n], f)|^2 + \lambda_1 \left((f - f_{\max}[n-1])^2 - \Delta f_{c1}^2 + s_1^2\right) \\ + \lambda_2 \left((f - f_F(t[n]))^2 - \Delta f_{c2}^2 + s_2^2\right)\end{aligned} \tag{3.14}$$

is minimised to find the frequency f that minimises the cost function $-|\Gamma[t, f]|^2$ and adheres to the constraints. The Lagrangian multiplier and slack variable associated with the ith constraint are denoted by λ_i and s_i, respectively. Equation (3.14) can be solved by considering the Karush-Kuhn-Tucker conditions, but the penalised unconstrained cost function in the form of (Arora, 2004),

$$\begin{aligned}\kappa(f, \boldsymbol{\rho}, n) = -|\Gamma(t, f)|^2 + \rho_1 \max\left[0, (f - f_{\max}[n-1])^2 - \Delta f_{c1}^2\right] \\ \rho_2 \max\left[0,\ (f - f_F(t[n]))^2 - \Delta f_{c2}^2\right]\end{aligned} \tag{3.15}$$

is minimised instead. The bandwidth describing the feasible region enforced by the two constraints are denoted by Δf_{c1} and Δf_{c2}, respectively. In this study the feasibility bandwidths (Δf_{c1} and Δf_{c2}) have the same value. The components of the vector $\boldsymbol{\rho} = [\rho_1, \rho_2]$, in equation (3.15), have the following values,

$$\rho_1 = \begin{cases} 10^9 & \text{if } n \leq 3 \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

and,

$$\rho_2 = \begin{cases} 10^9 & \text{if } n > 3 \\ 0 & \text{otherwise} \end{cases} \tag{3.17}$$

where only one of the constraints are active at a time due to the choice of $\boldsymbol{\rho}$. The estimate of the instantaneous frequency is obtained from,

$$f_{\max}[n] = \underset{f}{\operatorname{argmin}}\ \kappa(f, \boldsymbol{\rho}, n) \tag{3.18}$$

where $\kappa(f, \boldsymbol{\rho}, n)$ is given by equation (3.15). Note that the solution of equation (3.15) is also the solution of equation (3.3) if both constraints are incorporated into equation (3.3) similarly to equation (3.15). The method implemented by Urbanek et al. (2013) essentially uses $\rho_1 \neq 0$, while ρ_2 is always zero in equation (3.15).

3.1.3 The Vold-Kalman filter

The Vold-Kalman filter (VKF) is a bandpass filter with a varying centre frequency and bandwidth. If the centre frequency of the constraint in equation (3.3) is set equal to

the estimated IF (i.e. $\boldsymbol{f}_{max}$) and the bandwidth is set sufficiently fine, then a mono-component signal, denoted by $\mathbf{x}_{IF}$ can be extracted from a multicomponent vibration signal $\boldsymbol{x}$. Note that a discrete signal is investigated in this section and therefore the whole signal is denoted by $\boldsymbol{x}$ (i.e. a one-dimensional vector over all time steps) and the signal at increment n is denoted by $x[n]$.

Two generations of the VKF exist namely the angular-velocity (or first generation) and the angular-displacement (or second generation) VKF. The angular-velocity VKF aims to estimate the mono-component signal, while the angular-displacement VKF aims to estimate the envelope of the mono-component signal (Tuma, 2005). The signal containing the IF that needs to be extracted, $x_{\text{IF}}[n]$ is related to the vibration signal $x[n]$ in the data equation,

$$x[n] = x_{\text{IF}}[n] + \zeta[n], \tag{3.19}$$

where $\zeta[n]$ represents the other sinusoidal components and the background noise at time $t = n/f_s$. The structural equation for the angular-velocity VKF,

$$x_{\text{IF}}[n] - 2\cos(2\pi f_{\text{IF}}[n] f_s^{-1}) x_{\text{IF}}[n+1] + x_{\text{IF}}[n+2] = \varepsilon[n] \tag{3.20}$$

and of the angular-displacement VKF,

$$\beta[n] - 2\beta[n+1] + \beta[n+2] = \varepsilon[n] \tag{3.21}$$

describes the characteristic of interest and $\varepsilon[n]$ is known as the non-homogeneity term at time increment n which represents the error in the left-side of the structural equation. Equation (3.21) is for a two-pole angular-displacement filter, where the structural equation for other poles are found in the paper by Tuma (2005). The envelope $\beta[n]$ is related to the signal of interest by,

$$x_{\text{IF}}[n] = \beta[n] \exp\left(\frac{2\pi j}{f_s} \sum_{i=0}^{n} f_{\text{IF}}[i]\right) \tag{3.22}$$

where f_s is the sampling frequency of the signal x_{IF} and $j = \sqrt{-1}$. There are three sets of unknowns in equation (3.19), (3.20) and (3.21) namely $\boldsymbol{\varepsilon}$, $\boldsymbol{\zeta}$ and either $\boldsymbol{\beta}$ or $\boldsymbol{x}_{\text{IF}}$. This result in an underdetermined system of equations, which are solved by finding the characteristic of interest ($\boldsymbol{x}_{\text{IF}}$ or $\boldsymbol{\beta}$) that simultaneously minimises $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}$. This is achieved by solving the multi-objective optimisation problem with the weighted sum method in the form of (Tuma, 2005),

$$\kappa = \boldsymbol{\varepsilon}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{\varepsilon} + \boldsymbol{\zeta}^T \boldsymbol{\zeta} \tag{3.23}$$

where the weight matrix $\boldsymbol{A}$ is a diagonal matrix. The component at time increment n, denoted by $A[n, n]$, describes the importance of minimising $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ instead of $\boldsymbol{\zeta}^T \boldsymbol{\zeta}$ at time

increment. The choice of weighting factor at time increment n (i.e. $A[n, n]$) determines the rise time as well as the bandwidth of the filter and is extremely important to the success of the VKF (Pan and Lin, 2006). The weight factors proposed by Tuma (2005) are used in this study. Due to the nature of the data and structural equations, the optimal solution of the cost function, given by equation (3.23), can be obtained in closed form if only a single component is tracked (Tuma, 2005).

The angular-displacement VKF allows multiple components to be tracked, while the angular-velocity VKF is constrained to a single component (Tuma, 2005). Pan and Lin (2006) found that the angular-displacement is able to estimate the component of interest accurately in the presence of crossing orders. It obtains a better rise time and bandwidth combination than the angular-velocity VKF. The angular-displacement VKF does not have any frequency nor slew rate limitations as well (Blough, 2003). A brief investigation was conducted on the angular-velocity or angular-displacement VKF with the experimental signals introduced later in this section. The implementations are based on the paper by Tuma (2005). The results, for the investigated signals, indicate that the one-pole angular-displacement VKF obtains relatively low errors and it is very robust, since it is possible to use very small bandwidth factors. The bandwidth factor is defined by,

$$B_w = \frac{f_b[n]}{f_{\text{IF}}[n]} \tag{3.24}$$

where f_b denotes the bandwidth of the passband of the VKF at increment n. The angular-velocity VKF obtains relatively low errors as well, but larger bandwidth factors are required than the one-pole angular-displacement VKF to avoid numerical problems in the optimisation process. A two-pole angular-displacement VKF results in numerical problems when small bandwidth factors are used and is ill-suited for the investigated signals. The results of this investigation are included in section B.1 in Appendix B.

Even though a specific harmonic is tracked with the maxima tracking process, any harmonic can be filtered with the VKF. An investigation was performed on the aforementioned experimental signals to determine which harmonic to filter. If a small bandwidth factor is used, an arbitrary harmonic of the IF can be filtered without any change in the error occurring for the investigated signals. However, the harmonic that needs to be filtered must be chosen carefully if a large bandwidth factor is used and the results are dependent on the signal that is used as well. The results of this investigation are included in detail in section B.1. The angular-displacement VKF is used in this study from the results of the investigations and from the papers by Blough (2003) and Pan and Lin (2006). A bandwidth factor of 10^{-4} is used in this study since it is robust and the error does not depend on the harmonic that is filtered.

3.1.4 Instantaneous phase estimation and resampling

The phase estimation process is presented in this section for continuous signals, however it is easily extended to discrete signals obtained from the maxima tracking and the Vold-Kalman filtering processes. The mono-component signal obtained from the VKF in the previous section, denoted by x_{IF}, is transformed with the Hilbert transform,

$$H\left(x_{\mathrm{IF}}(t)\right) = \frac{1}{\pi}\int_{-\infty}^{\infty} x_{\mathrm{IF}}(\tau)\frac{1}{t-\tau}d\tau \tag{3.25}$$

and used to obtain the instantaneous phase of this signal,

$$\phi_{\mathrm{IF}}(t) = \arctan\left(\frac{H(x_{\mathrm{IF}}(t))}{x_{\mathrm{IF}}(t)}\right) \tag{3.26}$$

which is constrained to $[-\pi, \pi]$ due to the arctangent function. This phase, obtained from equation (3.26), is unwrapped to obtain the cumulative phase of the estimated IF component over time. The tracked IF component is related to the rotating shaft of interest with equation (3.2) if equation (3.5) is valid. The aforementioned relationship is used to obtain the instantaneous phase of the shaft,

$$\phi_r(t) = \frac{1}{kc_s}\mathrm{unwrap}\left(\phi_{\mathrm{IF}}(t)\right) \tag{3.27}$$

at time step t. The instantaneous phase of the shaft over time, obtained from equation (3.27), is used to resample the signal into equal angular increments.

3.2 Validation of proposed method

Two maxima tracking methods are evaluated and compared in this chapter on numerical data, numerical gearbox model data and experimental data. The maxima tracking method proposed by Urbanek et al. (2013) does not incorporate acceleration information into the maxima tracking process and is used as the benchmark maxima tracking algorithm. The method is indicated as (a) in all of the figures. The proposed method incorporates higher order (i.e. acceleration) information into the maxima tracking process and is indicated by (b) in all of the figures.

3.2.1 Numerical validation

Three numerical signals are investigated in this section. The signals are chosen specifically to highlight some of the deficiencies when performing maxima tracking without

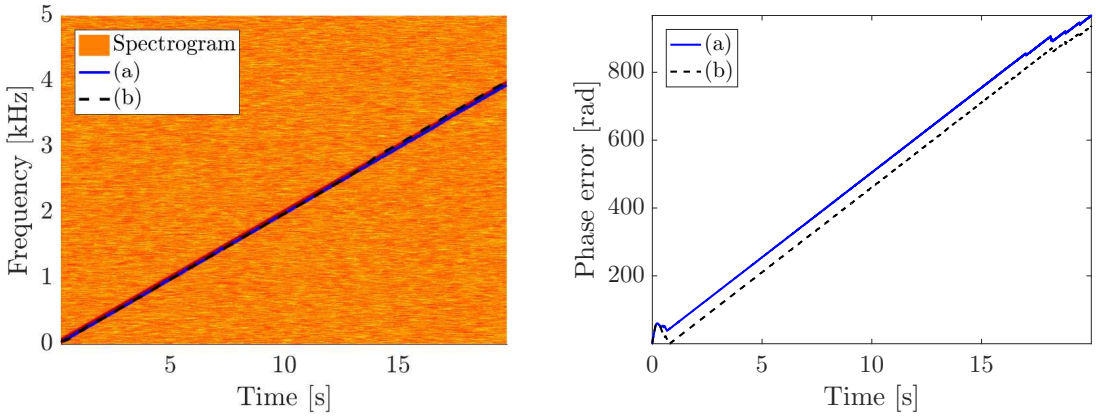
incorporating higher order information into the process. For the first signal, the shaft under consideration has the following rotational speed in Hertz,

$$f_r(t) = 200t + 50 \tag{3.28}$$

where f_r denotes the shaft speed that needs to be tracked in Hertz and t denotes time in seconds. The simulated shaft accelerates at 400π rad/s^2, which exceeds the standard operating range of common rotary machines. This is merely used to compare the two maxima tracking algorithms in the presence of large frequency gradients. The resulting signal,

$$x(t) = \sin\left(2\pi \int_0^t f_r(\tau)d\tau\right) \tag{3.29}$$

has a single IF, calculated from equation (3.28). Zero mean Gaussian noise with a standard deviation of 1.0 is added to the signal in equation (3.29). In figure 3.2a the spectrogram of the signal is superimposed with the two maxima tracking approaches considered in this study.



(a) (b)

Figure 3.2: The spectrogram and the maxima tracking results from the first numerical signal in figure 3.2a resulted in the phase error in 3.2b. The benchmark maxima tracking algorithm, proposed by Urbanek et al. (2013), is indicated by (a) and the proposed method is indicated by (b). Note that the final phase of the shaft component is 251324.899rad, which indicates that the relative phase error made by the maxima tracking process is small.

The phase error in figure 3.2b is computed by,

$$e(t) = |\phi_r(t) - \phi_a(t)| \tag{3.30}$$

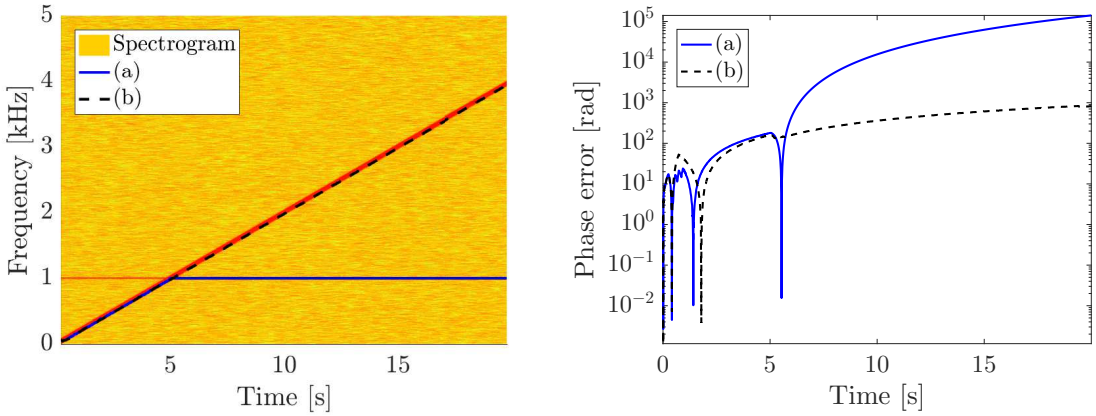
where ϕ_r, is the estimated phase in radians obtained from the process described in section (3.1.4). The analytical phase, denoted by $\phi_a(t)$ and used in equation (3.30), is

computed from $2\pi \int_0^t f_r(\tau)d\tau$. The exact same spectrogram, noise and Δf_c are used in the two maxima tracking algorithms with the same subsequent processing to obtain the instantaneous phase of the shaft. The phase error in figure 3.2b indicates that both maxima tracking algorithms made a constant frequency error in figure 3.2a. This is attributed to the fixed time-frequency resolution of the spectrogram which is unable to represent the IF accurately. Even though a linear phase error is present in figure 3.2b, the error is small compared to the total phase of 251324.899rad completed by the shaft over a 20 second period. The proposed maxima tracking method, indicated by (b), performs slightly better than the benchmark maxima tracking algorithm.

The second investigated signal contains the exact same rotational speed function (see equation (3.28)) as the first signal, but with an additional time invariant narrowband component,

$$x(t) = \sin\left(2\pi \int_0^t f_r(\tau)d\tau\right) + \sin\left(2\pi 10^3 t\right) \tag{3.31}$$

located at 1kHz. The purpose of this is to evaluate whether it is possible to track a frequency component with an intersecting component not related to the component of interest. The spectrogram of the signal in equation (3.31) is superimposed with the results from the two maxima tracking algorithms in figure 3.3a. The two components in



(a) (b)

Figure 3.3: The spectrogram and the maxima tracking results from the second numerical signal in figure 3.3a resulted in the phase error in 3.3b.

equation (3.31) have the same amplitudes, which make gradient information essential when determining which component to track. The maxima tracking results indicate that the proposed method, indicated by (b), performs significantly better than the method not incorporating acceleration information into the maxima tracking process, indicated by (a). If no acceleration information is incorporated into the maxima tracking process, then the constant frequency component is tracked because it satisfies the

assumption made when using equation (3.6). The phase error in figure 3.3b reflects the maxima tracking results. The proposed method resulted in a relative phase error, computed from,

$$e_R = \frac{\max\left(e(t)\right)}{\max\left(\phi_a(t)\right)} \tag{3.32}$$

of 0.1830%. The error $e(t)$ is computed from equation (3.30) and $\phi_a(t)$ is the analytical phase.

The third and last investigated numerical signal contains a shaft component with a rotational speed of,

$$f_r = \frac{1}{5}\cos(0.1\pi t) + \frac{1.25}{\pi} \tag{3.33}$$

in Hertz, which is used to generate a signal in the form of,

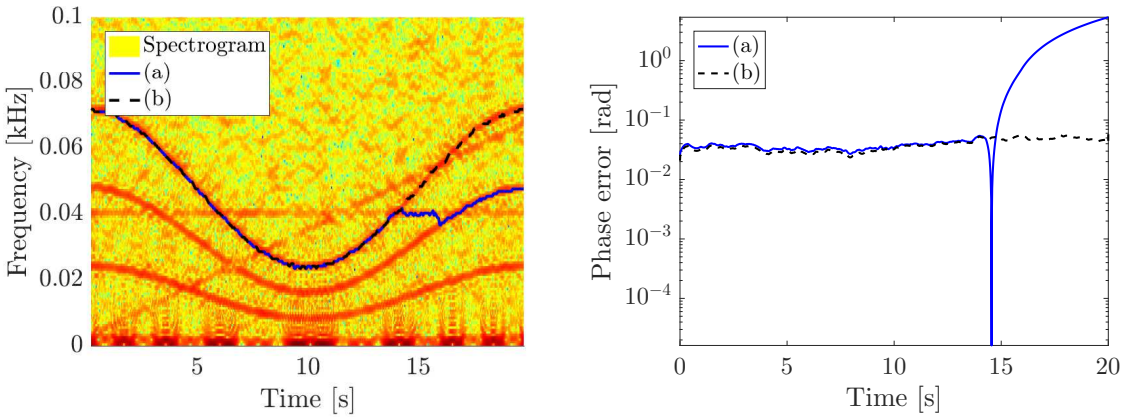
$$x(t) = \sum_{s=1}^{3}\left(1.5\sin\left(2.0\pi s\int_0^t f_r(\tau)d\tau\right) + \sin\left(80.0\pi s\int_0^t f_r(\tau)d\tau\right)\right) + 0.5\sin(2\pi f_1 t) + 0.5\sin(80.0\pi t) \tag{3.34}$$

where,

$$f_1(t) = 1.65t + 2.0 \tag{3.35}$$

is a linearly varying frequency component in Hertz. The signal in equation (3.34) contains three rotational speed and three gear mesh frequency, $40 \times f_r(t)$, harmonics. It contains two components which are independent of the rotational speed as well. In figure 3.4a the spectrogram of the signal in equation (3.34) and the maxima tracking results from the two approaches are presented. The benchmark method, using equation (3.6), is unable to distinguish the component of interest from the other frequency components in the signal. This causes the maxima tracking algorithm to drift off and follow the component with no acceleration. However, the proposed method is able to track the correct component in the presence of other misleading components, since it incorporates angular acceleration information into the maxima tracking process. The phase errors in figure 3.4b indicate that the proposed method performs significantly better.

The performance of the two maxima tracking approaches is summarised in table 3.1. The results indicate that the proposed method performs better than the benchmark maxima tracking approach. The seemingly large phase errors of the first two signals are small relative to the total phase completed by the shaft under consideration. It is concluded that the proposed method obtained significantly lower errors than the benchmark method, which will consequently result in low order tracking errors. This is investigated further on the numerical and experimental gearbox data that are discussed in chapter 2.



(a) (b)

Figure 3.4: The spectrogram and the maxima tracking results from the third numerical signal in figure 3.4a resulted in the phase error in 3.4b.

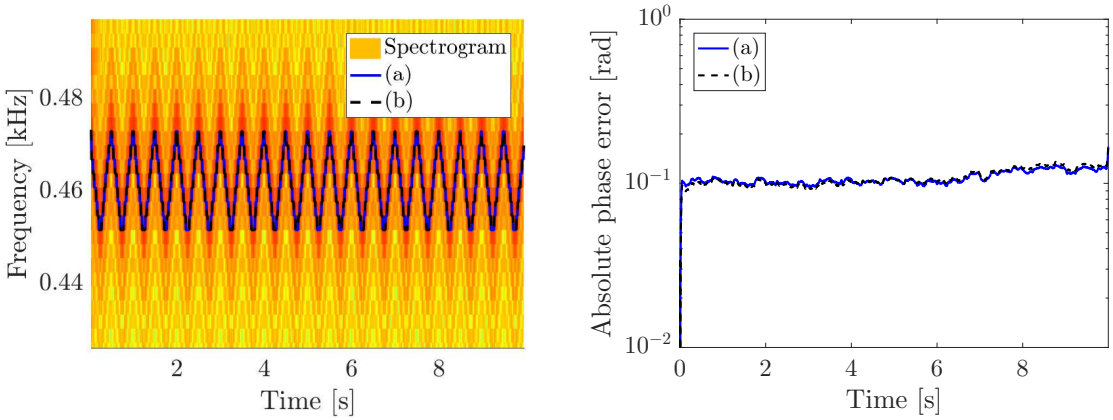
Table 3.1: Summary of the maxima tracking results from the numerical signals. All the phase units are in radians.

	Signal 1	Signal 2	Signal 3
Maximum phase completed	251324.899	251324.899	314.1593
Benchmark method by Urbanek et al. (2013):			
RMS phase error	576.3152	54636.0666	1.4576
Maximum phase error	967.3627	141189.2577	5.3663
Proposed method:			
RMS phase error	540.9771	459.872	0.039169
Maximum phase error	935.9197	829.7571	0.063935

3.2.2 Numerical gearbox model validation

The numerical vibration signal, obtained from the fifth degree of freedom of the gearbox model in section 2.1, is investigated in this section. The spectrogram of the vibration signal and the results from the two maxima tracking approaches are presented in figure 3.5a.

The phase calculated from the Vold-Kalman filtered vibration signal with the Hilbert transform resulted in the phase errors presented in figure 3.5b. The phase errors in figure 3.5b is calculated from equation (3.30). The output shaft of the gearbox completed 115.4570 rotations in 10sec. This resulted in a relative phase error of 0.0231% by the benchmark maxima tracking approach. The proposed maxima tracking method obtained a worse relative phase error of 0.0238%. The proposed method performs slightly



(a) (b)

Figure 3.5: The maxima tracking results are compared in figure 3.5a on the spectrogram, where the result of the benchmark and the proposed method are indicated by (a) and (b), respectively. The resulting phase errors are presented in figure 3.5b.

worse, because the angular acceleration of the gearbox shaft is discontinuous which violates the constant angular acceleration assumption made. Since the benchmark method does not incorporate any acceleration information into the maxima tracking process, it results in a better estimation of the IF. However, the operating conditions are not realistic for real machines, since real machines have inertia which does not allow such rapid operating condition changes. The resulting spectrum of the order tracked signal using the proposed method and the analytical phase is compared in figure 3.6.

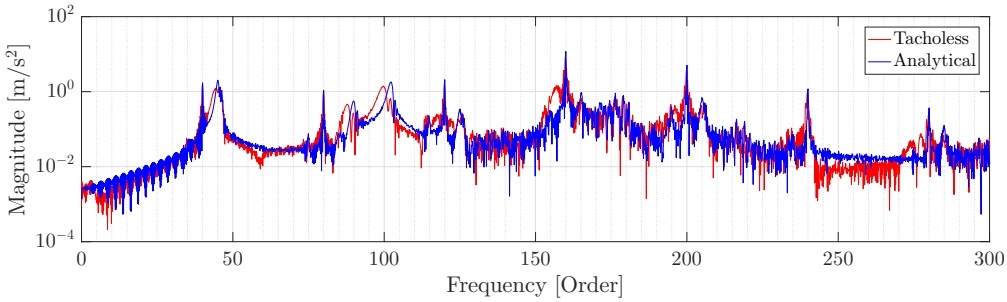


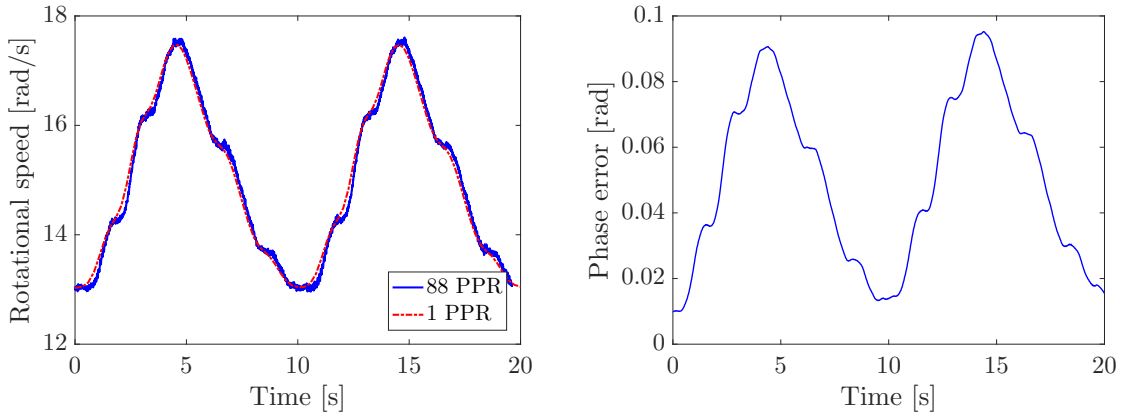
Figure 3.6: The spectrum of the order tracked numerical gearbox vibration signal, obtained from the analytical phase and the tacholess order tracking method.

The gear mesh frequencies, occurring at 40, 80, etc. orders, are approximately the same for the computed and the tacholess order tracking methods. However, the natural frequencies are different in figure 3.6 which indicate that rotational speed independent components are more sensitive to order tracking errors. The order tracking results, obtained from the proposed method, are sufficient for the purposes of this study.

3.2.3 Experimental validation

3.2.3.1 Operating conditions

Three operating condition profiles are investigated on the experimental setup, presented in chapter 2, to evaluate the performance of the proposed tacholess order tracking method. The rotational speed from the 88 PPR signal is superimposed with the rotational speed profile calculated from the 1 PPR signal obtained from the shaft key and proximity probe in figure 3.7a. The rotational speed profile from the 1 PPR signal is incapable of obtaining the correct speed profile due to the large speed fluctuations and its limited resolution. The phase that is calculated from the zebra tape shaft encoder is used as the analytical phase in this section. The phase difference between the two signals are compared in figure 3.7b, which highlights that the phase errors, from the 1 PPR signal, are rotational speed dependent.



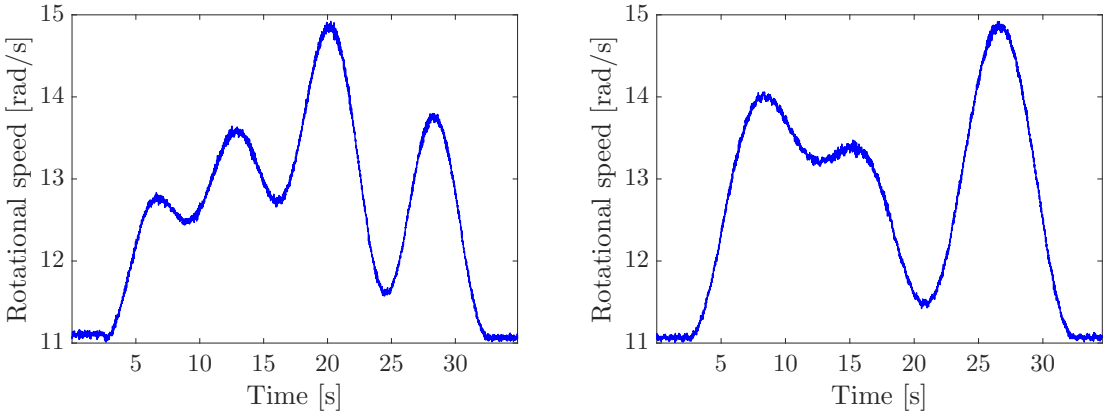
(a) (b)

Figure 3.7: The first experimental rotational speed profile. In figure 3.7a the calculated rotational speed of the two tachometers are compared and the resulting phase difference between them are presented in figure 3.7b.

Even though the equipment, required to obtain a single PPR tacho signal, is relatively easy to implement, it is not well suited for rotating machinery with relatively large angular accelerations. The phase errors in figure 3.7b result in additional phase distortion when performing computed order tracking. This emphasises that shaft encoders, that provide a large amount of PPR, must be used to ensure that large angular accelerations are fully captured. However, the shaft encoders also necessitate more expensive equipment, larger sampling frequencies and more data to be stored which increases the cost of performing CBM.

The second and the third experimental profiles in figure 3.8, obtained from the 88 PPR

shaft encoder, contain smooth rotational speed profiles which are aimed to reflect the profiles seen in wind turbines (He et al., 2016). Note that the second and third experimental rotational speed profiles are only investigated in this chapter. The rotational



(a) Profile 2 (b) Profile 3

Figure 3.8: The second and the third rotational speed profiles obtained from the zebra tape shaft encoder (88 PPR) and the optical probe.

speed calculated from the zebra tape shaft encoder tachometer signal is compared to the estimated rotational speed from the maxima tracking process as well as the speed calculated from the phase from the Hilbert transform later in this section.

3.2.3.2 Phase estimation results

The spectrogram, calculated from the vibration signal measured during the operating condition profile in figure 3.7a, is superimposed with the maxima tracking results of the two maxima tracking methods in figure 3.9. The spectrogram is obtained by applying equation (3.1) with a rectangular window and a 90% overlap. The time window length between centres is 0.0280s with a frequency bin width of 3.1250Hz. The true IF is masked by broadband noise at some instances, which impedes the maxima tracking process. The noise and other adjacent frequency components can easily cause the maxima tracking algorithm to drift off from the true IF. Therefore it is crucial to use small bandwidths in the constraints (i.e. Δf_c). The Δf_c in equation (3.15) is 6.4 times the frequency bin width. Five points, with a total time difference of 0.14s, are used for estimating the two coefficients of the first order polynomial in equation (3.10). It is assumed that the angular acceleration of the system will be approximately constant within 0.14s and therefore a first order polynomial is appropriate. If a larger number of points are used, then a larger acceleration change can possibly occur between the first and the last measurement which necessitates using higher order polynomials. If no acceleration information is incorporated into the maxima tracking process, the maxima

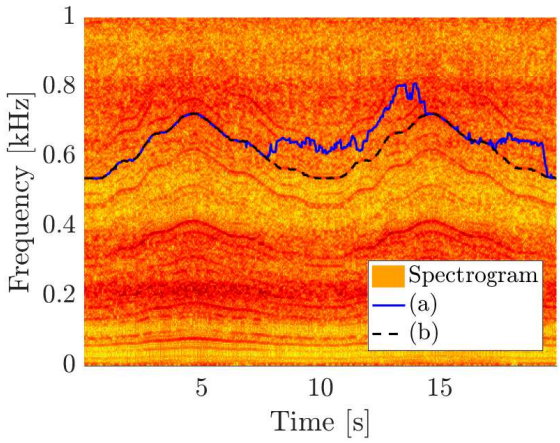


Figure 3.9: The maxima tracking results of the two algorithms are compared in this figure on the first experimental profile (see figure 3.7), with the benchmark method indicated by (a) and the proposed method indicated by (b).

tracking algorithm is unable to handle the large frequency changes and it drifts off from the true estimate in the presence of noise as shown in figure 3.9. The proposed method, indicated by (b) in figure 3.9, leads to a more robust maxima tracking procedure as compared to the benchmark method (indicated by (a)).

The spectrogram of the filtered signal obtained from applying a one-pole, angular-displacement VKF on the maxima tracked line obtained from the proposed method (see figure 3.9) is shown in figure 3.10. The mono-component signal in figure 3.10 indicates that the Hilbert transform can be used to obtain the phase of the frequency component over time. The frequency component in figure 3.10 is clearly related to the rotational speed profile in figure 3.7.

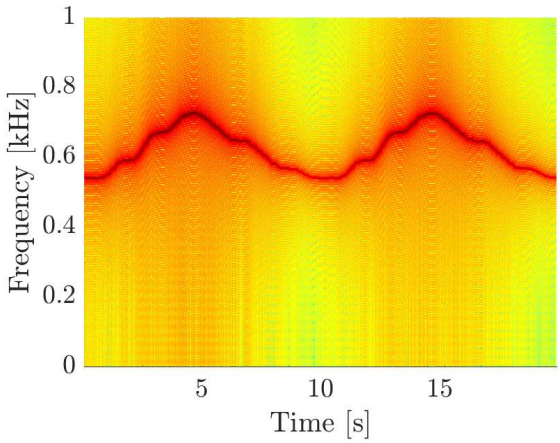
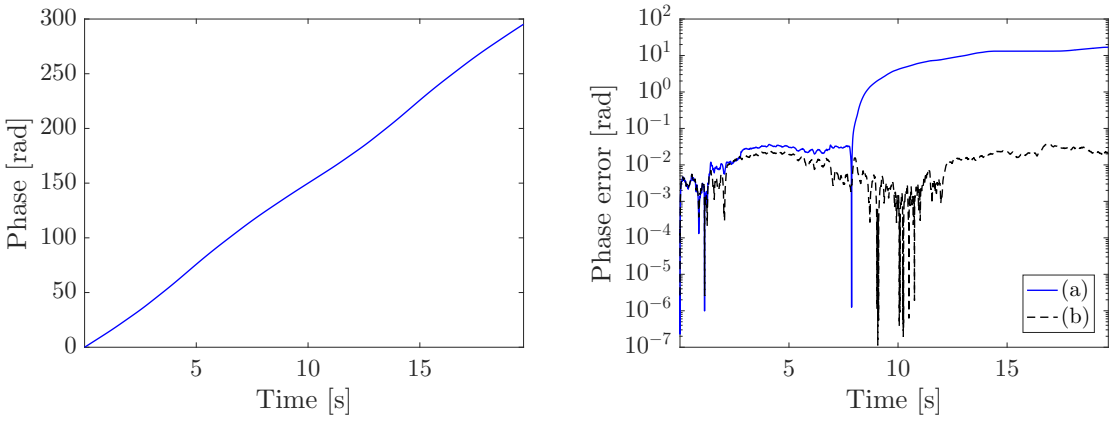


Figure 3.10: The spectrogram of the VKF filtered signal is presented in figure 3.10. A one-pole angular-displacement VKF with a centre frequency equal to the maxima tracked frequency (b) in figure 3.9 is used.

The phase in figure 3.11a is calculated from the 88 PPR shaft encoder located on the input shaft of the gearbox and is used to critically compare the errors in the phase estimation process. The phase profile in figure 3.11a serves as an example of the mapping between the time and the angle domain, which is used in the resampling process. The phase errors made by the benchmark method, in figure 3.9 are directly reflected in the phase errors in figure 3.11b and is significantly larger than the phase errors that resulted from the proposed method. The phase error of the proposed



(a) (b)

Figure 3.11: The instantaneous phase of the input shaft of the monitored gearbox, calculated from the zebra tape shaft encoder is given in figure 3.11a, while the absolute phase error by the two maxima tracking approaches are given in figure 3.11b. The results are presented for the first experimental rotational speed profile.

method is significantly smaller than the phase error that resulted from the 1 PPR signal in figure 3.7b as well. Note that all phase errors are computed from equation (3.30), with the analytical phase ($\phi_a(t)$) calculated from the tachometer signal obtained from the zebra tape shaft encoder and the optical probe.

The maxima tracking results as well as the phase estimation errors are given in figure 3.12 and figure 3.13 for the second and third experimental profiles. The proposed method performs slightly worse than the benchmark maxima tracking method for the second rotational speed profile in figure 3.12. However, in figure 3.13 a significant improvement is obtained by the proposed maxima tracking method compared to the benchmark maxima tracking method.

The resulting phase estimation errors from the two investigated maxima tracking methods are compared in table 3.2 for the experimental rotational speed profiles. The relative phase error, computed by equation (3.32), is 0.012769% for the first signal, 0.0096486% for the second signal and 0.01142% for the third signal. The aforementioned results and the results in table 3.2 prove that the proposed maxima tracking

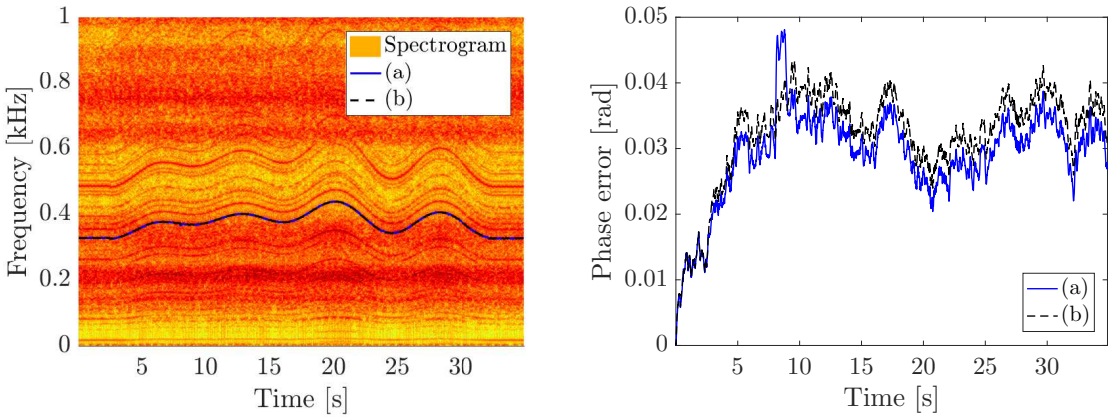


Figure 3.12: The spectrogram of the vibration signal resulted from the second rotational speed profile (see figure 3.8a) and the result of the two maxima tracking approaches in figure 3.12a. The resulting phase estimation error is given in figure 3.12b.

Table 3.2: Summary of the maxima tracking results from the experimental rotational speed profiles. All the phase units are in radians.

	Profile 1	Profile 2	Profile 3
Maximum phase completed	295.2322	439.753	439.7533
Benchmark method by Urbanek et al. (2013):			
RMS phase error	8.3951	0.030019	0.053025
Maximum phase error	16.8094	0.048112	0.069044
Proposed method			
RMS phase error	0.016662	0.033053	0.037182
Maximum phase error	0.037697	0.04343	0.050224

algorithm obtain very good phase estimation results and that it is an improvement of the benchmark method.

3.2.3.3 Order tracking results

The purpose of the proposed methodology is to order track a vibration signal in the presence of large rotational speed fluctuations with only an accelerometer. The spectrum of the vibration signal obtained from the first experimental operating condition profile, using the computed and tacholess order tracking methods, are compared in figure 3.14. The computed order tracking is performed by using the tachometer signal obtained from the zebra tape shaft encoder and the optical probe.

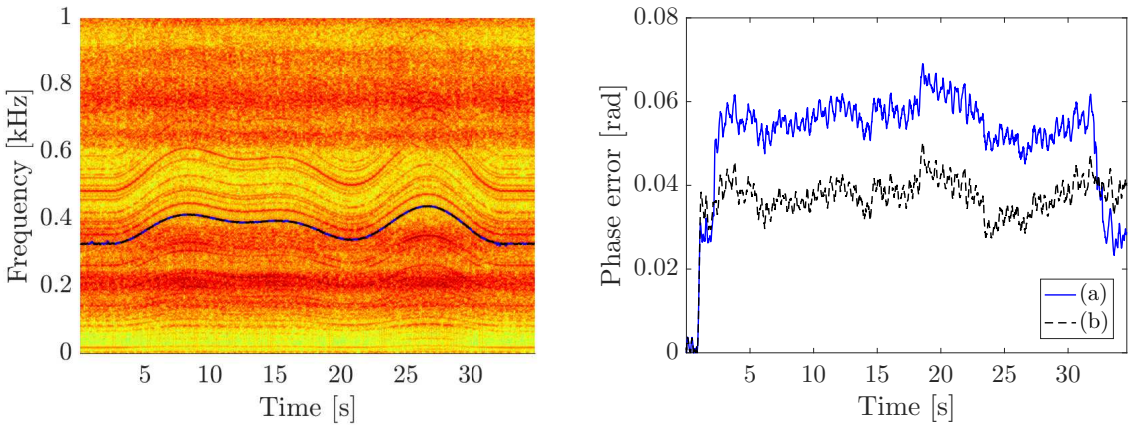


Figure 3.13: The spectrogram of the vibration signal resulted from the third rotational speed profile (see figure 3.8b) and the result of the two maxima tracking approaches in figure 3.13a. The resulting phase estimation error is given in figure 3.13b.

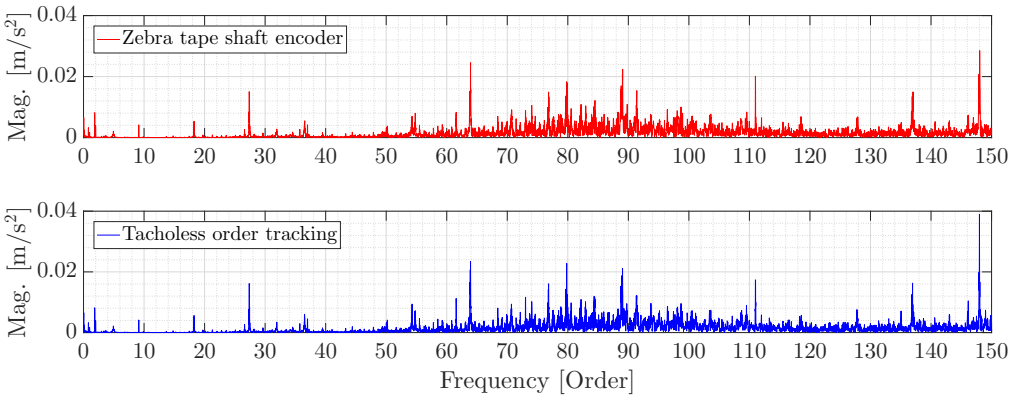


Figure 3.14: Comparison between the spectrum of the order tracked vibration signal using the proposed tacholess order tracking method and using the information from the zebra tape shaft encoder (i.e. computed order tracking).

The results in figure 3.14 indicate that a very similar spectrum is obtained from the computed and tacholess order tracking approaches. This is clearly seen in the low frequencies of the spectrum, presented in figure 3.15. The proposed method is able to capture the input rotational speed of the gearbox (1 order) and its harmonics, the output rotational speed (1.85 orders), the rotational speed of the motor (4.93 orders) and the rotational speed of the alternator shaft (9.1227 orders). Spectral smearing is avoided and the components of interest are clearly highlighted. The results indicate that the proposed methodology is able to order track the vibration signal well.

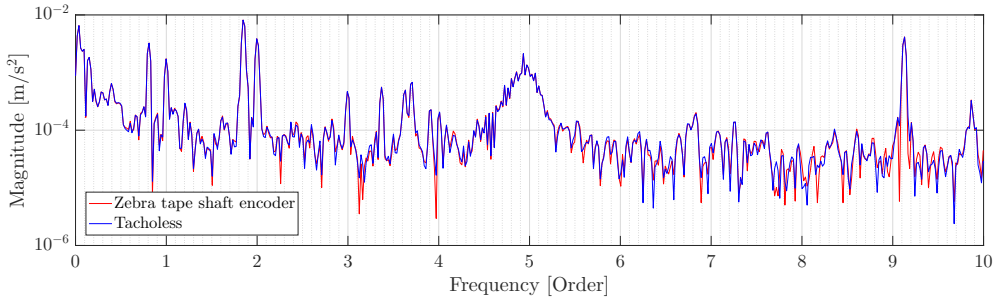


Figure 3.15: The low frequency range of the computed spectrum in figure 3.14.

3.2.3.4 Rotational speed results

The rotational speed can be estimated from the maxima tracked frequency or it can be estimated from the phase calculated from the Hilbert transform. The rotational speed estimated from the maxima tracking process differs from the true estimate due to the fixed resolution of the spectrogram. The results from maxima tracking, the Hilbert transform phase and the IAS from the zebra tape shaft encoder are compared in figure 3.16.

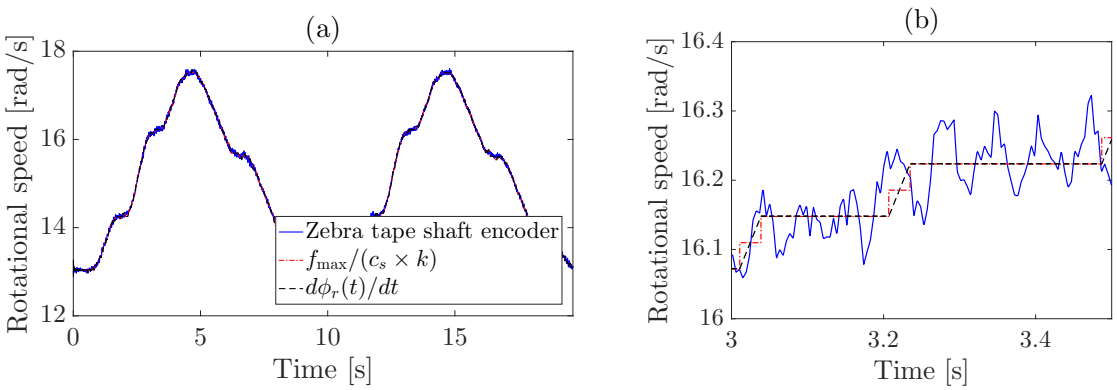


Figure 3.16: Comparison between the three rotational speed estimations for the first experimental profile. The full rotational speed profile is shown in (a) and a zoomed view is shown in (b).

The frequency content of the three estimations of the rotational speed (see figure 3.16) is compared to the frequency content of the measured vibration signal in figure 3.17. The characteristics of the system (i.e. rotational speed components) are present in the spectrum of (a) and (b) in figure 3.17. However, the frequency content of the maxima tracked signal, indicated by (c), and the frequency content of the numerically differentiated Hilbert transform phase, indicated by (d), contain only the input rotational speed of the monitored gearbox. A lot of noise is present in the spectra of (c) and (d) in figure 3.17 which indicates that the frequency content of the estimated rotational speed is limited as compared to the IAS calculated from a shaft encoder. Even though

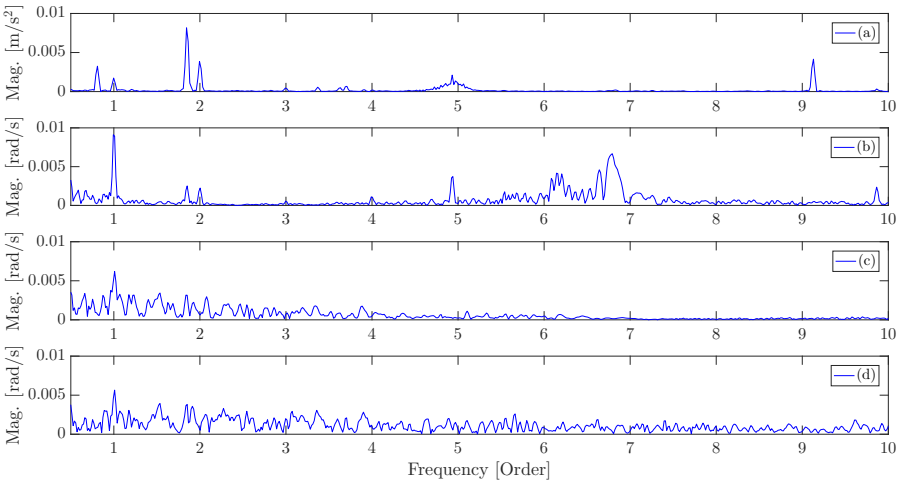


Figure 3.17: The spectrum of the axial component of the tri-axial accelerometer (a), the rotational speed calculated from the shaft encoder (b), the frequency content of the normalised maxima tracked estimate $f_{\max}/(c_s \times k)$ (c) and the derivative of the phase from the Hilbert transformed Vold-Kalman filtered signal (d) are compared.

the rotational speed estimates does not contain a lot of information of the system in its spectrum, the mean rotational speed remains useful for CBM tasks such as calculating the number of fatigue cycles completed by a rotary machine etc.

3.3 Conclusion

A tacholess order tracking methodology is developed which is robust to relatively large angular speeds. The tacholess order tracking uses an improved maxima tracking algorithm which incorporates acceleration information into the process by using a probabilistic approach. The spectrogram frequency content contains ridge smearing due to the fixed time-frequency relationship and therefore linear regression, using a maximum likelihood approach, is used to estimate the gradient characteristics of the instantaneous frequency of interest. The linear regression model of the frequency component is used to predict the region where the maxima tracking algorithm searches for the IF at the next time step. An angular-displacement Vold-Kalman filter is centred at the maxima tracked frequency and a relatively small bandwidth is used to filter the other components out. This mono-component signal is transformed with the Hilbert transform to obtain the instantaneous phase of a shaft over time. The instantaneous phase of the shaft is used to resample the vibration signal from the time to the angle domain. The proposed tacholess order tracking method is successful in the phase estimation and the rotational speed estimation process which can be used to perform CBM on rotary machines.

Chapter 4 Fault diagnostic methodology: Development

In this chapter a gearbox diagnostics methodology, with its process diagram in figure 4.1, is proposed and discussed.

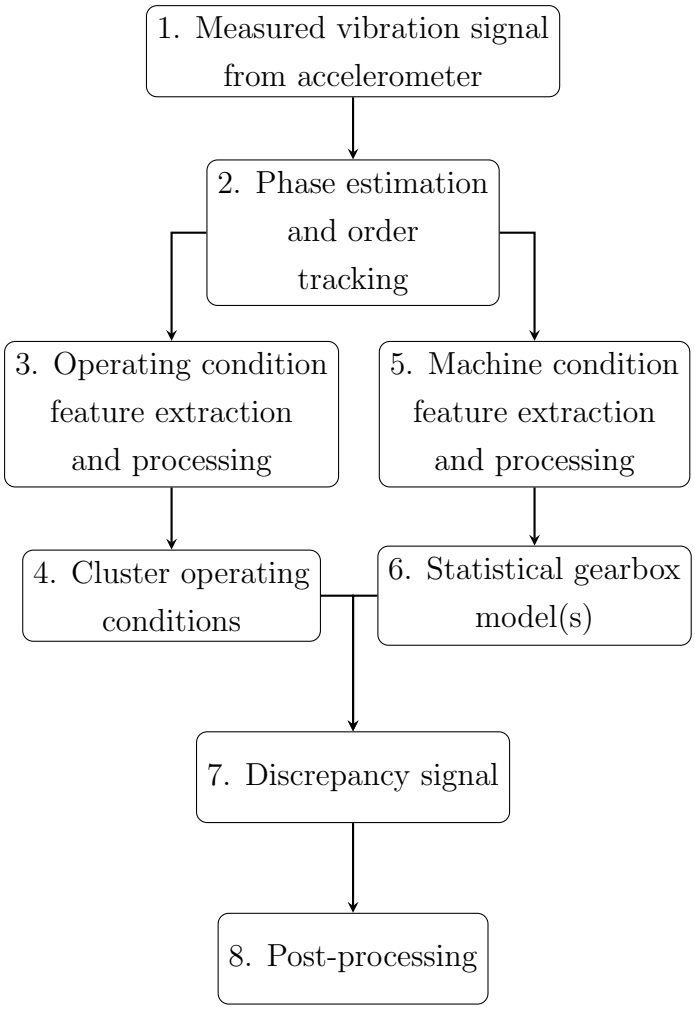


Figure 4.1: Process diagram for the proposed gearbox diagnostics methodology.

The aim is to develop a fault diagnostic methodology that adheres to the requirements

set out in section 1.3. It is essential to convert the time domain vibration signal to the angle domain so that the angle-cyclostationary properties can be preserved. This is performed with the proposed tacholess order tracking methodology in chapter 3 to ensure that the diagnostic methodology is cost-effective to implement and to apply. However, computed order tracking can be used if the required equipment is available or feasible to acquire.

Operating condition features are extracted from the order tracked signal and modelled with an operating condition model. Machine condition features are extracted from the order tracked vibration signal and the operating condition information is used to assign operating condition states (or labels) to the machine condition features. A machine condition model is optimised on the machine condition features from each operating condition state.

The information from the operating and machine condition models are combined using probabilistic approaches to generate a discrepancy signal which quantifies the deviation from the expected behaviour of the machine condition features for the present operating conditions. The information from the machine condition and operating condition models are combined by first considering the joint distribution over all machine and operating condition features, model parameters and operating condition model latent states. The states and the model parameters are marginalised out to obtain a marginal distribution over the machine and operating condition features. The discrepancy signal is obtained from the negative log-likelihood of the probability distribution of the machine condition features statistically conditioned upon the operating condition features. This process is thoroughly discussed in section 4.2.3. The discrepancy signal is subsequently post-processed so that fault detection, location and trending is possible in fluctuating operating conditions. Take note that in this diagnostic methodology the features generated from a healthy machine is modelled and not the machine itself (as opposed to physics-based models).

The training process follows steps 1 to 6, while the evaluation process follows steps 1 to 8 in figure 4.1. Each step in the process is explained in detail in the subsequent sections. The vibration signal obtained from the tri-axial accelerometer in figure 2.8 with the operating conditions in figure 2.6 is used to illustrate the concepts in this chapter unless stated otherwise.

4.1 Feature extraction

A two-phase feature extraction process is proposed to extract the operating and machine conditions of the gearbox. The necessity of this is illustrated in figure 4.2 where a revolution-frequency distribution is computed from an order tracked vibration signal. If machine condition features are extracted from the revolution-frequency distribution,

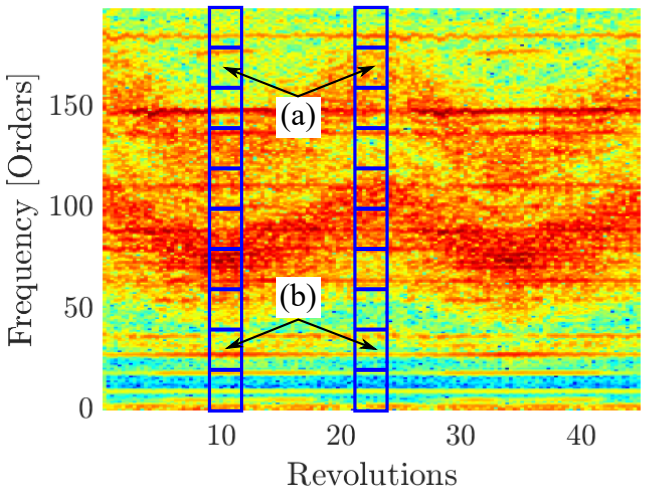


Figure 4.2: Illustration of ten features extracted from a revolution-frequency distribution at different shaft revolutions. A feature is located within a square at a specific revolution along the frequency axis. Some features contain dominant rotational speed information and some do not as indicated by (a) and (b), respectively.

then it will contain machine condition information as well as operating condition information (see (b) in figure 4.2). Hence if the machine condition features are considered in isolation, it is difficult to ascertain the cause of the magnitude changes in the features. If operating condition information is extracted and used when investigating the machine condition features, the cause of changes can be attributed either to changes in operating condition or changes in machine condition.

4.1.1 Fault localisation

A single set of machine condition features can be extracted from the raw vibration signal, but the features are insensitive to incipient faults and fault localisation is impossible. In the speech recognition field, which exhibits many similarities to the rotary machine diagnostic field, an audio signal is windowed into constant time segments from which features are extracted (Lee, 1990). Even though the speed at which the words are pronounced is unknown, most humans speak at approximately the same speed and

hence it is relatively simple to fix a window length in the time domain. However, it is not possible to fix a window in the time domain for rotary machine data measured in fluctuating speed conditions, since the characteristics within each window are dependent on the rotating speed. This motivates extracting features from the angle domain, since the rotary machines are inherently angle-cyclostationary (Antoni et al., 2004).

An example of a windowing process is illustrated in figure 4.3 to allow gear fault localisation. Each rectangular window in figure 4.3 contains the same characteristics

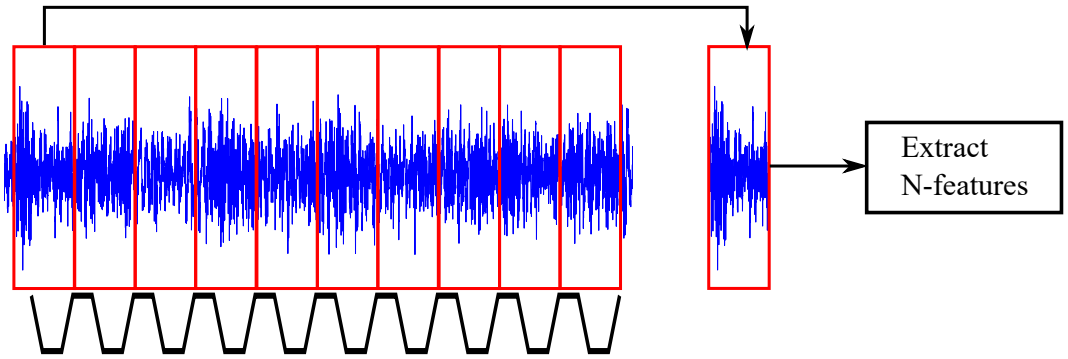


Figure 4.3: An illustration of the windowing process which allows angle localised features to be extracted from a signal. This rack gear emphasises that the windows have a physical significance in the angle domain for rotating machinery.

(i.e. data generated from the meshing tooth) and therefore the features from each window are supposed to be the same under constant operating conditions. However, if a tooth of a gear is damaged, then the features associated with the tooth differs at each rotation from the expected behaviour, which helps with fault localisation. To ensure that the information within the windows remain consistent, only small order tracking errors are allowed.

The window length is critical to the success of the feature extraction process. Features extracted from a short window is sensitive to non-condition related outliers (i.e. noise) and too long windows make fault localisation impossible. Window overlapping is beneficial, since this allows larger windows to be used with little loss in fault localisation capabilities. This windowing process is applied to extract operation and machine condition features from the order tracked vibration signal in the proposed methodology.

4.1.2 Operating condition features

The mean estimated rotational speed, from the maxima tracking process, within an operating condition feature window is extracted as the first operating condition feature. The aforementioned feature does not necessarily reflect the instantaneous load in the

system during data acquisition and therefore additional features need to be extracted. The load sensitive features are extracted from the spectrogram of the order tracked vibration signal. The amplitudes of the fundamental gear mesh frequency and its four harmonics are extracted from the spectrogram, as shown in figure 4.4.

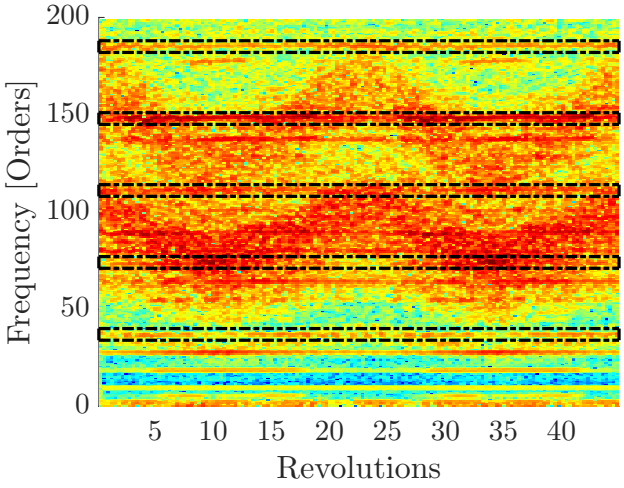


Figure 4.4: Operating condition features extracted from the spectrogram of the order tracked signal, where the abscissa denotes shaft revolutions.

It is expected that additional amplitude modulation from fluctuating loads is reflected in the operating condition features extracted from the spectrogram. Bartelmus and Zimroz (2009*a*) found that features from the short-time acceleration spectrum contain operating condition information. Bartelmus and Zimroz (2009*b*) used a filtered vibration signal only consisting of the gear mesh frequency to estimate the instantaneous load present in a bucket wheel excavator, which supports the idea that the features should be extracted from the gear mesh frequency of the monitored gearbox. The sensitivity of the spectrogram to impulses is influenced by the window length that is used. It is expected that by making the windows sufficiently long, the presence of impulses due to localised damage are masked by the operating condition effects. It is assumed that the operating conditions are approximately constant for one gear rotation and therefore the window length, used to extract the operating condition features, is one gear rotation. Another motivation for using long angular windows for the operating condition feature extraction is addressed in section 4.2.2.

The spectrogram features and the rotational speed features have different magnitudes which cause the features with larger magnitudes to dominate the operating condition modelling and clustering process. All the operating condition features are linearly scaled so that each dimension in the feature space has the same order of magnitude. Each feature is scaled independently between $[-10, 10]$, where the new dimension is chosen for convenience. The scaling coefficients for each dimension are obtained from

the training data and is applied to the training, validation and testing operating condition features. Redundant information is present in the spectrogram features, since the gear mesh frequency and its harmonics are extracted. This redundant information is removed by performing PCA on the data, where PCA is discussed in section 4.1.4.

4.1.3 Machine condition features

The wavelet coefficients have proven to be sensitive to the impulses generated by localised and distributed faults (see section 1.2.1.9) and therefore machine condition features are extracted from it. A wavelet coefficient is obtained at a scale (or frequency value) for each step along the order tracked vibration signal. Machine condition features are extracted from the windowed wavelet coefficients at each scale or frequency band. Two distinct machine condition feature sets are separately investigated in this study,

- Four features are extracted from the windowed continuous wavelet coefficients at each scale. Specific scales are investigated which correspond to frequency components with known diagnostic information.
- A single feature is extracted from the windowed wavelet packet coefficients at each frequency band of the whole vibration signal spectrum. It is assumed that the frequencies which contain diagnostic information are unknown.

4.1.3.1 Continuous wavelet transform features

The continuous wavelet coefficients are extracted from the discrete vibration signal, with equation (1.1), by using discrete time steps, but with preselected scales. The wavelet basis function as well as the scales, a_1, have to be selected before the CWT can be calculated. The Meyer wavelet basis function is investigated, since it was effective in the study by Jedliński and Jonak (2015) for gearbox fault detection. The CWT of the gearbox vibration signal in a healthy and a damaged condition in figure 4.5 is obtained by using equation (1.1) with the Meyer basis functions.

The resonance bands located between 300 and 800 orders in figure 4.5b contain operating condition information and indicates the possibility of damage within the gearbox if compared to figure 4.5a. It is difficult to ascertain the cause of the higher amplitude levels in the range of 300 and 800 orders (i.e. is it localised or distributed damage) as well as the extent of the damage from figure 4.5b. The resonance bands are a function of the system and with a novelty detection approach the information in figure 4.5b is not known beforehand. The gear mesh frequencies of a gearbox are easily calculated and

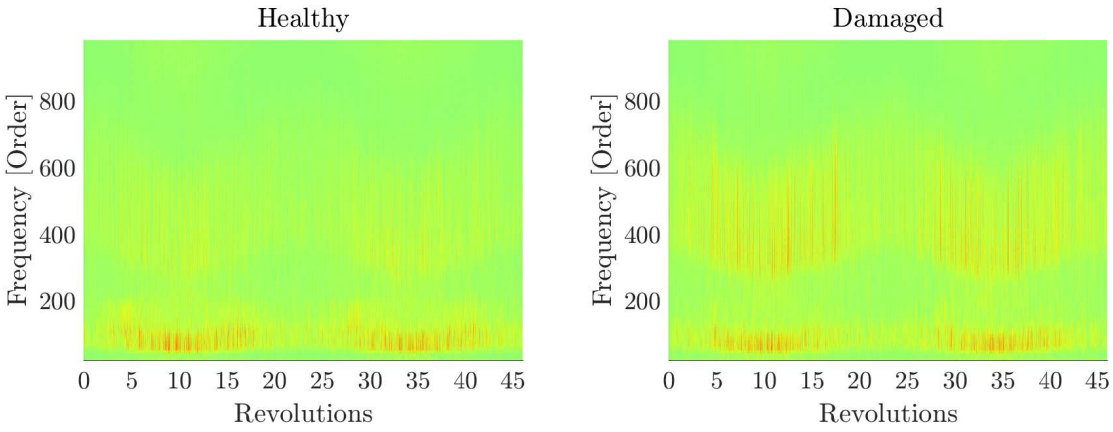


Figure 4.5: Continuous wavelet transform (CWT), with a Meyer wavelet basis function, of the vibration signals from a gearbox with a healthy and a damaged gear presented in figure 4.5a and 4.5b, respectively.

contain diagnostic information as well (Dalpiaz et al., 2000, Randall, 1981). Therefore the continuous wavelet coefficients are calculated for 20 scales in a bandwidth of three orders around the first five gear mesh frequencies so that features can be extracted from it. This results in 100 scales at each angular window.

The wavelet coefficients are windowed with a rectangular window so that features can be extracted from each window at each scale. The window length of the wavelet coefficients at each scale is $2\pi/N_{teeth}$ radians with a 50% overlap between windows, where N_{teeth} denotes the number of teeth on the gear of interest. The statistical functions, presented in table 4.1, are applied to the windowed wavelet coefficients at each scale and are used as machine condition features. The four features in table 4.1 result in a 400 dimensional machine condition feature space.

Table 4.1: The features extracted from the continuous wavelet coefficients W at scale $a_1[s]$. Note that the wavelet coefficients are windowed with rectangular windows and therefore the summation is performed over the N non-zero wavelet coefficients. The mean of the non-zero wavelet coefficients at scale $a_1[s]$ is denoted by $\mu_W[s]$.

Energy	$\sum_i W(a_1[s], a_2[i])^2$
Skewness	$\frac{\frac{1}{N}\sum_i (W(a_1[s],a_2[i]) - \mu_W[s])^3}{\left[\frac{1}{N}\sum_i (W(a_1[s],a_2[i]) - \mu_W[s])^2\right]^{1.5}}$
Kurtosis	$\frac{\frac{1}{N}\sum_i (W(a_1[s],a_2[i]) - \mu_W[s])^4}{\left[\frac{1}{N}\sum_i (W(a_1[s],a_2[i]) - \mu_W[s])^2\right]^{2}}$
RMS	$\sqrt{\frac{1}{N}\sum_i W(a_1[s], a_2[i])^2}$

4.1.3.2 Wavelet packet transform features

The wavelet packet transform (WPT) discretizes the frequency axis into fixed bands as opposed to the CWT. The WPT is implemented by low-pass and high-pass filtering the signal to obtain approximation and detail coefficients, respectively. The approximation and detail coefficients at the previous level are used to obtain the sets of approximation and detail coefficients at the next level. The bandwidth, as well as the sampling frequency, are halved with each level of decomposition. The WPT is discussed in section 1.2.1.9 and a more detailed overview is given by Addison (2002).

The WPT spectrum of the healthy and damaged gearboxes in figure 4.6 is obtained from the db1 wavelet basis function with a level four wavelet decomposition. This leads to the frequency band being decomposed into 2^4 bins. The db1 wavelet basis function is motivated by the results by Rafiee et al. (2010) and Jedliński and Jonak (2015).

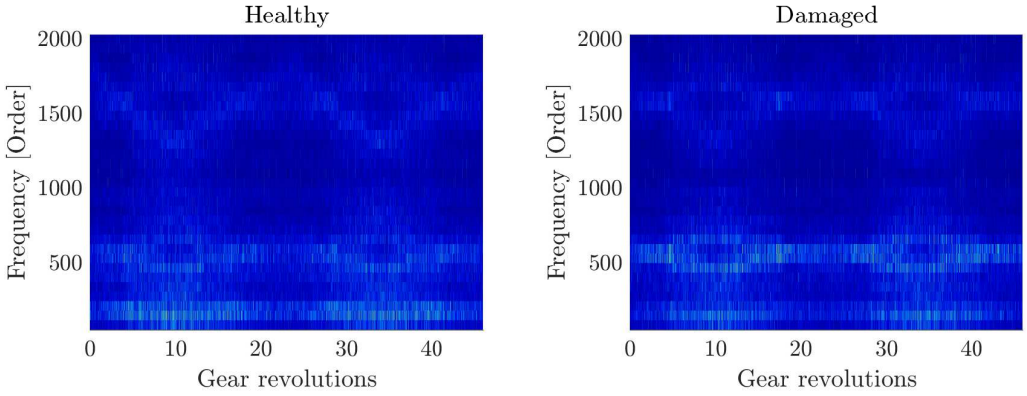


Figure 4.6: Wavelet packet transform (WPT), with a Daubechies db1 wavelet basis function using a level four decomposition, of the vibration signal from a gearbox with a healthy and a damaged gear.

The same resonance bands identified in the CWT of the vibration signal is present in the damaged spectrum in figure 4.6. It is evident from figure 4.6 that the frequency resolution of the WPT is poorer compared to the CWT. However, the WPT has the following advantages over the CWT:

- The coefficients are orthogonal.
- Since the coefficients are decomposed in terms of bands, some error is allowed in the phase estimation procedure in the tacholess order tracking process.
- The exact characteristic fault frequencies do not have to be known or determined beforehand.

The purpose of the WPT is to use a naive approach to select features where the location

of the fault information in the frequency spectrum is unknown. The WPT coefficients are windowed similarly to the CWT, but only the energy of the windowed wavelet coefficients, found in table 4.1, is calculated to minimise the feature extraction time.

4.1.4 Dimensionality reduction techniques

The large dimensional machine condition feature space as well as the possibility of redundant information within both the machine and operation condition feature spaces, motivate the use of dimensionality reduction techniques. For example, 80,600 parameters need to be estimated when modelling a 400 dimensional feature space with a single Gaussian distribution. This large dimensional feature space causes problems when performing model optimisation and when insufficient training data are available. Principal component analysis (PCA) is used in this study, since it has been effective on a range of problems (see section 1.2.2.2) and it does not require additional hyperparameter optimisation. This makes it ideal for novelty detection problems.

The eigenvalues and the eigenvectors of the covariance matrix of the features are obtained in the first step of PCA. The eigenvectors of the covariance matrix are orthogonal and are the new axes of the transformed feature space. Small eigenvalues indicate that the highly correlated features exist in the original feature space and therefore the new principal feature space is selected to consist only of the principal directions with large eigenvalues (i.e. with a large information contribution). The d-largest eigenvalues and corresponding eigenvectors are used to transform the dataset from the original D-dimensional feature space to a d-dimensional feature space with,

$$\boldsymbol{y}_n = \boldsymbol{V}_{1:d}^T (\boldsymbol{x}_n - \boldsymbol{\mu}_{\boldsymbol{X}}) \tag{4.1}$$

where the mean of the dataset $\boldsymbol{X}$ and the eigenvectors of the d-largest eigenvalues are denoted by $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{V}_{1:d}$, respectively. A D-dimensional sample $\boldsymbol{x}_n$ is transformed to a d-dimensional vector $\boldsymbol{y}_n$ with equation (4.1) at increment n. The transformed space can be whitened to reduce the covariance of the new feature space to an identity matrix by $y_{in} \leftarrow y_{in}/\sqrt{\xi_i}$, where ξ_i is the eigenvalue associated with the ith component (Bishop, 2006). The accumulative contribution rate (ACR) of principal component k is calculated from,

$$\text{ACR}_k = \frac{\sum_{j=1}^{k} \xi_j}{\sum_{i=1}^{D} \xi_i}, \quad \text{with } 1 \leq k \leq D, \tag{4.2}$$

where ξ_i is the ith eigenvalue and the ACR is used to select the dimensionality of the new feature space by using a predefined threshold.

4.2 Machine learning models

The operating and machine condition features, extracted from the processed order tracked vibration signal, are modelled using machine learning approaches in this study. Hidden Markov models (HMM), a probabilistic modelling approach introduced in section 1.2.2.3, are used to model the operating and the machine condition features and the motivation for this are given in the subsequent sections.

A HMM contains a latent variable $\boldsymbol{Z}$ representing discrete states which follows a Markov process. The joint distribution of the latent variable over N_T increments is decomposed with the product rule of probability,

$$p(\boldsymbol{Z}) = p(\boldsymbol{z}_1) \prod_{i=1}^{N_T} p(\boldsymbol{z}_i | \boldsymbol{z}_{1:i-1}) \tag{4.3}$$

and is simplified to,

$$p(\boldsymbol{Z}) = p(\boldsymbol{z}_1) \prod_{i=1}^{N_T} p(\boldsymbol{z}_i | \boldsymbol{z}_{i-1}) \tag{4.4}$$

if the latent variable $\boldsymbol{Z}$ follows a first order Markov process, where $p(\boldsymbol{z}_{1:i-1})$ is a compact form of $p(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_{i-1})$. The latent variable uses a 1-of-K coding scheme, where $z_{nk} = 1$ if the active hidden state at time increment n is k and $z_{ni} = 0$ if $i \neq k$. The transition probability, denoted by $p(z_{nj}|z_{n-1\,i})$, is discrete and time-invariant in this study.

The likelihood function of a HMM (Bishop, 2006),

$$p(\boldsymbol{X}) = \sum_{\boldsymbol{z}_1} p(\boldsymbol{z}_1) \sum_{\boldsymbol{z}_2} \cdots \sum_{\boldsymbol{z}_{N_T}} \left[\prod_{i=2}^{N_T} p(\boldsymbol{z}_i | \boldsymbol{z}_{i-1}) \prod_{n=1}^{N_T} p(\boldsymbol{x}_n | \boldsymbol{z}_n) \right] \tag{4.5}$$

takes the hidden state sequence as well as the position of the dataset $\boldsymbol{X}$ in the feature space into account. The $p(\boldsymbol{z}_i|\boldsymbol{z}_{i-1})$ is the transition probability of the HMM and $p(\boldsymbol{x}_n|\boldsymbol{z}_n)$ is the observation distribution. A Gaussian observation distribution, in the form of equation (1.2), is used to describe the relationship between a hidden state and its corresponding features. A continuous distribution is used in this study, because no quantisation errors occur as opposed to using a discrete distribution (Theodoridis and Koutroumbas, 2009).

It is significantly more complex to evaluate the likelihood function in equation (4.5), to infer the latent variable over all time increments and to optimise the parameters of a HMM as oppose to a GMM. However, computationally efficient methods have been developed to alleviate the aforementioned problems. The likelihood function in equation (4.5) is computed efficiently by exploiting the graph structure of the HMM, which

is significantly more efficient than a brute force approach. The optimal hidden state sequence is inferred by using the Viterbi algorithm, which maximises $p(\boldsymbol{z}_{1:N_T}|\boldsymbol{x}_{1:N_T}, \boldsymbol{\theta})$. The optimisation procedure is performed by using the expectation-maximisation procedure to find the maximum likelihood or maximum posterior solutions which is efficiently implemented using the forward-backward algorithm. More information on this can be found in the literature by Rabiner (1989) and Bishop (2006).

4.2.1 Operating condition model

The objective of the operating condition model is to identify clusters within the operating condition features which are approximately the same. The operating conditions contain strong sequential characteristics due to physical constraints such as the inertia of the rotary components and it is therefore expected that the operating condition features exhibit sequential properties as well. This makes a HMM well suited to model the operating condition features. The operating condition classification process is illustrated in figure 4.7.

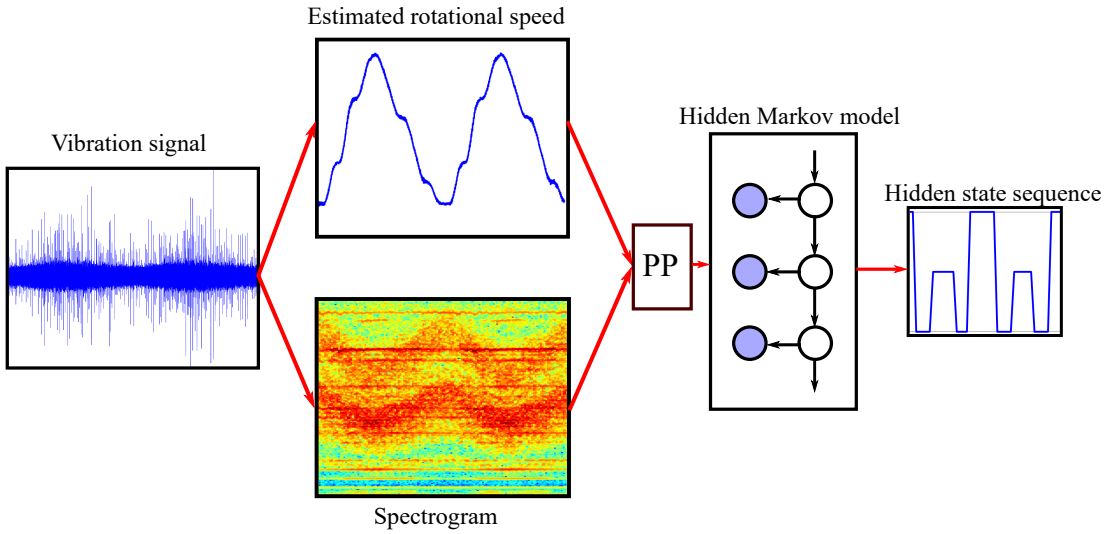


Figure 4.7: Illustration of the operating condition classification process, where pre-processing the features is denoted by PP.

The extracted operating condition features, discussed in section 4.1.2, are pre-processed using linear scaling methods and PCA. The processed features are evaluated with the trained operating condition HMM and a hidden state or operating condition state (OCS) sequence is inferred with the Viterbi algorithm. It is assumed that the operating condition features that are classified to a specific OCS, correspond to approximately the same operating conditions during data acquisition. It should be noted that the OCS sequence is obtained using unsupervised learning and therefore the operating condition state number is assigned arbitrarily to a cluster in the feature space.

4.2.2 Machine condition model

Each operating condition window, used to extract the operating condition features, are classified to a specific operating condition state with the information from section 4.2.1. The machine condition feature windows operate on different data (i.e. wavelet coefficients at each scale) than the operating condition feature windows, but both operate on data along the gear revolution axis. The machine condition features, extracted from windows that overlap with an operating condition feature window, are labelled with the same OCS as the operating condition features in that window. This is illustrated in figure 4.8.

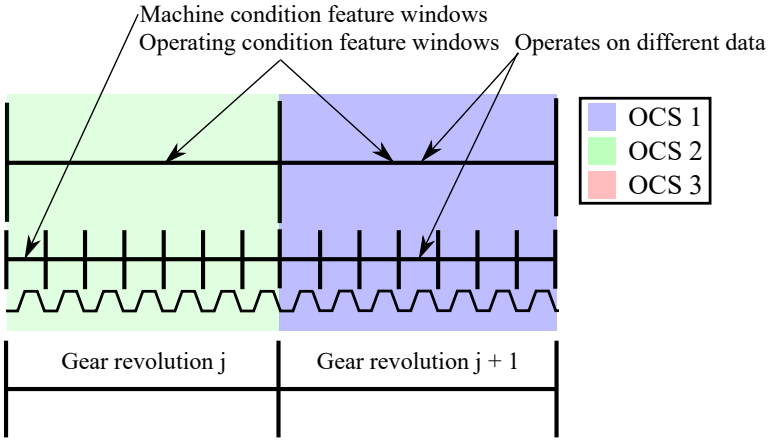


Figure 4.8: The operating condition and machine condition feature windows along the gear revolution axis are shown with the operating condition classification results.

It is assumed that the different sets of machine condition features associated with a specific OCS are similar for a gearbox in the same condition, since the operating conditions are similar according to the operating condition model. A machine condition model is created for each OCS and only trained with the machine condition features classified to the specific OCS. This makes the different machine condition models independent of operating conditions if the operating condition classification is correct. The training process of the machine condition models is illustrated in figure 4.9.

The machine condition features, processed with PCA, are modelled by a HMM, since the HMM has more discriminatory power than GMMs and Gaussian distributions, because of the additional temporal dimension. However the HMM requires features to be sampled consecutively during training and evaluation (see equation (4.5) for example), which requires that consecutively sampled machine condition features be from the same OCS. This restriction is satisfied by using long operating condition windows, where the operating condition windows are discussed in section 4.1.2.

The machine condition models can easily be evaluated on new features, but more work

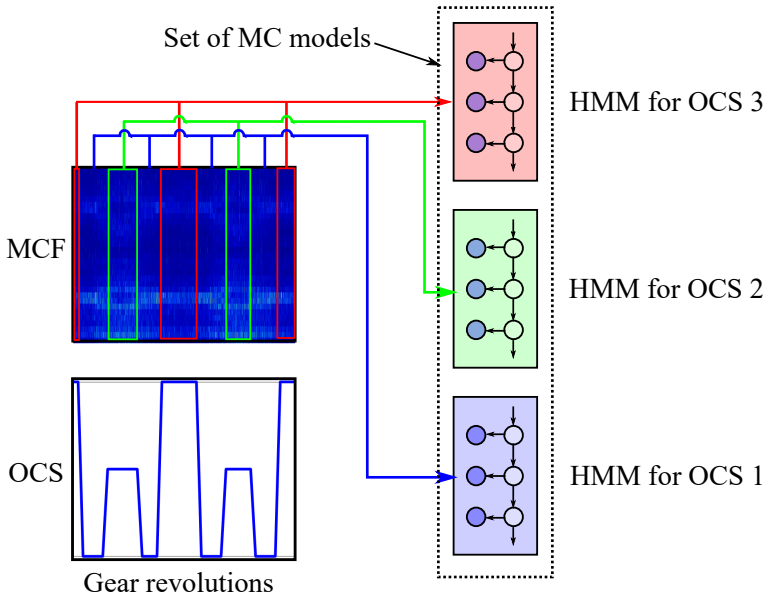


Figure 4.9: Illustration of the machine condition (MC) model training process using the operating condition states (OCS) and the machine condition features (MCF). The different machine condition features are located along the ordinate. Note that the feature processing is not included in the illustration.

is required to generate a discrepancy signal from the machine condition models.

4.2.3 Discrepancy signal generation

The expected behaviour of the machine condition features, extracted from a healthy gearbox, is captured by a machine condition model for each operating condition state. The deviation of the machine condition features from its expected behaviour can be quantified with a discrepancy measure at each machine condition feature window to generate a discrepancy signal. The discrepancy signal can be generated from the same process used during training (see section 4.2.2), similar to Vinson (2014). This means that the OCS classification process determines which machine condition model should generate the discrepancy measure at the corresponding machine condition window. However, this discrepancy signal generation procedure is not optimal, since it does not capture the uncertainty associated with the OCS classification process and gives false confidence in the generated discrepancy signal. It is expected that the discrepancy signal is more robust to detect faults in fluctuating operating conditions by conditioning the machine condition features on the operating condition features instead of the operating condition states. Some operating condition states are more probable than others for the specific operating condition features, which means that the output of some machine condition models are more correct than others. For example, when two operating condition states are approximately equally probable, instead of selecting the

most probable operating condition state to select the machine condition model, both machine condition models can be used to generate the discrepancy signal.

A probabilistic approach is used to generate the discrepancy signal in this section. The first step in deriving the discrepancy signal, is considering independent, identically distributed (i.i.d) machine condition features. The solution of this derivation is modified for HMMs later in this section. The joint distribution over all the random variables at machine condition feature window t,

$$\begin{aligned} p(\boldsymbol{b}_t, \boldsymbol{z}_t^{\text{o}}, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \boldsymbol{o}_t) =& p(\boldsymbol{b}_t|\boldsymbol{z}_t^{\text{o}}, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \boldsymbol{o}_t) p(\boldsymbol{z}_t^{\text{o}}|\{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \boldsymbol{o}_t) \times \\ & p(\{\boldsymbol{\theta}_b\}|\boldsymbol{\theta}_o, \boldsymbol{o}_t) p(\boldsymbol{\theta}_o|\boldsymbol{o}_t) p(\boldsymbol{o}_t) \end{aligned} \tag{4.6}$$

is considered, where the operating condition and the machine condition features are denoted by $\boldsymbol{o}_t$ and $\boldsymbol{b}_t$, respectively. The latent state of the operating condition model is denoted by $\boldsymbol{z}_t^{\text{o}}$ and contains the information of the OCS at machine condition feature window t. The latent states of the machine condition models are not present in equation (4.6) with the explanation provided later in this section. The parameters of the operating condition model and the set of machine condition models are denoted by $\boldsymbol{\theta}_o$ and $\{\boldsymbol{\theta}_b\}$, respectively. The set of machine condition models contain the parameters of the N_{ocs} machine condition models, $\{\boldsymbol{\theta}_b\} = \{\boldsymbol{\theta}_b^1, \boldsymbol{\theta}_b^2, \ldots, \boldsymbol{\theta}_b^{N_{\text{ocs}}}\}$, where $\boldsymbol{\theta}_b^i$ denotes the parameters of the ith OCS. There are N_{ocs} operating condition states in the operating condition model. The full joint distribution in equation (4.6) is simplified to,

$$p(\boldsymbol{b}_t, \boldsymbol{z}_t^{\text{o}}, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \boldsymbol{o}_t) = p(\boldsymbol{b}_t|\boldsymbol{z}_t^{\text{o}}, \{\boldsymbol{\theta}_b\}) p(\boldsymbol{z}_{tj}^{\text{o}}|\boldsymbol{o}_t, \boldsymbol{\theta}_o) p(\boldsymbol{o}_t|\boldsymbol{\theta}_o) p(\boldsymbol{\theta}_o, \{\boldsymbol{\theta}_b\}) \tag{4.7}$$

from the conditional independence properties of the different random variables. The only variables that are important for generating the discrepancy signal are the machine condition features and the operating condition features. Hence the unimportant variables are marginalised out,

$$p(\boldsymbol{b}_t, \boldsymbol{o}_t) = \sum_{j=1}^{N_{\text{ocs}}} \iint p(\boldsymbol{b}_t, z_{tj}^{\text{o}}, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \boldsymbol{o}_t) d\boldsymbol{\theta}_o d\{\boldsymbol{\theta}_b\} \tag{4.8}$$

to obtain a joint distribution over the machine and operating condition features. The integrals in equation (4.8) are complicated to compute and therefore it is assumed that the joint distribution $p(\boldsymbol{\theta}_o, \{\boldsymbol{\theta}_b\})$ is sharply peaked around the variables $\hat{\boldsymbol{\theta}}_o$ and $\{\hat{\boldsymbol{\theta}}_b\}$, which are obtained from maximising the likelihood function. This allows equation (4.8) to be simplified into a joint distribution over the machine and operating condition features,

$$p(\boldsymbol{b}_t, \boldsymbol{o}_t) \approx \sum_{j=1}^{N_{\text{ocs}}} p(\boldsymbol{b}_t, z_{tj}^{\text{o}}, \hat{\boldsymbol{\theta}}_b^j, \hat{\boldsymbol{\theta}}_o, \boldsymbol{o}_t) \tag{4.9}$$

from which the conditional distribution,

$$p(\boldsymbol{b}_t|\boldsymbol{o}_t) \approx \sum_{j=1}^{N_{\text{ocs}}} p(\boldsymbol{b}_t|z_{tj}^{\text{o}}, \hat{\boldsymbol{\theta}}_b^j) p(z_{tj}^{\text{o}}|\boldsymbol{o}_t, \hat{\boldsymbol{\theta}}_o) \tag{4.10}$$

is obtained. The machine condition features are conditioned on the operating condition features in equation (4.10) to generate a discrepancy signal in the form,

$$\eta[t] = -\log p(\boldsymbol{b}_t|\boldsymbol{o}_t) \tag{4.11}$$

which is robust to operating condition changes. The negative log-likelihood (NLL) or the error function over each machine condition window t, denoted by $\eta[t]$, is used as the discrepancy signal in this study. Note that $\eta[t]$ is a discrete signal at machine condition window t and $\boldsymbol{b}_t$ is a multidimensional feature at window t in equation (4.11). The assumption that the joint distribution is sharply peaked around the maximum likelihood estimates, to simplify equation (4.8) to (4.10), is valid if a lot of data is used (Bishop, 1995).

This whole derivation is based on the assumption that the machine condition features are i.i.d, however this is invalid for the machine condition features modelled by a HMM. Equation (4.10) is slightly modified for a set of HMMs, modelling the machine condition features,

$$p(\boldsymbol{b}_t|\boldsymbol{o}_t, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{t-1}) \approx \sum_{j=1}^{N_{\text{ocs}}} p(\boldsymbol{b}_t|z_{tj}^{\text{o}}, \hat{\boldsymbol{\theta}}_b^j, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{t-1}) p(z_{tj}^{\text{o}}|\boldsymbol{o}_t, \hat{\boldsymbol{\theta}}_o) \tag{4.12}$$

where the discrepancy is calculated similarly to equation (4.11). The likelihood of a machine condition model at a specific OCS is weighted with the probability of being at the specific operating condition state. The weights, denoted by $p(z_{tj}^{\text{o}}|\boldsymbol{o}_t, \hat{\boldsymbol{\theta}}_o)$, are easily calculated for latent variable models such as GMM and HMM. It is known as the responsibility function in GMM literature and is easily obtained by using the forward-backward algorithm for HMMs (Bishop, 2006). Since the operating condition features have a different sampling frequency than the machine condition features (see figure 4.8), $p(z_{tj}^{\text{o}}|\boldsymbol{o}_t, \hat{\boldsymbol{\theta}}_o)$ is not strictly available. However, it is assumed that the posterior distribution over the latent states at an operating condition window is the same for all coinciding machine condition windows, which makes $p(z_{tj}^{\text{o}}|\boldsymbol{o}_t, \hat{\boldsymbol{\theta}}_o)$ available.

The computation of the discrepancy signal is illustrated in figure 4.10 for i.i.d machine condition features. The operating condition model is evaluated with the sequence of operating condition features and each of the machine condition models are evaluated with the machine condition features at machine condition window t. A likelihood is generated from each machine condition model i with parameters denoted by $\boldsymbol{\theta}_b^i$ and it is weighted by the posterior distribution over the latent variable $\boldsymbol{z}^{\text{o}}$ of the operating

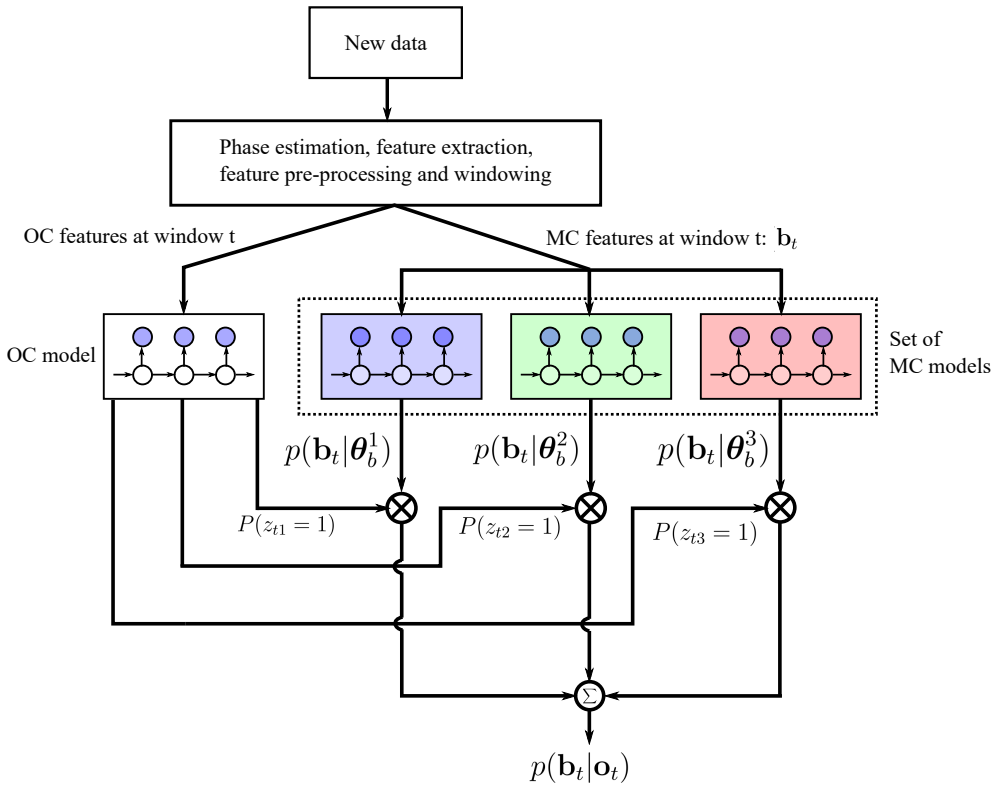


Figure 4.10: The discrepancy signal generation process with equation (4.10) and equation (4.11) is visualised for i.i.d machine condition features.

condition model. A weighted sum is computed to obtain the conditional distribution $p(\boldsymbol{b}_t|\boldsymbol{o}_t)$ where the discrepancy is calculated from equation (4.11).

From figure 4.10 it is concluded that it is important that the operating features must be insensitive to the presence of faults, since the posterior distribution over the latent variable of the operating condition model should only indicate which OCS is more probable. The latent states of the machine condition models are not present in equation (4.6), since it is marginalised out in the computation of the likelihood of the machine condition model $p(\boldsymbol{b}_t|z_{tj}^{\mathrm{o}}, \hat{\boldsymbol{\theta}}_b^j)$.

4.2.4 Model selection

The complexity of a probabilistic model (i.e. the number of parameters) needs to be determined before optimising the model. In the case of novelty detection, it is not possible to use the novelty data (i.e. the gearbox in different damaged conditions) to find the optimal model complexity which is able to detect faults well.

A complex model provides more flexibility to fit complicated data at the cost of a larger optimisation space and also the risk of overfitting on the data. A simple model has

less flexibility to fit complicated data and could result in modelling errors, but it is less likely to overfit on the data. This is known as the bias-variance trade-off (Bishop, 2006) and can be used to determine the appropriate model complexity. The model evidence,

$$p(\boldsymbol{X}|\mathcal{M}_i) = \int p(\boldsymbol{X}|\boldsymbol{\theta}, \mathcal{M}_i) p(\boldsymbol{\theta}|\mathcal{M}_i) d\boldsymbol{\theta} \tag{4.13}$$

for the multidimensional dataset $\boldsymbol{X}$ and model $\mathcal{M}_i$ with parameters $\boldsymbol{\theta}$, favours non-complex models which describe the data well. However equation (4.13) is analytically intractable and therefore approximation methods are used in practical implementations. The Bayesian information criterion (BIC) is computed from the Laplace approximation, a variational approach, which finds approximate solutions for equation (4.13) (Bishop, 2006). The BIC,

$$-2 \times p(\boldsymbol{X}|\mathcal{M}_i) \approx \text{BIC} = -2 \log p(\boldsymbol{X}|\hat{\boldsymbol{\theta}}) + N_\theta \log N_{\boldsymbol{X}} \tag{4.14}$$

needs to be minimised, which results in maximising equation (4.13) with respect to model complexity if the approximation is correct. The first term in equation (4.14) contains the negative log-likelihood of the optimised model parameters, denoted by $\hat{\boldsymbol{\theta}}$, and it penalises poor fits. The second term contains the number of parameters N_θ and the number of training data $N_{\boldsymbol{X}}$ and it penalises model complexity. The BIC penalises model complexity more than other similar methods such as Akaike information criterion (Bishop, 2006). Equation (4.14) is used as a guideline to select the appropriate model complexity in this study.

4.3 Post-processing the discrepancy signal

Three discrepancy signal post-processing techniques are proposed in this study which allow faults to be detected, localised and trended. Automatic fault detection is performed with the fault localisation technique and statistical theory. Therefore, the fault localisation section precedes the fault detection section.

4.3.1 Fault localisation

4.3.1.1 Synchronous average of the discrepancy signal

The discrepancy signal generated, in section 4.2.3, does not intuitively indicate the condition of the gearbox and requires further processing so that the condition of the

machine can be inferred. The synchronous average of the discrepancy signal η,

$$\mu_{\eta}^{(1)}[l] = \frac{1}{N_r} \sum_{i=0}^{N_r-1} \eta[l + iN_s], \text{ where } 1 \leq l \leq N_s \tag{4.15}$$

computed with N_s samples per gear revolution and over N_r rotations, enhances the synchronous components induced by localised faults etc. The synchronous average is used by Heyns, Heyns and De Villiers (2012) and Vinson (2014) to process the generated discrepancy signals.

4.3.1.2 Second synchronous average of the discrepancy signal

The ability of the synchronous averaging process to attenuate non-synchronous components depends on the number of rotations used for averaging (Stander and Heyns, 2005). In some cases a sufficient number if rotations are not available or spurious components, which adversely affects the quality of the synchronous average, are present in a single measurement. A second synchronous averaging process of the discrepancy signal, between the synchronous averages of consecutive measurements, are proposed to alleviate the aforementioned problems. The second synchronous average at position l on the gear,

$$\mu_{\eta}^{(2)}[l] = \left(\sum_{i=1+N_a}^{N_m+N_a} N_{r,i} \right)^{-1} \sum_{k=1+N_a}^{N_m+N_a} \left(N_{r,k} \mu_{\eta,k}^{(1)}[l] \right) \tag{4.16}$$

is computed over N_m synchronous averages $\mu_{\eta,k}^{(1)}[l]$ from a set of N_m consecutive measurements in the form of a weighted average. The synchronous average of measurement k, denoted by $\mu_{\eta,k}^{(1)}[l]$, is performed over $N_{r,k}$ gear rotations and is calculated from equation (4.15), where the measurement number is excluded. The second synchronous average is computed from measurement $1 + N_a$ to measurement $N_m + N_a$ and after the synchronous average is computed N_a is increased so that the next averaging process can ensue. If N_a is increased by N_m then a 0% overlap between measurements occur, and if N_a remains the same, then a 100% overlap between measurements occur. If the period between measurements is small, the condition of the gearbox is the same for a finite amount of consecutive measurements and therefore equation (4.16) attenuates only non-diagnostic information.

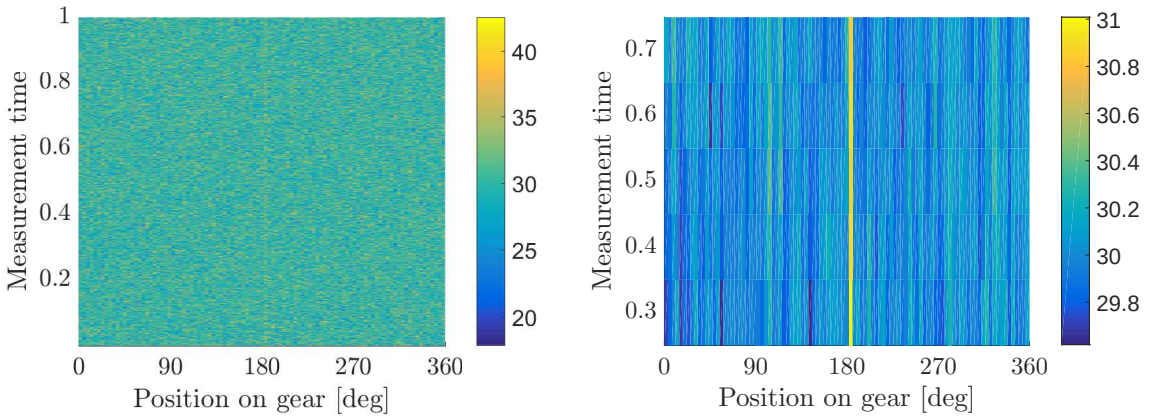
It is assumed that there is no phase difference between the N_m measurements when implementing equation (4.16). Non-synchronous averaging occurs in the second average, if the phase difference between measurements is non-zero. The phase difference between measurements cannot be obtained from a tachometer, since it is not available in this study. Instead the cross-correlation, denoted by the operator $\star$, between the

discrepancy signal from a reference measurement $\eta_{\text{ref}}[t]$ and measurement i, denoted by $\eta_i[t]$,

$$\eta_{\text{ref}} \star \eta_i[t] = \sum_{n=0}^{N_T-t-1} \eta_{\text{ref}}[t+n]\eta_i[n] \tag{4.17}$$

is maximised so that the time delay or phase difference between $\eta_{\text{ref}}[t]$ and $\eta_i[t]$ can be estimated. The phase difference, calculated from the cross-correlation maximisation process, is used to align the new measurement to the reference measurement. This is performed so that the phase difference between the two aforementioned measurements are zero. The best results are obtained for minimising the phase between the N_m measurements, when using the last measurement as a reference signal. See Appendix B for more details on the most appropriate reference signal. The presence of two faults may lead to the presence of local maxima in the objective function in equation (4.17). However, the cross-correlation maximisation process is a global optimisation strategy with the results unaffected by the presence of local minima. Therefore the cross-correlation process is theoretically unaffected by the presence of more than one fault.

The advantages of using the second synchronous averaging process between measurements are illustrated in figure 4.11 on simulated data. A synchronous average of a discrepancy signal is simulated for a gear that has a fault at 180 degrees. The discrepancy associated with the damage is 31, while the healthy part has a discrepancy of 30. A lot of synthetic noise is added to synchronous average to make it difficult to observe the damage on the gear. A 1,000 unique vibration signal measurements are available of this gear and therefore 1,000 synchronous averages are available. The 1,000 synchronous averages, of the discrepancy signal in figure 4.11a, are contained in the rows and it is presented over one full gear rotation. A feint gear fault is observed at 180 degrees, however it is immersed in a lot of noise which makes it difficult to infer the condition of the gear. A second synchronous averaging process over 500 measurements is completed with an overlap of 400 measurements between each average (i.e. N_a increases by 100 between averaging processes). The result of this second synchronous averaging process in figure 4.11b uses the mean measurement time of the N_m measurements in figure 4.11a with the simulated gear fault clearly observed at 180 degrees because the noise was averaged out and only the diagnostic information is retained in the signal. Note that this is merely an example to show that the second synchronous average is more robust and it makes it easier to observe the damage if the synchronous average contains spurious noise. The second synchronous average process is tested on experimental data in section 5.2.



(a) (b)

Figure 4.11: The advantages of using the second averaging process for incipient fault detection are illustrated on simulated data. The synchronous average of a simulated discrepancy signal, with a lot of noise, is presented in figure 4.11a for 1,000 measurements. The result of a second synchronous averaging process is presented in figure 4.11b with the noise content reduced and only the diagnostic information retained. Normalised time is used on the ordinate to show the reduction in sampling frequency in the measurement time.

4.3.1.3 Gear-pinion discrepancy distribution

The discrepancy signal, η, is generated at the meshing between the gear and the pinion. If the gear and the pinion information is investigated simultaneously, it can provide invaluable insight into the condition of the gears in the gearbox. A gear-pinion discrepancy distribution $J[i, l]$,

$$J[i,l] = \frac{1}{N_r^g} \sum_{k=1}^{N_r^g} \eta^g[i + kN_s^g]\eta^p[l + kN_s^p] \tag{4.18}$$

is proposed to investigate the information from the gear and the pinion simultaneously. The discrepancy signal with respect to gear and pinion revolutions are denoted by η^g and η^p, respectively. The incremental position on the pinion and the gear is denoted by l and i, respectively. The gear completes fewer rotations than the pinion and therefore the number of averages N_r^g is set to the number of rotations completed by the gear. The number of samples per revolution for the gear and the pinion discrepancy signals are N_s^g and N_s^p, respectively. This procedure is similar to calculating the synchronous average of the discrepancy signal in equation (4.15), but now the information from the pinion and the gear is treated simultaneously. Since the pinion and the gear discrepancy signals contain the same fault information and because they are multiplied in equation (4.18), the gear-pinion discrepancy distribution contains a large bias if discrepancies

are present. Therefore it remains critical to use the second average of the discrepancy signal and compare it to an alarm threshold to infer the correct condition of the gear.

4.3.2 Fault detection

This second synchronous average needs to be monitored by an expert or by automatically comparing it to an alarm threshold. The latter is used in this study, since it is more robust and cost-effective. The threshold of the alarm is obtained from a confidence bound generated from the healthy gearbox data, since it is based on sound statistical theory. The mean and the standard deviation of the second synchronous average are estimated from a maximum likelihood approach and therefore the Student-t distribution is appropriate to generate the confidence bound (Montgomery and Runger, 2011). An upper limit for the one-sided confidence bound for the second average $\mu_\eta^{\mathrm{CB}}[i]$,

$$\mu_\eta^{\mathrm{CB}}[i] = \hat{\mu}_\eta[i] + \mathcal{T}_{\alpha,N_h-1}\frac{\hat{\sigma}_\eta[i]}{\sqrt{N_h}} \tag{4.19}$$

is obtained from the estimated mean $\hat{\mu}_\eta[i]$ and the standard deviation $\hat{\sigma}_\eta[i]$ of the healthy dataset so that an alarm threshold can be set. The Student-t indicator $\mathcal{T}_{\alpha,N_h-1}$ is obtained from N_h healthy data sets with a $100 \times (1-\alpha)$ % confidence interval (Montgomery and Runger, 2011). Hence the second synchronous average from a set of N_m measurements is compared to the threshold, set by equation (4.20), to infer the condition as follows,

$$\text{Condition} = \begin{cases} \mu_\eta^{\mathrm{CB}}[i] - \mu_\eta^{(2)}[i] \geq 0 & \text{Expected behaviour} \\ \mu_\eta^{\mathrm{CB}}[i] - \mu_\eta^{(2)}[i] < 0 & \text{Novelty is observed at [i]} \end{cases} \tag{4.20}$$

where $\mu_\eta^{\mathrm{CB}}[i] - \mu_\eta^{(2)}[i]$ represents the degree of novelty and is calculated from equation (4.16) and equation (4.19). The i in equation (4.20) represents a value of: $1 \leq i \leq N_s$.

4.3.3 Fault trending

The second average of the discrepancy signal, computed with equation (4.16), can be represented over measurement time by using the mean measurement time over the N_m measurements as illustrated in figure 4.11. The discrepancy versus gear angle over time can be used to trend a fault in the gear as it progresses. If only a localised fault is present on the gear, it makes sense to plot the discrepancy value over time of the damaged tooth. This is performed automatically with a healthy-damaged decomposition.

It is assumed that the gear contains a healthy and a damaged portion and that this is reflected in a bimodal distribution in the second synchronous average of the NLL. The

two modes of the probability density can easily be estimated by using methods such as k-means, GMM etc. The k-means algorithm is utilised for this since it is simple, cost-effective and robust for this task. The healthy-damaged decomposition is implemented as follows:

1. Initialise a k-means algorithm with two clusters, since it is assumed that the gear contains a healthy and a damaged portion.

2. Find the two clusters within the data and classify the points according to those clusters.

3. Calculate the mean and the standard deviation of the two clusters.

4. The cluster with the larger and the smaller mean is labelled as damaged and healthy, respectively. The two clusters will be close to one another for a healthy gear.

Note that this can easily be modified to accommodate an arbitrarily amount of damage modes if the sampling frequency of the second synchronous average is sufficiently high and the number of damage modes is known.

4.4 Conclusion

The proposed fault diagnostic methodology requires a vibration measurement from a transducer connected to the gearbox of interest. The measurement from this transducer contains diagnostic (machine condition) and operating condition information which is extracted and modelled. A discrepancy signal is generated by combining the operating condition and machine condition model information which is subsequently processed by using a second synchronous averaging. The second synchronous averaged discrepancy signal is compared to an alarm threshold set by a confidence bound generated from healthy data to infer the condition of the machine. Fault trending can be performed from the second synchronous average or from the healthy-damaged decomposition. This procedure is tested in chapter 5 on the data in chapter 2.

Chapter 5 Fault diagnostic methodology: Results

The proposed diagnostic methodology in chapter 4 is validated on the numerical and experimental gearbox data described in chapter 2. The numerical gearbox data are investigated in section 5.1 and the experimental data are investigated in section 5.2. The investigation on the numerical gearbox data is focused on the performance of the machine condition features and corresponding machine condition models in their ability to locate and trend faults with the first synchronous average. The investigation on the experimental data focuses on the practical aspects of the diagnostic methodology, such as automatic and robust fault detection, localisation and trending.

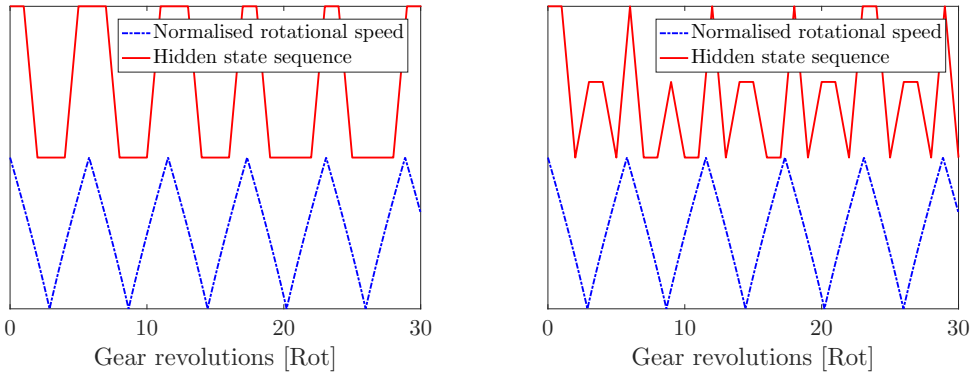
5.1 Numerical gearbox data

The numerical gearbox model, presented in section 2.1, is simulated in a healthy and in damaged conditions with the operating conditions presented in figure 2.4a. The acceleration of the fifth degree of freedom, located in the horizontal direction on the second mass block in figure 2.4b, is used as the measured signal in this section. The operating conditions result in severe amplitude modulation in the vibration signal, which tests the ability of the fault diagnostic methodology to handle complex fluctuating operating conditions. This temporal vibration signal is converted to the angle domain using the proposed tacholess order tracking method described in chapter 3.

The operating condition feature extraction and scaling is performed as described in section 4.1.2. The dimension of the feature space is reduced using PCA, as outlined in section 4.1.4, with the ACR threshold set to 80%. This results in an operating condition feature space of six dimensions, with the features modelled with a Gaussian observation distribution HMM. The more operating condition states (OCS) or hidden states are used in the operating condition model, the more machine condition models need to be optimised by using less training data, which can result in generalisation

problems etc. Some of the OCS may be rarely visited when using a complex HMM, which can result in the corresponding machine condition models to have insufficient training data as well.

Only HMMs with two and three hidden states are investigated to model the operating condition features for the numerical data in this study. The OCS sequence (or hidden state sequence) results are superimposed with the normalised rotational speed over gear revolutions in figure 5.1.



(a) (b)

Figure 5.1: The hidden state sequence with respect to normalised rotational speed for the numerical gearbox model over gear revolutions. The hidden state (or OCS) sequence of the HMM with the two and three hidden states are presented in figure 5.1a and 5.1b, respectively.

The hidden state (or OCS) sequence for the two state HMM in figure 5.1a is able to capture the trend in the operating conditions well. The two hidden states, in figure 5.1a, correspond to the maximum and minimum operating conditions (i.e. rotational speed and load). The hidden state sequence of the three state HMM, in figure 5.1b, provides a finer resolution to the operating conditions, since a hidden state associated with the transition region is added. However, this new hidden state (indicated as the first hidden state in figure 5.1b) is rarely visited due to its features' similarity with the features associated with the other hidden states. This results in insufficient data for optimising the machine condition model associated with the new hidden state. Therefore a two state operating condition HMM is used in this study.

The machine condition feature extraction process, explained in section 4.1.3, is preformed on the order tracked vibration signal with the CWT and WPT treated separately. The ACR for the CWT in figure 5.2a and the ACR of the WPT in figure 5.2b indicate that the dimensionality of the feature space is significantly lower than the apparent dimensionality of the original feature space. Only the first 20 principal

components are included for the CWT, since the ACR of the other principal components is close to unity. Both the WPT and CWT feature spaces are reduced to a six dimensional feature space with PCA, which results in little apparent information loss.

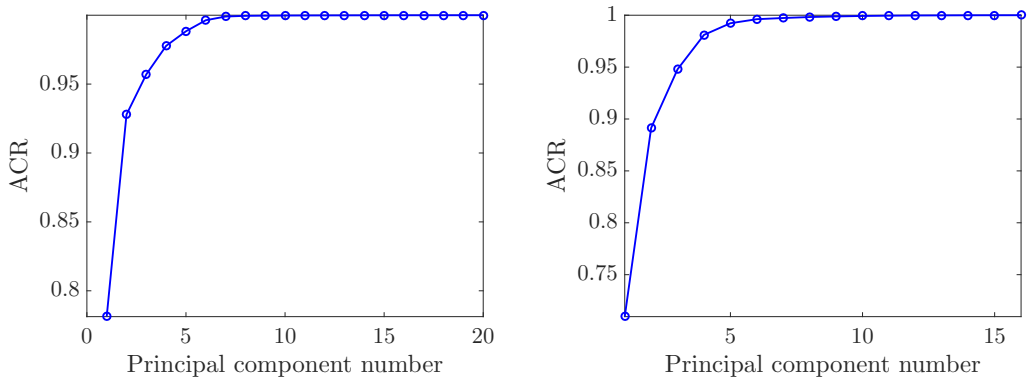


Figure 5.2: The accumulative contribution rate (ACR) of the principal components of the CWT and WPT machine condition features in figure 5.2a and 5.2b, respectively.

The machine condition features are labelled with the same operating condition states as the corresponding operating condition features as explained in section 4.2.2 and illustrated with figure 4.8. The labelled machine condition features are used to train two HMM machine condition models, with three hidden states in each model, with the process described in section 4.2.2. A discrepancy signal is generated from the procedure described in section 4.2.3 if new data are evaluated. The generated discrepancy signal, in the form of a negative log-likelihood signal (NLL), requires further processing before the condition of the machine can be inferred.

The synchronous average of the discrepancy signal from the numerical gearbox data is displayed in figure 5.3 for various gear tooth fault severities. The machine condition models are trained on the machine condition features extracted from the CWT. The localised faults are induced in the gearbox of the model by the procedure discussed in section 2.2 with the result shown in figure 2.2. A 0.05 fault severity indicates that the gear mesh stiffness of the damaged tooth is $(1-0.05)\times k_{gm}(t)$, where k_{gm} is the stiffness of a healthy tooth. The damaged tooth is centred at 180 degrees in all of the figures to make the comparison easier. A fault severity of 0.05 and larger is easily detected in the synchronous average of the discrepancy signal from the CWT machine condition features. The discrepancy generated by the damaged tooth looks very similar to the envelope of an impulse in figure 5.3.

The same investigation is performed with the machine condition features extracted from the WPT. The synchronous averages of the discrepancy signals generated from the numerical gearbox in different machine conditions are presented in figure 5.4.

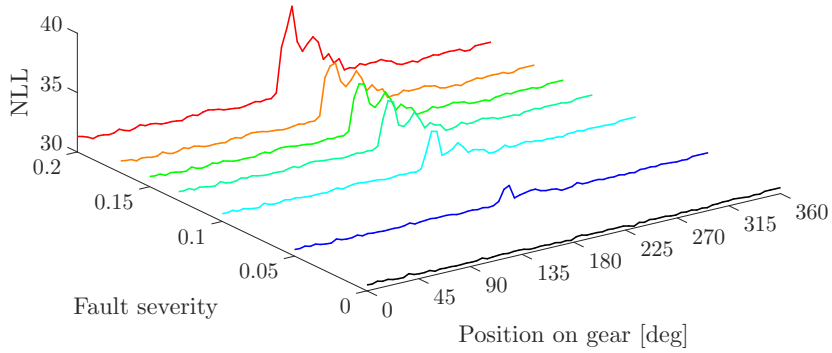


Figure 5.3: Synchronous average of the discrepancy signal for different fault severities using CWT machine condition features.

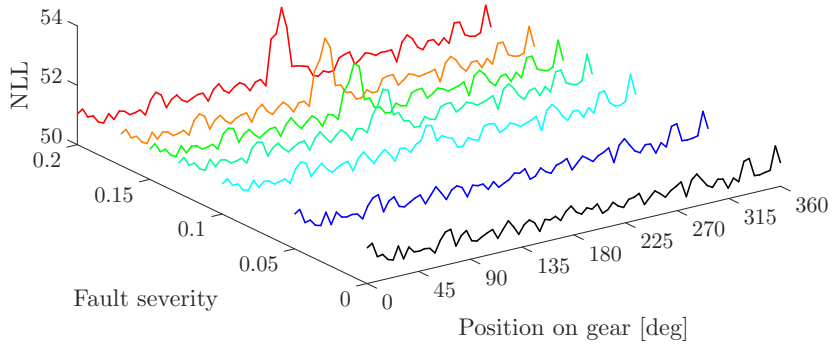
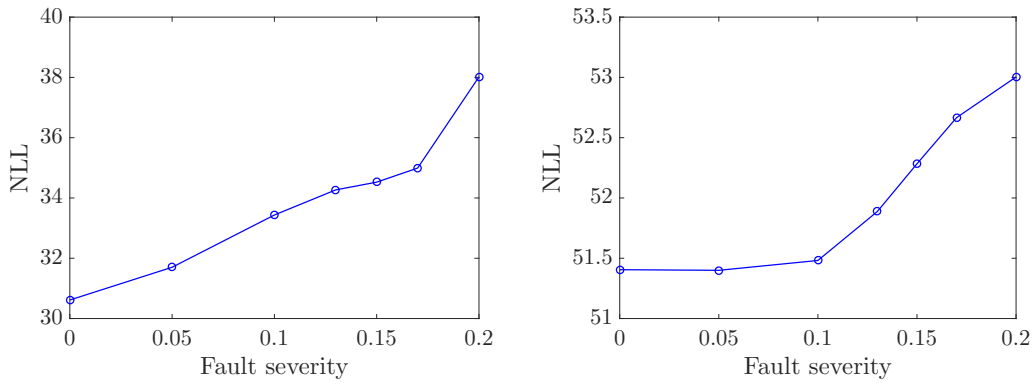


Figure 5.4: Synchronous average of the discrepancy signal for different fault severities using the WPT machine condition features.

A fault with a severity of 0.1 is observed in the synchronous average of the discrepancy signal in figure 5.4. However, the synchronous average of the data generated from a gearbox with a fault severity of 0.05, looks very similar to a healthy gearbox. The inability of the machine condition models, trained on WPT features, to detect a fault with a severity of 0.05 is attributed to the difference between the CWT and the WPT feature extraction approaches. A naive approach is used with the WPT machine condition features, since it is assumed that the nature of the damage is unknown. In the machine learning field, features are extracted that convey the information of interest (machine condition information), since unimportant information adversely affects the performance of the machine learning models. Hence the WPT features are able to detect localised damage from all of the available information at the cost of a decreased sensitivity to incipient faults. It is envisaged that the WPT features are appropriate for cases where the nature of the damage and the frequency range, in which the damage manifests within the spectrum, are unknown. This reduces the need for experts to

identify the frequency range that needs to be investigated to diagnose machine faults.

The NLL associated with the damaged portion of the gear in figure 5.3 and 5.4 is trended against fault severity in figure 5.5. The damaged portion is extracted from the healthy-damaged decomposition of the discrepancy signal.

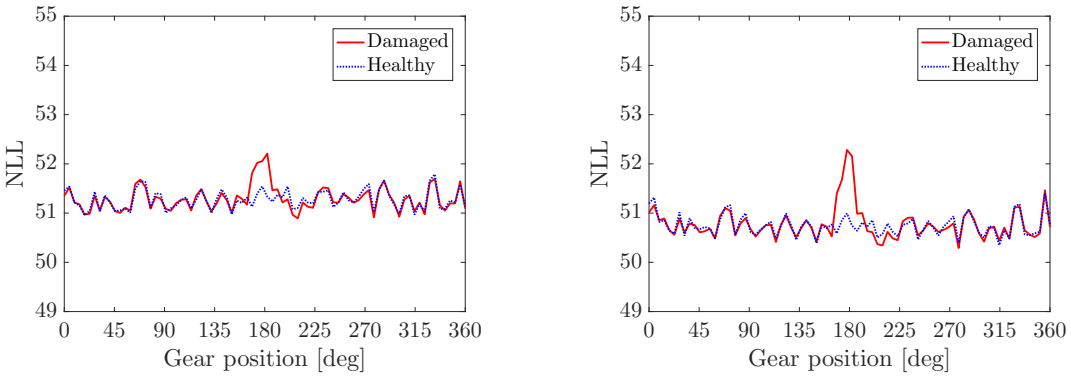


(a) CWT features (b) WPT features

Figure 5.5: The NLL, associated with the damaged portion of the gear, is trended against fault severity. The performance of the CWT and the WPT machine condition features are compared in the figure.

A change in fault severity is detected by both sets of machine condition features in figure 5.5, however, the CWT machine condition features are more sensitive than the WPT machine condition features to incipient faults.

Lastly, two different fault diagnostics methodologies are compared on the numerical gearbox data in figure 5.6. The first case, with results in figure 5.6a, does not incor-



(a) Neglecting OC information (b) Incorporating OC information

Figure 5.6: The result of incorporating operating condition information into the fault diagnostic methodology, with WPT machine condition features being used.

porate operating condition information into the process. Hence, only a single machine

condition model is trained with the WPT machine condition features. The second case, in figure 5.6b, follows the proposed diagnostics methodology. An evident improvement is observed in the ability of the synchronous average to detect faults if operating condition information is incorporated into the process.

The proposed methodology performs well on the data generated from the numerical gearbox model. The synchronous average, obtained from the proposed methodology, is sensitive to changes in machine condition caused by fault growth. The proposed discrepancy signal post-processing techniques in section 4.3 are not required for the numerical gearbox data and are only investigated on the experimental data as a result.

5.2 Experimental gearbox data

In this section, the proposed diagnostic methodology is evaluated with the experimental data introduced in chapter 2. The feature extraction and the model optimisation aspects of the diagnostic methodology are treated first in this section, whereafter the performance of the diagnostic methodology is investigated. The proposed tacholess order tracking procedure, developed in chapter 3, is used in this section. The operating and machine condition features are extracted from the order tracked vibration signal measured from the axial component of the tri-axial accelerometer.

5.2.1 Feature extraction and model optimisation

5.2.1.1 Operation condition feature extraction, processing and modelling

The two sets of operating condition features, discussed in section 4.1.2, result in a poorly scaled 18 dimensional feature space because the magnitude of the spectrogram features is larger than the rotational speed of the gear shaft. The principal component dimensionality reduction process would be dominated by the spectrogram features and therefore each operating condition feature is scaled independently as discussed in section 4.1.2. PCA analysis is performed on the scaled feature space to remove correlated (uninformative) features. The ACR of the operating condition feature space in figure 5.7 is calculated from equation (4.2).

A five dimensional operating condition feature space is obtained after setting the ACR threshold to 80%. The linearly scaled operating condition features are transformed to the principal component space with equation (4.1). The first two principal components of the operating condition features in figure 5.8 indicate that there is a strong sequential

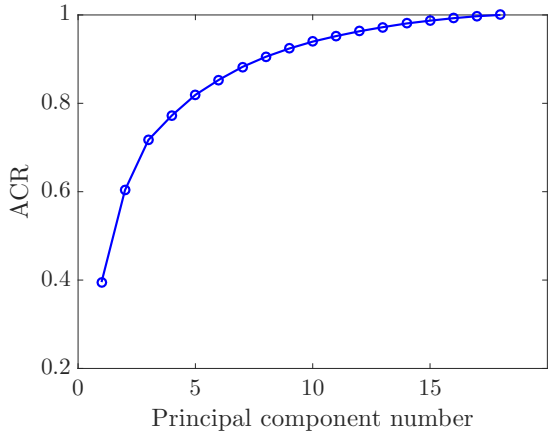
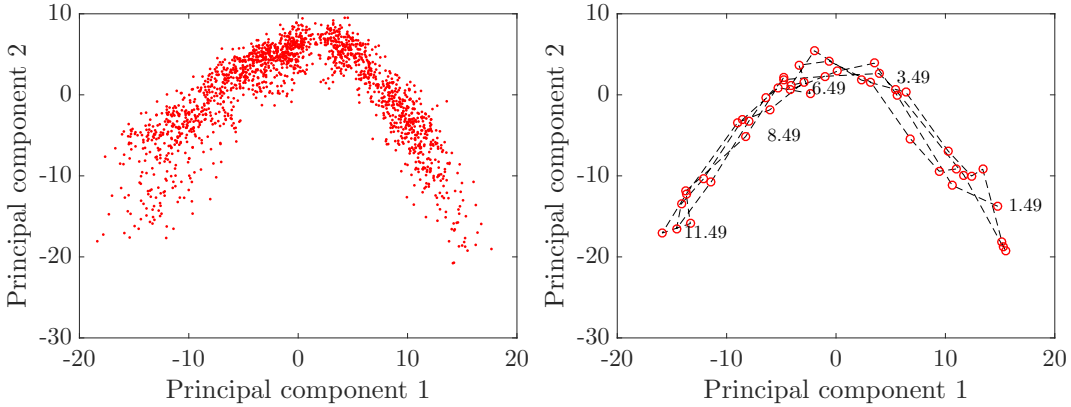


Figure 5.7: The accumulative contribution rate (ACR) of the linearly scaled operating condition features.

pattern along the manifold of the features. This makes a HMM well suited for modelling the operating condition features.



(a) All training OC features (b) First 45 OC features

Figure 5.8: The principal component space of the operating condition (OC) features of the undamaged gearbox with the gear revolutions (in rotations) also indicated in figure 5.8b.

The model complexity of the operating condition HMM describing the experimental data is chosen similarly to the numerical data in section 5.1. The three hidden state operating condition HMM provides the ideal compromise between model simplicity and good results and it is used further in this section. The rotational speed of the gearbox shaft is superimposed on the hidden state sequence from the aforementioned operating condition HMM in figure 5.9. It is concluded from the gear revolutions in figure 5.8b and the hidden state sequence in figure 5.9 that the operating condition features are sensitive to operating condition changes.

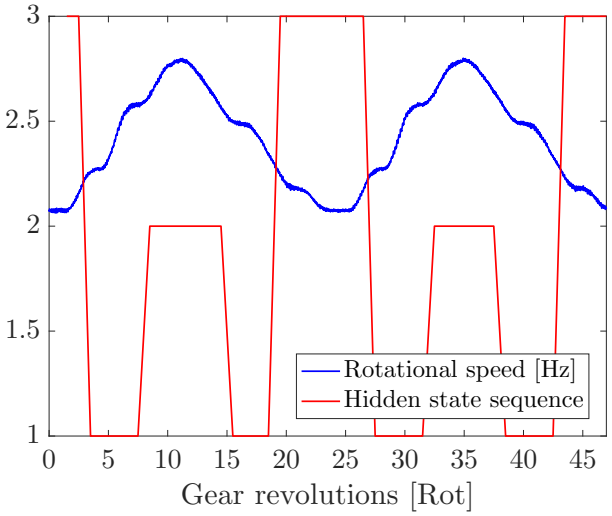


Figure 5.9: The OCS or hidden state sequence of the operating condition model with the corresponding rotational speed over gear revolutions.

5.2.1.2 Machine condition feature extraction, processing and modelling

The machine condition feature processing results are only illustrated for the CWT machine condition features in this chapter. The results for the machine condition models, optimised on the WPT machine condition features, are included at the end of this chapter. The ACR of the original CWT machine condition feature space in figure 5.10 indicates that the intrinsic dimensionality of the CWT features is significantly lower than the apparent dimensionality of 400. The first six principal components

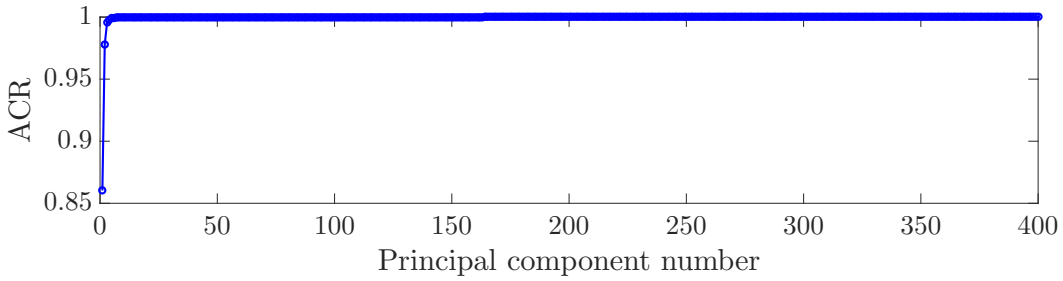
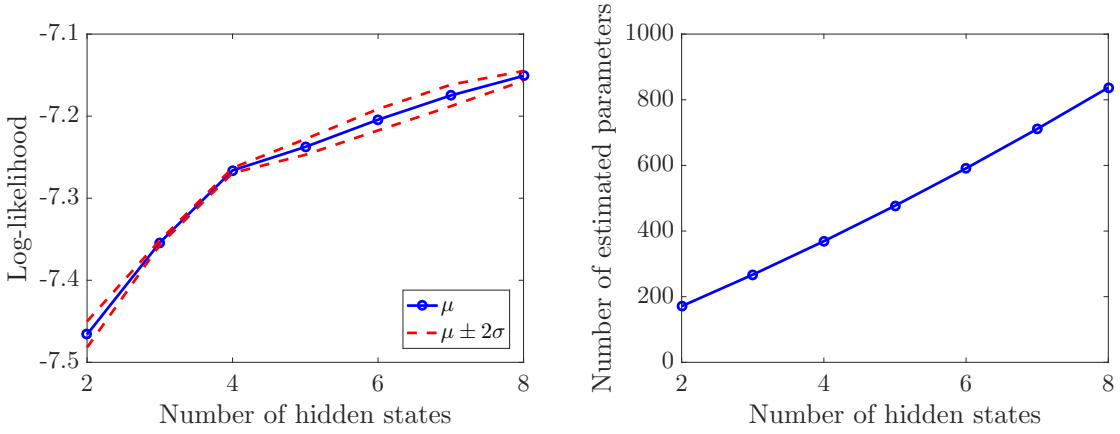


Figure 5.10: The ACR calculated from the CWT machine condition features. The first six principal components have an ACR of 0.999491.

describe the original 400 dimensional feature space well with negligible information loss. The continuous wavelet coefficients inherently contain redundant information as opposed to the orthogonal coefficients from the DWT and the WPT. The scales around the first five gear mesh frequencies of the monitored gearbox have very similar characteristics, which contribute to the large amount of redundant information as well.

The model complexity of the machine condition model HMMs needs to be determined

before the six dimensional machine condition feature space can be modelled and the discrepancy signal generation process can ensue. In this study, it is assumed that each machine condition model has the same model complexity (the number of parameters, the number of hidden states etc.). In equation (4.14) it is indicated that the BIC is minimised if the model complexity is low and the performance of the HMM (i.e. the likelihood) is high. The BIC is not computed directly in this section, since it is approximately inversely proportional to the model evidence under a set of assumptions, where the assumptions may be incorrect for the investigated models. The log-likelihood and the model complexity, contained within the two terms in the BIC, are computed and compared in figure 5.11 instead. The results in figure 5.11a are obtained from a



(a) (b)

Figure 5.11: The log-likelihood and the number of parameters that need to be estimated when optimising the three machine condition HMMs are presented in figure 5.11a and 5.11b, respectively.

statistical analysis over 20 runs with the same number of parameters. This ensures that the starting point of the expectation-maximisation algorithm does not adversely affect the validity of the results in figure 5.11a. A larger number of hidden states could not be investigated for the results in figure 5.11, since the covariance matrix became singular.

The log-likelihood keeps increasing as the model complexity increases in figure 5.11a and the number of parameters, that need to be estimated, grow monotonically with an increase in model complexity in figure 5.11b. The change in the log-likelihood gradient, that occurs at four hidden states in figure 5.11a, indicates the model's performance starts to saturate even if the number of parameters increases. Hence, a HMM with four hidden states is used to model the CWT machine condition features associated with an OCS.

5.2.2 Discrepancy signal and post-processing

A discrepancy measure is computed at each machine condition feature window and is used to generate the discrepancy signal in figure 5.12 over gear revolutions. The discrepancy is generated at the meshing point between the gear and the pinion's teeth and therefore the same discrepancy signal is obtained over the pinion revolutions as well. However, damage on the gear is synchronous with the gear revolutions and are non-synchronous with the pinion revolutions.

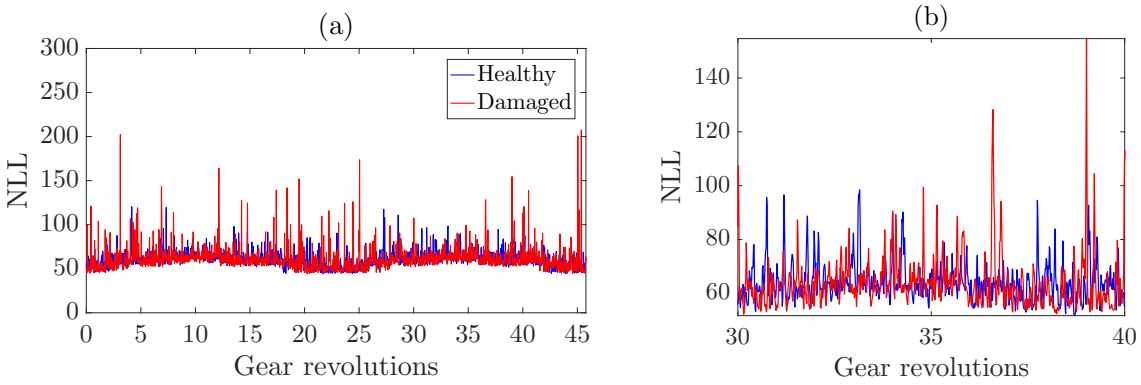


Figure 5.12: An example of the discrepancy signal over gear revolutions for a gearbox with a healthy and a damaged gear is shown in (a) with a zoomed view shown in (b).

The discrepancy signal for the gearbox with the damaged gear in figure 5.12 contains larger discrepancies than for the healthy gearbox, but it is difficult to ascertain the source of the larger discrepancies. The aforementioned larger discrepancies can be generated by the presence of localised faults or by other sources, which emphasises that the discrepancy signal requires further processing.

5.2.2.1 Synchronous averaging results and the alarm threshold

The synchronous averaging process of the discrepancy signal, with respect to the damaged gear, retains the synchronous components (i.e. localised faults on the gear) and attenuates the non-synchronous components (i.e. possible localised faults on the pinion, impulses generated from the bearing etc.).

The synchronous average from the validation set (i.e. data from a healthy gearbox not used during model optimisation) is computed and is subsequently compared to an alarm threshold, generated from equation (4.19) with healthy data, in figure 5.13. The synchronous average of the discrepancy signal associated with the validation data and the alarm threshold is denoted by μ_{v} and CB_{v}, respectively. Approximately 200 rotations (see N_h in equation (4.19)) are used to generate the confidence bound in

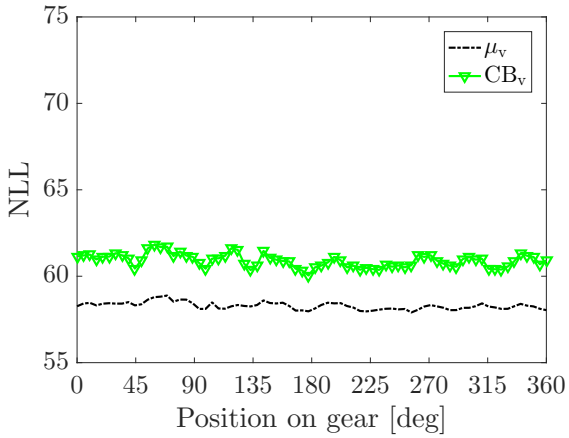
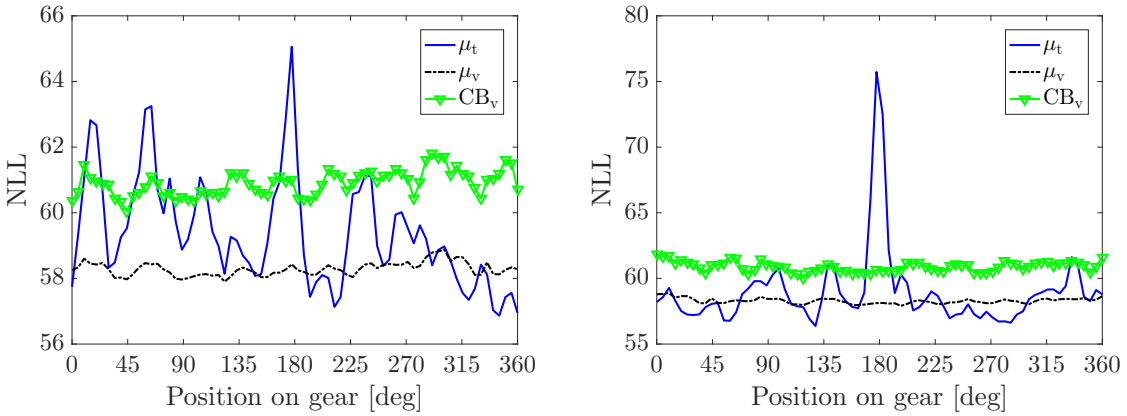


Figure 5.13: The confidence bound, denoted by CB_v, is used as the alarm threshold and is compared to the first synchronous average of the discrepancy signal from the validation data, denoted by μ_v.

figure 5.13 with α in equation (4.19) selected as 10^{-4}. Equation (4.20) indicates that the gear associated with the validation data in figure 5.13 is in a healthy condition. This validates that the selected confidence bound is appropriate and that the methodology works on the validation data.

The vibration data, from the damaged gear presented in figure 2.7, is investigated on the proposed fault diagnostic methodology. The first synchronous average of the discrepancy signal from the gearbox with a damaged gear, denoted by μ_t, is compared to the confidence bound, denoted by CB_v, and the synchronous average of the validation data, denoted by μ_v in figure 5.14. The synchronous average of the first measurement



(a) (b)

Figure 5.14: The discrepancy signal obtained from a single averaging process for the first measurement in figure 5.14a and for a measurement one week later in 5.14b. The position of the seeded slot is set to 180 degrees in all of the figures. Note the different y-axis scales. The synchronous average of the new measurement is denoted by μ_t.

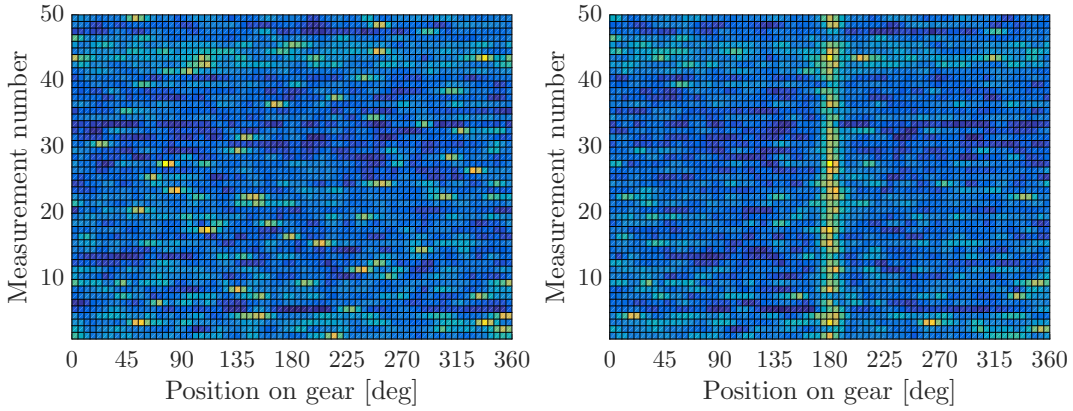
after the gearbox was reassembled, in figure 5.14a, exceeds the alarm on the position of the damage, but with five false alarms present as well. This synchronous average does not reflect the actual condition of the machine, since only one tooth is damaged. The synchronous average in figure 5.14b is from data measured at a later stage, with more prominent damage on the gear tooth. Spurious discrepancies, not related to the condition of the gear, exceed the confidence bound slightly in figure 5.14b which results in false alarms as well. The true condition of the gear is misrepresented in figure 5.14a and figure 5.14b. The spurious components that lead to the false alarms are attributed to the disassembling and reassembling process between the healthy and the damage datasets and the bearing discussed in chapter 2. The spurious components become less prominent over time in the synchronous average, however false alarms remain probable. The synchronous average requires further processing before it can be used as a diagnostic indicator.

5.2.2.2 The second synchronous average results

A second synchronous averaging process is proposed, over the discrepancy signals from consecutive measures, to reduce the influence of the non-diagnostic characteristics on the diagnostic indicator. The set of measurements which are averaged to obtain the second synchronous average must contain the same machine condition information (i.e. same gear condition, same position of the damage etc.) to ensure that the deterministic components are not averaged out. Hence, the phase difference between consecutive measurements has to be zero.

The phase difference between consecutive measurements is estimated by maximising the cross-correlation between them, as described in section 4.3. The synchronous average of the discrepancy signal from 50 consecutive measurements in figure 5.15a is aligned in figure 5.15b by minimising the phase difference between them. The cross-correlation maximisation process results in figure 5.15b contain errors, which results in an increased noise floor and standard deviation when performing the second synchronous average over the rows in figure 5.15b.

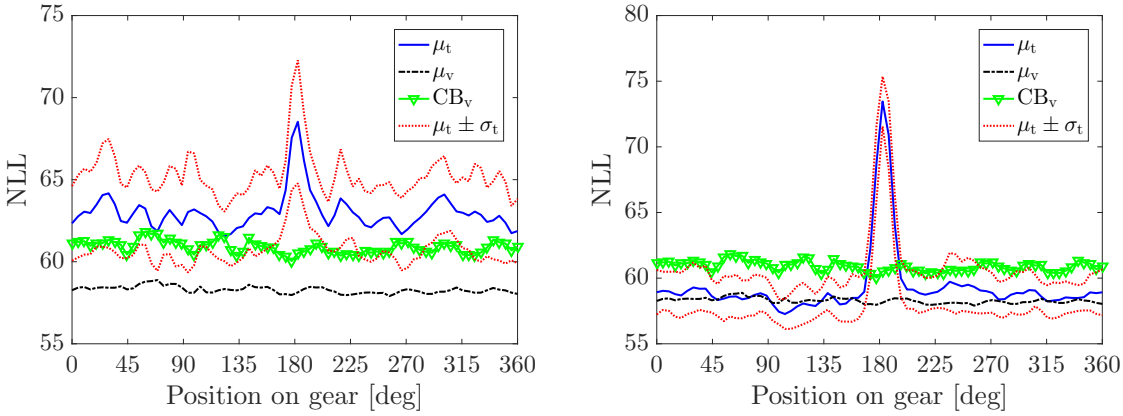
If the number of measurements used in the second averaging process is large, then the condition may change significantly between the first and the last measurement. This results in the discrepancy measure representing the average condition within the set of measurements, which results in a large standard deviation (i.e. large uncertainty in the true condition) at the location of the fault. The second average is computed over 20 measurements with an 80% overlap in this study. This ensures that the condition of the machine remains approximately the same between the measurements used in the second synchronous averaging process.



(a) Before (b) After

Figure 5.15: The set of synchronous averages, that is used to compute the second average of the discrepancy signal, before and after the phase difference minimisation process is completed. Each row represents a synchronous average of the discrepancy signal generated for a specific measurement.

The second averaging process results in figure 5.16a are obtained with the first set of 20 measurements, with the results in figure 5.16b obtained at approximately the same time as the results in figure 5.14b. The second synchronous average and the standard deviation of the discrepancy signal of the damaged gearbox is denoted by μ_t and σ_t, respectively. The synchronous average of the validation data and the confidence bound is denoted by μ_v and CB_v, respectively. Smoother results are obtained in figure 5.16



(a) (b)

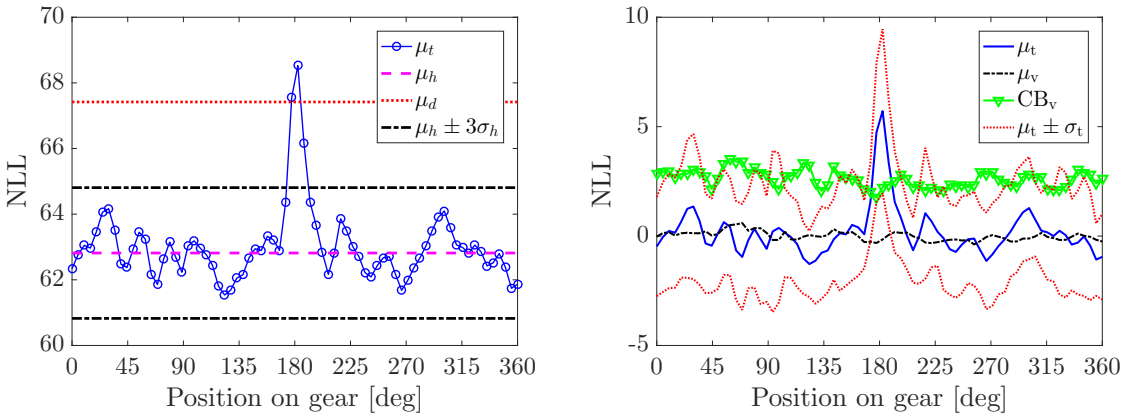
Figure 5.16: The results from the second averaging process on the discrepancy signal for the first set of measurements and the set of measurements taken approximately one week later are presented in figure 5.16a and 5.16b, respectively. The mean and the standard deviation of the discrepancy signal, from the second synchronous averaging process, are denoted as μ_t and σ_t, respectively.

than in figure 5.14. The raised noise floor or bias in figure 5.16a is attributed to the spurious discrepancies in the first synchronous average process as well as the errors in the cross-correlation maximisation process which are not properly averaged out. The bias decreases over time which is possibly due to the components in the gearbox which settle to its original positions and the cross-correlation process which estimates the phase difference better for well established damage. According to equation (4.20), the result in figure 5.16b has a novelty at 180 degrees. The small standard deviation in figure 5.16b provides confidence that the novelty at 180 degrees is from a deterministic phenomenon which is attributed to the presence of localised damage. However, the condition of the gearbox in figure 5.16b is misrepresented due to the high bias.

The increased bias in the second synchronous average, especially in the results of the initial set of measurements (see figure 5.16a), increases the risk for false alarms and misdiagnosing the condition of the gear. The following three remedies are proposed to avoid undesirable false alarms due to a high bias:

- Ignore the initial set of measurements and wait until the bias decreases to an acceptable level.
- Estimate and subtract the initial bias from the second synchronous average of the discrepancy signal. If the bias subtraction procedure is continuously used, then a method is required to determine whether distributed damage is present.
- Increase the alarm threshold temporarily or permanently. This avoids the risk of false alarms, but the presence of a developing fault may be detected too late if this alarm threshold is set too high.

A bias estimation procedure is used in this study, where the bias is automatically subtracted from the second synchronous average of the discrepancy signal as shown in figure 5.17a. The bias estimation procedure is very similar to the proposed fault trending procedure in section 4.3.3. The k-means clustering algorithm is used to determine the mean of second synchronous average of the discrepancy signal associated with the healthy and damaged portion of the gear. It is assumed that the mean associated with the healthy portion or cluster is the bias and can be subtracted from the second synchronous average. If a large overlap between the two clusters of the two portions exists, it indicates that the gear is healthy and then the mean of the entire discrepancy signal is subtracted instead. The mean from the validation data is subtracted from the confidence bound and is compared to the unbiased second averaged discrepancy signal when performing fault detection. The unbiased second synchronous average is compared to the confidence bound and the validation data, with zero mean, in figure 5.17b.



(a) Bias estimation

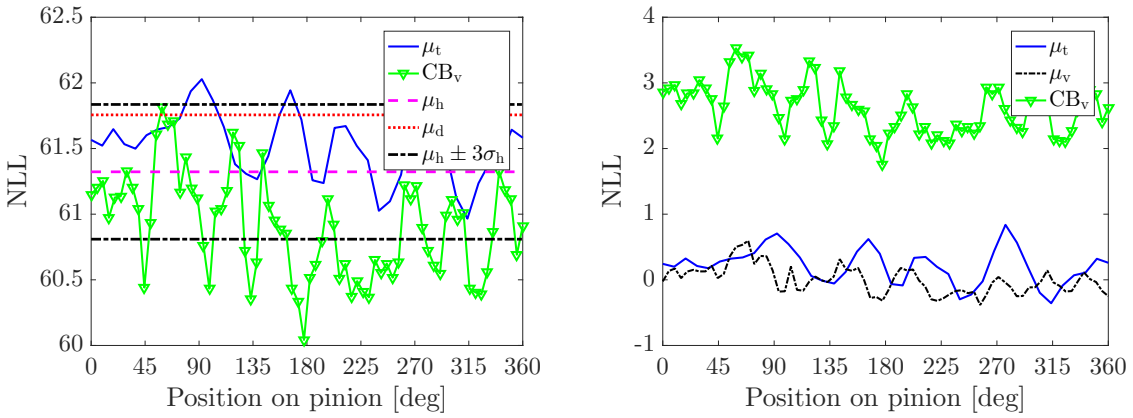
(b) Unbiased

Figure 5.17: The bias estimation process, displayed in figure 5.17a, is used to obtain the unbiased second synchronous average in figure 5.17b of the biased signal presented in figure 5.16a. In figure 5.17a, μ_{d} and μ_{h} denotes the mean of the damaged and healthy portions estimated from the k-means algorithm.

The mean of the damaged portion of the gear, μ_{d}, is outside the three standard deviation bound of the healthy portion of the gear in figure 5.17a. This indicates that the gear contains two machine conditions and the two cluster k-means algorithm is correctly applied. Therefore, μ_{h} is the true bias of the second synchronous average and is subsequently subtracted from the second synchronous average to obtain the result in figure 5.17b. The second synchronous averaging process results in a smoother discrepancy signal with a larger bias than the first synchronous averaging process. However, this bias is successfully estimated and subtracted to obtain a robust diagnostic indicator. Note that no false alarms occur in figure 5.17b as opposed to the result in figure 5.16a with the damage easily identified.

The second synchronous average of the pinion, in figure 5.18a, is computed with the gear in a damaged condition. The large discrepancies from the gear fault are attenuated in the synchronous averaging process over the pinion revolutions, which consequently results in the bias in figure 5.18a. The bias estimation procedure is performed on the pinion to ensure that its condition is correctly inferred.

The assumption that the pinion contains a healthy and a damaged portion failed in figure 5.18a, since the centre of the perceived damaged component, falls within the bounds of the healthy condition. This indicates that the whole signal is from a healthy source and therefore the mean of the entire second averaged signal (i.e. μ_{t}) is computed and subtracted to obtain the result in figure 5.18b. The desired result is obtained in figure 5.18b, since the pinion is healthy according to equation (4.20) with no false alarms.

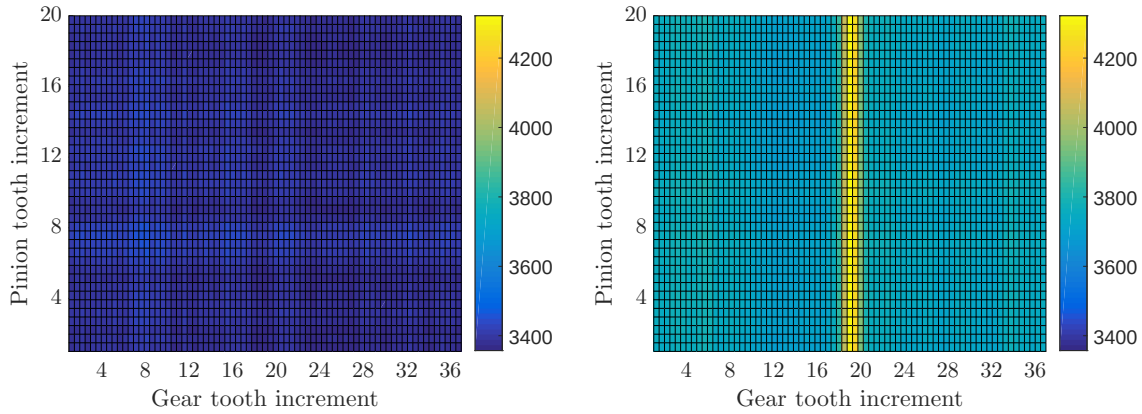


(a) (b)

Figure 5.18: The second synchronous average of the pinion, μ_t, contains a healthy and a damaged portion, denoted by μ_h and μ_d respectively. The bias is subtracted to obtain the unbiased second synchronous average in figure 5.18b.

5.2.2.3 Gear-pinion discrepancy distribution

Even though the unbiased second synchronous average is robust and can be used to infer the condition of the machine, it is beneficial to investigate the gear and the pinion information simultaneously. The gear-pinion discrepancy distribution, introduced in section 4.3, is used to represent the discrepancy information of the gear and the pinion simultaneously. The gear-pinion discrepancy distribution in figure 5.19 is obtained by evaluating equation (4.18) on the discrepancy signal generated from a gearbox with a healthy and a damaged gear. From figure 5.19b it is concluded that localised damage



(a) Healthy gear (b) Damaged gear

Figure 5.19: The gear-pinion discrepancy distribution of the gears of a gearbox in a healthy and damaged condition. The gear tooth increment emphasises that the number of teeth of the gear and the pinion is different.

is present on the gear, while the pinion is undamaged. If the healthy portion of the damaged gear in figure 5.19b is investigated and compared to the healthy gear in figure 5.19a, it can be observed that a large bias is present in the gear-pinion discrepancy distribution of the damaged gear. It is suspected that the gear-pinion distribution is very sensitive to changes in machine condition and it provides further insight into the condition of the gear.

5.2.2.4 Fault trending

The ability of the fault diagnostic methodology to trend localised faults in experimental conditions is vital for the proposed methodology to be successful. The damaged gear in figure 2.7 is overloaded with the operating conditions in figure 2.6 which resulted in the damaged tooth fail, as shown in figure 2.9, during experiments.

The second synchronous average of the discrepancy signal (NLL) over one gear rotation is presented over normalised time in figure 5.20. The change in the condition of the

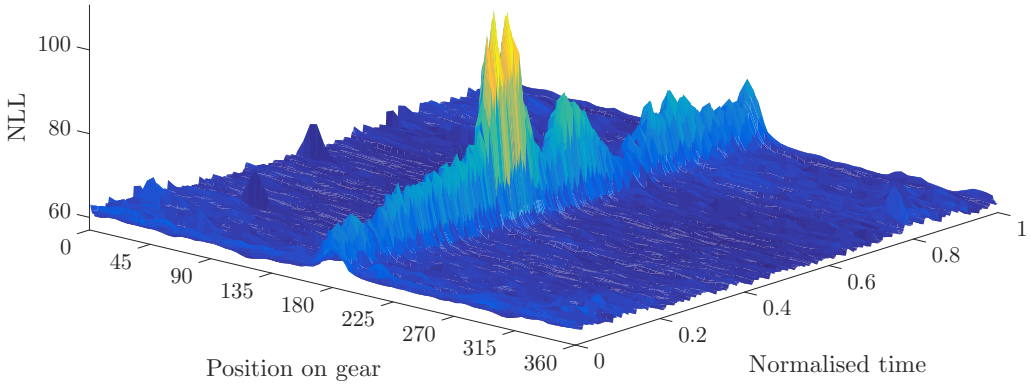


Figure 5.20: Second synchronous average of the discrepancy signal with respect to normalised time. Each average is performed over 20 measurements with an overlap of 80%.

machine, due to the deteriorating damaged tooth, is clearly observed in the second synchronous average of the discrepancy signal in figure 5.20. The sensitivity of the discrepancy signal to fault growth is further investigated by using the proposed healthy-damaged decomposition, presented in section 4.3.3.

The purpose of the healthy-damaged decomposition is to clearly visualise the progression of localised damage over time . This healthy-damaged decomposition is performed

on the first synchronous average and smoothed using a moving average window over 10 consecutive measurements to obtain the result in figure 5.21. Note that the smoothing process is essentially performing the second synchronous averaging process on the discrepancy signal, but it is only focused on two portions of the gear.

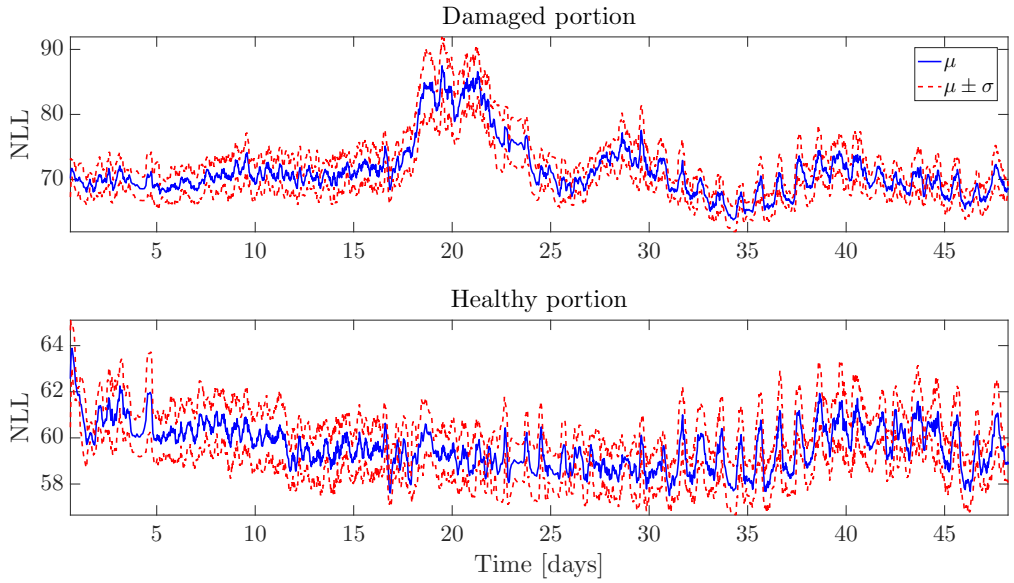


Figure 5.21: The discrepancy of the healthy and damaged portion of the damaged gear smoothed using 10 measurements and a 80% overlap between measurement windows. The smoothing process is similar to the proposed second synchronous average.

The damaged portion of the gear has an increasing trend in figure 5.21 from the start of the first experiment to the 19th day of the experiment. The discrepancy of the damaged portion decreases after this and another increase in the discrepancy is observed from the 26th day to the 28th day of the experiment. A similar phenomenon is observed at day 35 of the experiment. The damaged tooth of the gear broke off during the experiment, but the exact time of failure is unknown, because the gearbox has to be disassembled before the condition of the gear can be inspected. The disassembling process adversely affects the integrity of the data and is only performed if the gearbox has definitely failed. The presumed time of failure is the 19th day, since this is the time at which the discrepancy is a maximum in figure 5.20 and figure 5.21.

Helical gears have large contact ratios with respect to spur gears. Therefore the reduction in tooth stiffness, due to a damaged tooth, has a smaller effect on the total gear mesh stiffness for helical gears as opposed to spur gears. In the final stages of the damaged tooth's life, it does not contribute to the gear mesh stiffness, but it is present during meshing which increases the impulses within the vibration signal. The damaged tooth leads to additional impacts which disappear after failure. This can possibly explain the decrease after the presumed time of failure.

However, the phenomenon that occurs after the time of failure in figure 5.21 has not been explained. The discrepancy, associated with the damaged component, increases twice after the presumed time of failure and there is a periodic behaviour in the discrepancy signal of both components of the gear from the 30th day of the experiment. The periodic behaviour from the 30th day of the experiment has a maximum discrepancy in the afternoon which is proven in Appendix A. The daily temperature fluctuations are investigated to possibly explain the aforementioned phenomenon.

The minimum and maximum daily temperature, as well as the average daily temperature during the experiment, are superimposed on the log-likelihood (the negative of the discrepancy signal) in figure 5.22. Note that the log-likelihood data points in figure 5.22 are from the first synchronous averaging process, without any smoothing.

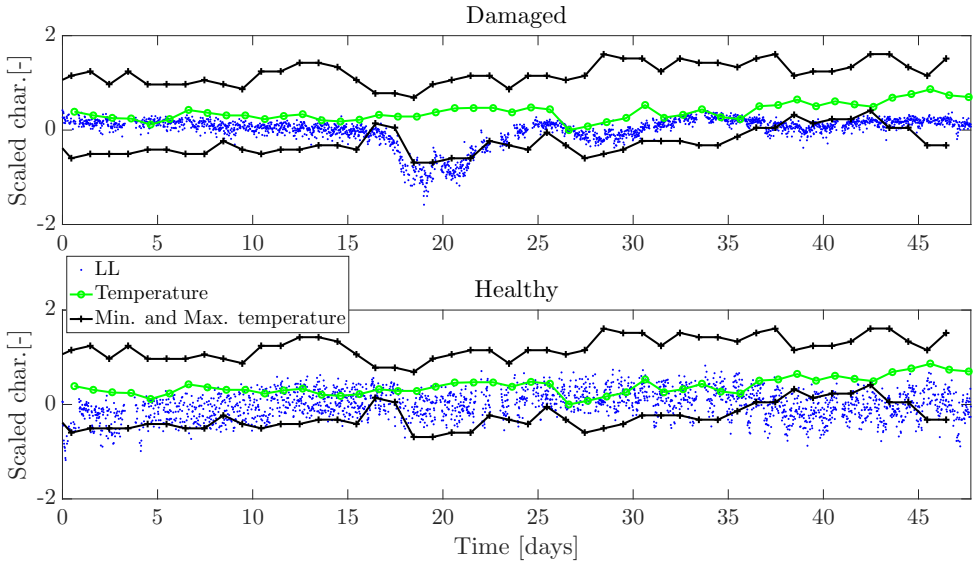


Figure 5.22: The scaled log-likelihood (LL) signal is superimposed with the scaled temperature characteristics over time. The maximum, minimum and the average temperature for each day are scaled consistently so that they are correct in proportion to one another.

There is a clear relationship between the temperature and the log-likelihood, especially in the damaged portion of the damaged gear. The minimum daily temperature dropped significantly at the 19th and the 26th day, while a sharp increase occurred on the 35th day to the 39th day of the experiment. The aforementioned temperature changes occur at the same time as the interesting behaviour observed in the discrepancy signal in figure 5.21. The viscosity of the oil is sensitive to temperature and affects the dynamic characteristics of the system (i.e. the damping characteristics). This can possibly explain the reason for the increase in discrepancy as well as the periodic behaviour that has a maximum in the warmest time of the day. However, it is unlikely that

the decrease in temperature at the 19th day is the only reason for the sharp increase in the discrepancy signal of the damaged portion. This supports the presumed time of failure at the 19th day. Even though the fluctuating temperature adversely affects the diagnostic indicator, it does not seem to dominate it, and therefore measuring temperature is not essential for diagnostic purposes.

The number of measurements that are averaged are increased from 10 (as in figure 5.21) to a 100 in figure 5.23. This allows the discrepancy of the damaged portion of the gear to be viewed over time until the presumed time of failure. The damaged portion is denoted by the subscript d and the healthy data is denoted by the subscript h. A significantly smoother increase is observed in the damaged portion as compared to figure 5.21. This indicates that by performing a second averaging process, the results are well suited for fault trending.

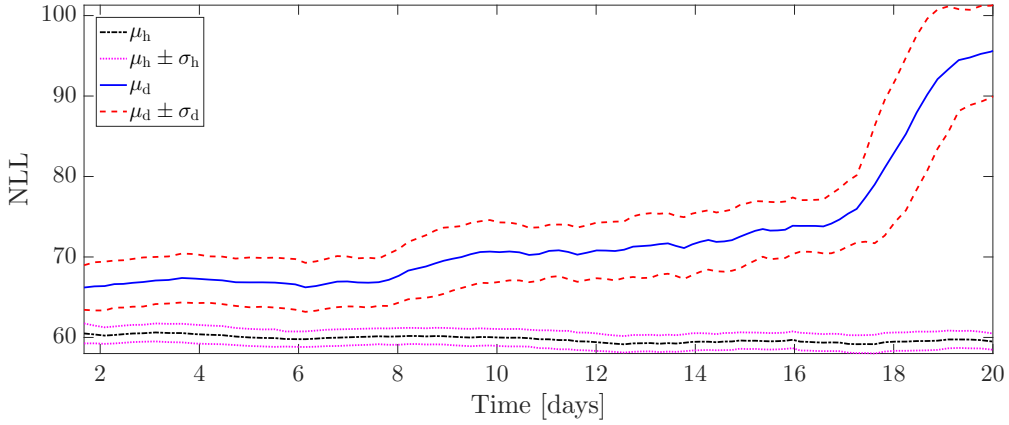


Figure 5.23: The discrepancy signal from the healthy, denoted by μ_h and damaged portion, denoted by μ_d, of the gear of the first experiment, smoothed using 100 measurements with an 80% overlap. The standard deviation is denoted by σ.

A second experiment was completed to prove that the presumed time of failure in the first experiment was the actual time of failure. More severe damage was seeded on the tooth of the gear, as discussed in chapter 2, which resulted in a shorter time to failure. The discrepancy signal is decomposed into the healthy and damaged portion again and is presented in figure 5.24 with the same characteristics as figure 5.23.

The experiment was stopped after the gradient of the diagnostic indicator started to decrease similar to the behaviour after the presumed time of failure of the first experiment. After opening the gearbox it was observed that the tooth had indeed failed. This smoothed discrepancy signal in figure 5.24 looks very similar to the results that are obtained in figure 5.23 for the first experiment. This gives confidence that the presumed time of failure in the first experiment is the actual time of failure and it also proves that this methodology is able to trend fault progression well.

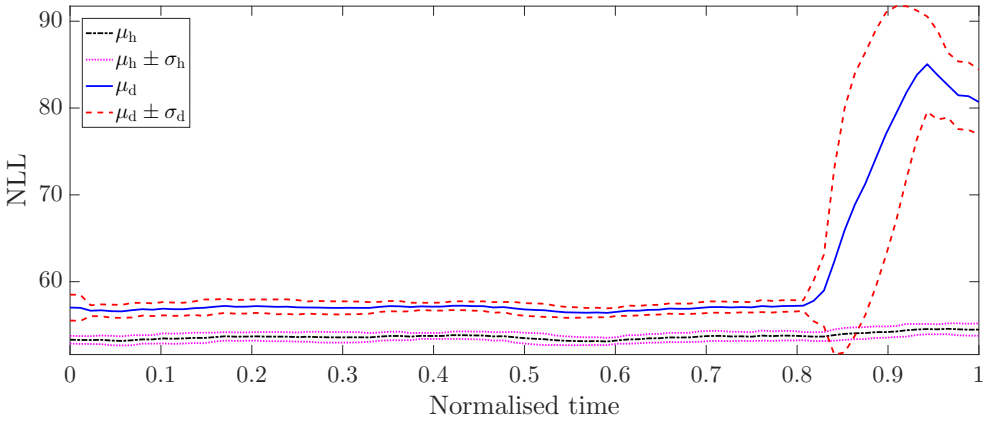


Figure 5.24: The second average of the discrepancy over normalised time from the data of the second experiment. A similar process and subscripts are used as in figure 5.23.

Lastly, the performance of the discrepancy signal generated from WPT machine condition features needs to be compared to the discrepancy signal from CWT machine condition features. Instead of showing all of the results from the WPT, only the final results of the two are compared. Note that the discrepancy signal is a relative measure and differs for different models and features. The discrepancy signals from the CWT and the WPT machine condition features are normalised and compared in figure 5.25. The same feature processing, modelling and discrepancy generation processes are used

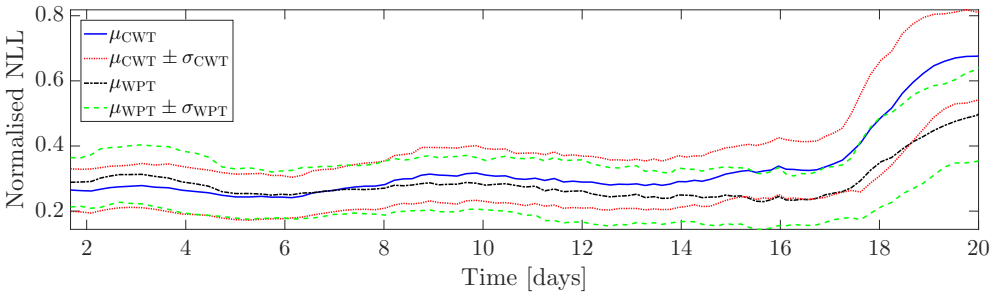


Figure 5.25: The comparison between the WPT-based and CWT-based discrepancy signal, associated with the damaged portion, obtained from the second averaging process. This is obtained over 100 measurements with an 80% overlap.

for the WPT and the CWT machine condition features. The healthy portion of the gear has very similar characteristics for the WPT and the CWT machine condition features and is therefore not presented in this study. Both machine condition approaches are able to detect and localise the fault on the gear, however only the discrepancy signal associated with the CWT machine condition features has a monotonic increasing trend. This indicates that the discrepancy signal associated with the WPT is unable to trend the fault as well as the CWT machine condition features.

5.3 Conclusion

The proposed fault diagnostic methodology in chapter 4 successfully detected, located and trended localised faults in the presence of fluctuating operating conditions with data only available from a single vibration transducer using the continuous wavelet machine condition features. The operating condition features and the operating condition model captures the operating conditions that are present within the data. The operating condition information is combined with the machine condition features with a probabilistic approach to generate a discrepancy signal which is subsequently processed so that the condition of the gear can be inferred. It is proven that gear faults can be detected using the confidence bound obtained from the healthy data. If this threshold is exceeded, the characteristic of the fault can be inferred by using the second average of the discrepancy signal and by using the gear-pinion discrepancy distribution. A bias estimation procedure is proposed which is used to obtain an unbiased second synchronous average which is robust and can be used for fault detection, localisation and trending. A healthy-damaged decomposition was performed to extract the mean associated with the damaged part as well as the healthy part and is used to trend damage until the time of failure. The healthy-damaged portion performs well for single faults, however multiple faults are not investigated in this study.

Chapter 6 Conclusion and recommendations

6.1 Conclusion

In this dissertation a diagnostic methodology was proposed that is able to perform fault detection, location as well as trending in fluctuating operating conditions. A robust tacholess order tracking method was proposed and used in this study to ensure that the proposed diagnostic methodology is practical and cost-effective to implement. In the proposed tacholess order tracking method, higher order frequency information is incorporated into the maxima tracking process by using probabilistic approaches. The angular-displacement Vold-Kalman filter and the Hilbert transform are subsequently utilised to obtain the instantaneous phase of the shaft of interest. The results, from a brief investigation, indicated that the angular-displacement Vold-Kalman filter is slightly more robust than the angular-velocity Vold-Kalman filter and is therefore used in this study. The proposed method was investigated on seven datasets in this study. The proposed maxima tracking method is robust in the presence of background noise and in applications with large accelerations as compared to the benchmark method proposed by Urbanek et al. (2013). The phase and rotational speed, without torsional information, were accurately estimated on the seven datasets using the proposed method, which consequently resulted in good order tracking results.

Machine condition features extracted from the continuous wavelet coefficients and from the wavelet packet transform were separately investigated in this study. The machine condition features, extracted from a healthy machine in a specific operating condition state, were modelled in the proposed fault diagnostic methodology using Hidden Markov models. Operating condition features, representative of the operating conditions, were extracted, modelled with a Hidden Markov model and used to determine the operating condition state of the machine. The information from the operating condition model and the machine condition model associated with each hidden state were

combined using probabilistic approaches. This is performed to quantify the deviation of the features from its expected value, in a specific operating condition state, in the form of a discrepancy signal.

The discrepancy signal is processed with synchronous averaging and additional post-processing so that the diagnostic information in the signal can be enhanced and the non-diagnostic information can be attenuated. The following novel post-processing techniques were proposed and successfully implemented in this study:

- A second synchronous averaging was performed over the synchronous averaged signals from multiple measurements. A bias estimation and subtraction process were used to obtain an unbiased second synchronous averaged signal which is a robust condition indicator.
- A gear-pinion discrepancy distribution was used to infer the condition of the gears within a gearbox. This distribution is very sensitive to changes in the data.
- A healthy-damaged decomposition is used to perform fault trending. The result of this is very intuitive to understand and can be used to track the stability of the developing fault.

A confidence bound, obtained from statistical theory, is used as a threshold for an alarm and is compared to the unbiased second synchronous average of the discrepancy signal for automatic fault detection. The unbiased second synchronous average was successfully used for fault detection, localisation and trending in this study. It is envisaged that the second synchronous average is sensitive to incipient faults if all of the measurements are correctly aligned (i.e. no errors in the cross-correlation maximisation process). The gear-pinion discrepancy distribution and the healthy-damaged decomposition provides invaluable information for gearbox diagnostics, since it intuitively represents the condition of the gearbox. The gear-pinion discrepancy distribution is very sensitive to changes reflected in the discrepancy signal, which makes it appropriate to detect the presence of incipient faults as well.

The fault detection and trending capabilities of the discrepancy signal from the continuous wavelet coefficients performed better than the discrepancy signal from the wavelet packet coefficients. This does not necessarily indicate that the continuous wavelet transform is better suited for fault diagnosis than the wavelet packet transform. It merely indicates that features that focus on very specific bands, using the Meyer basis functions, performs better than features extracted from the whole spectrum, using Daubechies db1 basis functions. However, the features extracted from the wavelet packet coefficients can potentially be used in cases where a lot of damage modes are present and the location of the damage modes, in the spectrum, are unknown. The

proposed diagnostic methodology was validated on numerical and experimental data and it satisfied all of the research objectives set out in section 1.3.

This methodology does not require historical fault data and is simpler than physics-based models, since it does not require the physics within the system to be modelled or compensated for. This novelty detection approach (i.e. only healthy data is available) combined with machine learning techniques provide a lot of advantages above standard fault detection techniques. The large amount of datasets, situated in large dimensional spaces, can be combined with expert knowledge and modelled so that small changes in the characteristics of the dataset can be automatically detected.

6.2 Recommendations

The following recommendations are made for future work:

- Distributed gear fault detection, localisation and trending need to be investigated, since it affects maintenance decisions differently than localised faults.
- A wider range of operating conditions needs to be investigated to evaluate the performance of the fault diagnostic methodology.
- Optimal machine condition features for general fault detection need to be investigated, since the success of the fault detection methodology depends on the quality of the features.
- Automatic harmonic selection techniques need to be investigated for the maxima tracking process. Techniques estimating the relationship between the tracked component and the shaft of interest needs to be investigated as well.
- This technique performed well on the experimental gearbox setup in the laboratory, however it needs to be investigated on gearboxes used in various industrial applications. This will help to evaluate the performance of the proposed methodology and validate it on those gearboxes.

Bibliography

Abboud, D., Antoni, J., Sieg-Zieba, S. and Eltabach, M. (2017), 'Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment', *Mechanical Systems and Signal Processing* **84**, 200–226.

Addison, P. S. (2002), *The illustrated wavelet transform handbook*, Institute of Physics Pub.

Antoni, J. (2009), 'Cyclostationarity by examples', *Mechanical Systems and Signal Processing* **23**(4), 987–1036.

Antoni, J., Bonnardot, F., Raad, A. and El Badaoui, M. (2004), 'Cyclostationary modelling of rotating machine vibration signals', *Mechanical Systems and Signal Processing* **18**(6), 1285–1314.

Arora, J. S. (2004), *Introduction to optimum design*, Elsevier/Academic Press.

Bartelmus, W., Chaari, F., Zimroz, R. and Haddar, M. (2010), '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.

Bartelmus, W. and Zimroz, R. (2009*a*), 'A new feature for monitoring the condition of gearboxes in non-stationary operating conditions', *Mechanical Systems and Signal Processing* **23**(5), 1528–1534.

Bartelmus, W. and Zimroz, R. (2009*b*), 'Vibration condition monitoring of planetary gearbox under varying external load', *Mechanical Systems and Signal Processing* **23**(1), 246–257.

Baydar, N. and Ball, A. (2000), 'Detection of Gear Deterioration Under Varying Load Conditions By Using the Instantaneous Power Spectrum', *Mechanical Systems and Signal Processing* **14**(6), 907–921.

Bishop, C. (1994), 'Novelty detection and neural network validation'.

Bishop, C. M. (1995), *Neural networks for pattern recognition*, Clarendon Press.

Bishop, C. M. (2006), *Pattern recognition and machine learning*, Springer.

Blough, J. R. (2003), 'A survey of DSP methods for rotating machinery analysis , what is needed , what is available', **262**(Journal of Sound and Vibration), 707–720.

Bonnardot, F., El Badaoui, M., Randall, R. B., Danière, J. and Guillet, F. (2005), '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.

Busby, R. L. (2012), *Wind power*, PennWell Corporation.

Chaari, F., Baccar, W., Abbes, M. S. and Haddar, M. (2008), '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.

Chaari, F., Bartelmus, W., Zimroz, R., Fakhfakh, T. and Haddar, M. (2012), 'Gearbox vibration signal amplitude and frequency modulation', *Shock and Vibration* **19**(4), 635–652.

Chaari, F., Fakhfakh, T. and Haddar, M. (2009), 'Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness', *European Journal of Mechanics, A/Solids* **28**(3), 461–468.

Combet, F. and Gelman, L. (2007), 'An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor', *Mechanical Systems and Signal Processing* **21**(6), 2590–2606.

Dalpiaz, G., Rivola, A. and Rubini, R. (2000), 'Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears', *Mechanical Systems and Signal Processing* **14**(3), 387–412.

Diamond, D. H., Heyns, P. S. and Oberholster, A. J. (2016), 'Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression', *Mechanical Systems and Signal Processing* **81**, 402–418.

Dowling, M. J., Lane, N., Park, L. and Pa, C. (1993), 'Application of non-stationary analysis to machinery monitoring', pp. 59–62.

Eggers, B., Heyns, P. and Stander, C. (2007), 'Using computed order tracking to detect gear condition aboard a dragline', *The Journal of The Southern African Institute of Mining and Metallurgy* **107**.

Fernández-Francos, D., Martínez-Rego, D., Fontenla-Romero, O. and Alonso-Betanzos, A. (2013), 'Automatic bearing fault diagnosis based on one-class ν-SVM', *Computers & Industrial Engineering* **64**(1), 357–365.

Figueiredo, E., Radu, L., Worden, K. and Farrar, C. R. (2014), 'A Bayesian approach based on a Markov-chain Monte Carlo method for damage detection under unknown sources of variability', *Engineering Structures* **80**, 1–10.

Fyfe, K. and Munck, E. (1997), 'Analysis of computed order tracking', *Mechanical Systems and Signal Processing* **11**(2), 187–205.

Gales, M. and Young, S. (2008), 'The Application of Hidden Markov Models in Speech Recognition', *Foundations and Trends® in Signal Processing* **1**(3), 195–304.

Gryllias, K. C. and Antoniadis, I. A. (2012), '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.

Gryllias, K. C. and Antoniadis, I. A. (2013), 'Estimation of the instantaneous rotation speed using complex shifted Morlet wavelets', *Mechanical Systems and Signal Processing* **38**(1), 78–95.

He, G., Ding, K., Li, W. and Jiao, X. (2016), 'A novel order tracking method for wind turbine planetary gearbox vibration analysis based on discrete spectrum correction technique', *Renewable Energy* **87**, 364–375.

Heyns, T., Godsill, S. J., De Villiers, J. P. and Heyns, P. S. (2012), 'Statistical gear health analysis which is robust to fluctuating loads and operating speeds', *Mechanical Systems and Signal Processing* **27**(1), 651–666.

Heyns, T., Heyns, P. S. and De Villiers, J. P. (2012), '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.

Heyns, T., Heyns, P. S. and Zimroz, R. (2012), '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.

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

Jedliński, Ł. and Jonak, J. (2015), 'Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform', *Applied Soft Computing* **30**, 636–641.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L. and Siegel, D. (2014), 'Prognostics and health management design for rotary machinery systems Reviews , methodology and applications', *Mechanical Systems and Signal Processing* **42**(1-2), 314–334.

Lee, K.-f. (1990), 'An Overview of the SPHINX Speech Recognition System', **38**(I), 35–45.

Lin, J. and Qu, L. (2000), 'Feature Extraction Based on Morlet Wavelet and Its Application for Mechanical Fault Diagnosis', *Journal of Sound and Vibration* **234**(1), 135–148.

Lin, J. and Zhao, M. (2014), A review and strategy for the diagnosis of speed-varying machinery, *in* 'Prognostics and Health Management (PHM), 2014 IEEE Conference on', IEEE, pp. 1–9.

Marwala, T., Mahola, U. and Nelwamondo, F. V. (2006), 'Hidden Markov Models and Gaussian Mixture Models for Bearing Fault Detection Using Fractals', *International Joint Conference on Neural Networks, Vancouver, Canada* pp. 3237–3242.

McFadden, P. D. (1985), 'Low frequency vibration generated by gear tooth impacts', *NDT International* **18**(5), 279–282.

McFadden, P. D. (1987), '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.

Mcfadden, P. D. and Smith, J. D. (1984), 'Model for the Vibration Produced by a Single Point-Defect In a Rolling Element Bearing', *Journal Of Sound And Vibration* **96**(1), 69–82.

Miao, Q. and Makis, V. (2007), 'Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models', *Mechanical Systems and Signal Processing* **21**(2), 840–855.

Montgomery, D. C. and Runger, G. C. (2011), *Applied statistics and probability for engineers*, John Wiley and Sons.

Ocak, H. and Loparo, K. a. (2005), 'HMM-Based Fault Detection and Diagnosis Scheme for Rolling Element Bearings', *Journal of Vibration and Acoustics* **127**(4), 299.

Pan, M.-C. and Lin, Y.-F. (2006), 'Further exploration of Vold-Kalman-filtering order tracking with shaft-speed information - I: Theoretical part, numerical implementation and parameter investigations', *Mechanical Systems and Signal Processing* **20**(5), 1134–1154.

Peng, Z. K. and Chu, F. L. (2004), '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.

Pimentel, M. a. F., Clifton, D. a., Clifton, L. and Tarassenko, L. (2014), 'A review of novelty detection', *Signal Processing* **99**, 215–249.

Purushotham, V., Narayanan, S. and a.N. Prasad, S. (2005), 'Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition', *NDT & E International* **38**(8), 654–664.

Qi, X., Yuan, Z. and Han, X. (2015), 'Diagnosis of misalignment faults by tacholess order tracking analysis and RBF networks', *Neurocomputing* **169**, 439–448.

Rabiner, L. R. (1989), 'A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition', **77**(2).

Rafiee, J., Rafiee, M. A. and Tse, P. W. (2010), 'Application of mother wavelet functions for automatic gear and bearing fault diagnosis', *Expert Systems with Applications* **37**(6), 4568–4579.

Randall, R. (1981), 'New Method of Modeling Gear Faults.', *American Society of Mechanical Engineers (Paper)* .

Randall, R. B. (2011), *Vibration-based condition monitoring*, Wiley.

Resor, B. R., Trethewey, M. W. and Maynard, K. P. (2005), 'Compensation for encoder geometry and shaft speed variation in time interval torsional vibration measurement', *Journal of Sound and Vibration* **286**(4-5), 897–920.

Samanta, B., Al-Balushi, K. R. and Al-Araimi, S. a. (2003), 'Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection', *Engineering Applications of Artificial Intelligence* **16**(78), 657–665.

Shao, R., Hu, W., Wang, Y. and Qi, X. (2014), '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.

Stander, C., Heyns, P. and Schoombie, W. (2002), 'Using Vibration Monitoring for Local Fault Detection on Gears Operating Under Fluctuating Load Conditions', *Mechanical Systems and Signal Processing* **16**(6), 1005–1024.

Stander, C. J. and Heyns, P. S. (2005), 'Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions', *Mechanical Systems and Signal Processing* **19**(4), 817–835.

Stander, C. J. and Heyns, P. S. (2006), 'Transmission path phase compensation for gear monitoring under fluctuating load conditions', *Mechanical Systems and Signal Processing* **20**(7), 1511–1522.

Staszewski, W. and Tomlinson, G. (1994), 'Application of the wavelet transform to fault detection in a spur gear'.

Sung, C. K., Tai, H. M. and Chen, C. W. (2000), 'Locating defects of a gear system by the technique of wavelet transform', *Mechanism and Machine Theory* **35**(8), 1169–1182.

Tandon, N. (1994), 'A comparison of some vibration parameters for the condition monitoring of rolling element bearings', *Measurement* **12**(3), 285–289.

Theodoridis, S. and Koutroumbas, K. (2009), *Pattern recognition*, Elsevier/Academic Press.

Timusk, M., Lipsett, M. and Mechefske, C. K. (2008), 'Fault detection using transient machine signals', **22**, 1724–1749.

Tuma, J. (2005), Setting the passband width in the vold- kalman order tracking filter, *in* 'Twelfth International Congresson Sound and Vibration', Lisbon, pp. 1–8.

Unal, M., Onat, M., Demetgul, M. and Kucuk, H. (2014), 'Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network', *Measurement* **58**, 187–196.

Urbanek, J., Barszcz, T. and Antoni, J. (2013), 'A two-step procedure for estimation of instantaneous rotational speed with large fluctuations', *Mechanical Systems and Signal Processing* **38**(1), 96–102.

Van Der Maaten, L. J. P., Postma, E. O. and Van Den Herik, H. J. (2009), 'Dimensionality Reduction: A Comparative Review', *Journal of Machine Learning Research* **10**, 1–41.

Vinson, R. G. (2014), 'Rotating machine diagnosis using smart feature selection under non- stationary operating conditions'.

Wan, X., Wang, D., Tse, P. W., Xu, G. and Zhang, Q. (2016), 'A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions', *Measurement* **78**, 138–150.

Wang, W. J. and McFadden, P. D. (1996), 'Application of wavelets to gearbox vibration signals for fault detection', **192**, 927–939.

Wang, W. Q., Ismail, F. and Farid Golnaraghi, M. (2001), 'Assessment of Gear Damage Monitoring Techniques Using Vibration Measurements', *Mechanical Systems and Signal Processing* **15**, 905–922.

Wang, X. Y., Makis, V. and Yang, M. (2010), 'A wavelet approach to fault diagnosis of a gearbox under varying load conditions', *Journal of Sound and Vibration* **329**(9), 1570–1585.

Wu, J.-D. and Liu, C.-H. (2008), 'Investigation of engine fault diagnosis using discrete wavelet transform and neural network', *Expert Systems with Applications* **35**(3), 1200–1213.

Wu, J.-D. and Liu, C.-H. (2009), '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.

Yang, C. Y. and Wu, T. Y. (2015), 'Diagnostics of gear deterioration using EEMD approach and PCA process', *Measurement: Journal of the International Measurement Confederation* **61**, 75–87.

Yen, G. G., Member, S. and Lin, K.-c. (2000), 'Wavelet Packet Feature Extraction for Vibration Monitoring', *IEEE Transactions on Industrial Electronics* **47**(3), 650–667.

Yiakopoulos, C. T., Gryllias, K. C. and Antoniadis, I. A. (2011), 'Rolling element bearing fault detection in industrial environments based on a K-means clustering approach', *Expert Systems with Applications* **38**(3), 2888–2911.

Zhao, M., Lin, J., Wang, X., Lei, Y. and Cao, J. (2013), 'A tacho-less order tracking technique for large speed variations', *Mechanical Systems and Signal Processing* **40**(1), 76–90.

Zienkiewicz, O. C. and Taylor, R. L. (2005), *The finite element method for solid and structural mechanics*, Elsevier Butterworth-Heinemann.

Zimroz, R., Bartelmus, W., Barszcz, T. and Urbanek, J. (2014), '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.

Zimroz, R. and Bartkowiak, A. (2013), 'Two simple multivariate procedures for monitoring planetary gearboxes in non-stationary operating conditions', *Mechanical Systems and Signal Processing* **38**(1), 237–247.

Chapter A Experimental setup

A.1 Characteristics of the setup

The schematic of the experimental setup in figure A.1 contains the equipment that is used during the experiments. The alternator applies a varying load to the system, while the motor applies a varying rotational speed to the system. Three gearboxes, denoted by GB in the schematic is used, to ensure that the maximum load is applied at the input shaft of the monitored gearbox (GB2). The different data streams are indicated in the figure with the characteristics of each data stream in figure A.2.

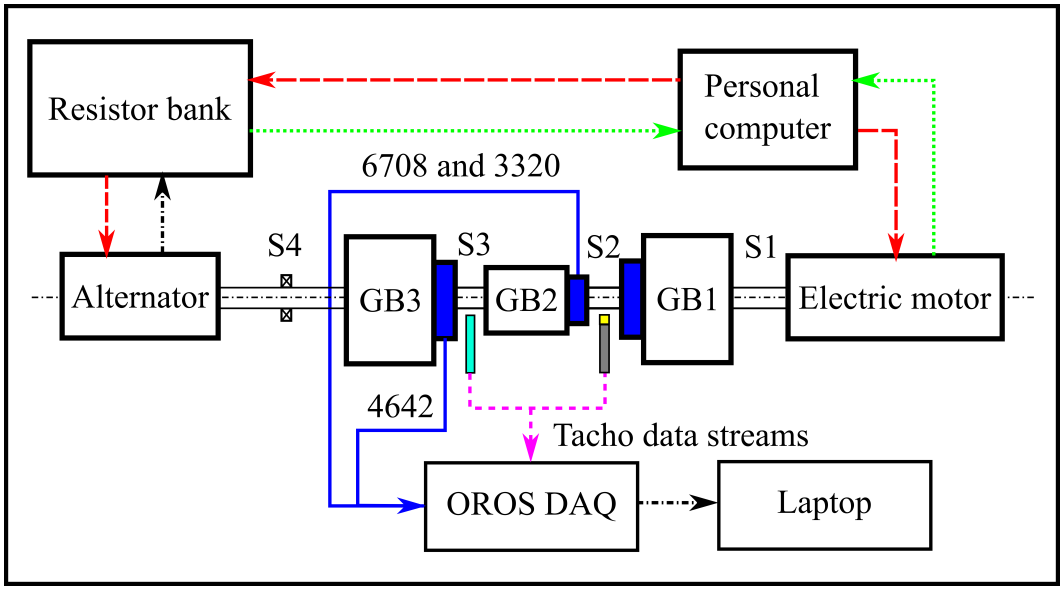


Figure A.1: Schematic of the experimental setup used in this study.

The personal computer has a National Instruments data acquisition card which is used to control the alternator as well as the motor. The alternating voltage and current from the alternator are converted to direct voltage and current before it is measured by the personal computer. The data streams from the accelerometers indicate the tri-axial accelerometer (6708) and the two 500mV/g uni-axial accelerometers measuring only in the vertical direction (denoted by 3320 and 4642). The axial-direction of the

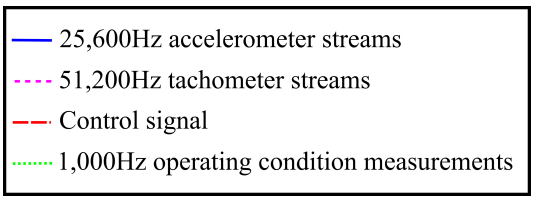


Figure A.2: Legend for experimental schematic in figure A.1.

accelerometer is used in this study. Two tachometers are used, the first tachometer is a proximity probe which is located on the key of the third shaft (indicated by S3) and the second is an optical probe measuring on a 88 pulse per revolution zebra tape shaft encoder. An eight channel OROS data acquisition system is used to acquire the five acceleration streams as well as the two tachometer streams. The characteristics of the gearboxes in table A.1 was obtained by the manufacturers, Siemens.

Table A.1: Characteristics of the gearboxes used in the experimental setup.

	E38-A-100	E68-A-100
Gearbox number in figure A.1	2	1,3
Gear ratio	1.85	4.93
Number of teeth:		
Pinion: N_p	20	15
Gear: N_g	37	74
Maximum output torque [N.m]	37.0	98.6
Maximum input torque [N.m]	20.0	∼20

The normalised characteristic frequencies of the shafts and the gears of the experimental setup are included in table A.2. The input shaft is located on the right of the gearbox in question in figure A.1.

Table A.2: Normalised rotation and gear mesh frequencies for the gearboxes.

	Gearbox 1	Gearbox 2	Gearbox 3
Rotational speed: Input	1.0	0.2027	0.375
Rotational speed: Output	0.2027	0.375	1.85
Meshing frequencies	15.0	7.5	27.75

A.2 Impulses in the vibration signal

The presence of impulses in the time domain signal was localised to the monitored gearbox (E38-A-100) by an experiment, but the cause of the impulses is unknown. A laser displacement transducer (LDT) was focused on the coupling attached to the input shaft of the monitored gearbox to detect the presence of axial movement. The experimental equipment is presented in figure A.3. The bearing that supports the input

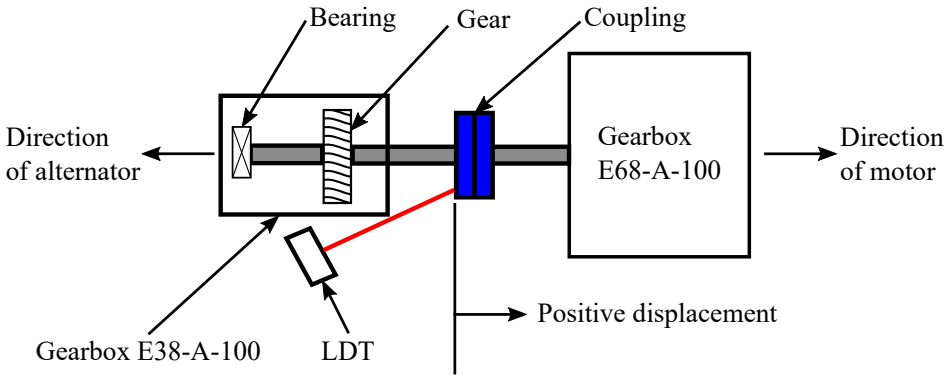


Figure A.3: The experimental equipment used to determine whether axial movement in the shaft is present during operation. The bearing which is suspected to cause the impulses is loose in its housing and is indicated in the schematic as well.

shaft of the monitored gearbox is loose in its housing. Circlips are the only constraints prohibiting large axial movement. Vibration measurements as well as measurements from the LDT were obtained for various rotational speeds and loads as indicated in figure A.4. The rotational speed associated with each measurement file is included

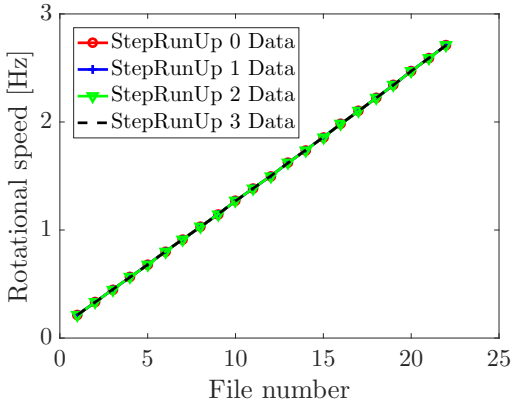


Figure A.4: The input rotational speed at the monitored gearbox versus the file number for the different load cases.

in figure A.4 with the load indicated in the legend. The convention that is used for a specific load is `StepRunUp` < `Load applied in volts` > `Data`. The applied load,

represented by a voltage value, was kept constant for a single set of measurements. The rotational speed of the system was constant for a measurement file and then it was increased incrementally for the next measurement file as indicated in figure A.4. The measured vibration signal in figure A.5 is obtained for the aforementioned operating conditions.

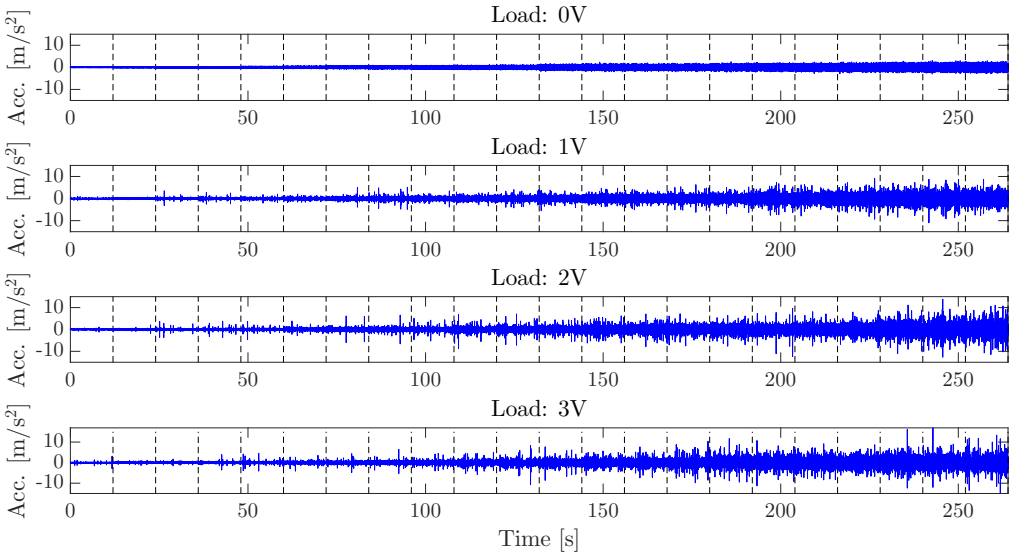


Figure A.5: The vibration signal against time for the different load cases. The vertical dashed lines indicate where new measurement files start.

The dashed lines indicate a new measurement file with the rotational speed indicated in figure A.4 and the load in the title in figure A.5. An interesting observation is made from the vibration signal between different load cases. No clear impulses are present if there is no external load from the alternator acting on the system. As the load increases (between sub-plots), the impulsive nature of the signal increases.

The results from the displacement transducer is presented in figure A.6. The displacement of the coupling (and the shaft) is dependent on the load as well as the rotational speed. A saturation is observed in the mean displacement presented in figure A.6 after the speed increases up to a point. Since the standard deviation between the measurements with a load applied to the system are relatively small and it is approximately constant, it is concluded that the mean displacements in figure A.6 are comparable.

The results from A.6 indicates that there are some axial movement present during operation and if the applied load is increased, the impulsive nature of the vibration signal increases. However, this does not directly prove that the axial movement is the cause of the impulses in the vibration signal. Another experiment was conducted with a new bearing used. The experiment was conducted continuously for longer than 45 days, with the bearing under consideration as well as continuously fluctuating loads. The

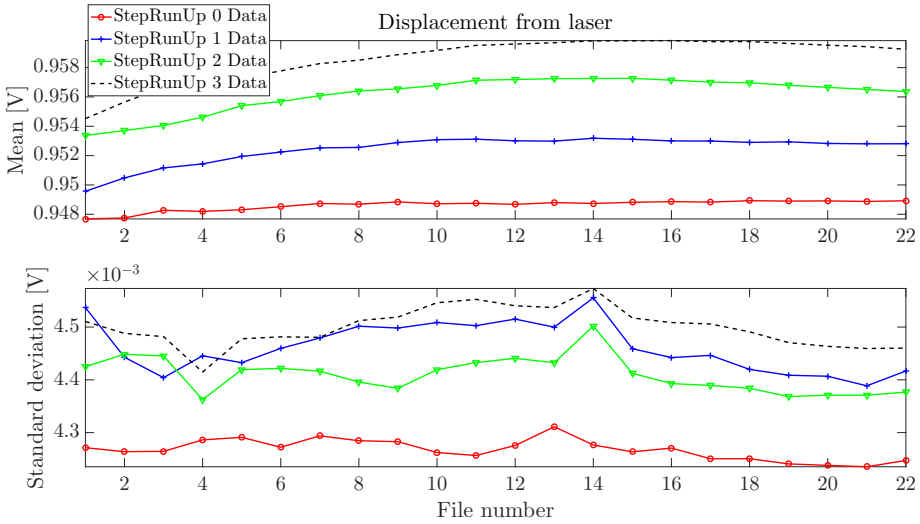


Figure A.6: The mean and standard deviation of the displacement of the coupling from the laser for the different load cases as the rotational speed increases. An increase in displacement indicates that the coupling moved further from the monitored gearbox and this indicates that the shaft moves as well.

outer surface of the bearing under consideration was investigated after the completion of the experiment with a photo included in figure A.7. There are some marks on the

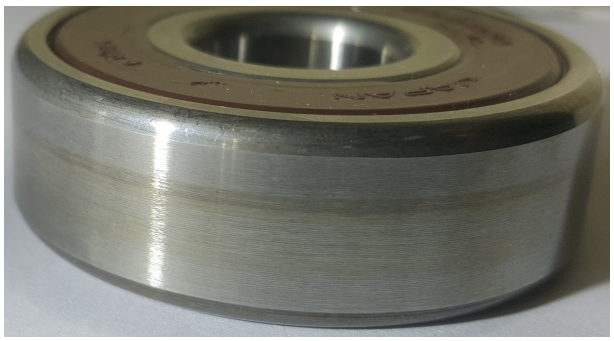


Figure A.7: The bearing with the marks on its outer ring.

outer surface of the bearing which was not there before the experiment was conducted. The marks were made in the radial direction of the bearing which indicates radial movement occurred as well. It is concluded from the results in this investigation that the bearing causes the impulses in the vibration signal. As the bearing is sliding within the housing it is expected that it will be momentarily stopped due to friction forces etc. which results in an impulse to occur. As the loads increases within the system, the effects of the phenomenon are more severe. These events may not occur periodically which explains the seemingly non-periodic behaviour of the impulses.

A.3 Temperature effects on the results

The synchronous average of the discrepancy signal over time is decomposed using the proposed healthy-damaged gear portion decomposition, with the result presented in figure A.8. Some very interesting characteristics are present within the data which

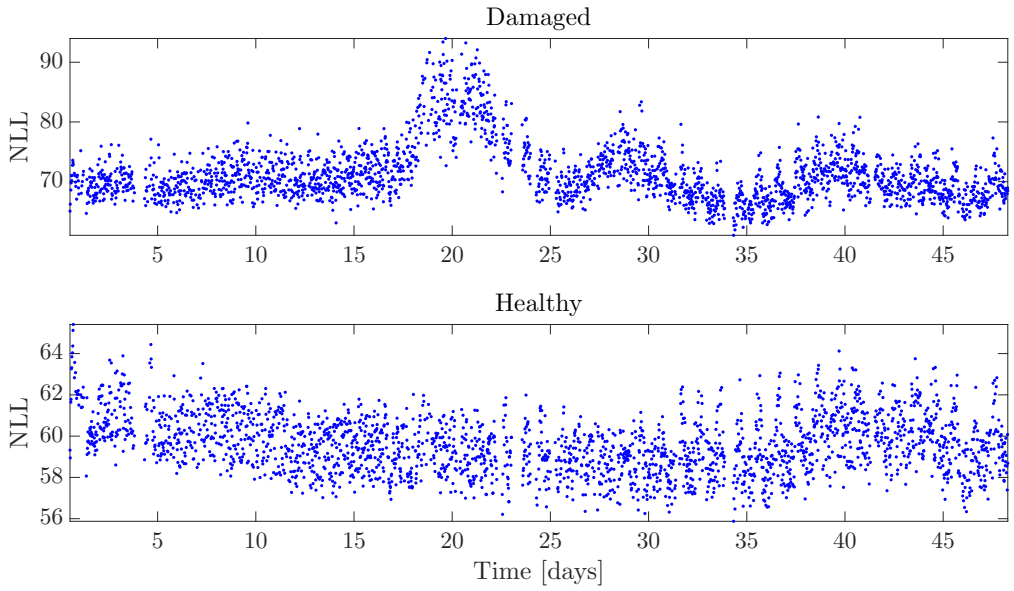


Figure A.8: The synchronous average of the discrepancy signal partitioned into the healthy and the damaged portion over measurement time. The mean of the healthy and the damaged portions.

is explained in the main report. The discrepancy within the healthy portion of the gearbox in figure A.9 indicates that there is periodic behaviour present during the last stages of the experiment. The x-axis is the form of $XX.YY$. The day from 1 July

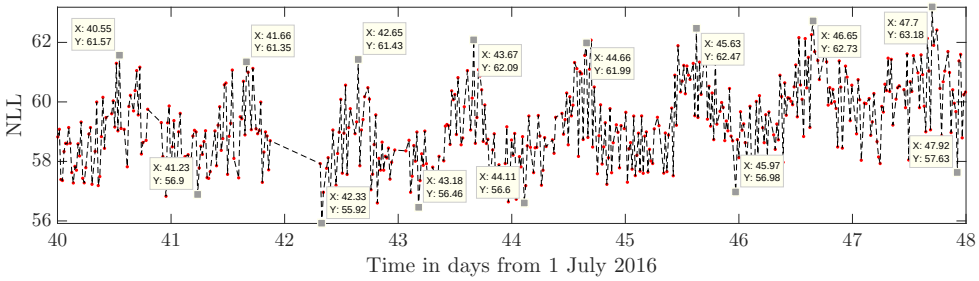


Figure A.9: Zoomed view of the mean discrepancy of the healthy portion of the gear. Note that the x-axis is from 1 July instead of the start of the experiments.

is indicated by XX and YY denotes the time from $00h00$. As an example, consider a x-axis reading of 41.66. This denotes that it is day 41 or 10 August and the 0.66 denotes that it is $0.66 \times 24 = 15.84$ hours from $00h00$. This can be converted to time

$15h50 + 00h00 = 15h50$. The 0.84 was converted to minutes. It can be observed that the maximum deviation occurred at approximately $15h00 - 16h00$. The minimum deviation occurs approximately in the middle of the night. Hence this indicates that there is a daily periodicity in the deviation in the healthy portion of the damaged gear and therefore the daily temperatures were investigated. The temperature data is obtained from the following website:
`http://www.timeanddate.com/weather/south-africa/pretoria/historic`
and plotted over time for different time frames in figure A.10. The average as well as the

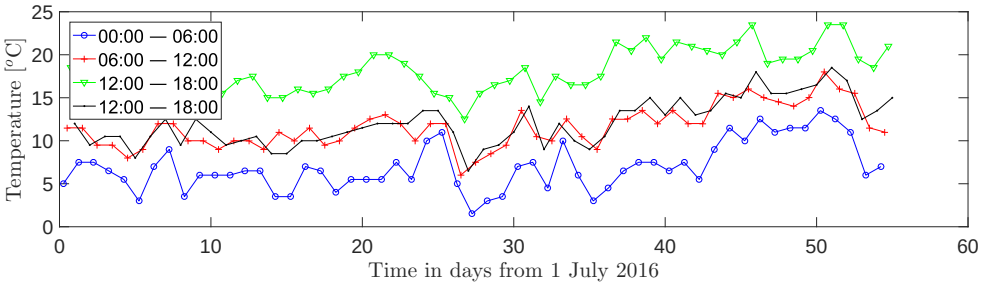


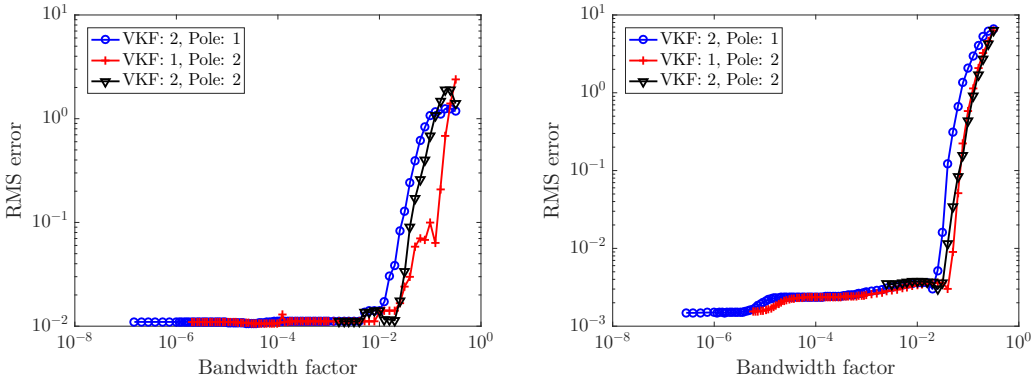
Figure A.10: Temperature over time for different times in Pretoria, South Africa.

minimum and maximum temperatures of each day was used instead of the temperature at the different periods of the day, to make the comparison with the discrepancy signal, easier. The temperature is compared to the discrepancy signal in figure A.8, in figure 5.22.

Chapter B Additional results

B.1 Vold-Kalman filter investigation

An investigation into the suitability of the angular-velocity and angular-displacement VKF is conducted on the different experimental signals. In figure B.1a and B.1b the RMSE are compared for different bandwidth factors for the angular-displacement (denoted by VKF 2) and the angular-velocity (denoted by VKF 1) VKF. The bandwidth factor is defined by equation (3.24). All of the results are based on the experimental profiles in figure 3.7a and 3.8a. The one-pole angular-displacement VKF is able to



(a) (b)

Figure B.1: Comparison between the RMSE from the different variations in the Vold-Kalman filters with experimental profile 1 in B.1a and profile 2 in figure B.1a. See section 3.2.3 for more details.

obtain very small differences to the actual rotational speed when very small bandwidth factors are used. The angular-velocity VKF is not suited for as small bandwidth factors as the one-pole angular displacement VKF, but it obtains low error estimates as well. The two-pole angular-displacement VKF is not capable of filtering signals with low bandwidth factors, however relatively low errors are obtained for its feasible bandwidth factors. It is concluded from the analyses investigated here that the one-pole angular-displacement VKF with a bandwidth factor of 10^{-4} to 10^{-6} is stable and cap-

able of obtaining low RMSE. This motivates the use of 10^{-4} as the bandwidth factor for the investigated signals.

Even though the kth harmonic is tracked, any of the other harmonics can be used to centre the Vold-Kalman filter. The most appropriate harmonic of the IF for centring the VKF is investigated in this section. In figure B.2 the results of the investigation for the profile in figure 3.7a is shown, which is investigated for the angular-displacement VKF.

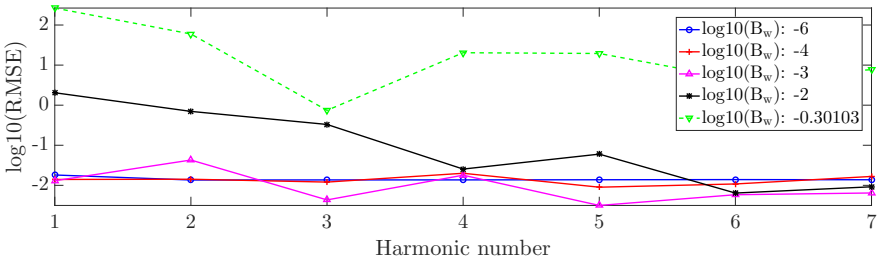


Figure B.2: Different harmonics of the estimated IF are used as the centre of the Vold-Kalman filter and different bandwidth factors, indicated by B_w, are investigated. This is performed for the same profile investigated in figure B.1.

The results of this analysis is shown for the second experimental profile investigated (see figure 3.8a) in figure B.3 as well. The large bandwidth factor of 0.5 has a relatively

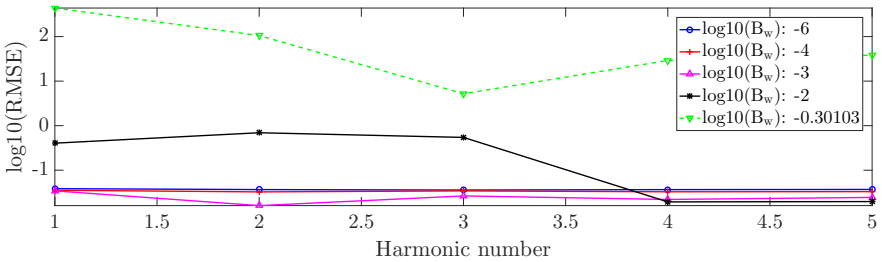


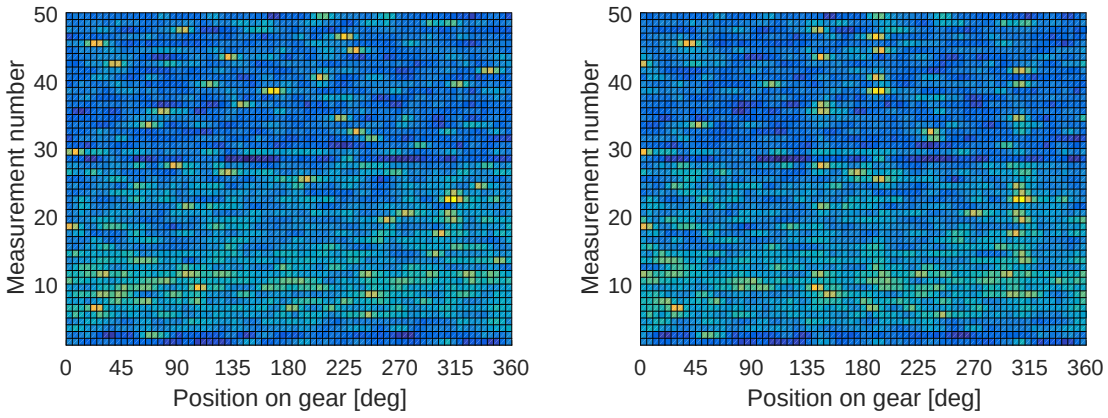
Figure B.3: The logarithm of the RMS error (RMSE) is computed from using different harmonics to centre the VKF and with different bandwidth factors, indicated by B_w. This is performed for the profile in figure 3.8a.

large error, since the large bandpass bandwidth includes other components not associated with the harmonic of the IF. The very small bandwidth factors (10^{-4} and 10^{-6}) obtain stable values for all the harmonics. The resulting bandwidths are so small that sidebands etc. are not included into the passband of the filter. As the bandwidth is increased the error can possibly increase or decrease between different harmonics, since it depends on the frequency characteristics that are present at the harmonics as well as the maxima tracking errors. Therefore, even though a global minimum error can be obtained when using a larger bandwidth for a specific signal, it is difficult to predict how it will react for other signals when the operating conditions changes etc. It is

therefore suggested that a very small bandwidth needs to be used on the investigated signals to ensure that the errors are predictable.

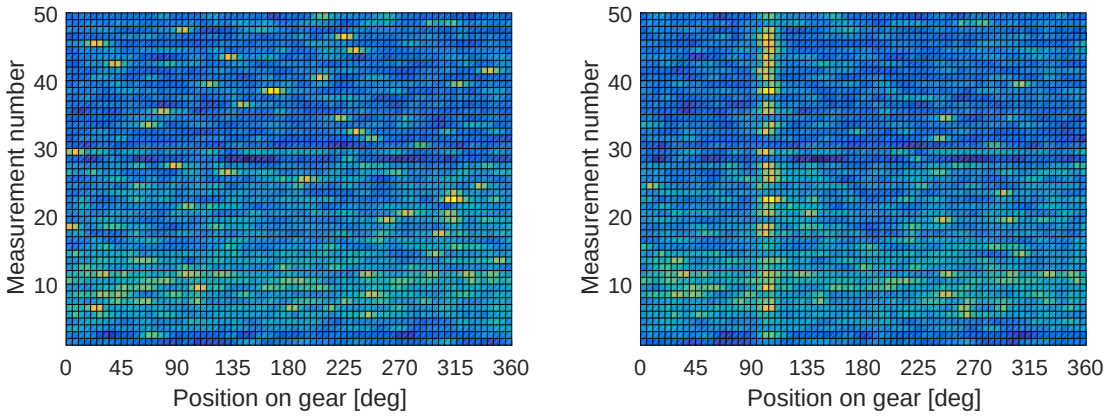
B.2 Cross-correlation maximisation

The results before and after performing the cross-correlation maximisation process using the first and last signal as a reference is presented in figure B.4 and B.5, respectively. The last signal is a better reference, since the fault is most probably more prominent at the last measurement than at the first measurement. This makes the task of cross-correlation maximisation easier. Each row represents the discrepancy obtained from a measurement.



(a) Before (b) After

Figure B.4: First signal as reference



(a) Before (b) After

Figure B.5: Last signal as reference