UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Engineering, Built Environment and Information Technology

Informative frequency band selection for performing envelope analysis under fluctuating operating conditions in the presence of strong noise and deterministic components

by **WN Niehaus**

1 November, 2019

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

Summary

Informative frequency band selection for performing envelope analysis under fluctuating operating conditions in the presence of strong noise and deterministic components

by

WN Niehaus

Supervisor:	Prof. PS Heyns
Co-supervisor:	Dr. S Schmidt
Department:	Mechanical and Aeronautical Engineering
Research group:	Centre for Asset Integrity Management
University:	University of Pretoria
Degree:	Master of Engineering (Mechanical Engineering)
Keywords:	Novelty information, fluctuating speed, envelope analysis, informative frequency band, deterministic components, blind separation

Condition-based maintenance is an important aspect in various industries to ensure reliable operation of machinery. To successfully execute maintenance responsibilities, it is required to know which components are healthy and which are in a damaged state. Thus, the need for effective incipient fault detection requires a method that can separate fault signatures from operating condition information. Conventional gearbox monitoring techniques assume that a change in the vibration signal is caused by the presence of a fault. Under constant operating conditions this assumption may be valid, but under fluctuating conditions the assumption does not hold.

Fluctuating operating conditions are inevitable for gearboxes in mining and wind turbine industries due to fluctuating ground and wind properties. The fluctuating conditions cause smearing of the signal frequency spectrum and valuable diagnostic information is lost when using classical condition monitoring techniques. More sophisticated signal processing techniques are therefore needed to effectively diagnose incipient faults to make informed asset management decisions.

In this dissertation, envelope analysis, which has long been recognized as one of the best methods for bearing fault diagnosis, is used as the primary diagnostic tool. A common precursor to envelope analysis is bandpass filtering which is aimed at emphasising bearing faults and removing background noise and deterministic components. Identification and optimal selection of the informative frequency band which contains damage related information is the focus area for research in this dissertation. Many automatic band selection techniques exist and have proven effective under constant speed conditions. However, it has been shown that these techniques occasionally identify frequency bands that contain non-damage related information, especially under fluctuating speeds and low damage

levels.

With this research, a new methodology is proposed which makes use of popular informative frequency band selection techniques, such as the Fast Kurtogram amongst others, to effectively identify damage under constant and fluctuating speed conditions. The proposed methodology uses both healthy and damaged vibration signals to identify novelty information. In doing so, the method can also identify damage earlier than existing methods. The technique is designed to ignore potentially dominant deterministic components which would lead to incorrect band selection for envelope analysis.

Furthermore, pre-whitening of vibration signals is a common technique to enhance the bearing signal-to-noise ratio. Without pre-whitening, random noise and deterministic components often dominate the bearing fault signatures and render existing diagnostic techniques ineffective. The proposed methodology is shown to be more robust than existing automatic band selection methods because it requires no pre-whitening. By using both healthy and damaged signals, the proposed methodology favours frequency bands that contain damage related information.

The findings in this dissertation are validated on a range of synthetic signals as well as on actual experimental data. The synthetic signals are constructed from a phenomenological gearbox model where the exact operating and bearing condition can be controlled. The experimental results are statistically compared for a wide range of signals and damage levels such that the robustness of the proposed method can be critically evaluated. It was found that the new method is capable of outperforming existing methods in terms of percentage classification of bearing signals with outer race damage and can detect damage with smaller fault severity.

Acknowledgements

First and foremost, I would like to thank my lord and saviour Jesus Christ for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without His blessings, this achievement would not have been possible.

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

- Prof. Stephan Heyns
- Dr. Stephan Schmidt
- Mr. George Breitenbach
- EPPEI for funding this research

Finally, a special word of gratitude to my family and friends for their unwavering support during my studies.

Contents

List of abbreviations

AR	Autoregressive
CBM	Condition-based maintenance
CF	Crest factor
COT	Computed order tracking
CPW	Cepstrum pre-whitening
CS1	First-order cyclostationary
CS2	Second-order cyclostationary
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
EDM	Electro discharge machining
EHNR	Envelope harmonic-to-noise ratio
FFT	Fast Fourier transform
HFRT	High-frequency resonance technique
IFB	Informative frequency band
IR	Inner race
NIC	Novelty information criterion
OR	Outer race
REB	Rolling element bearing
RMS	Root mean square
ROC	Receiver operating characteristic
RPM	Revolutions per minute
RUL	Remaining useful life
SE	Squared envelope
SES	Squared envelope spectrum
SNR	Signal-to-noise ratio
STFT	Short-time Fourier transform
TSA	Time synchronous average
WT	Wavelet transform
WVD	Wigner-Ville distribution
ZTSE	Zebra tape shaft encoder

1 Introduction

A vibration signal may contain vast amounts of information related to the condition and health of a rotating machine and its subcomponents. Timely identification of an incipient fault can save industry large sums of money and assists the maintenance team with making and performing strategic decisions and repairs on critical components. The difficulty lies in separating and extracting the useful diagnostic information from the baseline healthy state and interfering noise. Noise and non-related vibrations are picked up as a result of the vibration transducer itself along with the vibrations caused by other machinery in the vicinity. Therefore, the ability to autonomously extract the relevant signals and diagnose a machine's health with non-invasive techniques during normal operation has numerous benefits in the world of condition-based maintenance (CBM). Benefits include reducing the probability of infant mortality of newly and unnecessarily replaced components. Furthermore, CBM helps find the balance between plant health and plant availability by moving unscheduled maintenance to scheduled maintenance. Cost of maintenance can also be decreased by repairing and returning to service a gearbox prior to a catastrophic failure.

A perfect analysis would allow the identification of the type of damage, the exact location of the damage, size of the damage and finally the remaining useful life of the component before scheduled downtime must be planned. Unfortunately, there is no perfect processing technique which can accomplish all of the above ideal requirements. The choice of which method or combination of methods to use depends largely on the signal characteristics to select the optimum method. The method will depend on operating conditions, machinery-type, available budget which determines the type of equipment available to capture the data, as well as the expected fault types and levels of noise. Many methodologies have been developed over decades which build and improve on previous work. Stationary operating condition methodologies have proven very successful and current work largely focusses on improving more complicated non-stationary operating condition analysis techniques. Research reports are available on varying speed and varying load vibrations independently, but analysing both varying load and varying speed simultaneously has large scope for further research.

The field of signal processing in the context of mechanical engineering applications related to CBM has been developing momentum over the past few decades. The demand for improved product quality and increasing competition requires better methods to analyse system and component integrity. The purpose of this research dissertation is to highlight the pros and cons of the most popular signal processing techniques used in the vibration based condition monitoring industry, as well as present some of the ground-breaking techniques presented by leading researchers in the field. This dissertation pays specific attention to what will be referred to as the "Grams", such as the Fast Kurtogram, Infogram, Sparsogram, Autogram and a few others to be mentioned later in the dissertation. Each Gram attempts to isolate the informative frequency band (IFB) by finding a frequency band which maximizes a specific feature. Furthermore, a new diagnostic methodology is proposed which can be applied on any one of the popular Grams. The proposed methodology can be applied to any Gram and improves the diagnostic potential under stationary

and varying speed conditions.

1.1 Scope

Early detection of rotating machine faults in the presence of fluctuating operating conditions remains an important and challenging task. Much work has been done in developing methods that are capable of detecting and quantifying bearing damage and growth for constant operating condition applications. The methods, discussed in Section 2, have shown promising results even when bearing damage is very small and noise large. However, under fluctuating speed and load, most of these methods fail. In this dissertation, the fluctuating speed signals from the experimental setup remained within $\pm 25\%$ of the median speed.

When operating conditions change, the characteristics of the vibration signal naturally change with it too. Distinguishing whether the change in vibration signal is due to changing operating conditions or due to the presence of damage is thus the main focus of further research and the primary objective in this dissertation. It is not uncommon to have healthy vibration data with similar operating conditions to the vibration signals that are suspected of containing damage. As such, it is possible to use the healthy data as a benchmark for comparison when trying to detect damage in fluctuating operating conditions. The outcome of this dissertation is to develop a diagnostic methodology that,

- Can detect and locate bearing damage under fluctuating conditions,
- Uses no a priori information about the fault condition,
- Can be used for online fault diagnosis to trigger an alarm,
- Makes use of healthy data to identify damage.

The purpose of this dissertation is not to develop a method for optimal threshold selection, as it is beyond the scope of this dissertation. It is assumed that a suitable threshold is known and that any signals which trigger an alarm thus contain damage.

1.2 Layout of dissertation

This dissertation focuses on signal processing in the context of mechanical engineering and is structured as follows. Section 2 begins with a discussion about the need for CBM and common failure mechanisms so that the importance of further research is understood. Section 2 proceeds further to discuss and introduce the reader to the techniques commonly applied to vibration signals starting with constant operating conditions. Finally, methodologies related to varying load and speed are discussed ending with a phenomenological model which is used as the preliminary signal by which these methods are evaluated. Section 3 expands on the specific phenomenological model which is to be used and thoroughly explains each aspect of the model. The author then proceeds to the actual experimental setup, describing all components and the scenarios which will be considered for analysis. Photos from the actual experiment as well as schematics of the setup and damage

types are provided to allow the reader clear understanding of the setup and experimental database.

Section 4 is dedicated to the analysis of both phenomenological and experimental vibration signals with the intent of developing a method which outperforms current techniques in terms of bearing diagnostics under fluctuating operating conditions. Firstly, cepstrum pre-whitening and autoregressive models are discussed, while highlighting their potential applications and shortcomings. A new methodology is then developed which is designed not to change the statistical properties of a signal, like cepstrum pre-whitening does. The aim of this method is to automatically identify informative frequency bands in the presence of strong deterministic components which would otherwise dominate any spectral analysis technique. The proposed method is evaluated on fluctuating speed conditions and statistically compared to existing methods.

In this dissertation, only the most relevant methods which require no a priori knowledge of the damage condition are explored. The results are highlighted in Section 5 where the work of this paper is summarised as well as scope for future research is suggested. In this section, ideas which were not fully investigated are touched upon for further research. Section 5 can serve as a blueprint for future research which can make a valuable contribution to the field of mechanical systems and signal processing.

2 Literature Review

This section discusses relevant literature in the field of vibration based condition monitoring and provides the required information needed to understand the discussions presented in Section 4. A brief introduction to the field of maintenance is presented so that the importance of the context of this work can be fully appreciated. This dissertation is specifically focused on vibrations related to gearboxes. Of the many components in a gearbox, gears and bearings, along with their vibration signals are of particular interest in the early detection of damage in rotating machinery. Therefore, the mechanisms by which these components commonly fail are discussed and how they manifest in physical vibration signals so that the reader can understand the nature of the signals analysed. The remainder of Section 2 focuses on discussing the tried and tested methods that have been used to analyse vibration signals for decades, as well as some of the newest methods that have been gathering momentum over the last few years and have shown promising initial diagnostic potential.

2.1 Maintenance strategies and condition-based maintenance

Reliability is an important aspect in the assessment of industrial processes and equipment. Assessing the reliability of a component and the remaining life is equally important since all products and processes deteriorate over time due to their operating conditions (Jardine et al. 2006). Maintenance is thus the process of efficiently ensuring the reliability of equipment or processes during their useful life. Maintenance on machines in industry is an unavoidable yet critical task and can represent from 15 to 70% of total production costs (Bevilacqua & Braglia 2000). Unexpected breakdowns due to a lack of maintenance can be very expensive. Common maintenance techniques include replacing machines or components based on the amount of time that they have been in operation and also replacing system components upon failure. These techniques can result in unforeseen downtime and unnecessary expenses. A more efficient alternative is condition-based maintenance (CBM) which uses the condition of assets to determine which components require maintenance or replacement and which are still in an acceptable condition for operation. Traditionally, the main aim of CBM was early detection of incipient failure which permits scheduled maintenance and downtime instead of unexpected and potentially catastrophic failure. With more advanced signal processing methods and technology available, CBM programs are now equally concerned with predicting the remaining useful life (RUL). A CBM program consists of three steps: data acquisition, data processing and a maintenance decision-making step (Lee et al. 2004). This dissertation will mostly focus on the 2nd step, data processing or signal processing. Ideally, a CBM program would be able to exactly identify and predict the RUL of a structure, however this is a problematic task. Generally, in order of complexity, CBM programs aim to achieve any of the following levels of health monitoring (Doebling et al. 1998):

Level 1: Determine presence of damage

Level 2: Determine geometric location of damage

Level 3: Quantify severity and size of damage

Level 4: Predict RUL

In order to successfully identify faults, it is required to know something about the mechanisms that lead to failure and how they manifest in the vibrations of the machine.

2.1.1 Fault mechanisms of gears

A large part of CBM is focussed on the analysis of gears and rolling element bearings (REB) as they are found in most rotating machines. Knowing the different effects that each mechanism has on the vibration signal is the first step in analysing the potential faults. All rotating machines will have vibrations, even when in the healthy or undamaged state. Thus, for signals acquired under stationary operating conditions, any deviance from the healthy signal is assumed as a result of a fault. However, this is not the case for signals captured under non-stationary operating conditions which are subject to both fluctuating load and speed and make fault detection significantly more difficult. Nonetheless, the gear failure mechanisms are typically classed into the following categories (Davis 2005):

- Shock loading
- Fatigue failure
    - Bending fatigue
    - Pitting
- Wear
- Failure due to scuffing
- Misalignment

Shock loading occurs when a significantly large and sudden load is applied to the gear tooth surface causing the tooth to break instantaneously. This manifests in the vibration signal as a once per gear revolution high amplitude impulse at the gear rotational frequency and its harmonics.

Fatigue failure contains two common categories when considering gears. Bending fatigue is ever present due to the nature of the applied loads on gears. When this load is applied cyclically, cracks form at the root of the gear tooth and then propagate until a critical crack length where the tooth fails. Local faults such as fatigue cracks affect the sideband amplitudes.

Pitting is a form of surface wear caused by the compressive contact stresses on the gear tooth surface. Material is removed from the meshing surface causing small pits to be formed. Wear is the slow time dependent removal of material from the tooth meshing surface. Wear is not a local defect as it will most likely occur on all the teeth on a gear. Wear causes the gear teeth to become thinner. From an analysis perspective it is known that wear increases the amplitude of the sidebands of the gear meshing frequency (Vinson 2014). If gear wear causes an uneven tooth profile, the vibration signal may also

show impulses but these impulses may be difficult to detect due to uniform wear over all gear teeth. Both pitting and wear result in similar vibration profiles and may be hard to distinguish from the vibration signal alone.

Misalignment is when the gear meshing surfaces are not aligned, i.e. the profile of the gears do not follow on the meshing surfaces as designed. This may be caused by an axial shift of one of the gears and is characterised in the vibration signal by an increased amplitude at the gear meshing frequency. Finally, it has also been found that gear quality plays a significant role in the detection of damage. Low quality gears have a rougher and less precise gear tooth profile and may vary significantly between consecutive teeth. These manufacturing imperfections may appear in the vibration signal as damage but they also make quantifying the severity and size of the damage more demanding. For example, Wang et al. (2001) found that phase modulation is very sensitive to any imperfections in gears.

2.1.2 Fault mechanisms of bearings

Degraded bearings have unique vibration patterns that are indicative of different failure modes. A healthy bearing typically exhibits almost no vibration, whereas a failing bearing will display bearing tones related to the vibration patterns of the fault (Klein et al. 2011). Bearings can fail due to many reasons (Rubini & Meneghetti 2001), including:

- Initial production quality
- Contaminated lubrication
- Heavier loading than anticipated
    - This is often caused by imbalance, misalignment or a bent shaft.
- Improper installation
- Surface fatigue
- Plastic deformation
- Acid corrosion

Klein et al. (2011) presented a dephase algorithm derivative that is very successful in separating bearing tones from noise and concluded that it can be successfully implemented as an automatic diagnostic tool without prior knowledge of the nature of the bearing fault. They compared it to the basic dephase orders and envelope orders and found it superior to both existing methods. The key to the success of these methods, along with many other bearing diagnostic methods, is knowledge of the expected signal generated by specific faults in the REB. When a REB is damaged, either the inner race (IR), outer race (OR), cage or rolling element frequencies will be excited depending on where the damage is located. Statistically, 90% of ball bearing faults are related to either and IR or OR faults (Rubini & Meneghetti 2001). These characteristic frequencies are generally provided by the bearing manufacturer as a function of the operating speed in the bearing specifications. However, due to the nature of bearing damage, these diagnostic

frequencies cannot be detected from spectral analysis alone, discussed in Section 2.7.1, because of a phenomenon called amplitude modulation. Amplitude modulation is defined as the multiplication of two signals with one another. In terms of bearing signals, a low frequency modulating signal, typically indicative of a bearing fault, is multiplied by a carrier frequency such as the bearing housing resonant frequency (McInerny & Dai 2004). An example is shown in Figure 1 where a low frequency modulating signal is multiplied with a high frequency carrier signal to obtain the modulated signal.

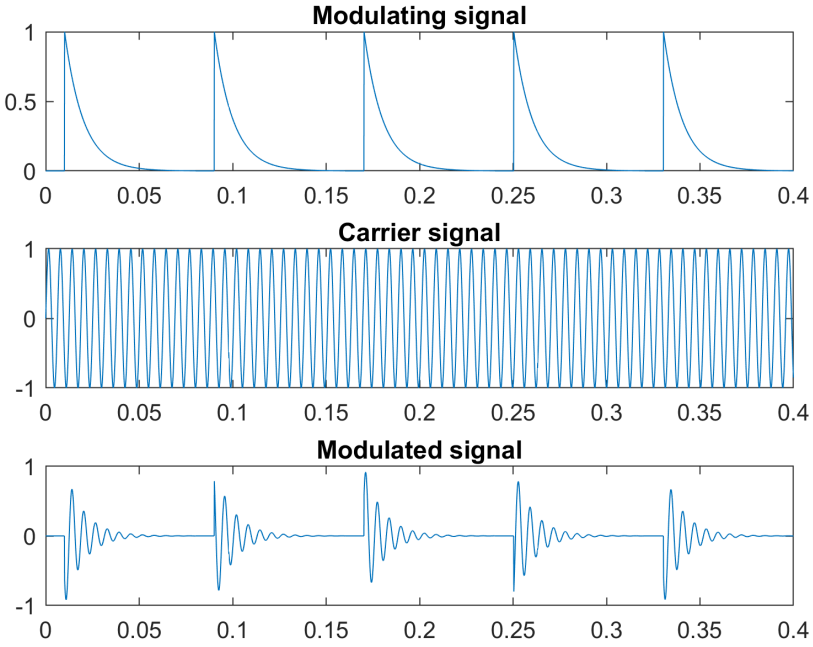


Figure 1: Illustration of amplitude modulation.

2.2 Vibration measuring methods

Vibration measurement is at the heart of CBM and selection of the correct tool is critical in capturing accurate and usable data. Vibration measurement can be used as a continuous and autonomous tool capable of diagnosing the real time condition of rotating machines in an effective, reliable and non-intrusive way (Goyal & Pabla 2016). As previously stated, a CBM program consists of three steps: data acquisition, data processing and a maintenance decision-making step. Vibration measurement, using vibration transducers, falls under the data acquisition step, and includes the following devices (Goyal & Pabla 2016):

- Displacement transducers
- Velocity transducers
- Accelerometers
- Laser Doppler vibrometers

2.3 Feature-based methods

The aim of feature based methods is to identify and accentuate specific features in a signal which may be indicative of a specific type of gear or bearing damage. Raw signals are noisy and offer little-to-no useful diagnostic information to an untrained eye. More often than not, vibration signals require a great deal of further processing and filtering in order to render features of interest. Hence, the selection of signal processing method and in which domain is a precursor to the features that can be extracted.

Dhamande & Chaudhari (2018) investigated the extraction of features from the time, frequency and time-frequency domain with the intention of discovering the potential of identifying compound bearing and gear faults. They found that in the time domain and frequency domain, very few features have diagnostic potential. However, in the time-frequency domain the wavelet transform, discussed later in Section 2.9, yielded excellent diagnostic results. Tandon (1994) investigated statistical time-domain features and their ability to detect bearing faults and it was found by Lin & Zhao (2014) that certain features were suitable but only to stationary operating conditions. Zimroz et al. (2014) expanded the work to consider features under harshly varying load/speed conditions. They did this by representing the data in the feature-operating condition space and used statistical processing of data, linear regression analysis, as fault sensitive features for CBM decision making.

2.4 Time-domain analysis

Time-domain analysis makes use of descriptive statistics such as mean, standard deviation, crest factor, kurtosis, peak-to-peak interval, time synchronous average (TSA) and many other metrics to identify characteristic features. Of particular interest is TSA which is an efficient method of noise reduction. By taking the waveform average signal over multiple time windows, random components such as noise will be eliminated because they are not synchronous with the average waveform (Braun 1975) and it is assumed that the noise is normally distributed. Some early vibration analysis was performed using TSA by Stewart (1977). TSA is defined by Jardine et al. (2006) as

$$\bar{s}(t) = \frac{1}{N}\sum_{n=0}^{N-1} s(t+nT), 0 \leq t < T, \tag{1}$$

where $s(t)$, T and N denote the signal, averaging period and number of samples for averaging respectively. TSA works particularly well for isolating gears of interest because all signals that are not synchronous with that gear will be removed. Working in the time domain can offer useful diagnostic or descriptive metrics which can be used to identify the nature of the signal or identify progressive damage. The TSA signal represents the meshing of teeth over one or more complete revolutions which highlights the pattern of tooth meshing and allows for the detection and location of a cracked tooth (Barszcz & Randall 2009). A thorough review of TSA algorithms was presented by Bechhoefer & Kingsley (2009), and he found that TSA is an essential algorithmic tool for determining

the condition of rotating equipment. McFadden & Smith (1985) proposed an improvement on traditional TSA which demodulates the amplitude and phase of the TSA signal to extract hidden features caused by varying load conditions.

2.4.1 Descriptive metrics

Working in the time domain can offer useful diagnostic or descriptive metrics which can be used to identify the nature of the signal or identify progressive damage.

- Peak values

  The peak values that are of concern are the peak value (maximum value of vibration) and the peak-to-peak value (difference between the extreme values of vibration). These values are often used to differentiate between acceptable and unacceptable levels of vibration (Lebold et al. 2000).

- Root mean square value (RMS)

  The RMS of a signal is calculated using Equation (2).

$$X_{RMS} = \sqrt{\frac{\int_0^T x(t)^2 dt}{T}} \tag{2}$$

- Crest factor

  The crest factor (CF) is the ratio between the peak vibration level, X_{max}, to the RMS level as defined below and is a measure of the signals impulsiveness,

$$CF = \frac{X_{max}}{X_{RMS}}. \tag{3}$$

  CF as a diagnostic tool has been found to be a poor indicator of incipient failure (Tandon 1994), and RMS is generally a better approach. However, CF can be a reliable measure for highly impulsive signals. As a rule of thumb, a CF between $2.5 < CF < 3.5$ is a healthy bearing, and a CF larger than 3.5 indicates damage.

- Kurtosis

  A common statistical method in machine diagnostics, specifically the fourth statistical moment, is kurtosis defined in Equation (4),

$$kurtosis = \frac{E[x^4]}{\sigma^4} = \frac{1}{\sigma^4}\int_{-\infty}^{\infty} x^4 p(x) dx = \frac{1}{\sigma^4 T}\int_0^T x^4 dt. \tag{4}$$

  Kurtosis indicates the spread of the distribution, where a value of 3 indicates a Gaussian distribution. Kurtosis can be a good indication of early damage, however as the damage grows and becomes physically distributed, the impulsive content will decrease and so will the kurtosis value, ultimately, resulting in false conclusions.

- Probability density

  The probability density function, $p(x)$, over a variable, x, describes the probability that data will assume a value within a defined range at any instant of time. Metrics such as the kurtosis, crest factor or RMS are often calculated on many healthy signals to define a distribution for the metric. In the presence of a fault, the calculated metric for that recording will hopefully fall outside of the distribution, triggering an alarm. Probability density has been found to be a good general indicator of bearing health under certain conditions where most bearing vibrations have a Gaussian distribution. Damaged bearings tend to have non-Gaussian distributions due to the presence of faults causing high-levels of time-localised acceleration, however, this is not always the case (Tandon & Choudhury 1999).

Acceleration is the second derivative of displacement, which makes it proportional to the shaft rate, squared. For this reason, condition indicators based on first and second moments are very sensitive to changes in operational speed. This supports the use of condition indicators like kurtosis which are based on the fourth moment, making them more stable to changes in operating speed.

2.5 Residual signal analysis

Residual signal analysis is one of the most widely used vibration diagnostic tools. The residual signal can be obtained by subtracting some modelled version of the signal from the original signal. Models that are used include autoregressive models, linear prediction, TSA and bandpass filtering around prominent gear-mesh harmonics (McFadden & Smith 1985). A very simple example is with TSA where the residual signal can be found by subtracting the regular (healthy) signal (obtained from TSA) from the original time domain average to leave only the signal that has diverged from the average. The aim is to remove all normal and healthy components so that only useful machine diagnostic components which are indicative of damage, remain. Thus, residual signal analysis can be seen as a digital filtering technique. McFadden (1987) demonstrated that by removing the tooth meshing components and their harmonics from a time domain average, evidence of a defect could be seen before it is detectable in the TSA signal.

Residual signals have been generated in many different ways in literature. Wang & Wong (2002) used the TSA of a healthy gear to establish an autoregressive (AR) model which is then used as a linear prediction error. Then by subtracting the filtered from the unfiltered signal, an AR model residual signal is generated. They found that the AR model technique is an effective tool in the detection and diagnosis of gear faults. Brie et al. (1997) developed a real-time implementation method utilising amplitude and phase demodulation which looks at the statistical parameters of the difference between a reference signal and the actual model. They found this to be a robust method insensitive to slow variations of amplitude and phase.

2.6 Discrepancy analysis

Discrepancy analysis is similar to residual signal analysis where the healthy components of a signal are removed, thus leaving only the components indicative of damage. Discrepancy analysis however, takes a statistical approach to the same idea and does not physically remove the 'healthy' components from the signal. Instead, a model based on the healthy signal is generated and a novel signal is statistically compared to this reference model. The discrepancy signal is formed by the likelihood with respect to this reference model (Vinson 2014). Therefore, since the reference model is based on a purely healthy signal, any discrepancy from that signal is assumed to be as a result of a fault. If the reference model takes non-stationary conditions into account, the effects of fluctuating load and speed can also be removed. Normal signal processing techniques can now be used on the discrepancy signal for further diagnosis.

2.7 Frequency-domain and related analysis

2.7.1 Spectrum analysis

Frequency-domain analysis is performed on a transformed signal in the frequency domain. Commonly this is done via spectrum analysis by means of a fast Fourier transform (FFT). Spectrum analysis entails looking at a specific or range of frequencies which can be used to extract features from the signal. An FFT analysis makes the assumption that the signal is stationary, periodic and can be approximated as the sum of many sinusoids of different frequencies, amplitudes and phase. Thus, by relating the amplitude of the frequencies to the physical parameters of the system, faults can be detected as the amplitude of the spectrum grows. It was shown by Stander et al. (2002) that damaged gears generally result in increased signal energy at the gear mesh frequency and its sidebands. For non-stationary signals or signals with a change in the operating condition, sidebands occur which result in spectral energy smearing into adjacent frequency bins rendering the spectrum impractical for diagnostic purposes (Troldborg & Sørensen 2014). The FFT can not interpret the change in frequency signal content over time and can not be used for monitoring structures under dynamic excitations (Goyal & Pabla 2016). Therefore, frequency-domain analysis is best applied to signals that operate under time-constant operating conditions. Methods used to analyse non-stationary signals include but are not limited to order tracking and time-frequency analysis which will be discussed later in this dissertation.

2.7.2 Cepstrum analysis

An in depth review of the history of cepstrum analysis and its application to mechanical problems are discussed in a recent paper by Randall (2017). The first main area of applicability is detecting, classifying and removing families of harmonics from a vibration signal. The second major application is for the blind separation of source and transfer functions which is also applicable to variable speed conditions. The power cepstrum of a

signal is mathematically defined as,

$$\tilde{s}(t) = FT^{-1}[ln|S(f)|], \tag{5}$$

with $S(f) = FT[s(t)]$, FT being the Fourier transform (El Badaoui et al. 2004). In simpler terms the power cepstrum of a signal is the inverse Fourier transform of the natural logarithm of the Fourier transform. El Badaoui et al. (2004) found that the power cepstrum is a robust fault indicator that can be used for rotating machine diagnosis and works best when signals are acquired under angular sampling using a shaft encoder. Cepstrum analysis is also independent of signal amplitude, signal-to-noise ratio and the position of the sensor and is ideal for monitoring and isolating many independent gear meshing frequencies. Cepstrum analysis highlights the periodicity in a spectrum and can therefore effectively recognise gear faults. However, for complicated systems with many different components, conventional cepstrum and spectrum analysis is less efficient (Wang et al. 2001). Dalpiaz et al. (2000) found from experimental analysis that another drawback of the power cepstrum is that it is insensitive to crack evolution and can therefore not be used to monitor crack growth. Historically, the power cepstrum has been used to detect the presence of echoes in signals and El Badaoui et al. (2004) give an intuitive description of this. The function of the natural logarithm in Equation (5) is to transform a convolution product, $s = s_1 * s_2$, into an addition, $\tilde{s} = \tilde{s}_1 + \tilde{s}_2$.

2.8 Time-frequency analysis

Time-frequency analysis is a three dimensional time, frequency and amplitude representation of a signal. Where spectrum analysis is only applicable to stationary signals, time-frequency techniques can be used to isolate features from non-stationary signals. Adaptations of the stationary FFT such as the short-time Fourier transform (STFT) can be applied to non-stationary signals. This is accomplished by using small chronological windowing functions over the whole signal to represent the amplitude in a time-frequency domain. The resolution of the STFT is a function of both time and frequency, so a trade-off is done between the accuracy in the time or frequency domain. Jardine et al. (2006) state that the most popular time-frequency analysis techniques are the STFT and spectrogram (the power of the STFT) as well as the Wigner-Ville distribution (WVD). Spectrograms are best applied to systems with slow changing dynamics. Each of these methods has its own drawbacks, and supplementary techniques have been developed to overcome these short-comings. The spectrogram has a time-frequency resolution issue and is thus best applied to signals with slow changes in their dynamics. Bilinear transforms such as the Wigner-Ville distribution overcome the resolution issue because they are not based on signal segmentation. However, these transforms have an interference issue due to the transformation itself. A Choi-Williams distribution was then developed to overcome the interference terms in the WVD (Peng & Chu 2004). Currently a very popular alternative time-frequency analysis technique is the wavelet transform.

2.9 Wavelet transform

Over the past two decades, much development has been done on wavelet transform techniques to aid with machine fault diagnostics. The continuous wavelet transform (CWT) uses shifted and scaled wavelets instead of sinusoids as the basis function to decompose signals (Vinson 2014). A continuous wavelet transform is defined by the following mathematical function,

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^*\Big(\frac{t-b}{a}\Big)dt, \tag{6}$$

where $x(t)$ is the waveform signal, a is the scale parameter, b is the time parameter and $\psi(.)$ is the chosen wavelet form. Popular wavelet basis functions include Morlet, Meyer, Haar, Mexican hat and Daubechies. Liu et al. (2008) state that the choice of an appropriate mother wavelet depends on the signal properties and applications. They found that in bearing fault detection, a Morlet wavelet gave superior results when compared to other wavelet functions. The complex Morlet wavelet, Equation (7), is a complex exponential carrier with a frequency, f_c, multiplied by a Gaussian window.

$$w(t) = ce^{-\sigma^2 t^2} e^{-i2\pi f_c t} \tag{7}$$

Figure 2 shows an example of a Morlet waveform.

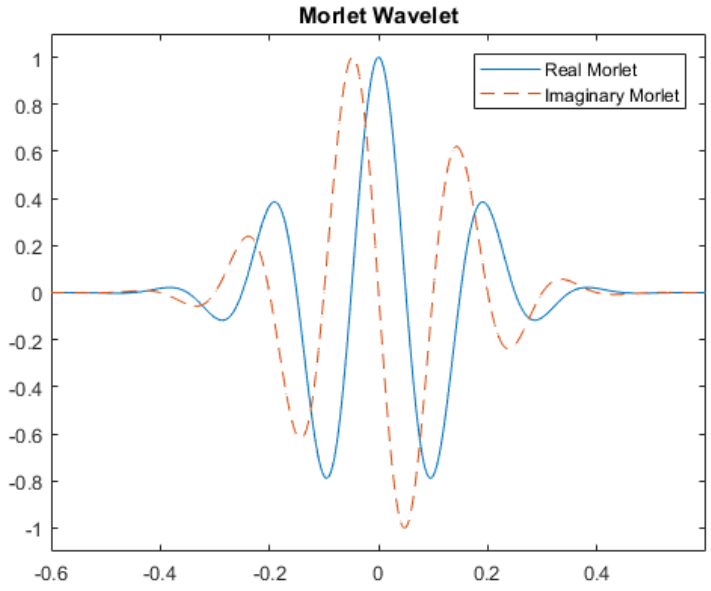


Figure 2: Morlet basis function example.

Wavelet analysis decomposes a signal into a weighted set of scaled wavelet functions. This makes wavelet analysis an attractive option for the analysis of signals with abrupt changes such as discontinuities and transients. The scaling allows for an adjustable frequency since the window lengths are a variable and not constant as with other time-frequency techniques. This allows long time intervals where precision for low frequency information

is required and short time intervals where high frequency information is required (Stander 2005). To achieve adequate resolution, analysis must be carried out at high frequencies such that the time resolution is comparable to the defect impulse duration (Rubini & Meneghetti 2001). Essentially, the wavelet transform (WT) solves the resolution issue encountered by the STFT due to its adjustable window sizes (Giurgiutiu & Lingyu 2003). The CWT has a wide range of uses ranging from denoising to spectrum analysis.

Gryllias & Antoniadis (2013) investigated the ability of shifted Morlet wavelets to estimate the instantaneous rotation speed of machinery and found that very good results were achieved as compared to the Hilbert transform. Dalpiaz et al. (2000) found that wavelet analysis is particularly useful when processing the residual part of a time-synchronous averaged signal.

Figure 3 is an academic example to highlight some of the benefits of the CWT as opposed to an FFT. The vibration signal consists of random noise along with a sinusoid of frequency 100 Hz at t<1, and a second sinusoid of frequency 50 Hz at t>3. Although the FFT is capable of detecting the two frequency components of the two signals, an FFT alone does not effectively deal with discontinuities and is limited in its application because the times at which each frequency is active can not be seen. Figure 4, however, shows the benefit of using the CWT and is able to identify the frequency components as well as the time where they are active. The white dashed line represents the cone of influence of the Scalogram. The area outside of the white dashed lines are potentially affected by edge-effects where the wavelets are stretched beyond the edges of the observation interval. Within the cone of influence we can be sure that the Scalogram is an accurate time-frequency representation.

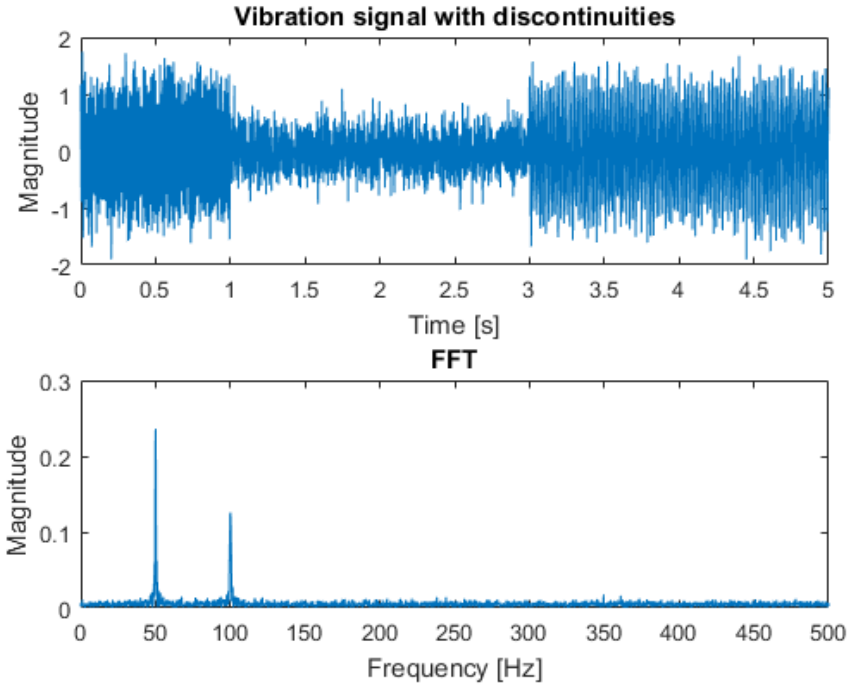


Figure 3: Academic example showing the inability of an FFT to highlight signal discontinuities.

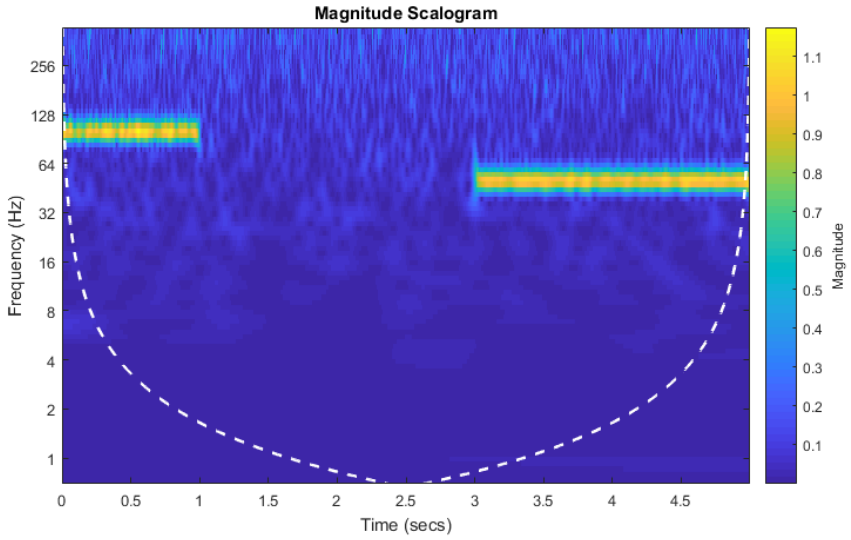


Figure 4: Scalogram using CWT on a signal with discontinuities.

Peng & Chu (2004) discuss some of the shortcomings related to the WT which suffers from energy leakage and the effects of border distortion. Also, the phase spectrum is not robust to noise and can lead to uncertainty if a signal becomes contaminated with noise. They also claim that aliasing can be a problem due to the fact that the WT is based on convolution which causes overlapping. Despite these shortcomings, the WT generally performs better than methods such as FFT or STFT due to the adaptive time-frequency ability. Peng & Chu (2004) touch on many of the abundant applications of the WT and its wide use for fault diagnostics. The WT is also an effective preprocessing tool often used to denoise and extract the weak parts of a signal.

The purpose of signal preprocessing is to increase the signal-to-noise ratio (SNR) which enhances the data's reliability and accuracy of signal analysis. Increasing the SNR is particularly important for early fault detection because the features, such as kurtosis, are normally weak and dominated by the noise from the environment and testing instrument itself. Some WT denoising techniques with proven success are the soft-thresholding technique by Donoho (1995) and wavelet shrinkage denoising by Zheng et al. (2000) along with many other methods. Peng & Chu (2004) did a thorough review of the WT and all its applications which cover almost every aspect of fault diagnostics. The review covers applications of the WT to time-frequency analysis, fault feature extraction, singularity detection, denoising and extraction of weak signals, the compression of vibration signals and system identification.

2.10 Order tracking

An alternate way of analysing non stationary signals is by utilising a technique known as order tracking. Spectral smearing caused by non-stationary vibration signals can be overcome by working in the angle domain as opposed to the time domain. This requires

the ability to capture the instantaneous angle of the rotating shaft to produce a tachometer signal which can be used to transform the time domain signal into the angle domain, known as order tracking. Fyfe & Munck (1997) defines order tracking as "a frequency analysis method that uses multiples of the running speed (orders), instead of absolute frequencies (Hz), as the frequency base". Popular methods of order tracking include conventional order tracking, computed order tracking (COT) and recently tacholess order tracking.

Conventional order tracking uses special analog hardware to sample at a rate synchronous to the shaft speed. Whereas with COT, the signal is sampled at a constant rate and then in a post processing step the data is re-sampled at constant angular increments. It may not always be feasible to install the required hardware, such as optical probes and zebra tape shaft encoders, to perform COT. Thus, methods have been developed to accurately estimate the phase of a shaft such as the method developed by Schmidt et al. (2018). In COT, the instantaneous shaft speed is not always known and thus interpolation is required to estimate the instantaneous angle of the shaft. Fyfe & Munck (1997) found that higher-order interpolating functions greatly improved the accuracy of the method. They also investigated the accuracy of COT with regard to keyphasor timing, filtering, rotational speed, interpolation method, noise and block size. Figure 5 illustrates the use of COT where an increasing frequency sine wave is re-sampled such that it has a uniform frequency in the order domain.

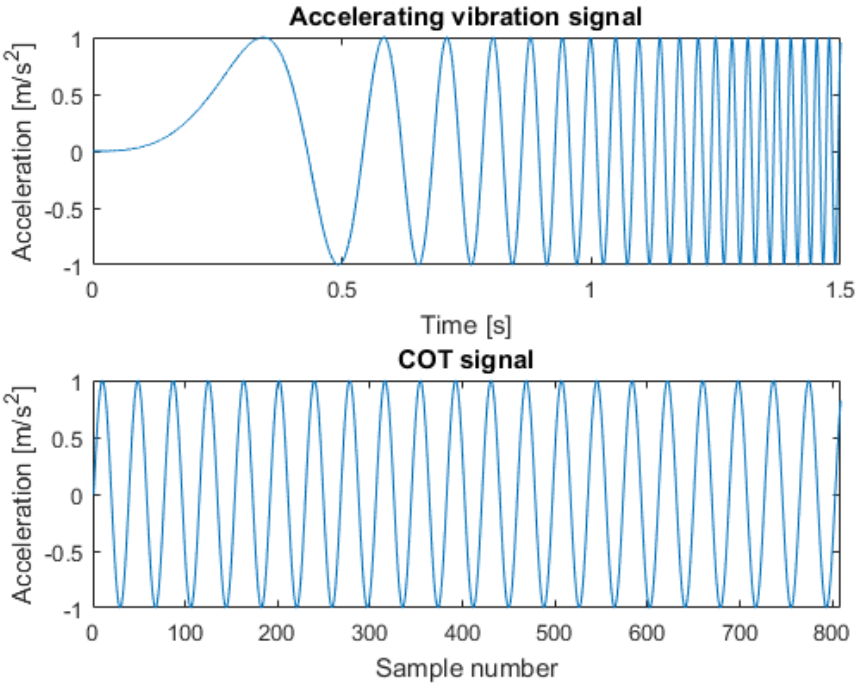


Figure 5: Accelerating sine wave vs a COT sine wave signal

2.11 Cyclostationary analysis

Often, non-stationary signals occur in a cyclic manner known as cyclostationary behaviour. Radar, sonar and telemetry were the first applications for the theory of cyclostationary processes. Since then, much work has been done on applying this theory to vibration signals. A rotating machine vibration signal is made up of a combination of periodic and random processes due to the machine's physical characteristics, rotation cycle and interaction with the real world. Dalpiaz et al. (2000) define cyclostationary signals as n-order cyclostationary if its statistical moments until the nth order are periodically time-dependent. Therefore, a cyclostationary process is a signal that has statistical properties that vary cyclically with time. A practical example is the temperature profile of a city. The temperature on a specific day is statistically different to the temperature on another day many months later, however, it is reasonable to assume that the same day over many years will be statistically identical on that specific day. This can be viewed as 365 stationary processes.

Mccormick & Nandi (1998) performed some good precursory work in the field of wide-sense cyclostationary statistics with special focus on estimation of features, time-frequency estimation of signal structures and a method for determining a machine's rotation speed from only the vibration signal. Antoni et al. (2004) define cyclostationarity as a statistical property that characterises periodically varying stochastic processes with regards to a generic variable such as time. The two most basic cyclostationary signals are first-order (CS1) and second-order (CS2) cyclostationary signals, also referred to as wide-sense cyclostationarity. CS1 vibrations are signals whose first-order moment or average (expected value) is periodic with T,

$$m_x(t) \triangleq E\{\boldsymbol{x}(t)\} = m_x(t+T), \tag{8}$$

where $E\{.\}$ means ensemble average.

A CS2 vibration is similarly a signal whose second-order moment (variance) is periodic with T, the period. Antoni et al. (2004) specifically refer to the autocorrelation function, $R_{2x}(t_1, t_2)$, as being periodic,

$$R_{2x}(t_1, t_2) \triangleq E\{\boldsymbol{x}^*(t_1)\boldsymbol{x}(t_2)\} = R_{2x}(t_1+T, t_2+T). \tag{9}$$

Higher order cyclostationarity also exists and when all moments up to infinity are periodic, the signal is strict sense cyclostationary. Antoni (2009) produced a convenient and thorough tutorial on cyclostationarity and its specific mechanical applications in machine diagnostics with many examples.

2.12 Envelope analysis

Envelope analysis, also commonly referred to as High-Frequency Resonance Technique (HFRT), is a demodulation technique commonly used on vibration signals. It is an effective tool in separating the useful spectra of a signal (the spectra indicating damage) from the background noise by separating the carrier frequency from the modulating frequency.

Through the use of the Hilbert transform, frequency analysis shifts to the low range of fault frequencies, from the high range of resonant frequencies excited by the faults. Envelope analysis helps identify bearing defects that may otherwise be hidden in the direct spectrum, by extracting characteristic frequencies. This technique is a powerful tool to extract vibrations with very low energy where the bearing fault is hidden in the total signal due to the magnitude of the background noise.

Envelope analysis generally consists of the following steps once the raw time signal has been captured:

- Bandpass filtering

  Take the raw input time domain signal and perform bandpass filtering. The purpose of bandpass filtering is to remove the vibrations at the frequencies that are known not to be indicative of a bearing fault. The low-frequency high amplitude signals associated with imbalance and misalignment as well as random noise outside the pass band are attenuated (McInerny & Dai 2004).

- Full wave rectification

  The signal is then demodulated or rectified. Rectification is often done using the Hilbert transform which is a linear operator on the bandpass filtered signal.

- Power spectrum

  The power spectrum of the absolute rectified bandpass filtered signal is then obtained to get the characteristic defect frequency.

Recent work on envelope analysis is done with the intent of extending the capabilities of envelope analysis from constant to variable speed conditions (Abboud et al. 2017). This is performed through the coupling of order tracking with the squared envelope spectrum (SES) to obtain an order domain representation. Work is also aimed at generalising existing cyclostationary tools with different approaches proposed by Urbanek et al. (2013) and D'Elia et al. (2010). Geropp (1997) showed the success of envelope analysis as a reliable method for detecting and isolating bearing defects but also that it can be used in many other applications such as gearbox diagnosis, diesel engines, hitting machine parts, and in extreme cases misalignment and imbalance. Randall & Antoni (2011) state that the benchmark method for bearing diagnostics is envelope analysis where the incoming vibration signal is bandpass filtered in a high frequency band in which the faults are amplified by the structural resonances. The signal is then amplitude demodulated which forms the envelope signal. This envelope contains the useful diagnostic information which can be used to distinguish between different fault types in a bearing. The Hilbert transform is typically the instrument of choice for demodulation and returns a complex analytic signal. It has also been shown that it is preferable to analyse the squared envelope rather than the traditional rectified absolute analytic signal (Randall & Antoni 2011). The challenge with envelope analysis is to strategically specify a bandpass frequency range that isolates the fault of interest.

2.13 Optimum frequency band selection

Identifying the most informative frequency band (IFB) that isolates the fault from heavy background noise and interferences has lead to the development of the Grams which attempt to automatically identify the IFB by selection of a frequency band which maximises a particular feature indicative of damage. The most well-known Gram is the Kurtogram which is also commonly used as the benchmark method when comparing new methods. The only Grams to be discussed and investigated in this dissertation are those that require no a priori knowledge about the fault, known as blind methods.

A blind method or blind source separation is the process of extracting individual vibration signals/sources from a mixture of observed signals and excitation forces (Antoni 2005). The benefits of such methods are that the results are independent of user input, no extensive pre-processing is required to identify hyperparameters and the model is more robust to user error. The required hyperparameters are however often difficult to calculate, which may drastically change the results. Due to the above reasons, this dissertation will focus on blind methods like the Grams which will be discussed in chronological order.

2.13.1 Kurtogram

Antoni & Randall (2006) define the Kurtogram as a fourth-order spectral analysis tool capable of characterising non-stationarities in a signal. In essence, the Kurtogram attempts to calculate a kurtosis value at each frequency of a signal. The term spectral kurtosis (SK) is termed as the kurtosis of a signals frequency components and is used in the development of the Kurtogram. SK is a spectral statistic just as time based kurtosis is a time based statistic. When a signal deviates from a Gaussian signal, its fourth-order cumulant increases and gives a measure of the impulsiveness or peakiness of a signal. Thus, SK is capable of detecting non-Gaussian components in a signal such as REB faults and cracked gear teeth. The spectral kurtosis can be defined by

$$K_x(f) = \frac{\langle |H(n,f)|^4 \rangle}{\langle |H(n,f)|^2 \rangle^2} - 2, \tag{10}$$

where $H(n,f)$ is the complex envelope of $x(t)$ at frequency f and level n. Antoni (2006) proved that the SK can be used to detect transients in the presence of strong background stationary noise. Antoni & Randall (2006) demonstrate how SK is used in the Kurtogram and that it is often used as a prerequisite to envelope analysis when the transients are buried in strong stationary background noise. This is done by finding the maximum value of the Kurtogram which provides both the optimal central frequency, f_c, and bandwidth, B_f, to be used in a band-pass filter.

Antoni (2007) developed a Fast Kurtogram algorithm that can be used for on-line industrial applications which calculates the SK for various parameters of frequency and bandwidth. The algorithm was found capable of accurately identifying an IR fault, ball fault and OR fault on several actual industrial signals (Antoni 2007). Barszcz & Randall (2009) used the SK to detect a tooth crack in a planetary gear of a wind turbine and found it was capable of detecting the crack several weeks before gear failure, where

contemporary methods failed. The process of using SK in analysing a signal is as follows (Barszcz & Randall 2009):

1. Calculate the Fast Kurtogram.
2. Design a bandpass filter with results from the Fast Kurtogram.
3. Analyse the filtered signal.

SK has been found to detect faults where many other methods fail, however, there are certain cases where SK too fails. According to Barszcz & JabŁoński (2011), in the presence of strong non-Gaussian noise containing high peaks or for a relatively high defect impulse repetition rate, SK will not accurately detect the presence of a fault. Since the development of the Fast Kurtogram, many other features have been proposed to overcome its limitations.

2.13.2 Enhanced Kurtogram

The Enhanced Kurtogram was suggested by Wang et al. (2013) and calculates kurtosis based on the power spectrum of the envelope signal. Wang et al. (2013) claim that the Enhanced Kurtogram performs better than the Kurtogram under low SNR and also non-Gaussian noise signals. The major difference of this technique is that the kurtosis is calculated in the spectrum domain as opposed to the temporal domain. The power spectrum more clearly shows sparsity (which is often indicative of a transient fault from a REB) and the kurtosis then measures the protrusion of the sparse representation. Wang et al. (2013) investigated 5 cases and compared the detection capability of the Kurtogram to the Enhanced Kurtogram:

- Case 1: Simulated bearing fault signals with a single resonant frequency
- Case 2: Simulated bearing fault signals with two resonant frequencies
- Case 3: A real bearing outer race fault signal obtained from an experimental motor
- Case 4: A real bearing inner race fault signal obtained from an experimental motor
- Case 5: A real bearing ball fault signal obtained from an experimental motor

For all 5 cases the Enhanced Kurtogram could detect the fault and correctly isolated the IFB. The Kurtogram on the other hand only identified the fault in the last 3 cases but with a lower visual inspection ability.

2.13.3 Sparsogram

The Sparsogram was recommended by Tse & Wang (2013*a*) and Tse & Wang (2013*b*) and calculates the sparsity of the power spectra to identify the resonant frequency band. When a REB suffers from a localized fault, the impacts caused by the fault excite a resonant frequency band. At the resonant frequency band, the sparsity value will be higher as a result of the definitive peaks caused by the fault at the characteristic frequencies. The

Sparsogram attempts to quantify the sparseness of a signal with the sparsity value defined by,

$$S = \frac{||d(f)||_2}{||d(f)||_1}, \tag{11}$$

where $||d(f)||_1$ and $||d(f)||_2$ are the L_1 and L_2-norms, respectively. It was found that bearing fault signatures could be extracted even at very low SNR values and that single as well as multiple combined faults could be extracted for real and simulated signals.

2.13.4 Infogram

Repetitive transients are often indicative of rotating machine faults. In the field of thermodynamics, transients are seen as departures from a state of equilibrium. Thus, Antoni (2016) proposed a new feature, negentropy (negative entropy), to detect rotating machine faults. The Infogram contains two components, the SE Infogram and the SES Infogram which calculate the negentropy of the squared envelope (SE) and the squared envelope spectrum (SES) respectively. The Infogram provides two measures of impulsiveness, one in the time domain (SE Infogram) and one in the frequency domain (SES Infogram). The SE and SES Infogram are additive, which allows for the calculation of the Average Infogram. The SE Infogram measures the time domain impulsive content and the SES Infogram measures the impulsiveness of the cyclic content. Spectral negentropy, $I_\varepsilon(f; \triangle f)$, is calculated as the negative of the spectral entropy, $H_\varepsilon(f; \triangle f)$, in the frequency band $[f - \triangle f/2,\ f + \triangle f/2]$. The SE Infogram values are thus calculated by,

$$I_\varepsilon(f; \triangle f) = -H_\varepsilon(f; \triangle f) = \left\langle \frac{\varepsilon(n; f, \triangle f)^2}{\langle \varepsilon(n; f, \triangle f)^2 \rangle} \ln \frac{\varepsilon(n; f, \triangle f)^2}{\langle \varepsilon(n; f, \triangle f)^2 \rangle} \right\rangle, \tag{12}$$

where $\langle \cdot \rangle$ indicates the averaged value and $\varepsilon(n; f, \triangle f)^2$ is the instantaneous flow of energy of the SE in the respective band. Similarly, the SES Infogram values are calculated by

$$I_E(f; \triangle f) = -H_E(f; \triangle f) = \left\langle \frac{E(n; f, \triangle f)^2}{\langle E(n; f, \triangle f)^2 \rangle} \ln \frac{E(n; f, \triangle f)^2}{\langle E(n; f, \triangle f)^2 \rangle} \right\rangle \tag{13}$$

with $E(n; f, \triangle f)^2$ representing the instantaneous flow of energy of the SES in the respective band. It was found that the SE Infogram and the Kurtogram produce similar results. Both are also found the be very sensitive to impulsive noise. However, the SES Infogram was found to be insensitive to random impulses and could easily detect series of transients. Another known downfall of the Kurtogram is that if the rate of repetition of transients is too high, the kurtosis vanishes. The same holds true for the SE Infogram where a very high repetition rate leads to failure in detecting transients. The SES Infogram does not have this limitation and successfully identifies the transients. Even though the SE Infogram is very similar to the Kurtogram, the added benefit of the SE Infogram is the fact that it is possible to systematically add the SE Infogram to the SES Infogram to form the Average Infogram, which can identify multiple frequency bands of interest.

2.13.5 EHNRogram

Xu et al. (2016) proposed a novel feature named envelope harmonic-to-noise ratio (EHNR) for periodic impulse detection in the frequency domain. The envelope is calculated via the Hilbert transform in order to separate the carrier from the modulating component. The autocorrelation of the envelope signal is calculated to find the position of maximum autocorrelation, (AC), in the lag domain. The EHNR values are calculated according to

$$\text{EHNR} = \frac{r_{Env_x}(\tau_{max})}{r_{Env_x}(0) - r_{Env_x}(\tau_{max})}, \tag{14}$$

where r_{Env_x} is the AC envelope spectrum which is maximized by the lag, τ_{max}. $r_{Env_x}(0)$ is the total energy of the AC envelope signal.

It was found that under conditions where there is large random bearing slip, the harmonics decrease and thus so does the EHNR value. Standard metrics such as kurtosis or RMS were not found to exhibit the same behaviour with an increase in bearing slip. However, the EHNRogram is robust against random impulses that show no periodicity. Random knock or even electricity interference have been shown to falsely trigger alarm thresholds for kurtosis based methods. With both analytical models and experimental signals, EHNR was shown to be positively correlated with damage growth, i.e. crack progression could be monitored. Xu et al. (2016) recommend that EHNR can not be applied directly on fluctuating speed signals, but with order tracking techniques, EHNR should be able to identify faults in the angle domain.

2.13.6 Autogram

High levels of second-order cyclostationarity are often exhibited by vibration signals from rolling element bearings with the presence of a localized fault. The Autogram, developed by Moshrefzadeh & Fasana (2018), takes advantage of this property as an improvement on the popular Fast Kurtogram. Instead of calculating the kurtosis of the filtered time signal, the Autogram calculates the kurtosis of the unbiased autocorrelation of the squared envelope of the demodulated signal. The unbiased autocorrelation is computed on the squared envelope by

$$R_{xx}(\tau) = \frac{1}{N-q} \sum_{i=1}^{N-q} X(t_i)X(t_i+\tau), \tag{15}$$

where X is the squared envelope of the signal, τ is the lag and $q = 0, ..., N-1$. Calculation of the autocorrelation function has the added benefit of removing uncorrelated components from the signal such as noise and random impulses unrelated to the bearing condition. From experimental tests, Moshrefzadeh & Fasana (2018) found positive results using the Autogram on the *Case Western Reserve University Bearing Data Center* (2011) and could classify bearing damage in the presence of non-Gaussian noisy data.

2.13.7 Grams conclusion

Various other Grams have also been proposed such as the Protrugram by Barszcz & JabŁoński (2011). It is different from the Fast Kurtogram because it is based on the kurtosis of the envelope spectrum amplitudes of the demodulated signal, rather than the kurtosis of the filtered time signal. The benefit of the Protrugram is that it has an improved ability to detect transients with smaller SNR but requires a priori knowledge of the sought fault and will therefore not be considered. Li et al. (2016) suggested the multiscale clustered grey infogram to extract repetitive transients but did not find it superior to either the SE or SES Infograms. Figure 6 shows the diagnostic method commonly used when working with one of the previously discussed Grams.

In general, the Grams have mostly been used on constant speed signals with very little information available on their application to fluctuating speed signals. One Gram that has shown promising results under fluctuating speed is the Fast Kurtogram where Wang et al. (2015) used it to find the frequency band with highest kurtosis under variable speed applications.

Each of the previously discussed Grams were designed to overcome specific shortcomings of other popular diagnostic methods. For example, the Fast-Kurtogram was designed as an estimate of the original Kurtogram so that it can be used for online applications due to faster computational time. The Sparsogram was designed to be more robust against random noisy impulses. Each of the Grams has specific applications where they are applicable. This dissertation will use the seven previously discussed Grams on variable speed applications, which is something that has not been discussed in depth and then propose a methodology that will improve their use on fluctuating operating conditions.

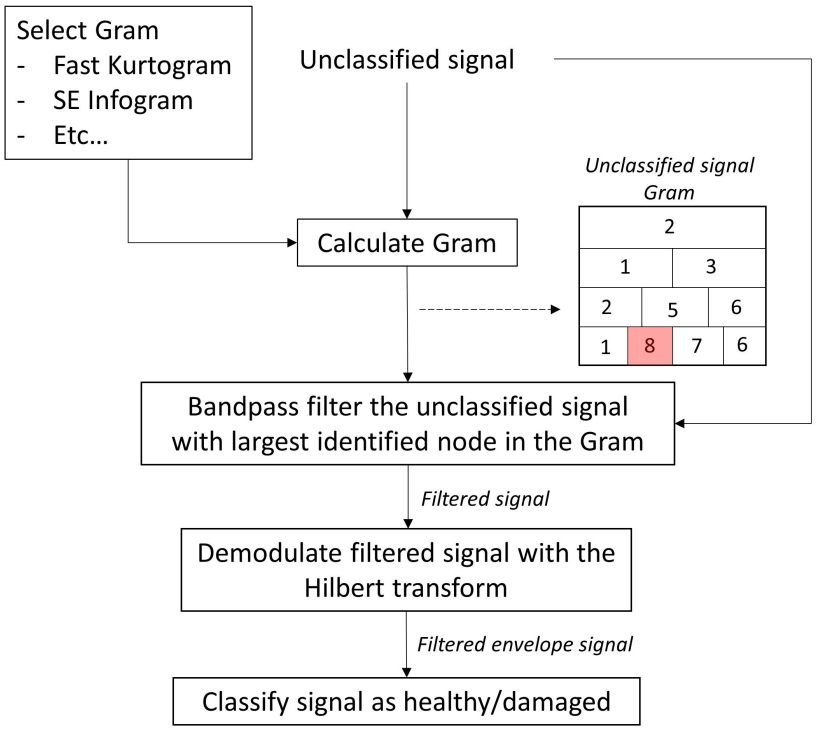


Figure 6: Automatic band selection process using the Grams and the node in the Gram with largest feature value highlighted.

2.14 Model-based methods

To test the effectiveness of the aforementioned diagnostic techniques, test data is needed. Ideally, actual experimental data should be used over a wide range of operating conditions in order to test each diagnostic technique or methodology. However, experimental databases are limited in availability and are costly and time consuming to generate. Thus there is a need to develop and make use of models which closely resemble that of an actual rotating machine vibration signal.

2.14.1 Lump mass models

Lump mass or mathematical models attempt to closely replicate the responses of a gearbox and the interaction between all the components within the system. It is desirable to test diagnostic methods on a range of scenarios such that the effectiveness of these methods can be compared against a wide range of parameters. These may include different types and sizes of bearing and gear damage and variations in speed and load. To develop a real database with a wide range of specific scenarios can be costly. Therefore, mathematical models allow for generating an infinite range of scenarios inexpensively. Omar et al. (2012) developed a nine degree-of-freedom model of a one stage gear system that

considers gear size, errors and faults, and found a good correlation between the simulations and actual experimental data. Feng et al. (2016) used a planetary gear fault signal model to simulate a range of planetary gear failures. Tadina & Boltezar (2011) developed a model for ball bearings during run up. Many different models exist that are capable of accurately modelling bearing and gear faults but these models are very complex and require many inputs such as the stiffness matrices of components and subcomponents, natural frequencies and a range of complicating assumptions. A much simpler approach to avoid such complicated simulations is to make use of phenomenological models.

2.14.2 Phenomenological models

Phenomenological models approximate complicated physical phenomena by describing simpler empirical relationships between these phenomena. These models are not derived from first principles, however they are consistent with fundamental theory in that the response is modelled from empirical evidence. McFadden & Smith (1984) developed a model incorporating bearing geometry, shaft speed, bearing load distribution, transfer function between the bearing and the accelerometer, and the exponential decay of vibration for a REB under constant radial load with a single point defect on its inner race. Abboud et al. (2017) presented a phenomenological model that also incorporates non-stationary conditions as well as gear and bearing information. The complete vibration signal that is typically recorded by an accelerometer can be modeled as the sum of all the contributions,

$$x_c(t) = x_{dg}(t) + x_{rg}(t) + x_b(t) + x_n(t), \tag{16}$$

where $x_d(t)$ is the deterministic gear component typically related to misalignments, gear vibrations and unbalance. $x_r(t)$ is the random and repetitive gear component as a result of factors like wear, friction forces, impacting forces and combustion forces. Finally, $x_b(t)$ is the bearing component caused by bearing faults and $x_n(t)$ the noise which is caused by factors such as environmental noise, sensor errors and other machinery.

Phenomenological models can very accurately simulate a vibration signal similar to what might be expected in industry (Parra & Vicuña 2017). D'Elia et al. (2018) developed a detailed bearing model that simulates a faulty bearing under constant and variable speed conditions with a range of different fault types. The benefit of such models is that many different scenarios can be analysed inexpensively which allow for a wide variety of techniques and methodologies to be rigorously tested under different operating conditions. Another benefit is that the exact conditions being tested are known as inputs to the problem, thus results can easily be validated against the expected outcomes. The model developed by Abboud et al. (2017) shall serve as starting point for vibration analysis tool comparisons in subsequent sections. It must be noted that simply because a methodology is applicable to one of these models, it does not imply that the methodology will work on real signals. Rather, results based on such models serve as a preliminary test to discover the potential of suggested methodologies and techniques.

3 Gearbox model and experimental setup

To design and compare a new diagnostic methodology to popular benchmark methods, an analytical model as well as real vibration data is used. Due to the uncertainties pertaining to the development of a new methodology as well as the infinite number of possible operating conditions, performing tests blindly with the hope of finding and improved method is naive. Creating a database of real vibration signals under such a wide variety of conditions is both costly and time consuming. A logical approach is to investigate new methods on controlled data sets where the exact condition and characteristics of the gearbox are known and where simulations are performed rapidly.

The chosen analytical model in this dissertation is the phenomenological method developed by Abboud et al. (2017). Phenomenological models describe observed physical phenomena and the relationships between them. The phenomenological model will be used under constant and varying speed conditions to validate the findings of the proposed methodology in detecting and locating damage. Furthermore, an experimental setup will be used to confirm the abilities of the methodology on real vibration signals.

3.1 Phenomenological model

The following work describes the phenomenological model developed by Abboud et al. (2017) and includes the model parameters, explanations and notation as described by Schmidt et al. (2019*a*). This model will be used with different operating speed profiles to highlight the strengths and shortcomings of common vibration signal analysis techniques under a wide range of operating conditions and fault parameters. It will also be used to investigate the methodology developed in this dissertation. Equation (16) is repeated below for convenience,

$$x_c(t) = x_{dg}(t) + x_{rg}(t) + x_b(t) + x_n(t). \tag{16}$$

where $x_{dg}(t)$ is the deterministic gear component typically related to misalignments, gear vibrations and unbalance. $x_{rg}(t)$ is the random and repetitive gear component as a result of wear, friction forces, impacting forces and combustion forces. $x_b(t)$ is the bearing component caused by bearing faults and $x_n(t)$ the noise which is caused by environmental noise and sensor errors.

The first three components of Equation (16) can be decomposed as follows:

$$x_{dg}(t) = h_{dg}(t) \otimes q_{dg}(t), \tag{17}$$

$$x_{rg}(t) = h_{rg}(t) \otimes q_{rg}(t), \tag{18}$$

$$x_b(t) = h_b(t) \otimes q_b(t), \tag{19}$$

where $h_i(t)$ is the impulse response function that physically expresses the transmission path from the signal source $q_i(t)$ to the sensor, and $\otimes$ is the convolution operator. The

impulse response function is written as a viscously damped single degree-of-freedom system,

$$h_i(t) = e^{-\zeta_i 2\pi f_{n,i} t} \sin\left(2\pi\sqrt{1-\zeta^2} f_{n,i} t\right), \tag{20}$$

where the damping ratio and natural frequency of component i are denoted by ζ_i and $f_{n,i}$ respectively. The signal sources of the deterministic and random components are described as follows,

$$q_{dg}(t) = M_{dg}(\omega_r(t)) \sum_{k=1}^{N_{dg}} A_{dg}^{(k)} \cos\left(kN_{t,g} \int_0^t \omega_r(\tau) d\tau + \phi_{dg}^{(k)}\right), \tag{21}$$

$$q_{rg}(t) = M_{rg}(\omega_r(t)) \varepsilon_{rg}(t) \sum_{k=1}^{N_{rg}} A_{rg}^{(k)} \cos\left(k \int_0^t \omega_r(\tau) d\tau + \phi_{rg}^{(k)}\right). \tag{22}$$

N_{dg} and N_{rg} are the number of gear mesh and random components considered respectively. $A_i^{(k)}$ and $\phi_i^{(k)}$ are the amplitude and phase of the kth component, and $N_{t,g}$ is the number of gear teeth. $M_i(\omega_r(t))$ is a monotonic function simulating the dependence of the magnitude of the signal on the rotational speed, $\omega_r(t)$, and has the following form:

$$M_i(\omega_r(t)) = a\omega_r(t) + b. \tag{23}$$

As seen from Equations (21) and (22) they contain similar components except for the number of gear teeth, $N_{t,g}$, which is only present in the deterministic signal and $\varepsilon_{rg}(t)$, a random function, which only contributes to the amplitude of the random component,

$$\varepsilon_{rg}(t) \sim \mathcal{N}(0, \sigma_{rg}^2). \tag{24}$$

$\mathcal{N}(0, \sigma_{rg}^2)$ denotes a zero-mean Gaussian distribution with variance σ_{rg}^2. The noise component is written as

$$x_n(t) = \varepsilon_n(t) M_n(\omega_r(t)), \tag{25}$$

with monotonic function, $M_n(\omega_r(t))$, and ε_n being a zero-mean Gaussian distribution with variance σ_n^2,

$$\varepsilon_n(t) \sim \mathcal{N}(0, \sigma_n^2). \tag{26}$$

Lastly, bearing damage is simulated as a train of Dirac functions (Schmidt et al. 2019*a*),

$$q_b(t) = F_{dam}.F_{const}.M_b(t) \sum_{i=1}^{N_\tau} \delta(t - \tau_i). \tag{27}$$

To simulate a gearbox with a healthy bearing, the contribution of the bearing to the casing signal, $x_c(t)$, must simply be modelled as $q_b(t) = 0$. The exact model parameters used in this dissertation are presented in Appendix A.

There are many different methods that can be used to scale the individual components of the phenomenological signal. The RMS, spectral energy, crest factor and many other metrics can be used to normalize and scale the components. In this dissertation each of the four signals from the phenomenological model, $[x_{dg}, x_{rg}, x_b, x_n]$, are normalized so that the peak amplitude of each is unity. The signals are then multiplied by a scaling constant in the time domain and summed to produce the vibration case signal, $x_c(t)$. If the scaling constants are $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [1.5, 0, 10, 2.3]$ this implies that the deterministic component has a maximum amplitude of 1.5, the random component has a maximum amplitude of 0, the noise dominates the signal with an amplitude of 10 and the bearing signal has a maximum amplitude of 2.3. The signal with these specific components is shown in Figure 7 and the corresponding fast Fourier transform (FFT) in Figure 8. First the components are scaled and added resulting in Figure 7 which is the vibration signal in the time domain. This signal can then be order tracked and the envelope signal FFT is shown in Figure 8 where clearly the deterministic component at order 40 dominates the spectrum similar to real vibration signals.

The operating speed profile is defined as Equation (28).

$$\omega = 2\pi \left(5\sin(0.4\pi \times t) - 20(0.1 \times t - 0.5)^2 + 15\right) \quad \text{rad/s}. \tag{28}$$

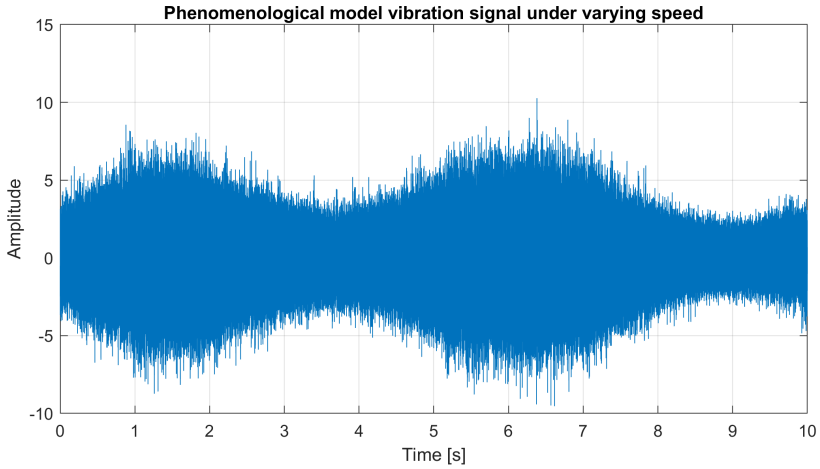


Figure 7: Vibration signal from the phenomenological model with component contributions $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [1.5, 0, 10, 2.3]$.

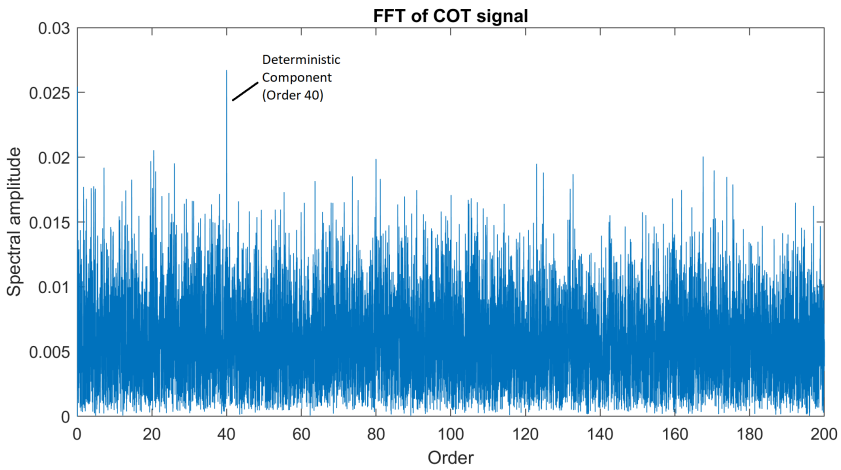


Figure 8: Order tracked signal FFT with components $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [1.5, 0, 10, 2.3]$.

3.2 Experimental setup

Real vibration signals from a helical gearbox setup was used to test and confirm the recommended diagnostic methodology on stationary and varying speed/load conditions. The setup is based on the design of Stander & Heyns (2005). The main components consisting of one electrical motor, three consecutive gearboxes and one alternator are shown in Figure 9.

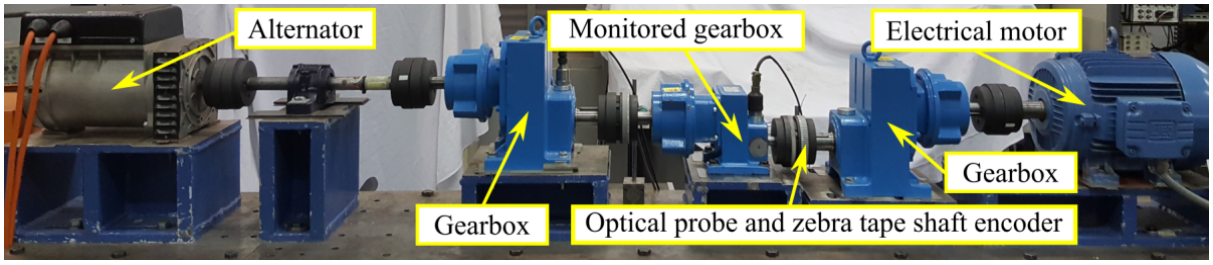


Figure 9: The helical gearbox setup used for capturing experimental datasets (Schmidt et al. 2019*b*).

A detailed schematic of this experimental setup is shown in Figure 10. The system is driven by the electrical motor followed by a speed reduction gearbox (GB 1 in Figure 10). The monitored gearbox, GB 2, has an input reference shaft with an optical probe and zebra tape shaft encoder which generates 88 pulses per revolution. The monitored gearbox, GB 2, then steps up the speed with a gear ratio of 1.85. The third gearbox, GB 3, further steps up the speed with a gear ratio of 4.93 such that the alternator shaft spins 9.12 times faster than the reference input shaft of the monitored gearbox. The electrical motor and alternator apply a varying speed and load to the system respectively.

The data is recorded using an eight channel OROS (OR 35) data acquisition device to capture the five vibration signals (three from the triaxial accelerometer and two uni-axial accelerometers) as well as the two tachometer signals. The sampling rate of each

recorded signal is also shown in Figure 10. To accurately capture the rotational speed of the system, the highest possible frequency of 51.2 kHz is used for the zebra tape shaft encoder (ZTSE) and the proximity probe, such that the order tracking can be performed successfully. The ZTSE is located on the input reference shaft of the monitored gearbox and the proximity probe is situated on the output shaft key of the monitored gearbox and thus has a once per revolution pulse. Furthermore, the ZTSE signal needs to be geometrically compensated using the method developed by Diamond et al. (2016).

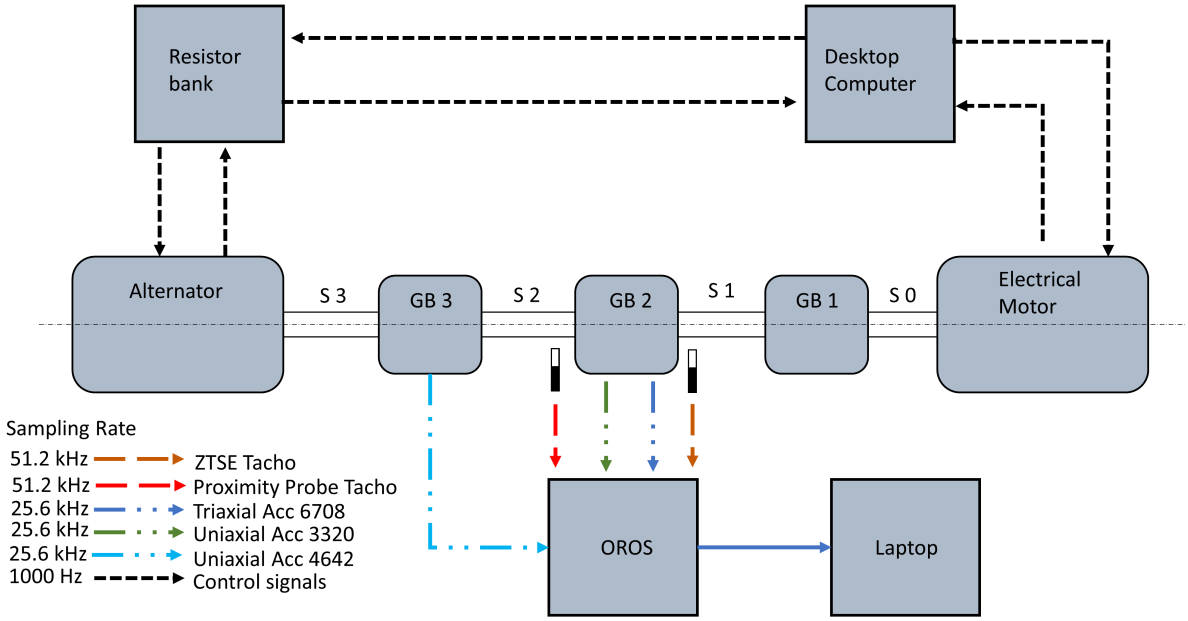
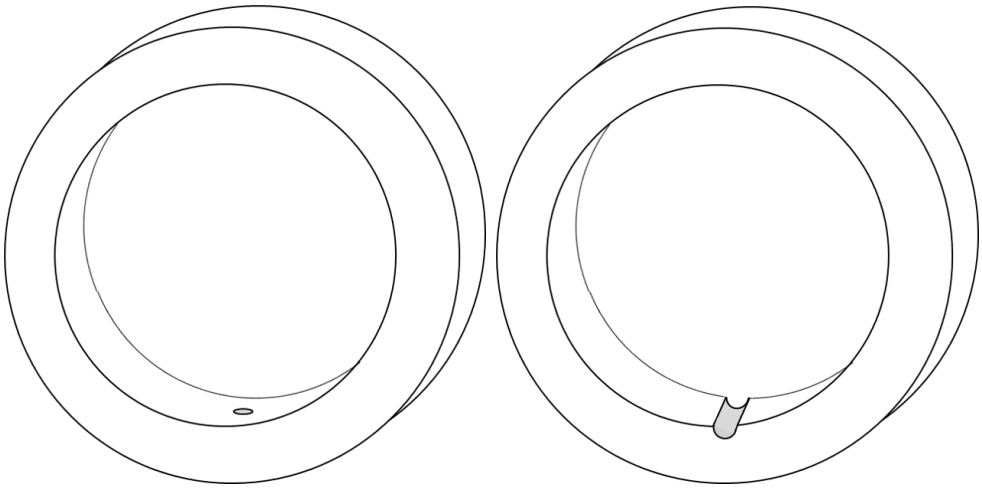


Figure 10: Experimental setup schematic.

The triaxial accelerometer has a sensitivity of 100mV/g and in this study, the axial direction vibration will be used because it was found to be most sensitive at detecting damage. The two uni-axial accelerometers (vertical direction) both have a sensitivity of 500mV/g and can be used for further investigations. They offer redundancy to verify results if required. The five accelerometer signals were measured at a rate of 25.6 kHz. The desktop computer which controls the operating conditions (speed and load) is part of a feedback loop involving the alternator and motor with a frequency of 1000 Hz.

The generated database using the above experimental setup consists of 350 healthy data recordings followed by 6 levels of discrete damage, each containing 210 separate recordings spaced at intervals of 10-15 minutes. Damage was induced artificially using a 2mm diamond tip Dremel drill bit attached to a flexible shaft. It is desirable to disturb the system as little as possible during each level of induced damage. Therefore, each time that damage was induced, the gearbox was never opened or removed from the setup as shown in Figure 9. Therefore, the system was minimally disturbed. Instead, damage was induced by removing the oil seal on the end of the monitored gearbox which allowed sufficient access to the bearing of interest. A small fault was then introduced on the outer race at the maximum load zone, with the position indicated in Figure 11a, using the

Dremel setup by drilling softly for 5-10 seconds. Afterwards, a new oil seal was installed and data recorded for 210 readings spread over 2 days. This process was repeated for damage levels 1-5. After damage level 5, the gearbox was opened and the bearing removed so that the damage could be quantified. In Figure 12, the 5th level of damage can be seen and measured to reveal a 1mm radius hole with a depth of approximately 0.5mm. Unfortunately due to the limited space, damage levels 1-4 can not be quantified, however, the damage is progressively becoming worse from the healthy state to damage level 5.



(a) D1-D5 Fault sketch. (b) D6 Fault sketch

Figure 11: Fault shape estimates after inspection at damage levels 5 and 6.

Figure 12: Visible damage on OR after damage level 5.

It was also desired to have a dataset with very large damage to use as a benchmark for comparing methods. Under this relatively large damage condition, it was hoped that all methods should be able to detect the damage. A slit was cut into the bearing using the same Dremel setup as described above. The final damage, level 6, measures 2mm wide,

1mm deep and goes across the full width of the outer race and can be seen in Figure 13. At damage level 6, a further 210 recordings were made over 2 days such that the total database from healthy to damage level 6 consists of 1610 recordings.

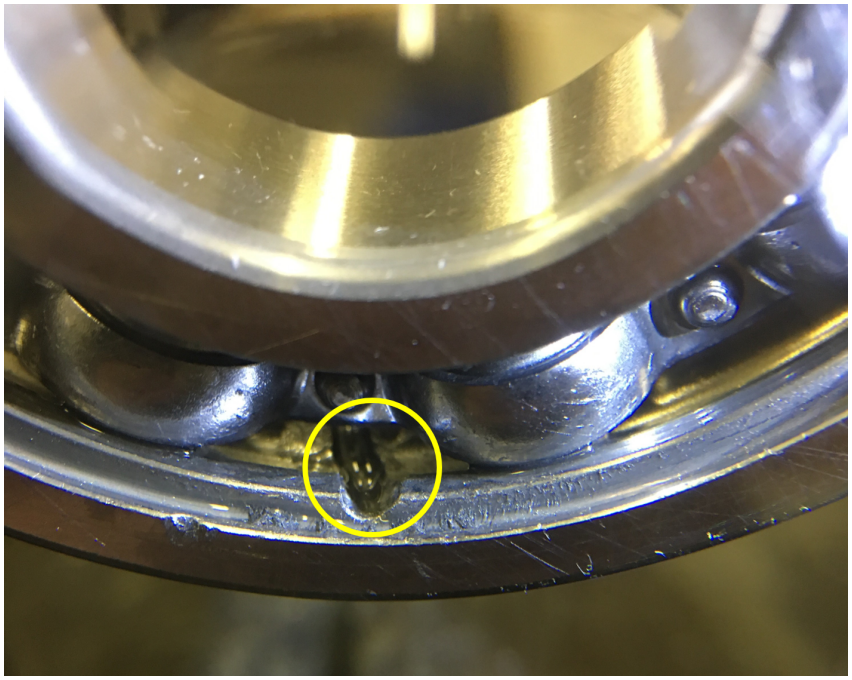


Figure 13: Visible damage on OR after damage level 6.

Under constant speed and load conditions, there are many different diagnostic techniques available. However, since the aim of this dissertation is to investigate bearing diagnosis under both varying speed and load conditions, it is required to have time varying conditions. At each bearing health condition, Healthy - damage 6, seven operating conditions were used. The seven operating profiles can be seen in Figure 14. Operating conditions 1-4 vary with time for both the speed and applied torque and operating conditions 5-7 have constant speed and torque but at different levels. The first four operating conditions are based on work done by Schmidt & Heyns (2020) which are sufficiently complex to critically evaluate the performance of different bearing diagnostic methodologies under varying speed and load conditions. Note that the colour of the speed profile lines in Figure 14 corresponds with the respective operating torque conditions lines. In this dissertation, only the varying speed signals are used to compare and validate each of the methods.

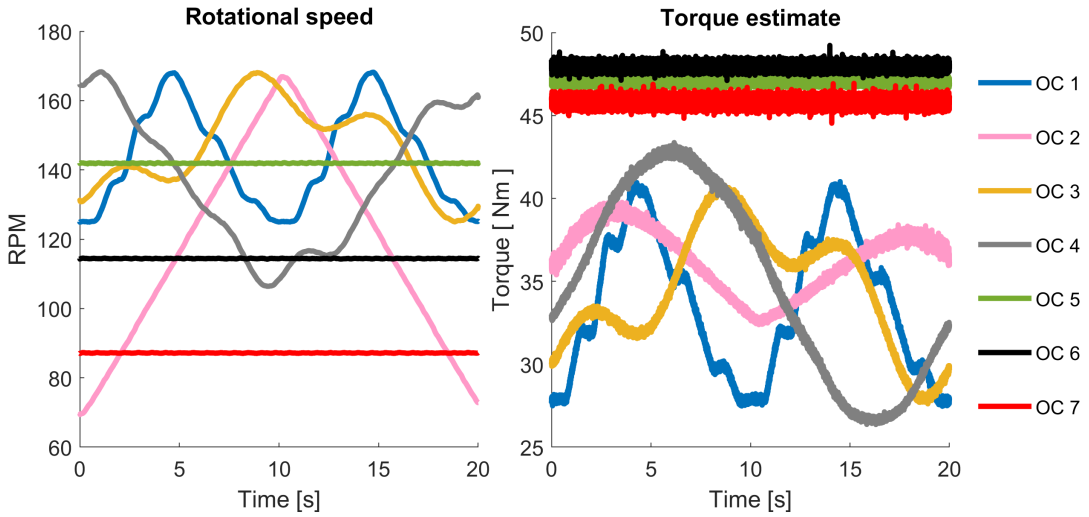


Figure 14: Operating conditions.

Table 1: Details for each level of damage in the experimental database.

Dataset	Recordings	Damage type	Damage size estimate
Healthy	350	-	0
Damage 1	210	Pit	*
Damage 2	210	Pit	*
Damage 3	210	Pit	*
Damage 4	210	Pit	*
Damage 5	210	Pit	1 mm radius, 0.5 mm deep
Damage 6	210	Slit	2 mm wide, 1 mm deep

* Damage size unknown due to not opening the gearbox for fear of disturbing gearbox properties.

A summary of the fault conditions and experimental dataset can be seen in Table 1. Note that care was taken not to disturb the gearbox between each consecutive damage increment. Only the oil seal nearest the bearing was removed so that the Dremel setup could be used to damage the bearing, but the gap was too small to allow for measuring of the exact damage size. After the tests were completed at damage level 5, the gearbox was opened and the bearing removed. With the bearing out of the gearbox, the damage could be measured using a vernier.

4 Investigative study

One of the main purposes of this dissertation is to highlight where different vibration analysis methodologies are applicable and where they fail. Section 4 starts by discussing the available databases that can be used and why it is needed to implement a phenomenological model. The author then proceeds with comparing different pre-whitening techniques in a logical manner on a range of signals that specifically highlight the abilities and limitations of each method. The main idea behind the order of the work is to start with popular techniques that are often used and then highlight the need for a new methodology. The work then moves towards proposing and validating a new method that can deal with the fluctuating operating conditions. The following techniques and methodologies are all discussed:

- Cepstrum pre-whitening
- Autoregressive models
- Grams
    - Fast Kurtogram
    - SE Infogram
    - SES Infogram
    - Sparsogram
    - Enhanced Kurtogram
    - EHNRogram
    - Autogram
- New proposed methodology

Finally, the proposed methodology is evaluated to highlight its superiority at detecting faults with higher classification accuracy and at an earlier stage for fluctuating speed conditions.

4.1 Vibration databases

Actual experimental vibration databases under a wide range of operating conditions would be the ideal scenario in which to test all of the methodologies to be discussed in this dissertation. Two common vibration databases which are currently available for analysis on bearings are:

1. Case Western Reserve University bearing data centre

    The Case Western database relates to bearing failures and contains data sampled at 12,000 and 48,000 Hz depending on the specific data file. Data files containing faults on the IR, Ball and OR are provided. The OR damage files have three

subsets, namely centred (6:00), orthogonal (3:00) and opposite (12:00) relative to the load zone. The faults were artificially induced using electro-discharge machining and so the exact dimensions of the faults could be controlled. The fault diameters available are 0.007", 0.014", 0.021" and 0.028". The motor load also ranges from 0,1,2 and 3 HP for each fault diameter and the approximate motor speed is also provided. For further information about the data and the apparatus and procedures, information is provided on their website, *Case Western Reserve University Bearing Data Center* (2011). Smith & Randall (2015) performed a benchmark study on this database which has become a standard reference for new methods.

2. Centre for Intelligent Maintenance Systems (IMS)

   This database contains bearing failure information and was generated by running the bearings until failure. A radial load of 6,000 lbs at 2000 rpm was applied and all failures occurred after exceeding the designed lifetime of the bearings. Since the damage was not artificially induced, the exact fault sizes are unknown for each data set but the type of fault is specified. 3 data sets are provided which describe a test-to-failure experiment. The individual files are 1 second vibration signal snapshots recorded at specific intervals with a sampling frequency of 20,480 Hz. For further information about the data, apparatus and experimental method, see (Lee et al. 2007).

Unfortunately, both of the bearing databases mentioned above operate under constant load and constant speed conditions. To the best of the author's knowledge, no database exists or is freely available containing damaged bearing signals under fluctuating operating conditions. Thus, the phenomenological model is used to test suggested methods on fluctuating speed profiles followed by an in-depth statistical analysis of the proposed methodologies on an actual experimental gearbox conducted by the author.

4.2 Constant vs fluctuating speed signals

The properties of a vibration signal are inherently linked to the operating conditions under which the signal was produced. Abboud et al. (2019) claim that a vibration signal comprises of two components, angular components and time-invariant components. The angular components are those linked to the actual system kinematics and are related to the machine rotational speed. If the machine speed increases, so does the repetition rate of the angular components since they are constant with respect to the angle variable and have a constant angular period. Consequently, these components are best analysed in the angle domain through popular methods such as order tracking. The time-invariant components arise from different phenomena all together. As time and speed varies, the time-invariant properties of the signal remain constant. A faulty bearing with outer race damage under fluctuating speed is a typical example highlighting both sets of components. Such vibration signals consist of the cyclic impacts when the rolling elements hit the fault which are locked to the machine rotating angle and also the excited structural resonances. Therefore, a temporal analysis of the signal will identify and emphasize the transients and excited resonance frequencies. An angular analysis would conserve angle periodicity but lose transient consistency.

4.3 Removing deterministic components from a vibration signal

It is common for a vibration signal to be dominated by deterministic components such as gear vibrations, unbalances and misalignment especially when the SNR of the bearing vibration signal is low. Thus, there exists the need to remove or attenuate the deterministic components from a vibration signal relative to the bearing signal. Common methods include Synchronous averaging, FFT filters, Cepstrum pre-whitening (CPW), Linear prediction and Autoregressive (AR) models (Antoni 2016). Synchronous averaging can easily be applied on constant speed signals and similarly on computed order tracked signals. However, synchronous averaging is most effective when a vibration signal consists of many shaft revolutions. Antoni & Randall (2006) recommend using the spectral kurtosis as a filter to maximise the output SNR. However, a high value of spectral kurtosis requires that the individual transients are separated, which in turn means that if their repetition rate is high, their damping must be sufficiently high that each dies away before the appearance of the next (Sawalhi et al. 2007). Recently, promising results have been obtained through the use of CPW and AR models.

4.3.1 Cepstrum pre-whitening

CPW is a blind preprocessing tool with remarkable capabilities of removing both harmonics and modulation side-bands of unwanted components. Deterministic excitations under constant speed are periodic but not sinusoidal and produce multiple peaks at harmonics of the first excitation frequency. Therefore, in the frequency domain, equally spaced peaks will be present. In the cepstral domain there will be a peak at a quefrency equal to the period between the harmonics. Removal of these deterministic components is done by liftering around the quefrencies of the deterministic components, and is calculated by

$$x_{cpw} = IFT\left\{\frac{FT(x)}{|FT(x)|}\right\}, \tag{29}$$

where FT is the Fourier transform of the signal, $|\cdot|$ represents the absolute value, and IFT is the inverse Fourier transform. Bearing damage components are second-order cyclostationary and their mean is not strictly periodic. Hence, the bearing components will not contain strong peaks in the cepstral domain and will remain relatively unchanged by the liftering process (Borghesani, Pennacchi, Randall, Sawalhi & Ricci 2013). CPW thus enhances the bearing component relative to the deterministic component.

Deterministic components need to be removed because they often dominate the vibration signal, making bearing fault diagnosis difficult. Under constant speed applications, the CPW process is very simple because the raw signal is simply passed through Equation (29), resulting in a pre-whitened signal as desired. For explanation purposes, Figure 15 explains the process of calculating the pre-whitened squared envelope spectrum from the raw constant speed signal and demonstrated in Figure 16. The signal used for the example in Figure 16 is taken from a constant speed recording at damage level 6 from the experimental gearbox setup.

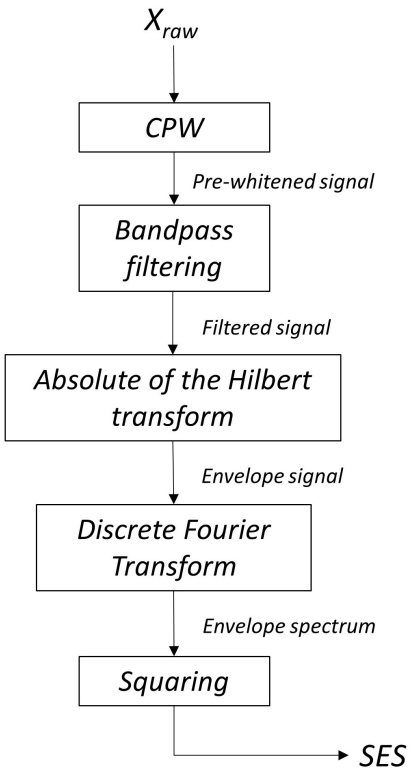


Figure 15: CPW process for constant speed applications.

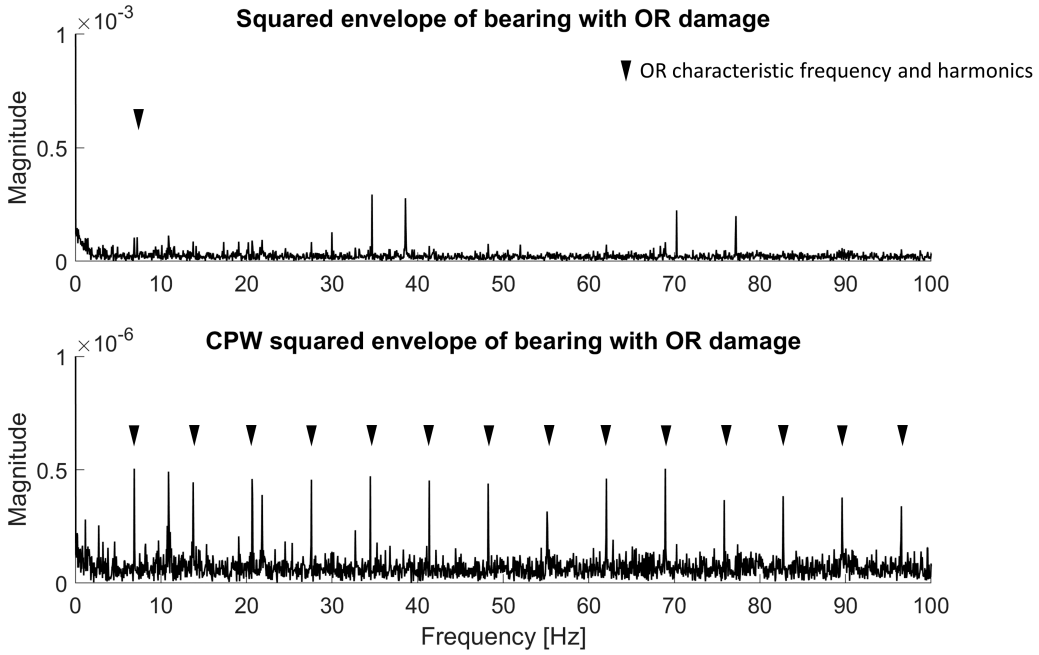


Figure 16: CPW applied to a real constant speed signal from the experimental gearbox setup where the bearing characteristic frequency content has been improved relative to the deterministic content and noise.

Under varying speed conditions the process is slightly more involved and Borghesani, Ricci, Chatterton & Pennacchi (2013) recommend pre-whitening in the order domain according to their proposed method in Figure 17. Fluctuating speeds cause the peak frequencies to be smeared and will thus not present clear harmonics. Deterministic components are angle-periodic and by working in the angle domain, clear peaks will be seen at the deterministic component characteristic orders which results in a signal that looks similar to a constant speed signal. Therefore, the signal must be order tracked and the signal then prewhitened in the angle domain. Once the deterministic component is removed in the angle domain, it is required to return to the time domain for the bandpass filtering process according to Figure 17.

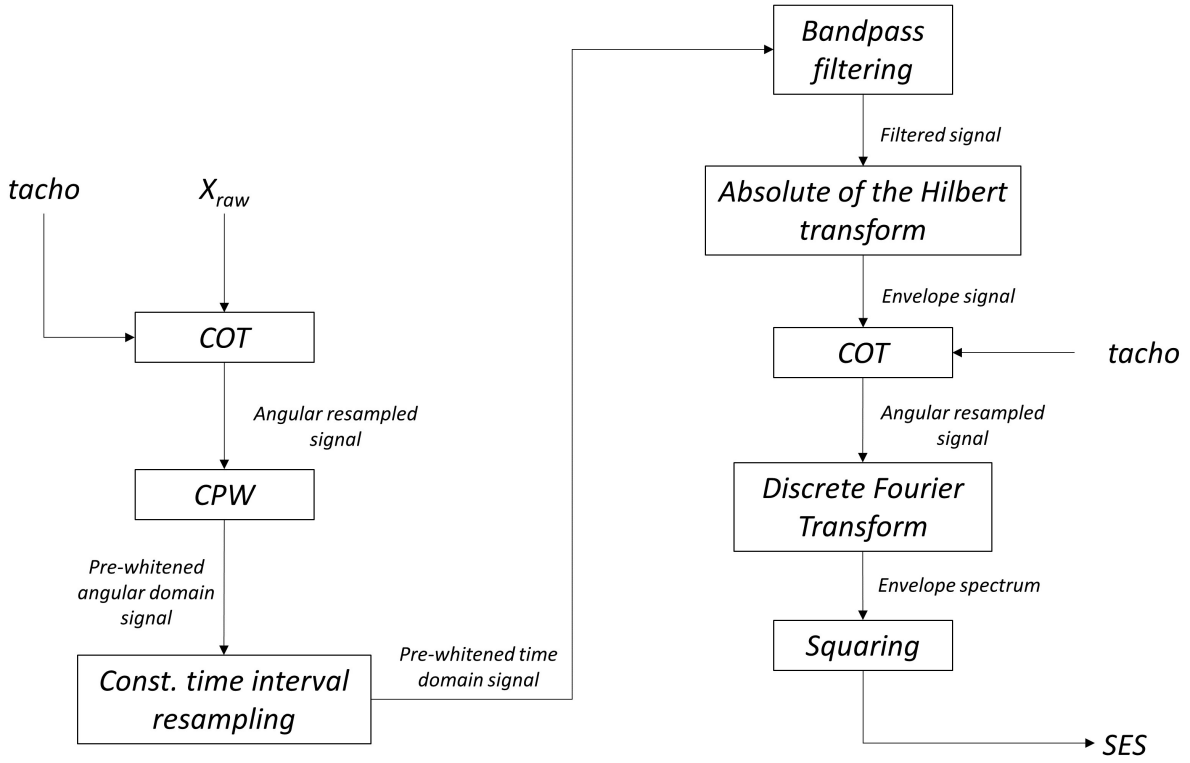


Figure 17: CPW process for varying speed applications adapted from Borghesani, Ricci, Chatterton & Pennacchi (2013).

In order to demonstrate the effectiveness of CPW on variable speed signals, an analytical signal based on the phenomenological model by Abboud et al. (2017) is used with the bearing characteristic order set as 13 times per revolution. The author proposes that the characteristic-frequency signal-to-noise ratio (CF SNR) will be used as a detection metric which either classifies a signal as healthy or damaged based on its value. An example calculation of the CF SNR is shown in Figure 18 with the threshold limit set to the 99th percentile value of the amplitude of the order spectrum. If the CF SNR is smaller than one, it implies that the fault is not identifiable and CF SNR values larger than one indicate a fault that is easily detected.

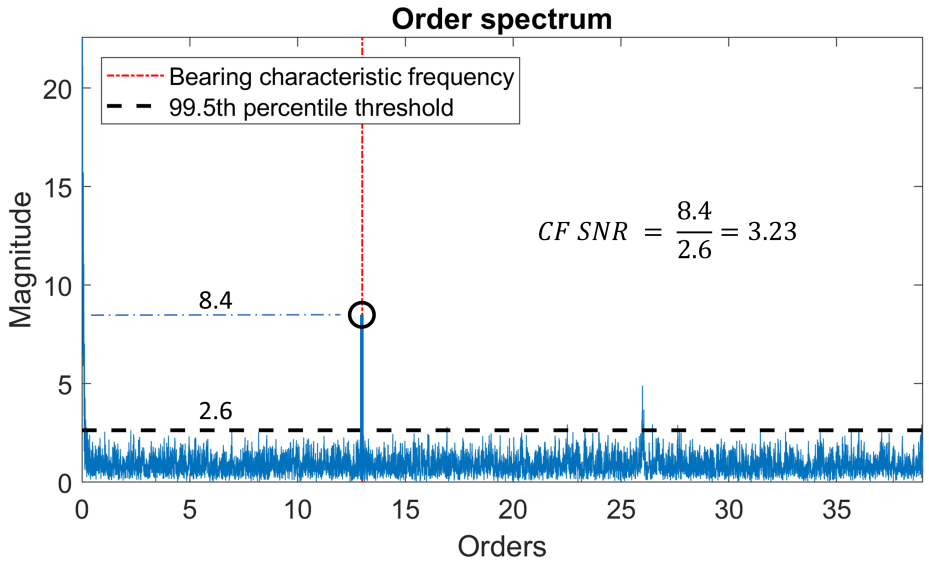


Figure 18: Example showing the bearing CF SNR.

The phenomenological model developed by Abboud et al. (2017) will now be used with the CF SNR performance metric to compare the diagnostic ability of various methods. The focus remains on diagnosing bearing faults under fluctuating speed profiles. Two CPW methods will be compared along with an autoregressive model, all aimed at increasing the bearing fault signature relative to the noise and very strong deterministic components. In the following examples, the CF SNR is determined in the order spectrum as the first harmonic spectral amplitude at the signal of interests' order over the spectral amplitude of the 99th percentile between order 0 and order 200 of the total order spectrum domain. Slip is accounted for by allowing a 1% deviation of the bearing characteristic frequency (Sawalhi & Randall 2008). Therefore, the amplitude value used for the bearing characteristic order is selected as the maximum amplitude in the range $[CF - 0.5\% \times CF,\ CF + 0.5\% \times CF]$.

In the original signal (Figure 20), the deterministic component (with fixed time domain amplitude of 3) dominates the spectrum until the bearing component has an amplitude of 5.5 or larger in the time domain. Thus, when the bearing signal is weak, which is generally the case, the deterministic component dominates the signal spectrum. However, after performing the pre-whitening process as recommended in Figure 17, the bearing component is now larger than the deterministic component for all cases as shown in Figure 21 by the yellow line (bearing) and the purple line (deterministic). The deterministic component has almost been completely removed and the bearing component is now dominant, but the SNR is still very low and thus makes detection of the bearing fault difficult, unless further processing is applied. However, following the process recommended in Figure 17, the pre-whitened signal should be bandpass filtered to remove the noise. The bandpass filter is designed according to the Spectrogram in Figure 19, with the filter limits set to include the gear mesh resonance frequency at 5000 Hz, and also the deterministic resonance frequency of 2000 Hz. Thus, to avoid losing any diagnostic information, the filter is designed slightly higher and lower than the mentioned frequencies and ranges from

1500Hz to 5500 Hz.

Finally, Figure 22 shows the result of filtering the pre-whitened signal between the suggested 1500 Hz and 5500 Hz (which removes the low and high frequency noise) according to the filter limits as explained from Figure 19. The bearing component of the filtered pre-whitened signal is now even more dominant than previously (green line in Figure 22). The following figures are generated by keeping the noise, deterministic and random component amplitudes constant and then incrementally increasing the amplitude of the bearing component.

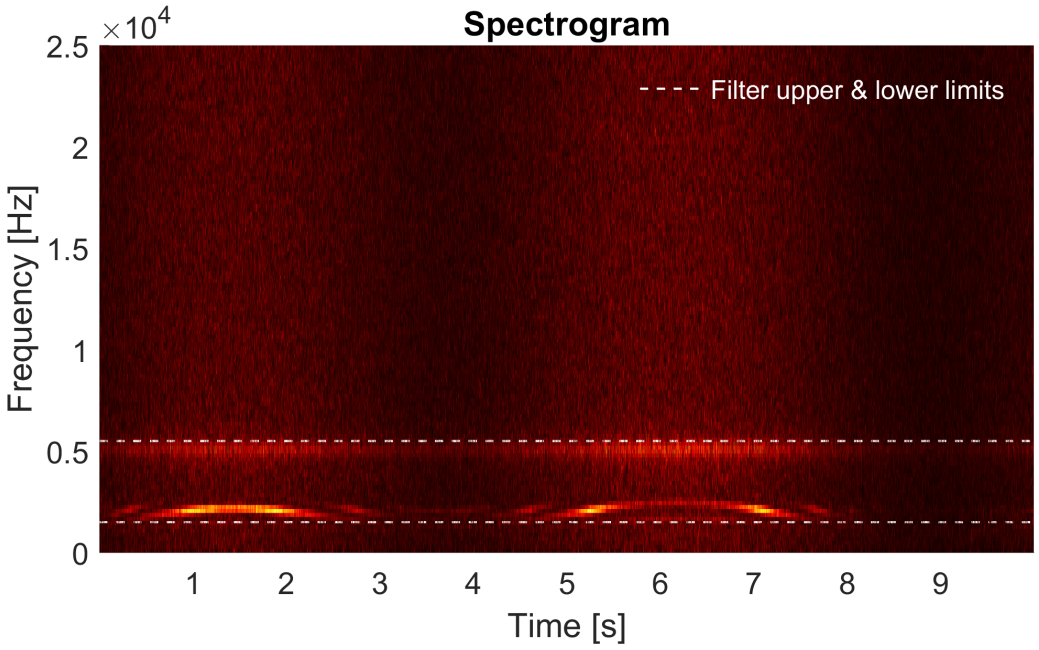


Figure 19: Spectrogram showing the upper and lower bounds of the applied bandpass filter on the fluctuating speed signal from the phenomenological model.

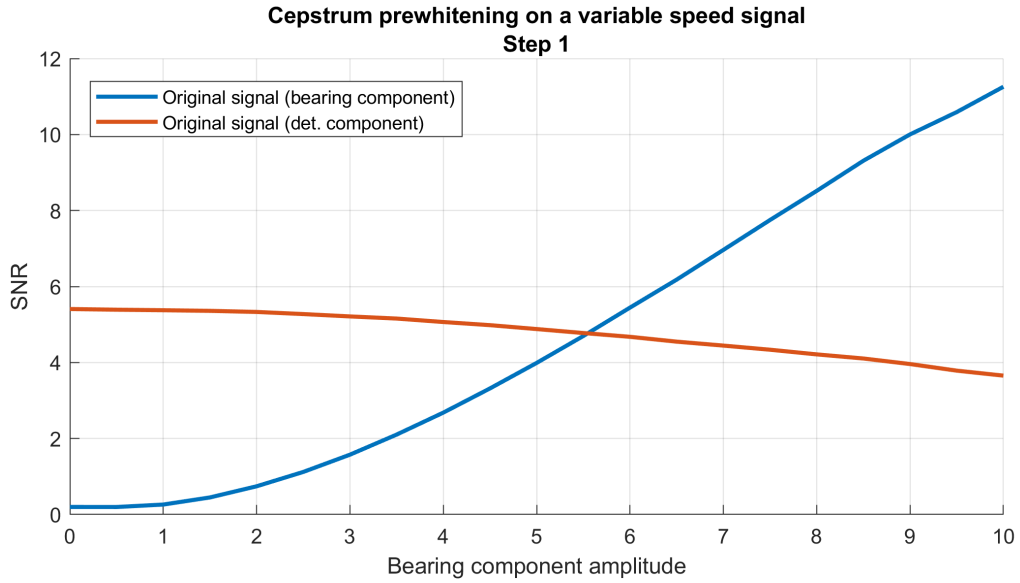


Figure 20: Variable speed phenomenological signal SNR with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

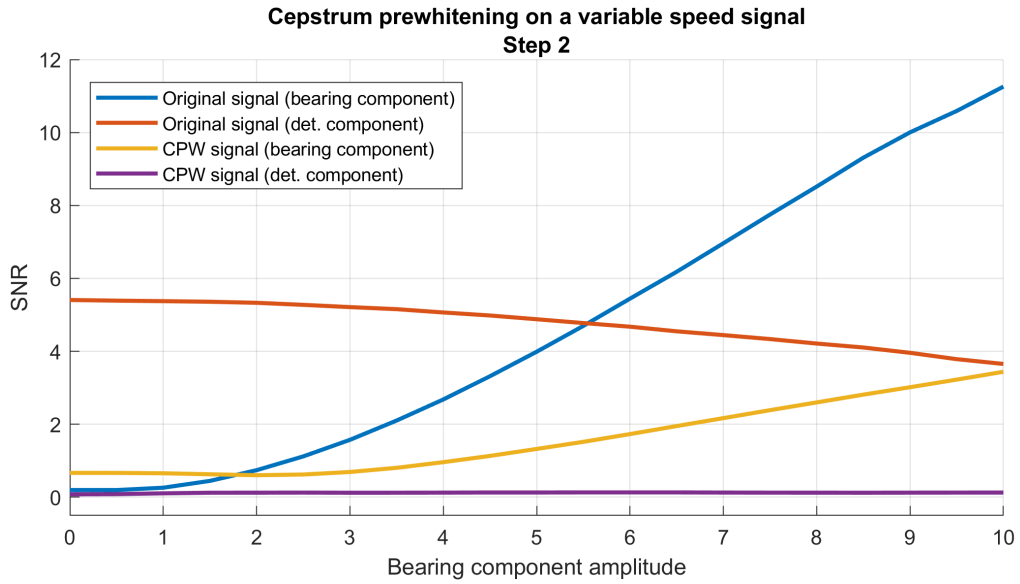


Figure 21: The effect of CPW according to the method proposed in Figure 17 on a variable speed signal with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

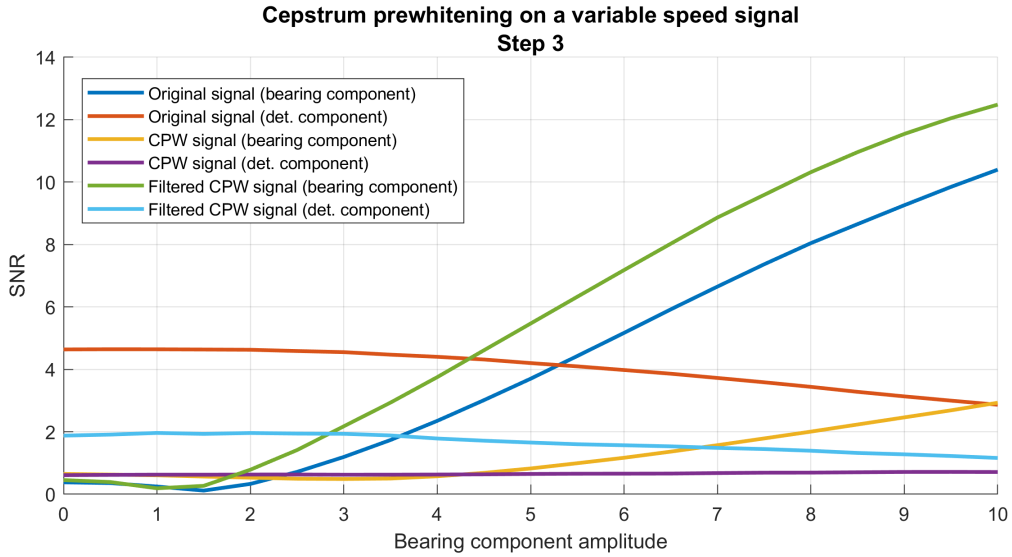


Figure 22: Filtered CPW on a variable speed signal with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

Other work involving CPW follows similar procedures to Figure 23 Method 1, where the time domain signal is first-order tracked and then pre-whitened in the angle domain. Interestingly, Barbini et al. (2018) found that CPW can also be performed in the time domain on variable speed signals prior to the COT operation. The argument presented by Barbini et al. (2018) is based on the fact that resonant frequencies excited by the impacts from a faulty bearing are constant or vary more slowly than that of the machine's rotational speed. Hence, if order tracking is performed based on the machine's rotational speed, the resonant frequencies will be smeared in the angle domain by the resampling. This alternative method for CPW is presented in Figure 23 Method 2.

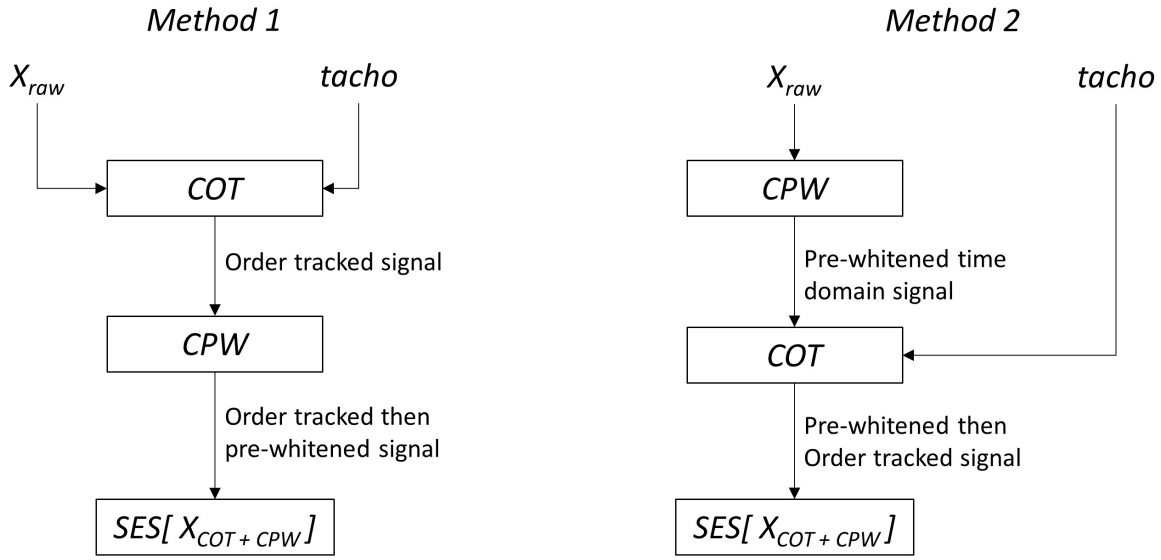


Figure 23: Two possible methods for CPW where Method 1 suggests CPW in the angle domain and Method 2 suggests performing CPW before order tracking.

Using the same phenomenological signals from the previous example (Figures 20 - 22), the method proposed by Barbini et al. (2018) in Figure 23 Method 2 will be compared. In Figure 24, the original SNR values are shown of the unfiltered, un-whitened bearing and deterministic components. In Figure 25, the signal was first cepstrum pre-whitened and only afterwards was it order tracked. One can notice that the deterministic component takes on a constant SNR value of approximately just under 2 (purple line).

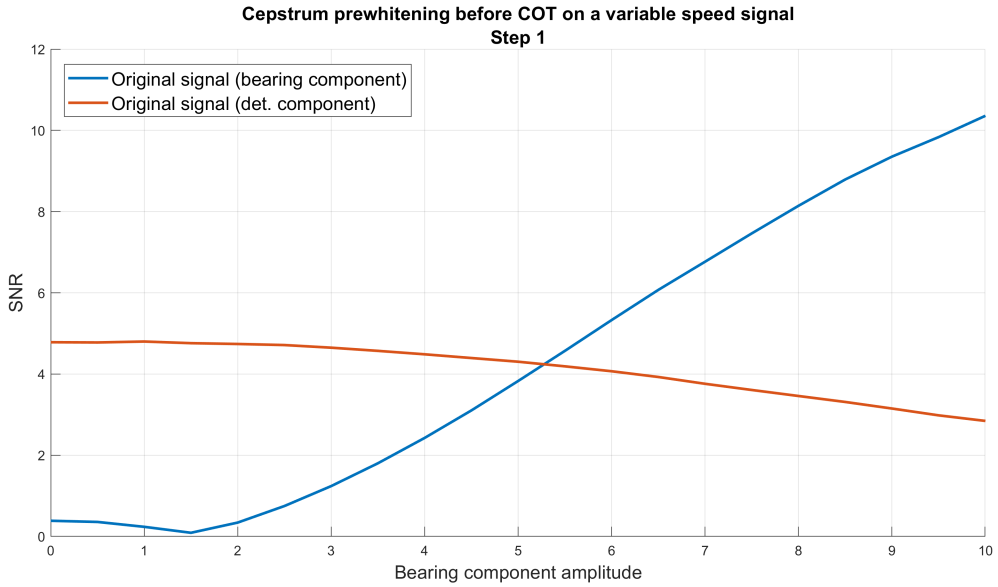


Figure 24: Variable speed signal SNR with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

Finally, in Figure 26 the result of bandpass filtering the signal between 1500 Hz and 5500 Hz after it has been prewhitened can be seen.

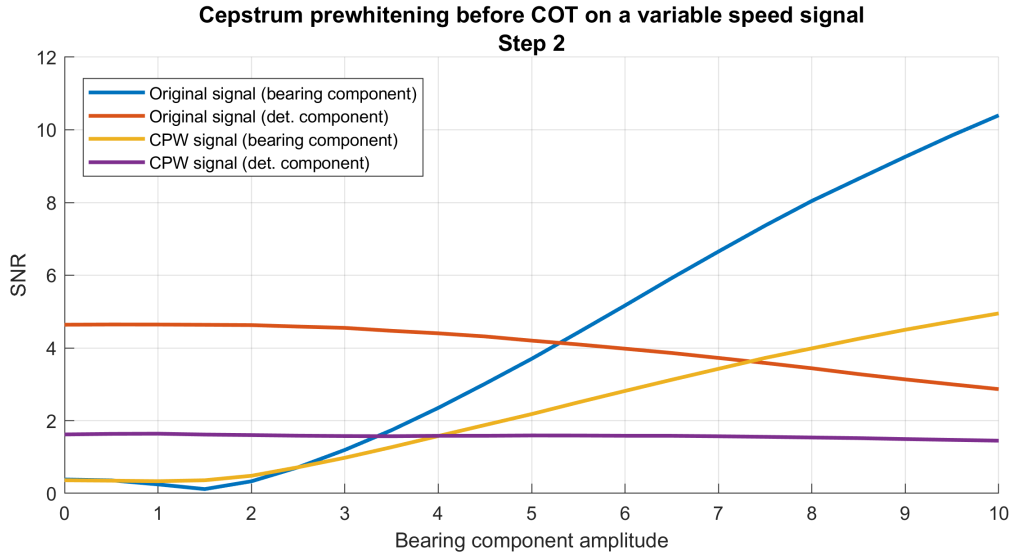


Figure 25: CPW on a variable speed signal with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

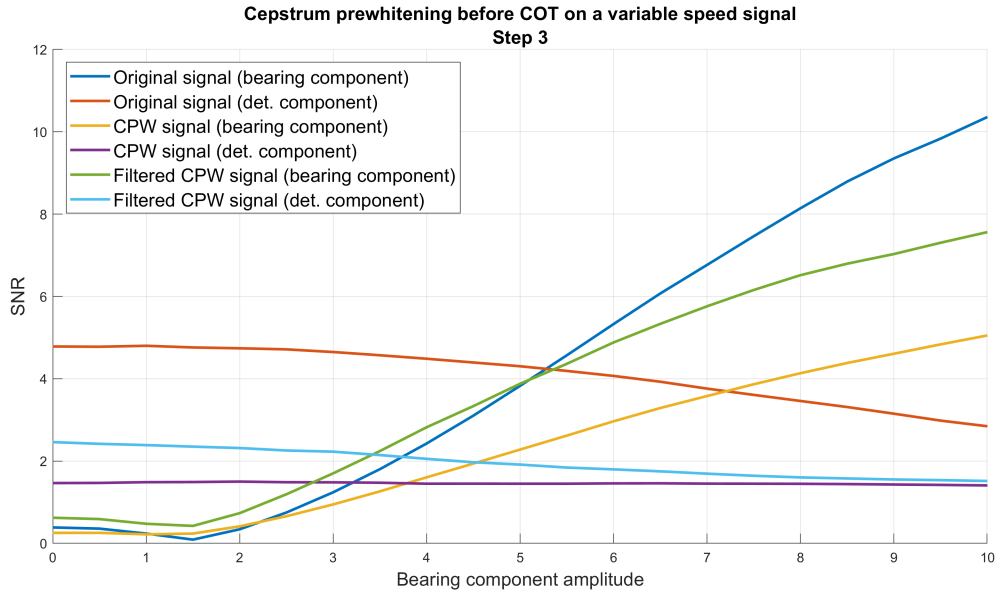


Figure 26: Filtered CPW on a variable speed signal with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_{n}] = [3, 3, 10]$.

From Figures 24, 25 and 26 it is evident that the CPW method proposed by Barbini et al. (2018), where CPW is performed prior to the COT operation, increases the SNR of the bearing component while simultaneously decreasing the deterministic component. To properly compare the two methods shown in Figure 23, the previous results are all

plotted in Figure 27 where the plain lines represent method a) or M1, and the lines with the diamond markers represent method b) or M2 as proposed by Barbini et al. (2018).

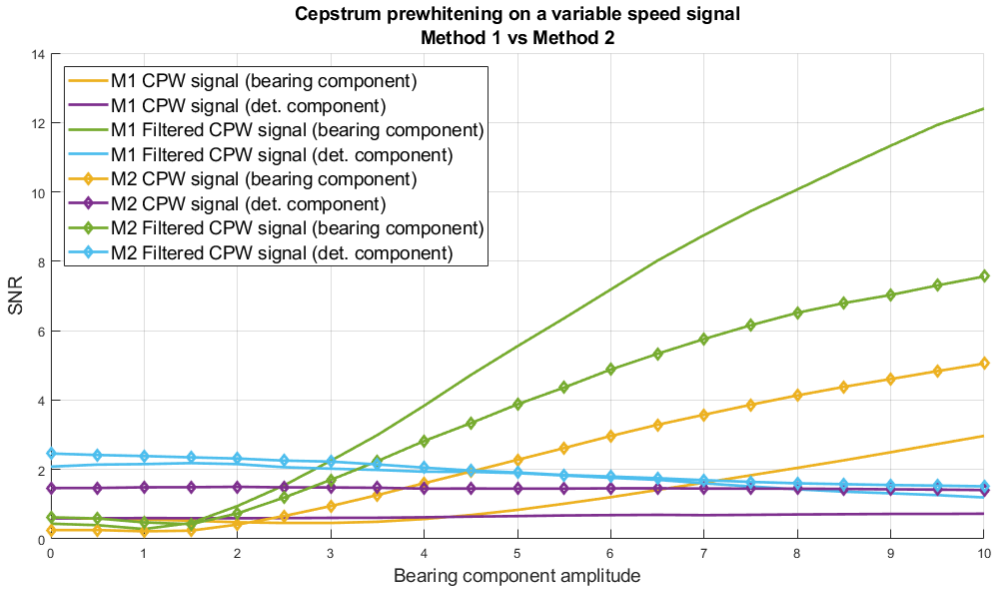


Figure 27: CPW processing method comparisons.

From the above results, when the bearing component amplitude is low, both method a) and b) result in similar SNR ratios after the final bandpass filtering step. The method proposed by Borghesani, Ricci, Chatterton & Pennacchi (2013) slightly outperforms the method suggested by Barbini et al. (2018), but the improvement, especially at lower bearing component amplitudes is not sufficient to justify the additional steps and computing time required. Therefore, for the remainder of this dissertation, the method proposed by Barbini et al. (2018) in Figure 27 b) will be used due its very simple approach, significant pre-whitening power and its computational efficiency.

4.3.2 Autoregressive models

Another popular method of increasing the bearing SNR is that of autoregressive (AR) models recommended by Sawalhi & Randall (2011) and Sawalhi et al. (2007). The AR filter as described in Sawalhi et al. (2007), predicts the deterministic pattern of the signal but is not capable of adapting to the sudden impulses caused by a localised fault. As a consequence, the fault impulses will be left in the residual, ε_k, of the AR filter. Mathematically, Wang & Wong (2002) describe the AR model with Equation (30). The optimum model order, p, should be less than the spacing between two consecutive impulses, as this guarantees that the model does not adapt to the impacts as being a part of the deterministic signal and that they will be contained in the residual signal (Sawalhi et al. 2007).

The same signals used in Figures 20 - 22 will be used to compare the effectiveness of the AR model. Based on the results from Figure 27 and Figure 28, both methods achieved

comparable results. The Autoregressive residual model performed only slightly better than CPW, however, the AR method requires the hyperparameter, model order p, to be selected which complicates the method. Due to the blindness, yet surprising effectiveness of CPW in reducing deterministic components, CPW will be used further in this dissertation.

$$x_k = -\sum_{i=1}^{p} a_i x_{k-i} + \varepsilon_k \tag{30}$$

where p is the model order, a_i are the autoregressive coefficients and ε_k is a Gaussian white noise series with zero mean and variance σ^2 (Wang & Wong 2002).

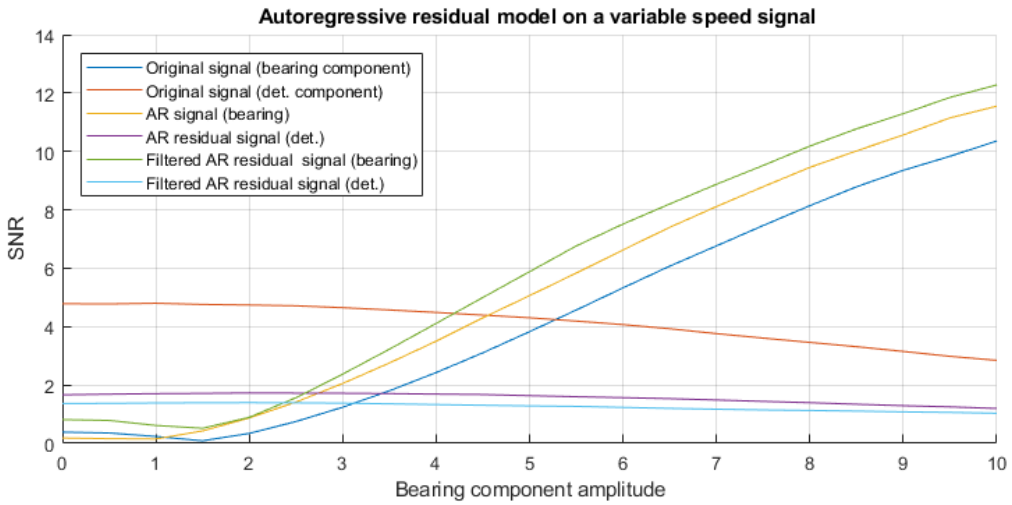
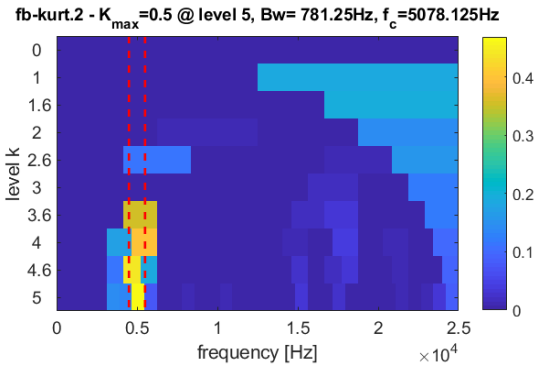


Figure 28: AR filter on a variable speed signal with increasing bearing component amplitude and fixed $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n] = [3, 3, 10]$.

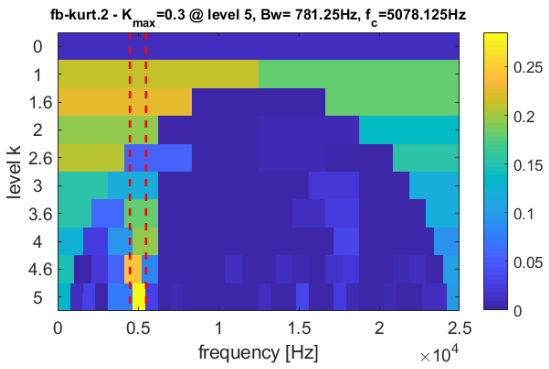
Even though both CPW and AR models performed similarly, the scope of this dissertation limits methods to those with no a priori knowledge of the fault, such as the optimal order for the AR model. Thus, only CPW Method 2, will be considered for removing the deterministic components from here on.

4.4 Cepstrum pre-whitening and the Grams

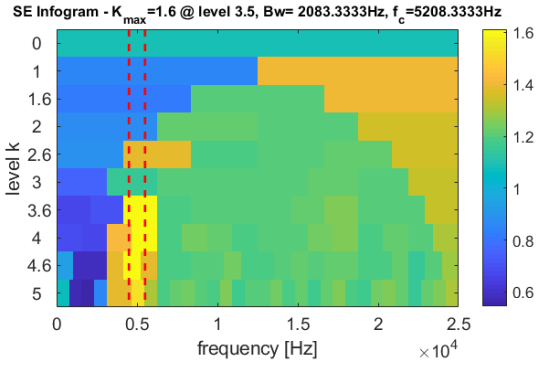
Without a doubt, CPW is capable of increasing the bearing signal-to-noise ratio. When visualizing a bearing fault by looking at the spectrum, CPW will increase the peak value at the bearing characteristic frequency relative to the noise and deterministic components in the signal. We wish to confirm that CPW will also increase the Grams ability at ignoring or removing deterministic components. To determine if CPW aids with the detection of bearing damage within the context of automatic IFB identification, CPW will be applied on both phenomenological and real signals before filtering with the Grams. Figure 29 shows the result of applying the Grams on a constant speed, 900 rpm, phenomenological signal containing scaling components $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [6, 2, 12, 3.1]$ with and without CPW.



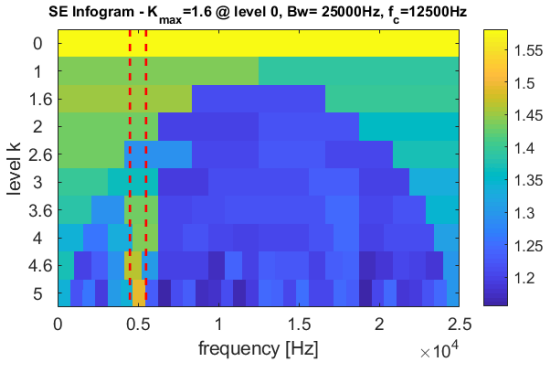
(a) Fast Kurtogram



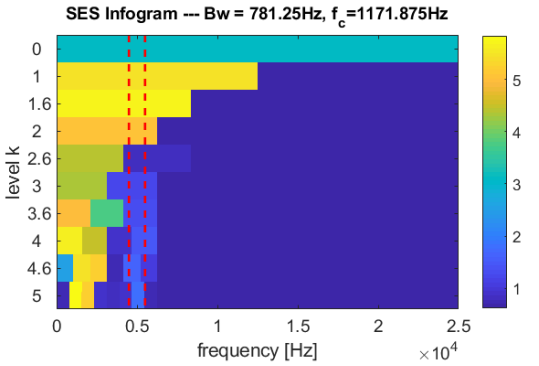
(b) CPW Fast Kurtogram



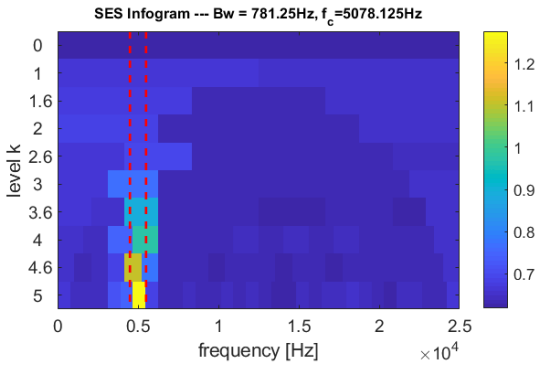
(c) SE Infogram



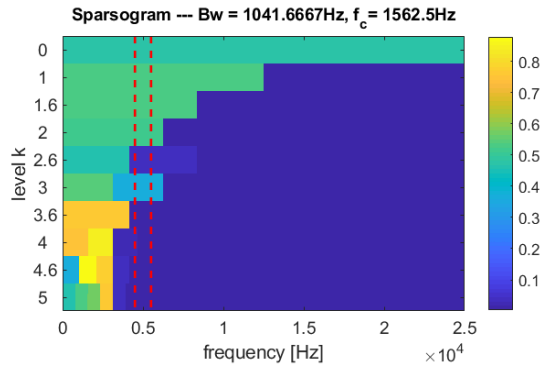
(d) CPW SE Infogram



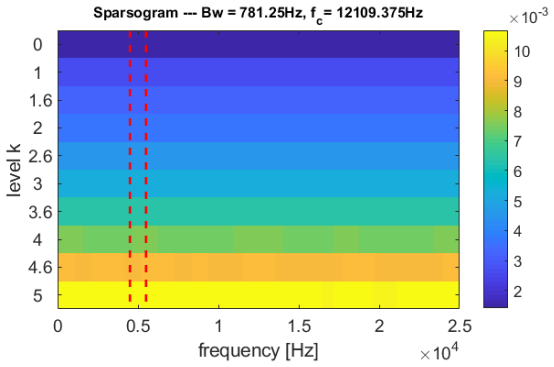
(e) SES Infogram



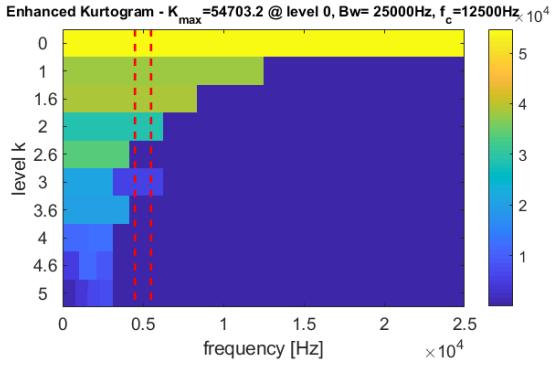
(f) CPW SES Infogram



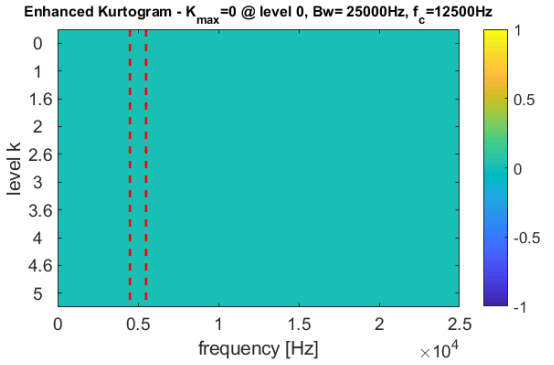
(g) Sparsogram



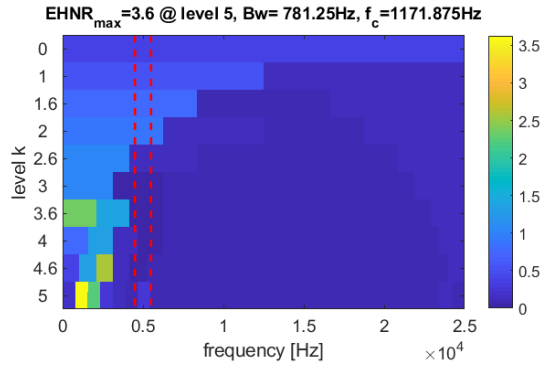
(h) CPW Sparsogram



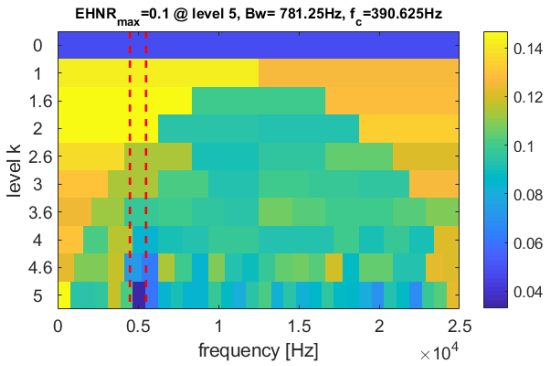
(i) Enhanced Kurtogram



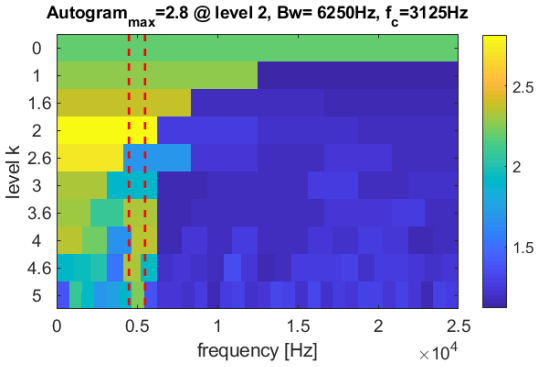
(j) CPW Enhanced Kurtogram



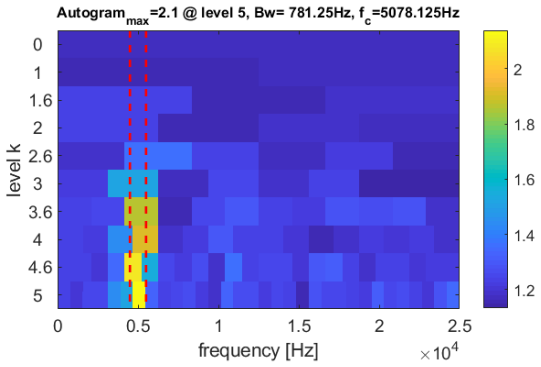
(k) EHNRogram



(l) CPW EHNRogram



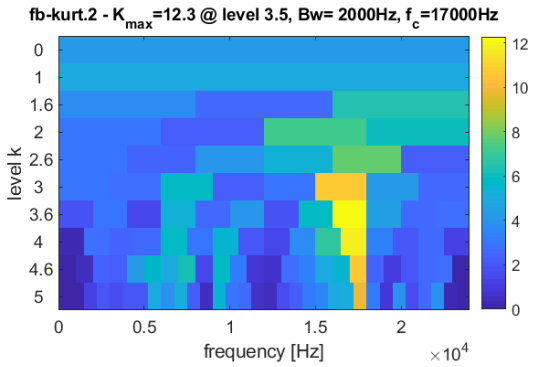
(m) Autogram



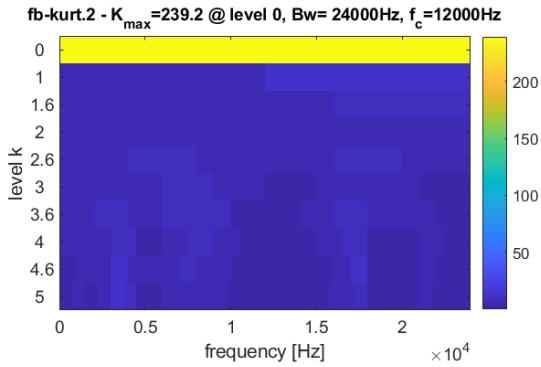
(n) CPW Autogram

Figure 29: Influence of CPW on popular Grams applied to a phenomenological signal under constant speed. The red dashed lines are for quick reference to the known ideal location of the bearing resonance centred at 5000 Hz.

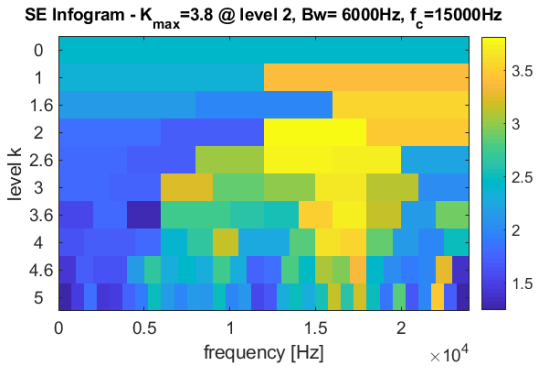
From Figure 29 it was found that of the 7 Grams under consideration, only the SES Infogram and Autogram had improved results with the use of CPW. In the case of the Sparsogram and Enhanced Kurtogram, neither rendered results of any use. The Enhanced Kurtogram had a feature value of zero at all decomposition levels and the Sparsogram consistently had an increase in the feature value with each decomposition level without great difference in the sub bands within each level. Similar results were found on a real signal with OR fault from the *Case Western Reserve University Bearing Data Center* (2011).



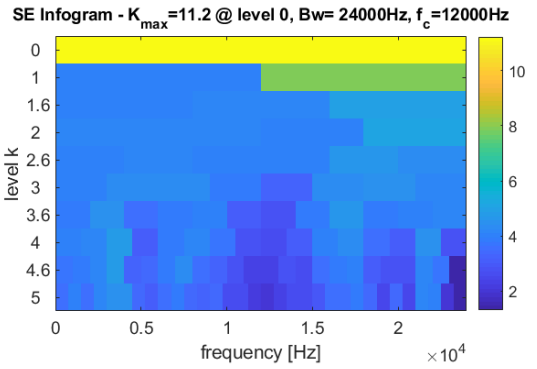
(a) Fast Kurtogram on real signal



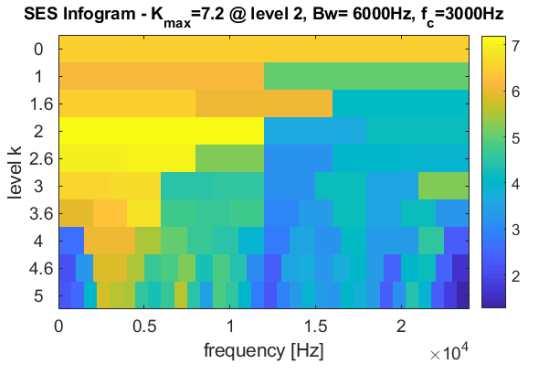
(b) CPW Fast Kurtogram on real signal



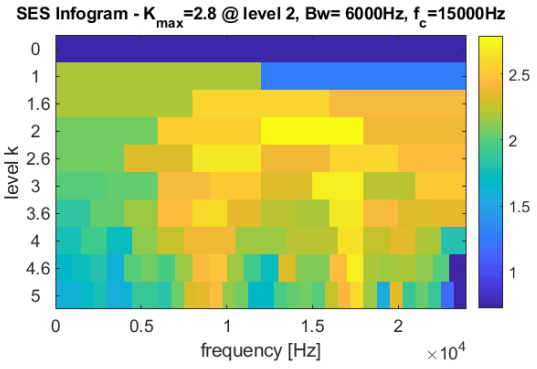
(c) SE Infogram on real signal



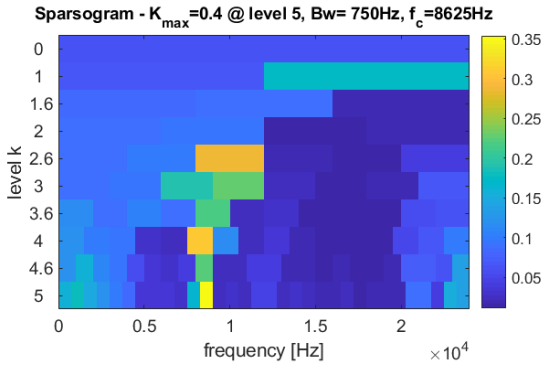
(d) CPW SE Infogram on real signal



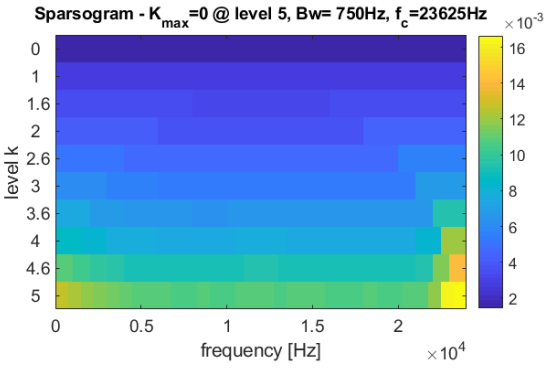
(e) SES Infogram on real signal



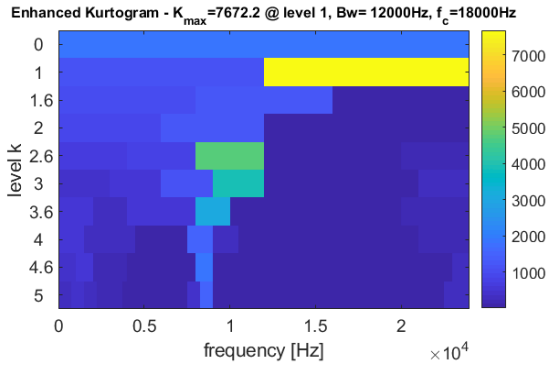
(f) CPW SES Infogram on real signal



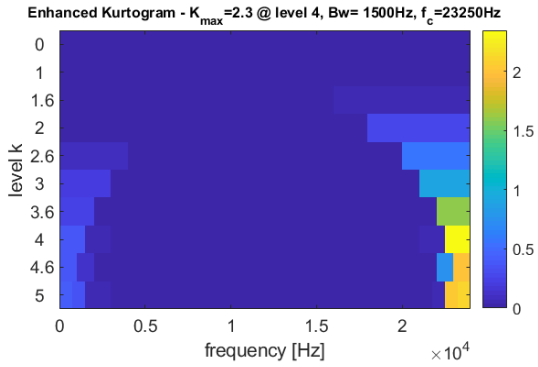
(g) Sparsogram on real signal



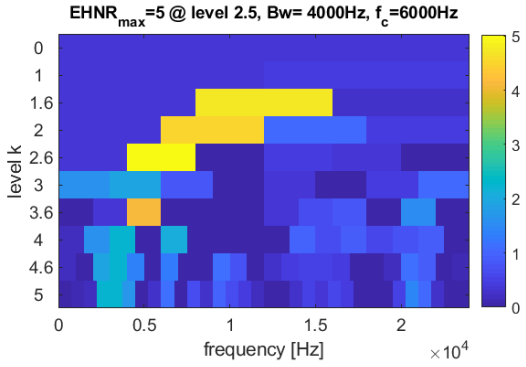
(h) CPW Sparsogram on real signal



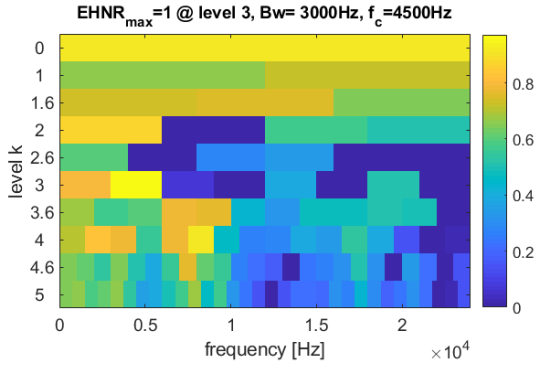
(i) Enhanced Kurtogram on real signal



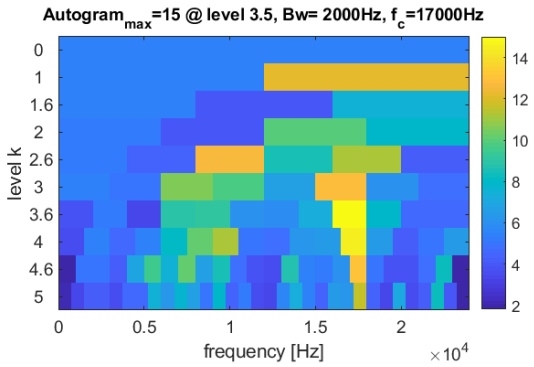
(j) CPW Enhanced Kurtogram on real signal



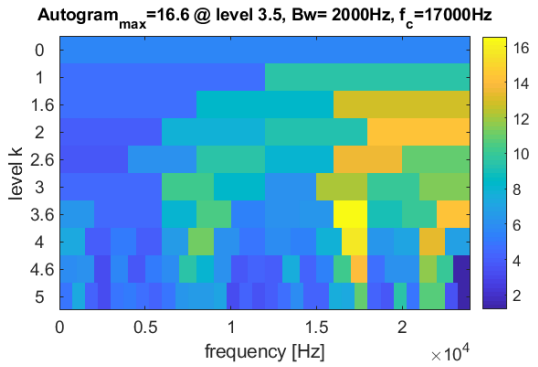
(k) EHNRogram on real signal



(l) CPW EHNRogram on real signal



(m) Autogram on real signal



(n) CPW Autogram on real signal

Figure 30: Influence of CPW on popular Grams on a bearing with OR fault operating under constant speed from *Case Western Reserve University Bearing Data Center* (2011).

From Figure 29 and Figure 30 it is evident that CPW is not generally applicable to the Grams, and does not warrant its blind use for removing deterministic components from a vibration signal. CPW affects the statistical properties of the signal and therefore feature values are affected. Visually, bearing damage components appear larger in the spectral domain when CPW has been applied, but the statistical properties are affected such that the Grams do not behave as with the raw signal. Thus, there exists the need

for a methodology that can blindly remove deterministic components from a vibration signal under constant and varying operating conditions.

4.5 Novelty Information Criterion - NICogram methodology

Data acquisition in a CBM plan generally occurs over a long period of time. During this time, machine components are assumed to go from a healthy state to some unknown damaged state with samples taken at set intervals, continuously or at whichever sampling rate the CBM plan calls for. This means that historical healthy data is readily available which may offer valuable insights into the condition of the machine. Currently, the Grams make no use of historical data, meaning that potentially valuable information is unused during the diagnostic process. Wang et al. (2016) then proposed using a spectral kurtosis ratio (SKR) between a single baseline signal from a healthy gearbox to that of an unclassified measured signal. A new SKRgram is constructed using this method by calculating the ratio between the baseline Fast Kurtogram and the unclassified Fast Kurtogram to highlight the presence of bearing damage in a planetary gearbox. This method was applied to both simulated and experimental data under constant speed applications and proved to be an effective damage detection method. Hence, in this dissertation the NICogram methodology is proposed which extends the work of Wang et al. (2016) to fluctuating speed conditions.

The NICogram methodology describes the process of normalizing an unclassified signal with known healthy historical data, by calculating NIC values which form a NICogram. Signals containing damage are assumed to have different properties from that of a healthy signal. The greater the damage, the larger the NIC value. The NICogram methodology is applied on the already existing Grams, mentioned earlier in this dissertation, and increases their diagnostic capabilities by favouring nodes containing novelty information. Hence, the method receives its name from the novelty information criterion (NIC) which it uses to improve the diagnostic ability of the conventional Grams. The method is an improvement on existing Grams in terms of damage detection and blind removal of deterministic components. It can be applied on any Gram according to the feature that is most relevant to the characteristics of a vibration signal. To calculate NIC values in the construction of a NICogram, healthy data must be available under the same conditions as the operating condition of the signal under inspection. The process of constructing a NICogram is described by the flow diagram in Figure 31.

Firstly, a Gram must be selected which the user deems most appropriate for the characteristics of the signal. For example, if it is known that there is a high chance of random impulse noise, select the Sparsogram. If the user is uncertain which Gram will be most appropriate, the author recommends using the popular Fast Kurtogram. For the purpose of this explanation, the Sparsogram will be used. Now, calculate a Sparsogram for both the healthy and unclassified signal. For each respective Sparsogram, normalize the Gram with respect to the node with the largest feature value (in this case sparsity). Normalizing of the Grams by the largest feature value reduces the effects of the changing operating conditions and was found to improve diagnostic results. Normalizing is also required so that NICogram can be calculated with values of similar magnitude. Now, divide the nor-

malized unclassified Sparsogram by the normalized healthy Sparsogram which will result in the construction of a NICogram. The procedure now follows that of any conventional Gram where the node with the largest NIC value is selected to design the bandpass filter after which the squared envelope spectrum is calculated and used for diagnosis. Since CPW visually improves the fault order in the spectrum, CPW can also be incorporated in the method just before calculating the SES of the signal although not shown in Figure 31. In this dissertation CPW is applied before the calculation of the SES for all examples and simulations since recent work by Borghesani, Pennacchi, Randall, Sawalhi & Ricci (2013) and Barbini et al. (2018) has shown it to be effective at enhancing bearing components in a noisy dataset.

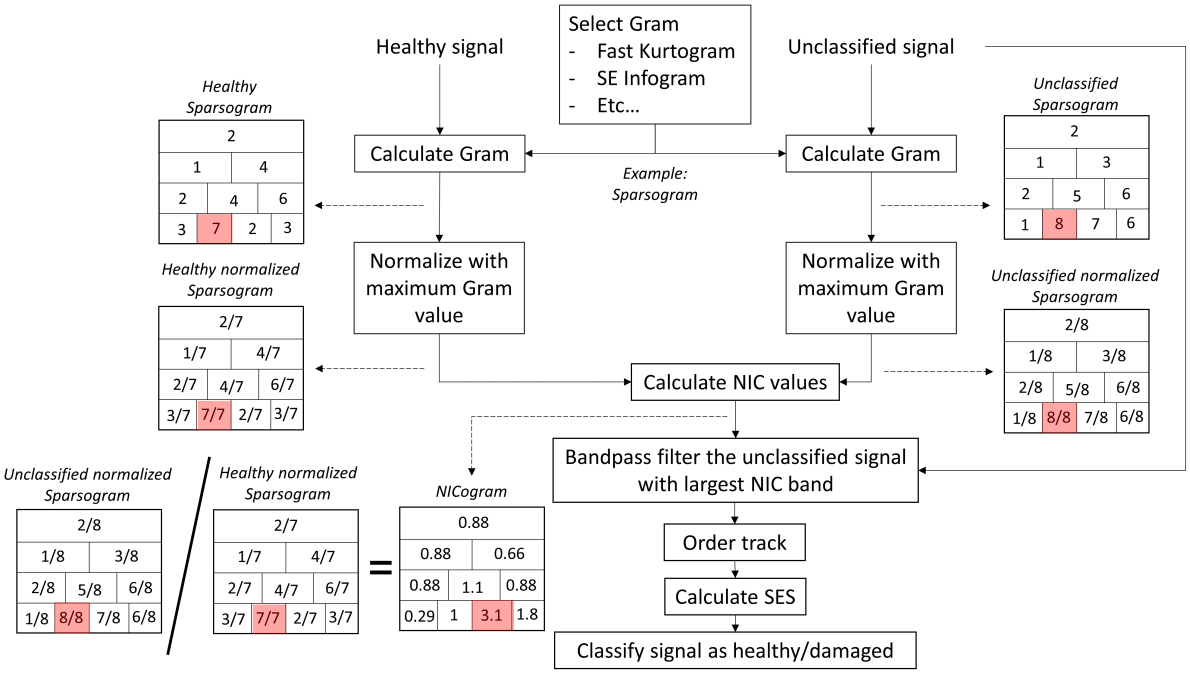


Figure 31: Band selection process using the NICogram methodology with the Sparsogram as an example and the nodes with largest feature value highlighted.

It is assumed that the only change in feature value within a specific band is due to the presence of a fault. Frequency bands that contain information related to the presence of damage are assumed to increase the feature value relative to the respective band of the healthy condition. Frequency bands that do not contain novel information related to a fault will thus either decrease or remain the same, resulting in an NIC value of less or equal to one. Thus, when comparing the healthy signal Gram to the unclassified signal Gram, only the bands that contain a positive change in feature value will result in NIC values greater than one according to Equation (31). The frequency band with the largest relative positive change will result in the largest NIC value and is assumed to be the band containing the fault resonance frequency. NIC values are calculated as the fraction of the new feature value over that of the healthy condition as follows,

$$NIC = \frac{FV_U}{FV_H}, \tag{31}$$

where FV_U and FV_H are the feature values of the unclassified and healthy signals respectively. The NICogram methodology can be applied to any one of the popular Grams aimed at finding the IFB. Figure 33 a) and b) show the Fast Kurtogram developed by Antoni (2007) for a healthy and damaged bearing signal respectively. The healthy signal has components $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [6, 2, 12, 0]$ and the unclassified signal has components $[\mathrm{X}_{dg}, \mathrm{X}_{rg}, \mathrm{X}_n, \mathrm{X}_b] = [6, 2, 12, 2.9]$. The resultant signals along with the operating profile are shown in Figure 32. The damage is sufficiently small that it is not detectable in the pure time domain signal.

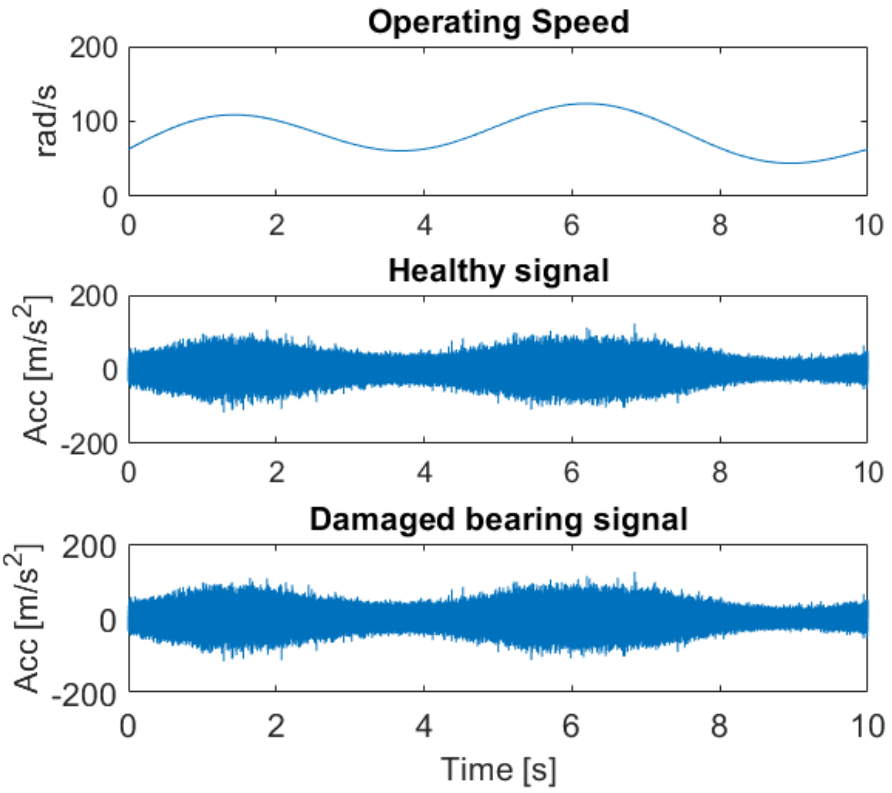


Figure 32: Operating profile and bearing signals from the phenomenological model.

The methodology is applied in Figure 33 on the signal that is presented in Figure 32. The bearing damage is centred at 5000 Hz as discussed in Section 4.3.1. It is clear that in Figure 33 b) the IFB containing bearing damage centred at 5000 Hz is present, but it is not the dominant band (highest kurtosis). Referring to Figure 34 b) we can see that the automatic band selection from the conventional Fast Kurtogram does not filter the signal as to maximise the bearing fault. Instead, the low frequency components related to the deterministic, random and noise components dominate the signal.

However, using the NICogram methodology on the popular Fast Kurtogram, the IFB centred at 5000 Hz is by far the most dominant band since it contains new information. Bandpass filtering with the selected band from Figure 33 c), the bearing fault at order 13 can be clearly identified in Figure 34 c) along with the second and third harmonics, while completely filtering out the deterministic and random components.

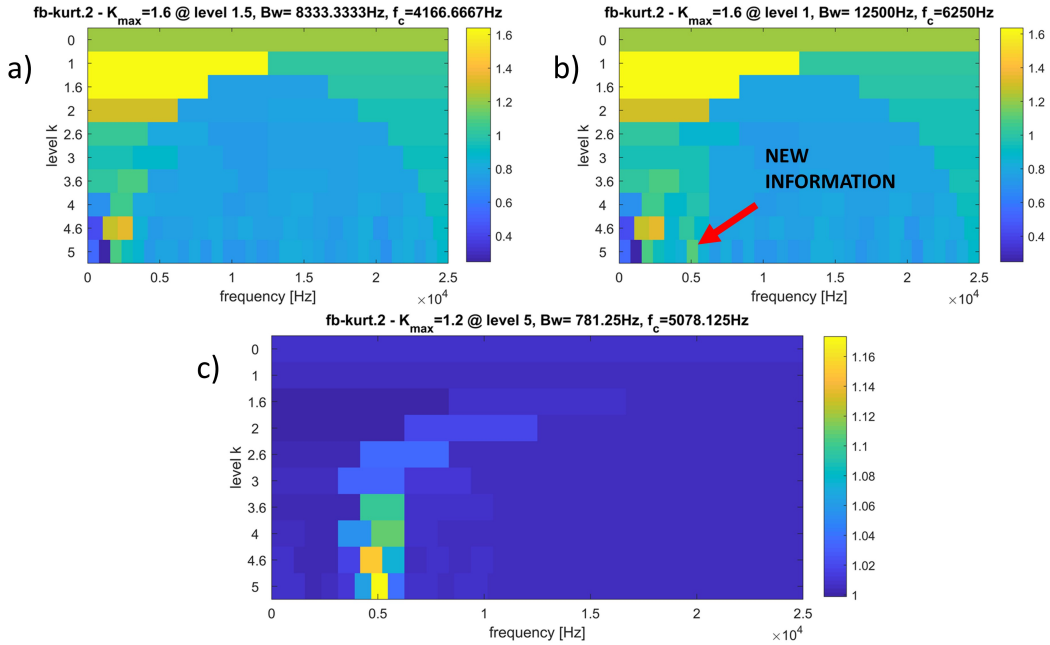


Figure 33: Band selection using the Fast Kurtogram for a vibration signal with a healthy bearing (a) and damaged bearing (b) and the result of using the NICogram method (c) to identify the new information between the healthy and damaged states on a varying speed signal.

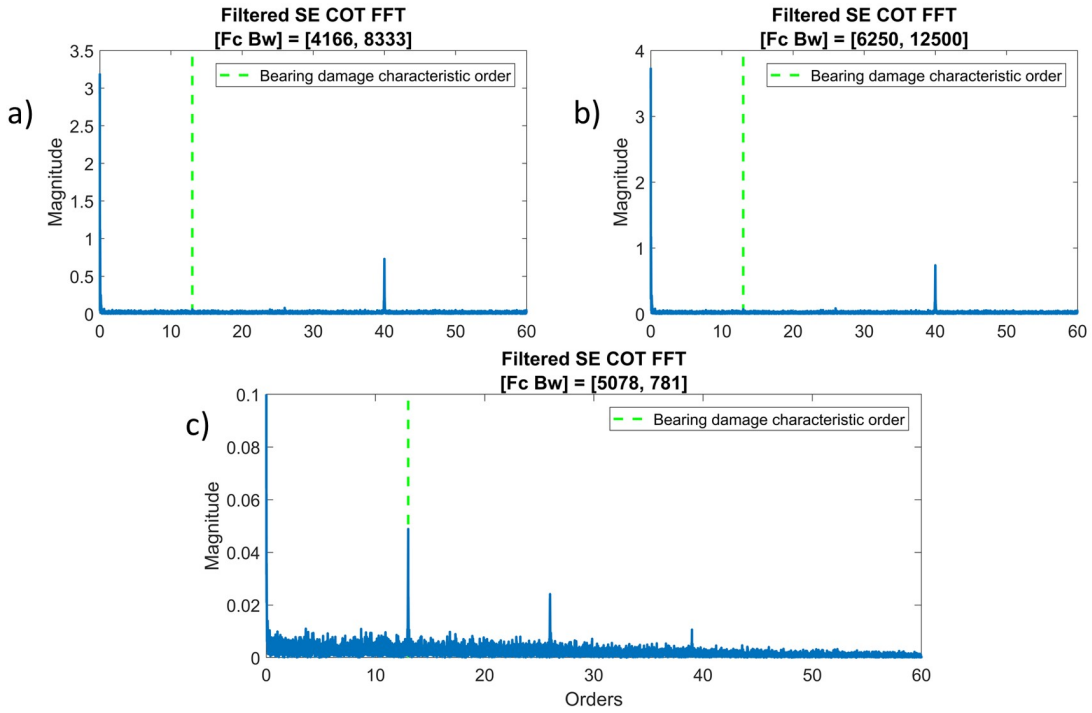


Figure 34: Squared envelope order spectrum obtained after bandpass filtering centred at the frequency selected by the Fast Kurtogram with a healthy bearing (a), damaged bearing (b) and the NICogram method (c) on a varying speed signal. No CPW was applied on any of these signals.

Deterministic components will be present at the same frequency bands in both the healthy and new signals. Therefore, when comparing the healthy Gram to the damaged Gram, the NIC value should theoretically remain the same at the frequencies excited by the deterministic components. The sub bands that contain the novelty information will thus be dominant and result in a higher NIC value. The resultant frequency band identified by the NICogram is sensitive to the choice of healthy signal. Not every healthy signal used in the calculation of a NICogram renders the same result. It is currently not possible to know beforehand which signal will yield the clearest results for diagnosis. Therefore, the author suggests using a probabilistic fault detection methodology. By independently normalizing against multiple healthy signals, a distribution of results can be obtained which can identify the presence of damage with greater likelihood.

4.6 NICogram evaluation

To evaluate the proposed methodology, both phenomenological and experimental data will be used. The phenomenological signal allows for the testing of many hypothetical scenarios without the expensive and time consuming procedure involved with collecting actual data. Then, once it has been shown that the proposed method is an effective approach, experimental data can be used to validate the application to real signals similar to what might be experienced in industry.

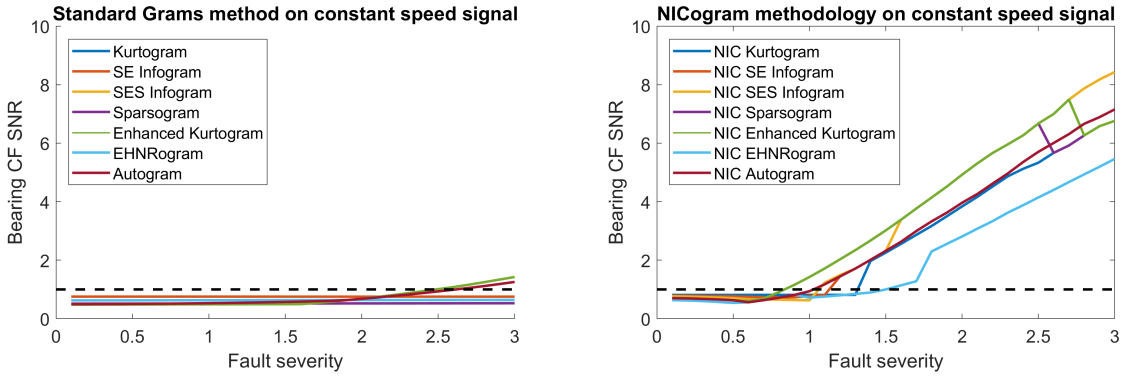
4.6.1 Phenomenological signal results

By not only looking at features that are generally indicative of damage, but also at the new information that is not present in the healthy signals, incipient damage detection is made possible by the NICogram method. Similar to the investigations performed using CPW, the bearing CF SNR, as explained in Figure 18, will be used to classify signals as healthy or damaged. The question remains as how to define the threshold limit such that a minimum number of healthy signals are classified as damaged while still maximising the correct identification of damaged signals. Appendix B shows the use of ROC (receiver operating characteristic) curves for finding the optimal spectrum percentile value to use as the threshold value for the signals from the experimental dataset. Appendix B is used to compare the performance of the different Grams in a systematic and fair way, and does not need to be calculated to construct a NICogram.

For all seven Grams, a threshold value set as the 50th percentile allowed for 100% correct classification of damaged data, but also falsely classified over 90% of the healthy data. On the other hand, a threshold value set to the 100th percentile meant that almost none of the Grams correctly identified any damage. Finally, it was found that a threshold value using the 99.5th percentile is the optimal value which allows for the highest true positive rate and the lowest false positive rate on the experimental datasets. The ratio between the spectrum value at the known bearing characteristic frequency and the 99.5th percentile are therefore used as the detectability metric for all further studies in this dissertation, including the phenomenological signals as well as the experimental signals. If the CF

SNR value is greater than one, it implies that the damage is likely detectable and the signal is then classified as damaged.

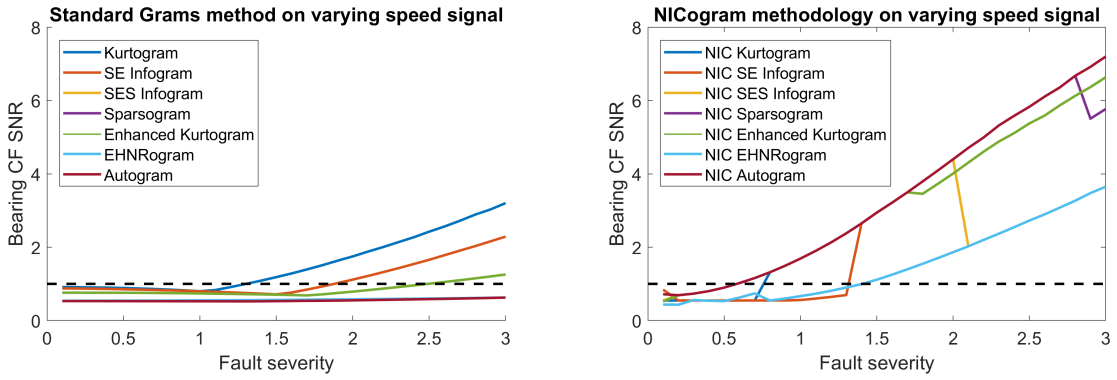
The greater the CF SNR value, the clearer the damage. This principle is used in Figure 35 (a) and (b) to show the CF SNR detectability level as the fault severity increases for both the conventional Gram procedure applied on all seven Grams and the NICogram method applied to the same Grams yet incorporating healthy data. The fault severity is increased by sequentially increasing the bearing component from the phenomenological model from zero to three. Only once the fault severity becomes relatively large do the Autogram and Enhanced Kurtogram detect the presence of a fault when using the conventional Gram procedure, while all other Grams fail to detect any damage at all. On the other hand, the NICogram methodology substantially improves the detection ability of all the Grams under consideration. The damage can be detected at much lower fault severity levels for all seven Grams using the NICogram method as shown in Figure 35 (b).



(a) (b)

Figure 35: Damage detection capability of conventional Grams vs the improvement by the NICogram method on a noisy constant speed signal with a threshold of 99.5th percentile.

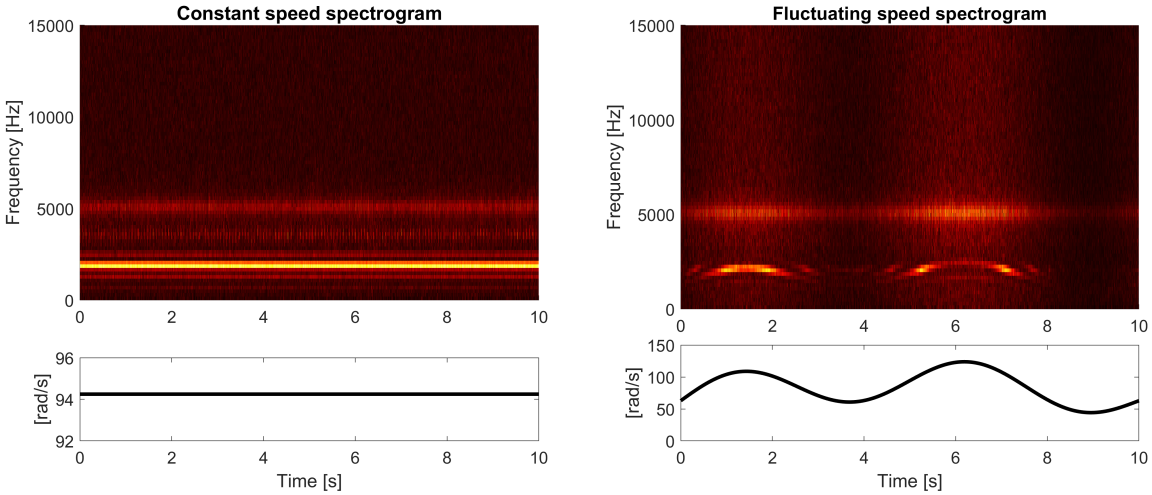
Similarly, we wish to compare the same methods under fluctuating speed conditions. Interestingly, only the Kurtogram, SE Infogram and Enhanced Kurtogram detect damage under varying speed conditions with damage severity levels from 0-6 when using the conventional Gram procedure in Figure 36 (a). On the other hand, when the NICogram methodology is applied on the varying speed signal, each of the Grams can detect damage even at low fault severity as shown by Figure 36 (b). It has long been known that the Grams can detect damage for constant speed applications, yet from the author's experience, little information is available regarding their application to fluctuating speeds. From the phenomenological model results, after applying the proposed methodology, it is clearly evident that the Grams can potentially be used for varying speed applications. Promisingly, from Figure 36 (b), it appears that the NICogram methodology greatly improves the Grams ability to detect damage even at low fault severity under fluctuating speed.



(a) (b)

Figure 36: Damage detection capability of conventional Grams vs the improvement by the NICogram method on a noisy varying speed signal with a threshold of 99.5th percentile.

It is intriguing to note that the conventional Fast Kurtogram and SE Infogram actually performed worse under the constant speed conditions in Figure 35 (a) than for the varying speed case in Figure 36 (a). The reason that the Fast Kurtogram and SE Infogram worked better on the fluctuating speed signal can be explained by looking at Figure 37 which shows both the operating profile and spectrogram of a constant and variable speed signal respectively. The deterministic component which often dominates a vibration signal is a lot stronger under constant speed. In both cases the bearing's resonance frequency at 5000 Hz can be detected, but at constant speed the deterministic component at 2000 Hz clearly dominates the signal. This makes it more difficult to identify bearing damage under constant speed when there is a strong presence of deterministic components and only using the conventional Grams. Nonetheless, when the NICogram method was used for both constant and fluctuating cases, the Grams were normalized with respect to the healthy condition and therefore all NIC Grams could identify damage.



(a) (b)

Figure 37: Effect of constant and varying speed on resonance frequencies from phenomenological model signal.

To verify the results on real signals, the experimental database with varying levels of damage will be used. The purpose is to test the improvement made by the NICogram methodology for fluctuating speed conditions and determine if the method truly is as effective as found using phenomenological signals.

4.6.2 Experimental and real signal results

The experimental setup described in Section 3.2 is now used to support the findings from the phenomenological signals in Section 4.6.1. The fluctuating speed profiles that were used to compare the conventional Grams with those improved by the NICogram method are shown in Figure 38. The operating profiles are of sufficient complexity to critically evaluate the effectiveness of the different features under fluctuating speed conditions. No pre-whitening or pre-processing was performed on any of the signals or results described below. Also, all Grams are used on the time-domain signal. Computed order tracking is only used to view the order spectrum of the varying speed signals, but is not used as a precursor to processing.

Preliminary analysis of the experimental signals highlighted a dominant component at roughly 8000 Hz which dominates the spectrum and also the Grams. This component is present in every recording, both healthy and damaged for constant and varying speed conditions. The spectrograms in Figure 40 highlight the presence of the 8000 Hz band as well as other less dominant energy bands in the spectrum of a healthy and damaged signal respectively. Figure 39 shows the unfiltered healthy order spectrum of a constant speed recording and highlights the spectral content of the signal along with the rotational frequencies (RF) of the different shafts.

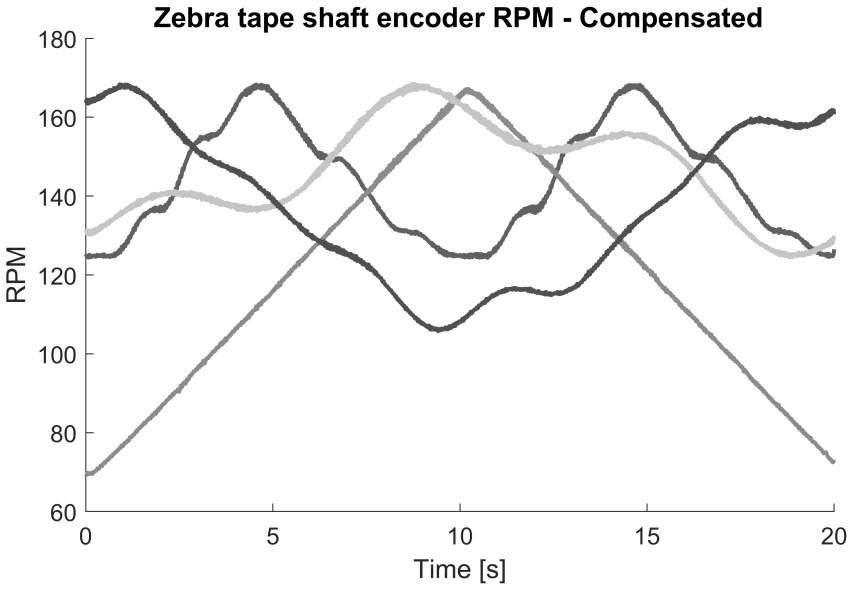


Figure 38: Speed profiles of fluctuating speed signals.

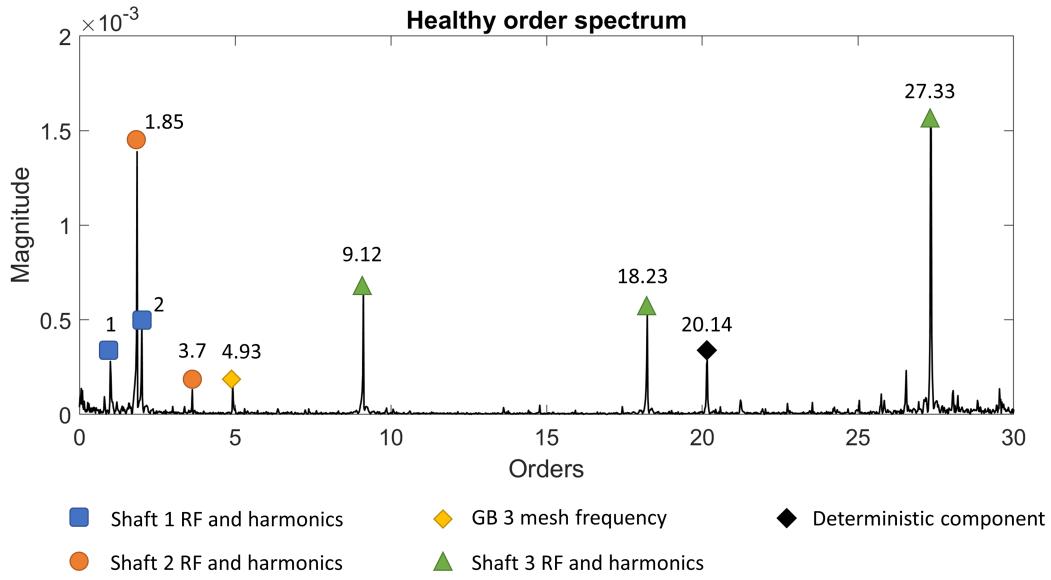


Figure 39: Healthy recording order spectrum highlighting the low frequency content of a constant speed signal.

The characteristic frequencies for the bearing of interest are shown in Table 2. Note that bearing slippage is generally between 1-2% (Sawalhi & Randall 2008), which means that the BPFO could lie anywhere between 3.51-3.65 orders. As expected, none of the bearing characteristic frequencies are present in the healthy order spectrum in Figure 39 since there is no damage present in the system.

Table 2: Bearing characteristic frequencies of a 6206C3 NTN bearing.

BPFO	**BPFI**	**FTF**	**BSF**
3.58	5.42	0.40	2.36

BPFO	Ball Pass Frequency of Inner Ring
BPFI	Ball Pass Frequency of Outer Ring
FTF	Fundamental Train Frequency
BSF	Ball Spin Frequency

In the healthy unfiltered order spectrum, the unbalance from shaft 3 is easily identified by the 9.12 order component and its harmonics. The rotational frequency of the first and second shaft is also easily identified by orders 1 and 1.85 respectively. Although not dominant, a deterministic component of the gearbox under investigation can also be seen at order 20. No other dominant components can be identified by looking at the spectrum as it currently is and the bearing clearly has no detectable fault on the outer race which would sit between 3.51-3.65 orders.

Figure 41 shows the order spectrum of the same healthy signal, but now filtered around the dominant frequency content at [Fc, Bw] = [8200, 200] as suggested by the spectrograms in Figure 40. As suspected, the dominant content in the signal is the resonance band excited by the deterministic frequency in the gearbox at order 20.18 and its harmonics. This component can be seen in the order spectrum as well as the squared envelope order spectrum. A second component, with stochastic properties, can also be seen in the filtered squared envelope order spectrum that wasn't noticeable in the unfiltered order spectrum itself. This stochastic component lies at 5.74 orders and its second harmonic. The signals recorded from the gearbox setup are therefore dominated by the deterministic and stochastic resonance frequencies around 8000 Hz. Thus, in many cases, the frequency band identified by the conventional Grams do not contain damage related information, but rather identify this dominant resonance band even at high levels of fault severity. It is only when the damage is extremely large that the conventional Grams procedure regularly detects damage.

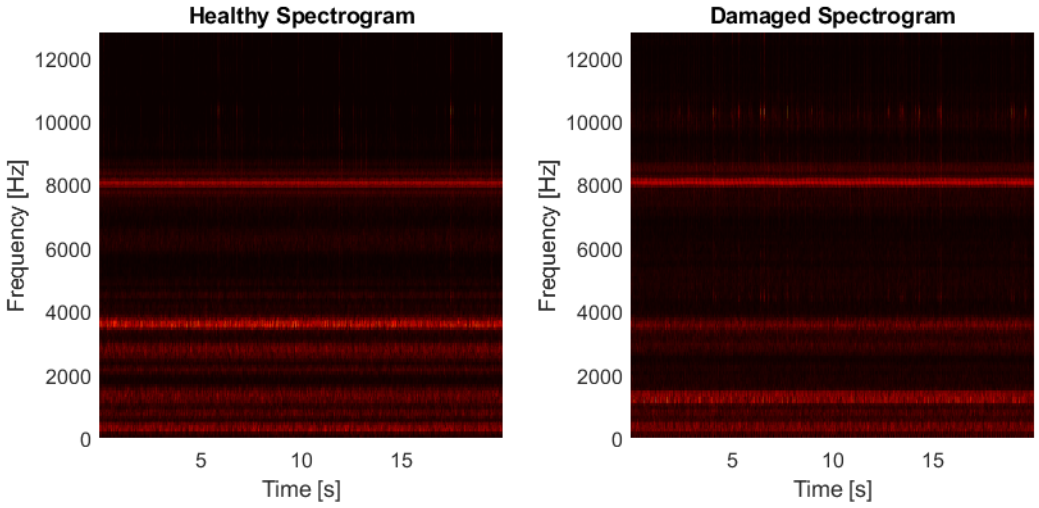


Figure 40: Spectrogram of a healthy and damaged signal showing the dominant spectral component at 8000 Hz under constant speed conditions.

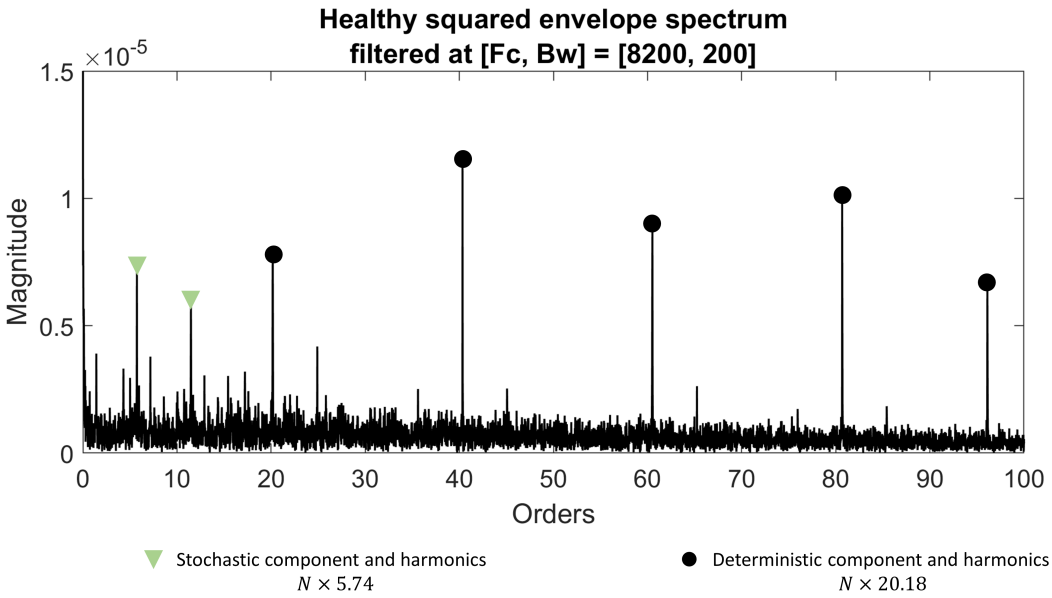
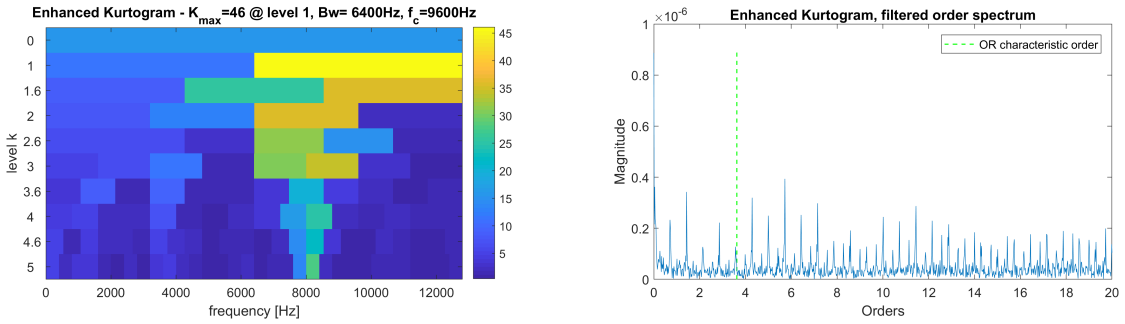
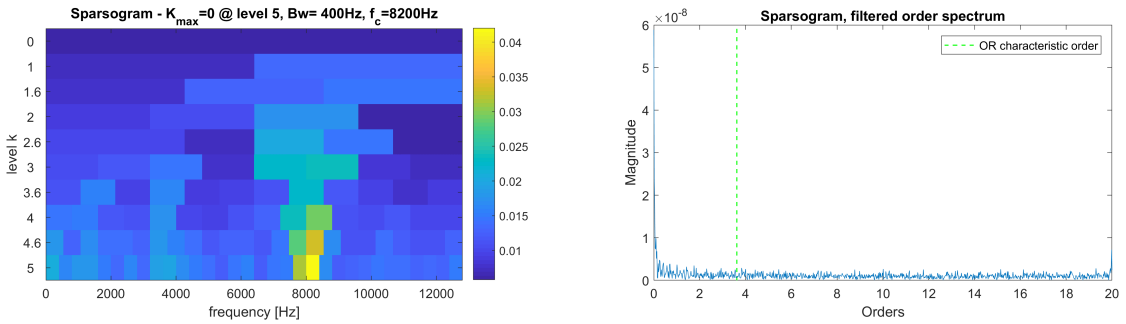


Figure 41: Healthy recording squared envelope order spectrum filtered at [Fc, Bw] = [8200, 200], highlighting the dominant content of the constant speed signal.

Figure 42 displays the resultant Enhanced Kurtogram and Sparsogram when applied on a damaged signal with fluctuating speed conditions. Clearly, the 8000 Hz component is dominant and renders little useful diagnostic information in either Gram. When looking at the filtered order spectrum from the Enhanced Kurtogram and Sparsogram, Figure 42 (b) and (c), it is difficult to detect any damage at the bearing BPFO. Neither Gram clearly identifies that an OR fault is present even though there is already significant damage at damage level 5.



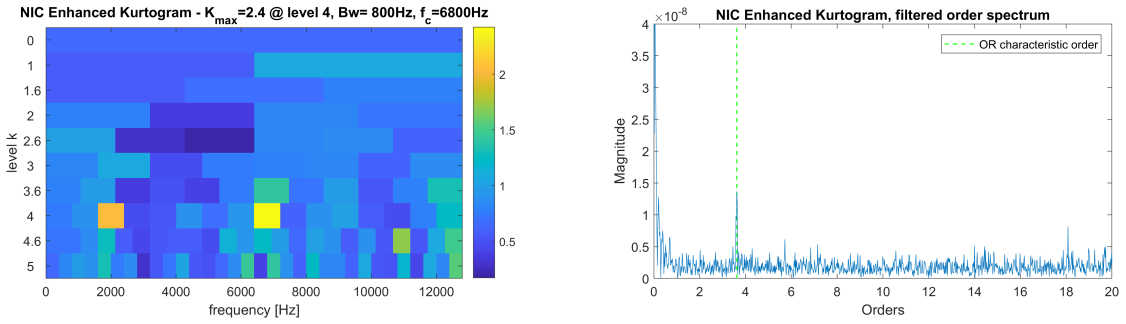
(a) (b)



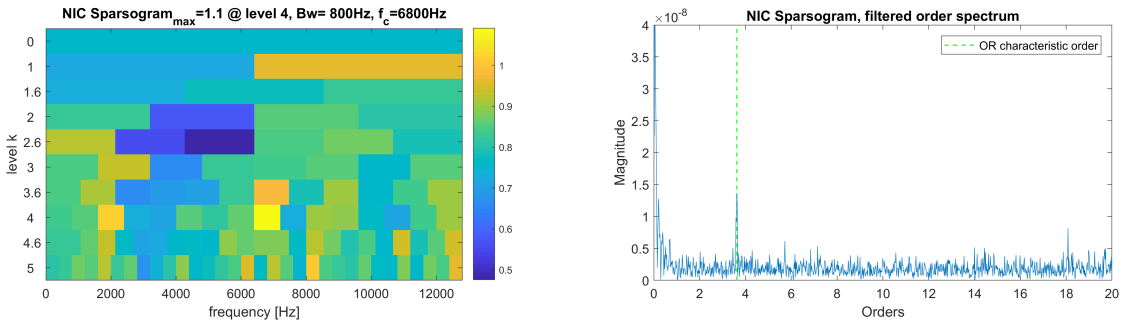
(c) (d)

Figure 42: Enhanced Kurtogram and Sparsogram dominated by 8000 Hz resonance band and the respective order spectrum when filtered around the 8000 Hz component for a fluctuating speed signal at damage level 5.

However, when the NICogram methodology is applied on the Sparsogram and Enhanced Kurtogram, both identify new information as seen in Figure 43. Although not dominant in the original Sparsogram and Enhanced Kurtogram, the information was still present but suppressed due to the very strong 8000 Hz component. By normalizing the Grams against the healthy condition as suggested in Section 4.5, a definitive peak at the OR first harmonic characteristic order can be seen for both Grams. We notice that the 8000 Hz component is almost completely removed since it does not contain novelty information. Figure 43 (b) and (c) show the resultant order spectrum after filtering around the suggested bands identified by the NIC Enhanced Kurtogram and the NIC Sparsogram. The bearing OR characteristic order is dominant and clearly better than the results from the conventional Grams in Figure 42. Thus, the NICogram method is clearly an efficient method at ignoring deterministic components and rather favouring novelty information.



(a) (b)



(c) (d)

Figure 43: NICogram methodology applied on the Enhanced Kurtogram and Sparsogram to the same fluctuating speed signal and now correctly identifying a frequency band containing fault information.

Having proved that the NICogram methodology can effectively ignore deterministic components not related to the presence of a fault, we now wish to statistically compare the conventional Grams performance to those improved by the NICogram methodology.

It must be noted that the damage at D3 is extremely small and hard to detect using existing methods. Figure 44 shows the squared envelope order spectrum of a fluctuating speed signal at D3 highlighting just how small the OR fault is. No visible damage can be seen by looking at the spectrum alone and one would assume that there is no damage in the system. Only the 3rd shaft's rotating frequency harmonics can be seen which is caused by the imbalance on the shaft.

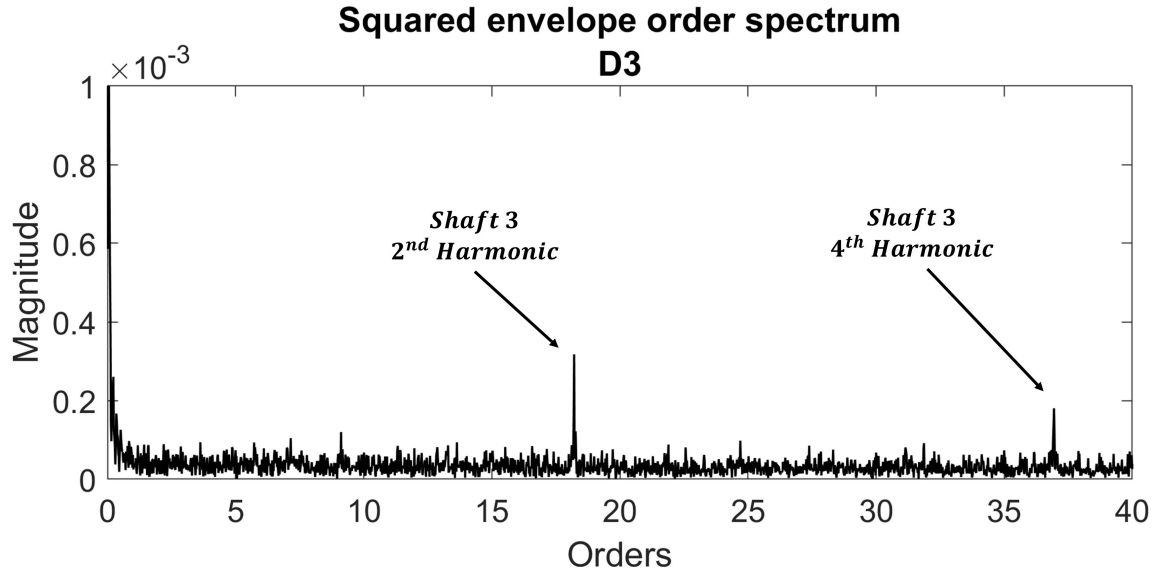


Figure 44: Squared envelope order spectrum with no visible damage at level 3 at the bearing outer race characteristic frequency.

Figures 45 - 51 compare the total classification accuracy of each of the seven popular

Grams investigated in this dissertation with those improved by the NICogram methodology. To further emphasize the achievable improvement through use of the NICogram methodology, the cepstrum pre-whitened squared envelope spectrum (CPW SES) accuracy, as proposed in the paper by Borghesani, Ricci, Chatterton & Pennacchi (2013), is also plotted as a benchmark. Each figure shows the classification accuracy at three different levels of damage, namely damage levels 3-5. Level 6 is not shown because the damage is so severe that a very high classification rate for all datasets is achieved regardless of the selected method. However, the results from damage level 6 can be seen in Appendix C. Since this dissertation is focused on early detection, only levels 3-5 are deemed necessary by the author. Damage level 1 and 2 are not shown because the damage is extremely small and the heavy noise and deterministic content completely mask the bearing damage. Also, the conventional Gram method as normally described in literature (Figure 6) and the CPW SES are used to emphasize that the NICogram methodology really does offer a substantial improvement for early damage detection in the presence of strong background noise and deterministic content.

The NICogram methodology relies on normalizing the Grams against a healthy signal of similar operating conditions to the unclassified signal under inspection. Therefore, normalizing against different healthy signals will play a role in the obtained classification accuracy. To best visualize the results, each damaged signal Gram from D3-D5 is normalized against every healthy signal with the same operating conditions to obtain a distribution of the classification results. The distributions consists of normalizing each damaged signal Gram against the 50 healthy signal Grams for each operating condition. The purpose of representing the data in this manner is so that the method sensitivity to the selected healthy signal can be visualized.

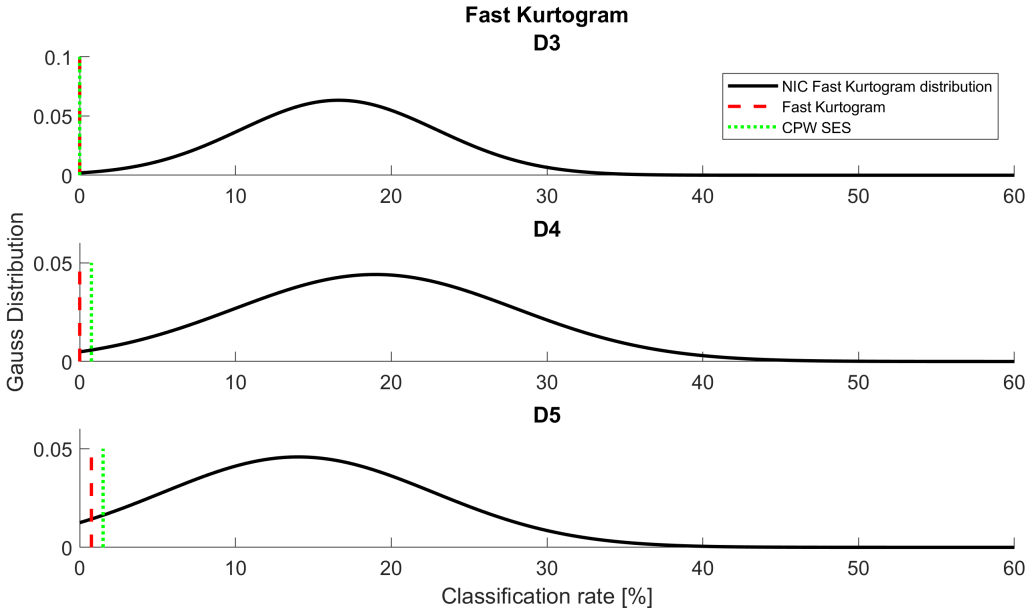


Figure 45: Comparison of the NIC Fast Kurtogram distribution with the normal Fast Kurtogram and popular CPW SES method.

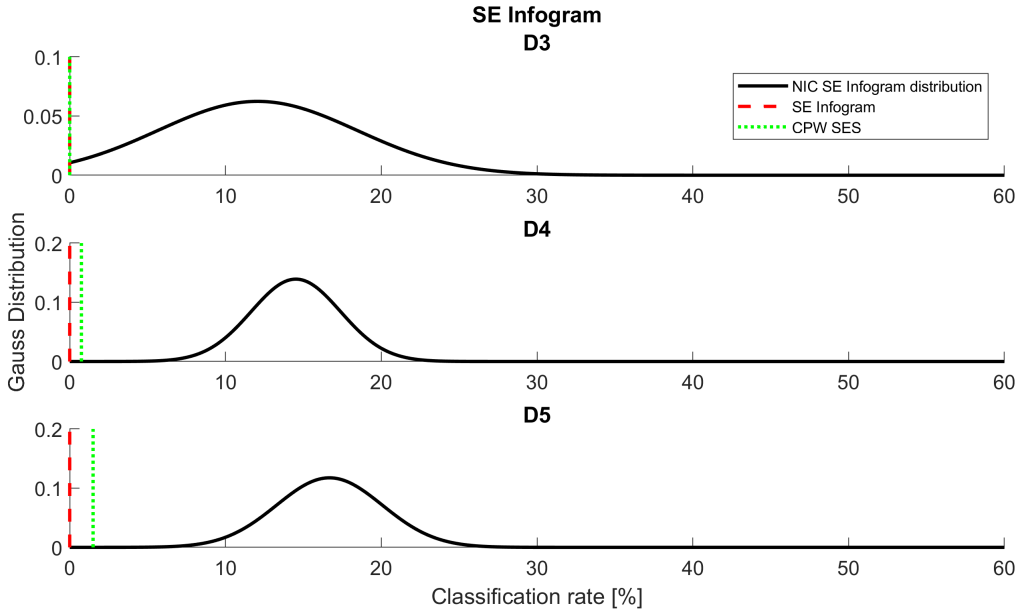


Figure 46: Comparison of the NIC SE Infogram distribution with the normal SE Infogram and popular CPW SES method.

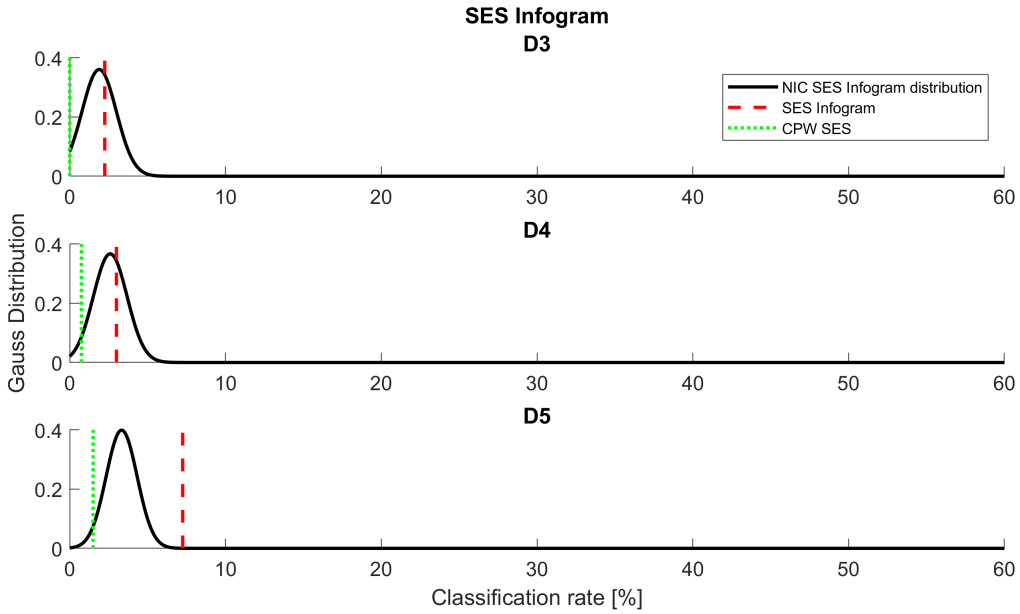


Figure 47: Comparison of the NIC SES Infogram distribution with the normal SES Infogram and popular CPW SES method.

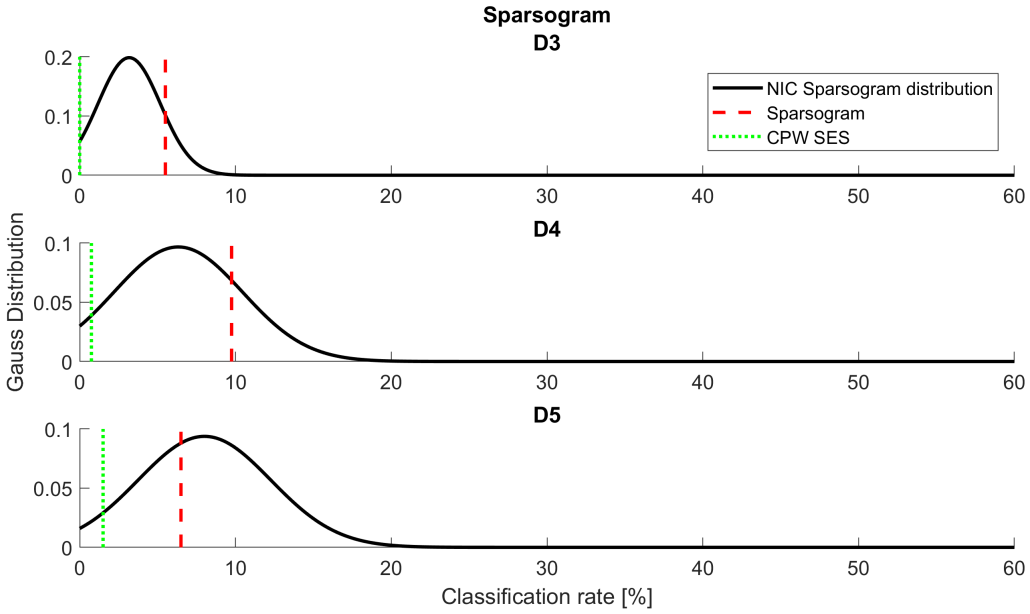


Figure 48: Comparison of the NIC Sparsogram distribution with the normal Sparsogram and popular CPW SES method.

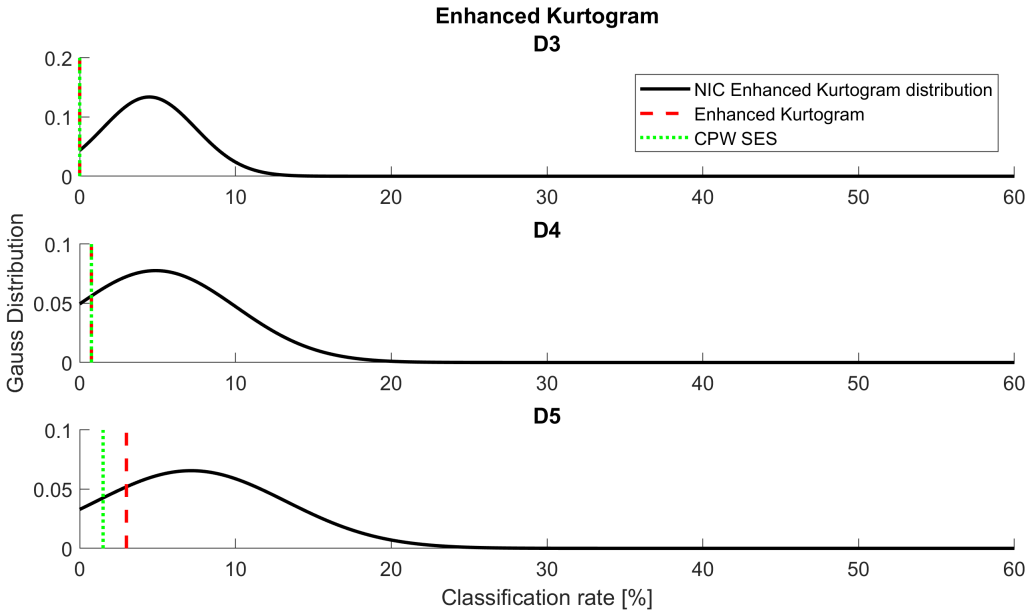


Figure 49: Comparison of the NIC Enhanced Kurtogram distribution with the normal Enhanced Kurtogram and popular CPW SES method.

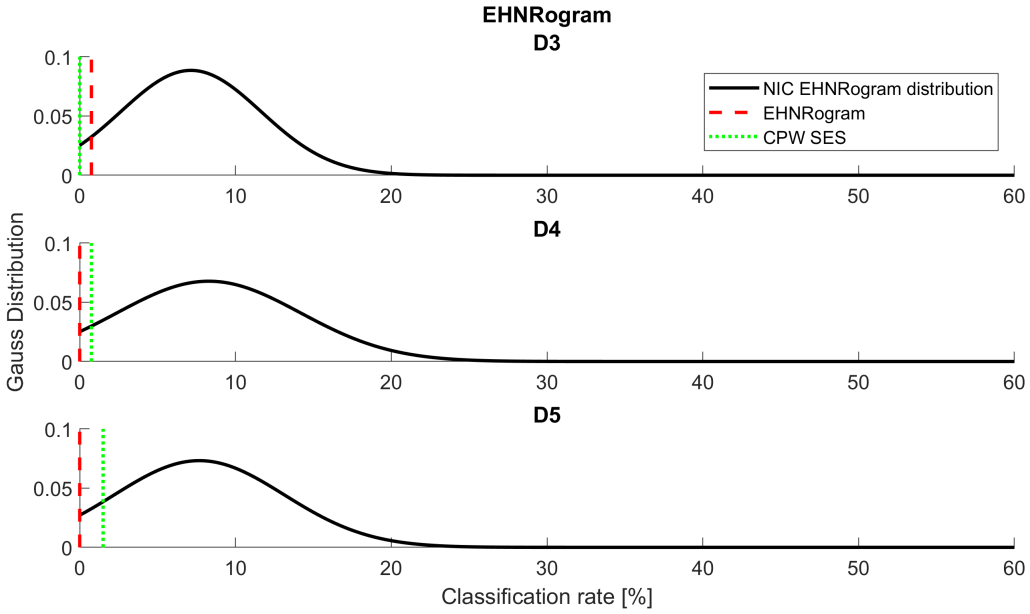


Figure 50: Comparison of the NIC EHNRogram distribution with the normal EHNRogram and popular CPW SES method.

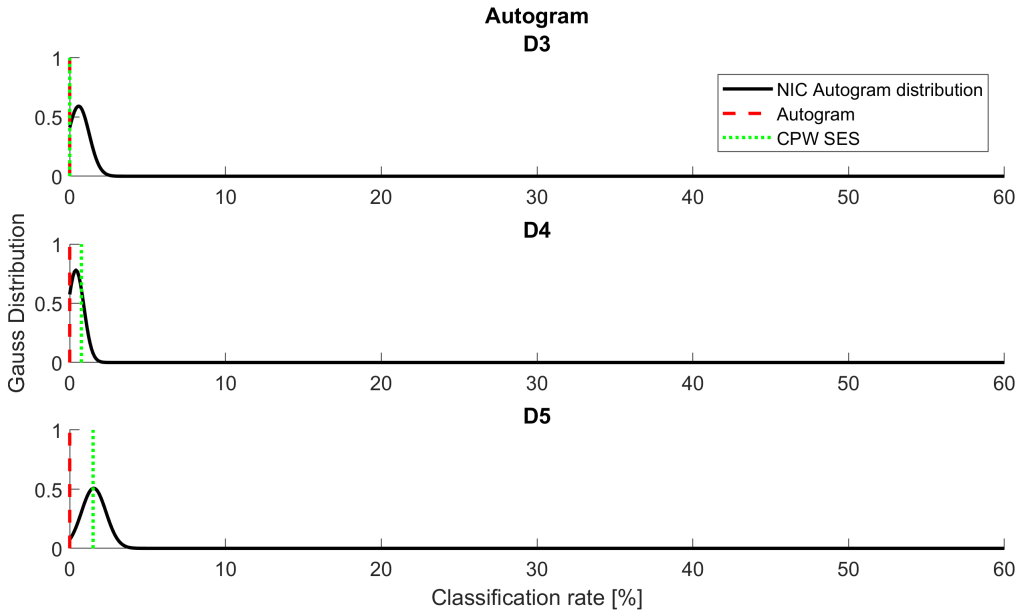


Figure 51: Comparison of the NIC Autogram distribution with the normal Autogram and popular CPW SES method.

To put the above results into perspective, it is important to note that the two existing methods of CPW SES and the conventional Gram procedure resulted in near zero percent classification for D3 on most of the Grams using the conventional methodology and that the damage at D3 is extremely small. Therefore, any classification greater than 10% can be considered significant.

It is interesting to note that the NIC Fast Kurtogram and NIC SE Infogram provide the best results of the 7 Grams under investigation. Likewise, the NIC Fast Kurtogram and NIC SE Infogram are the only Grams that calculate features in the temporal domain and not the frequency domain. Unless the damage is sufficiently large, features that are calculated in the spectral domain will be less pronounced due to spectral smearing under fluctuating speed conditions. As discussed by Abboud et al. (2019), a temporal analysis on a varying speed signal has exactly that effect of identifying transients and excited resonance frequencies. This explains why the Grams operating in the temporal domain are more successful at identifying damage under fluctuating speed.

4.7 Recommended diagnostic approach using NICogram methodology on fluctuating speed signals

Early diagnosis of incipient faults on fluctuating speed signals is best performed using either the NIC Fast Kurtogram or the NIC SE Infogram since both Grams use features that are calculated in the temporal domain. Furthermore, the NICogram methodology requires healthy signals under similar operating conditions as those of the signal which is under investigation. It is not possible to determine beforehand which healthy signal will yield the highest classification rate. Thus, it is recommended to use as many healthy signals as deemed sufficient by the user or as many healthy signals as are available to develop a distribution of classification results. From the distribution, an informed decision can be made as to whether or not a fault is present.

The NICogram methodology requires slightly more computing power than the conventional Grams procedure but yields a substantially higher classification rate for signals dominated by heavy background noise and deterministic components. Calculating a single NICogram takes twice as long as a conventional Gram since both a healthy and unclassified Gram need to calculated according to Figure 31. Calculation of the NICogram distribution time can be done according to,

$$t_{NIC} = t_{conventional} + n \times t_{conventional} \tag{32}$$

where $t_{conventional}$ is the time it takes to calculate a conventional Gram and n is the number of healthy signals used. Although the NICogram methodology takes slightly longer than the conventional Gram method, the process is fully automate-able because no a priori knowledge of a fault is required. This makes the NICogram methodology applicable for online fault diagnosis and outperforms the readily used conventional Gram methodologies.

5 Conclusion and recommendations

5.1 Conclusion

In this dissertation an automatic informative frequency band selection methodology was developed for use in envelope analysis under fluctuating operating conditions in the presence of strong noise and deterministic components. The proposed NICogram methodology was designed and compared to the popular Gram approach that makes use of features such as kurtosis, negentropy, sparsity, amongst many others, to identify incipient bearing faults. The method requires no a priori information regarding the fault and is applicable to any of the popular Grams. Initial results on a phenomenological model indicated that the proposed methodology can better identify damage at lower fault severity levels and that it is applicable on constant and fluctuating speed signals. Since the field of vibration diagnostics on constant speed signals is well researched, this dissertation paid special attention to fluctuating conditions.

An experimental gearbox setup was used to gather bearing fault data with discrete levels of damage on the outer race. Sufficient healthy data was captured for each operating condition as to critically analyse the method's sensitivity to the healthy dataset used for normalisation. It was found that even though the method is sensitive to the healthy datasets, using a distribution approach, a far higher classification rate could be achieved than the benchmark methods.

5.2 Recommendations

The following work is recommended as further research based on the content of this dissertation:

- Bearing fault detection in the presence of gear damage since gear damage often dominates vibration signals.
- A method for defining and selecting the optimal healthy signal to use for Gram normalization.
- A wider range of operating conditions.
- Selection of healthy datasets when exact operating conditions cannot be matched.
- Adapting the NICogram methodology for angle-domain methods in fluctuating speed conditions.

References

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

Abboud, D., Elbadaoui, M., Smith, W. A. & Randall, R. B. (2019), 'Advanced bearing diagnostics: A comparative study of two powerful approaches', *Mechanical Systems and Signal Processing* **114**, 604–627.

Antoni, J. (2005), 'Blind separation of vibration components: Principles and demonstrations', *Mechanical Systems and Signal Processing* **19**(6), 1166–1180.

Antoni, J. (2006), 'The spectral kurtosis: A useful tool for characterising non-stationary signals', *Mechanical Systems and Signal Processing* **20**(2), 282–307.

Antoni, J. (2007), 'Fast computation of the kurtogram for the detection of transient faults', *Mechanical Systems and Signal Processing* **21**(1), 108–124.

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

Antoni, J. (2016), 'The infogram: Entropic evidence of the signature of repetitive transients', *Mechanical Systems and Signal Processing* **74**, 73–94.

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

Antoni, J. & Randall, R. B. (2006), 'The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines', *Mechanical Systems and Signal Processing* **20**(2), 308–331.

Barbini, L., Eltabach, M. & du Bois, J. L. (2018), 'Application of cepstrum pre-whitening on non-stationary signals', *Applied Condition Monitoring* **9**, 275–283.

Barszcz, T. & JabŁoński, A. (2011), 'A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram', *Mechanical Systems and Signal Processing* **25**(1), 431–451.

Barszcz, T. & Randall, R. B. (2009), 'Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine', *Mechanical Systems and Signal Processing* **23**(4), 1352–1365.

Bechhoefer, E. & Kingsley, M. (2009), 'A review of time synchronous average algorithms', *Annual Conference of the Prognostics and Health Management Society, PHM 2009* pp. 1–10.

Bevilacqua, M. & Braglia, M. (2000), 'The analytic hierarchy process applied to maintenance strategy selection', *Reliability Engineering and System Safety* **70**, 71–83.

Borghesani, P., Pennacchi, P., Randall, R. B., Sawalhi, N. & Ricci, R. (2013), 'Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions', *Mechanical Systems and Signal Processing* **36**(2), 370–384.

Borghesani, P., Ricci, R., Chatterton, S. & Pennacchi, P. (2013), 'A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions', *Mechanical Systems and Signal Processing* **38**(1), 23–35.

Braun, S. (1975), 'The extraction of periodic waveforms by time domain averaging', *Acta Acustica united with Acustica* **32**(2), 69–77.

Brie, D., Tomczak, M., Oehlmann, H. & Richard, A. (1997), 'Gear crack detection by adaptive amplitude and phase demodulation', *Mechanical Systems and Signal Processing* **11**(1), 149–167.

Case Western Reserve University Bearing Data Center (2011), `http://csegroups.case.edu/bearingdatacenter/home`.

Dalpiaz, G., Rivola, A. & 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.

Davis, J. R. (2005), *Gear materials, properties, and manufacture*, ASM International.

D'Elia, G., Cocconcelli, M. & Mucchi, E. (2018), 'An algorithm for the simulation of faulted bearings in non-stationary conditions', *Meccanica* **53**(4-5), 1147–1166.

D'Elia, G., Daher, Z. & Antoni, J. (2010), A novel approach for the cyclo-non-stationary analysis of speed varying signals, *in* 'ISMA2010 International Conference on Noise and Vibration Engineering', pp. 2801–2814.

Dhamande, L. S. & Chaudhari, M. B. (2018), 'Compound gear-bearing fault feature extraction using statistical features based on time-frequency method', *Measurement: Journal of the International Measurement Confederation* **125**(January), 63–77.

Diamond, D. H., Heyns, P. S. & 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.

Doebling, S. W., Farrar, C. R., Prime, M. B. et al. (1998), 'A summary review of vibration-based damage identification methods', *Shock and vibration digest* **30**(2), 91–105.

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

El Badaoui, M., Guillet, F. & Dani, J. (2004), 'New applications of the real cepstrum to gear signals, including definition of a robust fault indicator', **18**, 1031–1046.

Feng, Z., Lin, X. & Zuo, M. J. (2016), 'Joint amplitude and frequency demodulation analysis based on intrinsic time-scale decomposition for planetary gearbox fault diagnosis', *Mechanical Systems and Signal Processing* **72-73**, 223–240.

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

Geropp, B. (1997), 'Envelope analysis - A signal analysis technique for early detection and isolation of machine faults', *IFAC Proceedings Volumes* **30**(18), 977–981.

Giurgiutiu, V. & Lingyu, Y. (2003), Comparison of short-time Fourier transform and wavelet transform of transient and tone burst wave propagation signals for structural health monitoring, *in* '4th International Workshop on Structural Health Monitoring', Stanford, CA, pp. 1–9.

Goyal, D. & Pabla, B. S. (2016), 'The vibration monitoring methods and signal processing techniques for structural health monitoring: A review', *Archives of Computational Methods in Engineering* **23**(4), 585–594.

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

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

Klein, R., Rudyk, E. & Masad, E. (2011), 'Methods for diagnostics of bearings in non-stationary environment', *The International Journal of Condition Monitoring* **2**(1), 562–573.

Lebold, M., McClintic, K., Campbell, R., Byington, C. & Maynard, K. (2000), Review of vibration analysis methods for gearbox diagnostics and prognostics, *in* 'Proceedings of the 54th meeting of the society for machinery failure prevention technology', Vol. 634, p. 16.

Lee, J., Abujamra, R., Jardine, A. K., Lin, D. & Banjevic, D. (2004), 'An integrated platform for diagnostics, prognostics and maintenance optimization', *Proceedings of the intelligent maintenance systems* pp. 15–27.

Lee, J., Qiu, H., Yu, G., Lin, J. & Rexnord-Technical-Services (2007), 'Nasa ames prognostics data repository', `https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/`. (Accessed on 09/26/2018).

Li, C., Cabrera, D., De Oliveira, J. V., Sanchez, R. V., Cerrada, M. & Zurita, G. (2016), 'Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram', *Mechanical Systems and Signal Processing* **76-77**, 157–173.

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

Liu, J., Wang, W. & Golnaraghi, F. (2008), 'An extended wavelet spectrum for bearing fault diagnostics', **57**(12), 2801–2812.

Mccormick, A. C. & Nandi, A. K. (1998), 'Cyclostationarity in rotating machine vibrations', *Mechanical Systems and Signal Processing* **12**(2), 225–242.

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. & 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.

McFadden, P. D. & Smith, J. D. (1985), 'A signal processing technique for detecting local defects in a gear from the signal average', *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **199**(4), 287–292.

McInerny, S. & Dai, Y. (2004), 'Basic vibration signal processing for bearing fault detection', *IEEE Transactions on Education* **46**(1), 149–156.

Moshrefzadeh, A. & Fasana, A. (2018), 'The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis', *Mechanical Systems and Signal Processing* **105**, 294–318.

Omar, F. K., Moustafa, K. A. F. & Emam, S. (2012), 'Mathematical modeling of gearbox including defects with experimental verification', *Journal of Vibration and Control* **18**(9), 1310–1321.

Parra, J. & Vicuña, C. M. (2017), 'Two methods for modeling vibrations of planetary gearboxes including faults: Comparison and validation', *Mechanical Systems and Signal Processing* **92**, 213–225.

Peng, Z. K. & 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.

Randall, R. B. (2017), 'A history of cepstrum analysis and its application to mechanical problems', *Mechanical Systems and Signal Processing* **97**, 3–19.

Randall, R. B. & Antoni, J. (2011), 'Rolling element bearing diagnostics - A tutorial', *Mechanical Systems and Signal Processing* **25**(2), 485–520.

Rubini, R. & Meneghetti, U. (2001), 'Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings', *Mechanical Systems and Signal Processing* **15**(2), 287–302.

Sawalhi, N. Ã. & Randall, R. B. (2008), 'Simulating gear and bearing interactions in the presence of faults Part I . The combined gear bearing dynamic model and the simulation of localised bearing faults', *Mechanical Systems and Signal Processing* **22**, 1924–1951.

Sawalhi, N. & Randall, R. B. (2011), 'Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size', *Mechanical Systems and Signal Processing* **25**(3), 846–870.

Sawalhi, N., Randall, R. B. & Endo, H. (2007), 'The enhancement of fault detection and

diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis', *Mechanical Systems and Signal Processing* **21**(6), 2616–2633.

Schmidt, S. & Heyns, P. S. (2020), 'Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring', *Measurement* **149**, 106964.

Schmidt, S., Heyns, P. S. & de Villiers, J. P. (2018), 'A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process', *Mechanical Systems and Signal Processing* **100**, 630–646.

Schmidt, S., Heyns, P. S. & Gryllias, K. C. (2019*a*), 'A discrepancy analysis methodology for rolling element bearing diagnostics under variable speed conditions', *Mechanical Systems and Signal Processing* **116**, 40–61.

Schmidt, S., Heyns, P. S. & Gryllias, K. C. (2019*b*), 'A pre-processing methodology to enhance novel information for rotating machine diagnostics', *Mechanical Systems and Signal Processing* **124**, 541–561.

Smith, W. A. & Randall, R. B. (2015), 'Rolling element bearing diagnostics using the Case Western Reserve University data : A benchmark study', *Mechanical Systems and Signal Processing* **1**(64-65), 100–131.

Stander, C. J. (2005), Condition monitoring of gearboxes operating under fluctuating load conditions, PhD thesis, University of Pretoria.

Stander, C. J. & 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., Heyns, P. S. & 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.

Stewart, R. (1977), Some useful data analysis techniques for gearbox diagnostics, PhD thesis, Institute of Sound and Vibration Research, Southampton University.

Tadina, M. & Boltezar, M. (2011), 'Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up', *Journal of Sound and Vibration* **330**(17), 4287–4301.

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

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

Troldborg, N. & Sørensen, J. (2014), 'A simple atmospheric boundary layer model applied to large eddy simulations of wind turbine wakes', *Wind Energy* **17**(April 2013), 657–669.

Tse, P. W. & Wang, D. (2013*a*), 'The automatic selection of an optimal wavelet filter and its enhancement by the new sparsogram for bearing fault detection: Part 2 of the two related manuscripts that have a joint title as "two automatic vibration-based fault diagnostic methods using the novel sparsity measurement—parts 1 and 2"', *Mechanical Systems and Signal Processing* **40**(2), 520–544.

Tse, P. W. & Wang, D. (2013*b*), 'The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "two automatic vibration-based fault diagnostic methods using the novel sparsity measurement - Parts 1 and 2"', *Mechanical Systems and Signal Processing* **40**(2), 499–519.

Urbanek, J., Barszcz, T. & Antoni, J. (2013), 'Time-frequency approach to extraction of selected second-order cyclostationary vibration components for varying operational conditions', *Measurement: Journal of the International Measurement Confederation* **46**(4), 1454–1463.

Vinson, R. G. (2014), Rotating machine diagnosis using smart feature selection under non-stationary operating conditions, PhD thesis, University of Pretoria.

Wang, D., Tse, P. W. & Leung, K. (2013), 'An enhanced Kurtogram method for fault diagnosis of rolling element bearings', *Mechanical Systems and Signal Processing* **35**(1-2), 176–199.

Wang, T., Han, Q., Chu, F. & Feng, Z. (2016), 'A new SKRgram based demodulation technique for planet bearing fault detection', *Journal of Sound and Vibration* **385**, 330–349.

Wang, T., Liang, M., Li, J., Cheng, W. & Li, C. (2015), 'Bearing fault diagnosis under unknown variable speed via gear noise cancellation and rotational order sideband identification', *Mechanical Systems and Signal Processing* **62-63**, 30–53.

Wang, W. Q., Ismail, F. & Golnaraghi, M. F. (2001), 'Assessment of gear damage monitoring techniques using vibration measurements', *Mechanical Systems and Signal Processing* **15**, 905–922.

Wang, W. & Wong, A. K. (2002), 'Autoregressive model-based gear gault diagnosis', *Journal of Vibration and Acoustics* **124**(2), 172.

Xu, X., Zhao, M., Lin, J. & Lei, Y. (2016), 'Envelope harmonic-to-noise ratio for periodic impulses detection and its application to bearing diagnosis', *Measurement: Journal of the International Measurement Confederation* **91**, 385–397.

Zheng, Y., Tay, D. B. H. & Li, L. (2000), 'Signal extraction and power spectrum estimation using wavelet transform scale space " ltering and Bayes shrinkage', *Signal Processing* **80**(8), 1535–1549.

Zimroz, R., Bartelmus, W., Barszcz, T. & 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.

Appendices

A Phenomenological model parameters

The model parameters that were used in the phenomenological model throughout this dissertation are shown in the following Tables 3-7. The notation and values used throughout this dissertation are similar to those used by Schmidt et al. (2019*a*) and result in realistic phenomenological signals.

Table 3: Impulse response function parameters.

	$f_{n,i}$ [Hz]	ζ_i
h_{dg}	2000	0.05
h_{rg}	3500	0.05
h_b	5000	0.05

Table 4: Deterministic component amplitude and phase parameters.

i	1	2	3	4	5	6	7	8	9	10
$A_{dg}^{(i)}$	1	1.5	2	1	0.5	0.3	0.2	0.1	0.1	0.05
$\varphi_{dg}^{(i)}$	0	0	0	0	0	0	0	0	0	0

Table 5: Random component amplitude and phase parameters.

i	1	2	3
$A_{rg}^{(i)}$	1	2	3
$\varphi_{rg}^{(i)}$	0	0	0

Table 6: Model parameters for the monotonic function of amplitude to speed dependence.

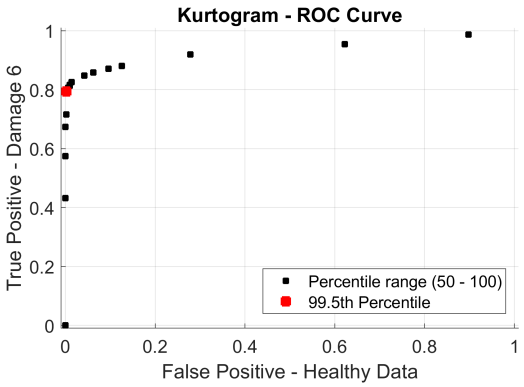
	a	b
$M_{dg}(w)$	1	0
$M_{dg}(w)$	1	0
$M_{dg}(w)$	1	0
$M_{dg}(w)$	1	0

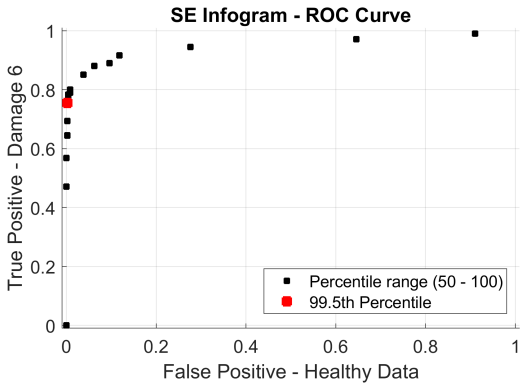
Table 7: Remaining values for the model parameters.

σ_n	0.1
σ_{rg}	1
$N_{t,g}$	40
F_{dam}	1
F_{const}	0.007958

B Threshold selection - ROC curves

It was required in this dissertation to compare the diagnostic capabilities of the conventional Grams to those improved by the NICogram methodology. The approach was treated as binary results, where the filtered signal was either classified as healthy or damaged through use of the bearing CF SNR metric as described in Section 4.3.1. A receiver operating characteristic curve, commonly known as an ROC curve, illustrates the change in diagnostic ability of a binary classifier as its discrimination threshold is changed. A very strict threshold will result in few signals correctly classified as damaged. On the other hand, a threshold that is set to low will result in many healthy signals incorrectly being classified as damaged. There lies some optimal threshold value that will result in a balance between the maximum true positive rate and minimum false positive rate as shown in the ROC curves below.





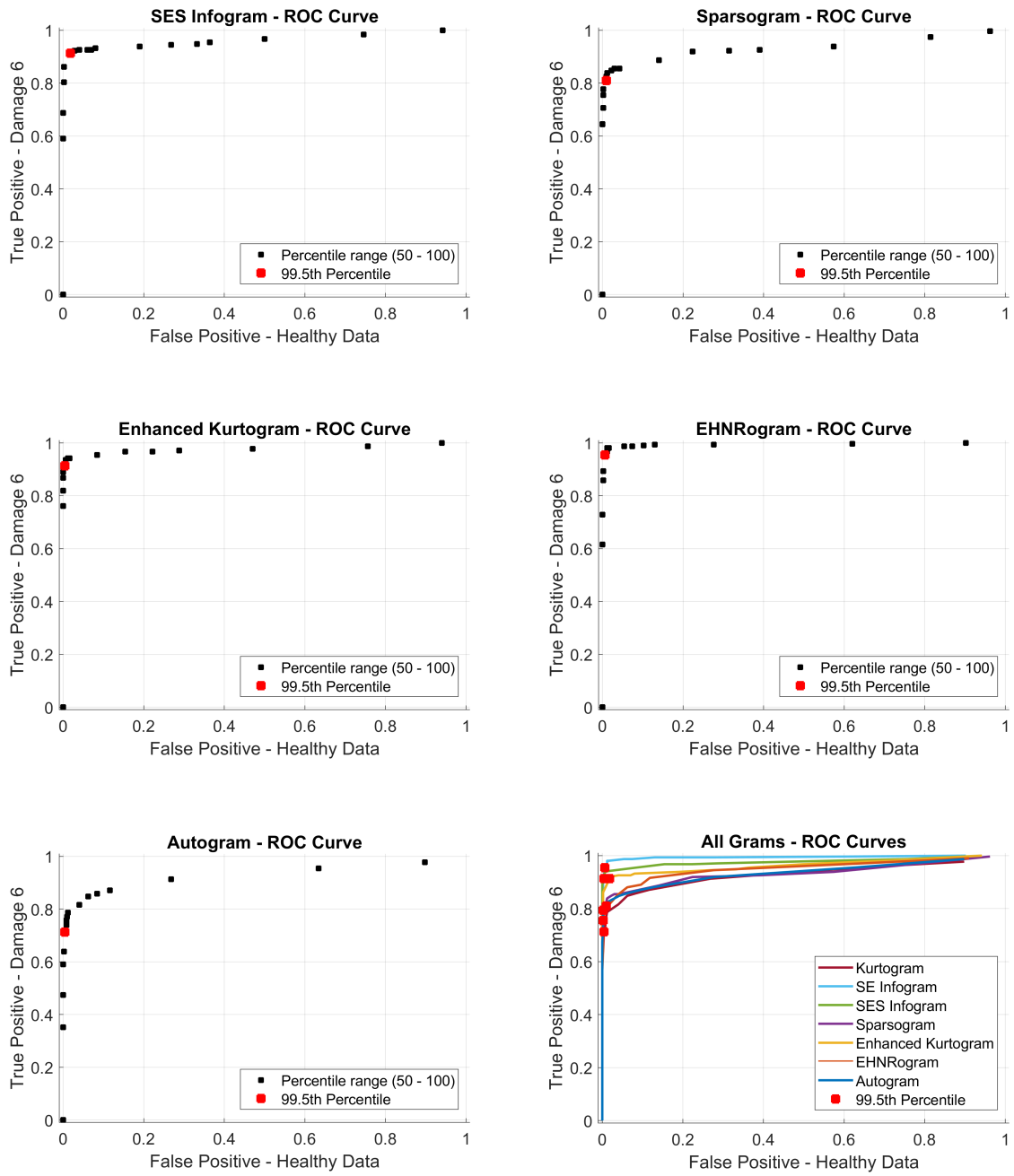


Figure 52: Threshold selection by finding the percentile with maximum true positive and minimum false positive rates for each Gram at both constant and fluctuating speeds.

From the graphs above it was found that the 99.5th percentile of the spectrum is a good choice for the threshold level. This ensures that a fair comparison between the different methods can be made. Using this selected threshold value will result in fewer false alarms, yet a maximum true positive rate for damage level 6. In the dissertation wherever experimental data were used, a threshold limit of the 99.5th percentile was used.

C Damage level 6 results

Figure 53 shows the relative performance of the conventional Grams, CPW SES, and NICograms calculated using each Gram and each of the healthy signals used for normalizing at damage level 6. As expected, the healthy signal that is used to normalize has a direct effect on the achievable accuracy. Each individual black bar represents the result of a NICogram that is normalized on a different set of healthy signals.

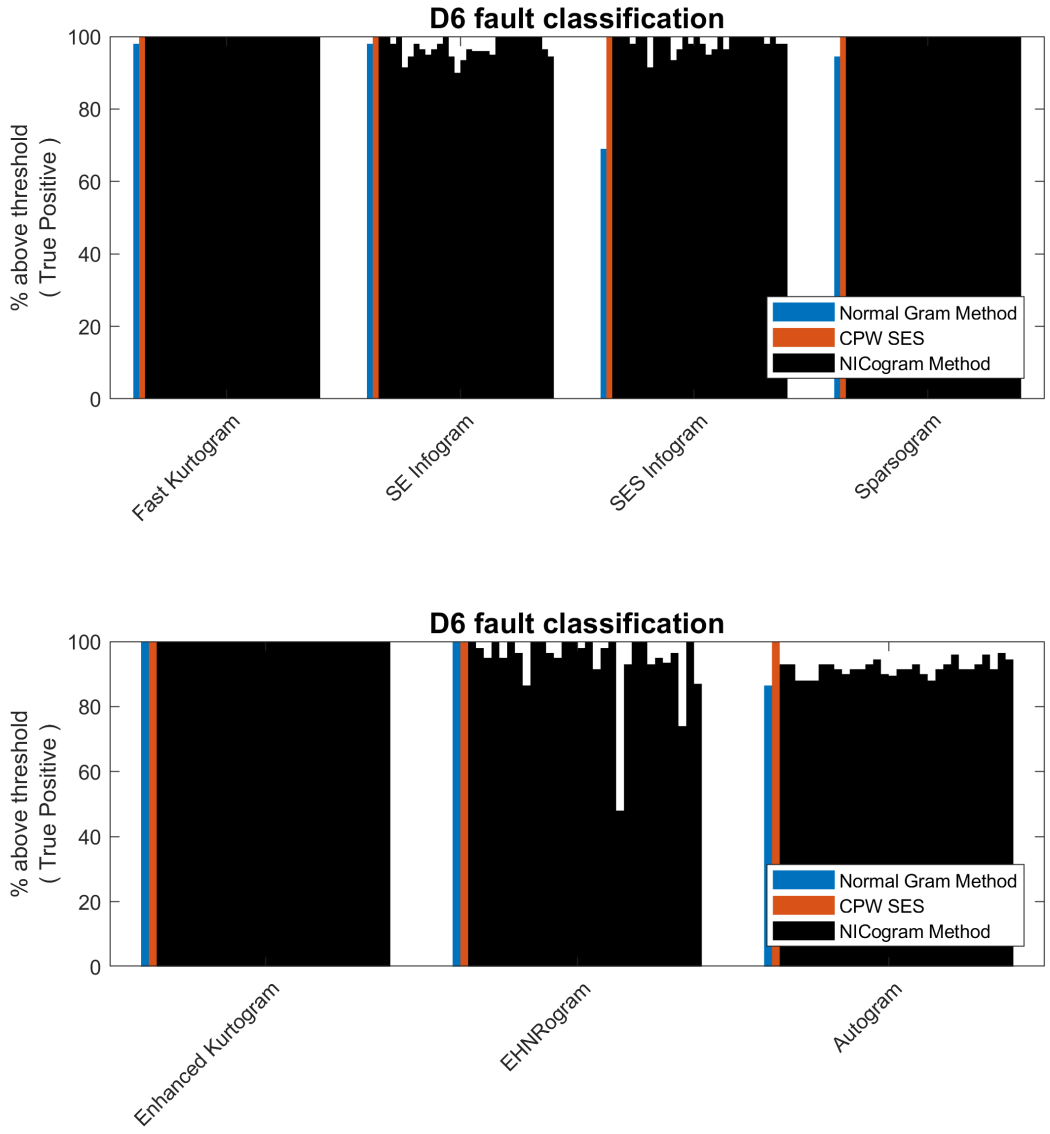


Figure 53: Damage detection at fault level 6.

The NICogram method applied to the Fast Kurtogram, Sparsogram and Enhanced Kurtogram all achieve 100% classification rate at damage level 6 regardless of the healthy dataset used. The other Grams also consistently achieve above 90% classification when used in the NICogram methodology. This data is not included in the body of this report because the damage at level 6 is so large that even the most trivial of methods should

achieve a high classification rate. This report is more interested in damage detection at small levels of damage such as levels 3-5.