Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring

Stephan Schmidt[a,*], P. Stephan Heyns[a]

[a] *Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa*

Abstract

Performing condition monitoring under time-varying operating conditions is challenging. The time-varying operating conditions result in amplitude and frequency modulation which mask the presence of incipient damage and make it difficult to distinguish between changes in the condition of the machine and changes in its operating conditions. In this work, the benefits of normalising the amplitude modulation caused by the varying operating conditions for condition monitoring are illustrated and a method is proposed to perform this normalisation. It is shown that the proposed method can be used as a pre-processing methodology for deterministic-random separation, it can be used to detect incipient damage and it can be used to reliably estimate the severity of the damage under time-varying operating conditions as well. The proposed method is investigated on numerical gearbox data and experimental gearbox data, where its benefits for condition monitoring under time-varying operating conditions are shown.

Keywords:

Amplitude modulation normalisation, Condition monitoring, Gearbox diagnostics, Time-varying operating conditions

1. Introduction

In a condition-based maintenance programme, processed condition monitoring data are used to infer the condition of the machine and to estimate its remaining useful life or the probability that the machine will fail before a specific time, e.g. typically the next maintenance period [1, 2]. In this process, vibration measurements are often processed to obtain intuitive representations of the condition of the components (e.g. synchronous average [3] and the envelope spectrum [4]) and condition metrics (e.g. Root-Mean-Square (RMS) [5] and the degree-of-cyclostationarity [6])

*Corresponding author.

Email address: stephan.schmidt@up.ac.za (Stephan Schmidt)

for diagnostics and prognostics.

However, vibration measurements and their indicators also contain information related to the operating conditions that were present during the measurement period. This is because, similarly to faults, time-varying operating conditions manifest in the vibration data as amplitude modulation and frequency modulation. This impedes the ability to distinguish between operating condition changes and machine condition changes when performing condition monitoring. The RMS for example is most frequently used for prognostics of rotating machinery [5, 7], however, it is also very sensitive to changes in operating conditions [8].

Condition monitoring under time-varying operating conditions has been studied for the past two decades. Baydar and Ball [9] found that spectral analysis did not perform well for gear diagnostics under varying load conditions and that the instantaneous power spectrum of the synchronous averaged signal performed much better. Stander et al. [3] found that the synchronous average performs well under non-cyclic stationary varying load conditions, but a pre-processing step, namely Load Demodulation Normalisation (LDN), needs to be performed to attenuate the amplitude modulation induced by cyclic stationary load conditions before the synchronous average can be used. Bartelmus and Zimroz [10] found that the relationship between a diagnostic metric and the rotational speed of a machine, estimated with a regression model, can be used to detect changes in the condition of the machine. Zimroz et al. [8] designed a diagnostic metric based on the regression coefficients, which captures the relationship between the features and the power generated by a wind turbine, and used it for gearbox condition monitoring. Urbanek et al. [11] found that the synchronous averaged instantaneous power spectrum is very effective for wind turbine bearing diagnostics under time-varying operating conditions. Abboud et al. [12–15] extended the classical time and angle-cyclostationary analysis techniques for cyclo-non-stationary signals and illustrated its efficiency on various datasets. Borghesani et al. [16] proposed a Cepstrum Pre-Whitening (CPW) procedure to attenuate the deterministic components from the signal, which enhances the Squared Envelope Spectrum (SES). The CPW is very easy to implement with no unknown parameters that need to be estimated from the data. The SES has more desirable properties when compared to the conventional envelope analysis method that utilises the Hilbert transform and is the preferred method for bearing diagnostics [14, 17].

The frequency modulation induced by varying speed conditions is typically compensated for by using computed order tracking [18] or tacholess order tracking [19, 20] techniques. However, varying operating conditions result in amplitude modulation as well [3, 13]. Stander et al. [3] illustrated the necessity of compensating for the amplitude modulation of cyclic stationary loads

when performing gear diagnostics. If the amplitude modulation is non-cyclic stationary, then it can be attenuated in a synchronous averaging process, with the number of required averages depending on the characteristics of the modulation [21]. Abboud et al. [22] developed a methodology that uses the Campbell diagram to attenuate the amplitude changes of different frequency components due to speed changes and referred to it as speed-spectral whitening. Abboud et al. [13] also noticed that the conventional synchronous average performs poorly when performing deterministic-random separation under time-varying operating conditions. It is very important to separate the deterministic and random components when performing envelope analysis, because the important diagnostic information for bearings manifests in the random part of the signal and the detection of bearing damage is impeded by the deterministic components [17]. Therefore, Abboud et al. [13] proposed the Generalised Synchronous Average (GSA) to alleviate the effect of the amplitude modulation due to varying speed conditions. The GSA, CPW and the improved synchronous average are compared by Abboud et al. [14]. The authors found that both the GSA and CPW can perform well when enhancing the squared envelope spectrum, with the appropriate choice depending on the dataset [14]. Schmidt et al. [23] proposed a discrepancy signal normalisation procedure for varying speed conditions. The mean and standard deviation of a healthy discrepancy signal, conditioned on the rotational speed of the system, are used to standardise a discrepancy signal to obtain a discrepancy signal that is robust to varying speed conditions, while still being sensitive to damage.

Therefore, amplitude modulation not only impedes the ability to perform deterministic-random separation, but also results in features that are inherently sensitive to varying operating conditions. In this paper, it is proposed that a Normalisation of the Amplitude Modulation caused by Varying Operating Conditions (NAMVOC) procedure needs to be used to alleviate the adverse influences of time-varying operating conditions. It is also recognised in this work that the LDN method [21] can be used as a NAMVOC procedure, but it is shown that it also attenuates the diagnostic information in the signal. Therefore, a new NAMVOC method is proposed for diagnostic and prognostics under time-varying operating conditions. In summary, the following contributions are made:

- The importance of using NAMVOC when performing condition-based maintenance under time-varying operating conditions is emphasised.
- The LDN procedure is revisited in this work and identified as a NAMVOC method. Its shortcomings for diagnostics and prognostics are also highlighted.

- A new NAMVOC method is proposed and its suitability for condition monitoring is compared to the LDN method.
- The benefits of using the proposed NAMVOC method are emphasised for deterministic-random separation, incipient fault detection and feature extraction for diagnostics and prognostics under time-varying operating conditions. An automatic novelty detection algorithm is also implemented and the benefits of the proposed method are shown.
- The proposed NAMVOC method is investigated and compared to the CPW for diagnostics.

The layout of the paper is as follows: In Section 2, an overview of condition monitoring under time-varying operating conditions is given and the proposed NAMVOC method is presented. Thereafter, the proposed NAMVOC method is investigated on numerical gearbox data in Section 3 and on experimental gearbox data in Section 4. Finally, the work is concluded and recommendations are made in Section 5. In Appendix A, supporting information regarding NAMVOC methods is given and additional information relevant to the numerical gearbox model is presented in Appendix B.

2. Condition monitoring under time-varying operating conditions

In this section, condition monitoring under time-varying conditions is discussed with the focus specifically placed on the amplitude modulation effects. The proposed NAMVOC method is motivated from the LDN perspective and some similar techniques such as CPW and the GSA are discussed as well.

2.1. Problem statement

It is assumed that the measured vibration signal

$$y(t) = M(\omega(t), l(t)) \cdot x(t), \tag{1}$$

can be decomposed in terms of a strictly positive function M, which is a function of operating condition variables such as the speed $\omega(t)$ and load $l(t)$ on the system, and $x(t)$ is for example a raw stationary or cyclostationary vibration signal . The modulation function $M(t)$, with the dependence of the operating condition variables removed for notational simplicity, is a time-varying function under non-stationary operating conditions and therefore $y(t)$ is generally cyclo-non-stationary. This has the following consequences:

- The assumption of cyclo-ergodicity is generally invalid and therefore the synchronous average is a poor representation of the first-order cyclostationary components in the signal. Its adverse influence on the deterministic-random separation task is illustrated by Abboud et al. [13, 14]. Hence, it impedes the ability to perform bearing fault diagnosis [14, 17].

- Time-varying operating conditions can mask weak damage components. This is illustrated by Stander et al. [3] for gear diagnostics under varying load conditions.

- It may be possible to detect damage under time-varying operating conditions, but it is not necessarily possible to quantify the damage or perform effective prognostics, i.e. changes in the condition indicator may indicate a change in machine condition, changes in operating conditions or both.

The latter point is illustrated in Figure 1 where the value of the diagnostic metric, denoted A is used to estimate the fault severity (e.g. crack length) of the damage under two operating states. If the diagnostic metric is sensitive to operating conditions, it can be difficult to distinguish between changes in operating conditions and changes in machine condition as shown by the multi-modal distribution $p(FS|DM = A)$ in Figure 1(i). If the diagnostic metric is robust to changes in operating conditions, then it is easier to infer the condition of the machine as shown in Figure 1(ii) with $p(FS|DM = A)$. The conditional distribution $p(DM|FS)$ can be used to determine

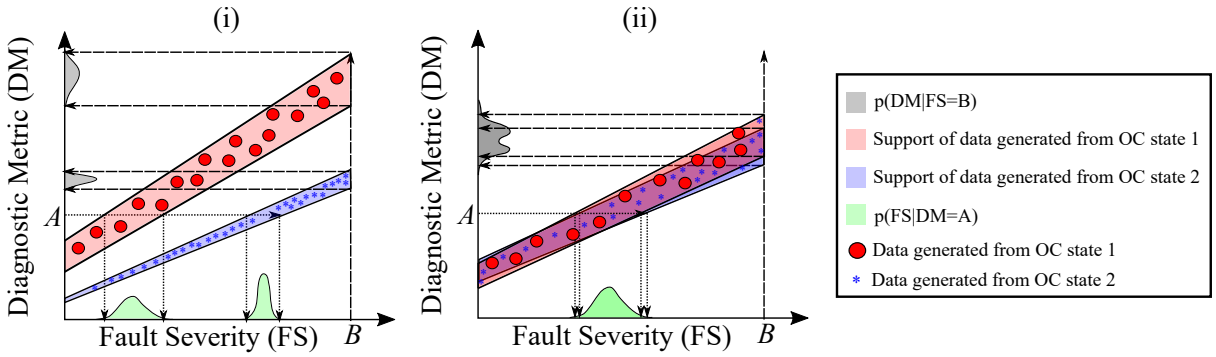

Figure 1: An illustration of the ambiguities that result from diagnostic metrics that are sensitive to changes in operating conditions (i) and the benefits of using diagnostic metrics that are robust to changes in operating conditions (ii) for diagnostics and prognostics. For the purposes of this example, it is assumed that the machine can operate in two distinct states, e.g. high load and low load states.

the sensitivity of the diagnostic metric to changes in operating conditions. If the conditional variance is too large, then it may not be possible to effectively detect changes in condition of the machine. If multi-modal behaviour is also observed for $p(DM|FS)$ (i.e. if multiple distinct

clusters are formed for measurements acquired from a machine in the same condition), then it is also indicative of the diagnostic metric being sensitive to changes in operating conditions.

It is possible to model the conditional distribution $p(FS|DM, OC)$ for more accurate fault diagnosis by using historical data, however, a more general method for condition monitoring is investigated instead. If the modulation function $M(t)$ could be estimated unsupervised and automatically removed from the signal, it would not only be beneficial for attenuating the influence of cyclic stationary operating conditions [3], but also for diagnostics and prognostics in general.

In this work, NAMVOC methods are proposed to attenuate the adverse influence of time-varying operating conditions. An overview of the application of a NAMVOC method is presented in Figure 2.

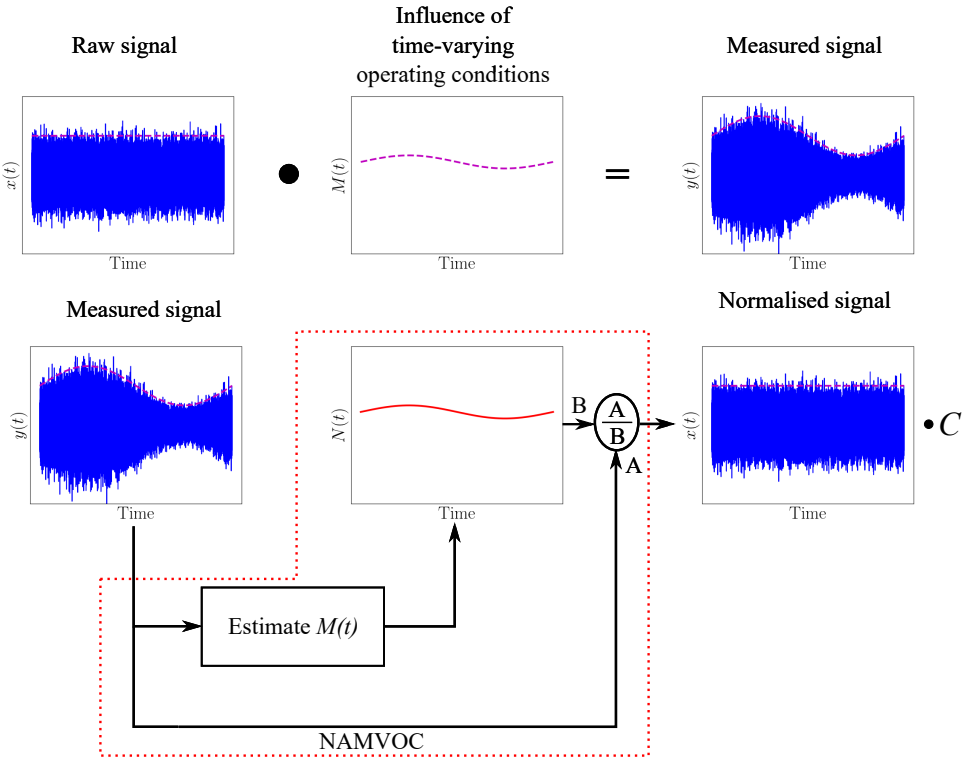

Figure 2: Illustration of the influence of time-varying operating conditions and the results obtained from a NAMVOC procedure. The amplitude modulation function caused by the time-varying operating conditions is denoted $M(t)$ and the estimated amplitude modulation function is denoted $N(t)$. The calibration constant C indicates that the normalised signal may be a scaled version of the raw signal.

In the next section, the LDN method is investigated as a NAMVOC method for condition monitoring under time-varying operating conditions.

2.2. Load Demodulation Normalisation (LDN) as a NAMVOC method

LDN is proposed by Stander et al. [3] to attenuate the modulation induced by cyclic stationary loads, i.e. loads varying synchronously with the rotation of the monitored gear. Even though it was proposed for cyclic stationary loads, it can be used as a general NAMVOC procedure as well. The LDN procedure is summarised in Figure 3. The amplitude modulation is estimated

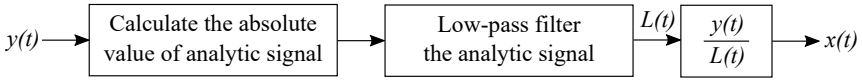

Figure 3: The LDN procedure proposed by Stander et al. [3] is shown without the filter passband optimisation process. The raw signal is denoted $y(t)$, the normalised signal is denoted $x(t)$ and the load demodulation normalisation function is denoted $L(t)$.

from the signal, whereafter it is filtered to extract the load information. The filtering step is important to ensure that the amplitude modulation induced by the varying load is separated from the amplitude modulation induced by potential gear damage. This is possible, because the cyclic frequency of the load is usually much lower than the cyclic frequency of the damage. Due to the importance of the filter passband selection, an optimisation procedure to select the passband is proposed by Stander et al. [3] as well.

In Appendix A.1, a mathematical investigation is performed which illustrates that the LDN is a biased estimator; it removes the modulation function due to the varying operating conditions, but it also scales the signal with a data-dependent factor (i.e. C in Figure 2) and it attenuates the diagnostic information as well. The amount of diagnostic information attenuation depends on the magnitude of the impulses caused by the damage and can impede effective fault trending and prognosis as shown in Appendix A.2. This is an undesirable property for a signal processing tool in the condition monitoring field and therefore an improvement is required.

2.3. Proposed NAMVOC method

A new procedure is proposed to alleviate the diagnostic information attenuation effect of the LDN. As shown in Appendix A.1, the LDN procedure can also be implemented by a moving average windowing procedure, which is a type of finite impulse response filter. However, the mean statistic is sensitive to outliers, which makes it sensitive to the presence of impulses in the signal and result in the diagnostic information to be attenuated. The median function is much more robust to outliers and is therefore more appropriate for use in a NAMVOC method. Therefore, the proposed method uses a moving median window instead of a conventional finite impulse response filter. The new amplitude modulation procedure is presented in Figure 4. The

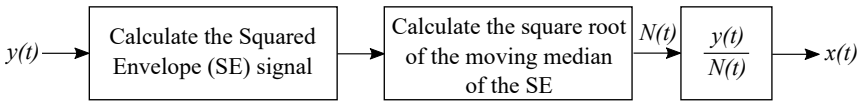

Figure 4: The proposed NAMVOC procedure. The normalising function is denoted $N(t)$.

proposed method is compared to the LDN method in Appendix A.2 on synthetic data, where it is shown that the proposed method is better suited for condition monitoring applications.

2.4. Implementation of the proposed method

In the estimation of the modulation function, a moving median window is applied to the squared signal, i.e. $|y|^2$, with a window length of ΔT_w seconds and an overlap of ΔT_o seconds. This windowing process results in the estimated modulation signal to have a lower sampling frequency than the vibration signal. Hence, the sampling frequency of the estimated modulation signal is increased by linearly interpolating the original estimated modulation signal at the time steps of the vibration signal under consideration to ensure that the two signals are compatible.

In terms of the two unknown parameters: Choosing the appropriate window length is important to ensure that the modulation function is properly estimated. If the window length is too short, the power of the impulses and other signal components is estimated and therefore the normalisation process would attenuate the important information for diagnostics. However, if the window length is too long, the modulation function may change too much within the considered window length. This means that the modulation function due to the time-varying operating conditions would not be adequately removed. The influence of the window length on the estimated local statistics of a signal is investigated and shown in Ref. [23]. Hence, the window length should be longer than the longest period of the signal components to ensure that the modulation function and not the cyclostationary statistics of the signal is estimated. However, in the windowing process, it is assumed that the modulation function is constant for the duration of the window, and therefore it should be sufficiently short to properly estimate the modulation function. The linear interpolation process, which can easily be implemented with Lagrange interpolating polynomials, alleviates the adverse effects of the constant modulation function assumption.

The window length should also be significantly longer than the time duration of impulses to ensure that the calculated median is unaffected by their presence. The latter restriction can easily be achieved, because the duration of the impulses are usually quite short because of damping in the system. The window overlap is chosen to be $\Delta T_o = 0.9 \cdot \Delta T_w$ throughout this work to ensure

that the estimated modulation function is smooth.

The optimal window length depends on the operating conditions as well as the system that is under consideration. It is possible to use the filter optimisation procedure which Stander et al. [3] used in their work on gear diagnostics, however, the following procedure can also be used as a general method:

1. For each candidate window length, calculate the variance of the localised Median Absolute Deviation (MAD) of the normalised signal. If there is a large variance between the MAD of different segments of the signal, it means that the modulation function is not properly estimated. The MAD is used instead of the standard deviation, to make the cost function more robust to outliers caused by damage.
2. Select the window length that corresponds to the local minima with the shortest window length for calculating the normalising signal.

An example of this process is shown in Figure 5. This was applied to two measurements of the experimental data discussed in Section 4 with the cost function (i.e. variance of the localised MAD) and the optimal solution shown. The procedure needs to be carefully applied to damaged

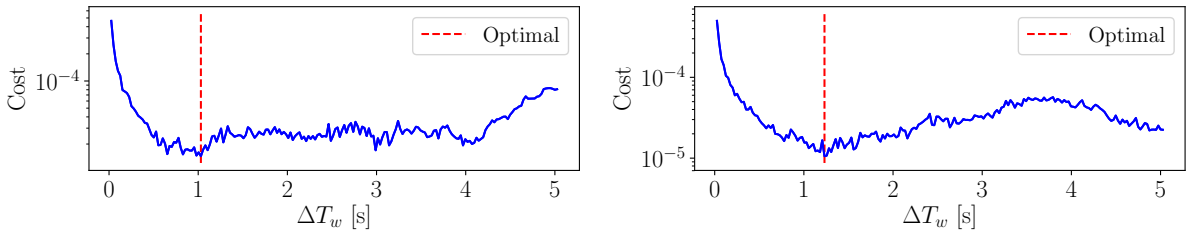

Figure 5: An example of the variance of the localised Median Absolute Deviation (MAD) of the normalised signals, referred to as the cost, is shown for two signals from Section 4. The optimal window length ΔT_w is shown as well.

data, since localised impulses can influence the cost function. The authors found that a time window of $\Delta T_w = 1$ seconds works quite well for many signals that were considered and is used throughout this work. It is significantly longer than the longest period in most signals and with the linear interpolation process, it is sufficiently short to estimate the varying modulation function well. It would be possible to scale the signal with the calibration constant, i.e. C in Figure 2, to ensure that it has the appropriate magnitude to compare to a standard for example, but for the purposes of these investigations $C = 1$.

2.5. Benefits of the proposed methodology

The proposed method is also well suited for general varying operating conditions (e.g. uncoupled load and speed variation), because it does not make any assumptions on the cause of the modulation during the estimation procedure. This is a beneficial property when compared to the methods in Refs. [22, 23] which explicitly segment the signals in terms of speed. It is also significantly simpler than using a regression model to estimate $M(t)$. This is because all variables affecting the modulation function (e.g. load, speed, temperature) need to be known a priori in a regression model and the non-linear relationship between the operating variables and the modulation function needs to be learned based on the available historical data.

One of the applications of this technique is deterministic-random separation under time-varying operating conditions. The GSA and CPW methods are specifically used to remove the deterministic components and to enhance the Squared Envelope Spectrum (SES) for detecting random components under varying speed conditions. If the proposed NAMVOC method is used prior to estimating the residual signal, the conventional synchronous average can be used for the deterministic-random separation task as opposed to the GSA. Even though it will suffer from most of the same shortcomings as the GSA [14], it overcomes the compromise between the resolution of the operating condition regimes and the number of cycles available for the estimation of the synchronous average of each regime in the GSA. It is also then easier to implement and the features extracted from the processed signal are more robust to varying operating conditions.

In the following investigations, the ability of the proposed method for deterministic-random separation and its fault trending capabilities are compared to the CPW.

3. Phenomenological gearbox dataset

In this section, the suitability of the proposed NAMVOC method for deterministic-random separation and diagnostics is investigated on numerical gearbox data and compared to the results obtained with the raw signal (i.e. without performing the normalisation procedure) and the cepstrum pre-whitened signal.

3.1. Overview of model

The phenomenological gearbox model proposed by Abboud et al. [14] is used in this section to evaluate the proposed procedure in a controlled environment. The measured signal

$$y(t) = y_{dg}(t) + y_{rg}(t) + \beta \cdot y_b(t) + y_n(t) \tag{2}$$

is decomposed in terms of a deterministic gear component $y_{dg}(t)$, a random gear component attributed to distributed damage of the pinion $y_{rg}(t)$, a bearing component attributed to localised outer race damage $y_b(t)$, an impulse magnification factor which increases the contribution of the bearing component β and a broadband noise component $y_n(t)$. The detailed equations of the signal components are given by Equations (B.1) to (B.4) in Appendix B.

Three speed profiles, are used to generate the data in this section, with random offsets added to each measurement. The rotational speed samples are presented in Figure 6(i) with the different profiles clearly indicated. Each signal component in Equation (2) also contains an explicit modulation function that is used to simulate the amplitude modulation induced by the varying operating conditions. The amplitude modulation functions used to generate the data in Figure 6(i) are presented in Figure 6(ii).

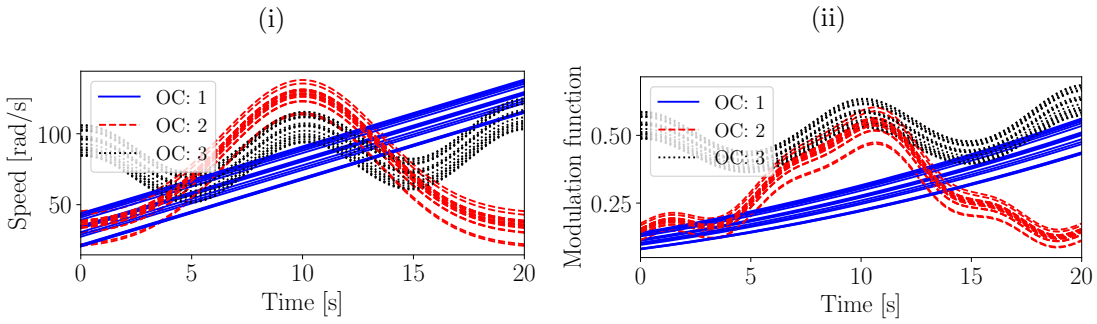

Figure 6: The rotational speed $\omega(t)$ (Figure 6(i)) and the corresponding modulation function (Figure 6(ii)) are presented for the three Operating Conditions (OCs). Equations (B.5) to (B.7) are used to generate the speed profiles and Equations (B.8) to (B.10) are used to generate the modulation functions.

The modulation functions of operating conditions 1 and 3 have one-to-one mappings with the corresponding rotational speed, however, they have different values for the same speed. This can for example simulate a case where a higher load is present for operating condition 3 than operating condition 1. Operating condition 2 does not have a one-to-one mapping, which means that the performance of techniques relying on the segmentation of the data in terms of speed (e.g. Schmidt et al. [23] and Abboud et al. [13]) will be adversely influenced.

The measured vibration signal and its components in Equation (2) were generated for a single realisation of operating condition 2 and shown in Figure 7 with the actual rotational speed used. The significant amplitude modulation induced by the varying operating conditions can clearly be seen in each signal component and will adversely affect the deterministic-random separation task and fault detection capabilities if the proposed NAMVOC method is not used.

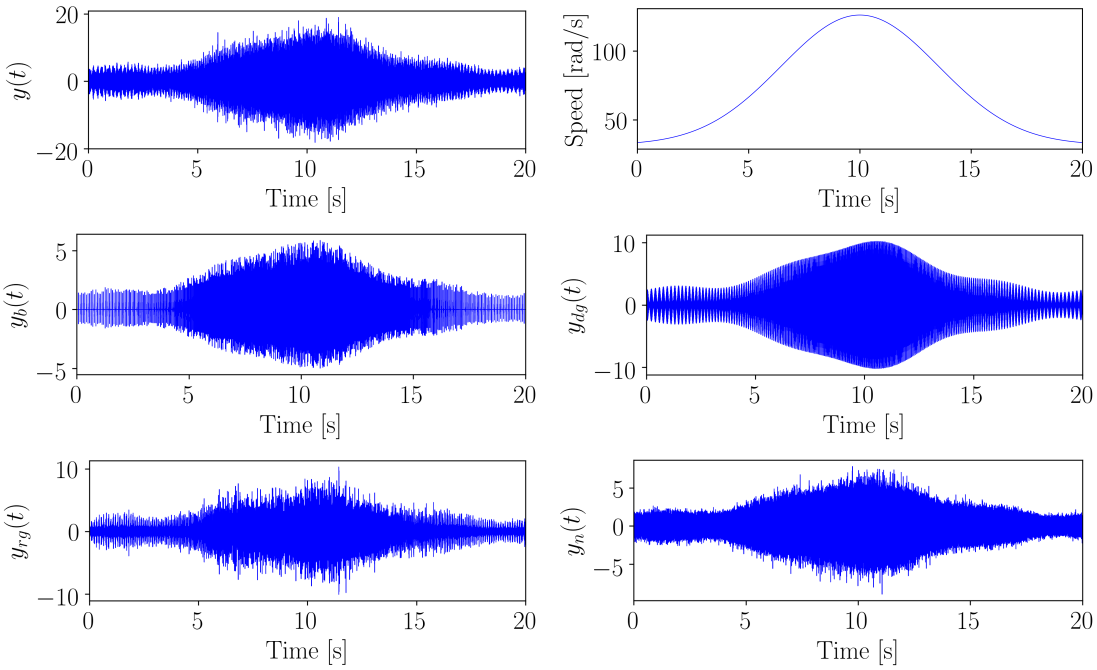

Figure 7: The different signal components of the phenomenological gearbox model are shown for a single realisation of operating condition 2.

3.2. Deterministic-random separation

The purpose of this investigation is to compare the residual signal of the original signal to the residual signal of the normalised signal. The residual signal of the angle domain signal $y(\varphi)$ is calculated by subtracting its synchronous average

$$\bar{y}(\varphi_0; \Phi_0) = \frac{1}{K} \sum_{k=0}^{K-1} y(\varphi_0 + k \cdot \Phi_0), \text{ with } \varphi_0 \in [0, \Phi_0), \tag{3}$$

from the signal. The angle of rotation is denoted by φ, the angular period is denoted by Φ_0 (e.g. gear rotation period). The populated synchronous average refers to the synchronous average, which has a domain of $\varphi_0 \in [0, \Phi_0)$, that is populated to have the same domain as the signal y.

In Figure 8, the populated synchronous averages of two operating conditions are superimposed on the raw signals y. The discrepancy between the amplitudes of the raw signals and their synchronous averages is evident over the time duration. This is attributed to the time-varying operating conditions. This is in contrast to the normalised signals, where the effect of the varying operating conditions cannot be seen in the amplitudes of the signal $x(t)$. This means that the amplitude difference of the signal $x(t)$ and its synchronous average remains consistent over time.

The SES is one of the most powerful techniques for rotating machine diagnostics [14, 24] and is used in this work to investigate the benefits of using the normalisation method before

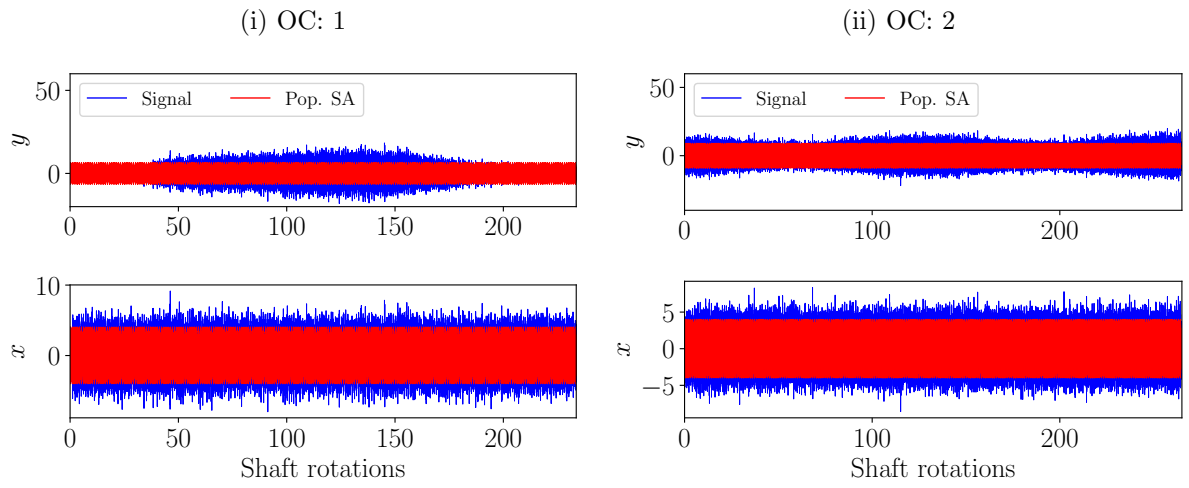

Figure 8: The raw vibration signal $y(t)$ and its populated synchronous average and the normalised vibration signal $x(t)$ and its populated synchronous average are presented for two Operating Conditions (OCs).

calculating the residual signal. The SES of the original signal (i.e. without any pre-processing) is presented in Figure 9 with the shaft rotational components, the fundamental distributed gear damaged component, the gear mesh components and the fundamental Ball-Pass Order Outer race (BPOO) components indicated for clarity. The deterministic gear component is prominent

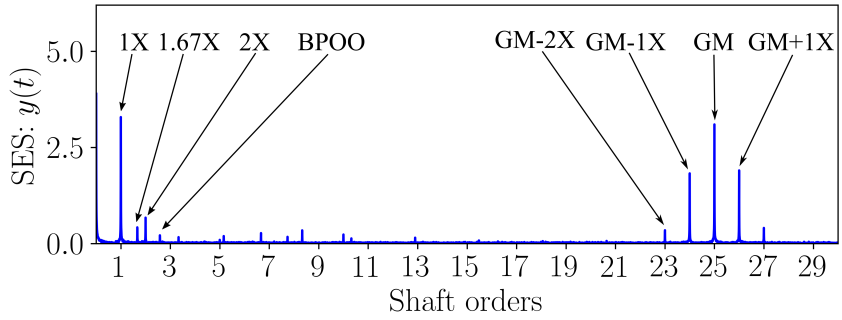

Figure 9: The SES of the original vibration signal is presented for operating condition 1. The following abbreviations are used: GM: Gear Mesh, BPOO: Ball-Pass Order Outer race and 1X refers to the fundamental shaft order of the gear and 1.67X refers to the distributed gear damage component on the pinion.

in the SES and can make it difficult to detect the bearing damage. Therefore, the SES of the residual signal is investigated to attenuate the deterministic components and to retain the random components attributed to the random gear component and the bearing damage component.

In Figures 10, the SES of the residual signal of the raw signal is shown, the SES of the residual signal, obtained by first normalising the signal with the proposed NAMVOC method, and the SES of the cepstrum pre-whitened signal are compared. The CPW signal is calculated with the

procedure described in Ref. [16], where the CPW is performed in the order domain to remove the deterministic gear components.

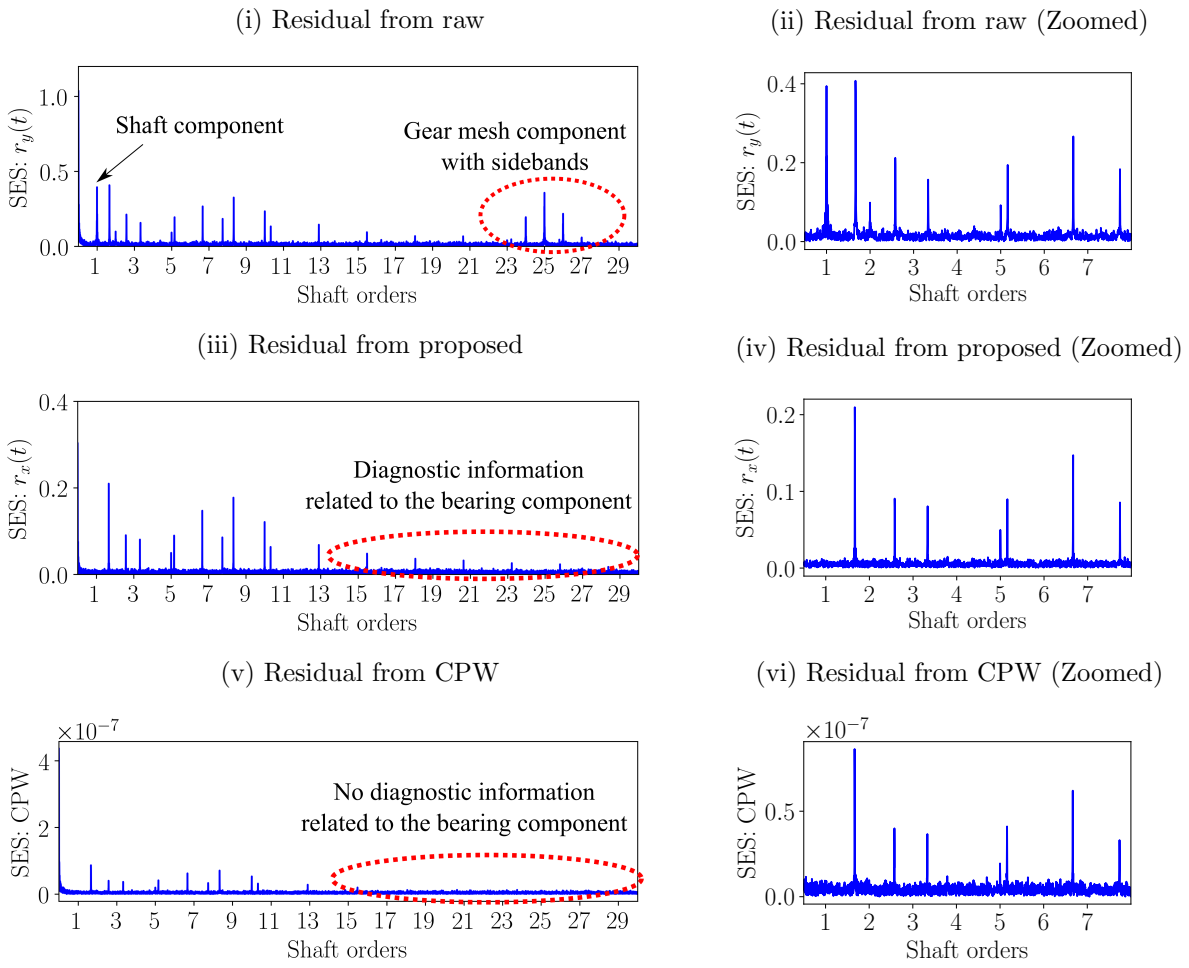

Figure 10: The Squared Envelope Spectra (SES) of the residual signal calculated of the raw vibration signal $r_y(t)$, the SES of the residual signal obtained from the normalised vibration signal $r_x(t)$ and the SES of the whitened signal obtained with the Ceptrum Pre-Whitening (CPW) signal are presented.

The deterministic components due to the gear damage is still very prominent in Figures 10(i) and 10(ii). This is attributed to the fact that the synchronous average is a poor estimate of the first-order cyclostationary components in the raw signal and therefore they cannot be properly removed. This is in contrast to the results obtained with the proposed NAMVOC method and the CPW procedure; the deterministic component due to the gear damage at 1 shaft order and the gear mesh component at 25 shaft orders are properly attenuated and the random signals are clearly observed. It is also seen that bearing components are more prominent in Figure 10(iii) compared to the results in Figure 10(v). This is especially seen at the higher orders (e.g. the harmonic at 12.9 orders and the higher harmonics). This is attributed to the whitening procedure

of the CPW where some of the diagnostic information can be distorted and low signal-to-noise ratio bands can be amplified [14].

The fault diagnosis capabilities of the proposed method using diagnostic metrics are also investigated in the next section.

3.3. Fault diagnosis with diagnostic metrics

It is very important to be able to detect faults early in the degradation process. However, it is also important to reliably estimate changes in the severity of the condition of the machine and to ultimately estimate the remaining useful life of the components. As illustrated in Figure 1, varying operating conditions could result in false alarms to be triggered and also adversely affect the accuracy of the fault severity estimate. In this section, the suitability of using the proposed NAMVOC method is investigated.

The Root-Mean-Square (RMS) and the kurtosis of the raw and the normalised signals are presented in Figure 11 over the impulse magnification factor for signals generated from the different operating conditions. These features are used because they are very popular condition indicators in the rotating machine diagnostics field [5].

The diagnostic metrics in Figures 11(i) and 11(ii) are very sensitive to the different operating condition profiles, which can easily be seen by the two clusters that are formed for an impulse magnification factor β value. This is very similar to the behaviour seen in Figure 1(i), where a bimodal distribution is seen for $p(DM|FS = B)$. The implication of this is that it is only possible to perform fault trending reliably if the operating condition profile is the same or result in metrics that are statistically similar for each measurement in the monitoring process.

The diagnostic metric obtained with the LDN NAMVOC method in Figure 11(iii) also performs very poorly. It is very sensitive to changes in operating conditions, while being insensitive to changes in the impulse magnification factor. This is attributed to the attenuation of the diagnostic information described in Appendix A.1 and shown in Appendix A.2. The kurtosis of the LDN signal performs much better than the kurtosis of the raw signal. It is only slightly sensitive to time-varying operating conditions as seen at the three clusters formed at the small impulse magnification factors (e.g. $\beta < 1.0$).

The diagnostic metrics obtained with the proposed NAMVOC method, shown in Figures 11(v) and 11(vi), display much more desirable characteristics. The conditional variance for a specific impulse magnification factor β (i.e. $p(DM|FS = A)$ in Figure 1(ii)) is small, while the metrics change significantly as the impulse magnification factor increases. It is therefore possible

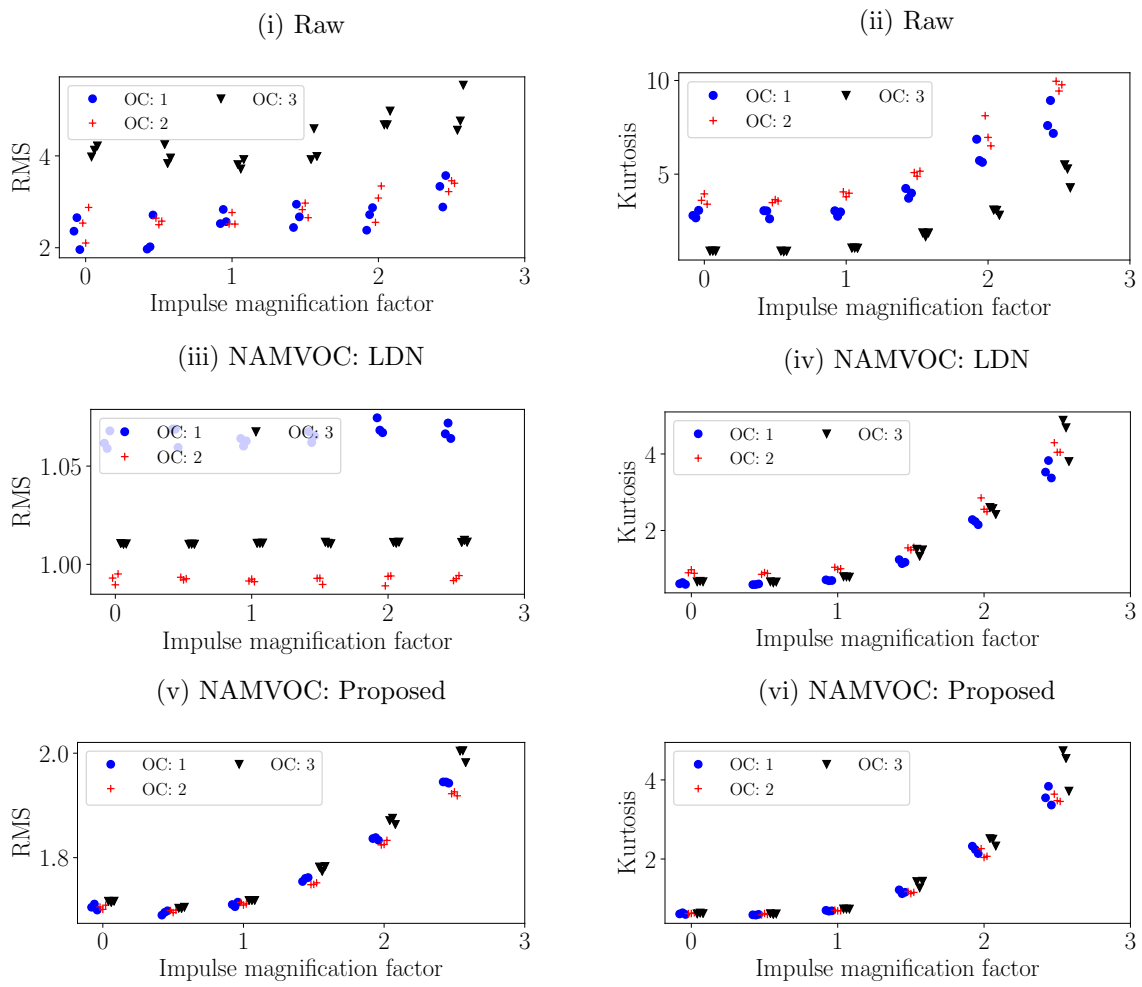

Figure 11: The Root-Mean-Square (RMS) and the kurtosis of the raw vibration signal and the normalised vibration signals, obtained with the LDN NAMVOC and proposed NAMVOC methods, are presented over an impulse magnification factor β, ranging from 0 to 2.5, for the numerical gearbox data. Six discrete impulse magnification factors are investigated, with plot jittering being performed on the β to make it easier to visualise the results.

to ascribe changes in the metrics to changes in the machine condition despite the machine operating under non-stationary operating conditions. Ultimately, the metrics of the normalised signal have a consistent definition in the presence of non-stationary operating conditions, which is very desirable for diagnostic and prognostic applications.

It is not only important to estimate the residual signal for calculating the SES, but it is also very important for calculating features such as NA4 [25] for gear diagnostics. Hence, the diagnostic metrics of the residual signals, after removing the deterministic components, are compared to the metrics obtained from the CPW signal in Figure 12.

The RMS of the residual signal obtained with the raw signal is more robust to changes in operating conditions than the RMS of the complete signal in Figure 12(i). This is attributed to

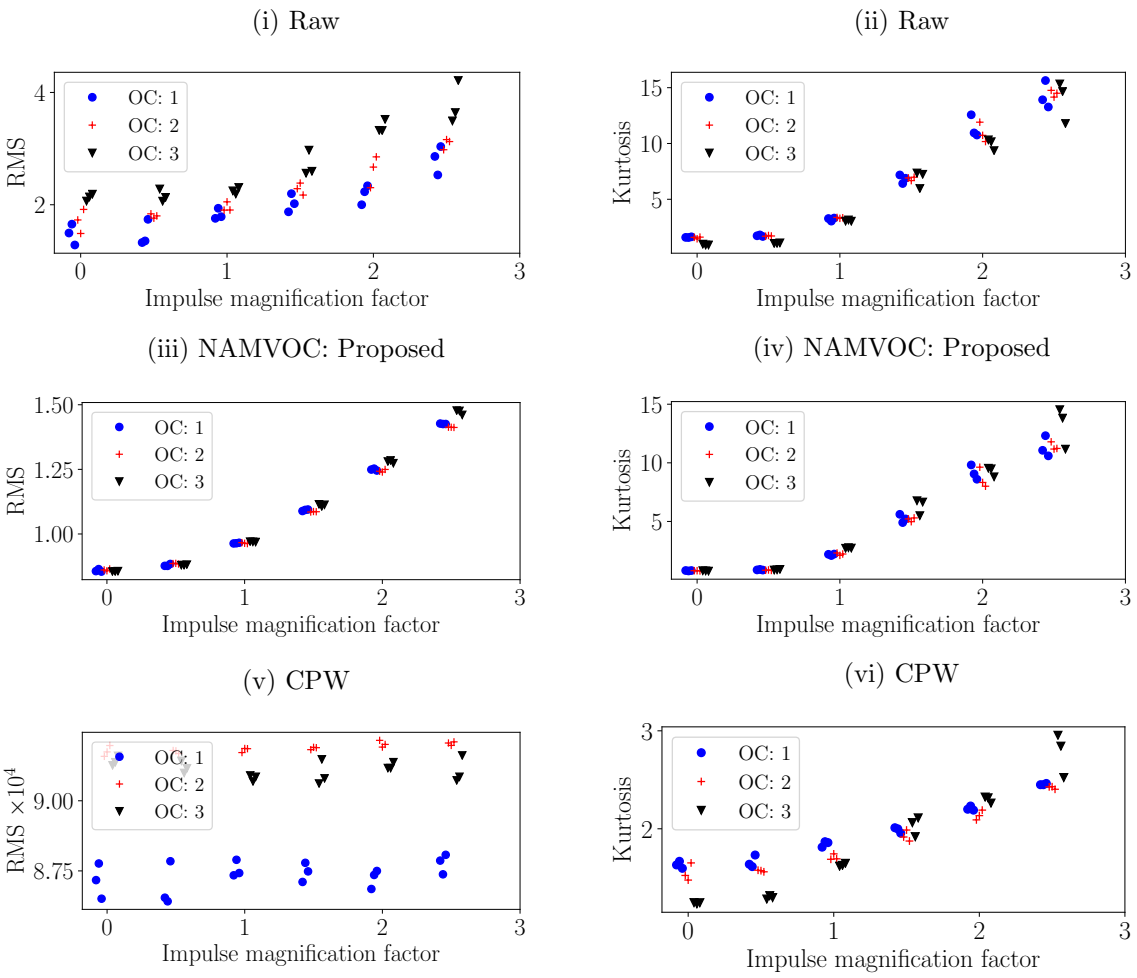

Figure 12: The diagnostic metrics of the residual signals of the raw signal (i.e. not normalised) and the normalised signal (i.e. obtained with the proposed method) are compared to the metrics obtained with the Cepstrum Pre-Whitened (CPW) signal.

the fact that the deterministic gear components dominate the signal and contain much operating condition information. However, the RMS is still quite sensitive to operating conditions. The kurtosis of the raw signal in Figure 12(ii) improved significantly by using the residual signal, with its sensitivity to operating conditions only seen by closely inspecting its values at the smaller impulse magnification factors (e.g. $\beta = 0$ and $\beta = 0.5$), where two clusters can be seen.

The metrics of the normalised signal in Figures 12(iii) and 12(iv) are even less sensitive to changes in operating conditions than the results in Figure 11. In spite of the improved properties of the kurtosis of the raw signal, the kurtosis of the normalised signal still performs better. This is especially prominent at the smaller impulse magnification factors, where the normalised signal only forms a single cluster.

The RMS of the CPW signal is very sensitive to changes in operating conditions, while being

fairly insensitive to changes in machine condition. The kurtosis is also sensitive to changes in operating conditions, as seen by the variation for specific impulse magnification factors, but is at least also sensitive to changes in machine conditions. The poor performance of the RMS is attributed to the fact that the CPW procedure only whitens the spectrum and does not remove the amplitude modulation due to varying operating conditions.

Hence, it is evident from the results that the proposed NAMVOC method offers significant benefits over conventional techniques for condition monitoring under time-varying operating conditions. It can be used to obtain an improved estimate of the deterministic components, which can subsequently be removed to obtain an enhanced envelope spectrum for bearing fault diagnosis. It also results in features that have a consistent meaning under time-varying speed conditions, which can result in performing bearing fault diagnosis more reliably. This method is further investigated on experimental gearbox data in the next section.

4. Experimental gearbox dataset

In this section, the suitability of the proposed method is investigated on experimental gearbox data and the importance of the normalisation of amplitude modulation is illustrated in an investigation where an automatic novelty detection methodology is used.

4.1. Overview of experimental dataset

The experimental gearbox data were acquired from a test bench in the Centre for Asset Integrity Management laboratory at the University of Pretoria. The experimental setup is shown in Figure 13 and consists of three helical gearboxes (one of which is the test gearbox), an electrical motor that drives the system and an alternator which dissipates the rotational energy.

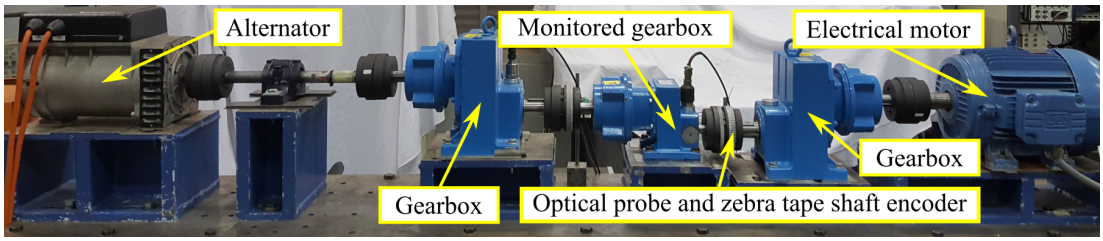

Figure 13: The experimental setup that was used to generate the data.

The axial vibration component of the tri-axial accelerometer, located on the bearing casing on the back of the test gearbox, is used for monitoring the test gearbox, while the geometrical shaft encoder and an optical probe is used to obtain the instantaneous angular speed of the

input shaft of the gearbox. The geometrical imperfections of the shaft encoder is compensated for by using the Bayesian geometrical compensation algorithm proposed by Diamond et al. [26]. The rotational speed applied by the electrical motor and the load applied by the alternator are changed independently with the four operating condition profiles used in this work shown in Figures 14(i) and 14(ii).

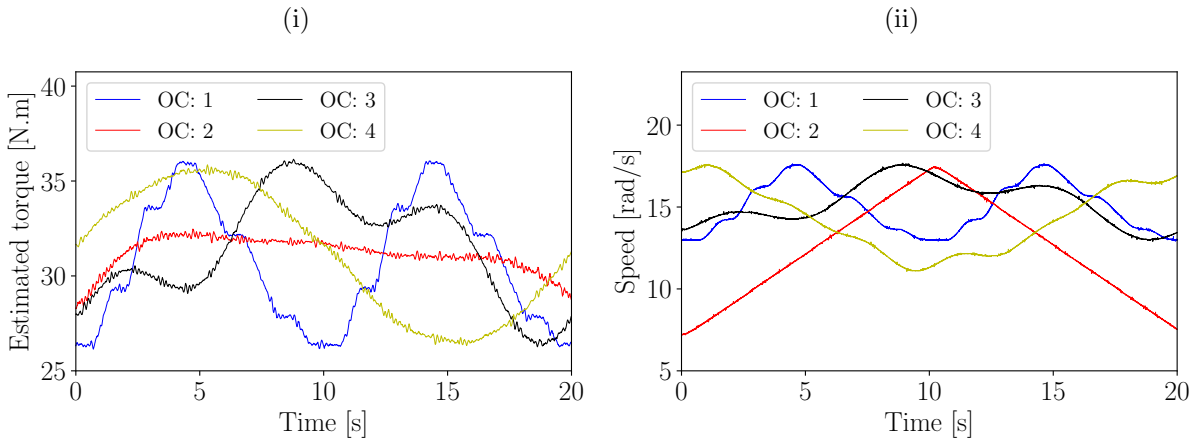

Figure 14: The operating conditions that were present at the input of the test gearbox during the experiments.

The helical test gearbox was operated in a healthy condition with the operating conditions shown in Figure 14 applied repeatedly to the gearbox. An initially healthy gear was also placed into a corrosive environment for approximately 1.5 years to introduce distributed damage on its surface and to decrease its toughness. The contact area between the gear and the shaft and the shaft key was partially protected from corrosion to ensure that it is not damaged in the corrosive environment. This damaged gear is shown in Figure 15, with its damaged surface clearly seen. The damaged gear was also lightly cleaned with a brush before the gear was tested. It was

Figure 15: The damaged gear before the experiment was started.

subsequently placed in the test gearbox and operated for approximately eight days, whereafter

the test was stopped due to excessive vibration. The gear after the experiments is shown in Figure 17. In the test with the damaged gear, 320 measurements were taken with each operating condition applied periodically as shown in Figure 16. The gear tooth broke off at measurement

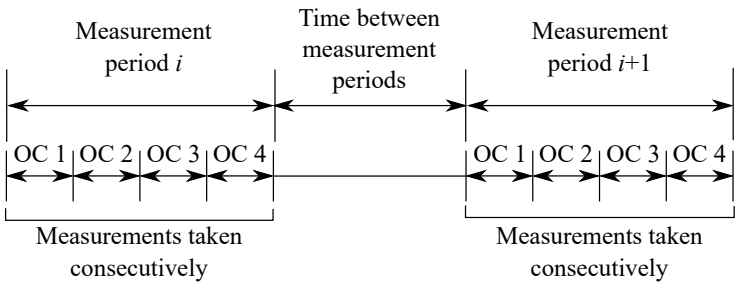

Figure 16: An illustration of the measurements used in the experimental gearbox dataset. A measurement period contains a measurement from each operating condition profile, i.e. OC 1, OC 2, OC 3 and OC 4. The average time between measurement periods was approximately between 10min and 15min.

number 148, whereafter the adjacent teeth deteriorated. At measurement number 280, large portions of the adjacent teeth broke off.

Figure 17: The damaged gear after the experiment was completed.

In the next sections, the performance of the proposed method is investigated on the healthy and damaged gearbox data.

4.2. Proposed NAMVOC method

The proposed NAMVOC method is used on this dataset with the procedure described in Section 2. The raw vibration signal from the damaged gearbox is shown in Figure 18 and denoted $y(t)$. The variation in the amplitude due to the varying operating conditions is clearly seen in the measured vibration signal. The amplitude modulation function $N(t)$ is estimated and used to obtain the normalised signal $x(t)$ with the procedure proposed in Section 2. The normalisation process attenuated the amplitude modulation caused by the time-varying operating conditions.

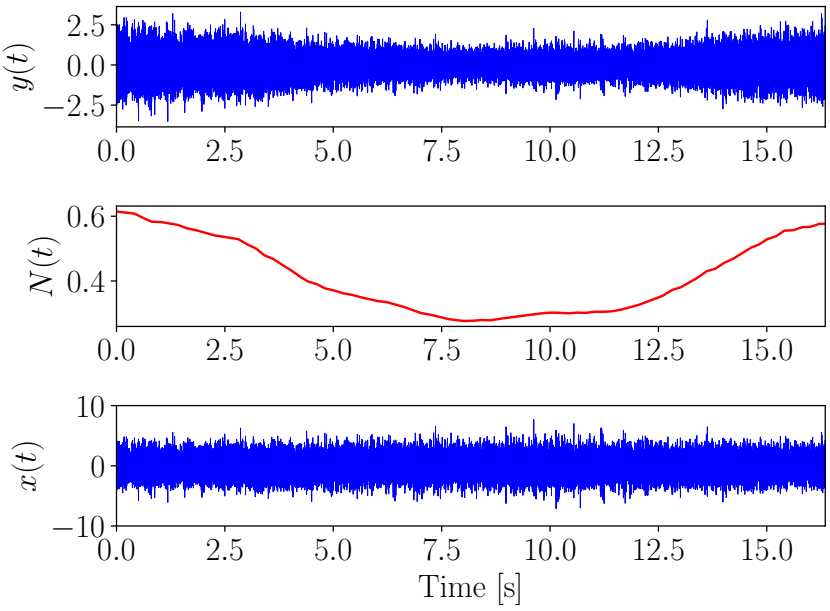

Figure 18: The raw signal $y(t)$, the estimated normalisation function $N(t)$ and the normalised signal $x(t)$ are presented for operating condition 4.

The SES obtained with the raw signal, the SES of the residual signal of the raw signal, the SES of the residual signal of the normalised signal, and the SES of the CPW signal were investigated similarly as in Section 3. However, in this dataset, the residual signal of the raw signal could remove all of the deterministic components sufficiently well. Therefore, no significant improvement could be obtained with the residual signal of the normalised signal nor the residual signal of the CPW signal. The results are therefore not included in this paper for the sake of brevity.

4.3. Fault diagnosis with diagnostic metrics

The RMS and the kurtosis of the raw vibration signal, the normalised signal obtained with the LDN NAMVOC method, the normalised signal obtained with the proposed NAMVOC method and the signal obtained from CPW, are calculated for each measurement of the damaged gearbox and presented in Figure 19. The first measurement corresponds to the gear in the condition shown in Figure 15, while the last measurement (i.e. measurement 320) corresponds to the gear in the condition shown in Figure 17.

The RMS of the raw vibration signal in Figure 19(i) is very sensitive to changes in the operating conditions, but less so to changes in the machine condition. The RMS mostly decreases with measurement number, even though the gear deteriorated significantly during the test. This may be attributed to the improvement of the corroded gear's surface during the test. The RMS

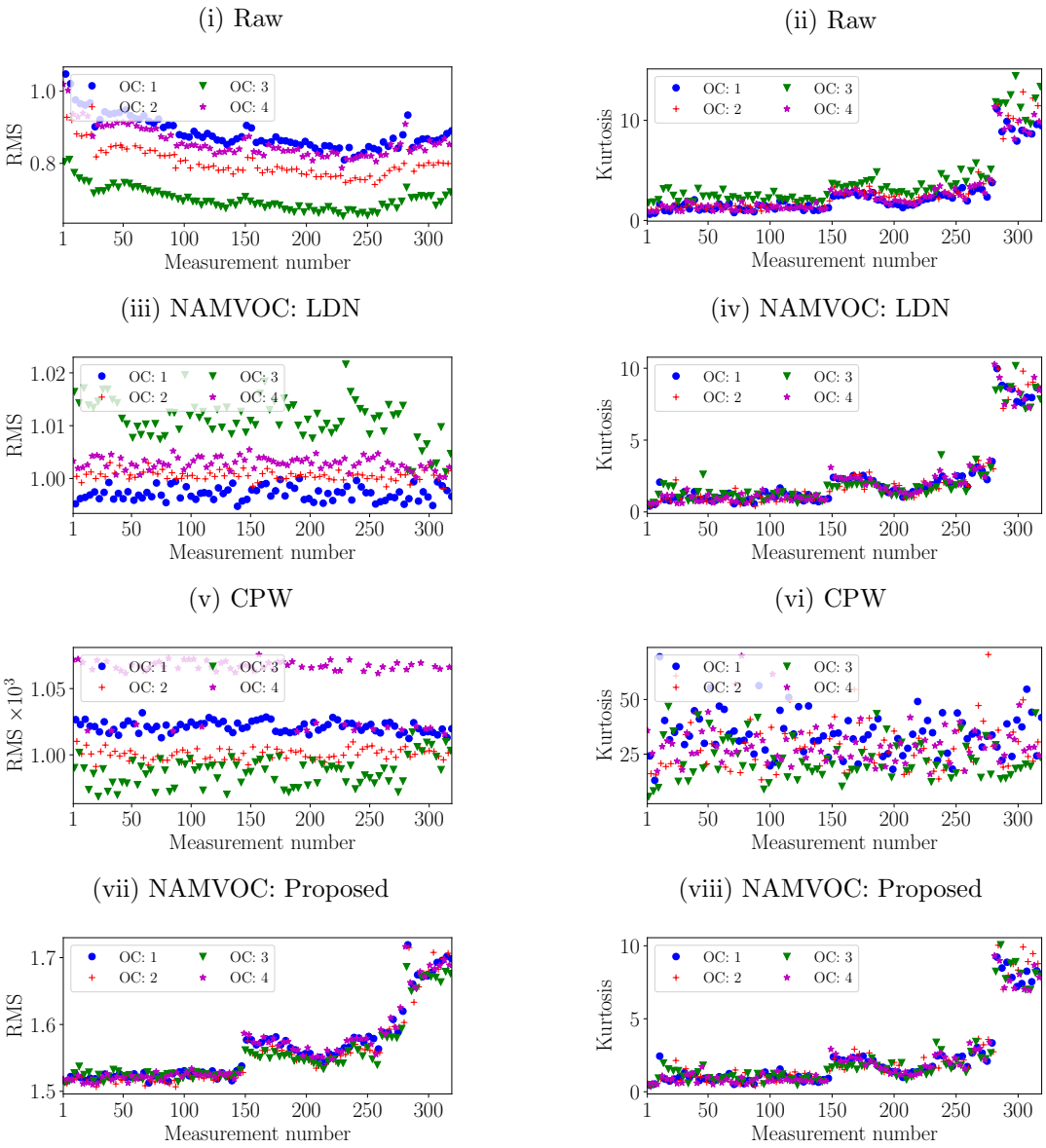

Figure 19: The Root-Mean-Square (RMS) and the kurtosis of the raw, normalised and CPW signals are presented for the experimental gearbox dataset.

of the LDN and CPW signals in Figures 19(i) and 19(v) have similar characteristics as in the numerical investigation; the RMS is insensitive to changes in the condition of the machine.

This is in contrast to the RMS obtained with the proposed method, shown in Figure 19(vii). It is very robust to changes in operating conditions, while being very sensitive to changes in the condition of the machine.

The kurtosis performs much better than the RMS, however, the kurtosis of the raw signal in Figure 19(ii) is still sensitive to the different operating conditions with OC 3 consistently having slightly larger values than the other OCs. The kurtosis of the LDN and proposed NAMVOC methods, shown in Figures 19(iv) and 19(viii) respectively, perform very well, because it is insensitive to changes in operating conditions, while being very sensitive to changes in the condition of the machine. The kurtosis of the CPW signal does not perform well with no changes in the condition of the machine being seen.

The benefits of using the features extracted from the normalised signal obtained with the proposed NAMVOC method are illustrated for automatic novelty detection in the next section.

4.4. Automatic gearbox novelty detection

Ultimately, these features will be combined with a statistical or machine learning model for automatic fault diagnostics or prognostics to ensure that the machine can be continuously monitored. It is therefore desired to illustrate the benefits of using the proposed NAMVOC method for automatic gearbox novelty detection. The results in this section also provides insights into the potential benefits of using the proposed method for classification (i.e. supervised learning) and prognostics (e.g. remaining useful life estimation).

In novelty detection, the assumption is made that we have sufficient data that describe the behaviour of a gearbox in some condition (usually a healthy condition) and then we use this data to train a model to identify changes in the system. The kurtosis and the RMS are used as features in this section. The features of the healthy and the damaged gearboxes are shown in Figure 20.

The sensitivity of the raw and LDN features to time-varying operating conditions are very clearly seen in Figures 20(i) and 20(iii), with well separate clusters formed for the different OCs, while the gear remained in the same condition during the healthy test.

The features of the proposed method are shown in Figure 20(v) and 20(vi). It is much more robust to operating condition changes when compared to the features of the raw signal and the LDN signal. The features of the healthy dataset in Figure 20(v) form a single cluster, while

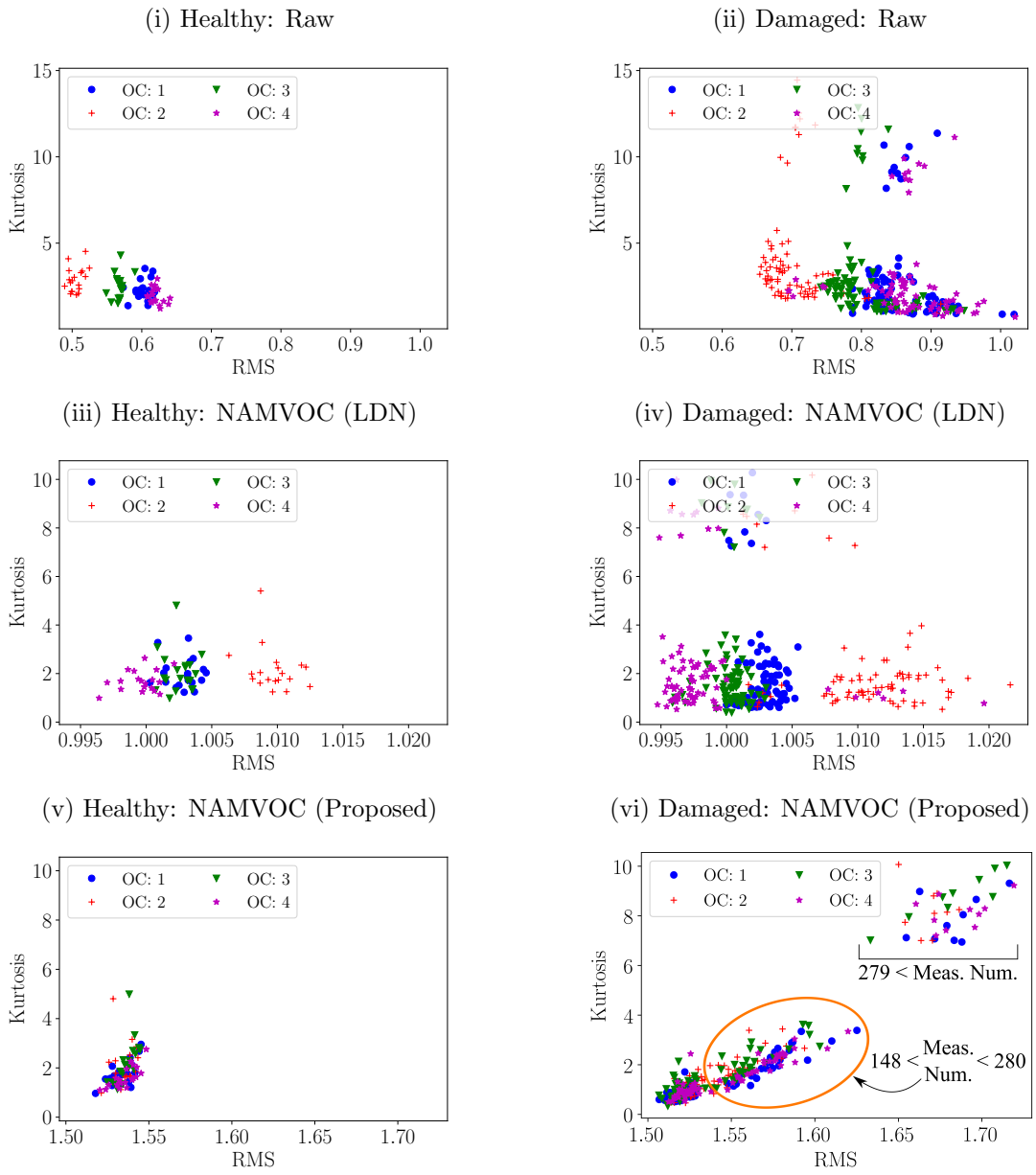

Figure 20: The feature space for the healthy and damaged experimental datasets.

the damaged dataset forms two clusters in Figure 20(vi). The second cluster is formed due to the severely damaged gear in the final measurements of the test. Hence, it is evident that the features are very robust to operating conditions, while still being sensitive to changes in the condition of the machine.

A model needs to be optimised on the reference data for automatic novelty detection. Many models such as Gaussian models, k-means clustering and autoencoders [27] can be used to learn the data from the gearbox in its reference condition, which can subsequently be used to detect changes in the data. The k-means clustering algorithm is investigated for automatic novelty detection in this work, because only the centres of the clusters need to be estimated and the only hyperparameter is the number of clusters.

In this section, the model is used for change detection and it is desired to illustrate the benefits of using features that are robust to changes in operating conditions. The first 52 measurements of the gearbox in a damaged condition are used for training, because the condition of the gearbox remained approximately constant for these measurements and provides enough time for monitoring before the event at measurement 148. Thereafter, the data associated with operating condition 2 are left out to form the training dataset. This means that the training dataset consists of 39 measurements that only comprise measurements from operating conditions 1, 3 and 4 and therefore the data of operating condition 2 are only used in the testing dataset. This will test the ability of the models to (a) detect changes in the condition of the gear and (b) their robustness to new operating conditions.

The number of clusters of the model is estimated with a five-fold cross validation procedure with three clusters being used, while scikit-learn [28] is used to optimise the models. The features are standardised to ensure that the different scales of the features do not adversely influence the performance of the model.

The novelty score

$$\eta\left(\boldsymbol{x}\right) = \min_{i} \left\{ \left(\boldsymbol{x} - \boldsymbol{\mu}_i\right)^T \left(\boldsymbol{x} - \boldsymbol{\mu}_i\right) \right\} \tag{4}$$

is the minimum distance between the features of a measurement $\boldsymbol{x}$ and the closest cluster centre where $\boldsymbol{\mu}_i$ is the centre of cluster i. A novelty is detected if the novelty score $\eta\left(\boldsymbol{x}\right) > \kappa_{thres}$, where κ_{thres} is a threshold obtained with the reference data. In this investigation, the threshold is set to the maximum value of the novelty score of the reference dataset.

The novelty score is calculated with Equation (4) for the damaged experiments and presented in Figure 21. The changes of the metrics at the events of measurement 148 and 280 are clearly

seen in the novelty score with the proposed method, while only the last event (i.e. at measurement 280) is clearly detected by the models using the raw and the LDN data.

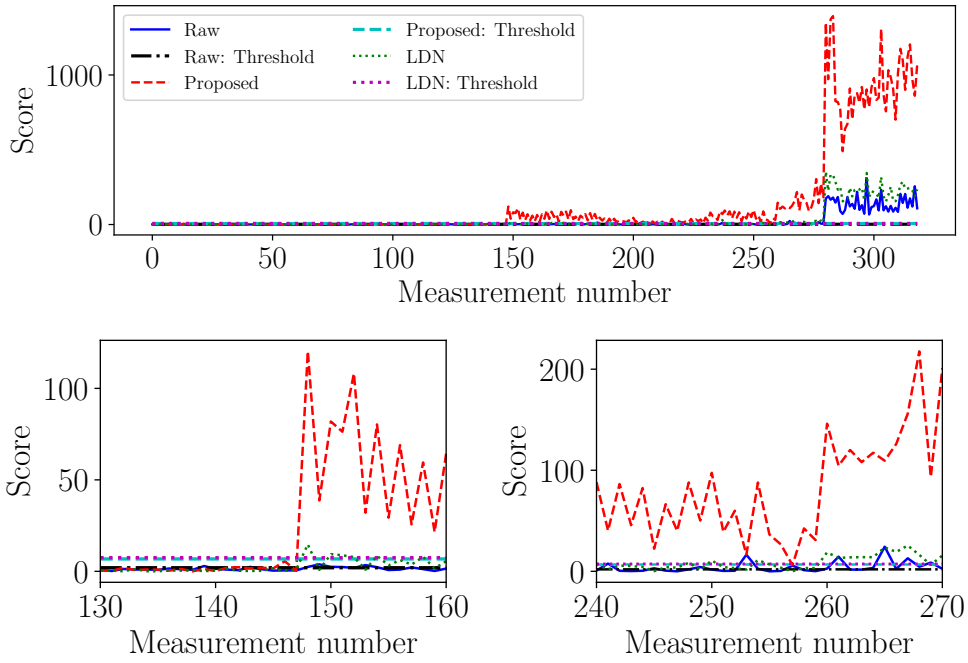

Figure 21: The novelty scores and thresholds that are obtained with the raw signal and the normalised signals, obtained with the LDN NAMVOC method and the proposed NAMVOC method, are compared.

A zoomed view of the novelty metrics is shown at measurement number 148 in Figure 21. It is clear from the results that the proposed method is significantly more sensitive to the event at measurement number 148 (i.e. where one of the teeth broke) compared to the results using the raw features or the LDN features.

The alarm, obtained by comparing the novelty score to the threshold, is presented in Figure 22(i) for the three signals.

The fluctuation of the alarm of the raw data is attributed to the sensitivity of the features to changes in operating conditions. It is clear from the results that the proposed method outperforms the other methods; it detects the event at measurement number 148 and then remains triggered for the duration of the test, while the fluctuation of the alarms of the raw and LDN alarms after measurement number 148 being attributed to their insensitivity to changes in the condition of the machine and their sensitivity to changes in operating conditions.

The RMS and the kurtosis of the raw signal result in a very unreliable model that cannot be used to infer the condition of the machine under time-varying operating conditions. However, if the proposed NAMVOC method is used before the RMS and the kurtosis are calculated,

(i)

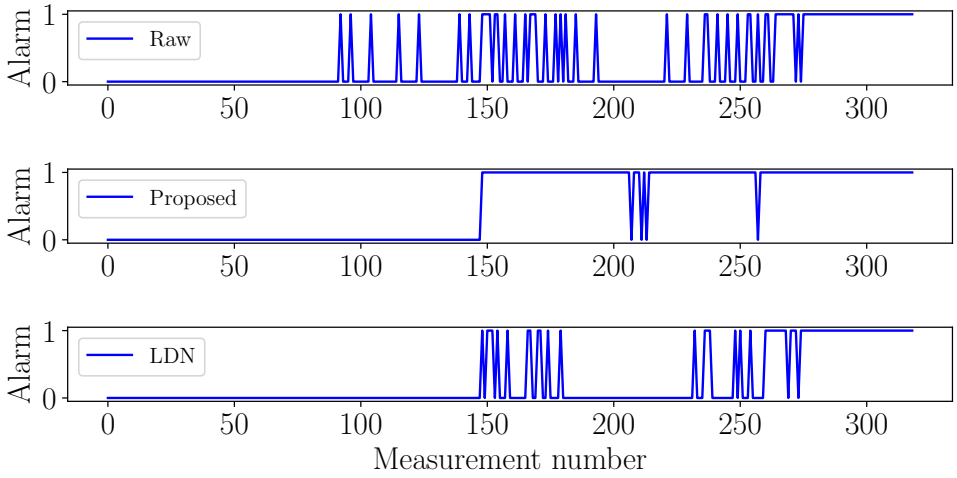

Figure 22: The alarm value obtained by comparing the data to the thresholds in Figure 21 is presented in Figure 22(i).

very desirable results can be obtained for condition monitoring under time-varying operating conditions.

5. Conclusion

In this work, the benefits of normalising the amplitude modulation caused by varying operating conditions are emphasised for diagnostics and prognostics. The load demodulation normalisation method, which was intended for attenuating cyclic stationary loads, is revisited and an improvement is proposed. The improved normalisation method uses a moving median window to estimate the modulation function whereafter the modulation function is attenuated with a normalisation process.

The proposed method is investigated on numerical gearbox data and experimental gearbox data, where the benefits of using the proposed amplitude normalisation method are illustrated for deterministic-random separation, incipient fault detection and automatic fault diagnostics under time-varying operating conditions. The proposed method is also compared to cepstrum prewhitening, where it can be seen that the squared envelope spectra obtained with the proposed method exhibits higher signal-to-noise ratios and the root-mean-square and kurtosis features extracted from the normalised signal are also more appropriate for fault diagnosis and prognosis. This is because it is very sensitive to changes in the condition of the machine, while being robust

to changes in the operating condition. The results indicate that prognosis methodologies that are based on the RMS measure can still be used under time-varying operating conditions if this methodology is used as a pre-processing step.

It is suggested that the method needs to be validated on data from machines such as wind turbines to investigate its capabilities on industrial datasets.

Acknowledgements

The authors gratefully acknowledge the support that was received from the Eskom Power Plant Engineering Institute (EPPEI) and the UP postdoctoral programme in the execution of the research.

Appendix A. Normalising Amplitude Modulation caused by Varying Operating Conditions (NAMVOC)

Appendix A.1. Conceptual motivation for proposed method

Consider the signal

$$y(t) = M(t) \cdot (\psi(t) + \epsilon(t)) \tag{A.1}$$

where $\epsilon(t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $\psi(t) \sim \mathcal{N}(0, \sigma_\psi^2) \cdot \mathcal{U}(t; t_\psi, T_\psi)$, and $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with a mean of μ and a standard deviation of σ. The slow-varying modulation function is denoted $M(t)$. The function $\mathcal{U}(t; t_\psi, T_\psi) = (u(t - t_\psi) - u(t - t_\psi - T_\psi))$ is used to simulate time-localised noise with $u(t)$ being the unit step function, i.e. $u(t) = 1$ for $t \geq 0$ and 0 otherwise. The signal $y(t)$ can be seen as a short segment of a vibration signal, where only a single localised impulse is present. The instantaneous mean of this function is zero and the instantaneous variance of the signal

$$\mathbb{E}\left\{|y(t)|^2\right\} = |M(t)|^2 \cdot \left(\sigma_\epsilon^2 + \sigma_\psi^2 \cdot \mathcal{U}(t; t_\psi, T_\psi)\right). \tag{A.2}$$

The different signal components of Equations (A.1) are shown in Figure A.23

Ultimately, the objective is to remove the contribution of the modulation function $M(t)$, i.e.

$$x(t) = \frac{M(t)}{\hat{M}(t)} (\psi(t) + \epsilon(t)) \approx (\psi(t) + \epsilon(t)), \tag{A.3}$$

where $x(t)$ is the function with the amplitude modulation normalised and $\hat{M}(t)$ is some estimate of $M(t)$. The form of the modulation function $M(t)$ is unknown and therefore it needs to be estimated unsupervised from the data. The amplitude modulation function results in slow

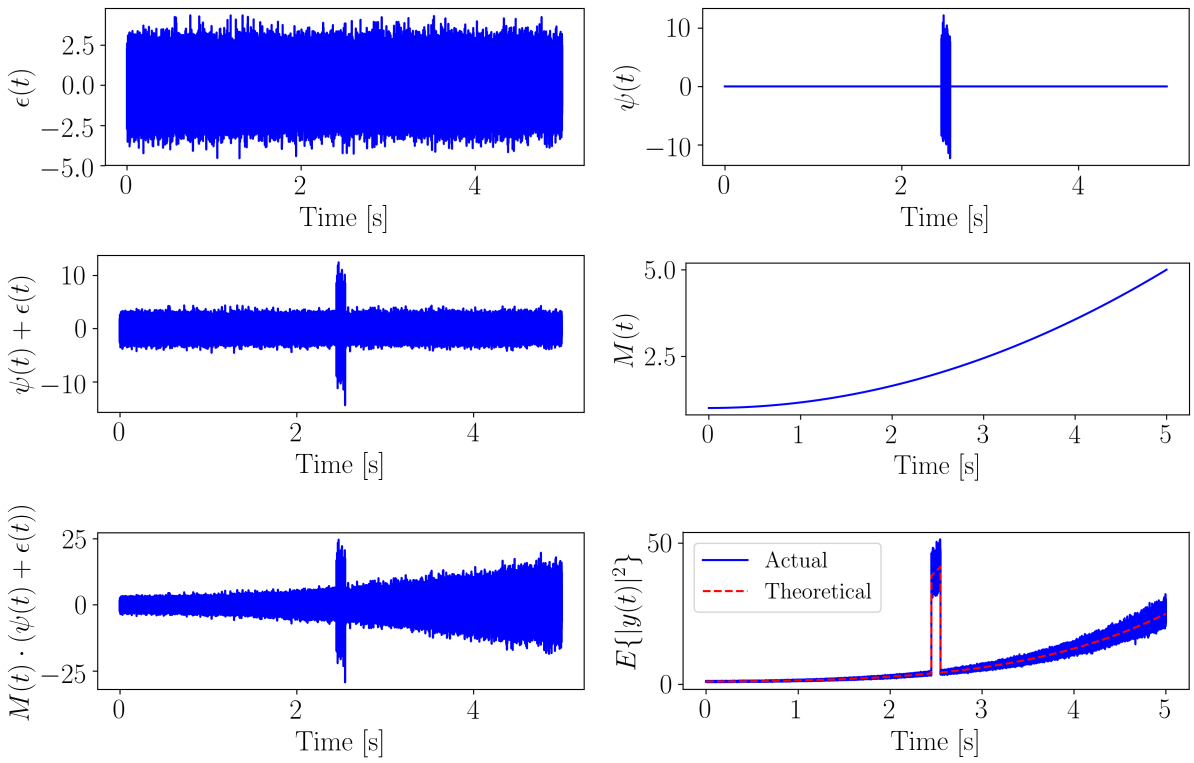

Figure A.23: The different signal components in Equation (A.1) that is used in this conceptual motivation.

changes in the amplitude of the data and therefore it is envisaged that the localised mean square of the signal can be used as an estimate of $|M(t)|^2$. If we use the localised mean square of the signal

$$e_y(t; t_0, T_0) = \frac{1}{T_0} \int_{t-T_0/2}^{t+T_0/2} |y(t)|^2 dt, \tag{A.4}$$

localised over a period of T_0, as an estimator of $|M(t)|^2$, then the bias of the estimator will be as follows:

$$b(t) = \mathbb{E}\left\{ \frac{1}{T_0} \int_{t-T_0/2}^{t+T_0/2} |y(t)|^2 dt \right\} - |M(t)|^2, \tag{A.5}$$

where it can be shown that

$$b(t) = \frac{1}{T_0} e_M(t; T_0) \cdot \sigma_\epsilon^2 - |M(t)|^2, \tag{A.6}$$

if there is no overlap between the time segment of the integration $[t, t + T_0]$ and the position of the impulse $[t_\psi, t_\psi + T_\psi]$. The bias can be further reduced to

$$b(t) = |M(t)|^2 \cdot \left(\sigma_\epsilon^2 - 1\right), \tag{A.7}$$

if $T_0 \to 0$. Hence, the localised mean square of the signal is a biased estimate of $|M(t)|^2$ unless the variance of the noise is equal to unity. Hence, this estimator scales the signal so that the

noise component has a variance of unity, which is not a problem under general circumstances. If we consider the case where the segment of the integration $[t, t+T_0]$ and the position of the impulse $[t_\psi, t_\psi + T_\psi]$ completely overlap

$$b(t) = \frac{1}{T_0} e_M(t_0; T_0) \cdot \sigma_\epsilon^2 + \frac{1}{T_0} e_M(t_\psi; T_\psi) \cdot \sigma_\psi^2 - |M(t)|^2 \tag{A.8}$$

and rearrange it to have a similar form as Equation (A.6)

$$b(t) = \frac{1}{T_0} e_M(t_0; T_0) \cdot \sigma_\epsilon^2 \cdot \left(1 + \frac{\frac{1}{T_0} e_M(t_\psi; T_\psi) \cdot \sigma_\psi^2}{\frac{1}{T_0} e_M(t_0; T_0) \cdot \sigma_\epsilon^2} \right) - |M(t)|^2, \tag{A.9}$$

it is seen that an additional factor $\left(1 + \frac{\frac{1}{T_0} e_M(t_\psi; T_\psi) \cdot \sigma_\psi^2}{\frac{1}{T_0} e_M(t_0; T_0) \cdot \sigma_\epsilon^2} \right)$ is present. This factor indicates that the bias is now also dependent on the magnitude of the impulse and the duration of the impulse and would therefore change as the severity of the damage changes. Hence, the normalisation process with this estimator would attenuate the modulation as well as the diagnostic information, where the attenuation of the diagnostic information depends on the severity of the damage. This is an undesirable property for an estimator in condition monitoring.

This localised average energy of the system is actually an estimate of the squared envelope of the signal as $T_0 \to 0$. The localised average energy can be estimated by a moving average window on the square envelope of the signal, with the moving average window being a finite impulse response filter. Hence, this estimator is the same as the Load Demodulation Normalisation (LDN) process, but instead of using a finite impulse response filter on the absolute of the analytic signal, a moving average finite impulse response filter is used on the squared envelope, whereafter the signal is normalised by its square root.

Hence, the LDN procedure attenuates the diagnostic information in the presence of localised impulses. This problem can be alleviated by using a moving median window instead of a moving average window; the moving median window is very robust to outliers and would therefore not be as severely affected by the presence of localised impulses. Hence, the proposed NAMVOC method uses the moving median window to estimate the modulation function.

Appendix A.2. Investigations

Investigations are performed to illustrate the benefits of using the proposed NAMVOC method for incipient damage detection and the benefits of using the proposed NAMVOC method instead of load demodulation normalisation. Other benefits of the proposed method can be seen in the main text of the article.

Appendix A.2.1. Modulation normalisation for incipient damage amplification

Varying operating conditions impede the ability of conventional signal analysis tools to detect changes in the condition of the machine. A classical example of this is the presence of cyclic stationary load modulation as illustrated by Stander et al. [3, 21]. In this work, this phenomenon is illustrated with $y(t) = M(t) \cdot x(t)$ with $M(t) = 1 + 0.8 \cdot \sin(2\pi \cdot 40 \cdot t)$ and

$$x(t) = \sin(2000 \cdot \pi \cdot t) + 4 \cdot \epsilon(t) + h(t) \otimes \sum_{k=1}^{10} \delta(t - 0.4 \cdot k), \tag{A.10}$$

where $h(t) = 1.3 \cdot \exp(-100 \cdot t) \cdot \sin(2000 \cdot \pi \cdot t)$ and $\epsilon(t)$ is a zero-mean Gaussian sample with a unit variance. The shaft rotates at 1 Hz. In Figure A.24, the signal and the scaled impulses (i.e. the last term of Equation (A.10)) are presented as well as the synchronous averaged angle-frequency instantaneous power spectra of the signals. The synchronous averaged angle-frequency instantaneous power spectrum is a very powerful tool that has been successful for wind turbine bearing fault diagnosis [11].

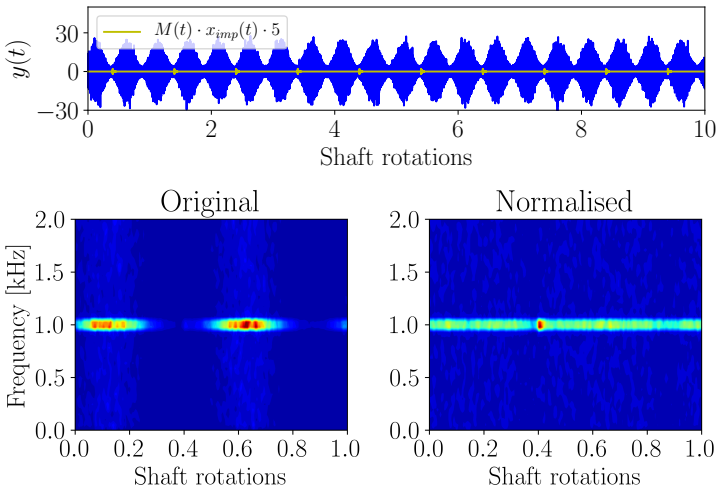

Figure A.24: The signal $y(t)$ is shown with the impulses that needs to be detected magnified in the plot. The synchronous average of the Angle-Frequency Instantaneous Power Spectrum (AF-IPS) is shown for the raw or original signal $y(t)$ and of the estimated normalised signal $x(t)$.

The damage is masked by the operating conditions in the raw or original signal, but is enhanced when using the normalised signal. This phenomenon will also be present if the synchronous averaging process does not attenuate the modulation induced by the varying operating conditions sufficiently well. Hence, the normalisation process can enhance the ability of fault diagnosis tools to detect incipient damage under time-varying operating conditions.

Appendix A.2.2. Fault size estimation

Inferring the state or the fault size of the machine is very important to ensure that the appropriate maintenance decisions can be made. In this section, the ability of the raw signal, the LDN processed signal and the proposed NAMVOC method are compared for inferring the extent of the damage under time-varying conditions. In this work, a signal was created of the form:

$$y(t) = M(t) \cdot (\beta \cdot p(t) \cdot \epsilon(t) + s(t) + n(t)) \tag{A.11}$$

where β is referred to as the impulse magnification factor. The data are generated over a time period of 100 seconds with a sampling frequency of 1000 Hz being used. The periodic function

$$p(t) = \exp(-10 \cdot t) \otimes \sum_{k=0}^{99} \delta(t - k) \tag{A.12}$$

consists of a train of impulses. The deterministic function has a form of:

$$s(t) = \sin(2\pi \cdot t \cdot 10) \tag{A.13}$$

while $\epsilon(t)$ and $n(t)$ are samples from a zero-mean Gaussian distribution with unit variance and a variance of 0.04, respectively. The modulation function

$$M(t) = c_a \cdot \cos(0.04 \cdot \pi \cdot t \cdot c_b) + c_d + 1 + c_c \cdot 0.01 \cdot t \tag{A.14}$$

is different for each dataset because c_a is a sample from a Gaussian distribution with unit variance and c_b, c_c, and c_d are samples from a uniform distribution with a domain $[0, 1]$.

In this investigation, a set of measurements is generated for different impulse magnification factors β, with 15 measurements created for each β. This is performed to compare the average magnitude of the impulses in the raw signal; the average magnitude of the impulses in the normalised signal, obtained by using the LDN procedure; and the average magnitude of the impulses in the normalised signal, obtained with the proposed method. The average impulse magnitude is shown in Figure A.25 for the different cases, with an isolated plot of the LDN signal shown in Figure A.25(i) for clarity.

A significant variation in the averaged magnitude of the impulses is seen when a specific impulse magnification factor is observed for the raw data. This significant variation is induced by the time-varying modulation function and makes it impossible to reliably infer the magnitude of the damage. The average impulse magnitude of the normalised signal has a linear relationship with the impulse magnification factor β and it has a very small conditional variance (i.e. the

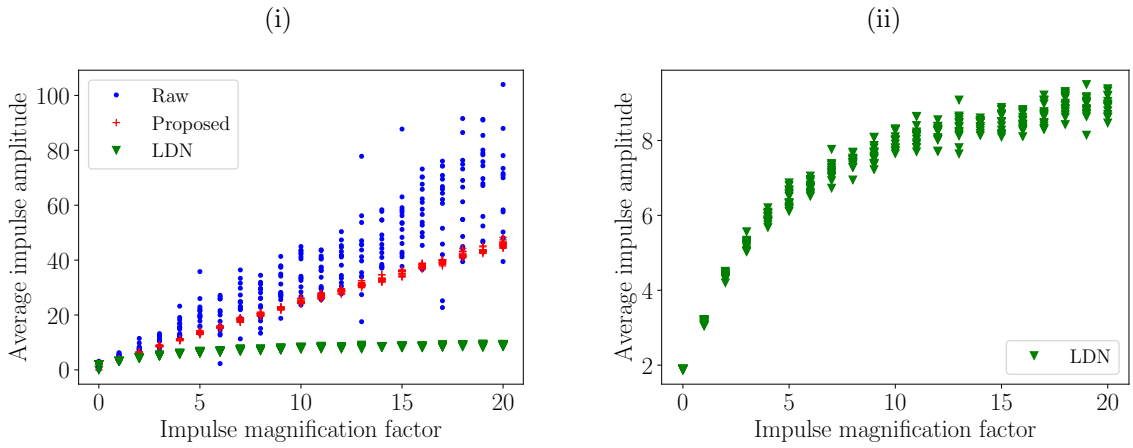

Figure A.25: The average amplitude of the impulses in the raw signal, of the normalised signal using the proposed NAMVOC method and of the normalised signal using the LDN NAMVOC method is presented. The data was generated for 15 measurements of each considered impulse magnification factor β. The LDN data in Figure A.25(i) is shown in Figure A.25(ii) for clarity.

variance of the amplitudes for a specific β). This means that the condition can be reliably inferred in the presence of a time-varying modulation functions. The diagnostic information attenuation effect of the LDN procedure result in the average impulse amplitude to have a non-linear relationship with the impulse magnification factor. This results in the average impulse amplitude to saturate which can also result in unreliable results when inferring the condition of the machine (i.e. the distribution $p(FS|DM = B)$ in Figure 1 has a very large variance). Hence, the proposed method is better suited than the raw signal and the LDN processed signal for fault diagnosis and prognosis under time-varying operating conditions.

Appendix B. Phenomenological gearbox model information

The different signal components in Equation (2) are given in this section. The deterministic gear component is given by

$$y_{dg}(t) = h_{dg}(t) \otimes \left(Q(t) \cdot b(t) \cdot \sum_{k=1}^{N_{dg}} A_{dg}^{(k)} \sin\left(k \cdot N_{teeth} \cdot \int_0^t \omega(\tau) d\tau + \phi_k \right) \right), \tag{B.1}$$

where h_{dg} is a single degree-of-freedom impulse response function and $Q(t)$ is the modulation function caused by the varying operating conditions. The function $Q(t) \neq M(t)$ in Equation (1), because the impulse response function for example may induce additional amplitude modulation in the system. The function $b(t) = 1 + 0.8 \sin\left(\int_0^t \omega(\tau) d\tau\right)$ is a deterministic component which simulates damage on the gear. The amplitude and phase of the kth harmonic of the gears are

denoted $A_{dg}^{(k)}$ and $\phi_{rg}^{(k)}$ respectively. The gear mesh order N_{teeth} is used to simulate the gear mesh components due to the gear interactions.

The random gear component

$$y_{rg}(t) = h_{rg}(t) \otimes \left(Q(t) \cdot \epsilon(t) \cdot \sum_{k=1}^{N_{rg}} A_{rg}^{(k)} \sin\left(k \cdot GR \cdot \int_0^t \omega(\tau) d\tau + \phi_k \right) \right), \tag{B.2}$$

has the same form as Equation (B.1), except for the zero mean, unit variance Gaussian component $\epsilon(t)$ and GR is the gear ratio to simulate damage on the pinion.

The bearing component

$$y_b(t) = h_b(t) \otimes \left(Q(t) \cdot \sum_{k=1}^{N_{imp}} a_k \cdot \delta(t - \mathcal{T}_k) \right), \tag{B.3}$$

consists of N_{imp} impulses, where the kth impulse has a time-of-arrival of $\mathcal{T}_k$. The arrival time of the kth bearing impulse depends on the instantaneous phase of the shaft as well as the slip of the component [14]. The slip is simulated by adding a random Gaussian variable with a mean of zero and a standard deviation of 0.05 to the arrival angle of each impulse. The kth bearing impulse is multiplied with a random function $a_k \sim U[1.0, 1.5]$, where U is a uniform distribution.

The noise component

$$y_n(t) = \epsilon(t) \cdot Q(t), \tag{B.4}$$

consists of a random Gaussian variable with a mean of zero and with a variance of unity. All modulation functions $Q(t)$ used in Equations (B.1) to (B.4) are the same.

The modulation function is not only parametrised in terms of the rotational speed of the system, but also in terms of other components, simulating for example varying load conditions. The data are generated for three operating conditions, where the rotational speed in radians of each operating condition

$$\omega_1(t) = 2\pi \cdot \left(\frac{15}{20} t + 5 + 4\,(\xi - 0.5) \right) \tag{B.5}$$

$$\omega_2(t) = 2\pi \cdot \left(5 + 15 \exp\left(-\left(\frac{t - 10}{5} \right)^2 \right) + 4\,(\xi - 0.5) \right) \tag{B.6}$$

$$\omega_3(t) = 2\pi \cdot \left(3\cos\left(2\pi t/10\right) + 3t/20 + 12 + 4\,(\xi - 0.5) \right), \tag{B.7}$$

contains a deterministic form and a random offset governed by the random variable ξ. The variable ξ is a sample from a uniform distribution with a domain $[0, 1]$. The corresponding

modulation functions of the three operating conditions are given by

$$Q_1(t) = \left(\frac{\omega}{80 \cdot \pi} + 0.2\right)^2, \tag{B.8}$$

$$Q_2(t) = \left(\frac{\omega}{80 \cdot \pi} + 0.2\right)^2 + 0.03 \cdot \sin(0.4 \cdot \pi \cdot t) + 0.03, \tag{B.9}$$

$$Q_3(t) = \left(\frac{\omega}{80 \cdot \pi} + 0.2\right)^2 + 0.2. \tag{B.10}$$

The important parameters of the model are given in Table B.1 and Table B.2.

Table B.1: Some important parameters of the model are given in this table. The natural frequency and the damping ratio of the single degree-of-freedom impulse response function associated with the signal component are denoted f_n and ζ respectively. The number of teeth associated with the gear mesh is N_{teeth} and the gear ratio is denoted by GR. The relative energy of each signal component is also given in the row "% energy" with the bearing component corresponding to a $\beta = 1$.

	x_b	x_{rg}	x_{dg}	x_n
f_n	2000	4000	3000	N/A
ζ	0.05	0.05	0.05	N/A
% energy	60	10	10	20
N_{teeth}	N/A	N/A	25	N/A
GR	N/A	1.667	N/A	N/A
BPOO	2.58	N/A	N/A	N/A

Table B.2: The amplitudes and phases of the harmonics of the deterministic gear and random gear components.

k	1	2	3
$A_{rg}^{(k)}$	1	2	3
$A_{dg}^{(k)}$	3	2	1
$\phi_{rg}^{(k)}$	0	0	0
$\phi_{dg}^{(k)}$	0	0	0

References

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

[2] A. Heng, S. Zhang, A. C. C. Tan, J. Mathew, Rotating machinery prognostics: State of the art, challenges and opportunities, Mechanical Systems and Signal Processing 23 (3) (2009) 724–739.

[3] C. J. Stander, P. S. Heyns, W. Schoombie, Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions, Mechanical Systems and Signal Processing 16 (6) (2002) 1005–1024.

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

[5] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: A systematic review from data acquisition to RUL prediction, Mechanical Systems and Signal Processing 104 (2018) 799–834.

[6] A. Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mechanical Systems and Signal Processing 22 (3) (2008) 574–587.

[7] K. Feng, P. Borghesani, W. A. Smith, R. B. Randall, Z. Y. Chin, J. Ren, Z. Peng, Vibration-based updating of wear prediction for spur gears, Wear 426-427 (September 2018) (2019) 1410–1415.

[8] R. Zimroz, W. Bartelmus, T. Barszcz, J. Urbanek, Diagnostics of bearings in presence of strong operating conditions non-stationarity - A procedure of load-dependent features processing with application to wind turbine bearings, Mechanical Systems and Signal Processing 46 (1) (2014) 16–27.

[9] N. Baydar, A. Ball, Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum, Mechanical Systems and Signal Processing 14 (6) (2000) 907–921.

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

[11] J. Urbanek, T. Barszcz, R. Zimroz, J. Antoni, Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions, Measurement 45 (7) (2012) 1782–1791.

[12] D. Abboud, J. Antoni, M. Eltabach, S. Sieg-Zieba, Angle\time cyclostationarity for the analysis of rolling element bearing vibrations, Measurement 75 (2015) 29–39.

[13] D. Abboud, J. Antoni, S. Sieg-Zieba, M. Eltabach, Deterministic-random separation in nonstationary regime, Journal of Sound and Vibration 362 (2016) 305–326.

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

[15] D. Abboud, J. Antoni, Order-frequency analysis of machine signals, Mechanical Systems and Signal Processing 87 (October 2016) (2017) 229–258.

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

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

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

[19] Q. Leclère, H. André, J. Antoni, A multi-order probabilistic approach for Instantaneous Angular Speed tracking debriefing of the CMMNO'14 diagnosis contest, Mechanical Systems and Signal Processing 81 (2016) 375–386.

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

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

[22] D. Abboud, J. Antoni, M. Eltabach, S. Sieg-Zieba, Speed-spectral whitening for enhancing envelope analysis in speed varying conditions, Vishno 2014 (2014) 2–5.

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

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

[25] P. D. Samuel, D. J. Pines, A review of vibration-based techniques for helicopter transmission diagnostics, Journal of Sound and Vibration 282 (1-2) (2005) 475–508.

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

[27] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.