

A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques

S. Schmidt^{a,*}, P.S. Heyns^a, J.P. de Villiers^{b,c}

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

^b*Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa*

^c*Defence, Peace, Safety and Security (DPSS), Council for Scientific and Industrial Research (CSIR), PO Box 395, Pretoria 0001, South Africa*

Abstract

In this paper, a fault diagnostic methodology is developed which is able to detect, locate and trend gear faults under fluctuating operating conditions when only vibration data from a single transducer, measured on a healthy gearbox are available. A two-phase feature extraction and modelling process is proposed to infer the operating condition and based on the operating condition, to detect changes in the machine condition. Information from optimised machine and operating condition hidden Markov models are statistically combined to generate a discrepancy signal which is post-processed to infer the condition of the gearbox. The discrepancy signal is processed and combined with statistical methods for automatic fault detection and localisation and to perform fault trending over time. The proposed methodology is validated on experimental data and a tachless order tracking methodology is used to enhance the cost-effectiveness of the diagnostic methodology.

Keywords:

Diagnostics, Gearbox, Discrepancy analysis, Fluctuating operating conditions, Hidden Markov Model, Probabilistic techniques

1. Introduction

Condition-based maintenance uses the current condition of the machine as basis for maintenance decisions and can be a cost-effective and more efficient alternative to run-to-failure and time-based maintenance procedures [1]. Rotating machines, such as gearboxes, frequently oper-

*Corresponding author.

Email address: stephanschmidt@zoho.com (S. Schmidt)

ate under fluctuating operating conditions due to its varying operating environment (e.g. ground properties for bucket wheel excavators [2], wind speed for wind turbines [3, 4] etc.). The fluctuating operating conditions lead to amplitude and frequency modulation [5], phase distortion when performing computed order tracking [6, 7] and varying signal-to-noise ratios [8] which complicate the condition monitoring process.

In recent years, machine learning techniques gained popularity in the engineering community due to its ability to solve difficult inference tasks such as problems found in the condition monitoring field [9–14]. Machine learning-based diagnostic methodologies, using supervised learning approaches, assume that historical fault data, of all relevant damage modes, are readily available for model optimisation. However, this assumption is rarely realised in industrial applications, which makes optimising the relevant models difficult. Novelty detection approaches in the machine diagnostic field, are attractive alternatives to supervised learning approaches, because the assumption is made that data from a healthy machine are abundant. A model of the healthy machine data is used to determine whether new data are from a healthy machine or not. Fernandez-Francos et al. [15] performed novelty detection for bearing diagnostics using a one-class support vector machine. Heyns et al. [16] used a statistical approach to average linear prediction models for gear fault detection under fluctuating operating conditions.

Discrepancy analysis is a novelty detection approach that uses a discrepancy measure to quantify the deviation of newly acquired data from the behaviour of data of a healthy machine. Heyns et al. [17] generated a discrepancy signal from the envelope of the residual signal obtained from a forward prediction made by a neural network. Heyns et al. [18] used a Gaussian mixture model to model the behaviour of a gearbox in a healthy condition. The negative log-likelihood, also known as the error function [19], was used to generate a discrepancy signal. Heyns et al. [20] developed a methodology using smart features and machine learning techniques for gearboxes operating under fluctuating operating conditions. A two-phase feature extraction approach was proposed using the concept of smart features, which was used to determine the instantaneous operating conditions and machine condition. Hidden Markov models (HMMs) and Gaussian mixture models modelled the operating and the machine condition features respectively, with its information combined using hard classification rules to generate a discrepancy signal. Heyns et al. [18] proposed synchronous averaging, and Schmidt et al. [21] proposed additional discrepancy

signal processing techniques which can be used to detect, locate and trend gear damage over the machine’s operational lifetime.

In this paper, a fault diagnostic methodology is proposed for gearboxes operating under fluctuating conditions with its process diagram presented in figure 1. It is assumed that only vi-

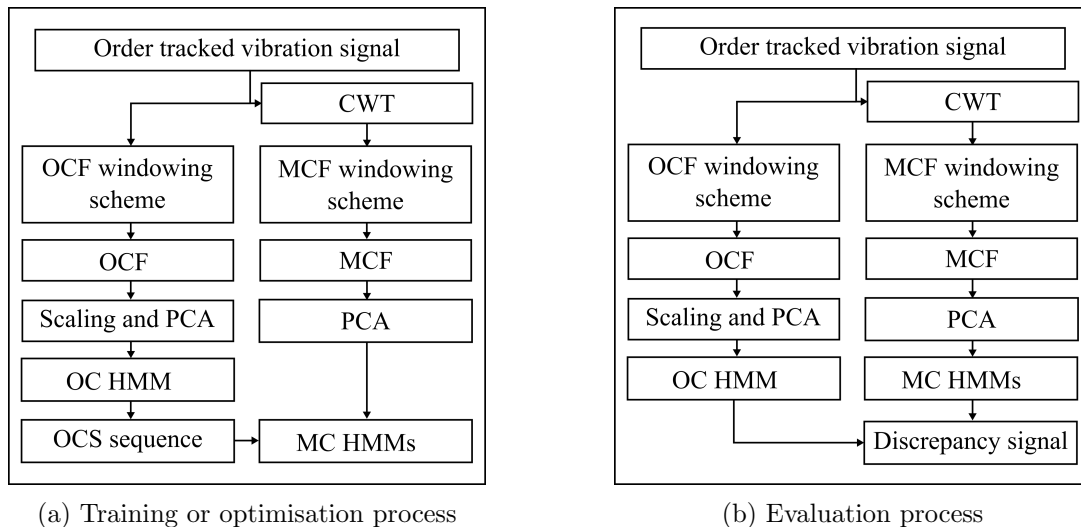


Figure 1: The optimisation and evaluating processes that are used in the proposed methodology. The following abbreviations are used: Operating condition (OC); Machine condition (MC); Operating condition feature (OCF); Machine condition feature (MCF); Principal component analysis (PCA); Continuous wavelet transform (CWT); Operating condition state (OCS); Hidden Markov Model (HMM).

bration data, measured from a single transducer on a healthy gearbox, are available for optimising the respective models. Operating and machine condition information are extracted separately from the order tracked vibration signal and modelled using separate HMMs. Information from the operating condition HMM is used to optimise a machine condition HMM for each operating condition state. The operating and machine condition information are statistically combined to automatically detect the relevance of each machine condition model, which is subsequently used to generate the discrepancy signal. The discrepancy signal is processed to detect, locate and trend damage automatically. The discrepancy generation process holds the advantage that the machine condition can be inferred in the presence of distinct operating condition states such as idling, full load and for transient states within a measurement. Another major advantage of this approach is that it does not require historical fault data and it is more flexible and simpler to implement than physics-based models with the condition being easily inferred from the processed discrepancy signal.

In this article, $x(t)$ denotes a continuous function, $x[t]$ indicates a scalar at instant t , \mathbf{x}_t

indicates a vector or multidimensional feature at instant t and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ indicates a multidimensional dataset over the N samples in the considered time period.

2. Proposed methodology

The key steps of the proposed methodology, with its process diagram in figure 1, are motivated and discussed in this section.

2.1. Order tracked vibration signal

The vibration signal, measured from a transducer on a rotating machine, is best represented in the angle domain due to the characteristics of rotating machines [22]. The signal is transformed from the time to the angle domain for example by using the instantaneous phase of the shaft of interest, acquired from a shaft key, shaft encoder or from using tacholess order tracking techniques. Tacholess order tracking is suggested here, since it reduces the implementation and running cost of the fault diagnostic methodology and it can be impractical or even impossible to install shaft encoders [23]. The order tracked vibration signal is processed further and moving windows with short angular lengths are applied to extract features from the data which allows localised changes in the signal to be detected. The angular lengths of the windows, associated with the machine and operating condition feature extraction processes, are different and discussed in the subsequent sections.

2.2. Operating condition feature extraction, modelling and classification

Some authors [2, 8, 24] have stated that it is essential to incorporate operating condition information into the diagnostic process, to ensure that the correct machine condition is inferred from the data. Operating condition information was incorporated by Bartelmus and Zimroz [2] and Zimroz et al. [3] to successfully diagnose the condition of a gearbox in fluctuating operating conditions. In this paper, it is assumed that the operating conditions cannot be measured directly and therefore the instantaneous operating conditions need to be inferred from representative features extracted from the data.

The operating condition feature extraction and modelling process, shown in figure 2, is used in the methodology to make the discrepancy signal more robust under fluctuating operating conditions. The operating condition features reflect changes in operating conditions and it is

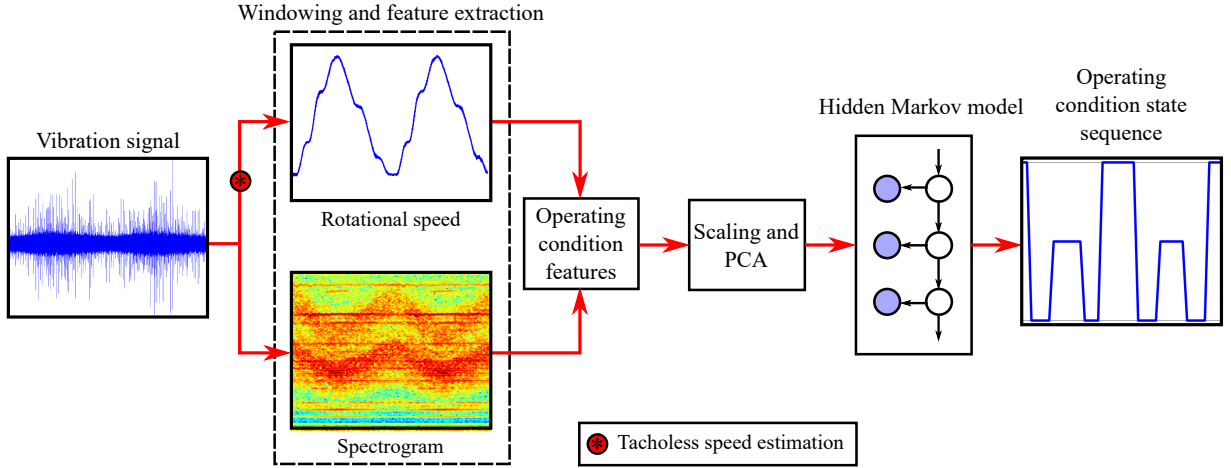


Figure 2: The operating condition feature extraction, modelling and state classification process. A tacholless (rotational) speed estimation process is required when a tacholless order tracking method is used. If the rotational speed obtained from a tachometer is available, it can be used instead.

expected that similar operating condition features will indicate that similar operating conditions are present. The operating condition states, at which the operating conditions are similar, are inferred from an operating condition HMM.

The first operating condition feature is the mean estimated rotational speed within one gear revolution. The second set of operating condition features are extracted from the spectrogram of the order tracked vibration signal. The spectrogram, is the squared magnitude of the short-time Fourier transform, and contains operating and machine condition information. However, if the operating condition feature windows are made a full gear revolution, it is assumed that the effects of localised faults will be masked by the mean operating condition and healthy machine condition effects. The spectrogram features are extracted in narrow bands of $2 \times k$ orders around the gear mesh frequency of the gearbox and its four harmonics, where $k = 1$ for the fundamental gear mesh frequency, $k = 2$ for its first harmonic etc.

The mean rotational speed and the spectrogram features have different magnitudes and the spectrogram features are expected to have redundant information contained within them. Linear scaling, is used to transform each feature to a new scale, consistently in the range of $(-10, 10)$, to ensure that the features are of the same order of magnitude. Linear scaling is performed on feature x

$$y = \frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}(x - x_{\min}) + y_{\min}, \quad (1)$$

to obtain a scaled representation y , with x_{\max} and x_{\min} obtained from the training set of the

specific operating condition feature and $y_{\min} = -10$ and $y_{\max} = 10$. Principal component analysis (PCA), is a linear dimensionality reduction technique, which is subsequently used to remove correlated or redundant information from the multidimensional operating condition feature space [19, 25].

In PCA, an eigenvalue analysis is performed to obtain the eigenvalues and the eigenvectors of the covariance matrix of the training feature set. The eigenvalues are used to determine the information content along the principal axes of the covariance matrix, with larger eigenvalues indicating that more information is present. Hence the d -eigenvectors associated with the d -largest eigenvalues, denoted by $\mathbf{V}_{1:d}$, are used to transform the original D -dimensional feature space to the d -dimensional feature space with [26]

$$\mathbf{y}_t = \mathbf{V}_{1:d}^T (\mathbf{x}_t - \boldsymbol{\mu}_x), \quad (2)$$

where \mathbf{x} and \mathbf{y} are D - and d -dimensional feature vectors and $\boldsymbol{\mu}_x$ is the mean of the healthy features. The relative information content of the D -principal components, estimated from the accumulative contribution rate (ACR) [14]

$$\beta_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{s=1}^D \lambda_s}, \quad (3)$$

is used as a guideline to determine the appropriate dimensionality of the new feature space. In equation (3), λ_i is the i th largest eigenvalue of the aforementioned covariance matrix, and β_k denotes the ACR associated with the first k -principal components.

A Hidden Markov model (HMM) is optimised on the processed operating condition features from the healthy gearbox. A HMM is a latent variable model which is able to model data with strong sequential characteristics [19, 27]. A HMM is used because the operating condition features are expected to contain strong sequential characteristics due to the rotational inertia of the machine. The latent variable in the operating condition HMM, denoted by \mathbf{z}_t^o , follows a Markov process and is linked to the operating condition features by Gaussian observation distributions. The HMM parameters are obtained by using the Baum-Welch algorithm [19, 27] with a maximum likelihood objective function.

The inferred hidden state sequence of the operating condition features, is obtained from the

Viterbi algorithm [19, 27], and is also referred to as the operating condition state sequence in this article. The operating condition states cluster similar features together which implies that similar operating conditions are present if the same operating condition state is present. The inferred operating condition state sequence is used in the machine condition feature extraction and modelling phase, considered in the next section.

2.3. Machine condition feature extraction and modelling

Machine condition features are required to be sensitive to changes in machine condition and need to be fairly insensitive to operating condition changes. The continuous wavelet transform, a form of wavelet analysis, has been very successful in the fault diagnostic field [13, 28–33] due to its sensitivity to impulses and singularities induced by damage in the rotating machine components [34]. The continuous wavelet transform (CWT) in the form of

$$W(a_1, a_2) = \frac{1}{\sqrt{a_1}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - a_2}{a_1} \right) dt, \quad (4)$$

is used to calculate the wavelet coefficient $W(a_1, a_2)$ at scale a_1 and translation a_2 for vibration signal x and the complex conjugate of the wavelet basis function, denoted by ψ^* . The problem with the CWT is that it contains much redundant information, as opposed to the discrete wavelet transform and the wavelet packet transform.

The performance of wavelet analysis is sensitive to the choice of wavelet basis function [28] with the Meyer basis function being used in this paper, from its performance in the article by Jedlinski and Jonak [13]. The CWT is evaluated with 20 scales, uniformly spaced with a bandwidth of 3 orders around the gear mesh frequency of the monitored gearbox and its four harmonics. The gear mesh frequency contains diagnostic information and is easily calculated for novelty detection applications. The wavelet coefficients of the 100 scales (20 scales at each of the five gear mesh frequency bands that are used), are windowed into equal angular windows from which the machine condition features, listed in table 1, are extracted. The features allow

Table 1: Machine condition features extracted from the windowed wavelet coefficients at each scale.

(1) Energy	(2) Skewness
(3) Kurtosis	(4) Root-mean-square (RMS)

changes in the characteristics of the wavelet coefficients to be detected. The angular length of the

machine condition windows are $2\pi/N_{teeth}$ rad with an overlap of π/N_{teeth} rad between adjacent windows, where N_{teeth} denotes the number of teeth on the gear of interest. The windows are made sufficiently long to ensure that the features, presented in table 1, are not sensitive to non-diagnostic related outliers, but are sufficiently short to allow fault localisation.

The four features extracted from the 100 scales, result in a 400 dimensional feature space which requires many parameters and much training data to be modelled. PCA is applied to the machine condition features so that the dimensionality of the feature space is reduced by removing the redundant information or correlated features from the final feature space.

The information from the operating condition classification process is used in the machine condition feature modelling process to ensure that its prediction is more robust to operating condition changes. The machine condition model optimisation process is illustrated in figure 3. The operating and machine condition feature extraction processes work on the same signal

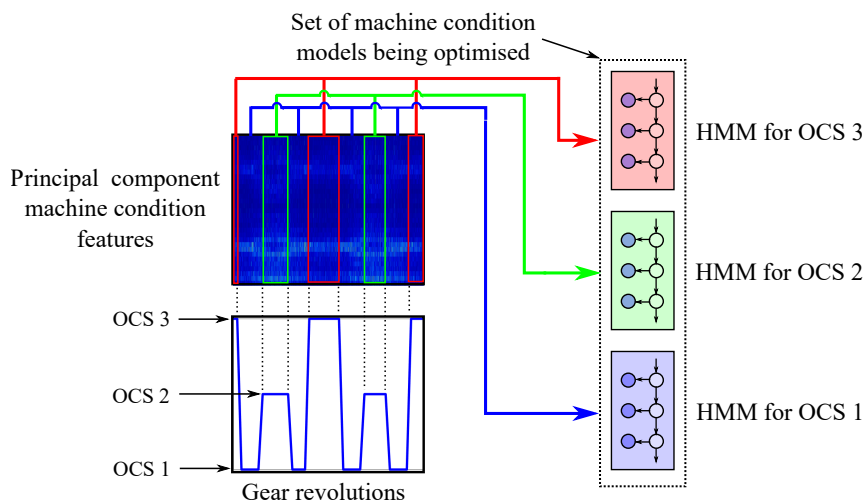


Figure 3: Machine condition model training process using the machine condition features trained with the operating condition information, in the form of the operating condition state (OCS) sequence.

and therefore the extracted operating condition states can be used to label the corresponding machine condition features as well. The idea is to label all the machine condition features of a revolution to a specific operating condition state and use these features to optimise the machine condition HMM associated with the operating condition state. It is assumed that the machine condition features, extracted from the same operating condition state, have similar characteristics for a healthy system and therefore it makes the output of the machine condition models operating condition independent. HMMs have been successful in the rotating machine diagnostics field [9, 31, 35–37] and it provides more discriminatory power than similar alternatives

such as Gaussian mixture models [37] and Gaussian distributions.

2.4. Discrepancy signal generation

The discrepancy signal generation process is performed for a N_{ocs} -state operating condition HMM with N_{ocs} machine condition HMMs being used. The discrepancy measure at instant t , is derived by considering the joint distribution over the machine condition features $\mathbf{b}_t \dots \mathbf{b}_1$, the operating condition features \mathbf{o}_t , the operating condition model latent variables \mathbf{z}_t^o , the operating condition model $\boldsymbol{\theta}_o$ and the set of the N_{ocs} machine condition model parameters, $\{\boldsymbol{\theta}_b\}$, written as

$$p(\mathbf{b}_t, \mathbf{z}_t^o, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \mathbf{o}_t, \mathbf{b}_{t-1}, \dots, \mathbf{b}_1) = p(\mathbf{b}_t | \mathbf{z}_t^o, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \mathbf{o}_t, \mathbf{b}_{t-1}, \dots, \mathbf{b}_1) p(\mathbf{z}_t^o | \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \mathbf{o}_t) \times \quad (5)$$

$$p(\{\boldsymbol{\theta}_b\} | \boldsymbol{\theta}_o, \mathbf{o}_t) p(\boldsymbol{\theta}_o | \mathbf{o}_t) p(\mathbf{o}_t) p(\mathbf{b}_{t-1}, \dots, \mathbf{b}_1),$$

where the machine condition features at the previous time instants are included because hidden Markov models are used to model the machine condition features. Note that $\{\boldsymbol{\theta}_b\}$ contains the model parameters of each machine condition model j which is denoted by $\boldsymbol{\theta}_b^j$. The model parameters and the operating condition latent variable are not important for diagnostic purposes and are as a result marginalised out to obtain

$$p(\mathbf{b}_t, \mathbf{o}_t, \mathbf{b}_{t-1}, \dots, \mathbf{b}_1) = \sum_{j=1}^{N_{\text{ocs}}} \iint p(\mathbf{b}_t, \mathbf{z}_{tj}^o, \{\boldsymbol{\theta}_b\}, \boldsymbol{\theta}_o, \mathbf{o}_t, \mathbf{b}_{t-1}, \dots, \mathbf{b}_1) d\boldsymbol{\theta}_o d\{\boldsymbol{\theta}_b\}, \quad (6)$$

which is used to obtain the conditional distribution

$$p(\mathbf{b}_t | \mathbf{o}_t, \mathbf{b}_1, \dots, \mathbf{b}_{t-1}) \approx \sum_{j=1}^{N_{\text{ocs}}} p(\mathbf{b}_t | \mathbf{z}_{tj}^o, \hat{\boldsymbol{\theta}}_b^j, \mathbf{b}_1, \dots, \mathbf{b}_{t-1}) p(\mathbf{z}_{tj}^o | \mathbf{o}_t, \hat{\boldsymbol{\theta}}_o), \quad (7)$$

using the conditional independence properties of the random variables and making the assumption that $p(\{\hat{\boldsymbol{\theta}}_b\}, \hat{\boldsymbol{\theta}}_o)$ is sharply peaked around the set of estimated machine and operating condition model parameters. The latter assumption is reasonable if much training data are used [26]. The output from each machine condition HMM is weighted by the posterior distribution of the operating condition HMM in equation (7). This means that the relevance of each machine condition model is automatically determined at each time step. Even though discrete latent states are used for the operating condition HMM, it is expected that as the number of latent

states is increased, larger operating condition fluctuations can be dealt with. The discrepancy measure, in the form of

$$\eta[t] = -\log p(\mathbf{b}_t | \mathbf{o}_t, \mathbf{b}_1, \dots, \mathbf{b}_{t-1}), \quad (8)$$

is evaluated at each machine condition window t to obtain the discrepancy signal. The discrepancy signal describes the deviation of the new data from the healthy machine's behaviour, given the inferred operating conditions at each machine condition window. The discrepancy signal contains information related to the gear and the pinion and therefore additional signal processing is necessary to determine the condition of the gears within the gearbox.

2.5. Processing the discrepancy signals

The impulses and subsequent discrepancies generated by gear damage are synchronous with the shaft to which the gear is fixed and are non-synchronous with the other shafts on the gear train. Hence, the synchronous average of the discrepancy signal in the form of

$$\mu_\eta[l] = \frac{1}{N_r} \sum_{i=0}^{N_r-1} \eta[l + iN_s], \text{ where } 1 \leq l \leq N_s, \quad (9)$$

is a useful tool for analysing discrepancy signals [18, 20]. The synchronous average $\mu_\eta[l]$ at position l on the gear is calculated over N_r rotations with N_s samples per rotation. The ability of the synchronous average to attenuate the non-synchronous components within a signal depends on the characteristics of the non-synchronous components as well as the number of shaft revolutions, N_r , that are used [38].

A second synchronous averaging process, over a set of consecutive N_m measurements, is useful when the number of revolutions is insufficient to attenuate the non-synchronous components [21]. The working principle of the second synchronous averaging process, proposed by Schmidt et al. [21], is illustrated in figure 4. The synchronous averages of the N_m discrepancy signals are aligned to have zero relative phase difference to ensure that the synchronous components are retained in the second synchronous averaging process. The signal alignment can be performed using tachometers or by using cross-correlation maximisation techniques if a tachless order tracking method is used [21]. The second synchronous average can be susceptible to an increased noise floor if phase estimation errors occur, additional noise are present during the experiments, etc. This problem can be circumvented by using a bias estimation and subtraction procedure on the

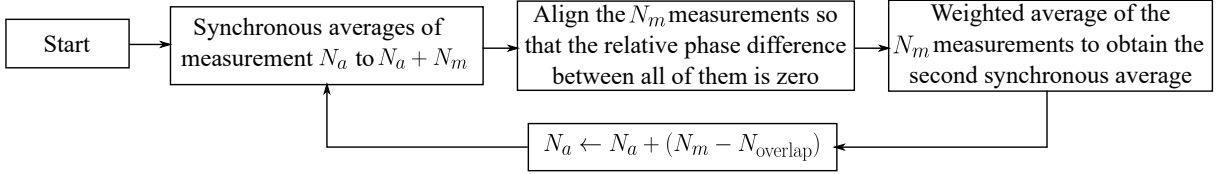


Figure 4: The second synchronous averaging process performed over a set of N_m measurements, where N_a is initially set to 0. The number of measurement overlap for two averaging processes are N_{overlap} . The mean and variance of the N_m synchronous averages are obtained for each unique N_a .

second synchronous average to avoid potential false alarms [21]. Note that the bias subtraction process is only performed when localised damage is present and other methods need to be used when investigating for distributed damage such as wear.

If the second synchronous average of the discrepancy signal is correctly implemented and the necessary provisions, such as bias estimation and subtraction, are made, it will only contain diagnostic information. A confidence bound is proposed, using statistical theory, on the synchronous average of the healthy gearbox to automatically generate an alarm threshold. The second synchronous average of the healthy vibration signal is estimated using point estimation methods. Hence, a confidence interval can be created which is expected to contain the true mean of a healthy synchronous averaged discrepancy signal with $100 \times (1 - \alpha)\%$ confidence. The upper confidence bound (CB) on the mean of the synchronous average (i.e. the second synchronous average)

$$\mu_{\eta}^{\text{CB}}[i] = \mu_{\eta}^{(2)}[i] + \mathcal{T}_{\alpha, N_m - 1} \frac{\sigma_{\eta}^{(2)}[i]}{\sqrt{N_m}}, \quad (10)$$

is used as an alarm threshold in this article and is calculated from the Student-t distribution because the population variance is unknown. The Student-t distribution, with $N_m - 1$ degrees of freedom, is calculated for a confidence bound of $100 \times (1 - \alpha)\%$ and is denoted by $\mathcal{T}_{\alpha, N_m - 1}$. The superscript (2), in equation (10), indicates the statistics of the synchronous average obtained from the second synchronous averaging process. The confidence bound, $\mu_{\eta}^{\text{CB}}[i]$, is calculated at position i on the gear using the sample mean and the sample standard deviation of N_m healthy measurements and is used with the following criteria

$$\text{Condition} = \begin{cases} \text{Expected behaviour,} & \text{if } \mu_{\eta}^{\text{CB}}[i] - \mu_{\eta}^{(2)}[i] \geq 0 \\ \text{Novelty is observed at } i, & \text{if } \mu_{\eta}^{\text{CB}}[i] - \mu_{\eta}^{(2)}[i] < 0 \end{cases}, \quad (11)$$

to infer the condition of the gear at position i .

The gear-pinion discrepancy distribution, proposed by [21], visualises the condition of the two gears within a single-stage gearbox simultaneously, and it aids with the condition inference process. It is calculated directly from the discrepancy signals of the gear and the pinion respectively and presents the joint discrepancy of a location on the gear and a location on the pinion. The discrepancy of the gear and the pinion are multiplied and averaged in the calculation process, which results in a bias in the distribution if localised damage is present [21].

Inferring the condition of the gear is important and if damage is detected it is important to analyse the stability of the damage growth as well. The second synchronous average of the discrepancy signal can be examined over time, and then changes in condition can be investigated. If localised gear damage is present, it is possible to utilise a healthy-damaged portion separation algorithm as proposed by Schmidt et al. [21]. This holds the advantage that the standard deviation of the damaged component can also be investigated over time, which is used to critically investigate the fault growth rate over time. The k -means clustering algorithm, as described in [19], is used to find two clusters within a second synchronous average. The clusters with the smaller and larger centres are labelled as healthy and damaged, respectively and trended over time. The results can be smoothed to reduce the noise content of the healthy and damaged portion [21].

3. Experimental setup

The developed fault diagnostic methodology is validated on data acquired from an experimental setup shown in figure 5. An electrical motor supplies rotational energy to the system, where an alternator, connected to a resistor bank, applies a counteracting load. The rotational

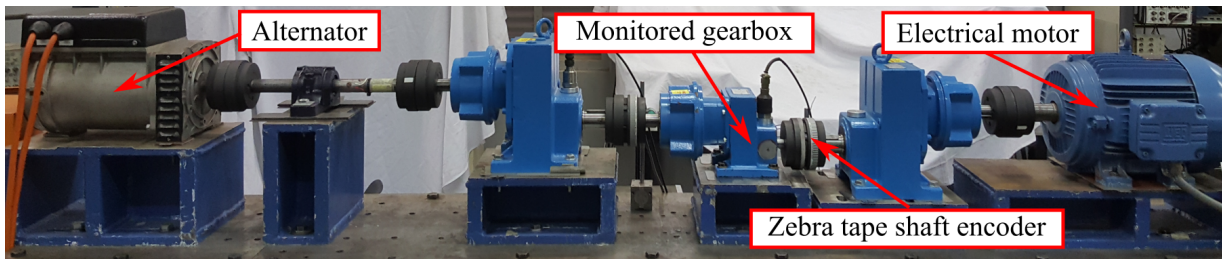


Figure 5: Experimental setup

speed of the electrical motor and the load applied by the alternator is controlled through a per-

sonal computer, which allows fluctuating loads and speeds to be applied. A zebra tape shaft encoder and an optical probe are located on the input shaft of the monitored gearbox and are used to measure the instantaneous angular speed of the input shaft of the monitored gearbox. The axial acceleration was measured on the bearing housing of the monitored gearbox with a 100mV/g tri-axial accelerometer. The experimental data were measured using an Oros OR35 data acquisition system.

The operating conditions, at the input shaft of the monitored gearbox, are presented in figure 6. The torque in figure 6a was estimated from the voltage and current generated from the

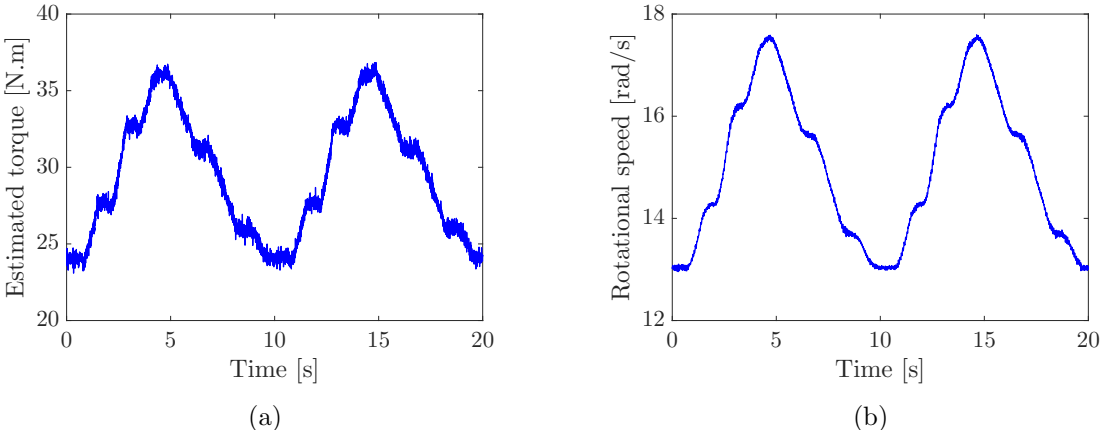


Figure 6: The experimental operating conditions

alternator and the speed in figure 6b was calculated from the zebra tape shaft encoder tachometer signal. The operating conditions in figure 6 were chosen to reflect a machine operating between various operating condition states with a large relative difference between the maximum and minimum states and are a significant simplification for the operating conditions seen typically in bucket wheel excavators and wind turbines for example.

After measurements were taken with the healthy gear, the gearbox was disassembled so that damage could be induced on the healthy gear. This was achieved by seeding a slot into the root of a gear tooth as shown in figure 7a. The slot was along the entire width of the tooth, was 50% of the tooth thickness deep and it had a height of 0.3mm. The experiment was performed continuously with regular measurements taken until the damaged tooth completely failed. The gear tooth failed approximately after 20 days of experiments, with the gear after failure shown in figure 7b.

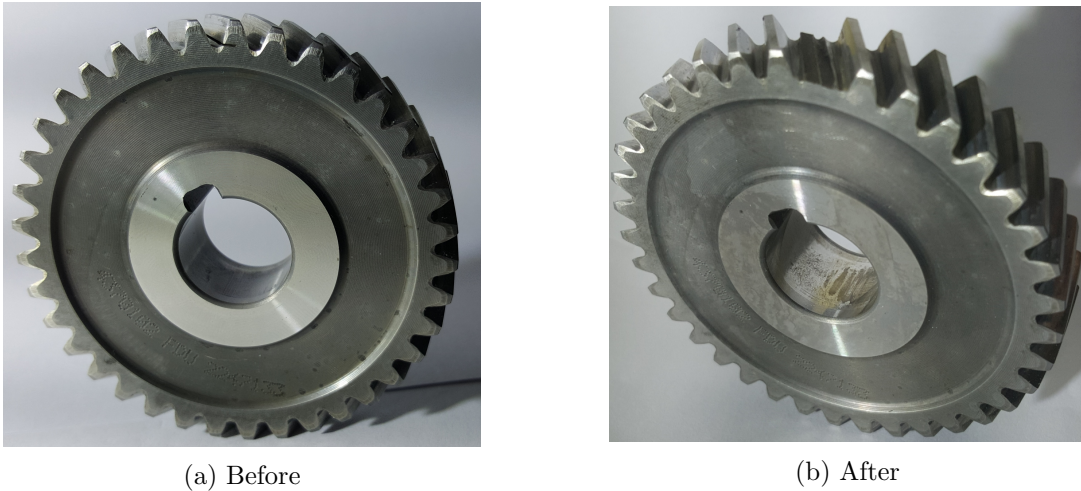


Figure 7: The gear before and after the experiment was completed.

4. Results

The proposed methodology is evaluated on the experimental data discussed in the previous section. The operating condition feature extraction and modelling results are presented and discussed, whereafter the machine condition feature extraction, -modelling, discrepancy generation and processing phases are presented and discussed.

4.1. Operating condition state classification

A set of vibration measurements was acquired from a gearbox in a healthy condition, whereafter it was order tracked using the tachless order tracking method proposed by Schmidt et al. [39]. Operating condition features were extracted from the order tracked vibration signal, scaled and transformed to a lower dimension using principal component analysis (PCA) as outlined in section 2.2. The ACR threshold for the 18 dimensional PCA feature space was set to 0.80, which resulted in a new feature space with a dimensionality of 5. The first two principal components of the operating condition features are shown in figure 8a for the training data. The sequential nature of the data is highlighted in figure 8b, where the first 45 data points are connected. A Hidden Markov model was optimised with three operating condition states, because it provided the ideal compromise between model complexity and prediction performance. It should be emphasised that if more operating condition states are used, a better resolution is obtained in the operating conditions. However, there is a risk of rarely visiting a specific operating condition state, which results in an insufficient amount of training data being available when optimising

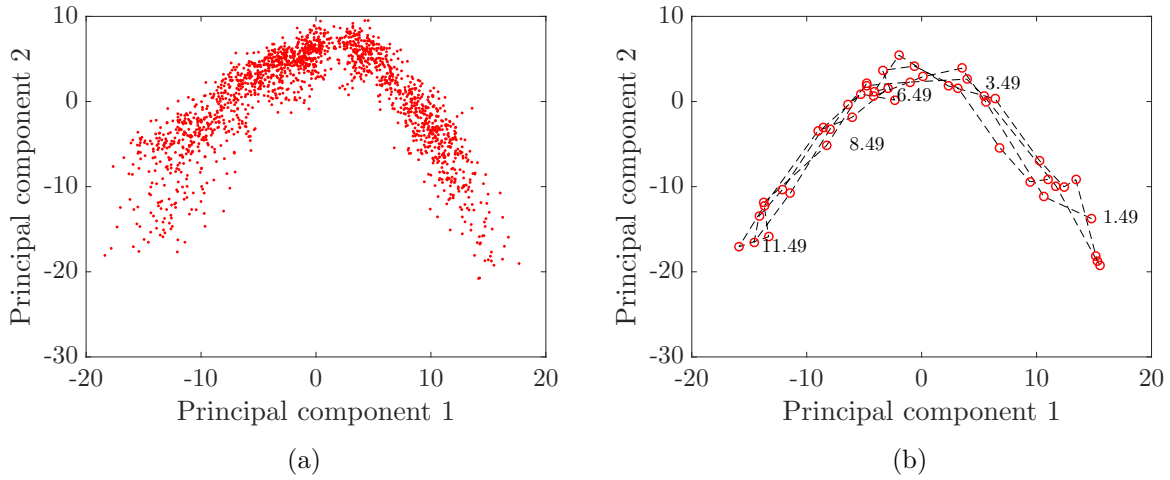


Figure 8: The operating condition principal component feature space of the training data is shown in figure 8a. The first 45 data points are connected, with the gear revolution at which the coordinates correspond to, shown in figure 8b.

the associated machine condition model. The inferred operating condition state sequence for the operating condition features set is superimposed in figure 9 with the corresponding rotational speed. If the results in figure 9 are compared to the result in figure 8b, it can be observed

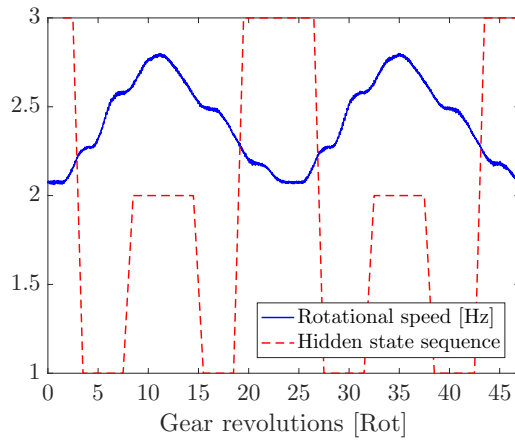


Figure 9: The operating condition state sequence, inferred by the operating condition HMM, is superimposed on the rotational speed of the input shaft of the gearbox.

that the coordinates in the PCA feature space correspond to the instantaneous operating conditions. This validates that the operating condition feature extraction, modelling and classification process works.

4.2. Machine condition feature extraction, modelling and discrepancy generation

The machine condition features are extracted as described in section 2.3. The first six principal components of the machine condition features have an ACR of 0.999491, which indicate that this captures most of the information content in the data. The processed machine condition features, with a dimensionality of six, are labelled with the same operating condition state as the overlapping operating condition features. The three machine condition models are optimised with the process shown in figure 3.

The model complexity of the machine condition models are determined by evaluating the likelihood of the machine condition features, from the validation dataset. The model optimisation may differ for different initialisation points and therefore a statistical analysis is performed. The average performance and the bound, obtained from the standard deviation, of 20 machine condition model optimisation runs are compared in figure 10. Note that the machine condition

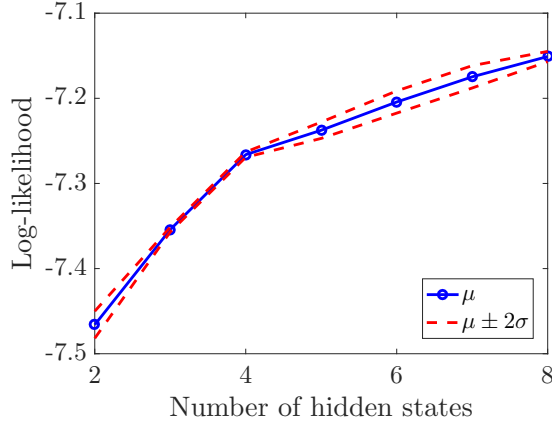


Figure 10: The log-likelihood of the training dataset versus the number of hidden states (i.e. model complexity) for the machine condition models.

model, associated with each operating condition state, has the same model complexity. It is observed in figure 10 that if more than four hidden states are used for the machine condition HMMs, the gradient of the performance of the HMM decreases significantly. This indicates that the model performance starts to saturate, even though the model complexity increases which motivates using four hidden state machine condition HMMs. A summary of the operating condition and machine condition model characteristics is presented in table 2.

The discrepancy signal is generated, as described in section 2.4, for the validation data as well as the new or testing data, and is indicated as the negative log-likelihood (NLL) in the

Table 2: A summary of the characteristics used in the operating condition (OC) and machine condition (MC) modelling process.

	Model type for each model	Number of models	Number of hidden states per model	Raw feature dimension	Processed feature dimension
OC	HMM	1	3	18	5
MC	HMM	3	4	400	6

figures. The discrepancy signal from the data of a healthy gearbox and from the data of a damaged gearbox, with a localised tooth fault, is compared in figure 11. It is observed that the

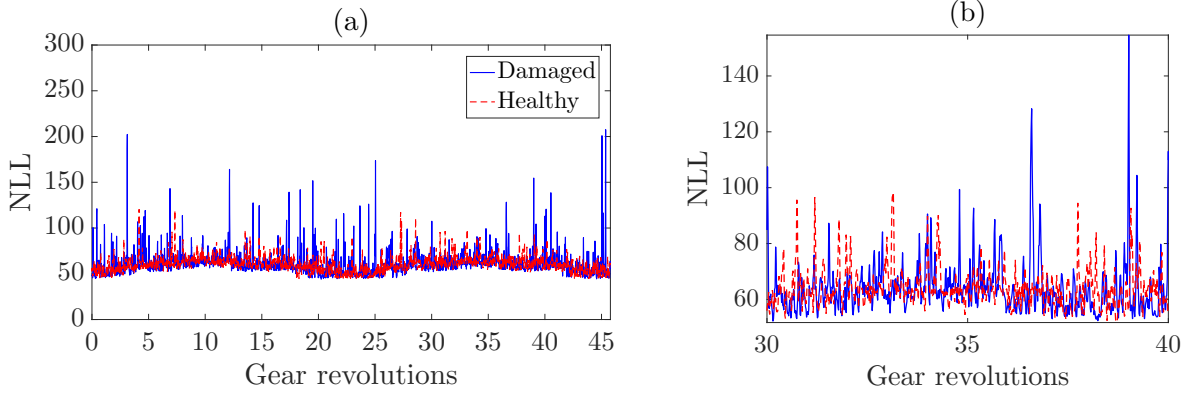


Figure 11: The discrepancy signals over gear revolutions of the gearbox, with healthy gears and the gearbox with the damaged gear, are compared. The full measurement period is shown in (a), while a zoomed view is shown in (b).

discrepancy signal, associated with the damaged gearbox, contains slightly larger values, but it is difficult to ascertain the characteristic or cause of the increase. Further processing techniques, as described in section 2.5, are required to correctly infer the condition of the gears within the gearbox.

4.3. Discrepancy signal processing results

The results of the processed discrepancy signal are given in this section. It should be noted that in all subsequent plots containing the discrepancy signal of the damaged gear over a single gear revolution, the damage on the gear is positioned manually to 180 degrees to make the comparison easier between different figures.

4.3.1. Synchronous averaging results

The first synchronous average of the NLL calculated from the healthy validation data, denoted by μ_v , and the resulting confidence bound from the NLL calculated from other validation data, denoted by CB_v , are superimposed in figure 12. The synchronous average of the NLL of the

validation data is fairly uniform and does not exceed the confidence bound. If equation (11) is used, it is observed that the gear is in a healthy condition and this validates that the proposed method is able to correctly diagnose a healthy gear.

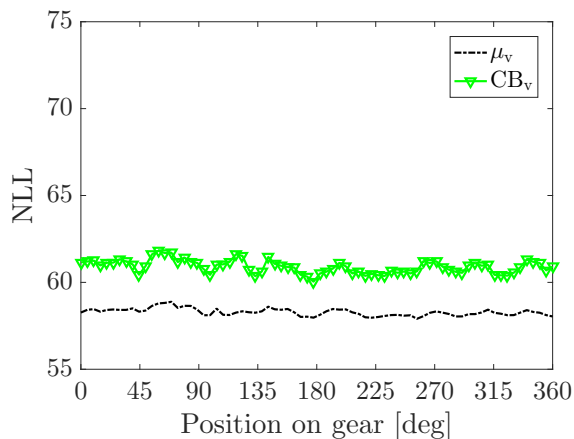


Figure 12: Synchronous average of the discrepancy signal generated from the validation data (i.e. healthy gearbox data not used during model optimisation), denoted by μ_v , and the generated confidence bound, denoted by CB_v which is used as the alarm threshold. In equation (10), $\alpha = 10^{-4}$.

The synchronous average of the NLL of two datasets, from a gearbox with a damaged gear during different measurement times, is denoted by μ_t and is compared in figure 13 to the results in figure 12. The data of the damaged gear, presented in figure 13a, are from the first set of

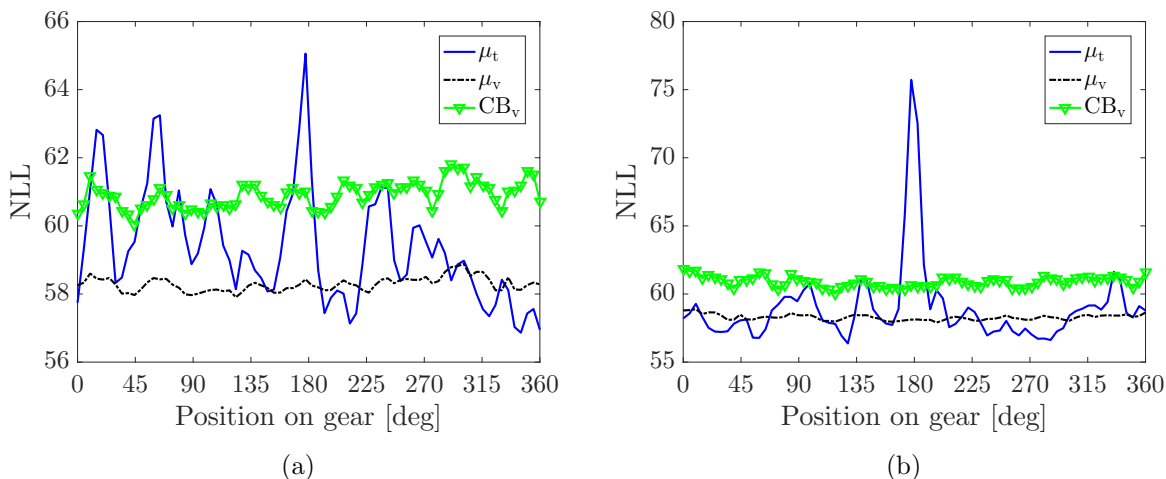


Figure 13: The first synchronous average of the first measurement is shown in 13a and the measurement taken a week later is shown in 13b of the damaged gearbox and is denoted by μ_t . It is compared to the confidence bound, CB_v , and the synchronous average of the validation data, μ_v , that is shown in figure 12 as well.

measurements after the gearbox was reassembled with the damaged gear and the data in figure 13b are from a measurement taken after a week of experiments. It is observed in figure 13a, that

various portions of the synchronous average of the tested dataset exceed the alarm threshold and if equation (11) is used, it seems that many novelties are present. This incorrectly indicates that more than one failure mode is present on the gear. This is an undesirable result, because the gear is incorrectly represented with the synchronous average in figure 13a.

The synchronous average in figure 13b has improved significantly, since the localised fault at 180 degrees is clearly observed. This is because the damaged have increased between the two measurements and possibly because the machine components settled to their original position after the disassembling and reassembling process. However, some portions of the synchronous average exceed the confidence bound, possibly due to noise etc., which incorrectly represents the true condition of the gear as well. In the next section, the second synchronous average is employed to circumvent the aforementioned problems and to obtain the true representation of the condition of the gear.

4.3.2. Second synchronous averaging results

The discrepancy signals associated with the 20 measurements, that are used in the second synchronous averaging process, are aligned using the cross-correlation maximisation procedure proposed by Schmidt et al. [21]. The statistics of the synchronous average (μ_t and σ_t) for the first 20 measurements, obtained from the second synchronous average process, are given in figure 14a. Note that the entire mean, μ_t obtained from the second synchronous average of the NLL,

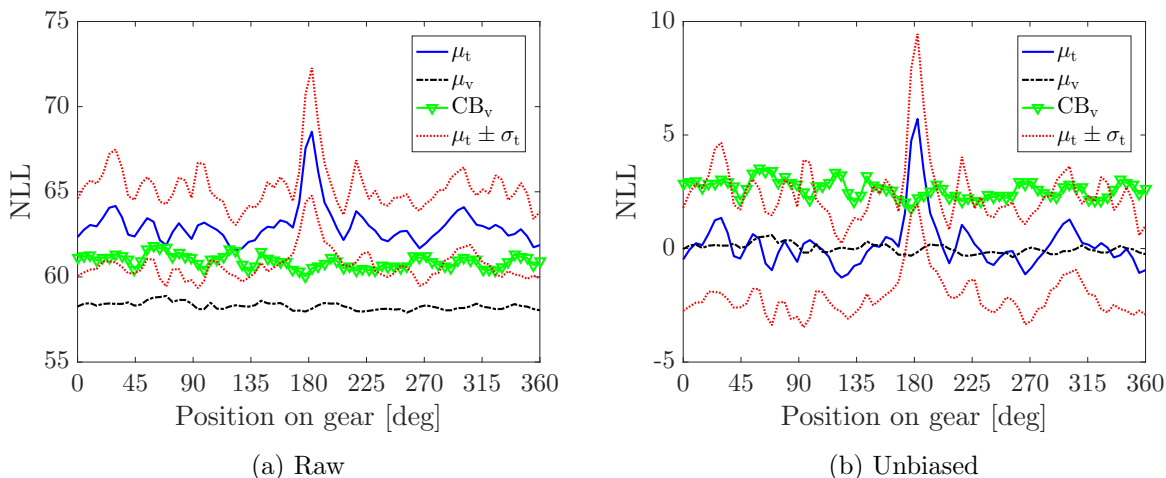


Figure 14: The second synchronous average and the unbiased second synchronous average of the first set of measurements are compared in this figure to the alarm threshold in the form of a modified confidence bound.

exceeds the alarm threshold, which is undesired. The bias is attributed to cross-correlation errors

because the damage is not well developed and the presence of noise in the initial measurements after the disassembling and reassembling process. Hence, this makes it necessary to calculate the unbiased second synchronous average, with the bias estimation and subtraction process proposed by Schmidt et al. [21].

The result of the unbiased second synchronous averaging process in figure 14b is a better representation of the actual condition of the gear. It can be observed that damage is located at 180 degrees, while the other portions of the gear are healthy. Note that the mean of the NLL of the validation data is subtracted from the NLL of the validation data and the confidence bound to ensure that the results can be compared. The bias subtraction process is allowed because it is expected that the bias in the NLL of a healthy gear and of the healthy portions of the damaged components need to be same. Hence, the unbiased second synchronous average is sensible when investigating localised damage on a gear and other techniques must be developed if distributed damage is present on the gear. The result of the second synchronous averaging process of 20 measurements, taken approximately a week after the experiments started, is shown in figure 15. The unbiased second synchronous average in figure 15 indicates that prominent localised damage

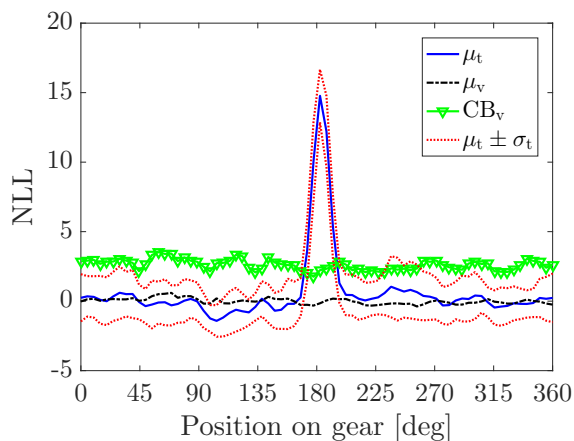


Figure 15: The unbiased second synchronous average of 20 measurements taken after a week of experiments.

is present on the gear and that the damage evolved from the result in figure 14b.

The unbiased second synchronous average of the healthy pinion is considered in figure 16, where it is observed that the entire pinion is healthy when compared to the confidence bound. The results indicate that it is possible to infer the presence of localised damage within a gearbox automatically, by using the second synchronous averaging process and the confidence bound on the gear and the pinion discrepancy signals, respectively. The second synchronous average is

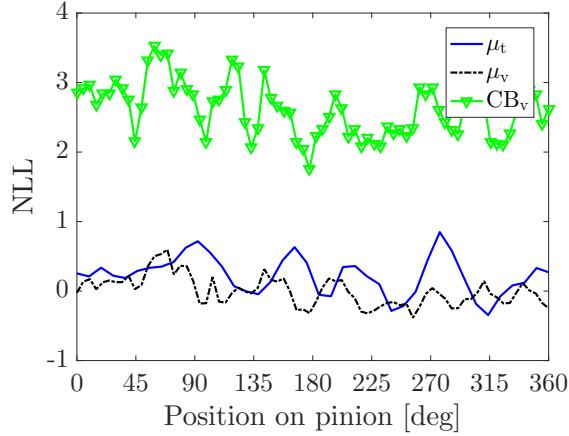


Figure 16: The unbiased second synchronous average of the pinion which is compared to the confidence bound and the discrepancy signal of the validation data.

significantly smoother than the first synchronous average result which makes it more robust and less prone to false alarms when inferring the machine condition.

4.3.3. Gear pinion discrepancy distribution

The gear-pinion discrepancy distribution, proposed by [21], is investigated to support the condition of the gearbox that was inferred in the previous section. The information of the condition of the gear and the pinion is contained within the discrepancy signal, which is used to calculate the gear-pinion discrepancy distribution. The gear-pinion discrepancy distribution of the healthy gearbox and the damaged gearbox is shown in figure 17. It is observed that a fairly

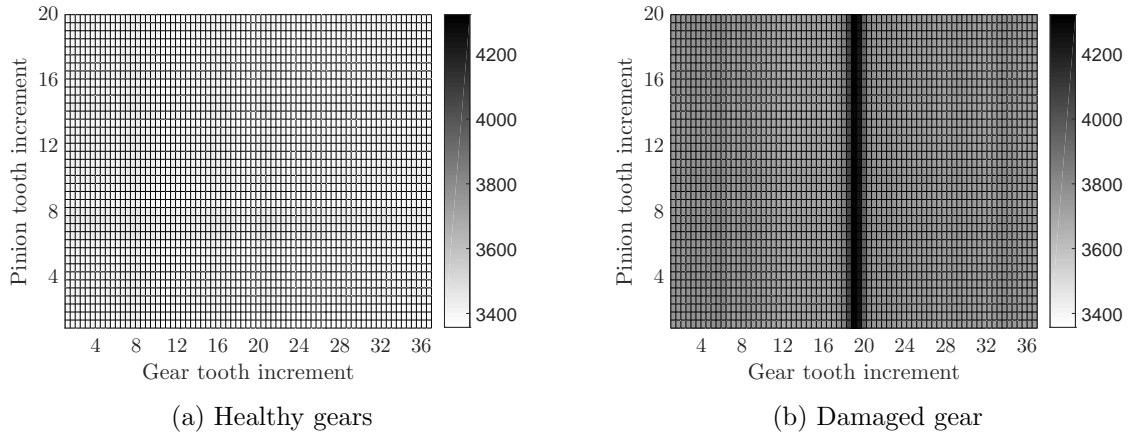


Figure 17: Gear-pinion discrepancy distribution for a gearbox with healthy gears and a gearbox with a damaged gear. A localised fault is present on one of the teeth of the damaged gear.

uniform distribution is obtained for the healthy gear and that the distribution, associated with the gearbox with the damaged gear, clearly indicates the presence of localised gear damage. The

bias in figure 17b is attributed to the presence of large discrepancies attributed to the damage within the discrepancy signals of the gear and the pinion as described in section 2.5 and by [21]. It is concluded from the distribution in figure 17a and 17b, that the pinion is in a healthy condition.

4.4. *Fault trending*

The moment damage is detected and located, its severity needs to be estimated. It is difficult to estimate the severity of the damage in terms of absolute quantities (i.e. dimension of the crack etc.), however it is possible to compare the discrepancy associated with the damaged portion to the healthy portion and to determine the stability of the damage evolution by comparing its value over time. The second synchronous average of the experimental data over normalised operational time is compared in figure 18. This three dimensional plot indicates the change in

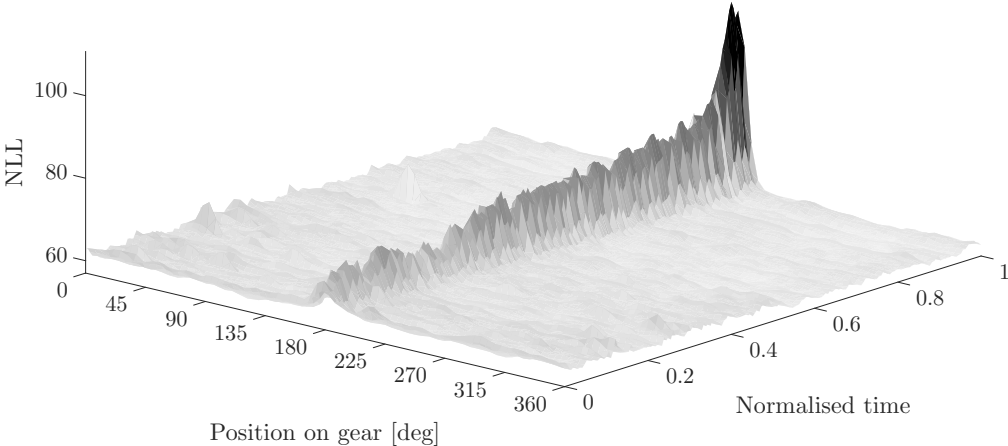


Figure 18: Second synchronous average of the damaged gearbox over measurement time using 20 measurements and an overlap of 80% between measurements in the second synchronous averaging process.

damage severity until the failure occurs, which validates that the presented approach is able to locate and trend damage until the failure occurs. It is often useful to investigate the higher order statistics, such as the variance, of a sampling distribution as well. This provides more insight into the deterministic nature of the damage and provides more confidence in the second synchronous average results. The healthy-damaged portion separation algorithm, proposed by Schmidt et al. [21], is used with the k -means algorithm to decompose the second synchronous average of the discrepancy signal into its healthy and damaged portions and is shown in figure 19. It is possible

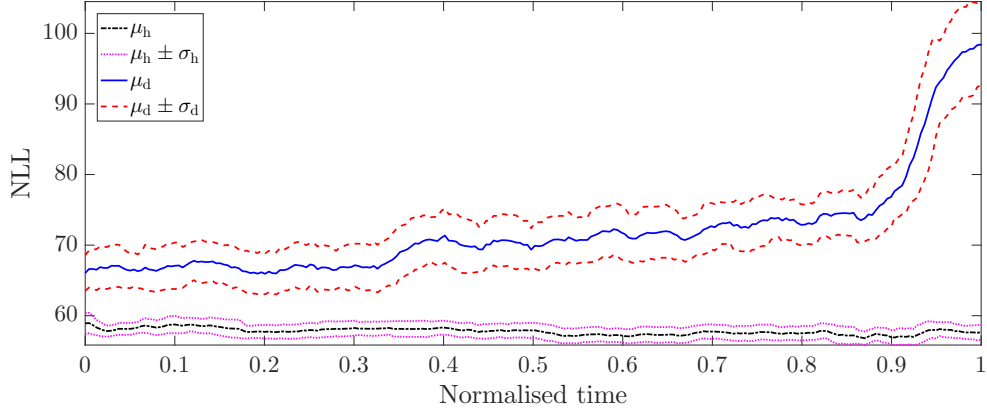


Figure 19: Healthy and damaged portion of the discrepancy signal compared over normalised time of the damaged experiment until the gear tooth failed. The mean μ and standard deviation σ of the healthy and damaged measurements are denoted by the subscripts h and d , respectively.

to determine the healthy and damaged portions within the discrepancy signal and it is easy to track the growth in severity of the damage until the time of failure. The results indicate that the healthy and damaged modes are well separated which indicates that the damage can be detected sufficiently early to make proper maintenance decisions.

5. Conclusion

The diagnostic methodology, presented in this article, is proposed for gearboxes operating under fluctuating conditions and is validated on experimental data. It is shown that it is possible to generate a discrepancy signal which is robust to fluctuating operating conditions and can be processed to detect, locate and trend gear damage over time. The methodology only requires healthy vibration data from a single vibration transducer which is placed on a rotating machine for model optimisation. This overcomes many of the practical limitations of supervised learning condition monitoring methodologies. The proposed methodology automatically determines the operating condition states after which the relevance of each machine condition model is automatically determined. The processed discrepancy results are easy and intuitive to interpret which holds many advantages in the condition monitoring field. The confidence bound and the unbiased second synchronous average are useful for performing automatic localised damage detection. The proposed methodology also holds the advantage that it is quite flexible and it can be extended to include other features which might provide a better performance than the features used in this investigation.

The authors envisage that this technique can be applied to machines which operate under operating condition states which differ significantly such as idling, full power and transient states. The implications of this is that the methodology will automatically determine, from the data, the relevance of the machine condition models which are used to generate a discrepancy signal. This makes the methodology robust to fluctuating operating conditions.

The authors suggest that future investigations, using the methodology, need to be focussed on data from industrial machines such as draglines, wind turbines etc. Investigations can be performed to have more optimal machine condition features, which can possibly make the methodology even more sensitive to changes in machine condition. The current focus is on localised gear fault detection and it needs to be extended to distributed gear and bearing fault detection, localisation and trending. This will have a positive impact on condition monitoring in many industrial environments such as the mining, energy and aeronautical industries.

Acknowledgements

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

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] 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.
- [3] 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.

- [4] 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.
- [5] R. Randall, A new method of modeling gear faults, *Journal of Mechanical Design* 104 (2) (1982) 259–267.
- [6] C. J. Stander, P. S. Heyns, Transmission path phase compensation for gear monitoring under fluctuating load conditions, *Mechanical Systems and Signal Processing* 20 (7) (2006) 1511–1522.
- [7] J. Lin, M. Zhao, A review and strategy for the diagnosis of speed-varying machinery, in: *Prognostics and Health Management (PHM), 2014 IEEE Conference on, IEEE, 2014*, pp. 1–9.
- [8] F. Chaari, W. Bartelmus, R. Zimroz, T. Fakhfakh, M. Haddar, Gearbox vibration signal amplitude and frequency modulation, *Shock and Vibration* 19 (4) (2012) 635–652.
- [9] H. Ocak, K. A. Loparo, HMM-based fault detection and diagnosis scheme for rolling element bearings, *Journal of Vibration and Acoustics* 127 (4) (2005) 299.
- [10] J.-D. Wu, C.-H. Liu, Investigation of engine fault diagnosis using discrete wavelet transform and neural network, *Expert Systems with Applications* 35 (3) (2008) 1200–1213.
- [11] M. Unal, M. Onat, M. Demetgul, H. Kucuk, Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network, *Measurement* 58 (2014) 187–196.
- [12] K. C. Gryllias, I. A. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, *Engineering Applications of Artificial Intelligence* 25 (2) (2012) 326–344.
- [13] Ł. Jedliński, J. Jonak, Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform, *Applied Soft Computing* 30 (2015) 636–641.
- [14] C. Y. Yang, T. Y. Wu, Diagnostics of gear deterioration using EEMD approach and PCA process, *Measurement* 61 (2015) 75–87.

- [15] D. Fernández-Francos, D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, Automatic bearing fault diagnosis based on one-class ν -SVM, *Computers & Industrial Engineering* 64 (1) (2013) 357–365.
- [16] T. Heyns, S. J. Godsill, J. P. De Villiers, P. S. Heyns, Statistical gear health analysis which is robust to fluctuating loads and operating speeds, *Mechanical Systems and Signal Processing* 27 (1) (2012) 651–666.
- [17] T. Heyns, P. S. Heyns, R. Zimroz, Combining discrepancy analysis with sensorless signal resampling for condition monitoring of rotating machines under fluctuating operations, *Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies* 2 (2) (2012) 52–58.
- [18] T. Heyns, P. S. Heyns, J. P. De Villiers, Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox, *Mechanical Systems and Signal Processing* 32 (2012) 200–215.
- [19] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2009.
- [20] P. S. Heyns, R. Vinson, T. Heyns, Rotating machine diagnosis using smart feature selection under non-stationary operating conditions, *Insight-Non-Destructive Testing and Condition Monitoring* 58 (8) (2016) 417–422.
- [21] S. Schmidt, P. S. Heyns, J. P. De Villiers, Discrepancy signal processing techniques for gearbox condition monitoring applications, *First World Congress on Condition Monitoring* (2017) Accepted.
- [22] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals, *Mechanical Systems and Signal Processing* 18 (6) (2004) 1285–1314.
- [23] 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.
- [24] W. Bartelmus, R. Zimroz, Vibration condition monitoring of planetary gearbox under varying external load, *Mechanical Systems and Signal Processing* 23 (1) (2009) 246–257.

- [25] S. Theodoridis, K. Koutroumbas, Pattern recognition, Elsevier/Academic Press, 2009.
- [26] C. M. Bishop, Neural networks for pattern recognition, Clarendon Press, 1995.
- [27] L. R. Rabiner, A tutorial on Hidden Markov Models and selected applications in speech recognition, Proceedings of the IEEE 77 (2).
- [28] W. J. Wang, P. D. McFadden, Application of wavelets to gearbox vibration signals for fault detection, Journal of Sound and Vibration 192 (1996) 927–939.
- [29] J. Lin, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, Journal of Sound and Vibration 234 (1) (2000) 135–148.
- [30] G. Dalpiaz, A. Rivola, R. Rubini, Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears, Mechanical Systems and Signal Processing 14 (3) (2000) 387–412.
- [31] Q. Miao, V. Makis, Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models, Mechanical Systems and Signal Processing 21 (2) (2007) 840–855.
- [32] J. Rafiee, M. A. Rafiee, P. W. Tse, Application of mother wavelet functions for automatic gear and bearing fault diagnosis, Expert Systems with Applications 37 (6) (2010) 4568–4579.
- [33] X. Y. Wang, V. Makis, M. Yang, A wavelet approach to fault diagnosis of a gearbox under varying load conditions, Journal of Sound and Vibration 329 (9) (2010) 1570–1585.
- [34] Z. K. Peng, F. L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography, Mechanical Systems and Signal Processing 18 (2) (2004) 199–221.
- [35] V. Purushotham, S. Narayanan, S. A.N. Prasad, Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, NDT & E International 38 (8) (2005) 654–664.
- [36] J. S. Kang, X. H. Zhang, Y. J. Wang, Continuous hidden Markov model based gear fault diagnosis and incipient fault detection, ICQR2MSE 2011 - Proceedings of 2011 International

Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (2011) 486–491.

- [37] T. Marwala, U. Mahola, F. V. Nelwamondo, Hidden Markov models and Gaussian mixture models for bearing fault detection using fractals, International Joint Conference on Neural Networks, Vancouver, Canada (2006) 3237–3242.
- [38] 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.
- [39] S. Schmidt, P. S. Heyns, J. P. de Villiers, A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process, Submitted to Mechanical Systems and Signal Processing (Reference number: MSSP17-78).