Localised gear anomaly detection without historical data for reference density estimation

Stephan Schmidt*, P. Stephan Heyns

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

Abstract

Performing condition monitoring on rotating machines which operate under fluctuating conditions remains a challenge, with robust diagnostic techniques being required for the condition inference task. Some of the developed diagnostic techniques rely on the availability of historical fault data, which is impractical or even impossible to obtain in many circumstances and therefore novelty detection techniques such as discrepancy analysis are used. However, discrepancy analysis assumes that the condition of the machine is the same throughout the signal in the model optimisation process i.e. no localised damage is present, which can pose problems if the training data unwittingly contain a component with localised damage. In this paper, an automatic procedure is proposed for diagnosing localised gear damage in the presence of fluctuating operating conditions, with no historical data being required to model the data used in the condition inference process. The continuous wavelet transform, principal component analysis and information theory are used to obtain divergence data of the gear under consideration. The divergence data are used with Bayesian data analysis techniques to automatically infer the presence of localised anomalies due to localised gear damage. The proposed technique is validated in two experimental investigations, with promising results being obtained.

Keywords: Diagnostics, Fluctuating operating conditions, Kullback-Leibler divergence, Bayesian

---

*Corresponding author.
Email address:* `stephanschmidt@zoho.com` (Stephan Schmidt)

1. Introduction

Gearbox vibration-based condition monitoring forms an integral part of condition-based maintenance when applied for rotating machines such as wind turbines [1, 2]. However, the performance of conventional vibration-based condition monitoring techniques is impeded when performed under unavoidable varying operating conditions and therefore more robust condition monitoring techniques are required [3–6]. Hence, much research has been conducted on signal processing and analysis methods to enhance the diagnostic information e.g. impulses induced by faults and to attenuate the non-diagnostic information e.g. operating condition changes for gear and bearing diagnostics [3, 7–16]. Examples of gear fault diagnosis techniques include synchronous averaging [17], residual signal analysis [18], the squared envelope of a narrowband signal [19], the synchronous variance [20], the empirical mode decomposition [21, 22], the instantaneous power spectrum [23] and wavelet analysis [24, 25]. Wang et al. [26] for example proposed a unified approach for rotating machine diagnostics where a residual signal is obtained by subtracting a reference signal from the vibration signal, whereafter a proposed condition indicator is used for rotating machine diagnosis.

However, the processed signals can be difficult to manually interrogate for the presence of damage and much data may need to be processed from a fleet of machines i.e. big data need to handled [27, 28]; this can impede the efficiency of the aforementioned techniques, especially when used by a non-expert. Hence, machine learning techniques are frequently combined with features extracted with signal analysis techniques to automatically infer the condition of the machine. Wavelet analysis [29–32], the empirical mode decomposition and its variants [22, 33], processed time-domain signal statistics [34, 35] and frequency domain statistics [35] have been used to obtain features. Recently, there are also approaches being developed where the features are extracted from the raw data by a model such as a convolutional neural network; the purpose of this is to avoid using techniques that are engineered for a specific problem [28, 36, 37]. Models such as neural networks [30, 32, 38], support vector machines [38, 39], radial basis function networks [29], and convolutional neural networks [37] have performed very well for automatically inferring the condition of rotating machines in a supervised learning framework. However, the performance of the

supervised learning approaches is sensitive to the amount and the quality of the available training data. For example, supervised learning approaches require historical data from all conditions to be available before the machine learning model can be optimised; this is impractical in many circumstances e.g. new machines do not have historical fault data [28, 35, 40, 41]. Physics-based [35] and novelty detection [40–48] approaches are therefore developed to circumvent the aforementioned problems in the condition monitoring field.

In novelty detection approaches, a model is optimised to represent the data of a machine in a reference condition, whereafter the optimised model is used to determine whether the new data are from a healthy source or not i.e. a novelty [40, 41]. Discrepancy analysis is a novelty detection approach, where a model of healthy gearbox data are used to assign localised novelty scores to vibration data, whereafter the discrepancy signal is analysed using conventional signal analysis techniques [43, 44, 47, 48]. The processed discrepancy signal is used to detect novelties due to damage and also to characterise the damage [47, 48]. However, in discrepancy analysis it is implicitly assumed that each portion of the gear is in the same condition when optimising the discrepancy model [47, 48] and therefore localised gear damage in the training dataset can adversely affect its performance. Hence, it is very important to ensure that localised damage is not present on the gears in the training dataset for discrepancy analysis.

It is also important to use efficient methods for localised gear damage detection, because it can be difficult to detect in industrial applications and it can significantly decrease the life of the gear [12, 24]. Ideally, a method is sought which can automatically infer the condition from the data, however, this requires at least historical data from a healthy machine to be available [47, 48]. Hence, it is necessary to develop a methodology which can automatically detect localised gear damage without knowing the distribution of the healthy data i.e. without having a reference density of the features available. This will allow localised gear damage, which can potentially result in an abrupt failure of a machine, to be automatically detected without historical data. Discrepancy analysis can then subsequently be used more effectively i.e. discrepancy analysis will not be erroneously implemented if localised damage is present.

Therefore, a novel diagnostic technique is proposed in this paper to automatically detect localised anomalies i.e. localised segments that differ statistically from the other

segments, using only the vibration data of a single measurement. This is performed by assuming that the data generated from the meshing of healthy gear should be statistically similar i.e. generated from the same distribution. After the gears are identified as being healthy, it then becomes possible to utilise more advanced approaches such as discrepancy analysis to detect localised and distributed changes in the condition of the gears. The purpose of this methodology is therefore not to replace techniques that utilise historical data such as discrepancy analysis or supervised learning approaches, but rather to address a special case where historical data are not available.

In summary, the proposed methodology offers the following advantages:

- Historical data are not required for model optimisation nor for automatic localised fault detection; this makes it suitable to be used in applications where no historical data are available and localised damage needs to be detected.

- It is simple to interpret the results i.e. manually detecting localised gear damage.

- It is robust to fluctuating operating conditions.

- The diagnostic metrics can be processed to detect, trend and visualise localised damage on the gear and it can consistently classify a healthy gear correctly as well.

The outline of the paper is as follows: The methodology is presented in Section 2, whereafter the proposed methodology is compared to conventional fault diagnosis techniques on experimental gearbox data in Section 3. In Section 4, conclusions are drawn from the results and recommendations are made for future work.

2. Methodology

The general methodology is presented in Figure 1(a), with the specific steps used in this paper shown in Figure 1(b). For the implementation used in this paper, it is assumed that an order tracked vibration signal is available, which was measured over N_r shaft rotations. A gear with N_{teeth} teeth is connected to the aforementioned shaft and is investigated for potential localised damage. Machine condition features are extracted from the windowed, processed vibration signal, which allow localised changes to be detected within the signal.
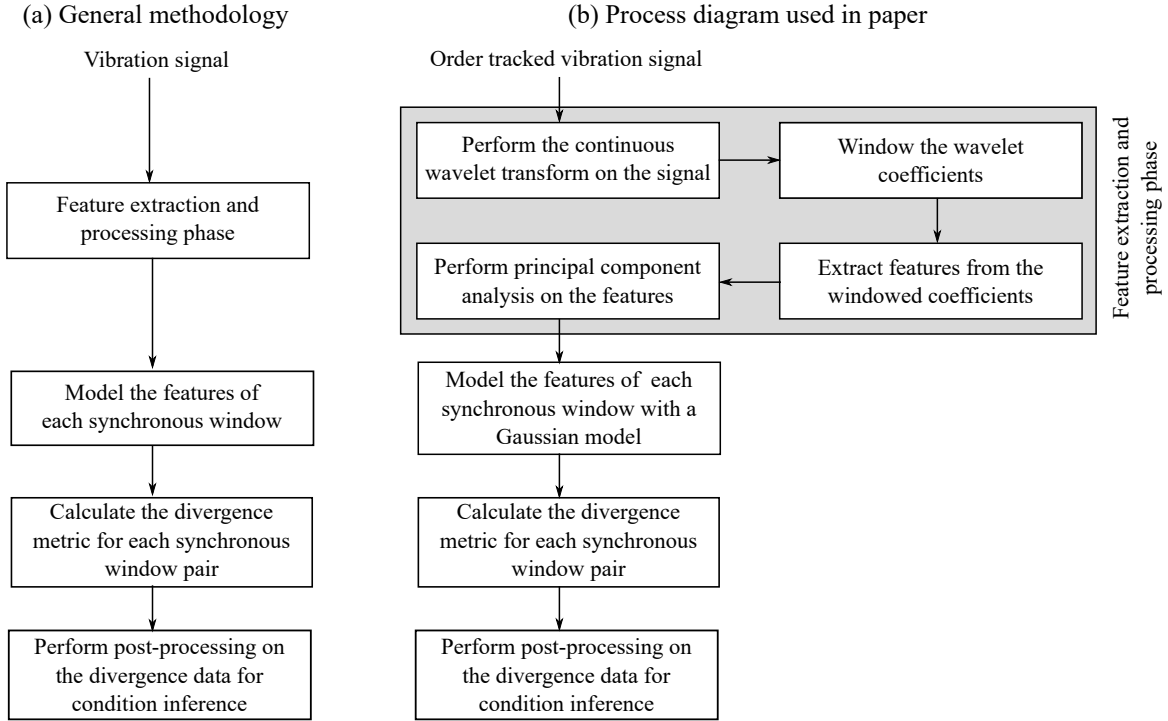
4

(a) General methodology (b) Process diagram used in paper

Vibration signal Order tracked vibration signal

```
┌─────────────────────┐      ┌──────────────────────┐    ┌──────────────────────┐
│ Feature extraction  │      │ Perform the continuous│───▶│ Window the wavelet   │
│ and processing phase│      │ wavelet transform on  │    │ coefficients         │
│                     │      │ the signal            │    │                      │
└─────────────────────┘      └──────────────────────┘    └──────────────────────┘
                             ┌──────────────────────┐    ┌──────────────────────┐
                             │ Perform principal     │◀───│ Extract features from│
                             │ component analysis on │    │ the windowed         │
                             │ the features          │    │ coefficients         │
                             └──────────────────────┘    └──────────────────────┘
```

Feature extraction and processing phase

```
┌─────────────────────┐      ┌──────────────────────┐
│ Model the features  │      │ Model the features of │
│ of each synchronous │      │ each synchronous      │
│ window              │      │ window with a         │
│                     │      │ Gaussian model        │
└─────────────────────┘      └──────────────────────┘

┌─────────────────────┐      ┌──────────────────────┐
│ Calculate the       │      │ Calculate the         │
│ divergence metric   │      │ divergence metric for │
│ for each synchronous│      │ each synchronous      │
│ window pair         │      │ window pair           │
└─────────────────────┘      └──────────────────────┘

┌─────────────────────┐      ┌──────────────────────┐
│ Perform post-       │      │ Perform post-         │
│ processing on the   │      │ processing on the     │
│ divergence data for │      │ divergence data for   │
│ condition inference │      │ condition inference   │
└─────────────────────┘      └──────────────────────┘
```

Figure 1: The process diagrams of the methodology and the specific implementation of the general methodology used in the paper are shown in Figure 1(a) and 1(b).

The processed data are windowed into N_w windows per gear revolution from which features are extracted and processed. The processed features are grouped together with other synchronous windows i.e. windows which correspond to the same angular position on the gear, whereafter the features of each synchronous window set are modelled with a probabilistic model. The windowing scheme's connection with the synchronous windows is illustrated in Figure 2. The windowing scheme and modelling approach result in N_w models, where each model captures the characteristics of the features within the respective synchronous window. The dissimilarity between the probability density functions of the models is used to generate a divergence matrix which is subsequently processed for fault detection.

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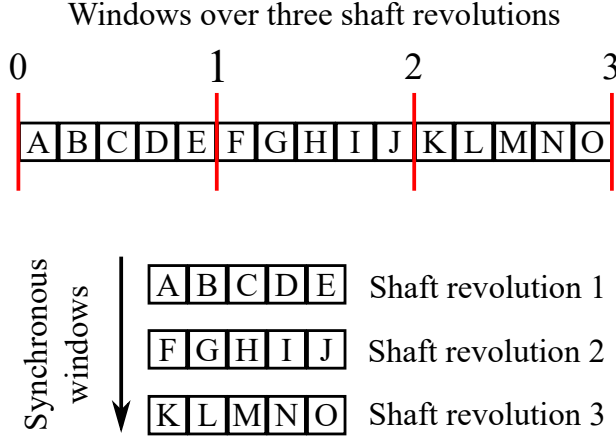
Windows over three shaft revolutions

Figure 2: The synchronous (equal angular) windows are illustrated over three shaft revolutions ($N_r = 3$), with five windows being used per shaft rotation i.e. $N_w = 5$. The $N_r \times N_w$ windows are distinguished by different letters. The different synchronous windows are: A-F-K, B-G-L, C-H-M etc.

2.1. Feature extraction and processing

Time-frequency analysis methods are important for analysing non-stationary signals found in rotating machine condition monitoring applications. The temporal signal is transformed into a two-dimensional time-frequency plane, which makes it possible to detect time-localised changes in the frequency content of the signals e.g. bearing damage excites specific frequency bands at specific cyclic frequencies [7, 9].

The short-time Fourier transform decomposes the signal into a time-frequency plane by assuming the signal is quasi-stationary, which results in the signal having fixed time and frequency resolutions. The fixed time and frequency resolutions make it difficult to analyse signals with characteristics that have different time lengths, which can be impede efficient fault diagnosis [7, 9].

The continuous wavelet transform is an alternative time-frequency analysis tool to the short-time Fourier transform [9]. The continuous wavelet transform [9]

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

decomposes the signal $x(t)$ into a time-scale representation by translating and dilating a wavelet basis function $\psi(t)$ with the translation b and scale a parameters, respectively. The conjugate of the wavelet basis function $\psi(t)$ is denoted by $\psi^*(t)$. The continuous wavelet transform is well-suited for non-stationary signals, is appropriate for extracting features with diagnostic information and is sensitive to the presence of singularities [7];

this makes it a popular analysis tool for rotating machine diagnostics [7, 9, 25, 30, 49, 50]. However, the performance of the continuous wavelet transform is sensitive to the wavelet basis function that is used [50]. The Daubechies db1 wavelet is used in this paper, due to its good performance in the gear and bearing fault detection field [30, 50].

A similar feature extraction process to Schmidt et al. [48] is used. The gear mesh frequency contains diagnostic information and therefore the continuous wavelet transform is applied at the scales situated around the fundamental gear mesh frequency of the gearbox and its four harmonics. Twenty scales at a bandwidth of $3 \times k$ orders around the gear mesh frequency are selected, with $k = 1$ for the gear mesh frequency, $k = 2$ for its harmonic, etc. Each scale is windowed with rectangular windows into N_w segments per rotation, from which the Root-Mean-Square (RMS) of the windowed wavelet coefficients is calculated. The window length is set to $360/N_{teeth}$ degrees with a 0 degree overlap between consecutive windows, which results in a window for each gear tooth in each shaft rotation (i.e. $N_w = N_{teeth}$). It is expected that the vibration signal segments, associated with the portions of the gear that are in the same condition, are similar and therefore the extracted features will be statistically similar.

The RMS features, extracted from the 20 scales at the five gear mesh frequencies, result in a 100 dimensional feature space with a total of $N_r \times N_w$ observations and N_r synchronous observations at each window. The dimensionality of the feature space and the small number of synchronous observations make the model optimisation process susceptible to overfitting. Principal component analysis, a linear dimensionality reduction technique [51], is used to transform the original feature space to a lower dimensional feature space, while retaining most of the information content. Principal component analysis has performed very well compared to other non-linear dimensionality methods for gear fault diagnosis and is therefore well-suited for this application [31]. The accumulative contribution rate [33] is used to select the appropriate dimensionality of the new feature space, so that the information loss in the dimensionality reduction process is minimal.

2.2. Kullback-Leibler divergence

The Kullback-Leibler (KL) divergence, a subset of f-divergence [52], is very popular in the fault detection field [45, 53–56] because of its sensitivity to changes in the density

7

of the data [53]. The KL divergence [51]

$$KL(p_i||p_j) = -\int_{\boldsymbol{x}} p_i(\boldsymbol{x}) \log \left(\frac{p_j(\boldsymbol{x})}{p_i(\boldsymbol{x})} \right) d\boldsymbol{x}, \tag{2}$$

describes the dissimilarity between the probability density function p_i and p_j, associated with synchronous window i and j respectively, over the data space $\boldsymbol{x}$. If p_i and p_j are the same, it means that the data in window i and j are from the same distribution and then $KL(p_i||p_j) = 0$, otherwise $KL(p_i||p_j) > 0$. The KL divergence has for example been used to detect gear damage [53, 54], distillation process anomalies [45, 53] and electrical motor damage [55]. Liu et al. [52] found that the symmetrised Pearson divergence outperforms the forward and the backward Pearson divergence for change point detection and therefore the symmetrised KL divergence

$$D_{ij} = KL(p_i||p_j) + KL(p_j||p_i), \tag{3}$$

over window pair i and j is used instead of the KL divergence given by Equation (2). Equation (3) is symmetric in the probability density functions p_i and p_j, which means that $D_{ij} = D_{ji}$ for any i and j. The symmetrised KL divergence is used to generate a divergence matrix over the gear, denoted by $\boldsymbol{D}$, which is a set of divergence measures over all synchronous window combinations. The divergence data of window i is represented by $\boldsymbol{d}_i \in \mathbb{R}^{N_w \times 1}$, which is used to construct the divergence matrix over all window combinations $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \dots \boldsymbol{d}_{N_w}]$, i.e. $\boldsymbol{D} \in \mathbb{R}^{N_w \times N_w}$.

The KL divergence only has a closed form solution for special probability density functions such as Gaussian distributions. If the datasets are non-Gaussian, the density ratio $p_j(\boldsymbol{x})/p_i(\boldsymbol{x})$ can be modelled using kernel density estimators to calculate the KL divergence [52, 53]. Ferracuti et al. [55] used kernel density estimators to model the probability density functions, whereafter the discrete form of the KL divergence was adopted. Equation (2) can also be evaluated using Monte Carlo integration, however it is very computationally expensive when performed over all window pair combinations i and j for each investigated dataset.

The objective is to find an efficient procedure to calculate the $N_w (N_w + 1)/2$ unique divergence values for each matrix $\boldsymbol{D}$ of a specific measurement. It is assumed that the processed features are described sufficiently well by a multivariate Gaussian distribution.

This results in the following closed form solution for Equation (3) [57]

$$D_{ij} = \frac{1}{2} \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right)^T \left(\boldsymbol{\Sigma}_i^{-1} + \boldsymbol{\Sigma}_j^{-1} \right) \left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j \right) + \frac{1}{2} \text{trace} \left(\boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_i^{-1} \boldsymbol{\Sigma}_j - 2\boldsymbol{I} \right), \qquad (4)$$

where the mean and the covariance matrix of the processed features, associated with synchronous window i, are denoted by $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ respectively and the identity matrix is denoted by $\boldsymbol{I}$. The Jarque-Bera test, available in Reference [58], is used later in this paper to determine whether the Gaussian assumption is correct. If the data are non-Gaussian distributed, more flexible models such as Gaussian mixture models [51] or density ratio estimation approaches [52, 53] can be used. However, due to the limited number of observations N_r in the dataset, it can be difficult to properly estimate and motivate the values of the hyperparameters of the more sophisticated modelling approaches.

In Figure 3, divergence matrices are presented for a healthy gear and a gear with localised damage; the matrix of each dataset is calculated by using Equation (4) over all window combinations. The divergence matrices are calculated from vibration data that was acquired during the gear fatigue experiments discussed in Section 3. The symmetry in
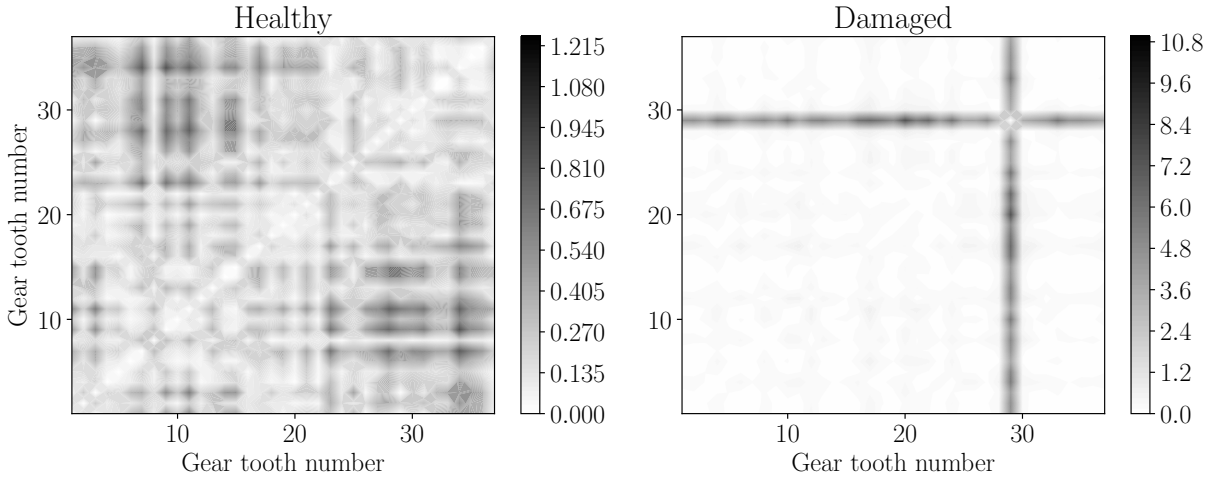
Figure 3: Divergence matrix data $\boldsymbol{D}$ for a healthy gear and a gear with localised damage obtained during experiments discussed in Section 3. Please note that different scales are used for the two plots and that a helical gear with 37 teeth is considered.

the distribution can easily be seen, with the diagonal having a divergence of zero due to the properties of the KL divergence. The divergence data of the healthy gear differ slightly for each window, with small differences indicating that the teeth are approximately in

the same condition. In the presence of localised damage, the data associated with the damaged tooth differ significantly from the data of the other healthy teeth, which results in dissimilar models that are easily identified in the image of the divergence matrix. The extent of the localised gear damage can be inferred from the techniques proposed in the next section.

2.3. Data analysis for automatic condition inference

The data in Figure 3 are shown in Figure 4(i) over a full gear rotation for a healthy and a damaged gear. The healthy gear divergence data are relatively close to one another when
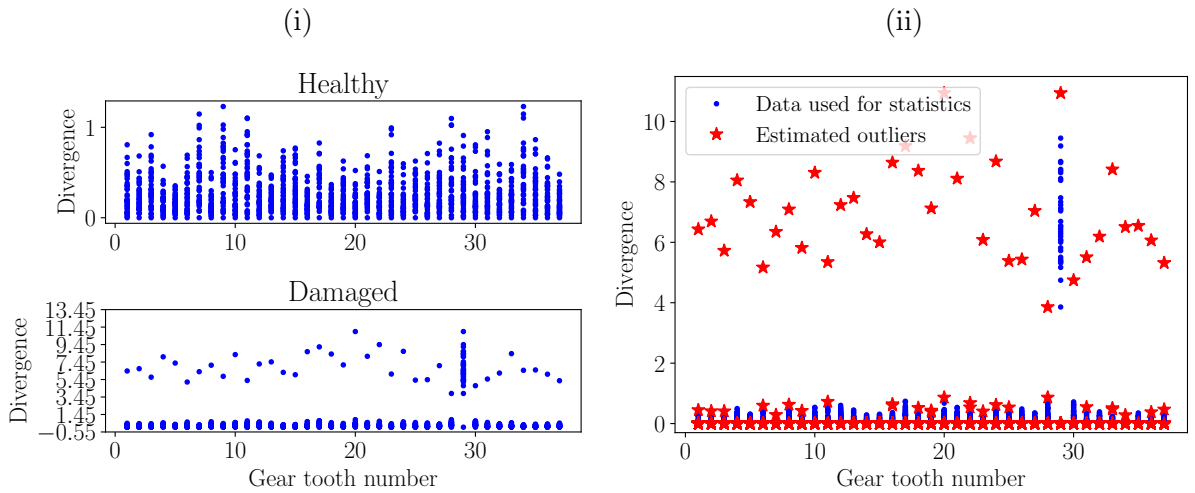
Figure 4: The synchronous divergence data are shown over a gear rotation for the healthy and the damaged gear in 4(i). The outliers are identified and shown in 4(ii) for the damaged gear using the proposed outlier removal process. The gear under consideration has 37 teeth.

considering the same tooth, with some slight variations being present between different teeth. The divergence data of the healthy portion of the damaged gear contains very similar characteristics to those of a healthy gear, with the exception that a large outlier is present. This outlier is due to the presence of localised damage on the gear, which manifests over all gear teeth as a result of the properties of D seen in Figure 3.

The outliers caused by localised damage and the zero divergence when $i = j$ in Equation (4) have to be removed from the dataset, given that they do not reflect the true condition of the portion of the gear associated with the window. The outliers in the divergence

10

data, due to the presence of localised damage, are automatically estimated in a three step process; this is performed separately on the data of each window and is as follows for the data $\boldsymbol{d}_i$ of window i:

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

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

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

The parameters υ and k_r, used in the outlier removal process, are equal to 5.0 and 3.0 respectively in this paper. Note that after the outlier removal process, each window may not have the same number of data points left and cannot strictly be written as a matrix. However, the processed divergence data is denoted by $\boldsymbol{D}^{proc}$ for the sake of notational simplicity where the superscript $proc$ highlights that the processed data are used.

The outliers that are present in the damaged gear divergence data, shown in Figure 4(i), were identified with the proposed method and shown in Figure 4(ii). The benefits of the outlier removal process are shown in Figure 5 on the point estimates of the data. It is observed that the point estimates calculated from the processed data are significantly better than the raw data, because it has a smaller variance and a mean without a bias. Hence, the processed data can lead to more robust results.

An automatic method is required to infer the condition of the gear i.e. to detect localised damage and evaluate changes in the extent of the damage. Bayesian analysis techniques [59, 60] are employed in this paper to perform this task. Bayesian analysis techniques use probability densities to quantify the uncertainty in the random variables, which is subsequently used to make inferences on the random variables. It is assumed that the mean μ and the precision λ or the reciprocal of the variance of the data, of a specific window that is under consideration, are unknown and need to be inferred. The data are denoted by X in this section, because two distinct datasets are investigated with
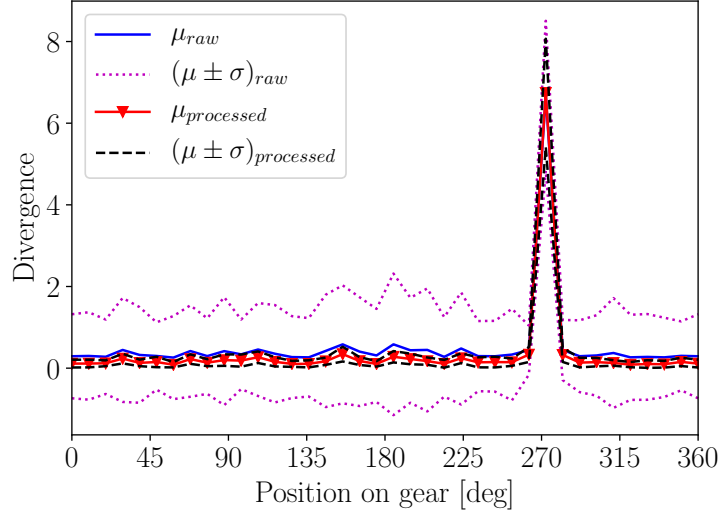
Figure 5: The point estimates of the processed data (using the outlier removal process) and the raw (unprocessed) data are compared for a gear with localised damage.

the approach outlined here. The prior distribution, incorporates prior knowledge of the parameters into the inference process and is chosen to be a Normal-Gamma distribution in the form of

$$p(\mu, \lambda | \mu_0, \kappa_0, \alpha_0, \beta_0) = \text{Gaussian}\left(\mu | \mu_0, (\kappa_0 \lambda)^{-1}\right) \text{Gamma}(\lambda | \alpha_0, \beta_0), \tag{5}$$

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

$$p(\mu, \lambda | X, \mu_0, \kappa_0, \alpha_0, \beta_0) = \text{Gaussian}\left(\mu | \mu_n, (\kappa_n \lambda)^{-1}\right) \text{Gamma}(\lambda | \alpha_n, \beta_n), \tag{6}$$

where [61]

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{x} \tag{7}$$

$$\kappa_n = \kappa_0 + n \tag{8}$$

$$\alpha_n = \alpha_0 + \frac{n}{2} \tag{9}$$

$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{1}{2}\frac{(\bar{x} - \mu_0)^2}{\frac{1}{n} + \frac{1}{\kappa_0}}, \tag{10}$$

12

are the parameters of the posterior distribution with n being the number of observations, and $\bar{x}$ being the sample mean of the data X of the considered window. The precision λ is a nuisance parameter and is marginalised out to obtain the posterior marginal distribution of the mean [51, 61]

$$p(\mu|X) = \int_\lambda p(\mu, \lambda|X)d\lambda \tag{11}$$

$$= \text{Student}_{2\alpha_n}\left(\mu\,\bigg|\,\mu_n, \frac{\beta_n}{\alpha_n \kappa_n}\right), \tag{12}$$

which is in the form of the Student-t distribution, with $2\alpha_n$ degrees of freedom, a variance of $\beta_n/(\alpha_n \kappa_n)$ and a mean of μ_n. The dependence on the fixed hyperparameters are neglected for future notational simplicity. The mode and mean of the posterior marginal distribution of the mean, presented in Equation (11), are the same. The marginal distribution in Equation (11) is used to infer the condition of the gear, by calculating the probability that the posterior mean exceeds an alarm threshold i.e.

$$P(\mu > \text{threshold}|X) = \int_{\text{threshold}}^\infty \int_\lambda p(\mu, \lambda|X)d\lambda d\mu, \tag{13}$$

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

3. Validation

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

3.1. Experimental setup

An experimental setup, designed by Stander and Heyns [8], was refurbished to conduct gear fatigue tests under fluctuating operating conditions. The vibration data were acquired from the experimental setup in Figure 6, which consists of an alternator, three helical

13

gearboxes and an electrical motor. The instantaneous load applied by the alternator and the instantaneous rotational speed applied by the electrical motor, were controlled with a personal computer. The axial acceleration from a 100mV/g tri-axial accelerometer,
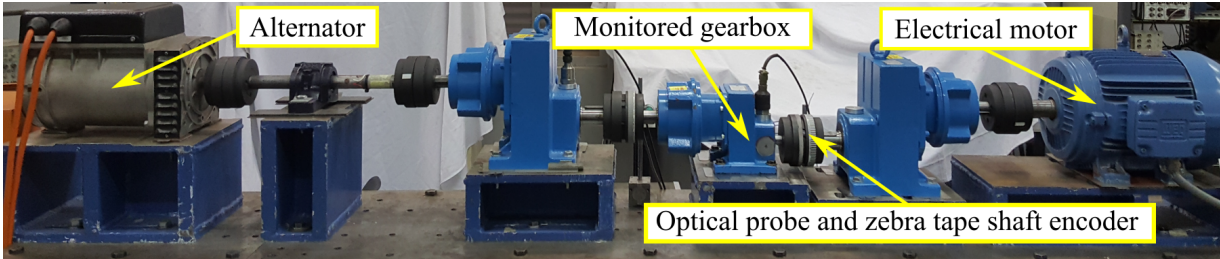

Figure 6: The experimental setup.

located on the bearing housing of the monitored helical gearbox, is used in the condition monitoring process. The acceleration vibration signal was sampled at 25.6kHz and the optical probe signal was sampled at 51.2kHz with an OROS OR35 data acquisition device, where the optical probe is used with an 88 pulse per revolution zebra tape shaft encoder to calculate the instantaneous angular speed of the input shaft of the monitored gearbox. Data were acquired from a healthy gearbox, whereafter the gearbox was disassembled so that localised damage could be seeded on the gear and thereafter the gearbox was reassembled. The gearbox, with the damaged gear shown in Figure 7(i), was subsequently operated for approximately twenty days under the operating conditions depicted in Figure 8, until the damaged tooth finally failed. The gear with the broken tooth is shown in Figure 7(ii).

3.2. Investigation 1

In the first investigation, the effectiveness of the methodology is investigated and compared to conventional fault diagnosis techniques on gearbox data obtained from three cases of different constant operating conditions. This investigation is performed to validate the robustness of the proposed methodology to detect localised anomalies in the data under different operating conditions.

The load and rotational speed properties of the three cases are presented in Table 1 and the operating conditions are present for measurements taken for a gearbox with a healthy gear and with a gear without a tooth (see Figure 7(ii)), respectively.

14

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

Figure 7: The gear with seeded damage before and after the fatigue experiment was completed.

Table 1: Operating conditions at the input shaft of the monitored gearbox for investigation 1.

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

3.2.1. Conventional fault diagnosis techniques

The Synchronous Average (SA) of the vibration signal have been successfully used for gear fault diagnosis [8] and is investigated on the signals described in Table 1. The synchronous averages of the investigated computed order tracked vibration signals are presented in Figure 9(i). The amplitudes of the synchronous averaged signals in Figure 9(i) change significantly with changes in operating conditions for gears in the same condition, with evidence of the broken tooth not being seen. Hence, it is difficult to distinguish between changes in operating conditions and changes in machine condition utilising the synchronous average for fault diagnosis under different operating conditions. This also makes it difficult to set a threshold for automatic fault diagnosis; the gear damage is not evident and the data are significantly influenced by the varying operating conditions.
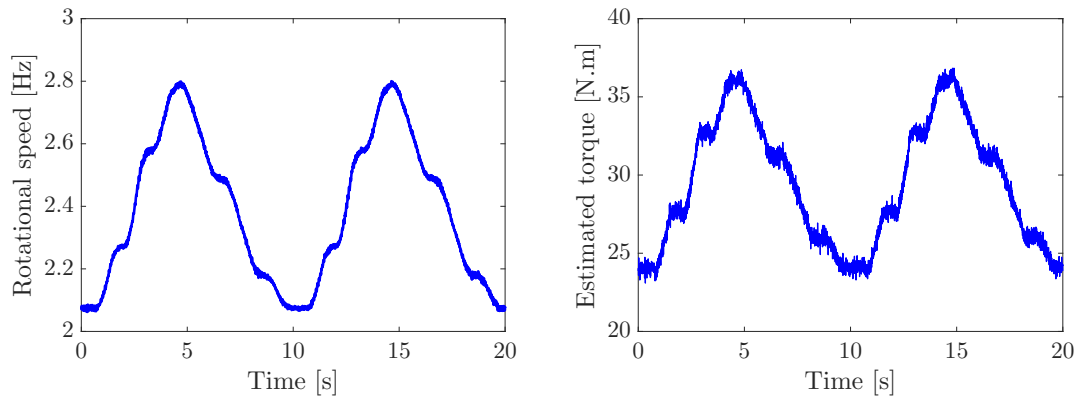
Figure 8: The operating conditions during the fatigue experiment at the input shaft of the monitored gearbox.

The residual signal i.e. the departure of the vibration signal from the average deterministic vibration components can be used for more effective fault diagnosis [16, 18]; for example, it has performed significantly better than the conventional synchronous average for gear crack detection [16]. The Synchronous Variance (SV) i.e. the synchronous average of the squared residual signal is a second-order cyclostationary technique that has been used for gear and bearing fault diagnosis [20] and is displayed in Figure 9(ii). The SV is estimated with the procedure in Ref. [20]. The gear damage is seen for operating condition Case 2 and Case 3, with the gear damage at Case 1 not being seen because no load is applied to the system. However, the results are also dependent on the operating conditions i.e. the variance of the synchronous variance is operating condition dependent and the damage is not very prominent for Case 2 and Case 3; the latter result is attributed to the fact that the helical gears have high contact ratios which mitigates the influences of the missing tooth on the gear mesh stiffness.

The Power Spectral Density (PSD) of the computed order tracked vibration signals is presented in Figure 10 similarly to the results in Figure 9. The variation of the gear mesh components with respect to operating condition changes is clearly observed, however, the differences between the gear conditions are not prominent. The results corroborate the observation that it is not easy to detect the damage using conventional techniques with the different operating condition levels adversely influencing the fault diagnosis task. It is also difficult to assign a threshold to automatically detect the presence of localised damage
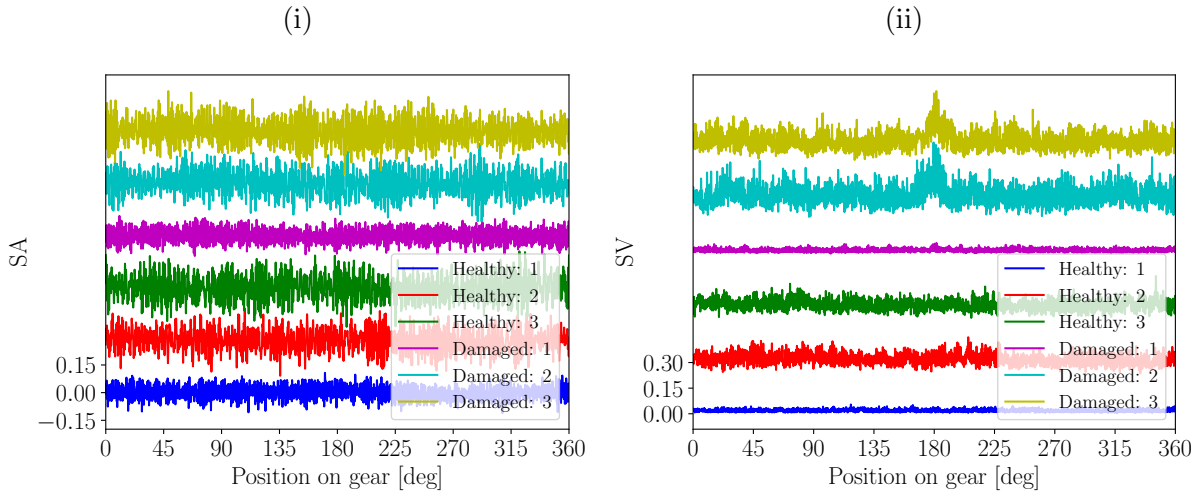
16

Figure 9: The Synchronous Average (SA) of the computed order tracked vibration signals of the gears and the Synchronous Variance (SV) of the gears in different conditions are compared for the operating conditions in Table 1. Each signal is given a unique offset to compare the results, with the damage being present at 180 degrees for the damaged gear cases. The legend indicates <Condition>: <Operating condition number>.

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

3.2.2. Proposed methodology

Hence, a more sophisticated approach needs to be used to automatically infer the condition of the machine and therefore the proposed methodology is investigated on the same dataset. For each measurement considered in this investigation, the following procedure is followed: The vibration signal is order tracked using the optical probe and zebra tape shaft encoder signal whereafter the features are extracted and processed as described in Section 2.1. The new dimensionality of the principal component feature space is selected as two, because the calculated accumulative contribution rate is approximately 0.95. A multivariate Gaussian model is to the principal component features of each set of N_w synchronous features, whereafter the divergence is calculated using Equation (4) for all synchronous window combinations. The outliers in $\boldsymbol{D}$ are removed with the procedure described in Section 2.3 to obtain the processed divergence matrix $\boldsymbol{D}^{proc}$, which is used in subsequent analyses.

17
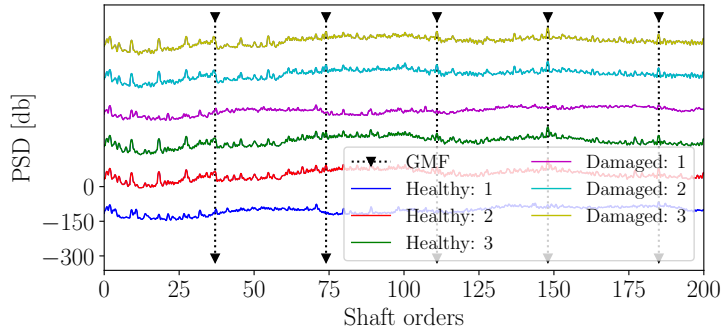
Figure 10: The Power Spectral Density (PSD) of the order tracked vibration signal for the different datasets described in Table 1.

Bayesian data analysis is used in the condition inference process as described in Section 2.3. The posterior distribution is a compromise between the prior information and the evidence obtained from the investigated dataset [51, 59]. The prior distribution conveys the prior beliefs in the data and are governed by a set of hyperparameters. The appropriate hyperparameters are unknown for the data under investigation and are set to $\mu_0 = 0.0$, $\kappa_0 = 0.01$, $\alpha = 10$, $\beta = 0.01$ so that the inferred posterior marginal distribution of the mean is dominated by the data. The consequences of the selected hyperparameter values are critically investigated in Appendix A.

It is necessary to specify an alarm threshold to infer the condition using Equation (13). If historical data of a healthy machine are available, it is possible to set an alarm threshold to perform automatic novelty detection and if historical fault data are available it is even possible to set alarm threshold for the different stages of degradation. However, this is not possible in the absence of historical data and a different strategy is required to perform this task.

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

$$\text{threshold} = \text{median}\left(\boldsymbol{D}^{proc}\right) + k_{thres} \cdot \text{median}\left(\boldsymbol{D}^{proc} - \text{median}\left(\boldsymbol{D}^{proc}\right)\right), \qquad (14)$$

18

where median $(\boldsymbol{X}) \in \mathbb{R}$ is the median of the data $\boldsymbol{X}$ and k_{thres} is a threshold factor and selected as five. Hence, it is possible to set a threshold with Equation (14) using only the current data under consideration i.e. historical data are not required.

The Posterior Mode of the Marginal Distribution of the Mean (PMMDM), for the healthy and the damaged gear divergence data of the three load cases, is superimposed in Figure 11 with the alarm threshold calculated with Equation (14). The alarm threshold for each dataset was approximately the same and therefore only their average is presented to keep the figures uncluttered. The results in Figure 11(i) indicate that in the absence

(i) Healthy gear   (ii) Gear with broken tooth
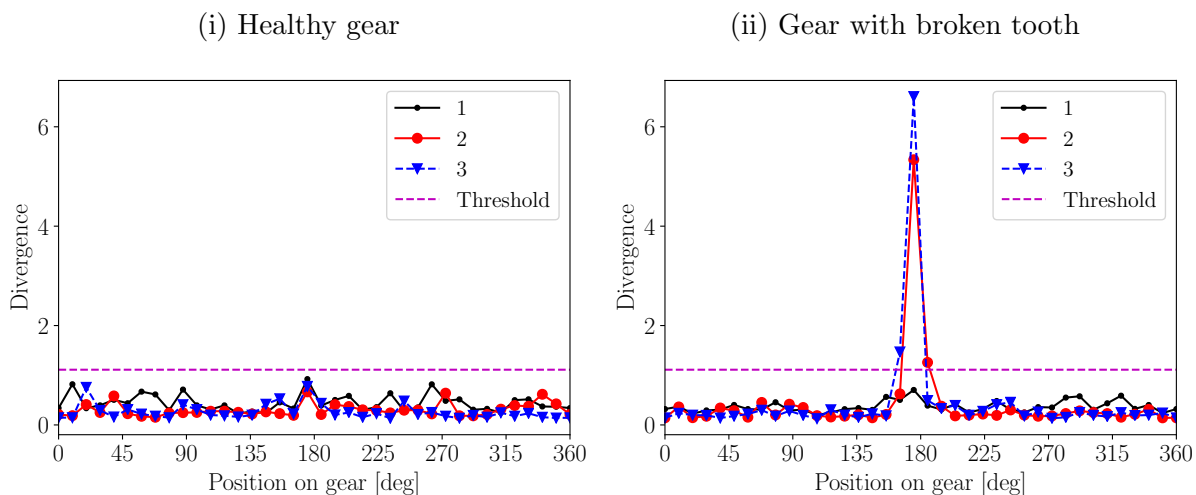
Figure 11: The Posterior Mode of the Marginal Distribution of the Mean (PMMDM) is shown for the healthy gear and the gear with a broken tooth for the operating condition cases described in Table 1. Only the average threshold of the different measurements is shown to ensure that the figure is not cluttered.

of localised gear damage, the divergence does not exceed the threshold. It is clearly possible from the results in Figure 11(ii) to detect localised gear damage using the proposed technique when loads are applied to the system; in the absence of loads acting on the system, the damage is not detected. The magnitude of the divergence associated with the broken tooth varies slightly with load and speed due to the fact that the physical impacts resulting from the missing tooth varies with operating conditions. The aforementioned results are reasonable, given the fact that helical gearboxes are investigated. Helical gears, that have large contact ratios, will always have one tooth in contact, even if a tooth is missing. Hence, in the absence of loads, the impulses due to a missing tooth will be very

19

small and may be undetected in many circumstances.

It is necessary to consistently compare the divergence data to the alarm threshold to perform automatic condition inference; this is performed by calculating the probability that the posterior mean is larger than the threshold using Equation (13) for each tooth. The results for the data in Figure 11 are shown in Figure 12. The results corroborate the
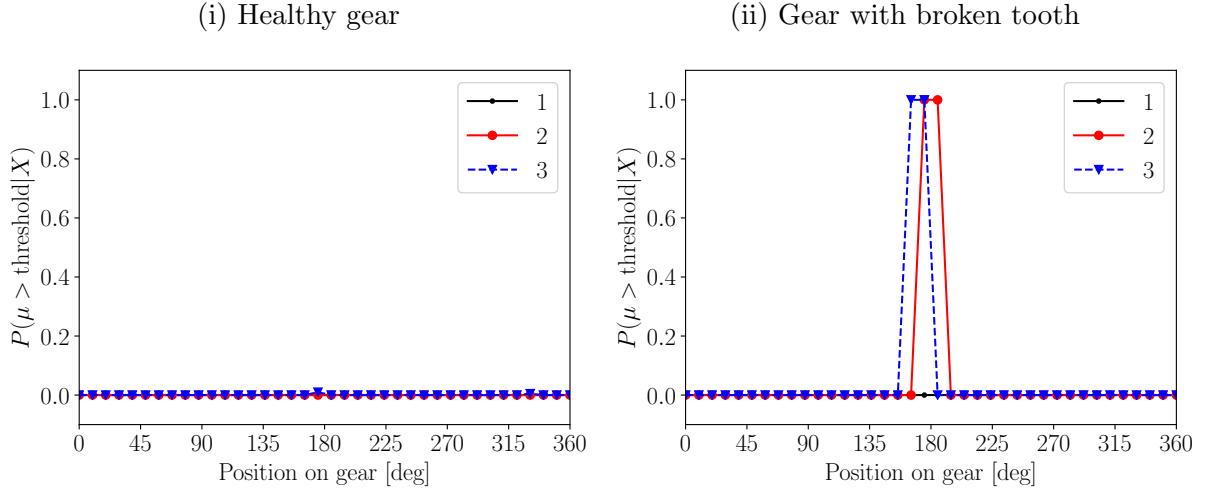

Figure 12: The probability that the posterior mean exceeds the threshold is calculated with Equation (13) for the data in Figure 11. The threshold, calculated for a specific measurement with Equation (14), is used with Equation (13) to infer the condition of the specific measurement i.e. the average threshold shown in Figure 11 is not used in the calculation procedure.

condition inferred from the results in Figure 11. The benefit of using the representation in Figure 12 is that it is easy to understand.

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

It is evident from the results in Figures 11 and 12 that the proposed methodology per-

forms significantly better when compared to the results of the conventional fault diagnosis techniques in Figures 9 and 10; the damage is significantly more prominent, it is possible to automatically infer the condition of the gears and the healthy portions of the gear are unaffected by the different operating conditions.

3.3. Investigation 2

In the second investigation, the performance of the proposed method is investigated and compared to conventional fault diagnosis methods for the gearbox fatigue dataset described in Section 3.1. Conventional fault diagnosis methods are investigated in Section 3.3.1, the methodology's fault detection and localisation capabilities are considered in Section 3.3.2, whereafter the damage trending potential of the methodology is illustrated in Section 3.3.3. The data considered in this section were acquired under the operating conditions presented in Figure 8 and the same properties as in Section 3.2, such as the dimensionality of the feature space and the values of the hyperparameters are used.

3.3.1. Conventional fault diagnosis techniques

The synchronous averaged order tracked vibration signals and the synchronous variance of the residual signals are presented in Figure 13(i) and Figure 13(ii), respectively. The residual signal, used to calculate the SV, is calculated by subtracting the generalised synchronous average from the synchronous average; the generalised synchronous average is a better estimation of the periodic part under varying operating conditions than the conventional synchronous average [15]. The damage is seen in the synchronous average as well as the synchronous variance for the damaged vibration signals. Due to the fact that the operating conditions are the same for each dataset in this section, the influence of varying operating conditions on the data is not evident.

The PSDs of the order tracked vibration signals are presented in Figure 14 for the measurements investigated in Figure 13 as well. It is difficult to detect the presence of localised gear damage in the PSD; it is even difficult when comparing the damaged gearbox data to the healthy gearbox data. It is only possible to observe changes in the harmonics of the gear mesh frequency for the damaged gearbox as the gearbox deteriorates.

It is also possible to compare the trending capabilities of conventional metrics and the metrics obtained from the proposed methodology i.e. the sensitivity of metrics to
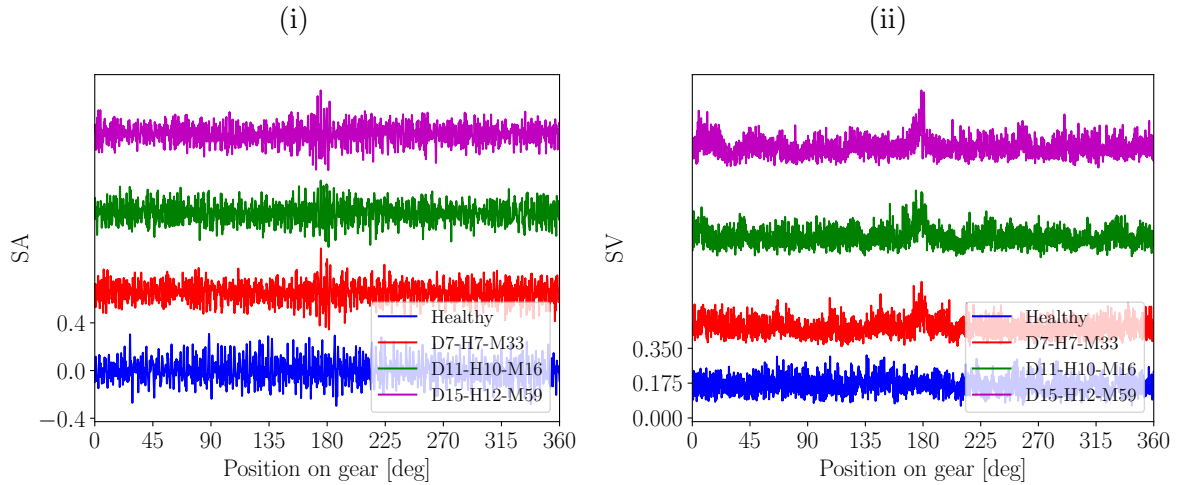
Figure 13: The Synchronous Average (SA) and the Synchronous Variance (SV) of the residual signal are compared for different datasets. The legend indicates the condition of the gearbox as well as the time from the start of the damaged experiment in Day-Hour-Minute format. Each signal is given a unique offset to make the comparison between the results easier, with the localised damage being present at 180 degrees.

changes in condition. The kurtosis, which is an indication of the impulsiveness of the vibration signal, and the RMS, which is an indication of the average energy in the signal, can be used for fault trending and are compared in Figure 15 on the fatigue data i.e. as the gear deteriorated from Figure 7(i) to Figure 7(ii). The healthy gear data are also inserted at the start of the measurement number, with the healthy gearbox and the gearbox with the seeded damage separated with the Change In Condition (CIC) vertical line. A prominent change in the RMS is observed at the CIC line, which is attributed to the disassembling-reassembling procedure to damage the gear. Thereafter, the RMS decreases with measurement number. The kurtosis increased slightly at the CIC line, whereafter it remains constant between measurement number 100 and 450. A change in the kurtosis and the RMS is observed at the 480th measurement when the gear failed, however, the deterioration of the gear is not prominent from the data.

Hence, the conventional analysis techniques can be used for fault detection, fault localisation and fault trending by manually investigating the data. However, the results are not ideal e.g. the damage is not very prominent in the synchronous averages and the
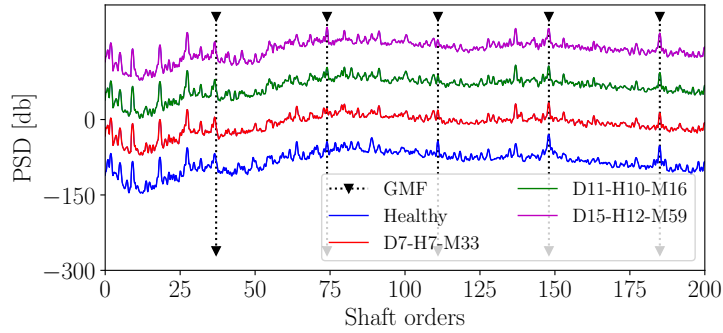
22

Figure 14: The Power Spectral Densities (PSDs) of the order tracked vibration signals, investigated also in Figure 13, are shown. The Gear Mesh Frequency (GMF) and its harmonics are also incorporated into the plot. Each signal is given a unique offset to make the comparison between the results easier.
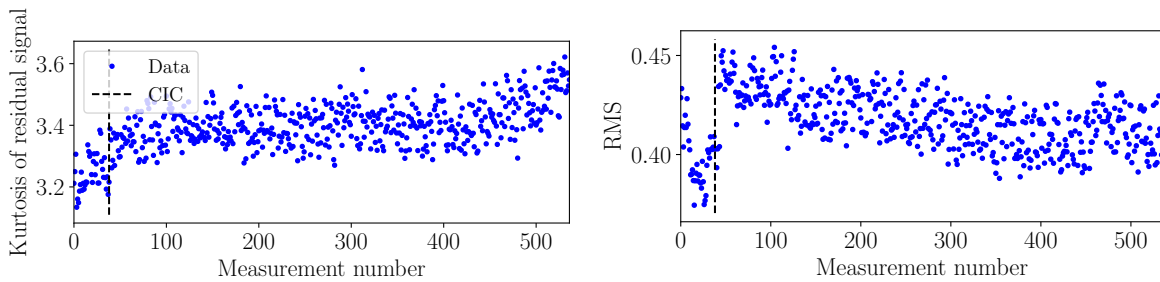

Figure 15: The kurtosis of the residual signal and the Root-Mean-Square (RMS) of the vibration signal are presented over measurement number.

deterioration of the gear is not clearly seen in the trended results. Therefore, more robust metrics are required for automatic condition inference.

3.3.2. Automatic fault detection and localisation without historical data

The proposed methodology is implemented with exactly the same procedure as Section 3.2. The PMMDM is shown in Figure 16(i) for different time stamps, in day-hour-minute format, from the start of the fatigue experiment with the damaged gear. The PMMDM, calculated from Equation (11) for the processed divergence of the healthy data, does not exceed the threshold; while the PMMDM of the damaged gear data exceed the threshold where the localised damage is situated (approximately 180 degrees).

The probability that the posterior mean exceeds the threshold, is presented in Figure 16(ii) and the results support the conclusions drawn from the PMMDM in Figure 16(i). Hence, this approach allows localised damage to be automatically detected without using historical data, with changes in damage severity observed as well.
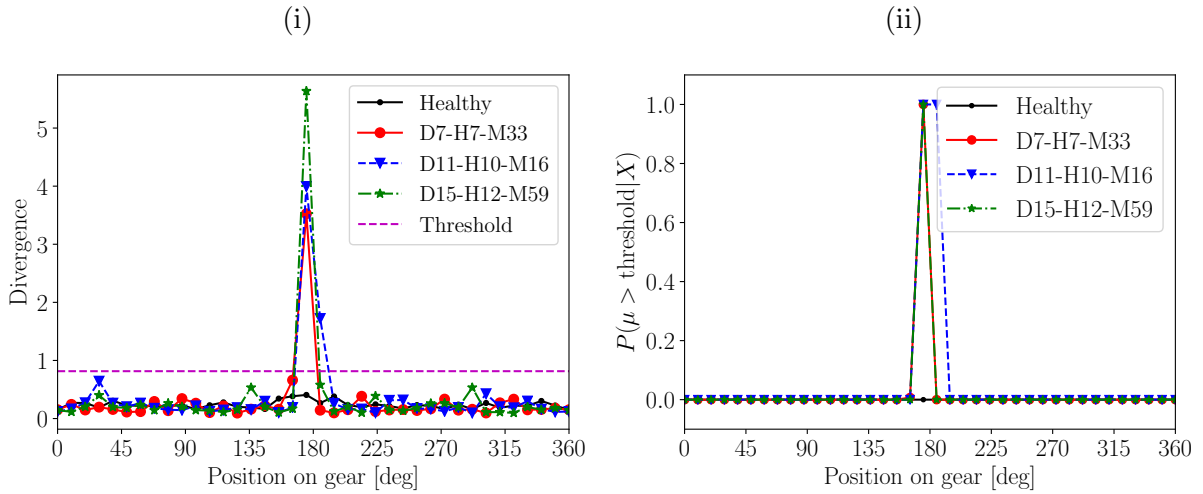
Figure 16: The Posterior Mode of the Marginal Distribution of the Mean (PMMDM) of the divergence for each window and the averaged threshold for the measurements are shown in Figure 16(i). The probability that the posterior mean of a specific measurement exceeds the threshold of the specific measurement is shown in Figure 16(ii). The legend indicates the day-hour-minute after the fatigue experiment was started with the damaged gear.

3.3.3. Sensitivity of metrics to machine condition changes

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

- The sensitivity fo the PMMDM to changes in machine conditions is investigated to determine whether it contains information on the severity of the damage.

- The performance of the automatic condition inference for different fault severities is investigated. This is performed by presenting the probability that the PMMDM exceeds the alarm threshold, calculated with Equation (14) by using the procedure described in Section 3.2, over measurement number.

24

- The sensitivity of the statistics of the divergence data to changes in condition is investigated for all of the measurements. The PMMDM of the statistics of the divergence data are also calculated by utilising historical reference data to highlight the sensitivity of the statistics to changes in condition.

- Lastly, fault trending by utilising historical reference data with Bayesian data analysis techniques is illustrated. This is to further support the benefits of using Bayesian data analysis techniques and to illustrate that this method can be used when historical data are available.

The PMMDM of the divergence is shown in Figure 17(i) for 38 healthy measurements and 498 damaged measurements, with the probability that the posterior mean exceeds the threshold shown in Figure 17(ii). The Change In Condition (CIC) line indicates the start of the damage measurements. The healthy and the damaged measurements are presented
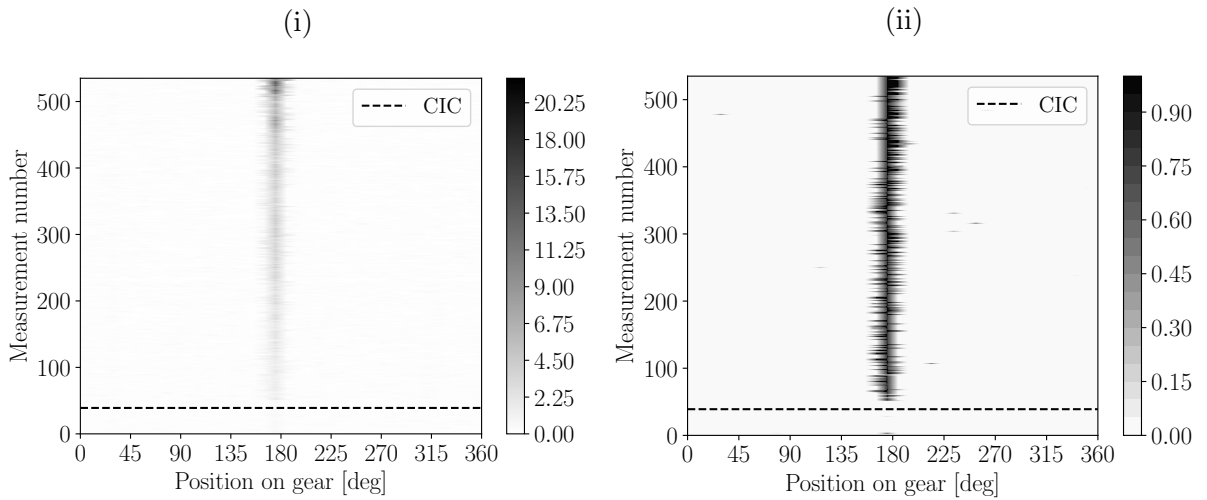

Figure 17: The surface plot of the PMMDM of the divergence data is shown in Figure 17(i) over measurement number and the probability that the posterior mean exceeds the threshold is shown in 17(ii). The threshold used to calculate the results in Figure 17(ii) is calculated separately for each measurement. The healthy and damaged data are separated by the Change In Condition (CIC) line.

together in Figure 17 to highlight the ability of the proposed approach to detect changes in machine condition. The change in the localised gear damage is evident from the results in Figure 17(i), especially in the last stages of the experiment.

In some instances in Figure 17(ii), it may seem that the angular arc of the damage is larger than in other cases. This occurs because the windows for feature extraction start arbitrarily between different datasets, but remain consistent for each dataset (if the order tracking is performed correctly). The consequence of this is that in some instances the damaged tooth is situated between two windows which result in the divergence data of two windows exceeding the threshold. This can potentially be alleviated by overlapping consecutive windows.

Four statistics, calculated from the divergence metric are presented over measurement number and compared to the metrics investigated in Figure 18. It is not only possible to
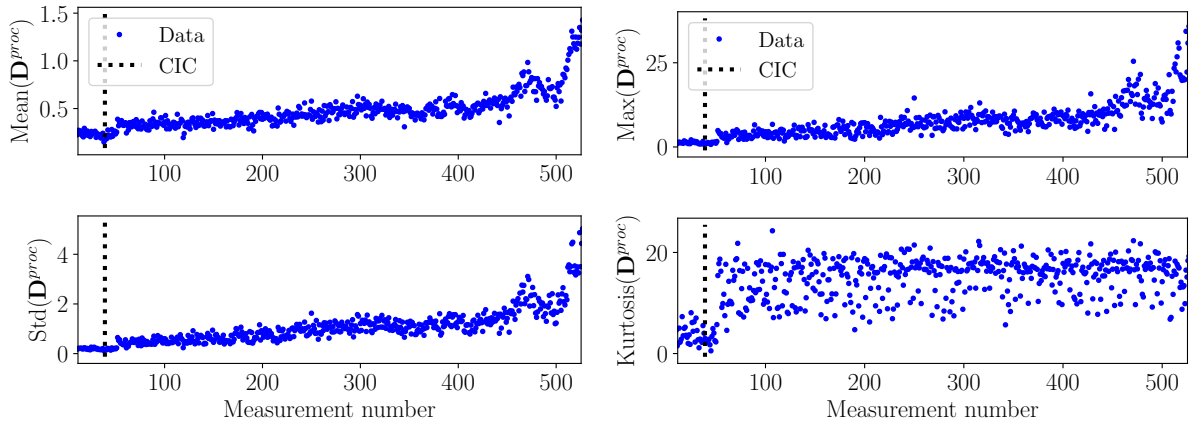
Figure 18: The mean, standard deviation (denoted by std), maximum (denoted by max) and the kurtosis of $\boldsymbol{D}^{proc}$ is shown over measurement number. The change in the gear condition is indicated by the Change In Condition (CIC) line.

observe the change in condition, but also the change from the seeded damaged gear to the gear with the broken tooth for the mean, standard deviation and maximum statistics in Figure 18. This is in contrast to the results in Figure 15, where only the change in condition is observed at the CIC line. Hence, the metrics are significantly more sensitive to changes in machine condition and therefore the proposed methodology is better suited for condition inference.

In this section, historical reference data are utilised for automatic fault trending to prove that the metrics do change with respect to measurement number as the condition of the gearbox changes. However, it is also investigated to emphasise that the metrics obtained from this methodology can be used for fault trending if historical data do become

available.

It is assumed that twenty healthy measurements are available to calculate an alarm threshold, whereafter the trending procedure for all subsequent data is as follows:

- Calculate a statistic, such as the mean, of the whole processed divergence matrix $\boldsymbol{D}^{proc}$ for each measurement as done in Figure 18.

- Window the one-dimensional statistic into windows with length N_L and an overlap of N_O between consecutive windows.

- Model the windowed data with a Gaussian likelihood function and impose a Normal-Gamma prior distribution on the parameters. Use Equation (7) to (10), with the same hyperparameters as the previous sections, to obtain the posterior distribution parameters.

- Determine the probability that the mean of the windowed data exceeds the threshold with Equation (13). The threshold is calculated with Equation (14) for the first 20 measurements.

The four statistics of $\boldsymbol{D}^{proc}$ used in Figure 18 are independently investigated and compared to an independently calculated threshold, with the results shown in Figure 19. The windows have a length of $N_L = 20$ and an overlap of $N_O = 18$ measurements in all investigations. The results in Figure 19 indicate that it is essential to analyse the data with probabilistic techniques due to the variation thereof. It can be observed that the PMMDM of all statistics changes at the change in condition line (i.e. where the healthy and the damaged measurements meet). Note that because the threshold is set far from the healthy data and windows are used, the threshold is not immediately exceeded for the first damaged measurements. A change in the metrics can however be observed.

The mean, standard deviation and the maximum value of $\boldsymbol{D}^{proc}$ contain less noise for the healthy measurements and therefore an alarm threshold is obtained which is close to the healthy data. The kurtosis, is very sensitive to the noise which is present in $\boldsymbol{D}^{proc}$. The implications of this is that the mean, standard deviation and the maximum of $\boldsymbol{D}^{proc}$ is more sensitive to changes in machine condition than the kurtosis, with the standard deviation and maximum values being the most sensitive.
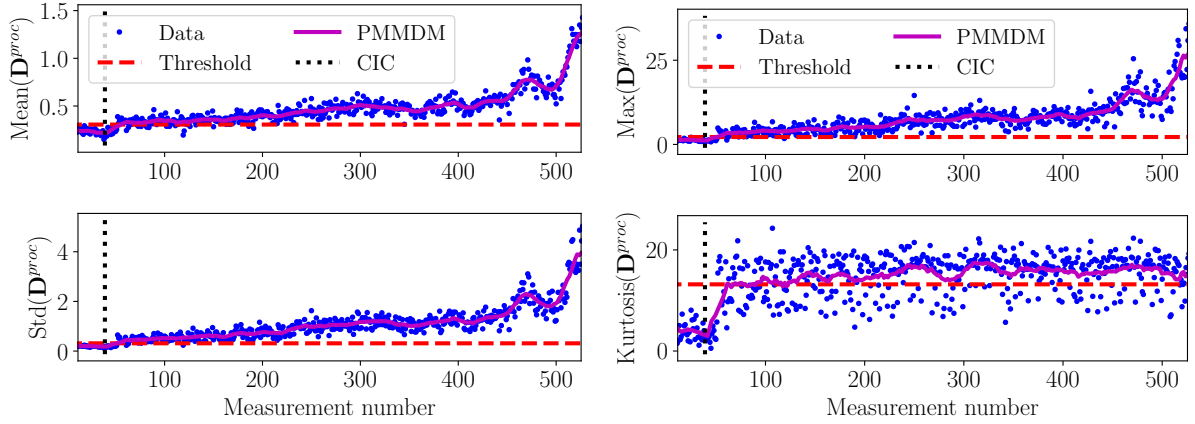
Figure 19: The PMMDM and the threshold of the data in Figure 18 are presented with the data in Figure 18.
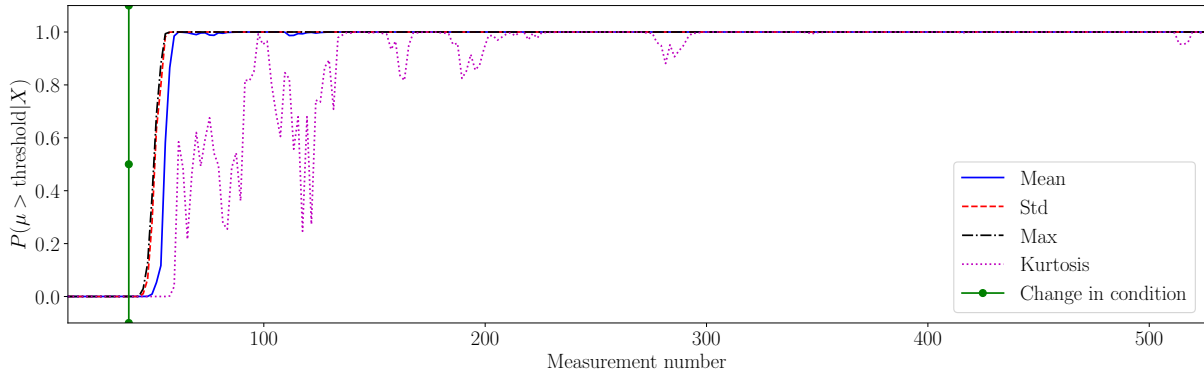

Figure 20: The probability that the mean of each statistic exceeds the threshold is shown in Figure 19. The change in condition line is shown as well.

If noise, for example from environmental effects or operating condition changes, is present in the trended statistics, it is uncertain whether the mean of the dataset exceeds the threshold. In Frequentist statistical approaches, the parameters are fixed [51] and therefore the mean will either be larger or smaller than the threshold. The Bayesian analysis approach allows the uncertainty in the parameters due to the noise to be quantified, which helps with the maintenance decision process. Another observation that is made, is that using the posterior mean with Equation (13) is better suited for the condition inference task than using the individual values in Figure 19. This is because the posterior mean is less susceptible to noise and other outliers and it will therefore cause fewer false alarms if for example a new operating condition is observed for a specific measurement.

Hence, from this investigation it can be concluded that the PMMDM of the syn-

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

3.4. Diagnosis of the pinion

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

The same process is performed as in Section 3.3.2 for the gear, with the exception that the number of windows (N_w) per revolution is 20 instead of 37, because the number of teeth on the gear and the pinion is different. The PMMDM and the probability that the posterior mean of a tooth exceeds the newly calculated threshold, is shown in Figure 21. Note that the exact same datasets are used as in Figure 16, with the same procedure performed as well. According to the results, the pinion is in a healthy condition and this
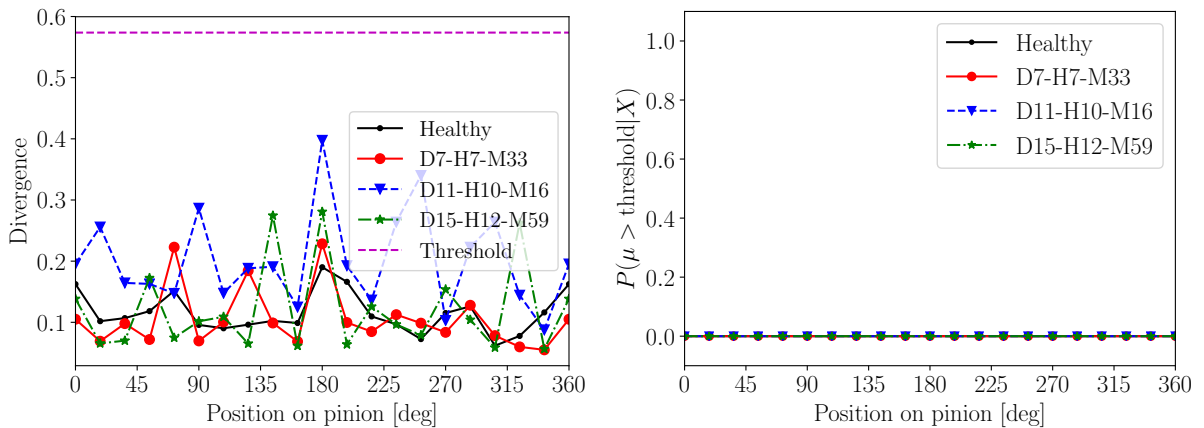

Figure 21: The PMMDM and the probability that the mean exceeds the threshold for the pinion.

proves that the condition of a healthy gear in a damaged gearbox can correctly be inferred by using the proposed technique.

The influence of the damaged gear on the statistics of the pinion's divergence data needs

to be assessed, therefore, the statistics of the pinion are calculated for each measurement in the dataset. The same process is used as in Section 3.3.3 with the results shown in Figure 22 for the pinion. The results indicate that the statistics of the pinion divergence
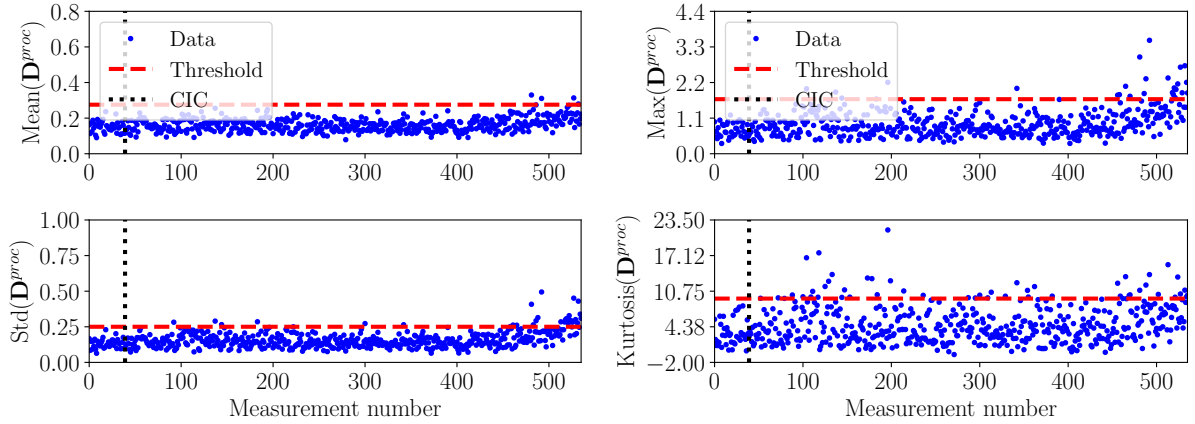

Figure 22: Statistics of $\boldsymbol{D}^{proc}$ over measurement number for the divergence data, obtained by considering the pinion of the gearbox instead of the gear. The same measurements are used as in Figure 19.

data contain much noise, however it remains fairly constant until the 480th measurement whereafter a slight increase is observed.

The probability that the mean of the windowed statistics exceed the threshold, calculated with the procedure in Section 3.3.3, is shown in Figure 23. The results indicate that the data reflect the true condition of the pinion for most of the measurements. Only in the last stages of the experiment, the damaged portion of the gear dominates the vibration data, which is ultimately reflected in the pinion statistics as well. Hence, the increase in the pinion's statistics and the results from Section 3.3.3 indicate that failure of the gear is imminent at the 480th measurement; albeit actually because of damage on the gear, rather than on the pinion. Figure 23 however illustrates that the proposed approach is able to largely separate the gear and the pinion effects so that the correct gear condition can be inferred.

3.5. Computational time of proposed methodology

The averaged computational time of each step in the proposed methodology was calculated for all the datasets investigated in this section and presented in Table 2. The pinion calculations are faster due to the fact that it has fewer teeth and therefore fewer windows
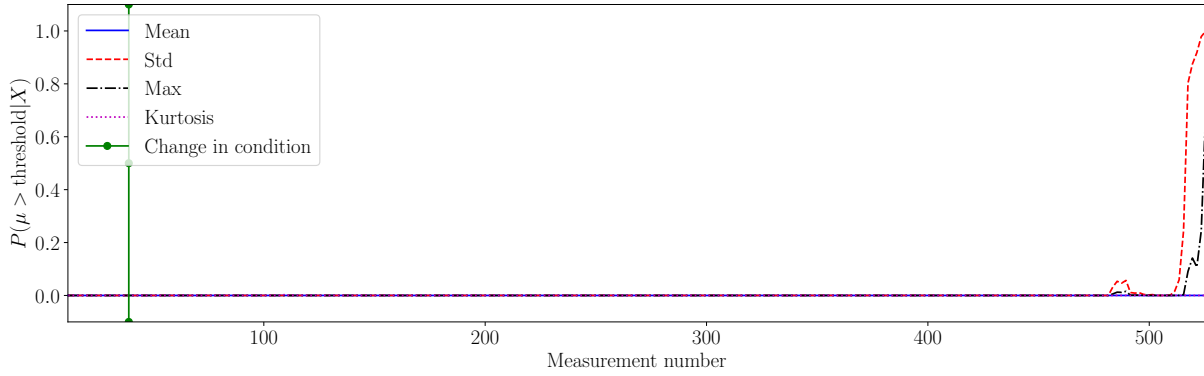
Figure 23: The probability that the mean of the statistics in Figure 22 exceeds the threshold is shown over measurement number.

are used throughout the procedure. The processed results for a single signal e.g. PMMDM can be obtained on average in under three seconds, which implies that much data can be analysed in a reasonable time. This is attributed to the fact that fast wavelet decomposition methods exist, the Gaussian model parameter optimisation consist of a closed form solution and the Bayesian inference results are also based on closed form solutions.

The results obtained from the conventional signal processing tools are presented in Table 3. The computational times of the conventional fault diagnosis techniques are less than the proposed method due to the fact that it is simpler e.g. features do not have to be extracted. However, it is believed that the benefits of using the proposed method for automatic condition inference without historical data will outweigh the computational gains of the conventional signal analysis techniques under many practical conditions.

4. Conclusion

In this paper, a localised gear anomaly detection methodology is proposed to automatically detect localised gear damage. It is based on the assumption that gear teeth in the same condition, generate data that are statistically similar and localised anomalies generate data that are statistically different from the other gear teeth. The continuous wavelet transform, principal component analysis, Gaussian models are utilised to obtain a robust representation that can be used for automatic condition inference by utilising Bayesian data analysis techniques.

The results obtained from the proposed methodology are also compared to the results

Table 2: The averaged computational time is given for a single signal for each process in the proposed methodology. Order tracking and the CWT was implemented in MATLAB 2016a student version, while the rest of the programs were implemented in Python 3.5.4. The computer has an Intel Core i7 2.50 GHz CPU with 16 GB of RAM.

Input	Output	Gear or Pinion	Avg. time [s]
Vibration signal	Raw features	Pinion	0.953581
Vibration signal	Raw features	Gear	1.702086
Raw features	Model parameters	Pinion	0.01457
Raw features	Model parameters	Gear	0.02738
Model parameters	Divergence	Pinion	0.014593
Model parameters	Divergence	Gear	0.048722
Divergence	PMMDM and $P(\mu >$ threshold)	Pinion	0.010684
Divergence	PMMDM and $P(\mu >$ threshold)	Gear	0.0202326
Divergence	Trending of metrics	Gear and Pinion	$< 10^{-3}$

obtained from conventional signal analysis tools i.e. the synchronous average, the synchronous variance of the residual signal and the power spectral density. The results attest to the ability of the proposed method to

- automatically determine the presence of localised gear damage without historical data, to

- obtain metrics which are robust to varying operating conditions, and to

- obtain metrics that are intuitive to understand and to interpret.

This is in contrast to the conventional fault diagnosis tools. The implication of the results is that this methodology can immediately be used to infer whether localised gear damage, which can significantly decrease the life of the machine, is present. The methodology can also be used as a training data quality check for discrepancy analysis i.e. it is not desired to optimise discrepancy analysis on data which contain localised gear damage.

Table 3: The averaged computational time for the conventional signal analysis tools, implemented in Python 3.5.4, is presented. The computer has an Intel Core i7 2.50 GHz CPU with 16 GB of RAM.

Input	Output	Gear or Pinion	Avg. time [s]
Vibration signal	Synchronous average	Gear	0.203121
Vibration signal	Residual signal	Gear	0.429919
Vibration signal	Residual signal (GSA)	Gear	1.895467
Vibration signal	PSD	N/A	0.326327

The advantages of using Bayesian data analysis techniques are also illustrated in this paper; the Bayesian techniques allow the uncertainty to be quantified which can be incorporated into the automatic condition inference procedure. It can also be used to obtain more robust estimates of the machine condition in the presence of noise, attributed for example to changes in operating conditions.

For future work, it is recommended that the methodology be further investigated on more gearbox datasets obtained under non-stationary operating conditions. The benefits of using this approach as a pre-processing step for training discrepancy analysis models need to be investigated as well. Lastly, different features obtained from cyclostationary analysis for example, more flexible models such as kernel density estimators and Gaussian mixture models and Frequentist inference techniques need to be compared to the results obtained in this paper.

Acknowledgements

Appendix A. Sensitivity analysis of the hyperparameters

The hyperparameters of the prior distribution can have a profound impact on the posterior distribution and therefore the choice of the hyperparameters, that were used in the previous analyses, need to be motivated. The hyperparameters were chosen so that the posterior marginal distribution on the mean is dominated by the data and not by the prior

information. However, the choice may not be appropriate and therefore the sensitivity of the hyperparameter choice on the resulting posterior distribution is investigated in this section.

Hierarchical or multilevel Bayesian analysis can be used to incorporate uncertainty in the hyperparameters into the inference process, by using hyperpriors [60]. Hence, the full joint distribution over the random variables is as follows

$$p(\mu, \lambda, \mu_0, \kappa_0, \alpha_0, \beta_0 | X, \boldsymbol{\phi}) = p(\mu, \lambda | \mu_0, \kappa_0, \alpha_0, \beta_0, X) p(\mu_0, \kappa_0, \alpha_0, \beta_0 | \boldsymbol{\phi}), \qquad (A.1)$$

where $p(\mu, \lambda | \mu_0, \kappa_0, \alpha_0, \beta_0, X)$ is the posterior distribution if the hyperparameters are fixed, $p(\mu_0, \kappa_0, \alpha_0, \beta_0 | \boldsymbol{\phi})$ is the distribution over the prior hyperparameters, known as the hyperprior, and $\boldsymbol{\phi}$ is the hyperprior parameters. The joint distribution over the posterior mean and precision is subsequently calculated as

$$p(\mu, \lambda | X, \boldsymbol{\phi}) = \int_{\alpha_0} \int_{\beta_0} \int_{\mu_0} \int_{\kappa_0} p(\mu, \lambda, \alpha_0, \beta_0, \mu_0, \kappa_0 | X, \boldsymbol{\phi}) \ d\alpha_0 \ d\beta_0 \ d\mu_0 \ d\kappa_0, \qquad (A.2)$$

which allows the uncertainty in the hyperparameters to be reflected in the posterior distribution parameters which are subsequently used in the condition inference process. The model with the hyperpriors in Equation (A.1) is called a three-level model and the approach in Section 2.3, with the fixed hyperparameters, is known as a two-level model [60]. Sampling methods are investigated to allow arbitrary hyperpriors to be investigated. A sample s from the hyperprior distribution

$$\mu_0^s, \kappa_0^s, \alpha_0^s, \beta_0^s \sim p(\mu_0, \kappa_0, \alpha_0, \beta_0 | \boldsymbol{\phi}), \qquad (A.3)$$

is used to obtain the posterior Gaussian-Gamma distribution parameters using the process described in Section 2.3. Note that the superscript s denotes that a sample of the variable is obtained or considered. The posterior precision sample is obtained from

$$\lambda^s \sim \text{Gamma}(\lambda | \alpha_n^s, \beta_n^s), \qquad (A.4)$$

and the posterior mean sample is obtained from

$$\mu^s \sim \text{Gaussian}(\mu | \mu_n^s, (\kappa_n^s \lambda^s)^{-1}), \qquad (A.5)$$

which is used for subsequent inference procedures. The sampling process is repeated multiple times to obtain a sufficient number of samples, which is used to represent the posterior distribution over the parameters of interest.
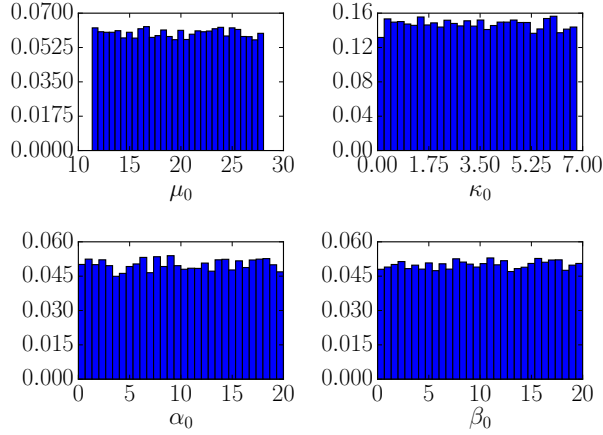
Figure A.24: Probability density function constructed from 20,000 samples taken from the hyperprior.

In this paper, the hyperparameters are assumed to be uncorrelated and uniformly distributed between a and b, which is denoted by $U[a, b]$. The true hyperparameter values are unknown and therefore all values within a sensible range are assumed to be equally probable. The hyperparameters for the processed divergence data $\boldsymbol{d}_i^{proc}$ are sampled from, $\kappa_0 \sim U[10^{-4}, n/5]$, $\mu_0 \sim U[\min(\boldsymbol{d}_i^{proc}), \max(\boldsymbol{d}_i^{proc})]$, $\alpha_0 \sim U[10^{-4}, 20]$, $\beta_0 \sim U[10^{-4}, 20]$, which is subsequently used to obtain the posterior parameters. The number of elements in $\boldsymbol{d}_i^{proc}$ is n and $\min(\boldsymbol{d}_i^{proc})$ and $\max(\boldsymbol{d}_i^{proc})$ denotes the minimum and maximum value of $\boldsymbol{d}_i^{proc}$ respectively.

The samples from the hyperprior is shown in Figure A.24. The posterior distribution from this three-level approach is compared to the two-level approach in Figure A.25. The hyperparameters of the two-level approach are the same as in Section 3.2 for the divergence data of a specific window. The posterior marginal distribution on the mean is very similar for the two approaches, with the probability density function of the two-level approach being slightly more localised. This means that the uncertainty in the posterior mean is larger for the three-level approach due to the additional uncertainty in the hyperparameters. The precision is more sensitive to the choice of hyperparameters, and therefore a broader probability density function is seen for the three-level approach. The mode of the three-level precision posterior is closer to the maximum likelihood data than the two-level approach. The posterior marginal distribution of the mean is used for condition inference and is very similar for the two-level and the three-level approaches, which validates that the choice of hyperparameters in the previous analyses is sensible. The
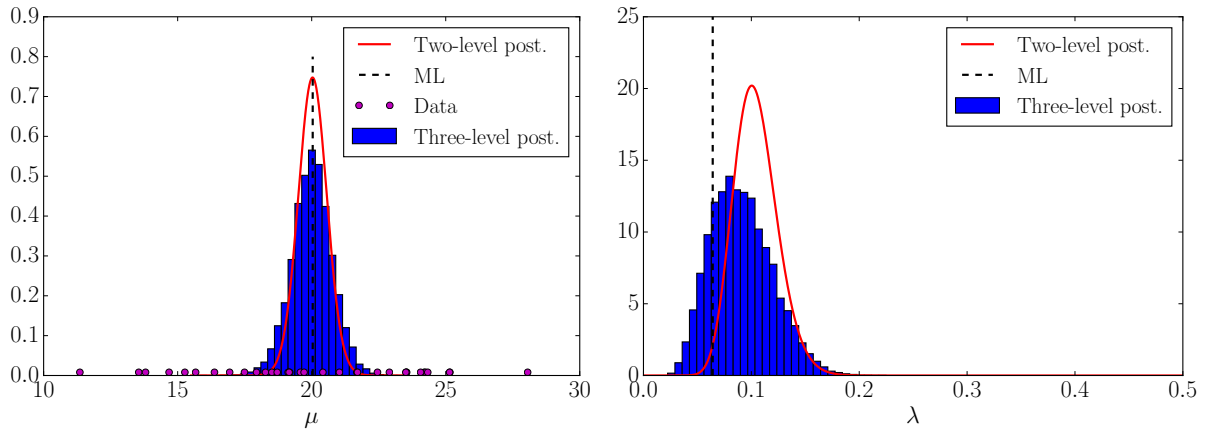
Figure A.25: The probability density function constructed from 20,000 samples, taken from the hyperprior of the three-level approach, is compared to the two-level approach and the maximum likelihood (ML) estimate.

three-level approach was not used for the other investigations, because similar inferences will be made if Equation (13) is used, but with a significant increase in computational time.

References

[1] J. P. Salameh, S. Cauet, E. Etien, A. Sakout, L. Rambault, Gearbox condition monitoring in wind turbines: A review, Mechanical Systems and Signal Processing 111 (2018) 251–264.

[2] F. Pedro, G. Márquez, A. Mark, J. María, P. Pérez, M. Papaelias, Condition monitoring of wind turbines: Techniques and methods, Renewable Energy 46 (2012) 169–178.

[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] 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.

[5] W. Bartelmus, R. Zimroz, Vibration condition monitoring of planetary gearbox under

varying external load, Mechanical Systems and Signal Processing 23 (1) (2009) 246–257.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.

[12] J. Obuchowski, A. Wylomanska, R. Zimroz, Selection of informative frequency band in local damage detection in rotating machinery, Mechanical Systems and Signal Processing 48 (1-2) (2014) 138–152.

[13] J. Yin, W. Wang, Z. Man, S. Khoo, Statistical modeling of gear vibration signals and its application to detecting and diagnosing gear faults, Information Sciences 259 (2014) 295–303.

[14] W. A. Smith, Z. Fan, Z. Peng, H. Li, R. B. Randall, Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference, Mechanical Systems and Signal Processing 75 (2015) 371–394.

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

[16] 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.

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

[18] W. Wang, A. K. Wong, Autoregressive model-based gear fault diagnosis, Journal of Vibration and Acoustics 124 (2) (2002) 172.

[19] W. Wang, Early detection of gear tooth cracking using the resonance demodulation technique, Mechanical Systems and Signal Processing 15 (5) (2001) 887–903.

[20] 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.

[21] S. J. Loutridis, Damage detection in gear systems using empirical mode decomposition, Engineering Structures 26 (12) (2004) 1833–1841.

[22] Y. Lei, J. Lin, Z. He, M. J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing 35 (1-2) (2013) 108–126.

[23] 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.

[24] 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.

[25] 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.

[26] W. Wang, B. D. Forrester, P. C. Frith, A unified approach to detecting and trending changes caused by mechanical faults in rotating machinery, Structural Health Monitoring 15 (2) (2016) 204–222.

[27] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mechanical Systems and Signal Processing 72-73 (2016) 303–315.

[28] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mechanical Systems and Signal Processing 107 (2018) 241–265.

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

[30] 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.

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

[32] J. D. Wu, J. J. Chan, Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network, Expert Systems with Applications 36 (5) (2009) 8862–8875.

[33] C. Y. Yang, T. Y. Wu, Diagnostics of gear deterioration using EEMD approach and PCA process, Measurement 61 (2015) 75–87.

[34] Y. Lei, M. J. Zuo, Gear crack level identification based on weighted K nearest neighbor classification algorithm, Mechanical Systems and Signal Processing 23 (5) (2009) 1535–1547.

[35] 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.

[36] S. Haidong, J. Hongkai, Z. Huiwei, W. Fuan, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mechanical Systems and Signal Processing 95 (2017) 187–204.

[37] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1–10.

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

[39] L. M. R. Baccarini, V. V. Rocha E Silva, B. R. De Menezes, W. M. Caminhas, SVM practical industrial application for mechanical faults diagnostic, Expert Systems with Applications 38 (6) (2011) 6980–6984.

[40] M. Timusk, M. Lipsett, C. K. Mechefske, Fault detection using transient machine signals, Mechanical Systems and Signal Processing 22 (7) (2008) 1724–1749.

[41] 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.

[42] N. Baydar, Q. Chen, A. Ball, U. Kruger, Detection of incipient tooth defect in helical gears using multivariate statistics, Mechanical Systems and Signal Processing 15 (2) (2001) 303–321.

[43] 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.

[44] 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) 1–8.

[45] F. Harrou, Y. Sun, M. Madakyaru, Kullback-Leibler distance-based enhanced detection of incipient anomalies, Journal of Loss Prevention in the Process Industries 44 (2016) 73–87.

[46] 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.

[47] 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.

[48] S. Schmidt, P. S. Heyns, J. P. de Villiers, A novelty detection diagnostic methodology for gearboxes operating under fluctuating operating conditions using probabilistic techniques, Mechanical Systems and Signal Processing 100 (2018) 152–166.

[49] J.-D. Wu, C.-H. Liu, An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network, Expert Systems with Applications 36 (3) (2009) 4278–4286.

[50] 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.

[51] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[52] S. Liu, M. Yamada, N. Collier, M. Sugiyama, Change-point detection in time-series data by relative density-ratio estimation, Neural Networks 43 (2013) 72–83.

[53] L. Xie, J. Zeng, U. Kruger, X. Wang, J. Geluk, Fault detection in dynamic systems using the Kullback Leibler divergence, Control Engineering Practice 43 (2015) 39–48.

[54] A. C. Bittencourt, K. Saarinen, S. Sander-Tavallaey, S. Gunnarsson, M. Norrlöf, A data-driven approach to diagnostics of repetitive processes in the distribution domain - Applications to gearbox diagnostics in industrial robots and rotating machines, in: Mechatronics, Vol. 24, 2014, pp. 1032–1041.

[55] F. Ferracuti, A. Giantomassi, S. Iarlori, G. Ippoliti, S. Longhi, Electric motor defects diagnosis based on kernel density estimation and Kullback Leibler divergence in quality control scenario, Engineering Applications of Artificial Intelligence 44 (2015) 25–32.

[56] J. Zeng, U. Kruger, J. Geluk, X. Wang, L. Xie, Detecting abnormal situations using the Kullback-Leibler divergence, Automatica 50 (11) (2014) 2777–2786.

[57] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 4th Edition, Academic Press, 2008.

[58] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, [Online; accessed 2017-04-20] (2001–).
URL http://www.scipy.org/

[59] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian data analysis, 3rd Edition, Chapman and Hall/CRC Texts in Statistical Science, 2013.

[60] B. P. Carlin, T. A. Louis, Bayesian methods for data analysis, CRC Press, 2008.

[61] K. P. Murphy, Conjugate Bayesian analysis of the Gaussian distribution (2007).