Measurement 222 (2023) 113515

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

An anomalous frequency band identification method utilising available healthy historical data for gearbox fault detection

Stephan Schmidt [a,*], Konstantinos C. Gryllias [b,c]

[a] Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
[b] Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300, 3001 Heverlee, Leuven, Belgium
[c] Flanders Make, KU Leuven, Belgium

ARTICLE INFO

Keywords:
Gearbox fault detection
Frequency band identification
Blind features/indicators
Squared envelope spectrum

ABSTRACT

Informative frequency band identification methods are used to automatically design bandpass filters to enhance fault signatures in vibration measurements. Blind and targeted features can be used to guide the frequency band selection process. Blind features' performance is impeded when there are dominant non-stationary extraneous components, whereas targeted features' performance is impeded when the characteristic frequency of the machine component is unknown, erroneously estimated or the damaged component is not targeted. An anomalous frequency band identification method is proposed that utilises the available historical data to detect weak damage components that deviate from the baseline or reference condition. This makes it possible to ignore dominant extraneous components that are also present in the historical dataset. The proposed method is analysed and compared against conventional and feature ratio methods on numerical and experimental datasets. The results demonstrate that the proposed method has much potential for identifying informative frequency bands for fault detection.

1. Introduction

Gearboxes are widely used in the power generation and mining industries [1,2]. Condition monitoring is employed on critical gearboxes for incipient fault detection, fault component identification and fault trending to detect a deterioration in their health [1,3]. Gear and bearing damage manifest as changes in the vibration signature of the system; and therefore vibration-based diagnostics and prognostics are much researched [4–7]. However, the performance of the vibration-based methods is impeded by varying operating conditions, by the presence of non-Gaussian or impulsive noise and by weak fault signatures masked by dominant signal components such as extraneous components (e.g. electromagnetic interference) [8–10].

The damage information often manifests in narrow frequency bands [11–13]. Therefore, the fault signatures can be enhanced by extracting the signals from the aforementioned frequency bands using bandpass filters or by applying matched filters [14]. The damage components often result in amplitude modulation and therefore the squared envelope spectrum can be used to identify the underlying periodicities in the amplitude modulation and the synchronous average of the squared envelope can be used to visualise the modulation caused by the damage (e.g. to determine whether the damage is potentially spread over multiple teeth or localised to a single tooth) [12].

Informative frequency band selection methods such as the kurtogram and its extensions [15–17], the protrugram [18], the sparsogram [19], the infogram [20], the autogram and its improvements [21–23], the distcsgram [24], the entrogram [25], the IESFOgram [26] and the PESOgram [27] are some of the methods that have been used to enhance weak fault signatures. Hebda-Sobkowicz et al. [28,29] proposed and compared methods for informative frequency band identification under non-Gaussian noise conditions. Ahsan and Bismor [30] used the harmonic search function with different kurtosis-based fitness functions for selecting the frequency band. Smith et al. [11] distinguished between blind and targeted demodulation-band selection methods and compared the performance of existing blind and targeted methods against the log-cycligram.

Targeted features extract the fault symptoms of specific cyclic components (e.g. the peak of the ball-pass outer race order in the envelope spectrum), whereas, blind features do not make any assumptions about the cyclic orders of the fault symptoms [11]. More specifically, blind features such as the spectral kurtosis [14], negentropy [20], Gini index [31] and its improvements [32], L2/L1-norm ratio [31,33], and Hoyer index [34] do not require prior knowledge about which mechanical component (e.g. inner race of a bearing) is damaged to guide the frequency band selection. For example, damaged signals are expected

* Corresponding author.
E-mail addresses: stephan.schmidt@up.ac.za (S. Schmidt), konstantinos.gryllias@kuleuven.be (K.C. Gryllias).

https://doi.org/10.1016/j.measurement.2023.113515
Received 28 May 2023; Received in revised form 18 August 2023; Accepted 31 August 2023
Available online 9 September 2023

to be sparse in the envelope spectrum, and therefore the sparsity in the squared envelope spectrum of a bandlimited signal [31,33] can be used to find informative frequency bands. In contrast, targeted features such as the indicator of cyclostationarity aim to identify fault components of specific pre-selected mechanical components [11]. This is specifically performed by targeting the characteristic frequencies or orders of the potentially damaged components, e.g. the inner race bearing is targeted by using the ball pass order of the inner race [11,12,26,35].

Blind methods can be used to identify multiple frequency bands with potential signal components of interest for further interrogation. However, it is desired to only isolate the damage component when performing automatic fault detection. The performance of blind methods for single frequency band identification is impeded by the presence of extraneous components [12,36]. Since targeted methods are calculated for specific pre-selected components (e.g. the inner race bearing damage component), they are much less affected by extraneous impulses or repetitive transients [12]. However, the implementation of the methods is impeded when the kinematics of the system is unknown, when the characteristic frequency of a damaged component is unknown, or when the damaged component is not targeted. This can lead to false negatives (i.e. damage is not detected), which ultimately would result in a delayed detection or even unexpected breakdowns. Therefore, there is an incentive to find features that do not require the fault components to be known a priori.

We expect that condition monitoring measurements are measured on a regular basis to enable early damage detection and to perform fault trending (e.g. [37]) and therefore it is assumed that historical data are available that can be utilised for improved frequency band identification and detection. Hong et al. [38] used the enhanced kurtogram, a Gaussian mixture model and principal component analysis to identify fault states for classification. Hou et al. [39] identified that the kurtogram is impeded by extraneous components and therefore proposed a data-aided method that uses healthy and damaged data for frequency band identification. Udmale and Singh [40] combined the kurtogram with an extreme learning machine for automatic fault classification.

Historical fault data are scarce and therefore methods are also required for cases where only healthy historical data or data from a reference condition are available. Hou et al. [41] used the difference spectrum decomposition between reference signals and the signal under consideration for fault enhancement. In the frequency band identification field, the SKRgram uses the ratio of the kurtogram feature plane of a gearbox in an unknown condition to the kurtogram feature plane of a healthy system to identify important frequency bands [42]. Different features have also been used in the development of similar feature planes (e.g. [43,44]). We refer to these methods as feature ratio methods in this work.

Feature ratio methods require the selection of a measurement for constructing the reference feature plane [42]. However, there are usually many historical measurements available that can be utilised to identify important frequency bands and therefore the selection of only a single measurement for constructing the reference feature plane needs to be improved. Therefore, a frequency band identification approach that utilises all of the available historical data is needed to perform automatic frequency band identification. Hence, in this work an anomalous frequency band identification method is proposed that uses all available healthy historical data to identify frequency bands with potential damage. It is intended for applications where historical data is available in a reference condition and deviations from this reference condition needs to be detected. More specifically, in this work, the proposed method is implemented to detect bearing damage in a numerical gearbox model and tested to detect gear damage in experimental data. The method is compared to conventional frequency band identification methods and to the feature ratio methods.

In summary, the contributions of this work are as follows:

- A new method is proposed to incorporate historical data in the frequency band identification process. This makes it possible to move from the conventional methods of finding frequency bands that maximise features, to finding frequency bands that shows the largest deviations from the reference condition.
- It is shown that the proposed method can naturally incorporate all of the reference historical data to support the frequency band identification process and does not require the a priori selection of a specific reference signal.
- The proposed method is compared to conventional frequency band identification methods and feature plane ratio methods on numerical and experimental datasets acquired under constant and time-varying operating conditions respectively. The results demonstrate the potential of the proposed method.

The layout of this work is as follows. The proposed method is presented in Section 2. Thereafter, the method is evaluated and compared to conventional and feature-ratio frequency band identification methods on numerical gearbox data in Section 3 and on experimental gearbox data in Section 4. The work is finally concluded in Section 5. Appendix A contains more information on the numerical gearbox model.

2. Anomalous frequency band identification method

Firstly, an overview is presented of a conventional frequency band identification method and of the proposed anomalous frequency band identification method. The different steps of the proposed anomalous frequency band identification method are discussed in detail in Sections 2.1–2.4.

The conventional frequency band identification method, shown in Fig. 1(a), constructs a feature plane by extracting features from a set of bandpass filtered signals. This feature plane, denoted $\boldsymbol{\Psi}$, is maximised to find the frequency band of interest. This process is described in Section 2.1.

In contrast to the conventional method, the proposed method contains two phases: In the training phase, shown in Fig. 1(b), it is assumed that N historical measurements are available from the machine in a reference condition. A feature plane is calculated of each signal using conventional methods, with the feature plane of the nth signal denoted $\boldsymbol{\Psi}_n$ whereafter the parameters of the proposed feature plane are estimated from the historical data, with the estimated parameters denoted $\{\hat{\theta}_h\}$. In the application phase shown in Fig. 1(c), the estimated parameters are used to calculate an anomalous feature plane $\boldsymbol{\Xi}$ of a new signal. The feature plane is subsequently maximised to identify the feature band of interest $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$. The training and application phases are discussed in Section 2.2 and the frequency band identification process is described in Section 2.3. The measured signal is subsequently bandpass filtered using the identified frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$ and analysed for damage as described in Section 2.4. The proposed method, in Figs. 1(b)– 1(c), uses deviations from the healthy signal for informative frequency band detection, which makes it more sensitive to damage and less sensitive to extraneous components. The benefits of the proposed method over the conventional method and feature ratio methods are discussed further in detail in Section 2.3.

2.1. Conventional feature plane decomposition

Conventional feature plane decomposition methods are also used to construct the frequency ratio methods and the proposed method. Therefore, a brief overview of the essential concepts of the conventional feature plane decomposition methods is first presented.

Before applying the conventional feature plane decomposition met hods, a feature, that is sensitive to damage, needs to be identified. Damaged rotating mechanical components often result (i) in an increase

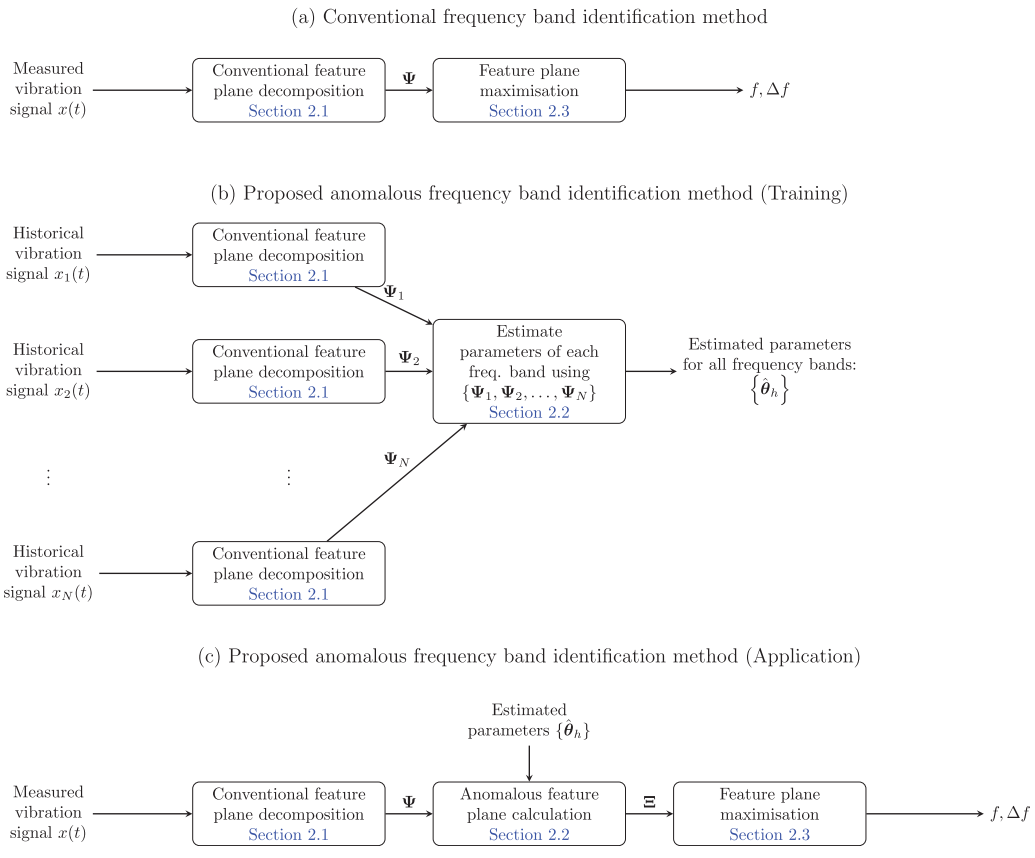

Fig. 1. The conventional and proposed anomalous frequency band identification methods are presented with the proposed method separated into a training and application phase.

in repetitive impulses in the time domain, (ii) in the presence of statistically significant components in the cyclic spectrum, and (iii) in the increase of the impulsiveness of the time domain signal. Therefore, measures such as the kurtosis, negentropy and sparsity indicators (e.g. L2/L1, Gini, Hoyer) are often utilised in the frequency band identification process [11,14,15,20,31,34]. Even though any feature can be utilised in this method, the proposed method will be implemented and compared using the Gini index, the negentropy and the Hoyer index. These blind features are selected since they have performed very well to improve the damage detection capabilities of frequency band identification methods [20,31,34,45]; the negentropy performed better than all the sparse measures considered in Ref. [46]; the Hoyer index is better suited than the L2/L1-ratio according to Ref. [47]; and the Gini index performed well relative to other the sparsity measures in Ref. [31].

The first step of the process is to decompose the signal into a time–frequency domain for a given frequency resolution Δf. This is performed using the Short-Time Fourier Transform (STFT) as described in Ref. [20] in this work. The STFT with a frequency resolution of Δf is denoted $\mathrm{STFT}(t; f, \Delta f)$ at a time step t and a frequency f. The STFT is used to calculate the features for different combinations of f and Δf. This makes it possible to identify the optimal frequency band $[f - \Delta f/2, f + \Delta f]$. The reader is referred to Refs. [11,15] for more information on different decomposition methods.

Subsequently, the features are extracted from the STFT. The Gini index measures the degree of inequality in a distribution or sparsity in data [31,34]. The Gini index is defined by [31]:

$$\psi(f, \Delta f) = 1 - 2 \cdot \sum_{n=1}^{N} \frac{\tilde{\mathrm{SE}}(t_n; f, \Delta f)}{\|\mathrm{SE}(t_n; f, \Delta f)\|_{L_1}} \left(\frac{N - n + \frac{1}{2}}{N} \right) \tag{1}$$

where the squared envelope $\mathrm{SE}(t_n; f, \Delta f)$ is estimated from the STFT, i.e. $\mathrm{SE}(t_n; f, \Delta f) = |\mathrm{STFT}(t_n; f, \Delta f)|^2$. The sorted squared envelope signal, denoted $\tilde{\mathrm{SE}}(t_n; f, \Delta f)$, is sorted so that $\tilde{\mathrm{SE}}(t_1; f, \Delta f) \leq \tilde{\mathrm{SE}}(t_2; f, \Delta f) \ldots \leq \tilde{\mathrm{SE}}(t_N; f, \Delta f)$. The L_1-norm of the squared envelope $\mathrm{SE}(t_n; f, \Delta f)$, calculated over the time axis, is denoted $\|\mathrm{SE}(t_n; f, \Delta f)\|_{L_1}$.

The negentropy [20]

$$\psi(f, \Delta f) = \left\langle \frac{|\mathrm{SE}(t_n; f, \Delta f)|^2}{\left\langle |\mathrm{SE}(t_n; f, \Delta f)|^2 \right\rangle_n} \cdot \log \left(\frac{|\mathrm{SE}(t_n; f, \Delta f)|^2}{\left\langle |\mathrm{SE}(t_n; f, \Delta f)|^2 \right\rangle_n} \right) \right\rangle_n \tag{2}$$

is the second indicator considered in this work. Antoni [20] proposed the SE infogram, the SES infogram and the weighted average of the SE and the SES infogram. The SE infogram and the SES infogram are calculated using the negentropy of the squared envelope and the squared envelope spectrum respectively. The SE infogram measures the impulsiveness in the signal and is used in this work [20].

Lastly, the Hoyer index [34]

$$\psi(f, \Delta f) = \frac{\sqrt{N}}{\sqrt{N} - 1} - \frac{1}{\sqrt{N} - 1} \frac{\sum_{n=1}^{N} |\mathrm{SE}(t_n; f, \Delta f)|}{\sqrt{\sum_{n=1}^{N} |\mathrm{SE}(t_n; f, \Delta f)|^2}} \tag{3}$$

of the squared envelope is considered. The Hoyer index is closely related to the L2/L1 norm; instead of maximising the L2/L1 ratio, the negative of the normalised L1/L2 ratio is maximised. The feature calculation process is summarised in Fig. 2 for a specific frequency bandwidth Δf. This is repeated for different Δf to construct the feature planes. We used the following window lengths $[4, 8, 16, 32, 64, 128, 256, 512]$ for the STFT calculation of all datasets.

In summary, each of the aforementioned features is evaluated for different combinations of centre frequencies f and frequency bandwidths Δf to obtain a feature plane $\boldsymbol{\Psi}$ as shown in Fig. 3(a). This

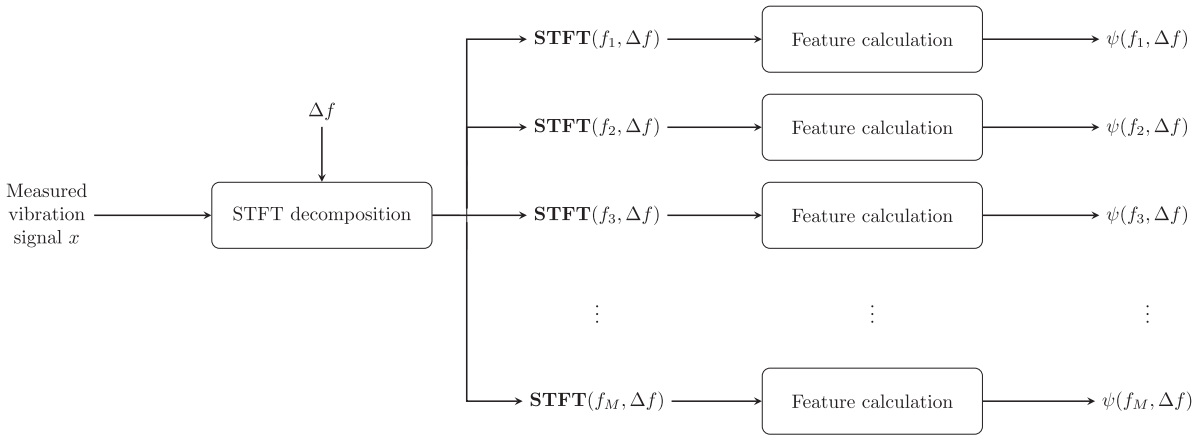

Fig. 2. The feature calculation process is summarised for a given Δf and feature (e.g. Eqs. (1)–(3)). The STFT decomposition of the signal is calculated, with the STFT data of the frequency band $[f_i - \Delta f/2, f_i + \Delta f/2]$ denoted $\mathbf{STFT}(f_i, \Delta f) = [\mathrm{STFT}(t_1; f_i, \Delta f), \mathrm{STFT}(t_2; f_i, \Delta f), \ldots \mathrm{STFT}(t_N; f_i, \Delta f)]$. The frequency resolution Δf is dependent on the window length. The process is repeated for all candidate Δf values to construct the feature plane in Fig. 3(a).

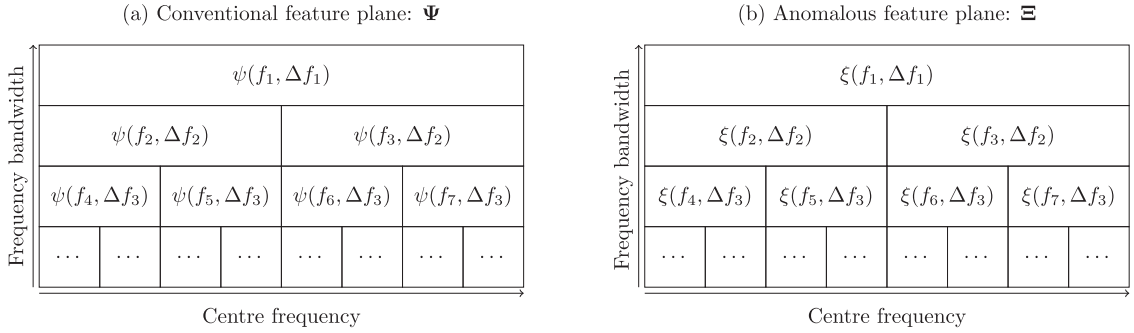

Fig. 3. Examples of conventional and the anomalous feature planes are presented with different centre frequencies denoted f_k and different frequency bandwidths denoted Δf_j, where $\Delta f_1 > \Delta f_2 > \Delta f_3$. The conventional feature associated with the centre frequency f_i and frequency bandwidth Δf_j is denoted $\psi(f_i, \Delta f_j)$ and used to construct the conventional feature plane $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\Psi}$ consists of all the centre frequencies and frequency bandwidths shown in (a). The anomalous feature plane construction process is described in Section 2.2. The anomalous feature plane, i.e. the anomalous feature associated at the centre frequency f_i and frequency bandwidth Δf_j, are denoted $\xi(f_i, \Delta f_j)$ and used to construct the anomalous feature plane $\boldsymbol{\Xi}$. Eq. (5) is used to calculate each $\xi(f_i, \Delta f_j)$. This is only an example of a feature plane; other feature plane decomposition methods are discussed in Refs. [11,15]. .

process is used to obtain the conventional feature planes in Fig. 1. The feature plane is also often used to visualise the objective function of the frequency band identification problem.

2.2. Anomalous feature plane calculation (training and application)

The objective is to find a method to identify anomalous frequency bands in the feature plane. Feature ratio frequency band identification methods (e.g. SKRgram), divide the feature of the signal in an unknown condition $\psi(f, \Delta f)$ by the feature from a measurement in the historical dataset $\psi_n(f, \Delta f)$ to construct the feature plane, i.e.

$$\kappa(f, \Delta f) = \frac{\psi(f, \Delta f)}{\psi_n(f, \Delta f)} \tag{4}$$

In contrast, the proposed anomalous feature plane of a signal $x(t)$ has the following form for the frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$:

$$\xi(f, \Delta f) = \frac{\psi(f, \Delta f) - m(f, \Delta f)}{s(f, \Delta f)} \tag{5}$$

where $\psi(f, \Delta f)$ is the conventional feature plane of the signal $x(t)$. The anomalous feature plane has two sets of parameters for each frequency band, the location $m(f, \Delta f)$ and scaling $s(f, \Delta f)$ parameters. The parameters $\theta_h(f, \Delta f) = [m(f, \Delta f), s(f, \Delta f)]$ are functions of the centre frequency f and bandwidth Δf respectively. Eq. (5) is used for all considered centre frequencies and frequency bandwidths to construct the anomalous feature plane $\boldsymbol{\Xi}$ as shown in Fig. 3(b).

In the training phase of the proposed method, the parameters of the anomalous feature plane, are estimated from the historical data in two steps. In the first step of the training phase, a feature plane (i.e. Fig. 3(a)) is calculated for each measurement in the historical dataset as shown in Fig. 1(b). The feature plane of the nth measurement is denoted $\boldsymbol{\Psi}_n$. Each feature plane consists of features calculated from a combination of centre frequencies f and frequency bandwidths Δf. The feature associated with the frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$, calculated from the nth measurement in the historical dataset, is denoted $\psi_n(f, \Delta f)$. The set of N features for the frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$ is denoted $\{\psi_n(f, \Delta f)\} = \{\psi_1(f, \Delta f), \psi_2(f, \Delta f), \ldots, \psi_N(f, \Delta f)\}$ and used to find the parameters $\theta_h(f, \Delta f) = [m(f, \Delta f), s(f, \Delta f)]$ of Eq. (5).

In the second step, after the conventional feature planes have been calculated, the parameters of each frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$ are estimated from the historical dataset $\{\psi_n(f, \Delta f)\}$ as follows: The location parameter is set to the expected value of the feature from a machine in the reference condition, i.e.

$$m(f, \Delta f) = \mathbb{E}_{\psi \sim p_{ref}(\psi; f, \Delta f)}\{\psi(f, \Delta f)\} \tag{6}$$

where $\mathbb{E}_{\psi \sim p_{ref}(\psi; f, \Delta f)}\{\psi(f, \Delta f)\}$ denotes the expected value of the feature under the (unknown) probability density function (pdf) from the machine in a reference condition denoted $p_{ref}(\psi; f, \Delta f)$ for the frequency band $[f - \frac{1}{2} \cdot \Delta f, f + \frac{1}{2} \cdot \Delta f]$. The squared scaling parameter can be estimated from

$$s^2(f, \Delta f) = \mathbb{E}_{\psi \sim p_{ref}(\psi; f, \Delta f)}\left\{(\psi(f, \Delta f) - m(f, \Delta f))^2\right\} \tag{7}$$

which is also the variance of the feature $\psi(f, \Delta f)$ from the machine in a reference condition. These parameters are estimated with Monte Carlo

integration as follows:

$$\hat{m}(f,\Delta f) = \frac{1}{N}\sum_{n=1}^{N}\psi_n(f,\Delta f) \tag{8}$$

and

$$\hat{s}^2(f,\Delta f) = \frac{1}{N}\sum_{n=1}^{N}\left(\psi_n(f,\Delta f) - \hat{m}(f,\Delta f)\right)^2 \tag{9}$$

where $\psi_n(f,\Delta f)$ denotes the feature of the nth historical measurement from the machine in the reference condition, which we assume are sampled from $p_{ref}(\psi; f, \Delta f)$. Eqs. (8) and (9) need to be used for all centre frequency–frequency bandwidth pairs in the feature plane (e.g. Fig. 3) and therefore a set of parameters $\{\hat{\theta}_h\} = \{\hat{\theta}_h(f_1,\Delta f_1), \hat{\theta}_h(f_2,\Delta f_2)\ldots\}$ are obtained, whereafter Eq. (5) is used to calculate the anomalous feature plane.

For finite measurements N, the Monte Carlo integration only provides estimates of the parameters. The estimator in Eq. (9) is biased for small N and therefore $\frac{1}{N}$ can be replaced with $\frac{1}{N-1}$ to correct for the bias. It is however recommended that the historical measurements should be representative of the machine under its normal operating conditions and a large amount of representative measurements should be used to ensure the Monte Carlo estimates in Eqs. (8) and (9) are accurate. We have, however, found that 40 historical measurements are adequate to obtain acceptable results for the considered datasets. The historical measurements should be acquired from the machine in a reference condition (e.g. a new machine, machine after corrective maintenance was performed), whereafter deviations from the reference condition are detected.

The proposed anomalous frequency plane has a few properties that are beneficial when compared against conventional and feature ratio methods.

Firstly, the expected value of the numerator $\psi(f,\Delta f) - m(f,\Delta f)$ is zero if the frequency band does not contain anomalous information. This is reminiscent of the classical spectral features such as the spectral kurtosis, the L2/L1 norm, and the Gini index where often a constant is subtracted to ensure that the expected value of the feature is zero under Gaussian noise, e.g. the spectral kurtosis has the form $\psi(f,\Delta f) - 2$ and the L2/L1 norm has the form $\psi(f,\Delta f) - \sqrt{2}$ if the signal is complex Gaussian. Hence, the baseline is usually Gaussian stationary noise with the conventional frequency band identification method. With the proposed method, the baseline is given by $m(f,\Delta f)$ which is based on the reference machine's expected behaviour in the frequency band and deviations from this reference is used for identifying the informative frequency band. This is expected to make it more sensitive to damaged components, which manifest as deviations from the reference condition.

Secondly, the negative log-likelihood, which is frequently used for anomaly detection, has the following form: $-\frac{1}{2}\log(2\pi) - \log s(f,\Delta f) - \frac{1}{2}\cdot\left(\frac{\psi(f,\Delta f)-m(f,\Delta f)}{s(f,\Delta f)}\right)^2$ if the feature is Gaussian distributed with a mean $m(f,\Delta f)$ and variance $s^2(f,\Delta f)$. The negative log-likelihood can be written in terms of the proposed feature as follows: $-\frac{1}{2}\log(2\pi) - \log s(f,\Delta f) - \frac{1}{2}\cdot\xi(f,\Delta f)^2$. The negative log-likelihood is insensitive to the sign of the deviation $\psi(f,\Delta f) - m(f,\Delta f)$ due to the squaring operation, however, we are usually interested in positive deviations when performing diagnosis (e.g. the signal usually becomes more leptokurtic or more sparse when the signal is damaged). This is the rationale behind using Eq. (5) instead of the negative log-likelihood and instead of $|\xi(f,\Delta f)|^2$.

Thirdly, in contrast to feature ratio methods, the proposed method incorporates the variance of the feature in the anomaly score calculation. More specifically, Eq. (5) normalises the distance of the observed feature from the baseline, i.e. $\psi(f,\Delta f) - m(f,\Delta f)$ by the standard deviation $s(f,\Delta f)$ of the reference feature in the frequency band. This means that frequency bands with large variations in the reference dataset will have a smaller effect on the anomaly score than frequency bands with smaller variations. This is expected to improve the reliability of the proposed method over equivalent feature ratio methods.

2.3. Frequency band identification

The objective of frequency band identification methods is to automatically find the frequency band $[f - \Delta f/2, f + \Delta f/2]$ with potential damage information. The classical informative frequency band identification approaches solve the following optimisation problem

$$\arg\max_{f,\Delta f}\ \psi(f,\Delta f) - C \tag{10}$$

for the preselected features, e.g. Eqs. (1)–(3). Since the feature plane contains by construction discrete centre frequencies and frequency bandwidths (e.g. Fig. 3), the maximisation process can be efficiently solved by searching for the optimal combination of centre frequencies f and bandwidths Δf on the grid. The constant C is usually the expected value of the feature for a Gaussian stationary signal. Since C is constant over all frequency bands, it is equivalent to solve the conventional frequency band optimisation problem without the constant, i.e. only maximising $\psi(f,\Delta f)$. A minimisation problem (e.g. if a feature needs to be minimised) can be converted to a maximisation problem by multiplying the feature with -1, i.e. $-\psi(f,\Delta f)$. The proposed method solves the following problem

$$\arg\max_{f,\Delta f}\ \xi(f,\Delta f) \tag{11}$$

to determine the informative frequency bands. The difference between the proposed method in Eq. (11) and the conventional method in Eq. (10) is highlighted by substituting Eq. (5) in Eq. (11) to obtain

$$\arg\max_{f,\Delta f}\ \frac{\psi(f,\Delta f) - \hat{m}(f,\Delta f)}{\hat{s}(f,\Delta f)} \tag{12}$$

which can be written in a similar form to Eq. (10):

$$\arg\max_{f,\Delta f}\ \frac{\psi(f,\Delta f)}{\hat{s}(f,\Delta f)} + \hat{C}(f,\Delta f),\ \text{with}\ \hat{C}(f,\Delta f) = -\frac{\hat{m}(f,\Delta f)}{\hat{s}(f,\Delta f)} \tag{13}$$

The proposed method uses the historical data's feature plane behaviour (e.g. its estimated expected value $\hat{m}(f,\Delta f)$ and estimated variance $\hat{s}^2(f,\Delta f)$ in a specific frequency band) to determine the importance of the frequency band $[f-\Delta f/2, f+\Delta f/2]$. In contrast to solving Eq. (10), the proposed method scales the feature by the reciprocal of the scaling parameter $\hat{s}(f,\Delta f)$ in the band and adds a frequency band dependent constant $\hat{C}(f,\Delta f)$ as shown in Eq. (13). Therefore, frequency bands with large variances are penalised. Ultimately, the proposed method finds the frequency band with the largest relative increase in feature from the reference condition as shown in Eq. (12), which makes it much less sensitive to extraneous components and therefore it is expected to perform much better than conventional methods.

In the next section, the processing and analysis of the measured signal using the data are discussed.

2.4. Processing and analysis

The measured vibration signal $x(t)$ can be filtered using the identified frequency band $[f - \Delta f/2, f + \Delta f/2]$ and subsequently analysed for damage. In this work, the Squared Envelope Spectrum (SES) and the Synchronous Average of the Squared Envelope (SASE) of the bandpass filtered signal are calculated [12].

The bandpass filtered signal, with a passband over the domain $[f - \Delta f/2, f + \Delta f/2]$, is denoted $|x_f(t; f, \Delta f)|^2$ in the time domain and the order tracked filtered signal is denoted $|x_f(\varphi; f, \Delta f)|^2$ where φ is the angle variable. The squared envelope spectrum is estimated with

$$\mathrm{SES}_x(\alpha; f, \Delta f) = \mathcal{F}_{\varphi\to\alpha}\{|x_f(\varphi; f, \Delta f)|^2\} \tag{14}$$

where is the Fourier transform $\mathcal{F}_{\varphi\to\alpha}$ of the squared bandpass filtered signal $|x_f(\varphi; f, \Delta f)|^2$. The Fourier transform converts the signal from the angle domain φ to the cyclic domain α. The squared envelope is powerful to identify the periodicity of the modulation in the signal and

can therefore be used to identify the characteristic frequency (or order) of the damaged components. The normalised SES [48]

$$\mathrm{NSES}_x(\alpha; f, \Delta f) = \left(\frac{\mathrm{SES}_x(\alpha; f, \Delta f)}{\mathrm{SES}_x(0; f, \Delta f)} \right)^2 \tag{15}$$

is used to compare the performance of the methods in this work. The normalised SES can be used to calculate an indicator of cyclostationarity $\mathrm{ICS}_x = \sum_\alpha \mathrm{NSES}_x(\alpha)$ for example.

The synchronous average of the squared envelope can be used to visualise the modulation and understand the nature of the damage (e.g. is the damage localised to one tooth or is it potentially distributed over multiple teeth). The synchronous average of the squared envelope is estimated with

$$\mathrm{SASE}_x(\varphi; f, \Delta f) = \frac{1}{K} \sum_{k=0}^{K-1} |x_f(\varphi + k \cdot \Phi; f, \Delta f)|^2 \tag{16}$$

where Φ is the cyclic period of the component of interest in the angle domain, e.g. the angle period of a gear rotation. The normalised SASE is used in this work as basis for comparison. The normalised SASE is defined as follows

$$\mathrm{nSASE}_x(\varphi; f, \Delta f) = \frac{\mathrm{SASE}_x(\varphi; f, \Delta f)}{\mathrm{med}\{\mathrm{SASE}_x(\varphi; f, \Delta f)\}_\varphi} \tag{17}$$

where $\mathrm{med}\{x(t)\}_t$ is the median of the function $x(t)$ over the variable t. The median is used instead of the average to ensure that the denominator is less affected by the damage information in the SASE.

In the next section, the method will be evaluated on numerical data, whereafter the method will be investigated on experimental data in Section 4.

3. Numerical investigation

In this section, the performance of the proposed method is evaluated on numerical gearbox data and its performance is compared to conventional and feature ratio frequency band identification methods. To highlight the contributions of the proposed method, a signal is generated that has dominant impulsive components that are unrelated to the damage of interest. Dominant extraneous impulses can be encountered when the operating environment or other mechanical components generate impulsive components that are consistently present in the vibration signal. The impulsive components will manifest in the baseline signal and in the new signals from the machine in an unknown condition. The impulsive components can mask weak damage components in the signal and delay detection.

The model by Abboud et al. [49] contains distributed gear fault and bearing fault signatures. To demonstrate the ability of the proposed method to detect weak damage components despite the presence of dominant extraneous components, the distributed gear fault component will be used to model an extraneous impulsive component. The distributed gear fault component is therefore purposefully included in the reference dataset, i.e. we would like to enhance the bearing damage despite the presence of this extraneous component. The model is discussed in more detail in the next section.

3.1. Numerical gearbox model

A phenomenological gearbox model is used in this work to evaluate the performance of the proposed anomalous frequency band method. The phenomenological gearbox model is based on the work of Abboud et al. [49]. The vibration measurement of the gearbox in its reference condition is modelled as follows:

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + x_n(t) \tag{18}$$

The casing signal comprises of a gear mesh signal component $x_{gmc}(t)$, a distributed gear damage component $x_{dgd}(t)$ and a noise component $x_n(t)$. The data from Eq. (18) is used to generate hundred measurements that will act as historical data. The distributed gear damage component, which was proposed in Ref. [49] is used in this work to model dominant extraneous impulsive components. We specifically include the distributed gear damage component in the reference dataset to demonstrate the performance of this method when dominant cyclic components are present in the baseline signal. We would like to detect frequency bands that deviate from the reference signal given in Eq. (18).

The casing vibration signal with bearing damage is modelled as follows:

$$x_c(t) = x_{gmc}(t) + x_{dgd}(t) + FS_{blo} \cdot x_{blo}(t) + FS_{bli} \cdot x_{bli}(t) + x_n(t) \tag{19}$$

where $x_{blo}(t)$ is the outer race bearing damage component and $x_{bli}(t)$ is the inner race bearing damage component. The model in Eq. (19) is used to evaluate the ability of the method to detect anomalies due to inner race and outer race bearing damage, despite the presence of the gear damage component. The fault severity functions of the outer race and the inner race, denoted FS_{blo} and FS_{bli} respectively, are used to model either outer race bearing damage (i.e. $FS_{bli} = 0$, $FS_{blo} = 1$) or inner race bearing damage (i.e. $FS_{bli} = 1$, $FS_{blo} = 0$) in this work.

The performance of the proposed method will be evaluated on two signals from the model; a signal with outer race bearing damage (i.e. $FS_{bli} = 0$, $FS_{blo} = 1$) and a signal with inner race bearing damage (i.e. $FS_{bli} = 1$, $FS_{blo} = 0$). The equations and parameters of the signal components are included in Appendix A. The distributed gear damage component excites a frequency band centred at 2.0 kHz and has a cyclic order of 1.0 shaft orders. The outer race bearing damage component excites a frequency band centred at 6.0 kHz and has a cyclic order of 4.12 shaft orders. Lastly, the inner race bearing damage component excites a frequency band centred at 9.0 kHz and has a cyclic order of 5.88 shaft orders. The sampling frequency of the simulated signal is 40.0 kHz. The signal-to-noise ratio of the different components are summarised in Table A.3 in Appendix A.

The normalised SES of the raw signals are presented in Fig. 4 for the two considered cases. The squared envelope spectra are dominated by the gear components, which makes it difficult to observe the evidence of damage (e.g. the BPFO and its harmonics in Figs. 4(a) and 4(b) and the BPFI, its harmonics and its sidebands in Figs. 4(c) and 4(d)). The results obtained by the conventional, feature ratio and proposed methods are compared against the raw signal's results in the subsequent sections.

3.2. Conventional frequency band identification

The conventional frequency band identification method is applied on the outer race and inner race bearing damage data of the numerical gearbox model in this section.

Firstly, the outer race bearing damage case is considered. The feature planes of the three features, i.e. the Gini index, the Negentropy and the Hoyer index, are presented in Figs. 5(a)–5(c) respectively for the outer race bearing damage signal. It is evident that there are two dominant frequency bands in the feature planes. The distributed gear damage component manifest at 2.0 kHz, while the outer race bearing damage manifest at 6.0 kHz. Since the distributed gear damage component is more dominant than the outer race bearing damage component, it maximises the feature planes and is extracted for further analysis. These results are reflected in the squared envelope spectra shown in Figs. 5(d)–5(f) and the zoomed squared envelope spectra in Figs. 5(g)–5(i). The squared envelope spectra only contain the gear components attributed to the distributed gear damage at the first (1.0) shaft order and its harmonics and does not contain the bearing damage components at 4.12 shaft orders. The inner race bearing damage case is considered in Fig. 6 using the same procedure as the outer race bearing damage. The feature planes of the Gini index, the negentropy and the Hoyer index in Figs. 6(a)–6(c) are again dominated by the

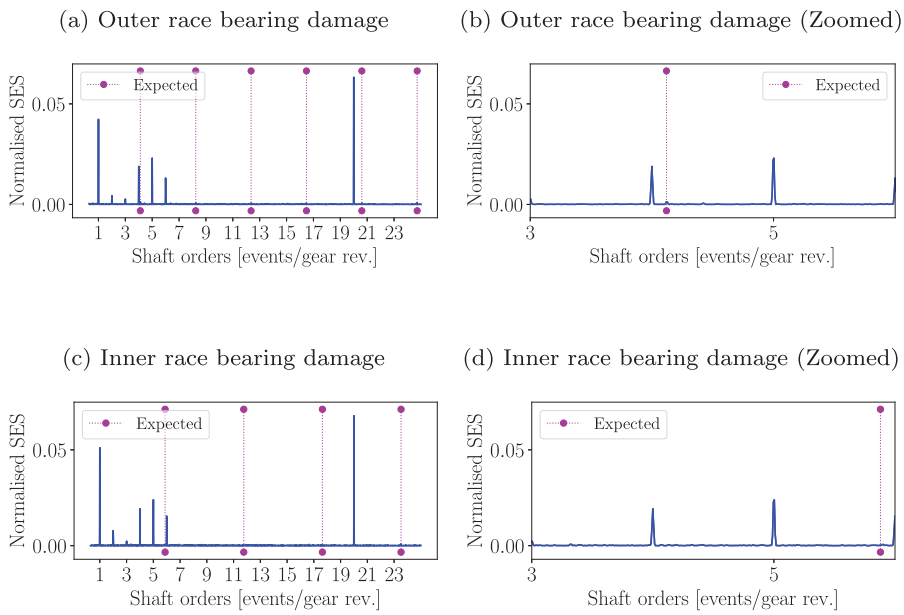

Fig. 4. The Squared Envelope Spectra (SES) of the two raw signals considered in this work.

dominant component and not by the inner race bearing damage component. Therefore, the dominant is extracted for demodulation as seen in the squared envelope spectra results in Figs. 6(d)–6(f); the first (1.0) shaft order component is very dominant, while the inner race bearing damage component at 5.88 shaft orders is missing. The zoomed views of the SES in Figs. 6(g)–6(i) highlight that the fundamental inner race bearing component is missing.

Therefore, the bearing damage is not detected using the conventional method with the Gini index, the Hoyer index, and the negentropy. The conventional methods perform as expected; the conventional methods select the most dominant non-stationary component in the signal (i.e. the distributed gear damage signal). However, the most dominant sparse components are not always related to the health of the machine under consideration (e.g. Ref. [36]). The proposed method is needed to detect large deviations from the baseline sparsity as opposed to the dominant sparse components itself.

3.3. Anomalous frequency band identification

The proposed method is applied with the procedure described in Section 2. Firstly, 100 historical measurements were generated using the model in Eq. (18), whereafter feature planes were extracted using the Gini index, the negentropy and the Hoyer index as described in Section 2.1. Thereafter, the feature planes of the historical measurements were used with the procedure described in Section 2.2 and maximised to find the anomalous frequency band with the procedure described in Section 2.3.

The proposed method is applied on the outer race bearing damage signal of the phenomenological bearing signal and the results are presented in Fig. 7. The feature planes of the Gini index, the negentropy and the Hoyer index in Figs. 7(a)–7(c) are dominated by the outer race bearing damage component as this is the novel component in the signal when comparing Eq. (19) to Eq. (18). This is reflected in the squared envelope spectra of the different methods; the square envelope spectra in Figs. 7(d)–7(f) contain dominant outer race bearing damage components at 4.12 shaft orders and its harmonics. When comparing the zoomed normalised squared envelope spectra in Figs. 7(g)–7(i) to 5(g)–5(i), we can conclude that it is possible to distinguish between the fourth harmonic of the distributed gear component (at 4.0 orders) and the outer race bearing damage component (at 4.12 orders). The damage is also significantly enhanced when compared against the raw signal's results in Figs. 4(a) and 4(b).

The proposed method is also applied on the inner race bearing damage signal and the results are presented in Fig. 8. The feature planes of the proposed method, using the Gini index, the negentropy and the Hoyer index, are shown in Figs. 8(a)–8(c). The feature planes are dominated by the inner race bearing damage components at a spectral frequency of 9.0 kHz. Therefore, the inner race bearing component is extracted and observed in the squared envelope spectra in Figs. 8(d)–8(f). The inner race bearing damage component of 5.88 shaft orders is dominant in the squared envelope spectra of the three methods. Since the inner race bearing damage rotates with the shaft, it will move in and out of the load zone with each shaft rotation. This causes the bearing impulses to be modulated by the shaft rotation and therefore the squared envelope spectra contain dominant sidebands spaced at 1.0 shaft orders. Figs. 8(g)–8(i) show the fundamental component at 5.88 shaft orders and two sidebands on the left of the fundamental component (i.e., at $5.88-1.0=4.88$ orders and $5.88-2.0=3.88$ orders). It is therefore possible to distinguish between the sixth distributed gear damage harmonic in Figs. 6(g)–6(i) and the first harmonic of the inner race bearing damage component in Figs. 8(g)–8(i).

This example was specifically used to demonstrate that the proposed method does not focus on the most dominant components, but the most anomalous components. Since the distributed damage component is present in the historical dataset used to establish the baseline, it means that we are already aware of this component and would like to detect deviations from this condition. Therefore, the dominant gear component is ignored and the weaker damage component can be extracted and characterised.

3.4. Feature ratio methods

The original SKRgram maximises the ratio of the kurtogram of a signal in an unknown condition to the kurtogram of a healthy signal. However, it is possible to generalise it to any feature, i.e.

$$\kappa(f,\Delta f)=\frac{\psi(f,\Delta f)}{\psi_n(f,\Delta f)} \tag{20}$$

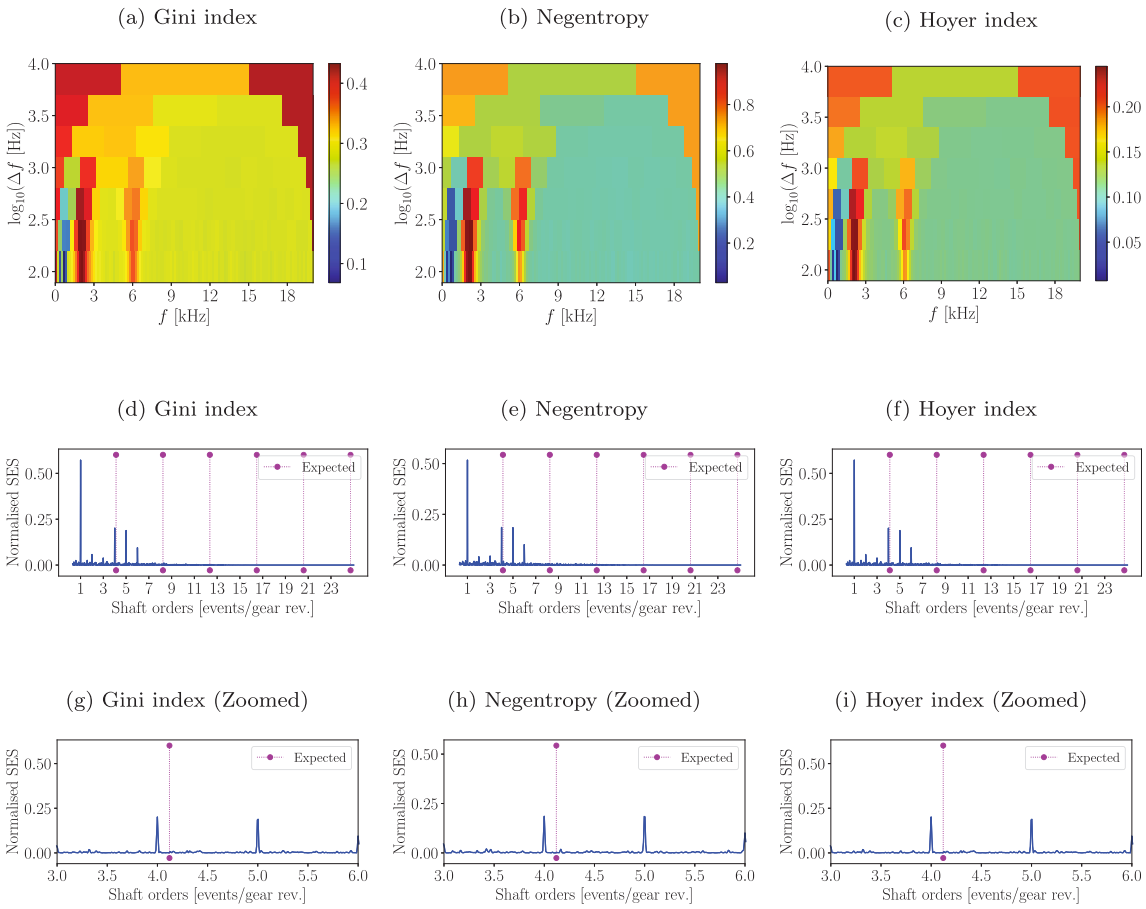

Fig. 5. The results of the conventional frequency band identification method are presented using the phenomenological bearing data with outer race bearing damage. Three features were used to obtain the feature planes in Fig. 5(a)–Fig. 5(c). The normalised Squared Envelope Spectrum (SES) of the bandpass filtered signals is obtained with Equation Eq. (15). The cyclic orders associated with the damage component is also superimposed on the SES and labelled as *Expected*.

where $\psi(f, \Delta f)$ is the feature calculated for the new signal and $\psi_n(f, \Delta f)$ denotes the feature of a measurement in the historical dataset. Therefore, the new feature plane, obtained with the feature ratio and denoted $\boldsymbol{K}$, can be maximised for demodulation. This feature ratio method has been applied for other features such as the Gini index [34], the negentropy [44], and the Teager energy entropy [43].

To apply the method, two aspects need to be addressed. Firstly, the denominator in the feature ratio, i.e. the reference feature in the denominator, i.e. $\psi_n(f, \Delta f)$, should not approach zero (0) to avoid numerical issues. Therefore, we did not subtract the constants from the features (e.g. subtract 2 from the kurtogram feature).

Secondly, a measurement should be selected from the historical dataset to obtain the feature in the denominator of Eq. (20). In this work, the following procedure is applied:

- Calculate $\psi_n(f, \Delta f)$ for all historical datasets to obtain a set of features for the frequency band $f, \Delta f$. The set of features is denoted $\{\psi_n^{(i)}(f, \Delta f)\}$.
- Apply Eq. (20) on all measurements to obtain a set of feature ratios. Thereafter, identify the healthy measurement n that maximises the feature ratio

$$\kappa(f, \Delta f) = \max_n \left\{ \frac{\psi(f, \Delta f)}{\psi_n(f, \Delta f)} \right\} \tag{21}$$

  to calculate the new feature plane denoted $\boldsymbol{K}$. The maximisation operator is used in Eq. (21) to ensure that feature plane is the most sensitive to anomalies.
- Maximise the feature plane $\boldsymbol{K}$ to find the optimal frequency band for demodulation.

The feature ratio method is applied on the outer race and inner race bearing damage signals of the phenomenological model using the same three features used in the previous investigations. The results of the outer race bearing damage is presented in Fig. 9. The feature planes of the three features, shown in Figs. 9(a)–9(c) are dominated by the outer race bearing damage. The distributed gear damage, that was part of the reference dataset, is absent in the feature planes.

The associated squared envelope spectra in Figs. 9(d)–9(f) corroborate the observations of the feature planes; the frequency ratio method is able to extract the bearing damage and therefore the squared envelope spectrum is dominated by the outer race bearing damage components. If the squared envelope spectra of the proposed method and the feature ratio method are compared, it is observed that the harmonics are low passed filtered in the feature ratio and therefore higher harmonics are not as pronounced.

The feature ratio methods are applied on the inner race bearing dataset and the results are presented in Fig. 10. The feature planes in Figs. 10(a)–10(c) are dominated by the inner race bearing damage component at 9.0 kHz. Therefore, the squared envelope spectra of the bandpass filtered signal shown in Figs. 10(d)–10(f) contains evidence of the damaged component.

When comparing the raw signals' results in Fig. 4, the conventional method's results in Figs. 5 and 6, the proposed method's results in

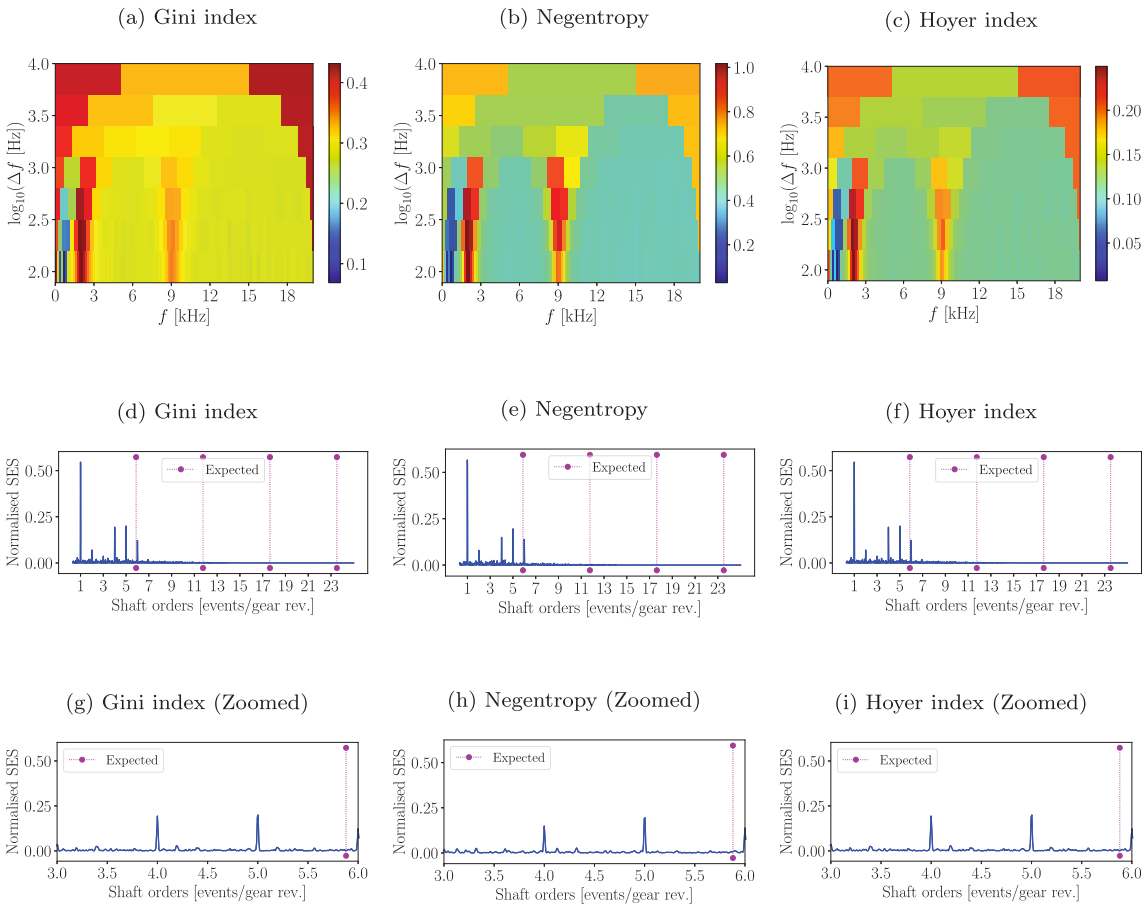

Fig. 6. The results of the conventional frequency band identification method are presented using the phenomenological bearing data with inner race bearing damage. Three features were separately applied and their respective feature planes and normalised Squared Envelope Spectra (SES) shown in this figure.

Figs. 7 and 8 and the frequency ratio method's results in Figs. 9 and 10, we can conclude that it is beneficial to incorporate historical data in the frequency band identification process if the features are blind, i.e. if a specific component is not targeted. Even though this investigation highlighted the benefits of incorporating historical data in the frequency band identification method (i.e. both the proposed and the feature ratio methods detected the weak damage component that was not present in the historical data), the benefits of the proposed method over the feature ratio methods still need to be demonstrated. In the next section, the proposed method is compared to the conventional and feature ratio frequency band identification methods on experimental data.

4. Experimental investigation

In this section, the proposed method will be compared against the conventional and the feature ratio frequency band identification methods on experimental data that were acquired under time-varying operating conditions. In the next section, the test-rig will be presented.

4.1. Experimental test-rig

The experimental test-rig is presented in Fig. 11. The test-rig contains three helical gearboxes, an alternator and an electrical motor. The electrical motor drives the system and the alternator dissipates the rotational energy from the system. The centre of the three gearboxes, referred to as the monitored gearbox in Fig. 11, is instrumented with sensors and its gear (as opposed to pinion) is damaged.

The gearbox was monitored for damage using a 100 mV/g tri-axial accelerometer. The axial component is monitored due to the dominant axial excitations of the helical gearboxes. The rotational speed of the system is measured using an optical probe and a zebra tape shaft encoder. The geometrical imperfections of the zebra type shaft encoder are corrected using the Bayesian geometry compensation method proposed by Diamond et al. [50]. The accelerometer was sampled at 25.6 kHz while the zebra tape shaft encoder was sampled at 51.2 kHz.

In the monitored gearbox, the pinion is healthy, while the gear is damaged during the degradation experiments. The gear is fixed to the reference shaft (i.e. the shaft with the zebra tape shaft encoder) and therefore the modulation due to a damage tooth will have a periodicity of 1.0 shaft orders.

Measurements were acquired from the healthy gearbox, whereafter the gearbox was disassembled, the gear was removed and damaged, and the gearbox was assembled and aligned. The dataset was acquired from an accelerated fatigue test of the gear tooth shown in Fig. 12. Damage was seeded in the gear tooth with an electric discharge machine and the gear tooth before the test is shown in Fig. 12(a). The gear was inserted in the gearbox and operated with the damaged tooth until it failed. The gear after the completion of the experiment is shown in Fig. 12(b).

The time-varying operating conditions for all measurements considered in this work are presented in Fig. 13. The rotational speed of the electrical motor and the applied load of the alternator were separately controlled.

4.2. Results

The result section is structured as follows: Firstly, the results using only the raw signal is presented in Section 4.2.1 to establish a baseline. Thereafter, the results of the conventional method, the feature ratio method and the proposed method are compared in Section 4.2.2.

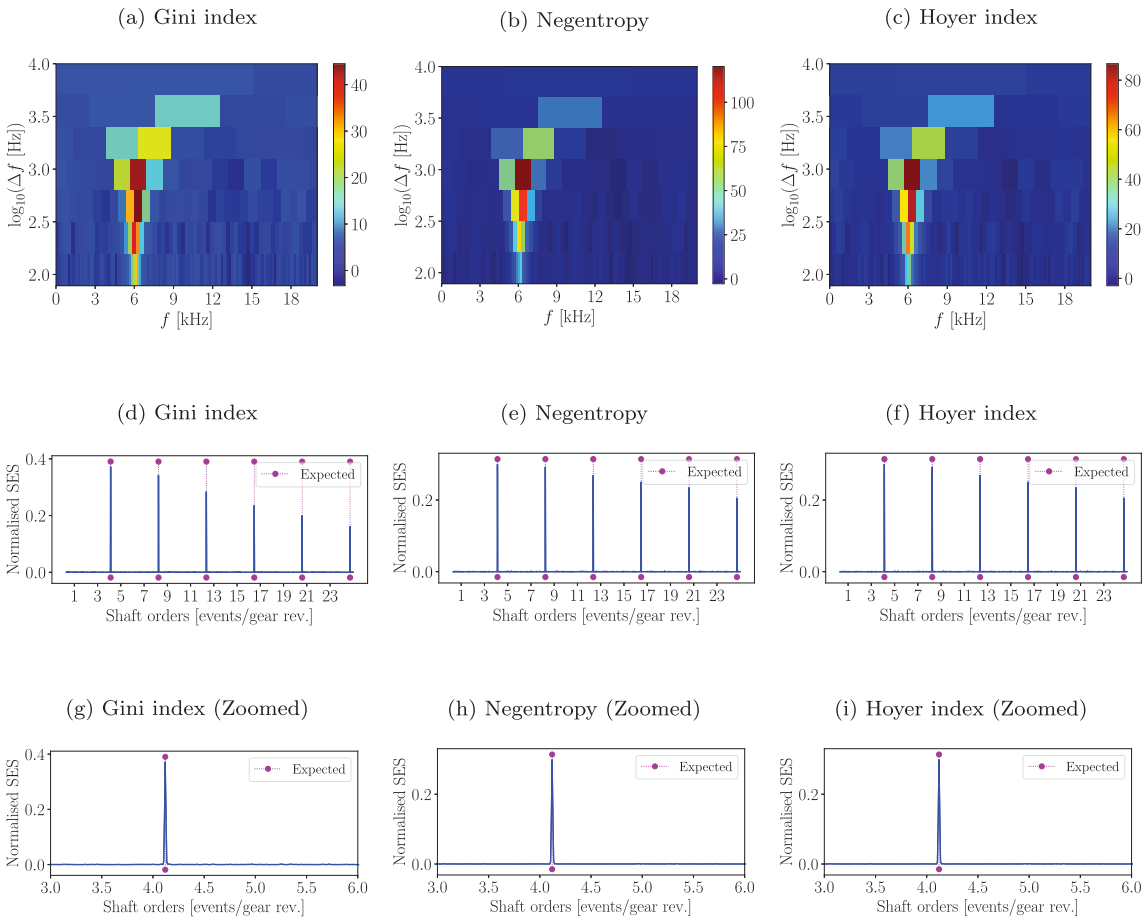

Fig. 7. The results of the proposed anomalous frequency band identification method are presented using the phenomenological bearing data with outer race bearing damage. Three features were separately applied and their respective feature planes and normalised Squared Envelope Spectra (SES) are shown in this figure.

4.2.1. *Raw signal results*

Four measurements of this dataset are considered in this work. Measurements one to four were acquired after 23%, 45% and 65% and 85% of the experiment were completed respectively. The normalised Synchronous Average of the Squared Envelope (SASE) and the normalised Squared Envelope Spectra (SES) are presented in Fig. 14 for the four measurements. The gear damage, which is located at 135 degrees in the normalised SASE, is not easily seen in Fig. 14. This is corroborated by the squared envelope spectra in Fig. 14; the SES does not contain clear evidence of the damaged components at 1 shaft order and its harmonics.

4.2.2. *Results of the different methods*

The feature planes, the squared envelope spectra and the synchronous average of the squared envelopes of the conventional, the anomalous and the feature ratio frequency band identification methods are compared in this section using the fourth measurement (i.e. the case where the damage was the most pronounced). The conventional frequency band identification method was applied on measurement four of the considered dataset and the results are presented in Fig. 15. The feature planes of the Gini index, the negentropy and the Hoyer index are shown in Figs. 15(a)–15(c). The feature planes are dominated by frequency bands in the higher spectral frequencies. The squared envelope spectra associated with the bandpass filtered signals of the feature planes are presented in Figs. 15(d)–15(f). The squared envelope spectra show evidence of dominant cyclic components at 5.72 shaft orders and its harmonics. These frequency components were even present in the data from the healthy gearbox and are seen as extraneous when monitoring the health of the gears. The synchronous average of the squared envelope signals shown in Figs. 15(g)–15(i) do not contain any useful information related to the health of the gears and therefore the gear damage is not detected in these measurements.

The proposed anomalous frequency band identification method is now considered. Sixty (60) measurements were used during the training phase of the method. Firstly, the parameters of the proposed method were obtained using the feature planes of 60 healthy measurements, whereafter the method was applied with the procedure described in Section 2.2 to obtain the anomalous feature plane. The results are shown in Fig. 16. More specifically, the feature planes are presented in Figs. 16(a)–16(c) for the Gini index, the negentropy and Hoyer index respectively. The three feature planes are dominated in the same spectral frequency region.

The squared envelope spectra of the bandlimited signals obtained by following the procedure described in Section 2.3 are presented in Figs. 16(d)–16(f). The squared envelope spectra show evidence of strong modulation at the gear shaft rotation. The source of the modulation is clear when observing the synchronous average of the squared envelope spectra of the gear shown in Figs. 16(g)–16(i). The damaged gear tooth results in modulation that is periodic with the rotation of the gear shaft. Therefore, by using the synchronous average of the squared envelope it is possible to determine that a small localised region (e.g. one tooth) of the gear is damaged.

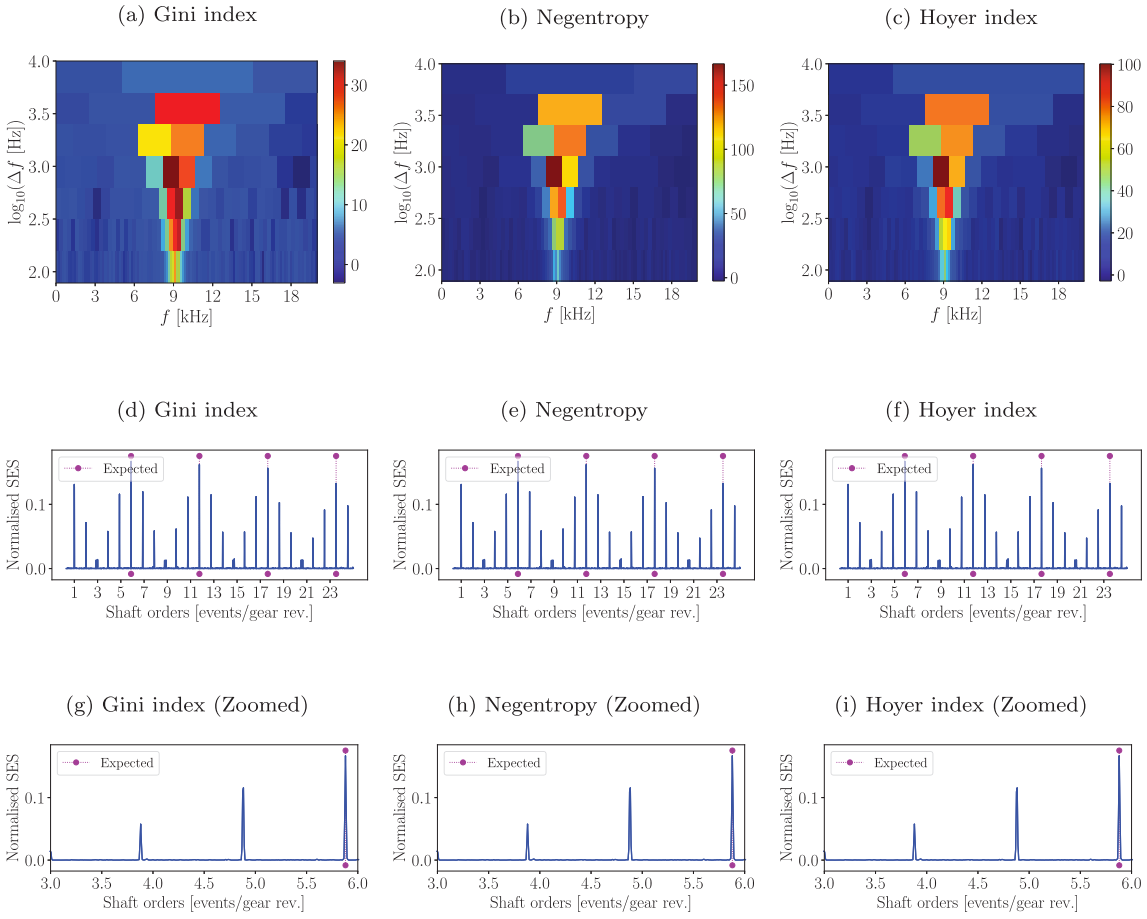

Fig. 8. The results of the proposed anomalous frequency band identification method are presented using the phenomenological bearing data with inner race bearing damage. Three features were separately applied and their respective feature planes and normalised Squared Envelope Spectra (SES) are shown in this figure.

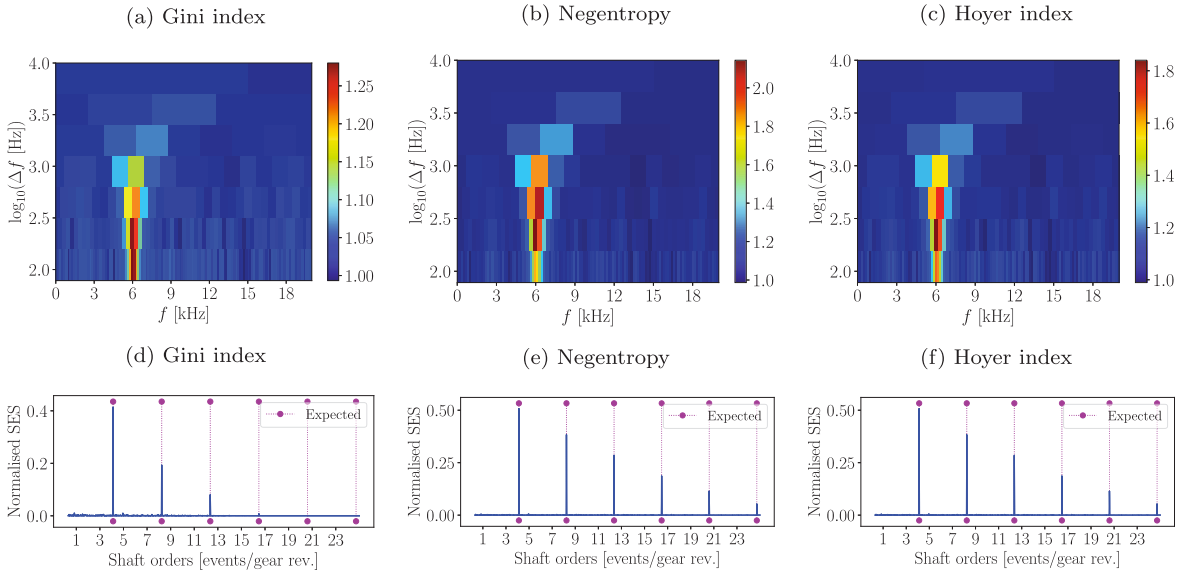

Fig. 9. The results of the feature ratio frequency band identification method are presented using the phenomenological bearing data with outer race bearing damage. Three features were separately applied and their respective feature planes and normalised Squared Envelope Spectra (SES) are shown in this figure.

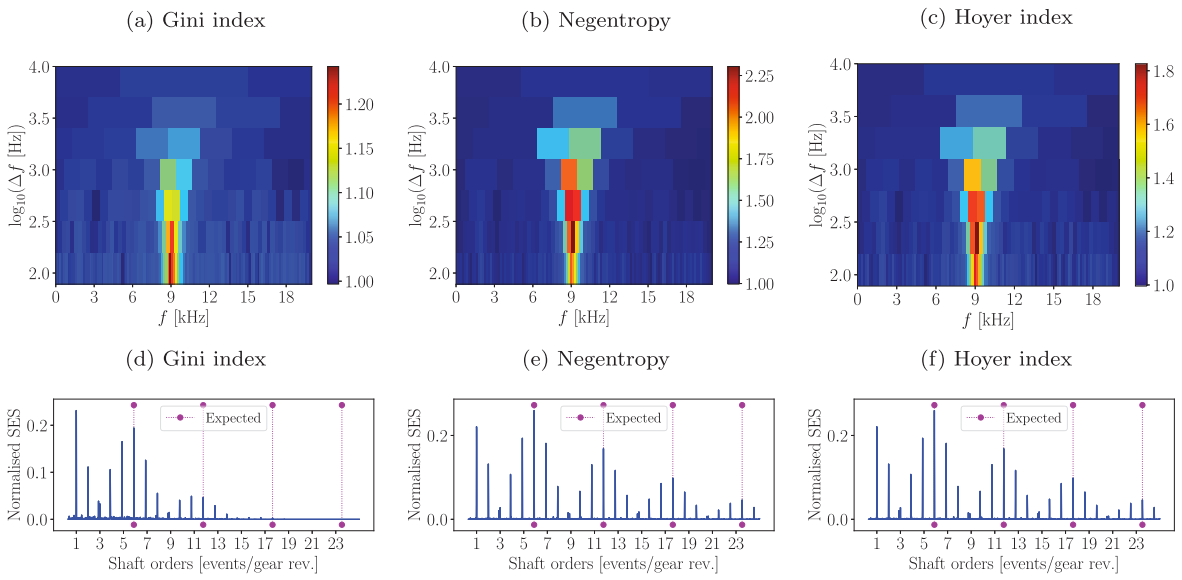

Fig. 10. The results of the feature ratio frequency band identification method are presented using the phenomenological bearing data with inner race bearing damage. Three features were separately applied and their respective feature planes and normalised Squared Envelope Spectra (SES) are shown in this figure.

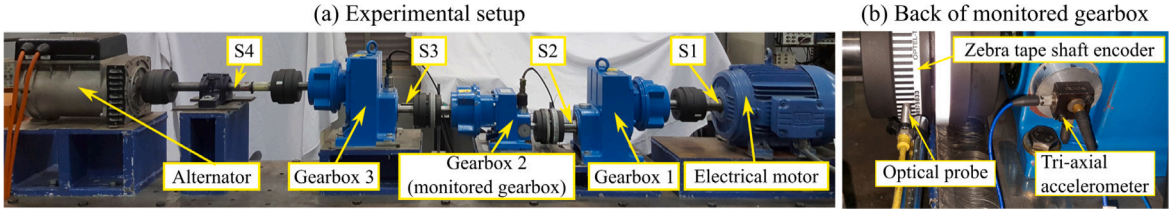

Fig. 11. The experimental test-rig is presented with the important components highlighted in (a). In (b), the input shaft is shown with the zebra tape shaft encoder and the tri-axial accelerometer used in this work.

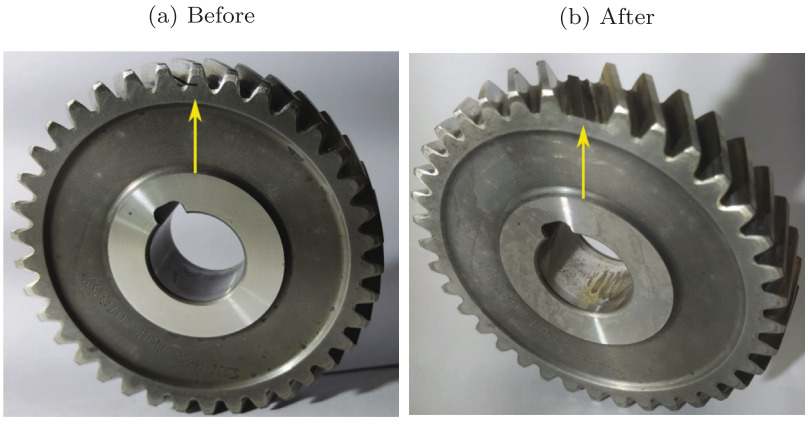

Fig. 12. The damaged gear before the start of the experiment and after the completion of the experiment are shown.

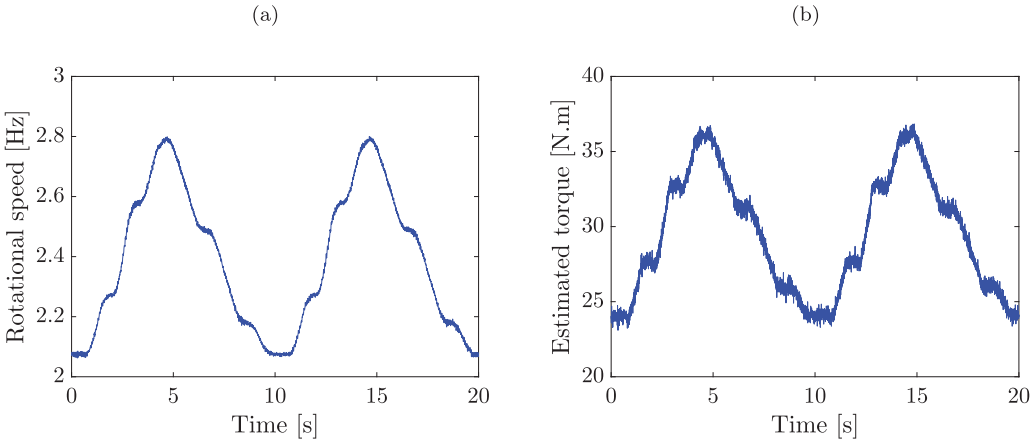

Fig. 13. The operating conditions at the input shaft of the monitored gearbox are presented.

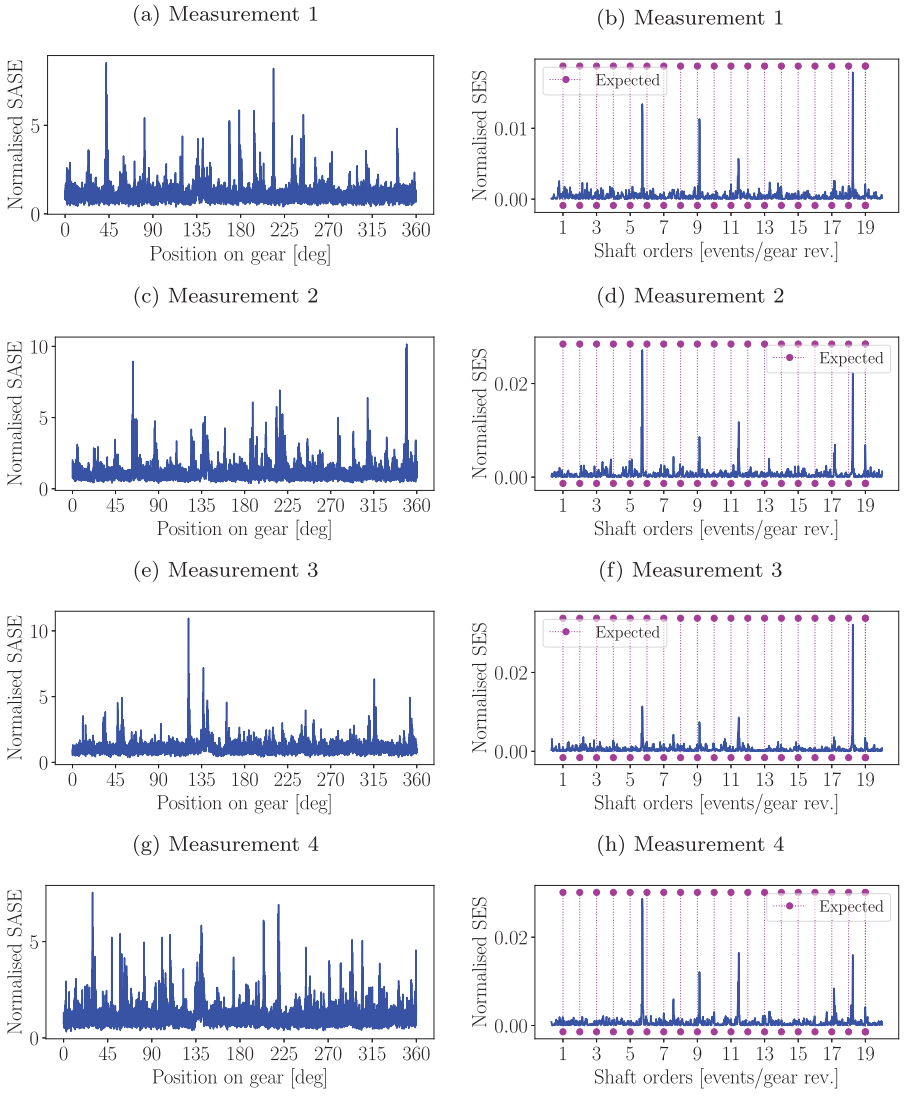

Fig. 14. The results of the raw signals for the four measurements, i.e. without applying the conventional, feature ratio and proposed methods.

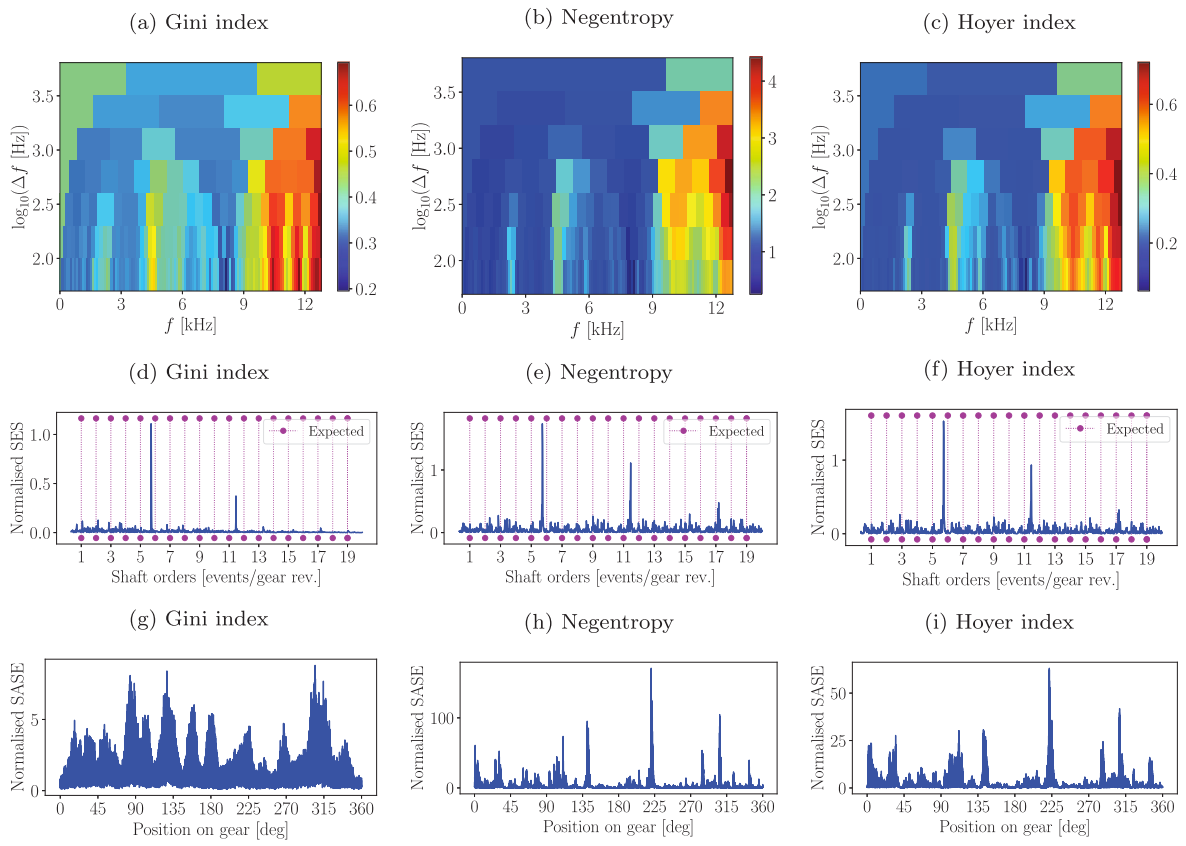

Fig. 15. The results of the conventional frequency band identification method applied to the fourth measurement of the experimental gearbox data are presented. The feature plane, the normalised Squared Envelope Spectrum (SES) and the normalised Synchronous Average of the Squared Envelope (SASE) are presented for three features.

The feature ratio methods are now applied with the procedure described in Section 3.4. The same measurements as the proposed anomalous frequency band identification method was used for training. The feature ratio method is applied on measurement four and shown in Fig. 17. The feature planes using the Gini index, the negentropy and the Hoyer index are presented in Figs. 17(a)–17(c) respectively. Even though the negentropy and the Hoyer index show some evidence of the gear damage in the lower frequency bands, the feature planes are maximised by a frequency band between 6 and 9 kHz. The squared envelope spectra shown in Figs. 17(d)–17(f) and the synchronous average of the squared envelope spectra shown in Figs. 17(g)–17(i) do not shown clear evidence of the gear damage.

To further understand the difference between the proposed method and the feature ratio methods' performance, the standard deviation in the frequency bands are quantified, with the results presented in Fig. 18. The standard deviation of the different frequency bands of the healthy measurements are presented in Figs. 18(a)–18(c) for the Gini index, the negentropy and the Hoyer index. According to these results, the higher frequency bands (6 kHz and higher) have larger variations between the measurements than the lower frequency bands. If this variance information is ignored, the frequency bands might be erroneously detected as informative frequency bands. While the feature ratio methods do not account for the variation in the frequency bands, the proposed method uses this variance to calculate the deviation from the reference condition. This highlights the benefits of the proposed method.

Furthermore, to compare the performance of the three methods on the additional measurements, the normalised synchronous average of the squared envelope applied to the bandpass filtered signals of Measurement 1–4 are presented in Fig. 19 for the three features.

The results further demonstrate the improved performance of the proposed method; the proposed method is able to identify the frequency band associated with the gear damage and therefore the gear damage is visible in the SASE. The conventional frequency band identification methods and the feature ratio methods are incapable of identifying the gear damage components.

Table 1
The computational time that was required for the numerical and experimental signals are presented in seconds using 50 measurements. A computer with the following properties were used: Processor Intel(R) Core(TM) i7-9750H CPU 2.60 GHz, 2592 MHz, 6 Cores, 12 Logical Processors. 32 GB RAM. The method was implemented in python using SciPy 1.10.0. [51] and NumPy 1.23.5 [52].

	Conv. feature plane decomposition (per signal)	Anomalous plane (training - Fig. 1(b)) (using 50 signals)	Anomalous plane (application - Fig. 1(c)) (per signal)
Numerical	4.249	212.468	4.269
Experimental	4.714	235.724	4.815

The computational time of the proposed method is summarised in Table 1 for the numerical and the experimental datasets. The table shows the conventional feature plane decomposition, present in all steps in Fig. 1, the completion of the training step in Fig. 1(b) using 50 signals and the application of the method on a new signal in Fig. 1(c). The most time-consuming part is the feature plane decomposition, which needs to occur for each signal in the historical dataset and for each new measurement that is evaluated. Fortunately, the historical data only need to be processed during the training phase of the method. There are also faster decomposition methods available (e.g. [15]) that would improve the computational time of the method, however, this is not in the scope of this work. The benefits of the proposed method justify the additional time associated with anomalous feature plane calculation.

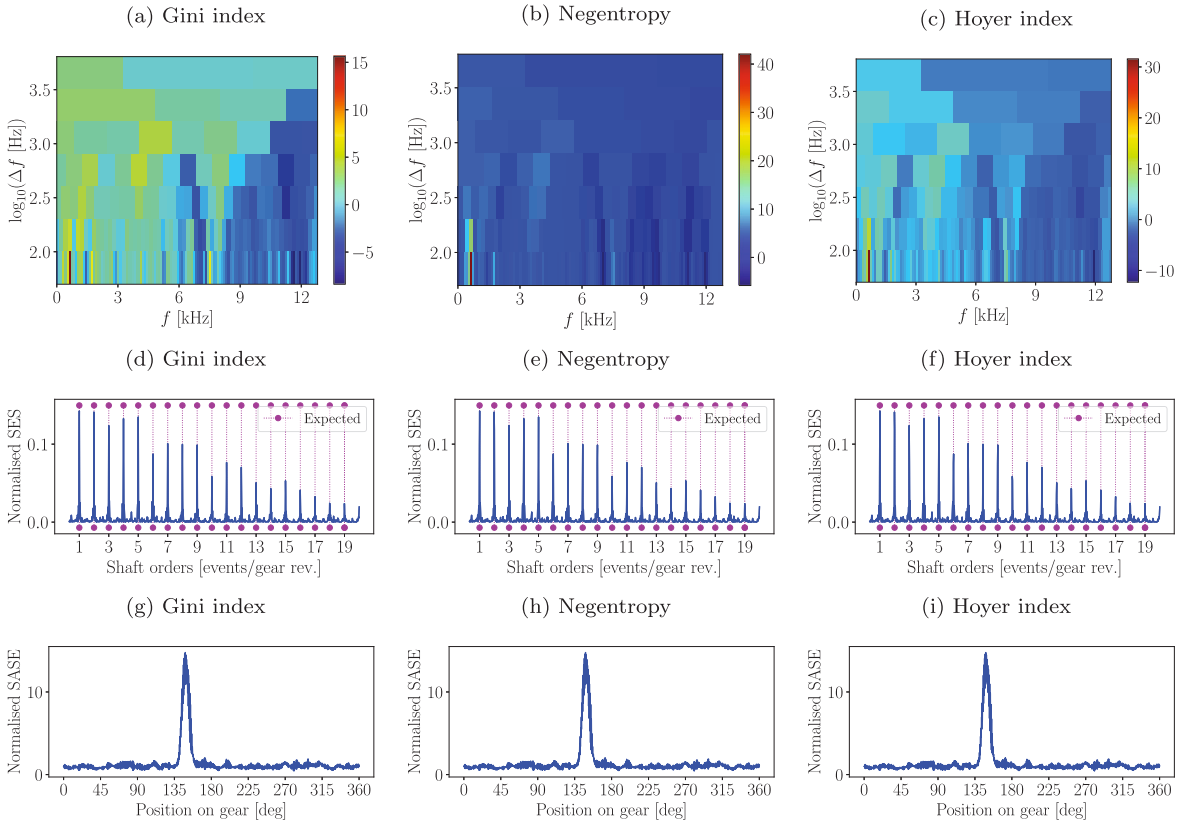

Fig. 16. The results of the proposed anomalous frequency band identification method applied to the fourth measurement of the experimental gearbox data are presented. The feature plane, the normalised Squared Envelope Spectrum (SES) and the normalised Synchronous Average of the Squared Envelope (SASE) are presented for three features.

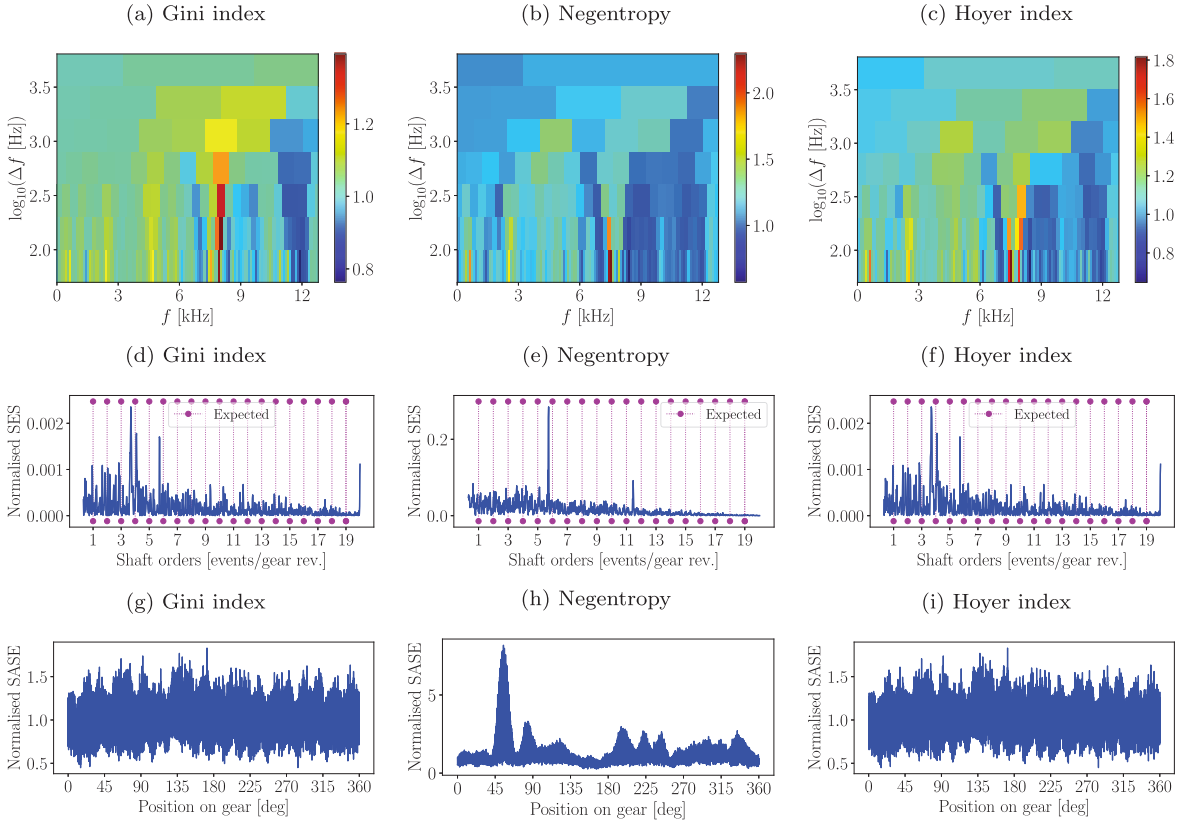

Fig. 17. The results of the frequency ratio frequency band identification method applied to the fourth measurement of the experimental gearbox data are presented. The feature plane, the normalised Squared Envelope Spectrum (SES) and the normalised Synchronous Average of the Squared Envelope (SASE) are presented for three features.

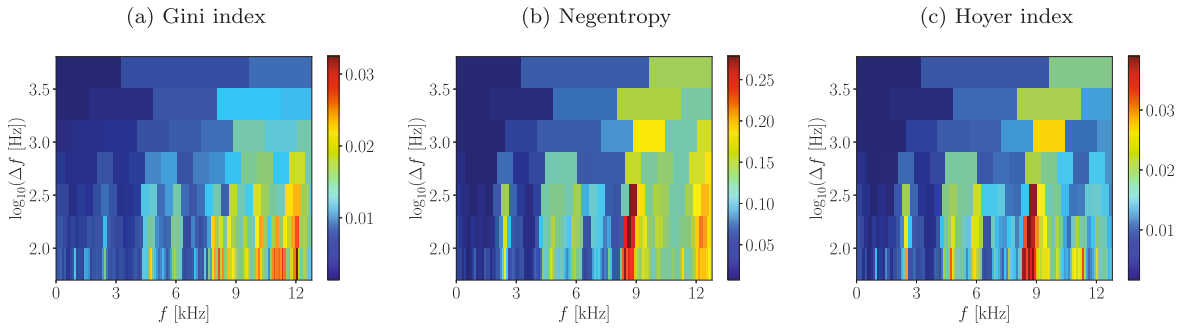

Fig. 18. The standard deviation of each frequency band of the healthy measurement dataset is presented for the three features.

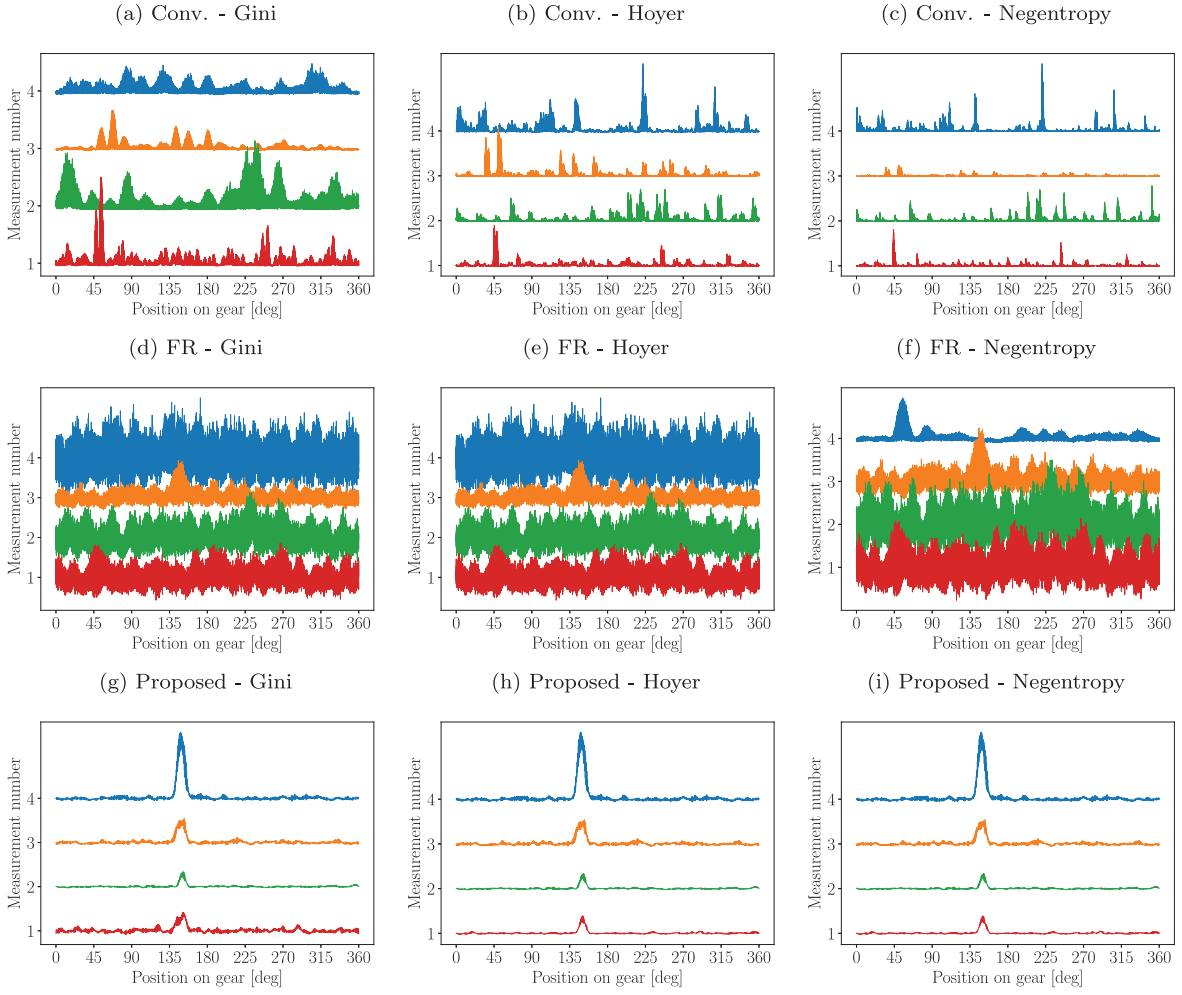

Fig. 19. The normalised Synchronous Average of the Squared Envelope (SASE) of the bandlimited signals obtained with the three features used in the three methods, namely, the conventional method, the Feature Ratio (FR) method and the proposed method. The results are presented for four measurements of the experimental gearbox dataset. The normalised SASE is obtained by applying Eq. (17) to the bandlimited signals.

5. Conclusion

A frequency band identification method is proposed in this work to identify the frequency band with the most anomalous behaviour. The method is compared to conventional frequency band identification methods and feature ratio frequency band identification methods. In numerical and experimental investigations, it is shown that the proposed method is better suited for detecting impulsive components that were not present in the reference dataset. Instead of using a Gaussian stationary signal as a reference, the proposed method uses the healthy historical data as a reference. Therefore, deviations from the healthy baseline is used for frequency band identification, which reduces the influence of extraneous signal components in the frequency band identification process and it makes the frequency band identification more sensitive to anomalous components. These results were demonstrated on the Gini index, the negentropy and the Hoyer index. Furthermore, the proposed method incorporates all of the available healthy historical data in the estimation of the anomalous frequency band, which results in an improved performance over feature ratio methods.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Stephan Schmidt gratefully acknowledges the research development programme of the University of Pretoria, South Africa for supporting the research and the VLIR-UOS Global Minds programme at KU Leuven for funding the research visit to the Department of Mechanical Engineering, KU Leuven, Belgium.

Appendix. Phenomenological model parameters

In this section, more detail on the phenomenological model is presented. The model is based on the work by Abboud et al. [49]. The measurement model given by Eq. (19) consists of a gear mesh component, a distributed gearbox component, an outer race bearing damage component, an inner race bearing damage component and a noise component.

The gear mesh component of the model is given by

$$x_{gmc}(t) = M\left(\omega(t)\right) \cdot h_{gmc}(t) \otimes \left(\sum_{k=1}^{N_{gmc}} A_{gmc}^{(i)} \cdot \sin\left(k \cdot N_{teeth} \cdot \int_0^t \omega(\tau) d\tau + \varphi_{gmc}^{(k)} \right) \right), \tag{A.1}$$

where $M\left(\omega(t)\right)$ scales the signal's amplitude according to rotational speed $\omega(t)$, $h_{gmc}(t)$ is the impulse response function of a single degree of freedom system; $A_{gmc}^{(i)}$ and $\varphi_{gmc}^{(k)}$ are the amplitude and phase of the kth harmonic, and N_{teeth} is the number of teeth on the gear. The convolution operator is denoted $\otimes$. N_{gmc} components are included in the signal. In this work, the model is applied under constant speed operating conditions, therefore, $M\left(\omega(t)\right) = \omega^2$ is a constant for all components. The following parameters are used: $A_{gmc} = [1, 2, 3]$, $\varphi_{gmc} = [0, 0, 0]$, $N_{gmc} = 3$, and $N_{teeth} = 20$. The parameters of the impulse response function $h_{gmc}(t)$ are given in Table A.2.

The distributed gear damage component

$$x_{dgd}(t) = M(\omega(t)) \cdot h_{dgd} \otimes \left(\varepsilon_\sigma(t) \cdot \sum_{k=1}^{N_{dgd}} A_{dgd}^{(k)} \cdot \sin\left(k \cdot \int_0^t \omega(\tau) d\tau + \varphi_{dgd}^{(k)} \right) \right), \tag{A.2}$$

Table A.2
The impulse response function parameters of the different components. The impulse response function is modelled as a single degree-of-freedom system and therefore two parameters need to be specified, namely, the damping ratio and the natural frequency.

	Damping ratio	Natural frequency [Hz]
Outer race bearing damage component	0.05	6000
Inner race bearing damage component	0.05	9000
Distributed gearbox damage component	0.05	2000
Deterministic gear component	0.05	500

has a similar form of the gear mesh component, except for the addition of a random variable $\varepsilon_\sigma(t)$ attributed to the distributed gear damage. The random variable $\varepsilon_\sigma(t)$ is a standardised Gaussian variable in this work. The parameters are as follows: $A_{dgd} = [1, 2, 3]$, $\varphi_{dgd} = [0, 0, 0]$, and $N_{dgd} = 3$. The parameters of the impulse response function $h_{dgd}(t)$ are given in Table A.2.

The stationary noise component

$$x_n(t) = M(\omega(t)) \cdot \sigma_n \cdot \epsilon_\sigma(t), \tag{A.3}$$

consists of the amplitude modification function $M(\omega(t))$, a standardised Gaussian random variable $\epsilon_\sigma(t)$ and the standard deviation of the noise σ_n.

The outer race bearing damage component

$$x_{blo}(t) = M(\omega(t)) \cdot h_{blo}(t) \otimes \left(\sum_{k=0}^{N_{imp}-1} A_{blo}^{(k)} \cdot \delta\left(t - T_{blo}^{(k)} \right) \right), \tag{A.4}$$

contains N_{imp} Dirac-impulses, with the kth impulse denoted by $\delta\left(t - T_{blo}^{(k)}\right)$. Each impulse is scaled with a random variable $A_{blo}^{(k)}$ and the time of arrival of the bearing impulse T_{blo} depends on the bearing outer race characteristic order, the rotational speed and slip. The impulses are filtered through the structure to the sensor and this transmission path is modelled as $h_{blo}(t)$. The magnitude of the impulses is also scaled with the amplitude modification function $M(\omega(t))$. The ball pass order of the outer race component is 4.12.

The inner race bearing damage component

$$x_{bli}(t) = z_{stribeck}\left(\int_0^t \omega(\tau) d\tau \right) \cdot M(\omega(t)) \cdot h_{bli} \otimes \left(\sum_{k=0}^{N_{imp}} A_{bli}^{(k)} \cdot \delta\left(t - T_{bli}^{(k)} \right) \right), \tag{A.5}$$

has a similar form to the outer race bearing damage, except for the Stribeck function $z_{stribeck}(\cdot)$. Since the inner race rotates with the shaft and the direction of the applied load is constant, the inner race bearing damage moves in-and-out of the load zone for each shaft rotation. This results in additional amplitude modulation, where the impacts in the load zone are larger than the impacts out of the load zone. The Stribeck equation [53]

$$z_{stribeck}(\varphi) = \begin{cases} z_0 \cdot \left(1 - \frac{1}{2\epsilon_{str}} \left(1 - \cos(\varphi) \right) \right)^{c_{str}} & \text{for } |\text{wrp}(\varphi)| < \varphi_{max} \\ 0 & \text{otherwise,} \end{cases} \tag{A.6}$$

models this modulation as the bearing damage moves in and through the load zone, where the constants used in this work are $\epsilon_{str} = 0.49$, $\varphi_{max} = 0.99 \cdot \pi/2$, $c_{str} = 1.5$, $z_0 = 1$ [53]. The $|\text{wrp}(\varphi)|$ constrains the shaft angle $-\pi \leq \varphi < \pi$.

The ball pass order of the outer race component and the ball pass order of the inner race bearing component are 4.12 and 5.88 shaft orders respectively. The impulse response function parameters are given in Table A.2. The SNR of the signals are included in Table A.3.

Table A.3
The variance of the different signal components and signal-to-noise ratio of the bearing signals for the two cases. The Signal-to-Noise Ratio (SNR) is calculated as follows: $10 \cdot \log_{10}\left(\frac{\sigma_i^2}{\sigma_n^2}\right)$ where σ_n^2 denotes the variance of the noise signal x_n and σ_i^2 denotes the variance of the either the outer race bearing component x_{blo} or variance of the inner race bearing damage component x_{bli}.

	var$\{x_{gmc}\}$	var$\{x_{dgd}\}$	var$\{x_{blo}(t)\}$	var$\{x_{bli}(t)\}$	var$\{x_n\}$	SNR$_{blo}$	SNR$_{bli}$
Outer race	1.040	0.667	0.061	0.000	1.006	−12.1	−inf
Inner race	1.040	0.644	0.000	0.049	1.003	−inf	−13.1

References

[1] Z. Liu, L. Zhang, A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings, Measurement 149 (2020) 107002.

[2] W. Bartelmus, R. Zimroz, A new feature for monitoring the condition of gearboxes in non-stationary operating conditions, Mech. Syst. Signal Process. 23 (5) (2009) 1528–1534.

[3] Z. Wang, C. Liu, Wind turbine condition monitoring based on a novel multivariate state estimation technique, Measurement 168 (2021) 108388.

[4] A. Kumar, C. Gandhi, Y. Zhou, R. Kumar, J. Xiang, Latest developments in gear defect diagnosis and prognosis: A review, Measurement 158 (2020) 107735.

[5] A. Jablonski, Z. Dworakowski, K. Dziedziech, F. Chaari, Vibration-based diagnostics of epicyclic gearboxes–From classical to soft-computing methods, Measurement 147 (2019) 106811.

[6] T. Wang, Q. Han, F. Chu, Z. Feng, Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review, Mech. Syst. Signal Process. 126 (2019) 662–685.

[7] B. Zhang, Y. Miao, J. Lin, H. Li, Weighted envelope spectrum based on the spectral coherence for bearing diagnosis, ISA Trans. 123 (2022) 398–412.

[8] W.A. Smith, Z. Fan, Z. Peng, H. Li, R.B. Randall, Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference, Mech. Syst. Signal Process. 75 (2016) 371–394.

[9] S. Schmidt, R. Zimroz, P.S. Heyns, Enhancing gearbox vibration signals under time-varying operating conditions by combining a whitening procedure and a synchronous processing method, Mech. Syst. Signal Process. 156 (2021) 107668.

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

[11] W.A. Smith, P. Borghesani, Q. Ni, K. Wang, Z. Peng, Optimal demodulation-band selection for envelope-based diagnostics: A comparative study of traditional and novel tools, Mech. Syst. Signal Process. 134 (2019) 106303.

[12] S. Schmidt, P.S. Heyns, K.C. Gryllias, An informative frequency band identification framework for gearbox fault diagnosis under time-varying operating conditions, Mech. Syst. Signal Process. 158 (2021) 107771.

[13] M.A. Ugwiri, M. Carratú, V. Paciello, C. Liguori, Benefits of enhanced techniques combining negentropy, spectral correlation and kurtogram for bearing fault diagnosis, Measurement 185 (2021) 110013.

[14] J. Antoni, The spectral kurtosis: a useful tool for characterising non-stationary signals, Mech. Syst. Signal Process. 20 (2) (2006) 282–307.

[15] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mech. Syst. Signal Process. 21 (1) (2007) 108–124.

[16] D. Wang, W.T. Peter, K.L. Tsui, An enhanced Kurtogram method for fault diagnosis of rolling element bearings, Mech. Syst. Signal Process. 35 (1–2) (2013) 176–199.

[17] Y. Lei, J. Lin, Z. He, Y. Zi, Application of an improved kurtogram method for fault diagnosis of rolling element bearings, Mech. Syst. Signal Process. 25 (5) (2011) 1738–1749.

[18] T. Barszcz, A. Jabłoński, A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram, Mech. Syst. Signal Process. 25 (1) (2011) 431–451.

[19] W.T. Peter, D. Wang, The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement–Parts 1 and 2", Mech. Syst. Signal Process. 40 (2) (2013) 499–519.

[20] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, Mech. Syst. Signal Process. 74 (2016) 73–94.

[21] A. Moshrefzadeh, A. Fasana, The autogram: An effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis, Mech. Syst. Signal Process. 105 (2018) 294–318.

[22] Z. Zheng, X. Li, Y. Zhu, Feature extraction of the hydraulic pump fault based on improved autogram, Measurement 163 (2020) 107908.

[23] X. Wang, J. Zheng, H. Pan, Q. Liu, C. Wang, Maximum envelope-based autogram and symplectic geometry mode decomposition based gear fault diagnosis method, Measurement 174 (2021) 108575.

[24] Q. Ni, K. Wang, J. Zheng, Rolling element bearings fault diagnosis based on a novel optimal frequency band selection scheme, IEEE Access 7 (2019) 80748–80766.

[25] K. Zhang, Y. Xu, Z. Liao, L. Song, P. Chen, A novel fast entrogram and its applications in rolling bearing fault diagnosis, Mech. Syst. Signal Process. 154 (2021) 107582.

[26] A. Mauricio, W.A. Smith, R.B. Randall, J. Antoni, K. Gryllias, Improved envelope spectrum via feature optimisation-gram (IESFOgram): A novel tool for rolling element bearing diagnostics under non-stationary operating conditions, Mech. Syst. Signal Process. 144 (2020) 106891.

[27] B. Chen, W. Zhang, J.X. Gu, D. Song, Y. Cheng, Z. Zhou, F. Gu, A.D. Ball, Product envelope spectrum optimization-gram: An enhanced envelope analysis for rolling bearing fault diagnosis, Mech. Syst. Signal Process. 193 (2023) 110270.

[28] J. Hebda-Sobkowicz, R. Zimroz, M. Pitera, A. Wyłomańska, Informative frequency band selection in the presence of non-Gaussian noise–a novel approach based on the conditional variance statistic with application to bearing fault diagnosis, Mech. Syst. Signal Process. 145 (2020) 106971.

[29] J. Hebda-Sobkowicz, R. Zimroz, A. Wyłomańska, Selection of the informative frequency band in a bearing fault diagnosis in the presence of non-Gaussian Noise—Comparison of recently developed methods, Appl. Sci. 10 (8) (2020) 2657.

[30] M. Ahsan, D. Bismor, Early-stage fault diagnosis for rotating element bearing using improved harmony search algorithm with different fitness functions, IEEE Trans. Instrum. Meas. 71 (2022) 1–9.

[31] D. Wang, Some further thoughts about spectral kurtosis, spectral L2/L1 norm, spectral smoothness index and spectral Gini index for characterizing repetitive transients, Mech. Syst. Signal Process. 108 (2018) 360–368.

[32] B. Chen, D. Song, Y. Cheng, W. Zhang, B. Huang, Y. Muhamedsalih, Igigram: An improved Gini index-based envelope analysis for rolling bearing fault diagnosis, J. Dyn. Monit. Diagn. (2022) 111–124.

[33] D. Wang, Spectral L2/L1 norm: A new perspective for spectral kurtosis for characterizing non-stationary signals, Mech. Syst. Signal Process. 104 (2018) 290–293.

[34] Y. Miao, M. Zhao, J. Lin, Improvement of kurtosis-guided-grams via gini index for bearing fault feature identification, Meas. Sci. Technol. 28 (12) (2017) 125001.

[35] P. Borghesani, P. Pennacchi, S. Chatterton, The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings, Mech. Syst. Signal Process. 43 (1–2) (2014) 25–43.

[36] A. Mauricio, J. Qi, W.A. Smith, M. Sarazin, R.B. Randall, K. Janssens, K. Gryllias, Bearing diagnostics under strong electromagnetic interference based on integrated spectral coherence, Mech. Syst. Signal Process. 140 (2020) 106673.

[37] 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, Mech. Syst. Signal Process. 46 (1) (2014) 16–27.

[38] Y. Hong, M. Kim, H. Lee, J.J. Park, D. Lee, Early fault diagnosis and classification of ball bearing using enhanced kurtogram and Gaussian mixture model, IEEE Trans. Instrum. Meas. 68 (12) (2019) 4746–4755.

[39] B. Hou, Y. Chen, H. Wang, Z. Peng, K.-L. Tsui, D. Wang, OSESgram: Data-aided method for selection of informative frequency bands for bearing fault diagnosis, IEEE Trans. Instrum. Meas. 71 (2022) 1–10.

[40] S.S. Udmale, S.K. Singh, Application of spectral kurtosis and improved extreme learning machine for bearing fault classification, IEEE Trans. Instrum. Meas. 68 (11) (2019) 4222–4233.

[41] B. Hou, D. Wang, T. Xia, Z. Peng, K.-L. Tsui, Difference mode decomposition for adaptive signal decomposition, Mech. Syst. Signal Process. 191 (2023) 110203.

[42] T. Wang, Q. Han, F. Chu, Z. Feng, A new SKRgram based demodulation technique for planet bearing fault detection, J. Sound Vib. 385 (2016) 330–349.

[43] S. Wan, X. Zhang, Teager energy entropy ratio of wavelet packet transform and its application in bearing fault diagnosis, Entropy 20 (5) (2018) 388.

[44] W.N. Niehaus, S. Schmidt, P.S. Heyns, NIC methodology: A probabilistic methodology for improved informative frequency band identification by utilizing the available healthy historical data under time-varying operating conditions, J. Sound Vib. 488 (2020) 115642.

[45] Y. Miao, J. Wang, B. Zhang, H. Li, Practical framework of gini index in the application of machinery fault feature extraction, Mech. Syst. Signal Process. 165 (2022) 108333.

[46] D. Wang, J. Zhong, C. Li, Z. Peng, Box-Cox sparse measures: A new family of sparse measures constructed from kurtosis and negative entropy, Mech. Syst. Signal Process. 160 (2021) 107930.
[47] Y. Miao, M. Zhao, J. Hua, Research on sparsity indexes for fault diagnosis of rotating machinery, Measurement 158 (2020) 107733.
[48] D. Abboud, M. Elbadaoui, W. Smith, R. Randall, Advanced bearing diagnostics: A comparative study of two powerful approaches, Mech. Syst. Signal Process. 114 (2019) 604–627.
[49] D. Abboud, J. Antoni, S. Sieg-Zieba, M. Eltabach, Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment, Mech. Syst. Signal Process. 84 (2017) 200–226.
[50] D. Diamond, P.S. Heyns, A. Oberholster, Online shaft encoder geometry compensation for arbitrary shaft speed profiles using Bayesian regression, Mech. Syst. Signal Process. 81 (2016) 402–418.
[51] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17 (2020) 261–272, http://dx.doi.org/10.1038/s41592-019-0686-2.
[52] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, Nature 585 (7825) (2020) 357–362, http://dx.doi.org/10.1038/s41586-020-2649-2.
[53] P. McFadden, J. Smith, Vibration monitoring of rolling element bearings by the high-frequency resonance technique—A review, Tribol. Int. 17 (1) (1984) 3–10.