Mechanical Systems and Signal Processing 224 (2025) 111974

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Generalised envelope spectrum-based signal-to-noise objectives: Formulation, optimisation and application for gear fault detection under time-varying speed conditions

Stephan Schmidt [a,*], Daniel N. Wilke [b], Konstantinos C. Gryllias [c,d]

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

ARTICLE INFO

Communicated by Z. Mao

Keywords:
Targeted optimal filter design
Gear fault detection
Varying operating conditions
Fault signature enhancement

ABSTRACT

In vibration-based condition monitoring, optimal filter design improves fault detection by enhancing weak fault signatures within vibration signals. This process involves optimising a derived objective function from a defined objective. The objectives are often based on proxy health indicators to determine the filter's parameters. However, these indicators can be compromised by irrelevant extraneous signal components and fluctuating operational conditions, affecting the filter's efficacy. Fault detection primarily uses the fault component's prominence in the envelope spectrum, quantified by an envelope spectrum-based signal-to-noise ratio. New optimal filter objective functions are derived from the proposed generalised envelope spectrum-based signal-to-noise objective for machines operating under variable speed conditions. Instead of optimising proxy health indicators, the optimal filter coefficients of the formulation directly maximise the envelope spectrum-based signal-to-noise ratio over targeted frequency bands using standard gradient-based optimisers. Four derived objective functions from the proposed objective effectively outperform five prominent methods in tests on three experimental datasets.

1. Introduction

Vibration-based condition monitoring is frequently used to monitor critical rotating assets such as gearboxes. Diagnostic information in collected vibration signals is often masked by signal components generated by high-energy interactions between healthy rotating components [1], operating conditions [2], and the operating environment [3]. These signal components impede the performance of conventional vibration-based condition monitoring methods and can result in delayed detection and unexpected breakdowns due to false negatives. Automatic filter design methods can be used to construct filters that enhance the fault signatures and attenuate the extraneous components [4,5].

Abbreviations: ACYCBD, Adaptive maximum second-order CYClostationarity Blind Deconvolution; C-AIM, Centre for Asset Integrity Management; CYCBD, maximum second-order CYClostationarity Blind Deconvolution; ES, Envelope Spectrum; FIR, Finite Impulse Response; GES2N, Generalised Envelope spectrum-based Signal-to-Noise; ICS2, Indicator of second-order CycloStationarity; LES, Log-Envelope Spectrum; LESa, Log-Envelope Spectra; MOMEDA, Multi-point Optimal Minimum Entropy Deconvolution Adjusted; MSESHIRD, Maximum Squared Envelope Spectrum Harmonic-to-Interference Ratio Deconvolution; SES, Squared Envelope Spectrum; SESa, Squared Envelope Spectra; VS-DFT, Velocity Synchronous Discrete Fourier Transform.

* Corresponding author.
E-mail address: stephan.schmidt@up.ac.za (S. Schmidt).

https://doi.org/10.1016/j.ymssp.2024.111974
Received 2 May 2024; Received in revised form 14 August 2024; Accepted 18 September 2024
Available online 22 October 2024
0888-3270/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Informative frequency band methods often use a specific filter parametrisation to formulate the filter design problem as a two-dimensional optimisation problem where the lower and upper bounds of the bandpass filters are the design variables. However, the exact fault signatures are unknown. Therefore, proxy objectives for damage such as the spectral negentropy [6], the L_2/L_1-norm [7], the ratio of the cyclic content [8], the Indicator of second-order CycloStationarity (ICS2) [4] and envelope spectrum-based signal-to-noise objectives (e.g., [9–11]) are used to derive objective functions to determine optimal filter parameters. The optimal filter design problem can also be formulated to determine all filter coefficients by optimising an objective function derived from proxy objectives for damage [5]. This optimisation-based approach is referred to as filter coefficient optimisation and provides the advantage that the filter's frequency response has more flexibility to extract the underlying fault information.

1.1. Blind objectives for filter coefficient optimisation

Time domain-based objectives such as the kurtosis [12] and the generalised L_p/L_q ratio [13] have been used to determine the filter coefficients for enhancing the fault signature. Since the amplitude modulation due to the damaged gear and bearing components manifest at sparse cyclic orders in the envelope spectrum, sparsity proxy measures such as the spectral negentropy [14], the Gini index [15], the L_2/L_1-norm [14], the generalised L_p/L_q-norm [16], and the modified smoothness index [17] have been used to construct objective functions for filter coefficient optimisation to detect damage.

1.2. Targeted objectives for filter coefficient optimisation

Instead of using metrics that measure a statistic of the entire time-domain signal or entire envelope spectrum, metrics can be defined as those that target the properties of specific components. These metrics are referred to as targeted metrics [4] and can be more robust to extraneous impulsive components [18]. McDonald and Zhao [19] proposed a Multi-point Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA) method that finds the optimal filter coefficients in closed form by maximising the multi-D-norm of the signal. Buzzoni et al. [20] proposed the maximum second-order CYClostationarity Blind Deconvolution (CYCBD), which uses a targeted objective function based on the ICS2. Soave et al. [21] replaced the Fourier transform of the CYCBD with a Fourier-Bessel expansion for improved performance. The CYCBD targets the theoretical cyclic orders of the fault. Chen et al. [22] investigated different periodicity detection techniques to circumvent the need to specify the targeted cyclic order a priori in the CYCBD. Zhang et al. [23] proposed an Adaptive maximum second-order CYClostationarity Blind Deconvolution (ACYCBD) method, which uses the envelope harmonic product spectrum to determine the actual cyclic order of the damaged component before calculating the ICS2 of the filtered signal. Wang et al. [24] proposed another ACYCBD method, which determines the targeted cyclic frequency and the filter length for CYCBD. Other targeted metrics have recently been proposed for optimal filter design formulations. Liang et al. [25] proposed maximising the average kurtosis of each fault period. Meng et al. [18] proposed the maximum cyclic Gini index deconvolution, which calculates the Gini index of specifically targeted components and overcomes the Gini index's sensitivity to extraneous components. Peng et al. [26] proposed maximising the generalised Gaussian cyclostationarity indicator to determine the filter's coefficients. Zhou et al. [27] proposed the Maximum Squared Envelope Spectrum Harmonic-to-Interference Ratio Deconvolution (MSESHIRD) method, which maximises the ratio of the sum of the amplitudes in specific frequency bands in the Squared Envelope Spectrum (SES) to the difference between the sum of all the amplitudes in the SES and the sum of the amplitudes in specific frequency bands in the SES.

1.3. Challenges with varying operating conditions

Many critical power generation and mining assets are exposed to time-varying speed [28,29] and impulsive[1] conditions [31] that impede the performance of filter optimisation methods. More research needs to be conducted on developing and testing methods well-suited for these conditions [29]. Liang et al. [25] applied the maximum average kurtosis under varying speed conditions. Buzzoni et al. [20] proposed and investigated an extension of the CYCBD for time-varying speed conditions utilising the Velocity Synchronous Discrete Fourier Transform (VS-DFT) proposed by Borghesani et al. [32]. Zhou et al. [27] used a cyclic order band to accommodate speed fluctuations.

Fault signature enhancement remains challenging under time-varying speed and impulsive operating conditions, and objectives that are better proxies for fault signatures could result in improved filters. Envelope-based signal-to-noise metrics have proven useful for performance quantification [33], and envelope-based signal-to-noise objectives have performed well for informative frequency band identification under complex operating conditions [9,10]. Furthermore, the prominence of the amplitudes in the envelope spectrum is often used to diagnose the machine, e.g., if the ball pass outer race component is more prominent than the noise floor in the envelope spectrum, there is potential outer race damage in the system [34]. It provides another avenue to address the filter coefficient estimation problem. It is expected to perform better for fault signature enhancement under complex (i.e., time-varying speed and impulsive) operating conditions.

[1] Impulsive noise is also referred to as leptokurtic, heavy-tailed, or non-Gaussian noise in condition monitoring applications [30,31].

1.4. Contribution

To address this gap, a generalised envelope spectrum-based Signal-To-Noise (GES2N) objective is proposed for optimal filter design under varying speed conditions. Similar to CYCBD, MSESHIRD, the ratio of cyclic content [8], ICS2gram [4] (which uses the same objective as CYCBD), IESFOgram [9], and IFBI$_\alpha$gram [10] our proposed objective is the ratio between two linear functions in the envelope spectrum. The GES2N objective can be used to derive existing objectives, such as the ICS2 and MSESHIRD. New objectives can be derived with sufficient flexibility to accommodate envelope spectra suited for impulsive noise conditions. The advantages of using these new objectives for gear damage enhancement under time-varying speed conditions are demonstrated when compared against established methods on a simulated dataset and three experimental datasets. In summary, the proposed GES2N method makes it possible to derive novel objective functions that can be optimised to find filter coefficients to enhance weak fault signatures under time-varying speed conditions.

This work establishes terminology regarding metrics, objectives, objective functions, optimisation formulations and problems. Firstly, the mathematical expression of the optimisation objective can differ from the objective function being optimised, which is often done to improve the optimisability of an optimiser. The objective represents a task's overarching goal, aim, or idea, e.g., maximising the signal-to-noise ratio, while the objective function is the function that is optimised to realise the objective. This function could be the original or some transformed objective that enhances the optimisability of the objective. The optimisation problem comprises the defined variables, objective, and constraint functions to be solved using an appropriately selected optimisation formulation (e.g. minimisation, maximisation, optimality criterion, or gradient-only) using relevant optimisers. Objective and constraint functions with underlying characteristics should be appropriately paired with optimisers' fundamental assumptions of objective functions [35]. Lastly, objectives, objective functions and constraint functions merely evaluated on data can be used as metrics for fault detection.

1.5. Overview and structure

The paper is structured as follows: Section 2 defines the proposed optimal filter design method and in Section 3 the study outline is presented with the eight novel objectives derived using the proposed method. In Section 4, the eight novel objectives are compared against five established methods on three signals from a foundational model, whereafter four of the novel objectives are compared against the five established methods on measurements from three experimental datasets in Section 5. A primary sensitivity analysis is conducted to identify the effect of the three main selected parameters in Section 6. Finally, the work concludes in Section 7 with recommendations for future work. Appendix A contains a link to the code of the proposed GES2N objective. Appendix B contains the gradient of the objective function. Appendix C contains a brief overview of the foundational model considered in Section 4.

2. Proposed optimal filter design method

The vibration signal $\boldsymbol{x}$ is first pre-processed in Section 2.1 to filter the signal with the filter coefficients $\boldsymbol{g}(\boldsymbol{h}) = \frac{\boldsymbol{h}}{\|\boldsymbol{h}\|_2}$, where $\boldsymbol{h}$ denotes the design variables. The pre-processed signal $\boldsymbol{y}(\boldsymbol{h})$ is then transformed, as shown in Section 2.2, to obtain a suitable envelope spectrum of the signal $\boldsymbol{b}(\boldsymbol{h})$ for the derivation of suitable signal $\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s(\boldsymbol{h})\boldsymbol{b}(\boldsymbol{h})$ and noise $\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n(\boldsymbol{h})\boldsymbol{b}(\boldsymbol{h})$ indicators as outlined in Section 2.3. These signal and noise indicators are then used to formulate Generalised Envelope Spectrum-Based Signal-to-Noise (GES2N) objectives $\psi(\boldsymbol{h})$ outlined in Section 2.4 from which objective functions are derived and the optimisation problem is defined in Section 2.5. The dependence of the variables on the design variables, e.g., $\boldsymbol{b}(\boldsymbol{h})$, is not shown in subsequent sections to simplify the notation. Fig. 1 summarises the pipeline for the proposed GES2N objective and the minimisation objective function, with the notation defined in the respective sections.

2.1. Pre-processing

The measured vibration signal $\boldsymbol{x} = [x[0], x[1], \dots, x[L-1]]^{\mathrm{T}} \in \mathbb{R}^{L\times 1}$ is sampled at a constant frequency of f_s, where $t[n] = n/f_s$ is the discrete time corresponding to $x[n]$ and n is the time index. The angular speed vector of the reference shaft, measured in radians per second, is denoted $\boldsymbol{\omega} = [\omega[0], \omega[1], \dots, \omega[L-1]]^{\mathrm{T}} \in \mathbb{R}^{L\times 1}$ and the corresponding instantaneous angle of the shaft is denoted $\boldsymbol{\theta} = [\theta[0], \theta[1], \dots, \theta[L-1]]^{\mathrm{T}} \in \mathbb{R}^{L\times 1}$, with $\theta[0] = 0$ radians. Using the trapezoidal integration rule, the instantaneous angle can be estimated by integrating the angular speed over time.

A filtered vibration signal $\boldsymbol{y} \in \mathbb{R}^{L_y\times 1}$, is obtained by convolving the measured signal $\boldsymbol{x}$ with the FIR filter's coefficients $\boldsymbol{g} \in \mathbb{R}^{D\times 1}$, with the filter's length denoted D and $L_y = L - D + 1$. The convolution process, denoted $\boldsymbol{y} = \boldsymbol{x} \otimes \boldsymbol{g}$, can be written as a matrix-vector product $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g}$ [20], where $\boldsymbol{X} \in \mathbb{R}^{L_y\times D}$ is defined as the following Toeplitz matrix of $\boldsymbol{x}$:

$$\boldsymbol{X} = \begin{bmatrix} x[D-1] & x[D-2] & \dots & x[0] \\ x[D] & x[D-1] & \dots & x[1] \\ \vdots & \vdots & \ddots & \vdots \\ x[L-1] & x[L-2] & \dots & x[L-D] \end{bmatrix}. \tag{1}$$

The aim is to find $\boldsymbol{g}$ such that the fault signatures are enhanced in the filtered signal $\boldsymbol{y}$ and the extraneous components are attenuated. The filter coefficients $\boldsymbol{g}$ are normalised to have a magnitude of 1, i.e., $\|\boldsymbol{g}\|_2 = 1$. Hence, $\boldsymbol{g}$ can be calculated from filter coefficients with arbitrary magnitude $\boldsymbol{h}$ as follows: $\boldsymbol{g} = \boldsymbol{h}/\|\boldsymbol{h}\|_2$, which makes it possible to use unconstrained optimisation algorithms for the filter design problem.

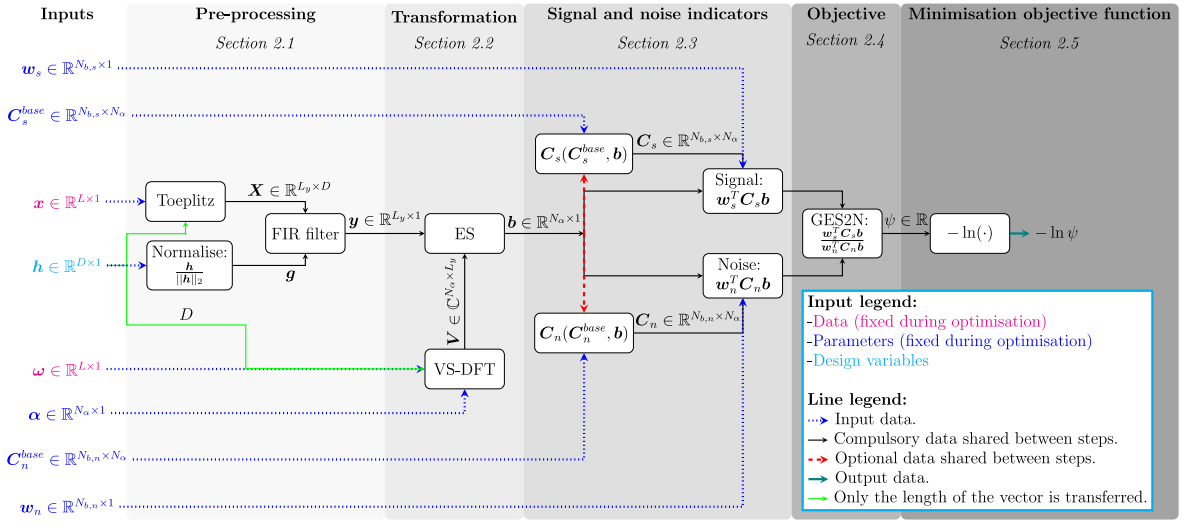


Fig. 1. An overview of the evaluation of the proposed Generalised Envelope Spectrum-Based Signal-to-Noise (GES2N) objective and associated objective function. Abbreviations: Finite Impulse Response (FIR); Envelope Spectrum (ES); Velocity Synchronous Discrete Fourier Transform (VS-DFT). The natural logarithm is denoted ln.

2.2. Transformation to the Envelope Spectrum (ES)

Damaged rotating components such as bearings and gears often generate fault signatures that manifest as prominent components in envelope spectra [10,34]. These components manifest at characteristic orders that are dependent on the kinematics of the damaged components (e.g., bearings) and the system (e.g., velocity ratio between the component's shaft and the reference shaft, i.e., the shaft with the tachometer) [34]. For example, outer race bearing damage typically generates periodic modulation governed by the ball pass outer race order of the bearing [34]. Gear damage often results in periodic modulation with the gear's rotation [36]. The component's characteristic cyclic order α_c can be targeted for damage detection and used in targeted objectives.

Variants of envelope spectra have been used in the literature, such as the SES estimated with the squared signal [14,20], the SES estimated using the squared analytic signal [11] and the squared spectrum of the log of the envelope, which is referred to as the Log-Envelope Spectrum (LES) [11,31]. The order-domain SES, denoted $\boldsymbol{b}^{(SES)} \in \mathbb{R}^{N_\alpha \times 1}$, with the corresponding cyclic orders $\boldsymbol{\alpha} = [\alpha[0], \alpha[1], \dots, \alpha[N_\alpha - 1]]^{\mathrm{T}} \in \mathbb{R}^{N_\alpha \times 1}$, is estimated with

$$\boldsymbol{b}^{(SES)} = (\boldsymbol{V}\,(\boldsymbol{y} \odot \boldsymbol{y}))^* \odot (\boldsymbol{V}\,(\boldsymbol{y} \odot \boldsymbol{y})), \tag{2}$$

where * denotes the element-wise conjugate operator, $\odot$ denotes the Hadamard product, and $\boldsymbol{V} \in \mathbb{C}^{N_\alpha \times L_y}$ denotes the velocity synchronous Fourier matrix [32], with the number of cyclic orders denoted N_α. Each element of the velocity synchronous matrix is calculated with [20,32]

$$V[m,a] = \frac{1}{f_s \cdot \theta[L_y - 1]} \cdot \omega[a + D - 1] \cdot e^{-j \cdot \alpha[m] \cdot \theta[a+D-1]}, \tag{3}$$

with the rotational speed of the shaft at time increment n denoted $\omega[n]$, the instantaneous angle of the shaft at time increment n denoted $\theta[n]$, $j = \sqrt{-1}$, $m \in \{0, 1, \dots, N_\alpha - 1\}$, and $a \in \{0, 1, \dots, L_y - 1\}$. The cyclic order resolution is fixed and equal to $\Delta\alpha = \frac{\alpha[N_\alpha - 1] - \alpha[0]}{N_\alpha - 1}$, with $\alpha[k] = k \cdot \Delta\alpha$, $\forall\, k \in \{0, 1, \dots, N_\alpha - 1\}$.

The Log-Envelope Spectrum (LES) is better suited than the SES for heavy-tailed signals [31]. The LES can be estimated with:

$$\boldsymbol{b}^{(LES)} = (\boldsymbol{V} \cdot \mathbf{log}\left((\boldsymbol{A}\boldsymbol{y})^* \odot (\boldsymbol{A}\boldsymbol{y})\right))^* \odot (\boldsymbol{V} \cdot \mathbf{log}\left((\boldsymbol{A}\boldsymbol{y})^* \odot (\boldsymbol{A}\boldsymbol{y})\right)) \tag{4}$$

where the transformation matrix is defined by $\boldsymbol{A} \in \mathbb{C}^{L_y \times L_y}$, and $\mathbf{log} : \mathbb{R}^{L_y \times 1} \mapsto \mathbb{R}^{L_y \times 1}$ denotes the array logarithm function. The Hilbert transform, often used in vibration-based condition monitoring, is an example of a linear transformation of $\boldsymbol{y}$ [37] and used to define $\boldsymbol{A}$ in Eq. (4). The envelope spectrum $\boldsymbol{b}$ can therefore be estimated using either Eq. (2) or Eq. (4), but the subsequent discussion will focus on the SES.

To establish terminology, an SES with a cyclic order resolution of $\Delta\alpha$ is shown in Fig. 2(a), with the signal components associated with a cyclic order of α_c and its harmonics highlighted in Fig. 2(a). Instead of monitoring a specific cyclic order (e.g., α_c), monitoring and extracting metrics from bands around the theoretically targeted components is possible. An example of cyclic order bands with a constant bandwidth around the targeted cyclic orders α_c is shown in Fig. 2(b). The bandwidth $\Delta\alpha_b$ is constant in this work, but can be selected to be a function of the cyclic order, e.g., $\Delta\alpha_b(\alpha_c)$. The cyclic order bands accommodate slight deviations from the expected cyclic order.

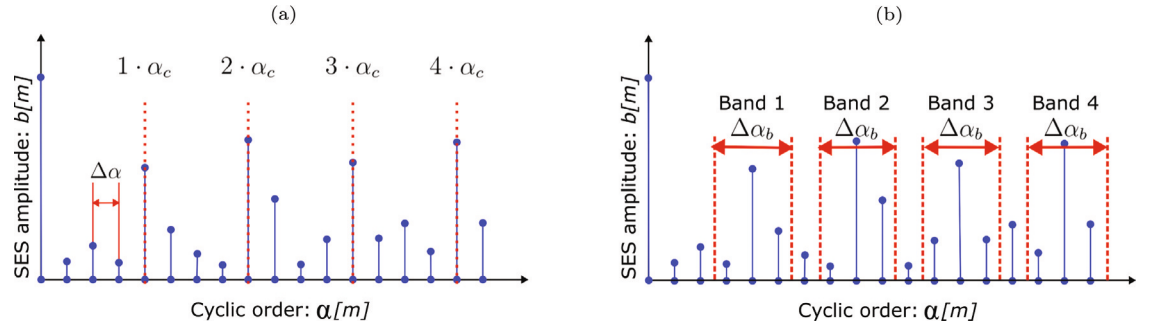


Fig. 2. The SES amplitudes $\boldsymbol{b}$ are presented against the corresponding cyclic orders $\boldsymbol{\alpha}$, with the amplitude corresponding to the cyclic order $\alpha[m]$ denoted $b[m]$. In (a), the cyclic order resolution $\Delta\alpha$ and the amplitudes of four targeted harmonics $k \cdot \alpha_c$ are shown, i.e., $N_h = 4$. In (b), four cyclic bands with a constant bandwidth $\Delta\alpha_b$ around the targeted harmonics $k \cdot \alpha_c$ are shown.

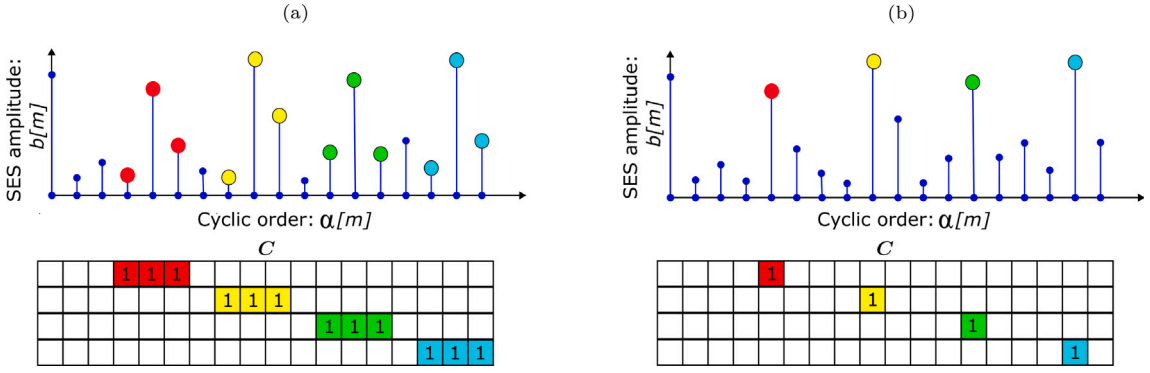


Fig. 3. Two examples of weighting matrices are shown for the Squared Envelope Spectrum (SES) in Fig. 2(b). The SES has four bands $N_b = 4$. If the sum of the components in each band is calculated (e.g., like the MSESHIRD objective function), the cyclic order weighting matrix in Fig. 3(a) can be used. If the maximum of each band needs to be calculated, the cyclic order weighting matrix has the form shown in Fig. 3(b).

2.3. Signal and noise indicators

The signal $\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}$ and noise $\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}$ indicators are derived from cyclic order weighting matrices $\boldsymbol{C}_i$ and cyclic band weighting vectors $\boldsymbol{w}_i$ with $i \in \{s, n\}$.

2.3.1. Cyclic order weighting matrices

The cyclic order weighting matrices assign a weight to each cyclic order $\alpha \in \boldsymbol{\alpha}$. The cyclic order weighting matrices $\boldsymbol{C}_i, i \in \{s, n\}$ are obtained by processing the baseline cyclic order matrices $\boldsymbol{C}_i^{base}$, with the processing denoted $\boldsymbol{C}_i(\boldsymbol{C}_i^{base}, \boldsymbol{b})$. The following baseline cyclic order weighting matrix is used for the signal indicator:

$$C_s^{base}[m,l] = \begin{cases} 1 & \text{if } (m+1)\cdot\alpha_c - \frac{1}{2}\Delta\alpha_b \le \alpha[l] \le (m+1)\cdot\alpha_c + \frac{1}{2}\Delta\alpha_b \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

with $l \in \{0, 1, \ldots, N_\alpha - 1\}$, $m \in \{0, 1, \ldots, N_{b,s} - 1\}$, and $N_{b,s}$ the number of bands in the signal indicator. A representative example of this weighting matrix is shown in Fig. 3(a) for the SES in Fig. 2(b), i.e., it is non-zero for each cyclic order $\alpha[l]$ that is within a band of $\Delta\alpha_b$ around the targeted cyclic order α_c.

Two cyclic order weighting matrices, obtained by processing the baseline cyclic order weighting matrix in Eq. (5), are considered for signal indicators. The first cyclic order weighting matrix calculates the mean of the amplitudes in all bands and is defined as follows:

$$C_s^{mean}[m,l] = \frac{C_s^{base}[m,l]}{\sum_{m=0}^{N_{b,s}-1}\sum_{l=0}^{N_\alpha-1} C_s^{base}[m,l]}. \tag{6}$$

The mean of the amplitudes in Eq. (6) is proportional to the amplitudes' sum and, therefore, will have the same optima. The second cyclic order weighting matrix calculates the maximum amplitude in each targeted band:

$$C_s^{max}[m,l] = \begin{cases} 1 & \text{if } C_s^{base}[m,l]b[l] = \max_q \left\{C_s^{base}[m,q]b[q]\right\} \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

A representative example of the cyclic order weighting matrix $\boldsymbol{C}_s^{max}$ is shown in Fig. 3(b) and can be derived using the cyclic order weighting matrix in Fig. 3(a) using Eq. (7). Hence, either $\boldsymbol{C}_s^{max}$ or $\boldsymbol{C}_s^{mean}$ is used to define the signal indicators in the objectives.

The noise indicator is defined as a baseline cyclic order weighting matrix using a single band $N_{b,d} = 1$ as follows:

$$C_n^{base}[0,l] = \begin{cases} 1 & \text{if } \alpha_{n,min} \le \alpha[l] \le \alpha_{n,max} \text{ and } \sum_{m=0}^{N_{b,s}-1} C_s[m,l] = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where $\boldsymbol{C}_n^{base} \in \mathbb{R}^{1\times N_\alpha}$ and $\alpha_{n,min}$ and $\alpha_{n,max}$ denotes the minimum and maximum cyclic order for estimating the noise. It is non-zero for all cyclic orders in the range $\alpha_{n,min} \le \alpha[l] \le \alpha_{n,max}$, excluding the cyclic orders that fall in the numerator's bands, i.e., excluding cyclic orders where $\sum_{m=0}^{N_{b,s}-1} C_s[m,l] \neq 0$. It is recommended that $\alpha_{n,min} < \alpha_c$ and $\alpha_{n,max} > N_{b,s}\cdot\alpha_c$ to ensure a representative estimate of the noise floor is obtained. More complex weighting matrices, where the noise is estimated around the targeted bands [10], can be used with this formulation but are not investigated in this work. Instead, the cyclic order weighting matrix of the noise component is processed as follows:

$$C_n^{mean}[m,l] = \frac{C_n^{base}[m,l]}{\sum_{m=0}^{N_{b,n}-1}\sum_{l=0}^{N_\alpha-1} C_n^{base}[m,l]}, \tag{9}$$

i.e., the mean of the amplitudes is calculated with $\boldsymbol{C}_n^{mean} = \boldsymbol{C}_n(\boldsymbol{C}_n^{base}, \boldsymbol{b})$.

2.3.2. Cyclic band weighting matrices

The cyclic band weighting vectors $\boldsymbol{w}_s \in \mathbb{R}^{N_{b,s}\times 1}$ and $\boldsymbol{w}_n \in \mathbb{R}^{N_{b,d}\times 1}$ are used to assign a weight to each respective cyclic order band in the signal indicator (i.e., the weight of each row in Fig. 3) and the noise indicator, respectively. This provides additional flexibility to increase the relative weight of specific cyclic bands, e.g., the fundamental targeted component's cyclic band can be weighted more than its harmonics using $\boldsymbol{w}_s$. In this work, the weight of all cyclic order bands is equal, i.e., therefore $\boldsymbol{w}_s = \boldsymbol{1} \in \mathbb{R}^{N_{b,s}\times 1}$ and $\boldsymbol{w}_n = \boldsymbol{1} \in \mathbb{R}^{N_{b,d}\times 1}$, with the investigation of other variants recommended for future work.

2.4. Generalised Envelope spectrum-based Signal-to-Noise (GES2N) objective

The following generalised objective is proposed using the signal and noise indicators[2]:

$$\psi\left(\boldsymbol{g};\boldsymbol{x},\boldsymbol{w}_s,\boldsymbol{C}_s,\boldsymbol{w}_n,\boldsymbol{C}_n\right) = \frac{\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}}{\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}}, \tag{10}$$

where $\psi : \mathbb{R}^{D\times 1} \mapsto \mathbb{R}$ and $\boldsymbol{b}$ is the envelope spectrum of the signal filtered with $\boldsymbol{g}$. The numerator in Eq. (10) contains the signal indicator, i.e., the targeted information that needs to be maximised such as the fault signatures, and the denominator contains the noise indicator, i.e., the information that needs to be attenuated such as the noise floor in the SES and the extraneous components. The objective in Eq. (10) is subsequently denoted $\psi(\boldsymbol{g};\boldsymbol{x})$ to simplify the notation, with the objectives used in this work summarised in Section 3.

2.5. Optimisation problems

The optimal filtering problem can be casted as the following unconstrained minimisation problem:

$$\min_{\boldsymbol{h}} -\ln\psi\left(\boldsymbol{g}(\boldsymbol{h});\boldsymbol{x}\right), \tag{11}$$

where $\boldsymbol{g}(\boldsymbol{h}) = \frac{\boldsymbol{h}}{\|\boldsymbol{h}\|_2} \in \mathbb{R}^{D\times 1}$ is constrained to be a unit vector, i.e., it has the magnitude $\|\boldsymbol{g}\|_2 = 1$ and the design variables $\boldsymbol{h}$ can have an arbitrary magnitude. Hence, the objective function is always evaluated with normalised filter coefficients. The minimisation objective function is denoted $-\ln\psi(\boldsymbol{g};\boldsymbol{x})$, where $\psi(\boldsymbol{g};\boldsymbol{x})$ is defined in Eq. (10). We found that the logarithm of the objective in Eq. (10) was easier to optimise than the objective itself. However, finding the best problem for a given optimiser is not in the scope of this work. The natural logarithm of the objective in Eq. (10), evaluated with the normalised filter coefficients $\boldsymbol{g}$, is given by

$$\ln\psi(\boldsymbol{g};\boldsymbol{x}) = \ln\left(\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}\right) - \ln\left(\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}\right), \tag{12}$$

with the filtered signal obtained using $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g} = \boldsymbol{X}\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|_2}$.

Gradient-based minimisers with analytical gradients are used, as these methods are better suited than gradient-free minimisers and population-based solvers for high-dimensional problems. The gradient-based minimisers perform well for filter coefficient optimisation problems, and several well-established minimisers are available in the literature [35] and programming language libraries (e.g., NLopt [38], Julia's Optim [39] and Python's SciPy modules [40]). Eq. (12)'s gradient is defined in Appendix B.

We found that the Conjugate Gradient (CG) method in Julia's Optim [39] performs sufficiently well to solve the optimisation problem in Eq. (11) when using linear predictive coding [20] to initialise the filter. Finding the best combination of optimisation problem, optimiser, and initialisation strategies is beyond the scope of this work. The functions were implemented in Julia [41].

[2] This representation is not unique, for example, we can write $\frac{\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}}{\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}} = \frac{\boldsymbol{c}_s^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{c}_n^{\mathrm{T}}\boldsymbol{b}}$ where $\boldsymbol{c}_i$ contains the weighting of the cyclic order components, i.e., $\boldsymbol{c}_i = \boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{C}_i, \quad i \in \{s,n\}$. However, the specific objective separates the cyclic order and cyclic band weightings to provide more flexibility.

2.6. Summary of method

To summarise, the vibration signal $\boldsymbol{x}$ is firstly filtered with L_2-normalised filter coefficients $\boldsymbol{g} = \boldsymbol{h}/\|\boldsymbol{h}\|_2$ to obtain the filtered signal $\boldsymbol{y}$, where $\boldsymbol{h}$ denotes the design variables as described in Section 2.1. After that, the ES of the filtered signal $\boldsymbol{b}$ is calculated using the VS-DFT matrix $\boldsymbol{V}$, the rotational speed of a reference shaft $\boldsymbol{\omega}$ and the cyclic order vector $\boldsymbol{\alpha}$ as described in Section 2.2. In Section 2.3, the signal $\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}$ and noise $\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}$ indicators are calculated using cyclic band weighting vectors $\boldsymbol{w}_i$ and cyclic order weighting matrices $\boldsymbol{C}_i$, with $i \in \{s, n\}$. The cyclic order weighting matrices are obtained by processing base matrices $\boldsymbol{C}_i^{base}$ that remain constant during the optimisation process, where the general processing function is denoted $\boldsymbol{C}_i(\boldsymbol{C}_i^{base}, \boldsymbol{b})$, with $i \in \{s, n\}$. Therefore, the ES can be used in the base matrix processing and is included in the general processing function. The signal and noise indicators are combined as ratios to derive the GES2N objective in Section 2.4. In Section 2.5, the minimisation objective function is defined as the objective's negative logarithm, and the optimisation problem is defined.

3. Study outline

This section presents the study outline used in subsequent investigations. Section 3.1 summarises the datasets considered, and Section 3.2 describes the novel objectives, established methods considered, and the optimisation settings used. Section 3.3 describes the performance metrics used in the investigations and the filter length selection process.

3.1. Datasets

3.1.1. Simulated data

In Section 4, three simulated signals from a foundational model are used to compare the SES-based GES2N objectives, the LES-based GES2N objectives and established methods under time-varying speed under different impulsive noise conditions. This dataset allows for comparing methods in a controlled environment, demonstrates the impact of impulsive noise conditions on the methods, and demonstrates the impact of using SES and LES to estimate $\boldsymbol{b}$ in the GES2N objectives.

3.1.2. Experimental datasets

In Section 5, experimentally measured datasets are used to compare SES-based GES2N objectives against established methods to enhance damage in SES under time-varying speed conditions. Since the focus is on comparing against established SES-based methods and their metrics, the LES-based GES2N objectives are not considered but will be considered in future work. Three experimental datasets are considered: A localised gear damage dataset with the same varying speed profile for five measurements and two distributed gear damage cases with different speed profiles for the four measurements used.

Additional datasets were also considered but not incorporated in this paper for brevity's sake. The results are supplied in Appendix A.

3.2. Methods and parameter selection

Eight objectives are derived using the proposed GES2N objective and used on the datasets outlined in Table 1. Two groups of objectives are used, namely, SES-based objectives and LES-based objectives. Within each group, four objectives use the following naming convention:

$$\text{GES2N-\{Signal: Max| Mean\}-\{Noise band: Nf} \implies \alpha_{n,\min} = 0 \mid \text{Np} \implies \alpha_{n,\min} = 0.5 \cdot \alpha_c\}. \tag{13}$$

The naming convention indicates whether the Maximum or Mean cyclic order weighting matrix is used to construct the signal indicator and whether the noise indicator's $\alpha_{n,min}$ uses either the full spectrum with the 0 cyclic order component included (Nf) or only includes a part of the cyclic orders (above $0.5 \cdot \alpha_c$) (Np). The GES2N-Mean-Nf objective is consistent with the MSESHIRD method [27]. Still, instead of summing all the amplitudes in the bands in the numerator, the mean of the amplitudes is calculated. Furthermore, only the cyclic orders associated with the first $N_{h,s}$ harmonics are utilised in the denominator, not the full envelope spectrum. The 0 cyclic order component calculates the squared energy of the angle domain signal and can be sensitive to extraneous components. Hence, the GES2N-Mean-Np objective is also investigated, which ignores the low cyclic order components (below $0.5 \cdot \alpha_c$ cyclic orders). The mean in the numerator can be insensitive to sparse components. Therefore, the maximum amplitudes in each cyclic order band are utilised to extract the amplitudes in the GES2N-Max-Nf and the GES2N-Max-Np objectives.

The proposed objectives are compared against CYCBD [20], ACYCBD [23], MOMEDA [19], L_2/L_1 norm of the envelope spectrum [14] and spectral negentropy of the SES [14]. For CYCBD, the angle domain formulation was utilised with the first 10 harmonics used and a cyclostationary order of 2 was maximised. For ACYCBD, the angle formulation was utilised due to the time-varying speed conditions. A cyclostationary order of 2 was maximised with no frequency limit imposed. MOMEDA was applied to the time domain signal, with the average period of the targeted component used with the default windowing function.

The parameters for the proposed GES2N variants are as follows unless stated otherwise: $\Delta\alpha_b = 0.1$ shaft orders, $N_{h,s} = 10$, and $\Delta\alpha = \pi/(\theta[L-1] - \theta[0])$. Further information on the required cyclic order resolution can be found in Refs. [32,42]. The following parameters are used for all the algorithms unless stated otherwise: Convergence tolerance: 10^{-12}, Maximum number of iterations: 1500, Filter initialisation method: Linear predictive coding [20]. Three filter lengths were applied $D \in \{16, 64, 256\}$ with only the results of the best performing filter lengths reported.

Table 1
Four SES-based and four LES-based objectives are derived using the proposed GES2N objective. The GES2N objective's naming convention is defined in Eq. (13). The number of harmonics in the numerator of Eq. (10) is denoted $N_{h,s}$, and the target cyclic order is denoted α_c. The simulated and experimental datasets are discussed in Sections 3.1.1 and 3.1.2.

Objective	Signal	Noise	Noise band		b	Datasets	
	Eq.	Eq.	$\alpha_{n,min}$	$\alpha_{n,max}$	Eq.	Sim.	Exp.
GES2N-Mean-Nf (SES)	(6)	(9)	0.0	$N_{h,s} \cdot \alpha_c + 1$	(2)	✗	✗
GES2N-Mean-Np (SES)	(6)	(9)	$0.5 \cdot \alpha_c$	$N_{h,s} \cdot \alpha_c + 1$	(2)	✗	✗
GES2N-Max-Nf (SES)	(7)	(9)	0.0	$N_{h,s} \cdot \alpha_c + 1$	(2)	✗	✗
GES2N-Max-Np (SES)	(7)	(9)	$0.5 \cdot \alpha_c$	$N_{h,s} \cdot \alpha_c + 1$	(2)	✗	✗
GES2N-Mean-Nf (LES)	(6)	(9)	0.0	$N_{h,s} \cdot \alpha_c + 1$	(4)	✗	
GES2N-Mean-Np (LES)	(6)	(9)	$0.5 \cdot \alpha_c$	$N_{h,s} \cdot \alpha_c + 1$	(4)	✗	
GES2N-Max-Nf (LES)	(7)	(9)	0.0	$N_{h,s} \cdot \alpha_c + 1$	(4)	✗	
GES2N-Max-Np (LES)	(7)	(9)	$0.5 \cdot \alpha_c$	$N_{h,s} \cdot \alpha_c + 1$	(4)	✗	

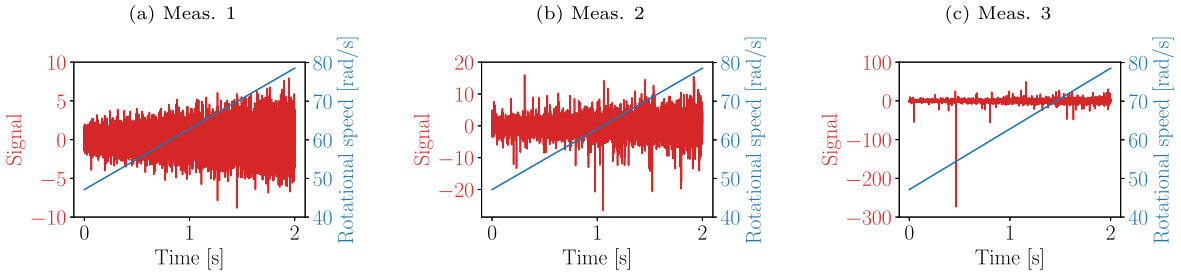


Fig. 4. The three signals x_s considered in the foundational example (red) with the linear varying speed profile superimposed on the figure (blue). The signals have the same train-of-impulses and extraneous component, but with different noise distributions. Figures 4(a), 4(b) and 4(c) use a Student-t noise variable with 1000.0, 3.0 and 2.5 degrees-of-freedom respectively. Note: The following time-varying speed is used: $\omega(t) = 2 \cdot \pi \cdot 10 \cdot ((t - 1.0) \cdot 0.25 + 1)$ in rad/s.

3.3. Performance comparison and filter length selection

The median normalised ES (SES or LES) of the filtered signals and the L_2-normalised filter coefficients' frequency response functions are compared between the different methods. The spectra are normalised to aid the comparison. To quantitatively evaluate and compare the performance of the methods, the following quantitative performance metrics of the ES of the filtered signals are used in subsequent sections:

M1: The first metric calculates the logarithm of the ratio between the mean of the amplitudes of the first $N_{h,s}$ targeted harmonics in the ES and the median of the ES. Hence, it is a different estimate of the signal-to-noise ratio in the ES and is large if the damage is prominent relative to the noise floor in the ES.

M2: The second metric calculates the logarithm of the ratio between the mean of the amplitudes of the first $N_{h,s}$ targeted harmonics in the ES and a known dominant extraneous component in the signal. Hence, this metric measures the prominence of the fault components relative to the extraneous component in the ES.

The metrics were normalised over all the measurements of each datasets to simplify the comparison. For each method, the filter length $D \in \{16, 64, 256\}$ that maximises the average normalised metrics was selected and used in subsequent analyses. This reduces the impact of the filter length on the results.

4. Simulated dataset

The foundational model used to generate the signals is described in Appendix C, briefly summarising the signals here. Each simulated signal $x_s(t)$ contains three components:

$$x_s(t) = x_{dam}(t) + x_{ext}(t) + x_{noise}(t), \tag{14}$$

where $x_{dam}(t)$ is a damaged component that consists of a periodic train of impulses with a cyclic order of 2.0 events/revolution, $x_{ext}(t)$ is an extraneous periodic impulsive component that manifests at a cyclic order of $\alpha_{ext} = 15.0$ events/revolution and impulsive noise $x_{noise}(t)$. The $x_{dam}(t)$ is targeted and therefore $\alpha_c = 2.0$. The impulsive noise is simulated using a Student-t distribution, with its Degree-Of-Freedom (DOF) parameter used to control the impulsivity of the noise. The DOF is inversely related to the impulsiveness of the generated samples. The three simulated measurements are shown in Fig. 4 for different Student-t DOFs. Figs. 4(a), 4(b) and 4(c) use a Student-t noise variable with 1000 DOF (Gaussian-like noise), 3 DOF and 2.5 DOF (highly impulsive noise) respectively.

The objective functions were optimised as described in Section 3.2 to obtain the optimal filter coefficients for the two signals. After that, the filtered signals were obtained, and either the Squared Envelope Spectra (SESa) or the Log-Envelope Spectra (LESa) of the filtered signals were calculated.

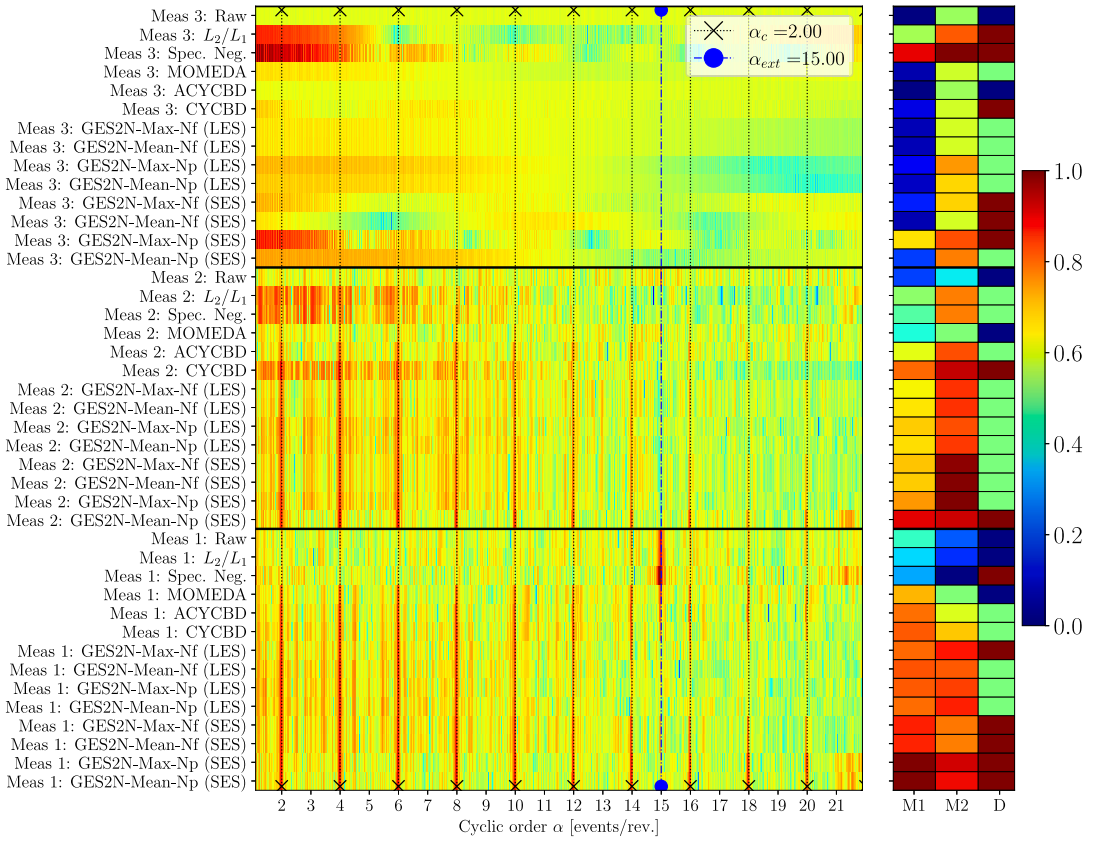


Fig. 5. The logarithm of the median normalised Squared Envelope Spectra (SESa) of the filtered signals $\boldsymbol{y}$ for the signals in Fig. 4 are presented on the left side with the logarithm of the normalised SESa of the raw signals (denoted Raw). The SESa are normalised with their respective medians. The damage manifests at $\alpha_c = 2.0$ and its harmonics, and $\alpha_{ext} = 15.0$ shaft orders and its harmonics are associated with extraneous impulsive components. The two performance metrics, described in Section 3.3, are shown on the right with the filter length $D \in \{16, 64, 256\}$. The logarithm of each respective metric and the logarithm of the filter length D are normalised between 0 and 1 across the rows. The SES plot on the left side is also normalised to have a magnitude between 0 and 1, and therefore, both plots use the same colour bar.

4.1. Squared Envelope Spectra (SESa) of the filtered signals

The SESa of the filtered and raw signals are shown in Fig. 5. To condense the results, the SESa of the filtered signals in Fig. 4 are presented as three-dimensional plots in Fig. 5; the x-axis shows the cyclic orders, the y-axis shows the measurement number and the name of the objective, and the colours quantify the logarithm of the median normalised SES's amplitude. The SESa are normalised to make them easier to compare. In Fig. 5, two cyclic order components are highlighted, namely, the targeted damaged component ($\alpha_c = 2.0$) and its harmonics and the extraneous component ($\alpha_{ext} = 15.0$). The two metrics discussed in Section 3.3 and the selected filter length are summarised in Fig. 5. The metrics were normalised to be within the range of 0 to 1 along the rows, while the logarithm of the normalised SESa was also adjusted to fall within the same range of 0 to 1, ensuring uniformity in colour representation.

For measurement 1, the SESa and metrics show that all the methods enhance the damage components at 2.0 orders and its harmonics except for the spectral negentropy and the L_2/L_1 objectives that maximise the extraneous component. The performance of the spectral negentropy and L_2/L_1 is expected, as these methods do not target a specific cyclic component and the extraneous component is a dominant sparse component in the SES. The GES2N-Mean-Np (SES) and GES2N-Max-Np (SES) perform the best in enhancing the fault signatures and attenuating the extraneous component according to M1 and M2, with M1 equal to 1.0 and their M2s equal to 0.90 and 0.93 respectively. The other SES-based GES2N objectives (M1 between 0.87–0.88), the LES-based objectives (M1 between 0.79–0.82), the CYCBD (M1 of 0.82), the ACYCBD (M1 of 0.79) and MOMEDA (M1 of 0.72) are slightly less effective in enhancing the damage. The GES2N objectives (M2 between 0.78 and 0.90) attenuated the extraneous component better relative to the damage components when compared to the CYCBD (M2 of 0.70), ACYCBD (M2 of 0.61) and MOMEDA (M2 of 0.5).

It is more challenging to enhance the fault signatures in Measurement 2. The GES2N-Mean-Np (SES) performs the best in enhancing the fault signatures (M1 of 0.91), with the other SES-based GES2N objectives (M1s between 0.69–0.75), LES-based GES2N objectives (M1s between 0.65–0.69), the CYCBD (M1 of 0.8), and the ACYCBD (M1 of 0.63) still able to enhance the fault signatures. The MOMEDA (M1 of 0.39), the spectral negentropy (M1 of 0.45) and the L_2/L_1 (M1 of 0.52) are not able to enhance the fault

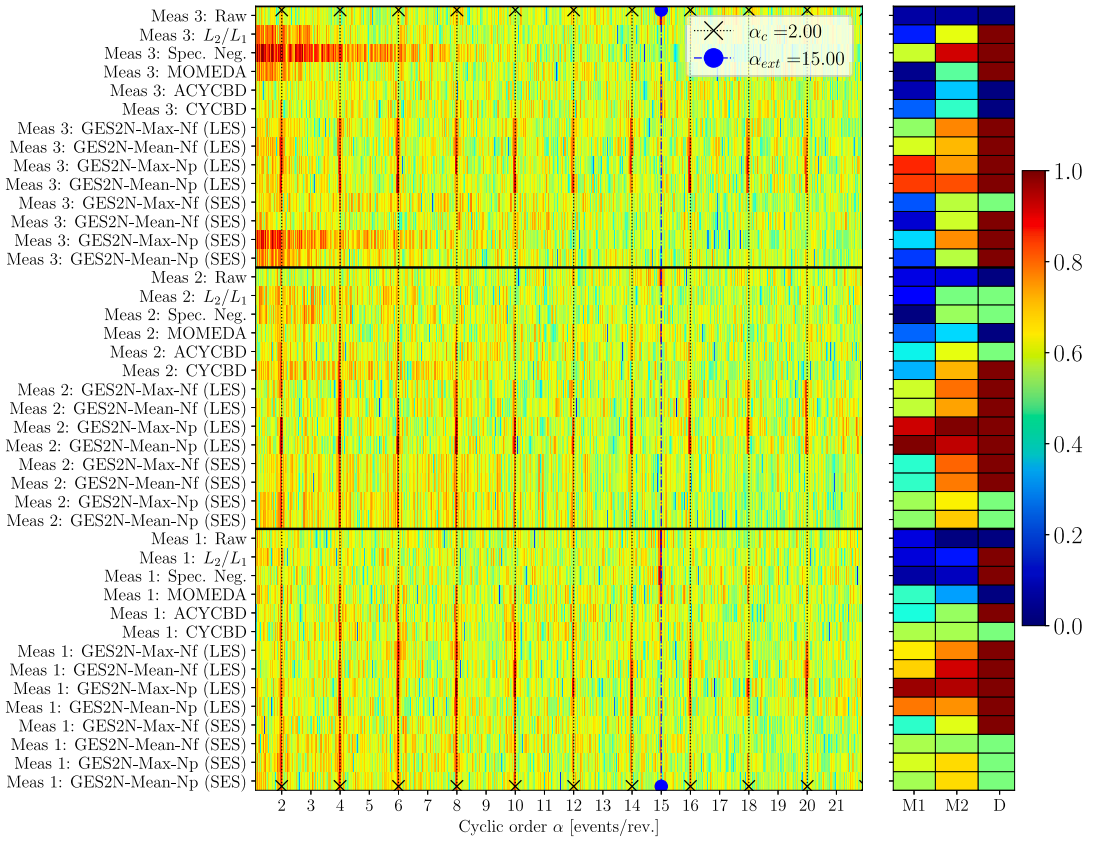


Fig. 6. The logarithm of the median normalised Log-Envelope Spectra (LESa) of the filtered signals and the corresponding normalised metrics M1 and M2, and the normalised filter length are presented for the signals in Fig. 4. The same presentation style as Fig. 5 is used.

signatures. All the methods, except MOMEDA (M2 of 0.50) performed well in attenuating the extraneous components with M2s between 0.77−1.00. Using M2, the CYCBD (0.94) and the SES-based GES2N objectives (0.93–1.00) performed better than the ACYCBD (0.83) and the LES-based GES2N objectives (0.82–0.86).

For Measurement 3, the impulsive noise impeded the efficacy of the squared envelope spectrum. The spectral negentropy, the L_2/L_1 and the GES2N-Max-Np (SES) enhance lower cyclic orders likely caused by the impulsive noise that overlaps with the fault's cyclic orders.

The longer filter length $D = 256$ is selected in most cases (64.1% of the cases) in Fig. 6, followed by $D = 64$ (25.6% of the cases). Some objectives consistently use the same filter length (e.g., the LES-based GES2N objectives), but some variation is seen (e.g.,. CYCBD selected three different Ds). This variation highlights the importance of considering different filter lengths when comparing objectives, the same filter length is not the best for all cases.

4.2. Log-Envelope Spectra (LESa) of the filtered signals

Instead of visualising the SESa of the filtered signals and corresponding metrics, the LESa are presented in Fig. 6 with their metrics. For Measurement 1, the GES2N-Max-Np (LES) enhanced the fault signatures (M1 of 0.97) and attenuate the extraneous components (M2 of 0.95) the best, followed by the other LES-based GES2N objectives (M1s and M2s ranging between 0.66−0.71 and 0.76−0.82 respectively), the SES-based GES2N (M1s and M2s ranging between 0.41−0.61 and 0.53−0.63 respectively), the CYCBD (M1 and M2 of 0.56 and 0.55 respectively), the ACYCBD (M1 and M2 of 0.38 and 0.53 respectively) and the MOMEDA (M1 and M2 of 0.41 and 0.28). The spectral negentropy and the L_2/L_1 did not enhance the damage.

For Measurement 2, the GES2N-Max-Np (LES) and GES2N-Mean-Np (LES) perform the best (M1s of 0.93 and 1.0 respectively and M2s of 1.0 and 0.94 respectively). The other GES2N objectives enhance the fault signatures and attenuate the extraneous component, but are less effective with M1s and M2s between 0.40−0.60 and 0.65−0.80 respectively. The ACYCBD (M1 of 0.37), CYCBD (M1 of 0.30) and MOMEDA (M2 of 0.23) enhance the damage, despite them being less effective than the GES2N-based methods. The spectral negentropy and L_2/L_1 enhanced lower cyclic orders, potentially caused by the impulsive conditions.

For Meas. 3, only the LES-based GES2N could enhance the fault signatures (M1s between 0.53−0.87) and attenuated the extraneous impulsive component (M2s between 0.71−0.83). The GES2N-Max-Np (SES), the GES2N-Mean-Np (SES), the spectral negentropy and the L_2/L_1 enhanced lower cyclic orders likely caused by the impulsive noise.

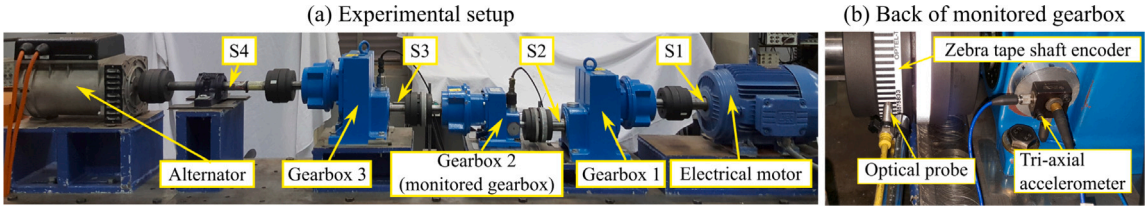


Fig. 7. The helical gearbox test bench in the Centre for Asset Integrity Management (C-AIM) laboratory is presented. (a) The components in the test bench are shown. (b) The sensors that were utilised are shown.

4.3. Discussion

The results demonstrate that for Gaussian-like noise (Measurement 1), all the targeted methods could enhance the fault component of interest. Still, the GES2N-based objectives perform better according to the metrics (e.g., on average 13.3% (M1) and 11.2% (M2) better than the CYCBD). Using the SES-based metrics, the GES2N-Max-Np (SES) performed on average the best, 15.37% (M1) and 5.76% (M2) better than the second best, the GES2N-Mean-Np (SES), and 39.78% (M1), 12.16% (M2) better than the GES2N-Max-Nf (SES), the third best. Using the LES-based metrics, the GES2N-Max-Np (LES) performed the best, 5.53% (M1), 6.74% (M2) better than the GES2N-Mean-Np (LES) and 47.54% (M1), 13.50% (M2) better than the GES2N-Mean-Nf (LES) (the third best).

The SES-based GES2N performed on average 25.3% (M1) and 6.7% (M2) better than the corresponding LES2-based GES2N methods when using the SES-based metrics, but the LES-based GES2N objectives performed on average 47.0% (M1) and 18.7% (M2) better than the SES-based GES2N objectives when using the LES-based metrics. This shows the envelope spectrum used during the optimisation should also be used when performing the analyses. Lastly, the SES-based objectives' performance deteriorates as the noise becomes more impulsive. In contrast, the LES-based GES2N objectives perform better and should rather be used (at an increased computational cost). For the remainder of this work, the focus will only be on SES-based GES2N objectives to make the comparisons more focused, with LES-based GES2N objectives considered for future work.

5. Experimental results

The proposed method is compared to five optimal filtering methods on three experimental datasets with the process described in Section 3 and used in Section 4. An overview of the experimental test bench used to generate the vibration data is presented in Section 5.1, the results of the three datasets are presented in Sections 5.2–5.4, whereafter the results are discussed in Section 5.5.

5.1. Overview of setup

The helical gearbox test bench in the Centre for Asset Integrity Management (C-AIM) laboratory of the University of Pretoria is used. The test bench, shown in Fig. 7, consists of a 5.5 kW electrical motor that drives an alternator. Three helical gearboxes are used in the system, with the centre gearbox (gearbox 2) instrumented with accelerometers. The axial component of the tri-axial accelerometer, shown in Fig. 7(b), is used in subsequent sections. An optical probe with an 88 pulses per revolution zebra tape shaft encoder is used on gearbox 2's input shaft, as shown in Fig. 7. Geometrical defects in the zebra tape shaft encoder were corrected using Bayesian geometry compensation to find the rotational speed of the input shaft [43]. The gear connected to the gearbox 2's input shaft is damaged, so modulation from the gear damage is expected at 1.0 shaft orders. Hence, $\alpha_c = 1.0$ for this dataset. The optical probe signal is sampled at 51.2 kHz, and the tri-axial accelerometer signal is sampled at 25.6 kHz using an OROS OR35 data acquisition system.

The following three datasets are considered: Localised gear damage, distributed gear damage (case 1), and distributed gear damage (case 2). For the localised gear damage case in Section 5.2, a slot is seeded in a tooth of the gear. For the distributed gear damage cases, gear damage is distributed across the faces of multiple gear teeth, with case 2's damage in Section 5.4 being more severe than case 1's damage in Section 5.3.

In the next section, the dataset with localised gear damage is considered.

5.2. Localised gear damage

The experimental test rig acquired data from a gear with localised gear damage. The slot, shown in Fig. 8(a), was seeded in one of the teeth of gearbox 2's gear before the tests were started. The gear after the tests is shown in Fig. 8(b). Five signals with a duration of 5 seconds, acquired over the life of the gear, are considered from this dataset. The gearbox was not disassembled or inspected during the experimental study. Therefore, the exact condition of the gearbox for the measurements is unknown, but the severity of the damage was the least for measurement 1 and the most for measurement 5. Fig. 9 shows these five acceleration and rotational speed signals. The system operates under quasi-stationary regimes, with an approximate linear ramp-up between the regimes as shown in Fig. 9. Dominant extraneous impulses in the time domain signal manifest in the raw signals' SES. These

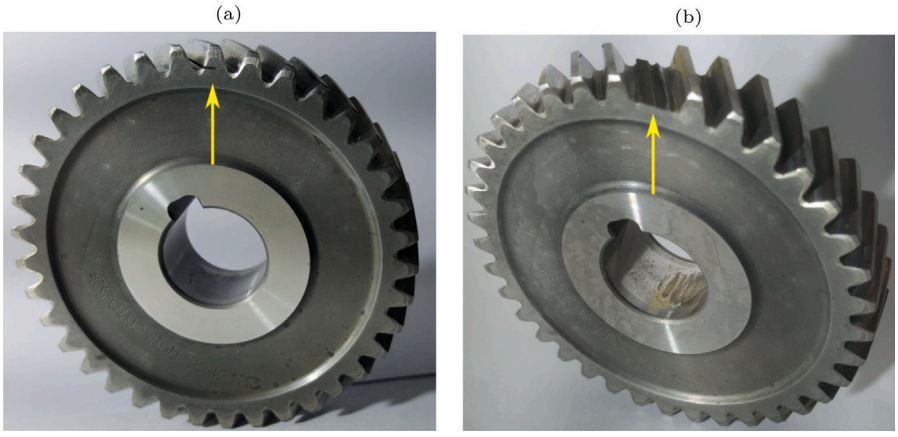


Fig. 8. The gear with localised damage is shown in 8(a) before the experiment was started and in 8(b) after the experiment was completed.

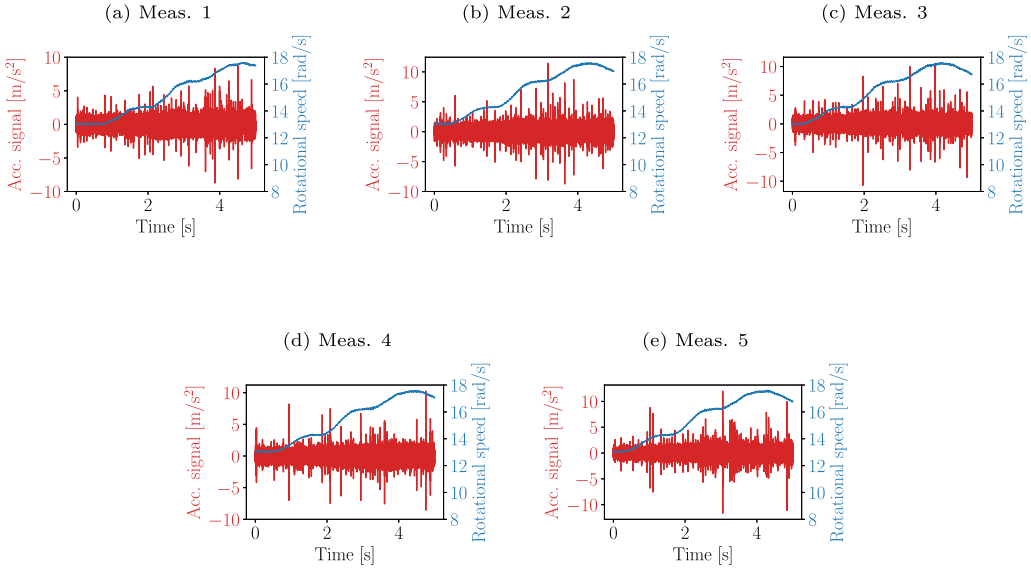


Fig. 9. The five measurements acquired from the gearbox test bench are shown for the localised gear damage case. The time domain vibration signal is shown in red with a corresponding *y*-axis on the left side, while the rotational speed signal is in blue and has a corresponding *y*-axis on the right side.

extraneous impulses occurred at approximately 5.72 shaft orders, and their exact source is unknown. The extraneous component was also present in the healthy measurements. This component impedes the performance of conventional time domain and envelope spectrum-based metrics.

The SES-based GES2N objectives and established methods were optimised on the signals in Fig. 9. The logarithm of the normalised Squared Envelope Spectra (SESa) of the filtered signals are presented in Fig. 10 with the corresponding metrics and selected filter length. The targeted component ($\alpha_c = 1.0$), the extraneous impulsive component ($\alpha_{ext} = 5.72$) and their harmonics are shown. The raw signals are dominated by the extraneous component at 5.72 orders, and the gear damage components at 1.0 order and its harmonics not seen. In contrast, the gear damage symptoms are visible in the SESa of all measurements when using the filter coefficients obtained with the GES2N-Max-Np and the GES2N-Mean-Np-based objective functions. For all measurements, both objectives' filtered results contain dominant components at 1.0 orders and their harmonics with M1 and M2 metrics between 0.75−1.00 and 0.64−1.00, respectively. The GES2N-Max-Nf and GES2N-Mean-Nf-based objective functions can only enhance the gear damage components for measurements 4 and 5 (i.e. when the damage is more pronounced), with M1s and M2s between 0.02−0.12 and 0.05−0.18 for measurements 1−3 and M1s and M2s between 0.51−0.84 and 0.43−0.70 for measurements 4 and 5. For measurements 4 and 5, the GES2N-Max-Np and GES2N-Mean-Np performed the same (M1s and M2s within 1%), but the GES2N-Max-Np performed using M1 and M2 on average 40.67% and 48.63% better than the GES2N-Max-Nf and 41.56% and 51.45% better than the GES2N-Mean-Nf.

The ACYCBD could only enhance the damage in measurement 5, with an M1 of 0.73 (the GES2N-Max-Np performed 36.97% better) and an M2 of 0.83 (the GES2N-Max-Np performed 18.18% better). The L_2/L_1 could only enhance the first harmonic for

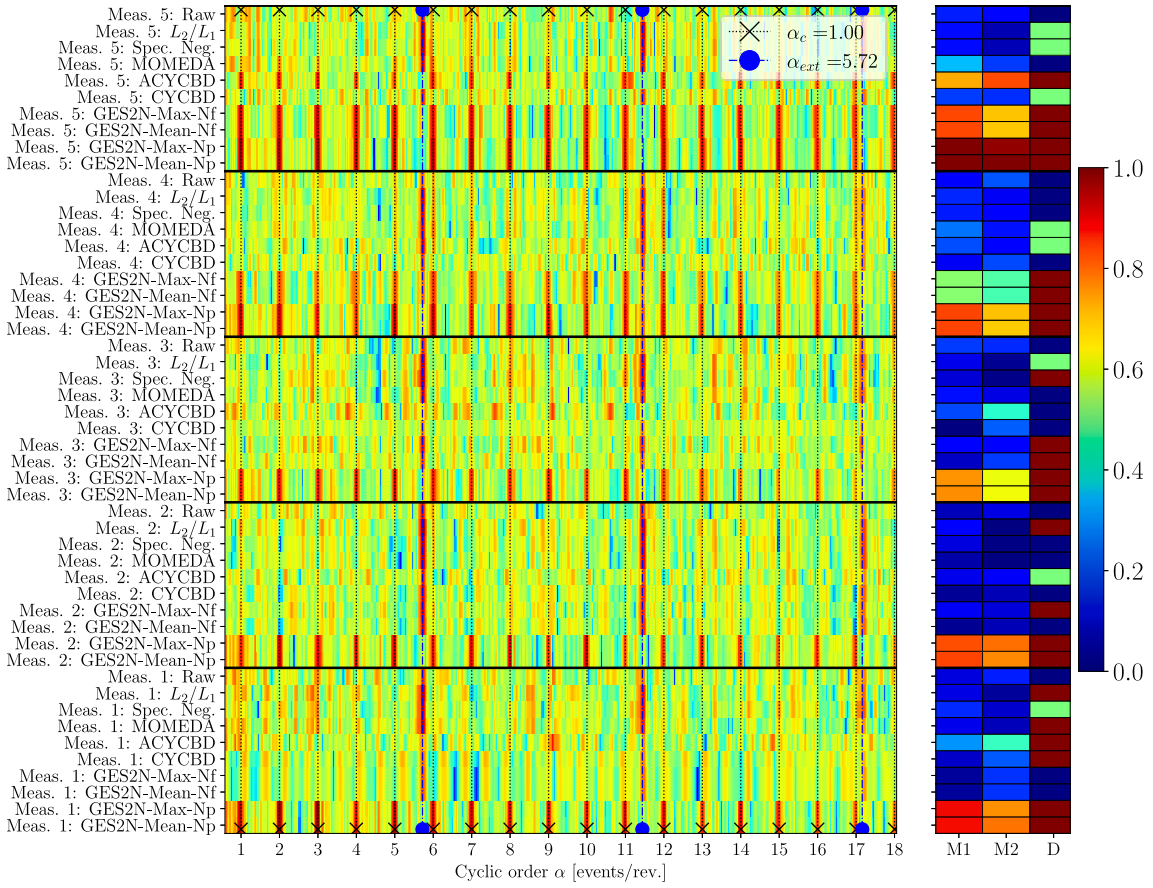


Fig. 10. The logarithm of the median normalised Squared Envelope Spectra (SESa) of the filtered signals $\boldsymbol{y}$ for the gear in Fig. 8 are presented using the same procedure as Fig. 5. Only the SES-based GES2N objectives are presented; therefore, explicit reference to SES is avoided for brevity's sake.

measurement 2 (M1 of 0.12, 85.22% worse than the GES2N-Max-Np), but enhanced the extraneous component also which resulted in the lowest M2 score (0). The CYCBD (M1s and M2s ranging between 0.00–0.19 and 0.05–0.22), MOMEDA (M1s and M2s ranging between 0.04–0.31 and 0.00–0.18) and the spectral negentropy (M1s and M2s ranging between 0.08–0.17 and 0.01–0.12) could not enhance the damage in any of the measurements. The spectral negentropy and L_2/L_1 objectives enhance sparse components in the SES and do not target an a priori selected component, which enhanced the extraneous component (5.72 orders).

The magnitude of the L_2-normalised optimal filter coefficients' spectra can be used to understand which spectral frequency bands' content maximises the objective function (i.e., which bands contain the diagnostic information). The magnitude of the optimal filter coefficients' spectra is shown in Fig. 11 with the power spectral densities of the raw signals also shown as a baseline. Some dominant frequency bands are highlighted with vertical markers.

The raw signals' power spectral densities in Fig. 11 (denoted Raw) highlight multiple high-energy frequency bands. The spectra of the best-performing filter coefficients (i.e., GES2N-Max-Np and the GES2N-Mean-Np) in Fig. 11 contain multiple dominant frequency bands that are not dominant in the power spectral density of the raw signal (denoted Raw). Hence, the damage does not manifest in the dominant frequency bands. Furthermore, the GES2N-Max-Np and GES2N-Max-Np objective functions' filters contain prominent amplitudes at 0.6 kHz, 4.0 kHz and 6.7 kHz. When enhancing the 0.6 kHz band, the ACYCBD (measurement 5) and the GES2N-Max-Nf and GES2N-Max-Nf (measurements 4–5) could enhance the gear fault signatures.

The GES2N-Max-Nf and GES2N-Max-Nf enhanced the higher frequency bands (above 11 kHz) for measurements 1–3, which the CYCBD also enhances for all measurements. This is likely caused by the inclusion of the mean squared energy in the objectives' noise indicator (or denominator). The L_2/L_1 enhanced the 6.7 kHz band for measurement 2 resulting in slight enhancement of the first gear harmonic in Fig. 10. In Fig. 10, there is a variation in the filter length selected for the objectives. However, the objectives that resulted in the enhanced damage used the longest considered filter length ($D = 256$).

The next section considers the first distributed gear damage case.

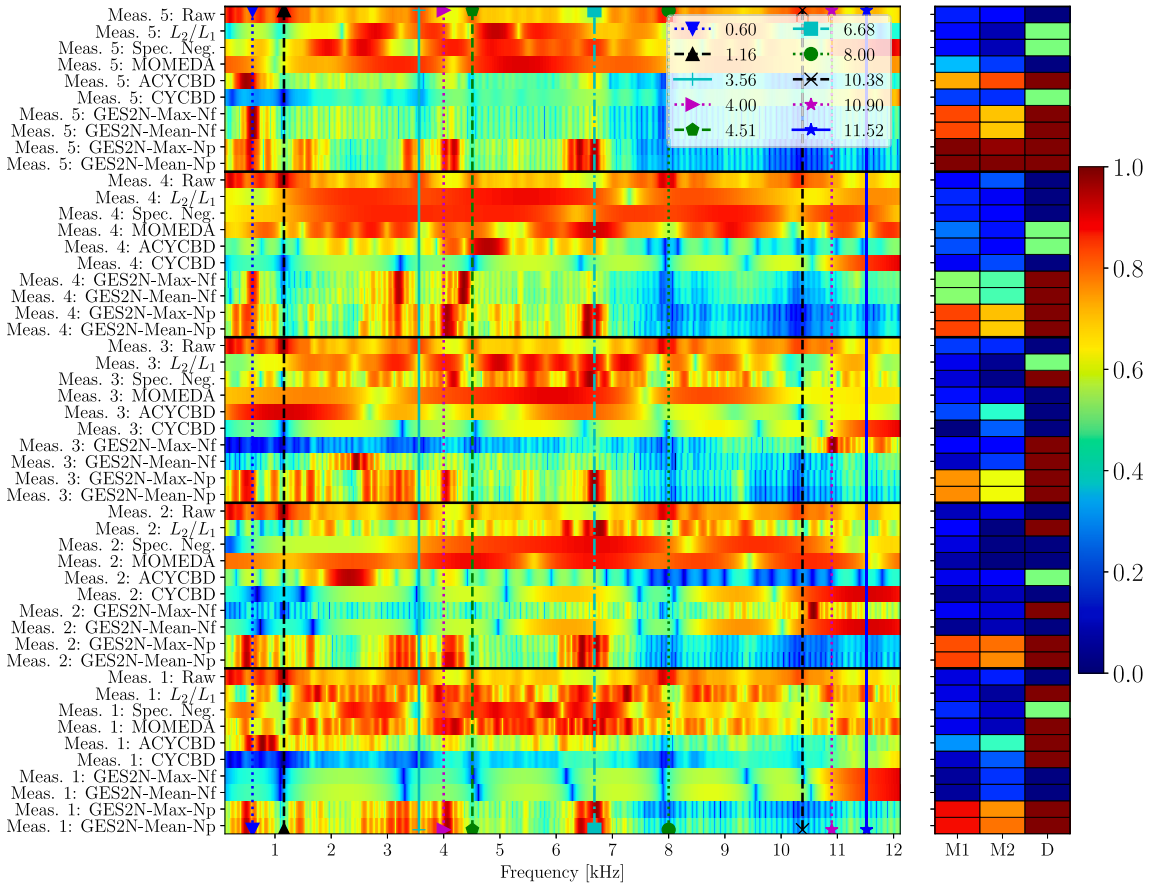


Fig. 11. The normalised spectra of the optimised filter coefficients are presented. The figures refer to the normalised power spectral densities of the raw signals as *Raw*. The L_2-norm of the spectrum was used to normalise the respective spectrum. Dominant frequencies (in kHz) are identified and marked in the legend to aid the discussion. The metrics are repeated here for the reader's convenience.

5.3. Distributed gear damage: Case 1

Experimental measurements were acquired from the gear shown in Fig. 12. The damage was distributed over half of the gear teeth as shown by the highlighted section in Fig. 12(a). The damage was introduced across the respective teeth's faces with a Dremel rotary tool, and the teeth did not deteriorate for the considered measurements. The damaged teeth are shown in Fig. 12(b). Fig. 13 shows the four measurements considered in this dataset, each with its own distinct rotational speed profile. The first measurement's speed profile is similar to the speed profiles in the previous section; the second measurement contains a linear ramp-up, the third measurement contains a slight speed fluctuation, and the last one considers a predominantly speed-down scenario. Since the gear teeth are in the same condition, the results will provide insights into the effect of speed on the performance of the methods. The impulses with a cyclic order of 5.72 are visible in the raw signal and more dominant than the weak damage components and, therefore, are expected to impede the performance of the designed filters.

The same procedure as Sections 4 and 5.2 is used to obtain the SESa of the filtered signals. The logarithm of the median normalised SESa of the filtered signals, their corresponding metrics and the selected filter lengths are shown in Fig. 14 for the different measurements.

The GES2N-Max-Np and GES2N-Mean-Np performed the best, with the damage enhanced for all measurements and with M1s and M2s between 0.72–1.00 and 0.75–1.00 for measurements 1, 3 and 4 and M1s and M2s between 0.56–0.57 and 0.49–0.55 for measurement 2 (the most challenging case). GES2N-Max-Nf and GES2N-Mean-Nf only show evidence of damage symptoms for measurement 4 (M1 of 0.74 and M2 of 0.81) but performed 25.57% and 18.74% worse than the GES2N-Max-Np and GES2N-Mean-Np objectives using M1 and M2 respectively. The L_2/L_1 has weak evidence of the first gear harmonic in measurement 1's envelope spectrum, with an M1 of 0.36 (62.74% worse than GES2N-Max-Np) and an M2 of 0.11 (85.71% worse than GES2N-Max-Np). The first harmonic is also present in the SES of measurement 2 for the MOMEDA, the spectral negentropy and the L_2/L_1 with M1s of 0.29, 0.32 and 0.26 respectively and M2s of 0.11, 0.16 and 0.1 respectively. The GES2N-Max-Np performed 75.29% better for M1 and 222.98% better for M2 of the spectral negentropy (the best-performing method between MOMEDA and the L_2/L_1). This highlights that despite enhancing the damage components (M1), the extraneous component is also dominant relative to the fault



Fig. 12. The gear with distributed damage (case 1) is shown. (a) The section of teeth that are damaged is highlighted in red, i.e., half the gear's teeth are damaged. (b) The damaged teeth are shown.

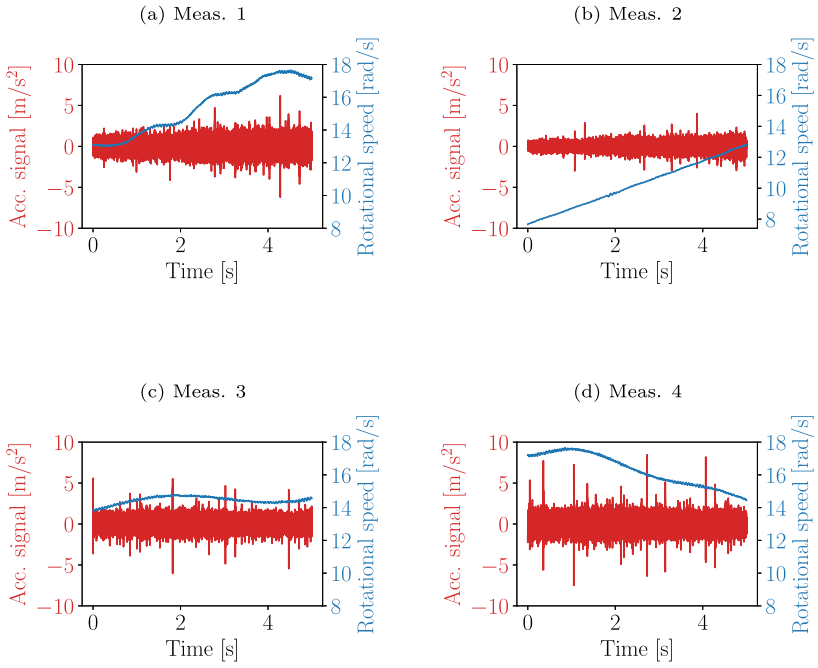


Fig. 13. Four measurements acquired from the gearbox with the gear's condition shown in Fig. 12. The time domain vibration signal is shown in red, with its y-axis shown on the left side and the corresponding rotational speed of the input shaft is shown in blue with a corresponding y-axis on the right side.

harmonics (M2). The CYCBD (M1s and M2s between 0.00 − 0.17 and 0.03−0.22 respectively) and the ACYCBD (M1s and M2s between 0.13−0.24 and 0.04−0.58 respectively) did not perform well in enhancing the gear damage for any measurements.

In Fig. 10, the longest filter length ($D = 256$) is used in the cases where the damage is enhanced. However, the results highlight that the same filter length did not consistently perform the best for the methods (e.g., GES2N-Max-Nf alternated between window lengths of 256 and 64).

For brevity's sake, the frequency spectra of the filter coefficients are not included. However, the results corroborate the previous dataset's results; the best-performing filter coefficient's spectra contain dominant spectral frequencies that do not align with the power spectral density dominant frequency bands, highlighting that the gear damage manifests in weaker spectral frequency bands.

5.4. Distributed gear damage: Case 2

The second distributed gear damage dataset was obtained with the gear shown in Fig. 15. This gear is the same gear considered in the previous section, but the severity of the damaged teeth was further increased with a Dremel rotary tool, i.e., the healthy teeth were not damaged. Hence, the gear damage is much more prominent than the previous dataset. Four measurements are considered in this section with the same rotational speed profiles as the previous dataset, and therefore, the time domain signals are not included for brevity's sake. Since the gear damage is much more prominent than the distributed gear damage in case 1, the different objective functions are expected to perform much better on this dataset.

The methods were applied to the four measurements, with the logarithm of the normalised SESa of the filtered and raw signals shown in Fig. 16 with their respective metrics. The raw signal has weak damage symptoms in measurements 1, 3 and 4 with the 1.0 shaft order and some of its harmonics visible.

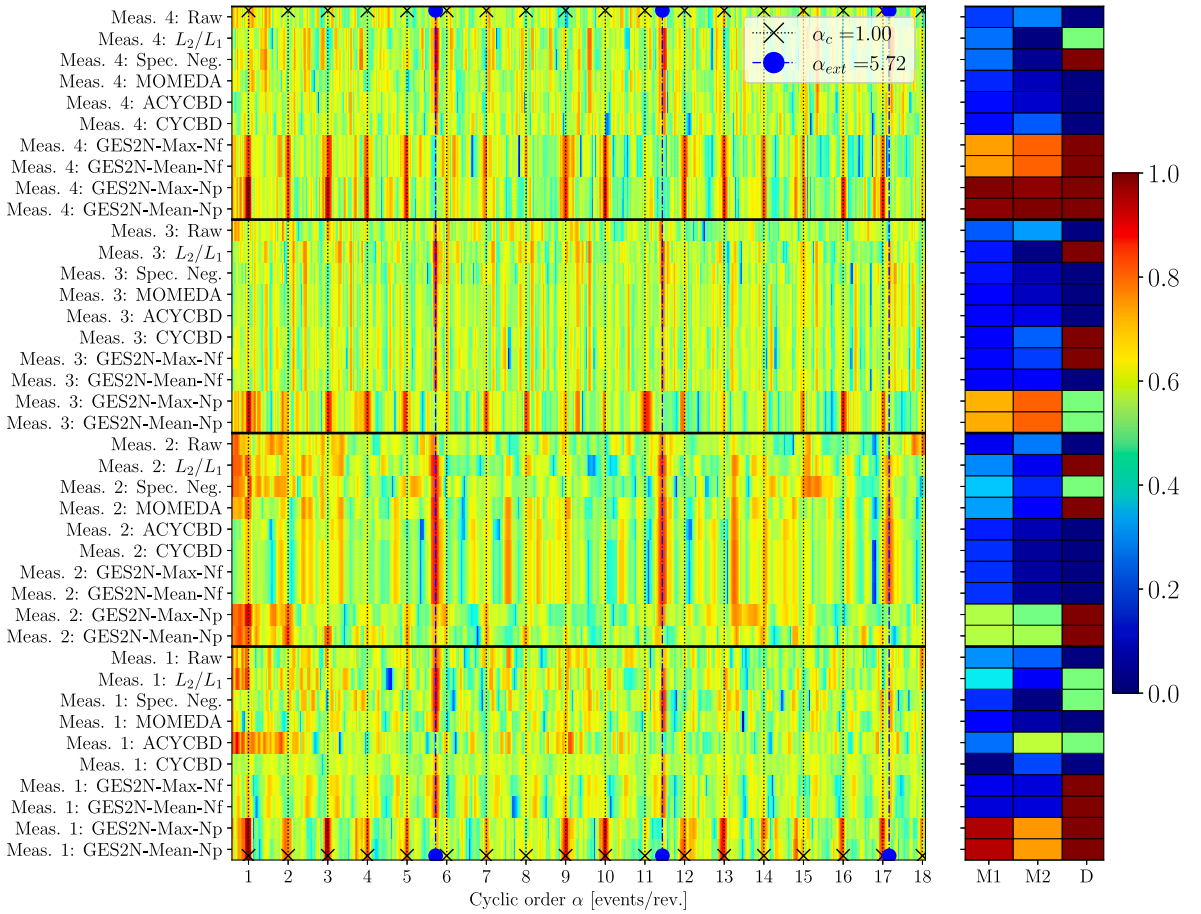


Fig. 14. The logarithm of the normalised Squared Envelope Spectra (SESa) of the filtered signals $\boldsymbol{y}$ for the gear in Fig. 12 are presented with the logarithm of the normalised SESa of the raw signals (denoted Raw). The same representation is used in Fig. 10.



Fig. 15. The gear with distributed damage (case 2) is shown. The gear in Fig. 12 was reused, but the teeth's damage increased.

GES2N-Max-Np and GES2N-Mean-Np perform the best; these objectives enhanced the damage for the four measurements with their metrics ranging between 0.65−1.00 (M1) and 0.85−1.00 (M2) over the four measurements. The GES2N-Max-Np and the GES2N-Mean-Np's metrics differ on average less than 2%. GES2N-Mean-Nf also enhanced the damage in all measurements, but its M1 and M2 metrics are on average 7.91% and 7.36% lower than GES2N-Mean-Np. GES2N-Max-Nf only enhanced the damage for measurements 1, 3 and 4 with the metrics M1 and M2 for these measurements ranging between 0.83−0.98 and 0.85−0.96, respectively. GES2N-Max-Nf's M1 and M2 metrics are on average 5.34% and 5.18% lower than the GES2N-Max-Np for measurements 1, 3 and 4. The ACYCBD could only enhance damage for Measurement 3 (0.82 and 0.85 respectively), whereas the CYCBD could only enhance the damage for

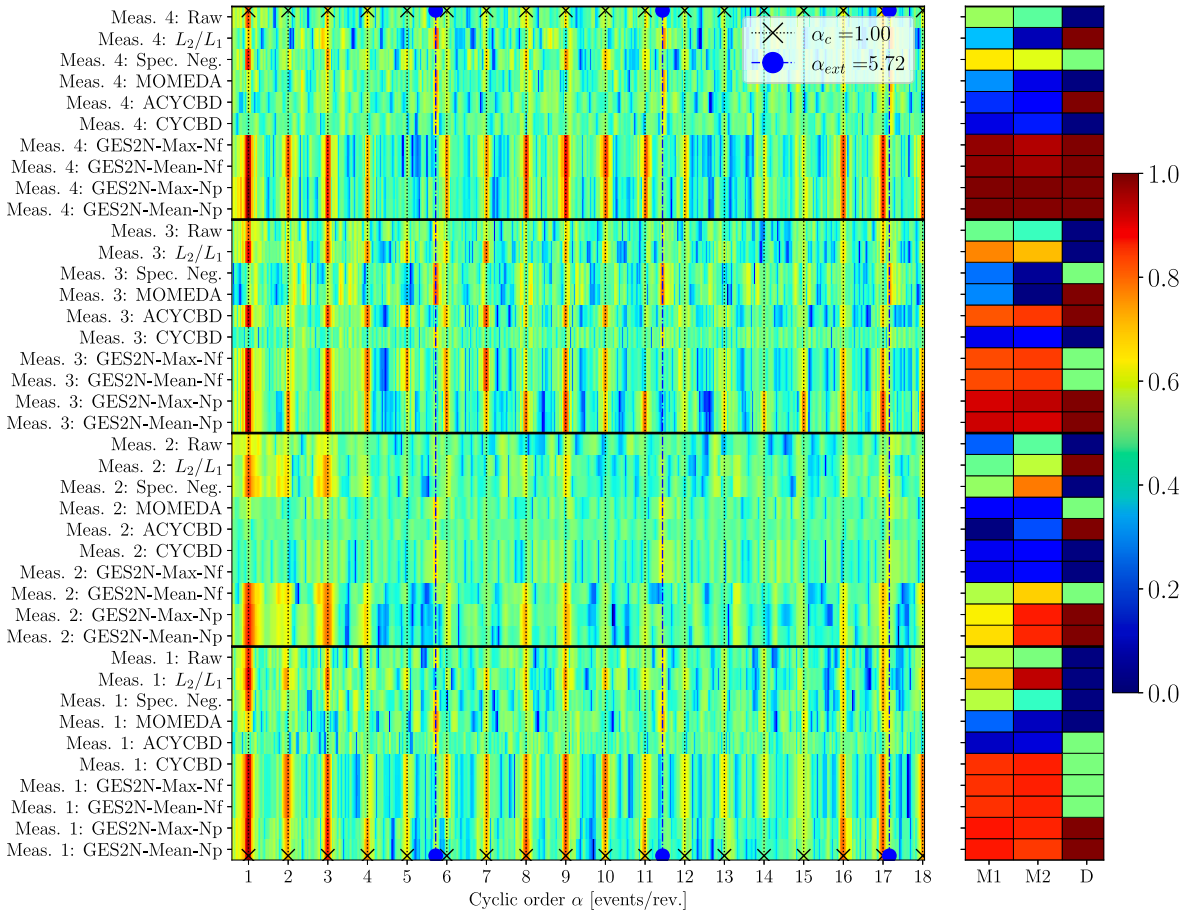


Fig. 16. The logarithm of the normalised Squared Envelope Spectra (SESa) of the filtered signals $\boldsymbol{y}$ for the gear in Fig. 15 are presented with the logarithm of the normalised SESa of the raw signals (denoted Raw). The damage manifests at $\alpha = 1.0$ and its harmonics. The same representation is used as Fig. 10.

Measurement 1 (M1 and M2 of 0.86 and 0.88). The L_2/L_1 could detect damage for measurements 1–3, with M1s and M2s between 0.47–0.77 and 0.58–0.95, respectively. On average this is 21.25% and 16.83% lower than GES2N-Max-Np. The spectral negentropy only enhanced the damage for measurement 1–2 and 4 with the metrics M1 and M2 between 0.53–0.66 and 0.41–0.78. This is on average 29.49% and 33.98% lower than GES2N-Max-Np. Lastly, MOMEDA could not enhance the damage for any cases with its M1 and M2 metrics between 0.12–0.27 and 0.00–0.13.

The filter lengths in Fig. 16 show that different filter lengths are optimal for different methods. For example, for Measurement 3 a filter length of 16 is optimal for L_2/L_1, while GES2N-Max-Nf performed better with a filter length of 64, and GES2N-Max-Np with 256. All of them still enhanced the fault signatures.

5.5. Discussion

GES2N-Mean-Np and GES2N-Max-Np enhance the fault signatures for all measurements and consistently perform the best with the metrics. GES2N-Mean-Np and GES2N-Max-Np only have slight differences between them (on average less than 1% difference for metrics M1 and M2), but perform on average at least 60% better than GES2N-Max-Nf (the next best method) for experiments 1 and 2. For experiment 3, the GES2N-Mean-Np and GES2N-Max-Np perform 7% better than the GES2N-Mean-Nf (the next best).

The results demonstrate that using the limited cyclic orders (GES2N-Mean-Np and GES2N-Max-Np) is better for weak damage enhancement, with the GES2N-Mean-Nf performing on average 49.39% (M1) and 48.47% (M2) worse than the GES2N-Mean-Np and GES2N-Max-Nf performing on average 53.18% (M1) and 53.08% (M2) worse than GES2N-Max-Np.

CYCBD, ACYCBD, and MOMEDA generally performed worse than the GES2N-based objectives, with the fault signatures only enhanced for some of the measurements when the damage was more prominent. Averaged over all experiments, the M1 and M2 metrics for CYCBD are 82.06% and 74.15% lower than the metrics for GES2N-Max-Np, whereas the ACYCBD metrics are 72.29% and 63.62% lower and the MOMEDA metrics are 77.84% and 90.00% lower. Similarly to GES2N-Max-Nf and GES2N-Mean-Nf, these methods use the mean energy of the signal in the denominator, highlighting that incorporating lower cyclic orders could impede weak fault signature enhancement.

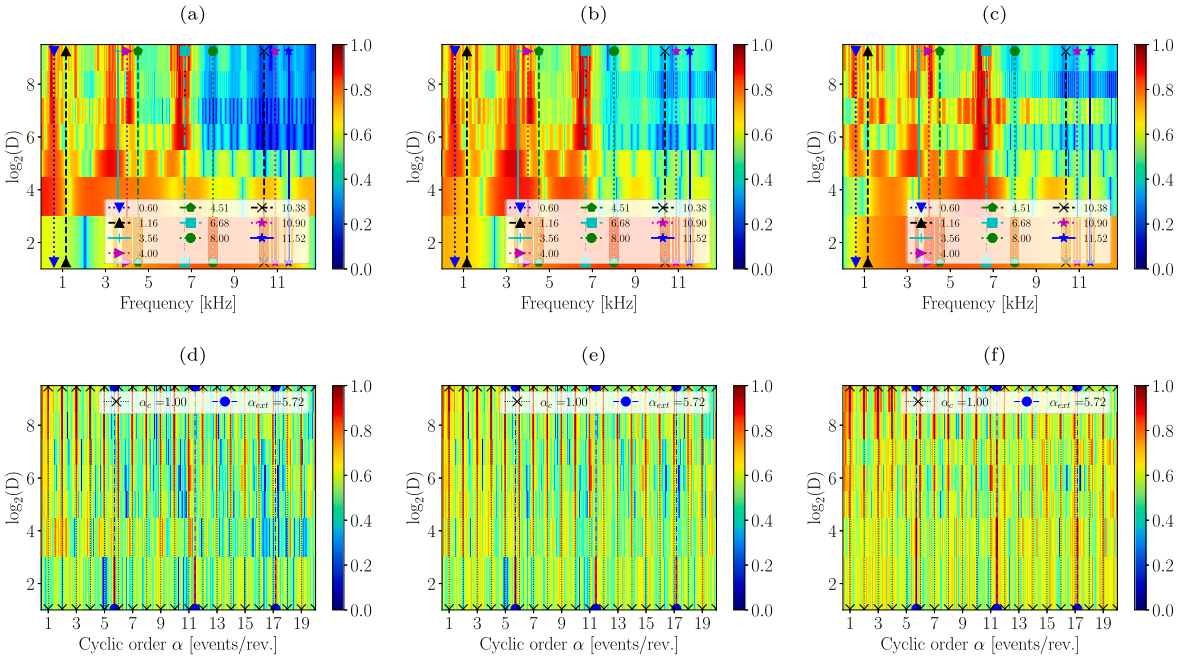


Fig. 17. The performance of the GES2N-Max-Np objective function using different parameters for the cyclic order band's width $\Delta\alpha_b$ and the resolution of the cyclic orders $\Delta\alpha$ as a filter length D. The top row shows the frequency response of the optimised filtering coefficients, and the bottom row shows the SES of the filtered signal. The following settings are shown: 17(a) & 17(d): $\Delta\alpha_a = 0.1, \Delta\alpha = \Delta\alpha_{baseline}$; 17(b) & 17(e): $\Delta\alpha_a = 0.1, \Delta\alpha = 0.5\Delta\alpha_{baseline}$; 17(c) & 17(f): $\Delta\alpha_a = 0.62, \Delta\alpha = 0.5\Delta\alpha_{baseline}$.

The L_2/L_1 and spectral negentropy objectives enhanced the damage for only some of the measurements, typically where the damage was more prominent. The envelope spectra corresponding to the L_2/L_1 objectives' M1 and M2 metrics are on average 64.29% and 75.74% lower than the GES2N-Max-Np and the spectral negentropy's M1 and M2 metrics are on average 67.31% and 78.97% lower than GES2N-Max-Np. These methods did not target prior-selected components and often enhanced the extraneous components. The performance of these objectives for the most severe damage case (Section 5.4) highlights that these methods can be utilised to detect more severe damage components if the fault component's cyclic order is unknown beforehand. It is challenging to define all the targeted cyclic orders beforehand (for example, the cyclic orders of all the failure modes of the components). It is also computationally expensive to design filters for each targeted cyclic order. Hence, this work highlights that further work is needed to utilise blind objectives for filter design.

6. Primary sensitivity study

For the proposed approach, numerous selections are required, including the filter length, filter initialisation, the targeted cyclic order, the number of harmonics N_h, the cyclic order bands used to estimate the signal and noise components respectively (i.e., cyclic order range and bandwidth), the cyclic order resolution, the weighting of the cyclic orders and the cyclic order bands. Furthermore, the method's sensitivity to these selections should be quantified over multiple measurements. Therefore, an extensive sensitivity study is needed to investigate all parameters. However, this requires a dedicated study and will be done in future work. In this section, a limited but primary sensitivity study is performed of the two best-performing methods, namely, the GES2N-Max-Np and the GES2N-Mean-Np, to the filter length N_w, the cyclic order resolution of the SES $\Delta\alpha$ and the cyclic order bands' width $\Delta\alpha_b$.

In Figs. 17 and 18, the filter coefficients' frequency response and the SESa of the filtered signal are shown for measurement 2 of the localised gear damage dataset treated in Section 5.2. The results for GES2N-Max-Np and GES2N-Mean-Np are included in Figs. 17 and 18 respectively. Two cyclic order resolutions and cyclic order bands' bandwidth are considered: The cyclic order resolution of the SES is either equal to $\Delta\alpha = \Delta\alpha_{baseline}$ (where $\Delta\alpha_{baseline} = \pi/(\theta[L-1] - \theta[0])$ which was used for all previous studies) or it is equal to $\Delta\alpha = 0.5 \cdot \Delta\alpha_{baseline}$. Furthermore, the cyclic order bands' width is equal to 0.1 (i.e., localised closely to the target component) or $\Delta\alpha_b = 0.62$ so that the extraneous component at 5.72 shaft orders is included in the sixth harmonic's cyclic order band.

Fig. 17 shows that the GES2N-Max-Np performs better with longer filter lengths than shorter ones. Fig. 17(a) shows that if the filter length is short, the filter's frequency response has limited flexibility and therefore cannot effectively enhance the narrow frequency bands and attenuate the extraneous frequency bands. The SES corroborates this in Fig. 17(d), which contains extraneous amplitudes and the targeted amplitudes are not dominant. If the filter length is longer, the filter's frequency response has sufficient flexibility to enhance the narrow frequency bands as shown in Fig. 17(a), which results in fault information rich SESa in Fig. 17(d)

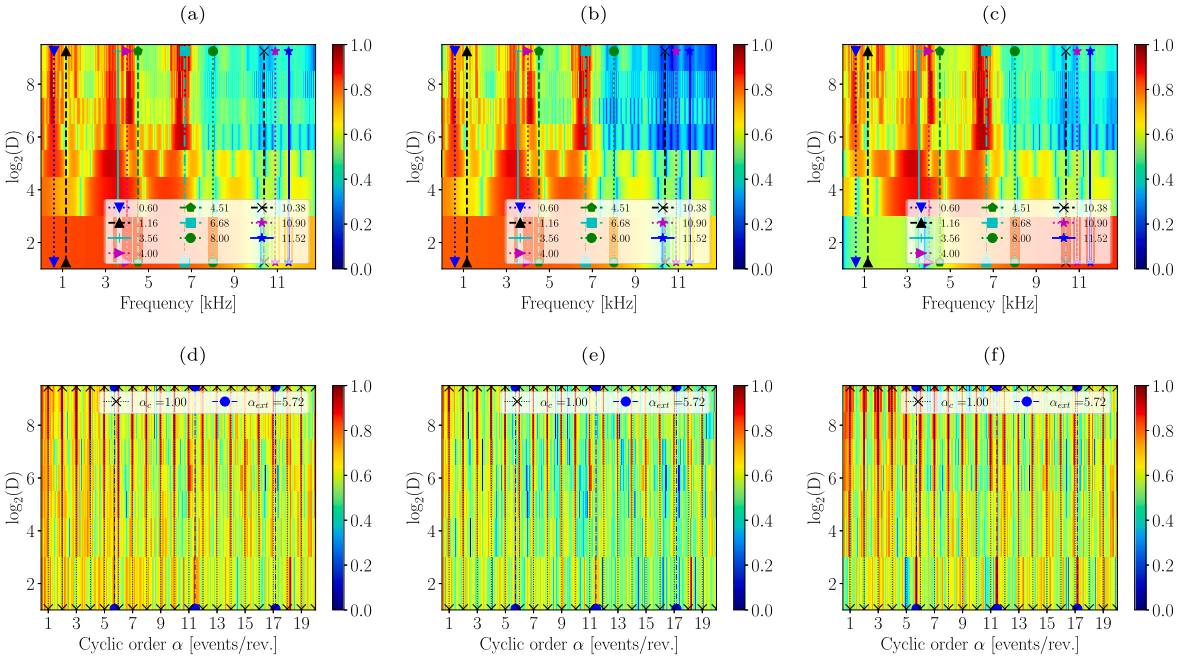


Fig. 18. The performance of the GES2N-Mean-Np objective function using different parameters for the cyclic order band's width $\Delta\alpha_b$ and the resolution of the cyclic orders $\Delta\alpha$ as a filter length D. The top row shows the frequency response of the optimised filtering coefficients, and the bottom row shows the SES of the filtered signal. The following settings are shown: 18(a) & 18(d): $\Delta\alpha_a = 0.1, \Delta\alpha = \Delta\alpha_{baseline}$; 18(b) & 18(e): $\Delta\alpha_a = 0.1, \Delta\alpha = 0.5\Delta\alpha_{baseline}$; 18(c) & 18(f): $\Delta\alpha_a = 0.62, \Delta\alpha = 0.5\Delta\alpha_{baseline}$.

at the expense of increased computational cost. When comparing the results in Figs. 17(a), 17(b), 17(d), and 17(e), the cyclic order resolution used in the study is sufficient to improve the gear damage components effectively.

The cyclic bands' width has a more observable impact on the results in Figs. 17(c) and 17(f). If the bands' width is too large, the extraneous impulse at 5.72 shaft order is in the band used in the numerator. Hence, if the maximum amplitude in the spectrum is used in the objective function and the filter length is short, the optimal filter coefficients maximise the extraneous component. However, if the filter length is sufficiently long, the gear damage components and the 5.72 shaft order components manifest in the SES. This was not investigated, but if the cyclic bands' width is too narrow, it can impede detection if the analytical cyclic order α_c differs from the actual cyclic order in the signal and is not within the targeted band.

The results of GES2N-Mean-Np are shown in Fig. 18, with similar observations made relative to GES2N-Max-Np regarding the filter length and the cyclic order resolution in Figs. 18(a), 18(d), 18(b) and 18(e). However, its performance differs from the GES2N-Max-Np when using broad cyclic order bands. Since the GES2N-Mean-Np calculates the mean amplitude in the numerator, the extraneous impulse at 5.72 orders and the 6th harmonic of the damage component at 6.0 orders contribute to the objective function, and, therefore, the damage information is also enhanced. This is seen in Fig. 18(f) where the gear components are much more enhanced than Fig. 17(f) for short filter lengths.

7. Conclusions

A generalised envelope spectrum-based signal-to-noise objective is proposed, from which various objective functions can be derived. The proposed generalised objective is derived for time-varying speed conditions. It enables the derivation of new objective functions, such as the GES2N-Max-Np, and it is simple to accommodate other envelope spectra (e.g., LES). The performance of four SES-based GES2N objectives and four LES-based GES2N objectives is compared against five established methods (CYCBD, ACYCBD, MOMEDA, the spectral negentropy and L_2/L_1 of the envelope spectrum) using signals from a foundational model. The results demonstrate that the proposed GES2N-based objectives perform better in enhancing the targeted fault signatures under impulsive and speed-varying conditions. Furthermore, the four SES-based GES2N objective functions are compared against established methods for gear damage detection under time-varying speed conditions. The squared envelope spectra, the filters' frequency response functions, and metrics extracted from the squared envelope spectra are used for comparison. These results demonstrate that the proposed functions perform better than the other methods in enhancing gear damage under time-varying operating conditions. This is because the proposed generalised objective quantifies the signal-to-noise ratio in the squared envelope spectrum, thereby capturing the prominence of the fault signatures in the squared envelope spectrum.

For future work, we recommend a more extensive comparison against other objectives, where the objective functions can be applied to different damage modes and damaged components. The work was focused on gears and although the method is,

in principle, well-suited for bearings, the parameters might need to be updated for cases where deviations from the theoretical values are expected. Further investigations are also needed to investigate the selection of robust and efficient optimisation settings (e.g., optimisers and initialisation strategies), and an extensive sensitivity study of the parameters and settings. More specifically, we recommend that the benefits of using the log-envelope spectrum in the GES2N objective be evaluated for non-impulsive and impulsive noise conditions. Furthermore, we recommend the following investigations for future work: The impact of different cyclic band weighting vectors selection strategies, the impact of different cyclic order band forms, e.g., $\Delta_b = k \cdot \Delta_b^{\text{base}}$, automatic filter length selection strategies, and the impact and selection of the number of harmonics for fault enhancement.

CRediT authorship contribution statement

Stephan Schmidt: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Daniel N. Wilke:** Writing – review & editing, Visualization, Methodology, Investigation, Conceptualization. **Konstantinos C. Gryllias:** Writing – review & editing, Conceptualization.

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

S. Schmidt gratefully acknowledges the Research and Development Programme of the University of Pretoria, South Africa for supporting the research. K.C. Gryllias gratefully acknowledges the financial support provided by the FWO - Fonds Wetenschappelijk Onderzoek, Belgium - Vlaanderen under project G0A3123N.

Appendix A. Code and additional results

The following repository contains an implementation of the proposed objective function: https://gitlab.com/vspa/snrocl and additional results related to this paper.

Appendix B. Gradients

The objective function's gradient is defined in Appendix B.1 and the gradients of envelope spectra variants are defined in Appendix B.2.

B.1. Gradient of the objective function

The objective function, defined in Eq. (11), is repeated here for the reader's convenience:

$$\ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right) = \ln\left(\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}\right) - \ln\left(\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}\right). \tag{B.1}$$

The analytical gradient of this objective function with respect to the design variables, i.e., the filter coefficients, is defined as follows:

$$\frac{d}{d\boldsymbol{h}} \ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right) = \frac{d\boldsymbol{g}}{d\boldsymbol{h}}\left(\frac{d}{d\boldsymbol{g}} \ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right)\right), \tag{B.2}$$

where $\frac{d}{d\boldsymbol{h}} \ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right) : \mathbb{R}^{D\times 1} \mapsto \mathbb{R}^{D\times 1}$. The first gradient component in Eq. (B.2) is given by

$$\frac{d\boldsymbol{g}}{d\boldsymbol{h}} = \frac{\|\boldsymbol{h}\|_2^2 \cdot \boldsymbol{I} - \boldsymbol{h}\boldsymbol{h}^{\mathrm{T}}}{\|\boldsymbol{h}\|_2^3}, \tag{B.3}$$

with $\frac{d\boldsymbol{g}}{d\boldsymbol{h}} : \mathbb{R}^{D\times 1} \mapsto \mathbb{R}^{D\times D}$, and the identity matrix denoted $\boldsymbol{I} \in \mathbb{R}^{D\times D}$. The second gradient component in Eq. (B.2) is given by

$$\frac{d}{d\boldsymbol{g}} \ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right) = \frac{d\boldsymbol{b}}{d\boldsymbol{g}}\left(\frac{\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s}{\boldsymbol{w}_s^{\mathrm{T}}\boldsymbol{C}_s\boldsymbol{b}} - \frac{\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n}{\boldsymbol{w}_n^{\mathrm{T}}\boldsymbol{C}_n\boldsymbol{b}}\right)^{\mathrm{T}}, \text{ with } \frac{d}{d\boldsymbol{g}} \ln \psi\left(\boldsymbol{g};\boldsymbol{x}\right) : \mathbb{R}^{D\times 1} \mapsto \mathbb{R}^{D\times 1}, \tag{B.4}$$

if $\boldsymbol{w}_i$ and $\boldsymbol{C}_i$ are not dependent on the filter coefficients, i.e., $\frac{d\boldsymbol{w}_i}{d\boldsymbol{g}} = \boldsymbol{0}$ and $\frac{d\boldsymbol{C}_i}{d\boldsymbol{g}} = \boldsymbol{0}$ for the numerator and the denominator. Hence, for cases where the vectors and matrices are functions of the data, the sensitivity in Eq. (B.4) is incomplete. In Eq. (B.4), the gradient of the squared envelope spectrum is denoted $\frac{d\boldsymbol{b}}{d\boldsymbol{g}} \in \mathbb{R}^{D\times N_\alpha}$ and defined in Appendix B.2 for the variants of envelope spectra.

B.2. Gradients of envelope spectrum variants

The squared envelope spectrum of a linearly transformed signal is defined as follows:

$$\boldsymbol{b}^{(LT)} = \left(\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y})^* \odot (\boldsymbol{A}\boldsymbol{y})\right)\right)^* \odot \left(\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y})^* \odot (\boldsymbol{A}\boldsymbol{y})\right)\right), \tag{B.5}$$

with $\boldsymbol{A} \in \mathbb{C}^{L_y \times L_y}$. Eq. (B.5)'s gradient with respect to normalised filter coefficients is denoted $\frac{d\boldsymbol{b}^{(LT)}}{d\boldsymbol{g}} \in \mathbb{R}^{D \times N_\alpha}$ and calculated as follows:

$$\begin{aligned}\frac{d\boldsymbol{b}^{(LT)}}{d\boldsymbol{g}} = &\left[\left(\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y})^* \odot (\boldsymbol{A}\boldsymbol{y})\right)\right) \odot \left[\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y}) \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)^* + (\boldsymbol{A}\boldsymbol{y})^* \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)\right)\right]^*\right]^{\mathrm{T}} + \\ &\left[\left(\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y})^* \odot \boldsymbol{A}\boldsymbol{y}\right)\right)^* \odot \left[\boldsymbol{V}\left((\boldsymbol{A}\boldsymbol{y}) \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)^* + (\boldsymbol{A}\boldsymbol{y})^* \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)\right)\right]\right]^{\mathrm{T}},\end{aligned} \tag{B.6}$$

where $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{g}$, $\frac{d\boldsymbol{y}}{d\boldsymbol{g}} = \boldsymbol{X}^{\mathrm{T}}$ and $\boldsymbol{y} \odot \boldsymbol{X}$ denotes the Hadamard product between $\boldsymbol{y}$ and each column of $\boldsymbol{X}$, i.e., $\boldsymbol{y} \odot \boldsymbol{X} \in \mathbb{R}^{L_y \times D}$. The gradient of the SES in Eq. (2) is obtained by setting $\boldsymbol{A} = \boldsymbol{I}$ in Eq. (B.6).

The gradient of the log-envelope spectrum in Eq. (4), with respect to the normalised filter coefficients is denoted $\frac{d\boldsymbol{b}^{(LES)}}{d\boldsymbol{g}} \in \mathbb{R}^{D \times N_\alpha}$ and calculated as follows:

$$\begin{aligned}\frac{d\boldsymbol{b}^{(LES)}}{d\boldsymbol{g}} = &\left[\left(\boldsymbol{V}\,\boldsymbol{e}^{(LES)}\right) \odot \left[\boldsymbol{V}\left(\boldsymbol{u} \odot \left((\boldsymbol{A}\boldsymbol{y}) \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)^* + (\boldsymbol{A}\boldsymbol{y})^* \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)\right)\right)\right]^*\right]^{\mathrm{T}} + \\ &\left[\left(\boldsymbol{V}\,\boldsymbol{e}^{(LES)}\right)^* \odot \left[\boldsymbol{V}\left(\boldsymbol{u} \odot \left((\boldsymbol{A}\boldsymbol{y}) \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)^* + (\boldsymbol{A}\boldsymbol{y})^* \odot \left(\boldsymbol{A}\left(\frac{d\boldsymbol{y}}{d\boldsymbol{g}}\right)^{\mathrm{T}}\right)\right)\right)\right]\right]^{\mathrm{T}},\end{aligned} \tag{B.7}$$

where the log-envelope is given by $\boldsymbol{e}^{(LES)} = \mathbf{log}\left((\boldsymbol{A}\boldsymbol{y})^* \odot \boldsymbol{A}\boldsymbol{y}\right)$ with $\boldsymbol{e}^{(LES)} \in \mathbb{R}^{L_y \times 1}$. The vector $\boldsymbol{u} = \left(\mathbf{diag}\left((\boldsymbol{A}\boldsymbol{y})^* \odot \boldsymbol{A}\boldsymbol{y}\right)\right)^{-1}\mathbf{1}$ where $\boldsymbol{u} \in \mathbb{R}^{L_y \times 1}$, $\mathbf{diag}(\boldsymbol{x})$ denotes a diagonal matrix with diagonal elements given by $\boldsymbol{x}$ and $\mathbf{1} \in \mathbb{R}^{L_y \times 1}$ denotes a vector with each element equal to 1. The gradients were checked with the finite difference method.

Appendix C. Foundational model

The foundational model is defined in Eq. (14) with three components. The damage component is defined as follows:

$$x_{dam}(t) = \left(\frac{\omega(t)}{2 \cdot \pi \cdot 10}\right)^2 \cdot \left(\kappa \cdot \left(\sum_{k=1}^{K} a_i \cdot \delta\left(t - t_k^{toi}\right)\right) \otimes q(t)\right), \tag{C.1}$$

where $\kappa = 3$ is a scaling constant that controls the prominence of the component of interest and $\delta(x) : \mathbb{R} \mapsto \mathbb{R}$ denotes the Dirac delta function, i.e., $\delta(0) = 1$ and $\delta(x) = 0 \,\forall\, x \neq 0$, a_i is a zero-mean Gaussian distributed with a standard deviation of 0.1. The time-of-arrival of the kth impulse t_k^{toi} is dictated by the rotational speed $\omega(t)$ and the cyclic order, which is 2.0 events/rev. Slip is not simulated. The following impulse response function is used: $q(t) = e^{-500 \cdot t} \cdot \sin(2 \cdot \pi \cdot t \cdot f_s \cdot 0.25)$, where the sampling frequency is $f_s = 10$ kHz. The term $\left(\frac{\omega(t)}{2 \cdot \pi \cdot 10}\right)^2$ simulates amplitude modulation due to the varying speed. The extraneous stochastic component is defined as follows:

$$x_{ext}(t) = \left(\frac{\omega(t)}{2 \cdot \pi \cdot 10}\right)^2 \cdot \left(1.25 \cdot \sin\left(7.5 \cdot \int_0^t \omega(\tau) d\tau\right) \cdot z(t)\right), \tag{C.2}$$

where $z(t)$ is a standardised Gaussian random variable. The noise variable

$$x_{noise}(t) = \left(\frac{\omega(t)}{2 \cdot \pi \cdot 10}\right)^2 \cdot \left(\epsilon_{\mathrm{DOF}}(t)\right), \tag{C.3}$$

where $\epsilon_{\mathrm{DOF}}(t)$ is zero mean Student-t distributed with a scaling parameter of 1.0 and DOF degrees-of-freedom, with DOF > 0. Decreasing the DOF makes the noise variable more leptokurtic.

References

[1] P. Borghesani, P. Pennacchi, R. Randall, N. Sawalhi, R. Ricci, Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions, Mech. Syst. Signal Process. 36 (2) (2013) 370–384.
[2] S. Schmidt, P.S. Heyns, Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring, Measurement 149 (2020) 106964.
[3] J. Hebda-Sobkowicz, R. Zimroz, A. Wyłomańska, J. Antoni, Infogram performance analysis and its enhancement for bearings diagnostics in presence of non-Gaussian noise, Mech. Syst. Signal Process. 170 (2022) 108764.
[4] 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.

[5] Y. Miao, B. Zhang, J. Lin, M. Zhao, H. Liu, Z. Liu, H. Li, A review on the application of blind deconvolution in machinery fault diagnosis, Mech. Syst. Signal Process. 163 (2022) 108202.
[6] J. Antoni, The infogram: Entropic evidence of the signature of repetitive transients, Mech. Syst. Signal Process. 74 (2016) 73–94.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] C. Luo, Z. Mo, Q. Miao, Cyclic harmonic ratio defined in squared envelope spectrum and log-envelope spectrum for gearbox fault diagnosis, IEEE Trans. Instrum. Meas. 69 (12) (2020) 9568–9577.
[12] N. Sawalhi, R. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mech. Syst. Signal Process. 21 (6) (2007) 2616–2633.
[13] X. Jia, M. Zhao, Y. Di, P. Li, J. Lee, Sparse filtering with the generalized lp/lq norm and its applications to the condition monitoring of rotating machinery, Mech. Syst. Signal Process. 102 (2018) 198–213.
[14] C. Peeters, J. Antoni, J. Helsen, Blind filters based on envelope spectrum sparsity indicators for bearing and gear vibration-based condition monitoring, Mech. Syst. Signal Process. 138 (2020) 106556.
[15] 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.
[16] L. He, D. Wang, C. Yi, Q. Zhou, J. Lin, Extracting cyclo-stationarity of repetitive transients from envelope spectrum based on prior-unknown blind deconvolution technique, Signal Process. 183 (2021) 107997.
[17] B. Chen, W. Zhang, D. Song, Y. Cheng, F. Gu, A.D. Ball, Squared envelope sparsification via blind deconvolution and its application to railway axle bearing diagnostics, Struct. Health Monit. (2023) 14759217231151585.
[18] Z. Meng, Z. Cui, J. Liu, J. Li, F. Fan, Maximum cyclic Gini index deconvolution for rolling bearing fault diagnosis, IEEE Trans. Instrum. Meas. (2023).
[19] G.L. McDonald, Q. Zhao, Multipoint optimal minimum entropy deconvolution and convolution fix: application to vibration fault detection, Mech. Syst. Signal Process. 82 (2017) 461–477.
[20] M. Buzzoni, J. Antoni, G. d'Elia, Blind deconvolution based on cyclostationarity maximization and its application to fault identification, J. Sound Vib. 432 (2018) 569–601.
[21] E. Soave, G. D'Elia, M. Cocconcelli, M. Battarra, Blind deconvolution criterion based on Fourier–Bessel series expansion for rolling element bearing diagnostics, Mech. Syst. Signal Process. 169 (2022) 108588.
[22] B. Chen, W. Zhang, D. Song, Y. Cheng, Blind deconvolution assisted with periodicity detection techniques and its application to bearing fault feature enhancement, Measurement 159 (2020) 107804.
[23] B. Zhang, Y. Miao, J. Lin, Y. Yi, Adaptive maximum second-order cyclostationarity blind deconvolution and its application for locomotive bearing fault diagnosis, Mech. Syst. Signal Process. 158 (2021) 107736.
[24] Z. Wang, J. Zhou, W. Du, Y. Lei, J. Wang, Bearing fault diagnosis method based on adaptive maximum cyclostationarity blind deconvolution, Mech. Syst. Signal Process. 162 (2022) 108018.
[25] K. Liang, M. Zhao, J. Lin, J. Jiao, C. Ding, Maximum average kurtosis deconvolution and its application for the impulsive fault feature enhancement of rotating machinery, Mech. Syst. Signal Process. 149 (2021) 107323.
[26] D. Peng, X. Zhu, W. Teng, Y. Liu, Use of generalized Gaussian cyclostationarity for blind deconvolution and its application to bearing fault diagnosis under non-Gaussian conditions, Mech. Syst. Signal Process. 196 (2023) 110351.
[27] Q. Zhou, C. Yi, L. Yan, C. Huang, J. Lin, A blind deconvolution approach based on spectral harmonics-to-noise ratio for rotating machinery condition monitoring, IEEE Trans. Autom. Sci. Eng. 20 (2) (2022) 1092–1107.
[28] 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.
[29] D. Liu, L. Cui, H. Wang, Rotating machinery fault diagnosis under time-varying speeds: A review, IEEE Sens. J. (2023).
[30] J. Wodecki, A. Michalak, R. Zimroz, Local damage detection based on vibration data analysis in the presence of Gaussian and heavy-tailed impulsive noise, Measurement 169 (2021) 108400.
[31] P. Borghesani, J. Antoni, CS2 analysis in presence of non-Gaussian background noise–Effect on traditional estimators and resilience of log-envelope indicators, Mech. Syst. Signal Process. 90 (2017) 378–398.
[32] P. Borghesani, P. Pennacchi, S. Chatterton, R. Ricci, The velocity synchronous discrete Fourier transform for order tracking in the field of rotating machinery, Mech. Syst. Signal Process. 44 (1–2) (2014) 118–133.
[33] S. Schmidt, P.S. Heyns, K.C. Gryllias, A pre-processing methodology to enhance novel information for rotating machine diagnostics, Mech. Syst. Signal Process. 124 (2019) 541–561.
[34] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—A tutorial, Mech. Syst. Signal Process. 25 (2) (2011) 485–520.
[35] J.A. Snyman, D.N. Wilke, Practical Mathematical Optimization: Basic Optimization Theory and Gradient-Based Algorithms, in: Springer Optimization and Its Applications (SOIA), vol. 133, 2018.
[36] R.B. Randall, A new method of modeling gear faults, J. Mech. Des. 104 (1982) 259–267.
[37] V. Cizek, Discrete Hilbert transform, IEEE Trans. Audio Electroacoust. 18 (4) (1970) 340–343.
[38] S.G. Johnson, The NLopt nonlinear-optimization package, 2007, https://github.com/stevengj/nlopt.
[39] P.K. Mogensen, A.N. Riseth, Optim: A mathematical optimization package for Julia, J. Open Source Softw. 3 (24) (2018) 615, http://dx.doi.org/10.21105/joss.00615.
[40] 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.
[41] J. Bezanson, A. Edelman, S. Karpinski, V.B. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. 59 (1) (2017) 65–98.
[42] D. Abboud, S. Baudin, J. Antoni, D. Rémond, M. Eltabach, O. Sauvage, The spectral analysis of cyclo-non-stationary signals, Mech. Syst. Signal Process. 75 (2016) 280–300.
[43] 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.