ISA Transactions 95 (2019) 346–357

Contents lists available at ScienceDirect

ISA Transactions

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

Practice article

ACCUGRAM: A novel approach based on classification to frequency band selection for rotating machinery fault diagnosis

Zhiliang Liu [a,*], Yaqiang Jin [a], Ming J. Zuo [a,b], Dandan Peng [a]

[a] School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR China
[b] Department of Mechanical Engineering, University of Alberta, Edmonton T6G2G8, Canada

HIGHLIGHTS

- Importance of health reference has been revealed in FBS.
- Health reference, impulsiveness and cyclostationarity are firstly integrated.
- Classification accuracy is newly developed to indicate FBS for fault diagnosis.

ARTICLE INFO

Article history:
Received 16 May 2018
Received in revised form 5 May 2019
Accepted 10 May 2019
Available online 15 May 2019

Keywords:
Fault transient
Frequency band selection
Health reference
Classification
Rotating machinery fault diagnosis

ABSTRACT

Frequency band selection (FBS) in rotating machinery fault diagnosis aims to recognize frequency band location including a fault transient out of a full band spectrum, and thus fault diagnosis can suppress noise influence from other frequency components. Impulsiveness and cyclostationarity have been recently recognized as two distinctive signatures of a transient. Thus, many studies have focused on developing quantification metrics of the two signatures and using them as indicators to guide FBS. However, most previous studies almost ignore another aspect of FBS, i.e. health reference, which significantly affect FBS performance. To address this issue, this paper investigates importance of a health reference and recognize it as the third critical aspect in FBS. With help of the health reference, the frequency band where the fault transient exists could be located. A novel approach based on classification is proposed to integrate all three aspects (impulsiveness, cyclostationarity, and health reference) for FBS. Classification accuracy is developed as a novel indicator to select the most sensitive frequency band for rotating machinery fault diagnosis. The proposed method (coined by accugram) has been validated on benchmark and experiment datasets. Comparison results show its effectiveness and robustness over conventional envelope analysis, the kurtogram, and the infogram.

1. Introduction

Rotating machinery fault diagnosis is a multidisciplinary topic in the field of prognostics and health management. Vibration response has been proved of having a strong link with various faults of rotating machinery in operation [1]. Therefore, vibration based methods have been widely studied in the past decades for rotating machinery fault diagnosis. Among them, spectral analysis plays a central role to build relationship between a fault and its associated signatures. A conventional process of vibration based spectral analysis is implemented as follows. First, collect a piece of raw vibration signal by a data acquisition system settling on a target machine; second, extract envelope of this piece signal; third, obtain the envelope spectrum by applying Fourier transform to the extracted envelope signal; finally, fault characteristic frequencies are observed over the envelope spectrum to determine whether their associated faults occur or not. It is worth pointing out that the fault characteristic frequencies are usually modulated to a high frequency, e.g. resonance frequency or mesh frequency. Thus, fault diagnosis could be significantly improved if this fault associated frequency band is recognized and extracted out of the full band spectrum. This problem is termed as *frequency band selection* (FBS) in this paper. Some other terminologies are also used in the literature, e.g. selection of informative frequency band [1] and optimal band selection [2,3].

In essence, the task of FBS is to find a frequency band that has an intrinsic relationship with a machinery fault. In rotating machinery, incipient faults excite vibration transients when rotating, which leaves good hope of using transient characteristics to guide FBS. The two recognized characteristics of a transient are impulsiveness and cyclostationarity. Antoni firstly presented a representative work, i.e. spectral kurtosis [4,5] and kurtogram [6] with a fast implementation, which uses the temporal kurtosis

* Corresponding author.
E-mail address: Zhiliang_Liu@uestc.edu.cn (Z. Liu).

https://doi.org/10.1016/j.isatra.2019.05.007
0019-0578/

to measure the impulsiveness characteristic. Later on, Barszcz et al. [2] proposed protrugram that uses the kurtosis of the envelope spectrum to measure the cyclostationarity characteristic. More recently, Antoni [7] summarized evidence studies of transients and proposed a new FBS method, namely infogram, which considers both of the two characteristics and uses negentropy of the square envelope and its spectrum to depict the two characteristics of transients. The above studies are representative contributions made in FBS. Besides, improvements and extensions of the kurtogram, the protrugram, and the infogram can be found in many other papers, e.g. the improved kurtogram with robust local mean decomposition [8], the improved kurtogram with wavelet packet transform [9], the improved kurtogram with wavelet packet transform and power spectrum [10], the adaptive spectral kurtosis [11], PAR-based kurtogram [12], and FBS with genetic algorithm and fast kurtogram [13].

From the above literature review, most of the previous studies actually believe that an absolute value of their proposed indicator can lead to a successful FBS for fault diagnosis. However, it is not fully true because the indicators may be affected by other fault-irrelevant matters. That is, a fault occurrence can change the indicators in their absolute value, but an absolute change of the indicators may not be caused by the fault. Therefore, FBS may be misleading and further results in a wrong decision of fault diagnosis. Wang et al. [14] provided a promising solution that use a relative value rather than an absolute one in FBS. The reported SKRgram [14] uses spectral kurtosis ratio between the test signal and its baseline signal in FBS. However, this idea does not attract much attention mainly due to (1) it does not clearly point out importance of health reference and (2) it lacks of thorough validation and discussion about the health reference, where the terminology of *health reference* refers to vibration signal collected under operating condition without any faults. In rotating machinery fault diagnosis, FBS should consider three aspects: impulsiveness, cyclostationarity, and their health references. However, to our best knowledge, previous studies almost ignore the third aspect and most effort focuses on partial consideration of the three aspects in FBS; thus, rare studies have been found to unite all three aspects for a comprehensive FBS.

To address this challenge, this paper introduces health reference as a baseline to further enhance performance of FBS for rotating machinery fault diagnosis. With the aid of the health reference, changes from the health reference to the test signal could be more properly evaluated than the case without this information, and thus true frequency band related to machinery faults could be more accurately recognized. Keeping this spirit, we extensively investigate importance of a health reference and recognize it as the third critical aspect for FBS. A new method that can fuse all three aspects is proposed to make a robust evaluation on information amount of a frequency band candidate. The most sensitive frequency band can be selected based on their sensitivity scores.

This paper is organized as follows. Section 1 introduces background, motivation and a brief literature review of FBS. Section 2 introduces idea and implementation of the proposed method from four steps. Numerical validations on plenty of benchmark and experimental datasets are carried out in Section 3. Finally, Section 4 provides conclusions and remarks for this paper.

2. Methodology

The key principle of our method is to evaluate frequency band candidates out of a frequency spectrum and find the most sensitive one that is supposed to have the most significant fault evidences. Usually, a fault makes its intrinsically associated frequency band much different from others. If this frequency band is rightly recognized, fault diagnosis could benefit from less influence from other irrelevant frequency bands. In the proposed method, the sensitivity of a frequency band is measured by difference between signals under healthy and test conditions, and the concept of classification is then first employed to quantify this difference. In other words, the most sensitive frequency band mostly results in the largest classification accuracy (defined in Section 2.3), and vice versa. It is worth pointing out that a classification method is used to generate accuracy guiding the selection of frequency band; that is, the accuracy metric quantifies fault information in each band and it is not directly related to decision making of what fault mode occurs in rotating machinery. From the above statements, the proposed method introduces classification into frequency band selection and it has the ability to make use of information from healthy condition data.

According to Fig. 1, the proposed method consists of four steps: signal segmentation (Step 1), frequency band partition & feature extraction (Step 2), accuracy representation & frequency band selection (Step 3), and fault diagnosis (Step 4). Step 1 obtains enough number of segments for classification. Step 2 generates a set of frequency band candidates and then calculates specific features for each candidate. Step 3 calculates classification accuracy for every frequency band and then selects the most sensitive frequency band according to accuracy criterion. The test signal is filtered with the selected frequency band. Step 4 is the last step that looks for fault evidences (e.g. fault characteristic frequency) by means of squared envelope analysis and provides a decision of fault diagnosis. The details of the four steps are introduced as follows.

2.1. Signal segmentation

The main purpose of this step is to divide a whole piece of raw signal into a number of segments with the same time period. We next introduce a sliding segmentation approach as an option to implement this idea. Suppose we have a collected signal $x[n]$ with N samples ($n = 1, 2, \ldots, N$) at a sampling frequency f_s, we then define another three parameters: $L_{overlap}$ = the length of sample overlap for two neighbor segments; n_{seg} = the number of segments; and L_{seg} = the length of each segment. Their relationship is shown in (1). Once the aforementioned four parameters (N, $L_{overlap}$, n_{seg}, and L_{seg}) are specified, the sliding signal segmentation can be applied to the whole signal $x[n]$. After that, n_{seg} individual signals are generated for the following processing of the proposed method. Obviously, this step is very useful if the number of signal segments is limit; on the other hand, if you already have segmented individual signals, this step is not that necessary.

$$N = (n_{seg} - 1) \times (L_{seg} - L_{overlap}) + L_{seg} \tag{1}$$

2.2. Frequency band partition & feature extraction

2.2.1. Frequency band partition

This step aims to generate frequency band candidates for the following evaluation. Frequency band is generated by low-pass/band-pass/high-pass filters with specified central frequency and bandwidth, which are defined in Eqs. (2) and (3), respectively. The input of the frequency band partition is the full spectrum, i.e. from zero to a half of the sampling frequency (f_s). We follow the same frequency band partition as in the kurtogram [6] for a fast implementation. The principle of the partition algorithm is described in the case of composite filtering structure based on bisection and trisection decompositions. The following decomposition are based on the two elementary decompositions. It is repeated in a pyramidal manner to produce the tree of filterbands shown in Fig. 2. We define the level j that is equal to

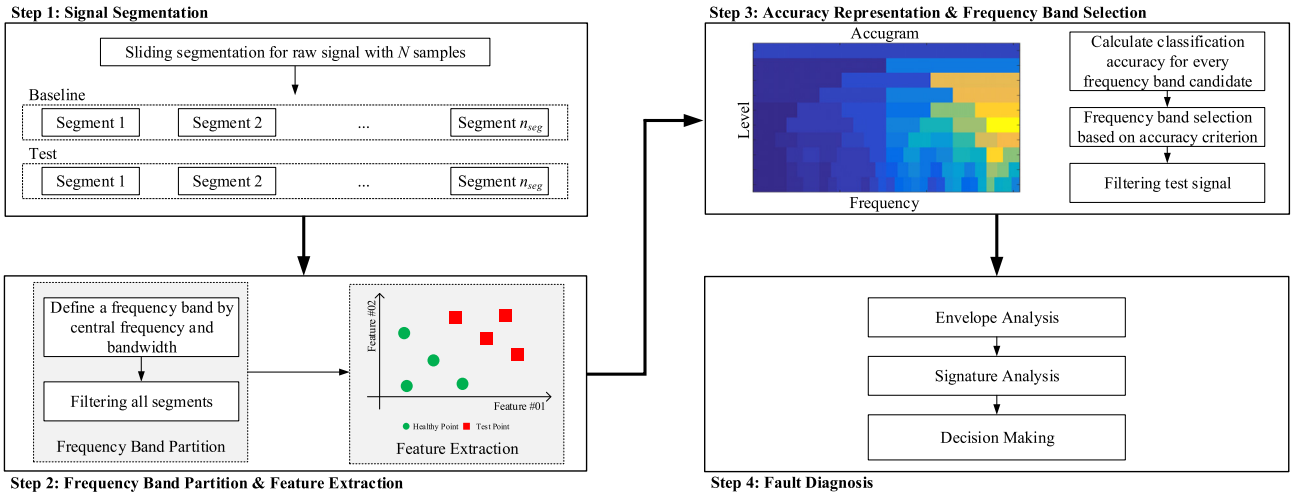

Fig. 1. Scheme diagram of the proposed method.

base-2 logarithm of the number of subsections splitting in the full spectrum. Taking the level 1 for example, it requires to divide the full spectrum into quartered sections. It applies the bisection decomposition to the lower and upper sections obtained from the level 1, respectively. And then X_2^0 and X_2^1 are generated from X_1^0; X_2^2 and X_2^3 are generated from X_1^1. Note that this algorithm only filters the signal that locate in the defined decomposition layers. In the jth level, the central frequency f_c and its associated bandwidth Δf of each frequency band can be calculated by

$$f_c = (i + \frac{1}{2}) \times 2^{-j-1} \times f_s, \text{ and} \tag{2}$$

$$\Delta f = 2^{-j-1} \times f_s, \tag{3}$$

where i = the band subsection index of each level; f_s = the sampling frequency; and j = the level of the filtering structure

2.2.2. Feature extraction

The main purpose of this step is to extract transient evidences containing in a signal segment as a set of features. Feature extraction is a crucial preprocessing step for a successful accuracy representation. There already have many features available in literature for rotating machinery fault diagnosis, e.g. clearance factor and FM4 [15]. Among them, entropy is a versatile indicator that measures disorder of a system. Negentropy, a variation of entropy, is defined as negative entropy that reaches its maximum value if energy condenses to a single impulse. To demonstrate the proposed idea, we hereby extract two features, i.e. negentropy of the squared envelope (SE) and negentropy of the squared envelope spectrum (SES). The two features have been recently proposed to characterize transients by Antoni [7]. They can reveal information hidden in the signal from the time domain (using the squared envelope) and the frequency domain (using the envelope spectrum) viewpoints. The two features are most recently advanced to characterize impulsiveness and cyclostationary of a transient. Their effectiveness has been proved in many reported papers [16–18]. Performance evaluation of other features are out of scope of this paper. This paper mainly focuses on how to integrate impulsiveness, cyclostationary, and health reference for frequency band selection. Of course, more advanced features could be developed in the future. If having these more advanced features, we can still use them in the proposed method. Since it does not relate to the contribution of this paper, we have not compared performance with different features.

We next introduce how to extract negentropy of SE and SES for a segment filtered to a specific frequency band. SE negentropy (denoted by NE_{SE}) [7] is introduced first and its definition is expressed in (4). It is used to measure single transient impulsiveness of a signal. The transient impulse occurrence in the signal leads to a high NE_{SE} value. Therefore, NE_{SE} is a promising indicator to distinguish the faulty signal from its health reference.

$$NE_{SE} = \frac{1}{L_{seg}} \sum_{n=1}^{L_{seg}} \frac{\varepsilon_x[n]}{\overline{\varepsilon_x}} \ln \frac{\varepsilon_x[n]}{\overline{\varepsilon_x}}, \tag{4}$$

where ε_x is the square of squared envelope and $\overline{\varepsilon}_x$ is the arithmetic mean of the ε_x. The squared envelope is equal to $|a[n]|^2$, where $a[n]$ is the analytic signal of $x[n]$ defined by Hilbert transform.

SES negentropy (denoted by NE_{SES}) [7] is the other extracted feature and it is defined in (5) NE_{SES} is similar to NE_{SE}. It is used to reflect cyclic transient impulsiveness in a signal. The cyclic transient impulses in the time domain, in fact, is a single transient impulse in its transformed frequency domain. Therefore, NE_{SES} value increases sharply if the signal presents the cyclic property of repetitive transient impulses. It is another promising indicator to distinguish the faulty signal from its health reference.

$$NE_{SES} = \frac{1}{L_{seg}} \sum_{k=1}^{L_{seg}} \frac{E_x[k]}{\overline{E_x}} \ln \frac{E_x[k]}{\overline{E_x}}, \tag{5}$$

where E_x is energy spectrum of ε_x and $\overline{E}_x$ is the arithmetic mean of the E_x. The energy spectrum can be estimated by squared amplitude spectrum from discrete Fourier transform.

2.3. Accuracy representation & frequency band selection

2.3.1. Accuracy representation

As we already mentioned in the beginning of Section 2, the proposed method assumes that monitoring data under healthy and test conditions should be available. Section 2.2 introduces how to generate frequency band candidates and how to extract information in the two types of data. This section shows how to fuse their information to evaluate sensitivity of a frequency band candidate. Classification techniques are used to achieve this purpose. Classification accuracy (0–1) then becomes an indicator to measure how much information a frequency band contains for rotating machinery fault diagnosis. This is the most distinguishable characteristic of the proposed method comparing to other reported methods. Thank to this idea, it will show later that it

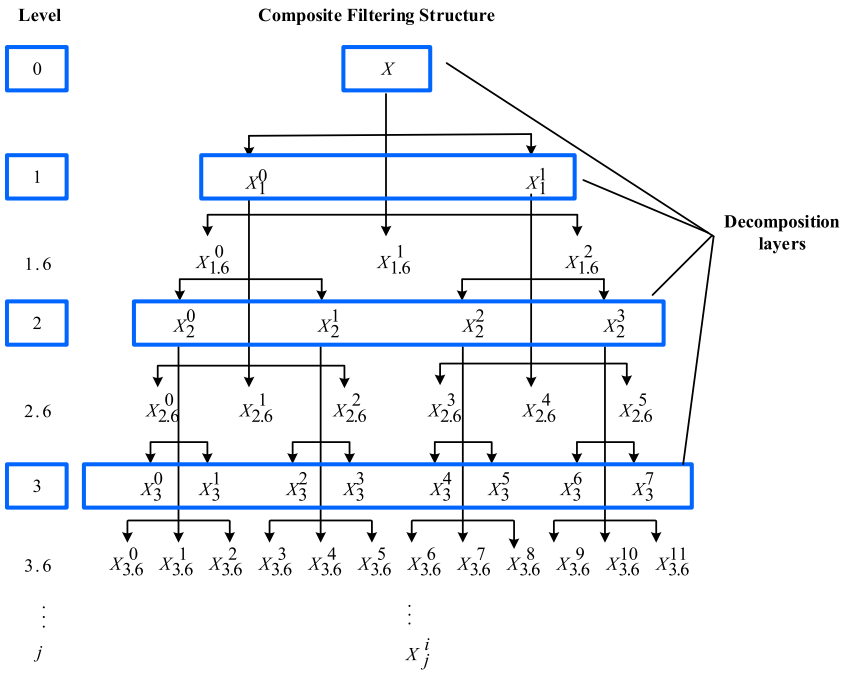

Fig. 2. The composite filtering structure based on the two elementary decompositions.

selects a more sensitive frequency band that more likely leads to a successful fault diagnosis.

Classification accuracy is calculated as follows. Suppose the healthy condition signal carried the true healthy condition and the test signal having unknown health condition. n_{seg} segments are generated after applying the same sliding segmentation process for each signal. The n_{seg} segments are corresponding to n_{seg} classification samples. Frequency band candidates can be generated by the frequency band partition introduced in Section 2.2. For one frequency band candidate, we can extract feature values using the features introduced in Section 2.2 from each of the two datasets described in the previous paragraph. Therefore, we can generate a dataset: a $2\times n_{seg}$-by-2 feature matrix and its associated $2\times n_{seg}$-by-1 label vector. In the feature matrix, $2\times n_{seg}$ is the number of classification samples, and 2 is the number of extracted features. We assume the same number of samples of the two classes to avoid the so-called class imbalance problem [19]. In the label vector, we use 0 and 1 to represent healthy condition and test condition, respectively. Obviously, this is a two-class classification problem due to two classes existing in the label vector. *k*-nearest neighbor (*k*NN) with standard *K*-fold cross validation is employed to solve this two-class classification problem and obtain its associated classification accuracy. *k*NN is a non-parametric method used for classification. *k* is the only tuning parameter and it is set by one according to [20–22]. *K* is set by 10 for cross validation, that is, 90% samples are used for training and the rest 10% samples are used for testing.

Because the true label of the test signal is unknown, we need to justify the above assumption of defining a different label from the health reference. The label of the test signal has two cases: (1) the true label is 0, which means the assumption wrong; (2) the true label is 1, which means the assumption correct. In the case 1, the training data of the two labels are very similar; the classification model will be definitely confused; and thus the test accuracy will be low for every frequency band candidate. No matter which band we select, the fault characteristic frequency will not be found in the envelop spectrum and thus the diagnosis result is "no fault", which matches the true label for this case. In the case 2, the classification model will be confused for the fault-irrelevant frequency bands, but it is assumed to perform well for the fault-relevant frequency band where data from the two classes are quite different. Therefore, the test accuracy is high for the fault-relevant frequency bands and is low for the fault-irrelevant frequency bands. Based on the above analysis, we demonstrate that the assumption does not affect FBS strategy and the following decision making of fault diagnosis. Based on this valid assumption, the classification accuracy is defined with the predicted labels and our assumed labels (not the true labels).

2.3.2. Frequency band selection

Frequency band selection is introduced as follows. After we complete accuracy calculation for all frequency band candidates that are defined in Section 2.3, a classification accuracy matrix is generated as shown in Fig. 3, where $Accuracy_j^i$ is the classification accuracy for the *i*th partition in the *j*th level. This matrix can be interpreted as a diagram as the kurtogram does. So the diagram is coined by *Accugram* (short for accuracy diagram) in this paper. The strategy to determine the most informative frequency is very straightforward. The most informative frequency band is corresponding to the highest classification accuracy. That is, classification accuracy is the proposed indicator for FBS. A tie problem may exist if there are more than one frequency bands having the same accuracy, e.g. the upper bound of classification accuracy (i.e. 100%). If this problem occurs, further evaluation needs to be done for all frequency bands with same accuracy. We calculate Euclidean distance between center arithmetic means of cluster of the healthy state and cluster of the test state. The tied frequency band with a larger Euclidean distance is selected as the most sensitive frequency band.

We hereby discuss influence of classification methods to FBS performance. According to the accuracy criterion, the proposed method uses only ordinal accuracy relationship between frequency bands rather than their real accuracy values. Although classification methods may generate different real accuracy values for the same frequency band, it does not change the final selection result if their ordinal relationships still hold. It means that the proposed method is not sensitive to classification method. We next use an example to further illustrate this point. The dataset

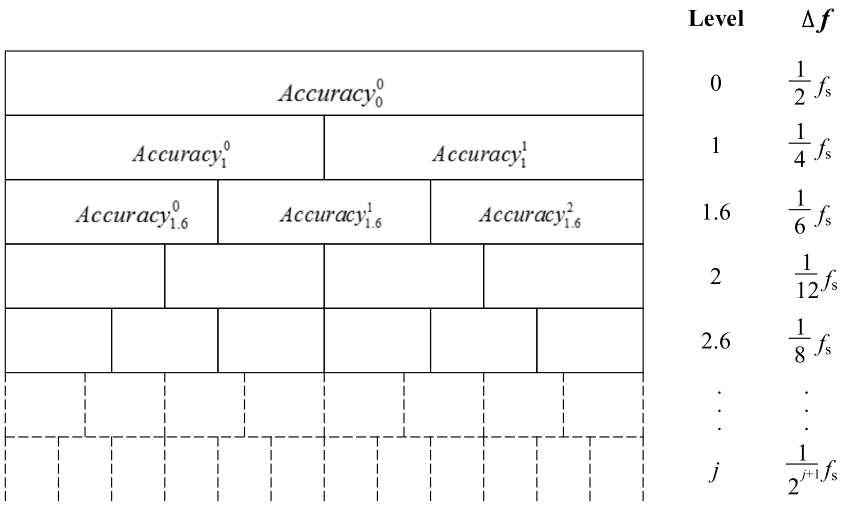

Fig. 3. Classification accuracy matrix.

used here is the dataset #177 (FE) that will be introduced in Section 3. We generate 108 frequency band candidates following the procedure of the proposed method. The 108 frequency bands are evaluated by three classifiers that are *k*NN, support vector machine (SVM), and Naive Bayes classifier. The experimental results are presented in Fig. 4. According to this figure, accuracy for all frequency bands varies a bit between the three classifiers. The ordinal relationship of the maximum accuracy does not change for the three classifiers. Thus, no matter which classifier is selected, the optimal frequency band is always the 33th candidate. This example shows that classifier has no influence on the proposed method.

At the end of frequency band selection, the optimal combination of central frequency and bandwidth is selected. The test signal is filtered with the optimal parameters and then it is ready for the following fault diagnosis.

2.4. Fault diagnosis

Many machinery faults, e.g. rolling element bearing and gear defects, excite vibration resonance at a specific frequency [2]. The resonance frequency and its associated sidebands are the so-called most sensitive frequency band. The previous sub-sections introduce how to recognize this frequency band and obtain a narrow-band signal. However, it usually cannot make a decision for machinery fault diagnosis, and a further step is still needed.

Envelope analysis is quite useful to signature analysis for rotating machinery fault diagnosis. It aims to analyze envelope of a signal rather than the signal itself. Because fault characteristic frequency is usually modulated to a high resonance frequency, envelope analysis can be used to demodulate the fault characteristic frequency. The general idea of envelope analysis based signature analysis is as follows. First, Hilbert transform is used to obtain envelope of the narrow-band signal obtained from Section 2.2. Squared envelope spectrum, introduced by Ho and Randall [11], can highlight the defect frequency in the spectrum and degrade the other noise components. Therefore, in the second step, the squared envelope spectrum is generated by applying discrete Fourier transform to the squared envelope signal. Third, signature analysis is conducted on the extracted squared envelope signal. A conventional approach to implement signature analysis is to find fault characteristic frequencies on the squared envelope spectrum. If amplitude of a fault characteristic frequency increases, it indicates a fault associated to this frequency occurs in the system. For example, if one finds a significant increase of ball pass frequency of outer ring (BPFO), occurrence of a local fault on the outer ring of this bearing can be interpreted for this system. Of course, further signature analysis can be done to make a more accurate diagnosis decision. However, this is out of scope of this paper.

3. Validation

This section aims to validate the proposed method on two kinds of datasets, i.e. benchmark dataset and experimental dataset. The benchmark dataset is downloaded from Case Western Reserve University Bearing Data Center [23]. It is open accessed so that readers of this paper can easily repeat reported results that will be given in the following subsections. The second experimental dataset is from a planetary gearbox test rig located in Equipment Reliability, Prognostics and Health Management (ERPHM) Laboratory at University of Electronic Science and Technology of China (UESTC). It helps to further validate generalization performance of the proposed method. We compare the proposed method with three reported methods [4,7,24] to show its superiority over existing development in rotating machinery fault diagnosis. Table 1 gives main profiles of the four comparison methods.

3.1. Benchmark bearing dataset

In this section, ball bearing data from the Case Western Reserve University Bearing Data Center are benchmark datasets that we use to validate the proposed method. More details about the benchmark datasets are available in [23]. Here, we randomly select datasets associated to inner race fault with two fault diameters (0.18 and 0.36 mm) at the drive end (DE) for validation. Their dataset numbers are 109, 110, 111, 112, 174, 175, 176, and 177. In addition, the proposed method requires datasets of health reference under four load conditions. Their dataset numbers are 97, 98, 99, and 100. Thus, 12 datasets in total are used in this section and their numbers are consistent with [23].

Based on the test bearing specifications, the characteristic frequency of the inner race fault (f_{BPFI}) at the DE is equal to $5.4152\times f_n$, where f_n is the shaft rotating frequency. According to the conventional theory of bearing fault diagnosis, it aims to locate the characteristic frequency and observe its change in amplitude. Next, we apply all four methods in Table 1 to find the most sensitive frequency band and then observe the characteristic frequency of the inner race fault and its corresponding amplitude

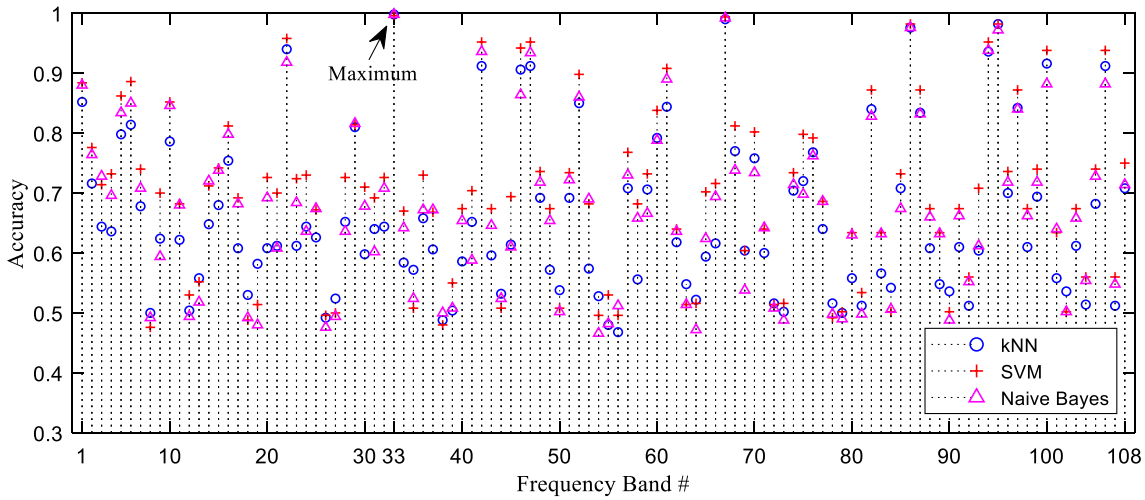

Fig. 4. Comparison of accuracy representation by three classifiers.

Table 1
Four comparison methods of frequency band selection.

Notation	Method	Descriptions
M1	Envelope analysis of the raw signal	Squared envelope spectrum of the full bandwidth raw signal without any frequency band selection.
M2	Kurtogram [4]	Squared envelope spectrum of the selected frequency band with maximum spectral kurtosis.
M3	Infogram [7]	Squared envelope spectrum of the selected frequency band with maximum negentropy of the squared envelope (ΔI_ε), the squared envelope spectrum (ΔI_E), and their average ($\Delta I_{1/2}$).
M4	The proposed method (Accugram)	Squared envelope spectrum of the selected frequency band with maximum classification accuracy.

Table 2
Three diagnosis category.

Diagnosis category	Diagnosis success	Interpretation
Y	Yes	Data clearly diagnosable and showing unambiguous characteristic frequency in the spectrum.
P	Partial	Data probably diagnosable, e.g., the characteristic frequency and its harmonics are buried or not dominant in the spectrum.
N	No	Data not diagnosable for the specified bearing fault and virtually indistinguishable from noise in the spectrum.

in an envelope spectrum. We use three categories to qualitatively evaluate performance of fault diagnosis. Descriptions about the three categories are listed in Table 2.

In the following, we analyze and compare all four methods in terms of two aspects, fault diagnosis result and selected frequency band, from a statistical viewpoint. Table 3 summarizes bearing fault diagnosis results for all four methods. We count numbers of "Y", "P", and "N" for each method in this table. From the results, infogram, a powerful tool recently proposed by Antoni, does make a step forward for frequency band selection because it has seven "Y", a larger number than that in M1 and M2. One may notice that ΔI_E and $\Delta I_{1/2}$ work better than ΔI_ε in M3. A possible reason is that the cyclostationary characteristic is more important than impulsiveness in rotating machinery fault diagnosis. Another evidence of the above statement is that spectral kurtosis, considering only the impulsive characteristic, performs worse among the four methods on the tested benchmark datasets even though it can help to narrow down frequency band. Results for the dataset #174 are meaningless because this dataset likely has corrupted due to some unknown reasons, e.g. mechanical looseness interpreting in [24]. In general, the proposed method makes a further progress over existing development because of the fact that (1) it leads the most number of "Y" for all datasets, that is, it has a better generalization ability than other three methods; and (2) it make all datasets clearly diagnosable from either DE or fan end (FE) side except the dataset #174, that is, it can achieve accurate fault diagnosis. This conclusion shows a benefit from health reference consideration.

Table 4 summarizes selected frequency bands from the four methods in terms of center frequency (f_c) and bandwidth (Bw). Because the sampling frequency is 48 kHz, the frequency band limit is 24 kHz in spectrum and combination of (12 000, 24 000) actually means a full frequency band with no frequency band selection. From the results, M1 obviously performs squared envelope spectrum with a full bandwidth signal. One interesting observation is that M2 and M3 sometime select a full frequency band in the datasets #109 and #176, and these cases are highlighted with underlines in the table. No doubt, they have the same fault diagnosis results as M1 in Table 3. These results indicate that M2 and M3 sometime fail to recognize the most sensitive frequency band as initially expected, but the proposed method can always narrow down the full frequency band. From the perspective of selected bandwidth, the proposed method possesses the smallest value (1054 Hz) for all the datasets except the dataset #174. The average bandwidth for M2, M3 (ΔI_ε), M3 (ΔI_E), and M3 ($\Delta I_{1/2}$) are 4536 Hz, 1107 Hz, 11 321 Hz, and 8393 Hz, respectively. Therefore, we conclude that the proposed method can locate a more accurate and narrow frequency band than other three methods on the premise that it shows better diagnosis results as summarized in Table 3. This conclusion shows another benefit from health reference consideration.

At the end of benchmark validation, we use the dataset #177 (FE) as an example to further demonstrate the proposed method. According to Fig. 5, four diagrams from M2 and M3 shows location (highlighted by red rectangles) and associated information of their selected frequency bands. As already displayed in Table 4, M2 and M3 (ΔI_ε) select a relatively small bandwidth than that in M3 (ΔI_E) and M3 ($\Delta I_{1/2}$). Fig. 6 shows accugram from the

Table 3
Bearing fault diagnosis results (DE/FE).

Dataset #	Fault diameter (mm)	Motor load (HP)	Motor speed (rpm)	M1	M2 [4]	M3 [7]			M4 (Proposed)
						ΔI_ε	ΔI_E	$\Delta I_{1/2}$	
109	0.18	0	1796	Y/Y	P/Y	P/P	Y/Y	Y/P	Y/Y
110	0.18	1	1772	Y/P	P/Y	P/P	Y/P	Y/P	Y/Y
111	0.18	2	1748	Y/P	P/Y	P/P	Y/P	Y/P	Y/Y
112	0.18	3	1721	Y/Y	Y/Y	Y/P	Y/Y	Y/P	Y/Y
174	0.36	0	1796	N/N	N/N	N/N	N/N	N/N	N/N
175	0.36	1	1772	P/P	P/P	P/P	P/Y	P/Y	Y/Y
176	0.36	2	1751	P/P	P/P	P/P	P/P	P/Y	Y/Y
177	0.36	3	1726	P/P	P/P	P/P	P/P	Y/P	P/Y
Summary		"Y" Count		6	5	1	7	7	**13**
		"P" Count		8	9	13	7	7	**1**
		"N" Count		2	2	2	2	2	**2**

Notes: A/B: A = diagnosis category based on the DE data; B = diagnosis category based on the FE data.

Table 4
Selected frequency bands (DE/FE).

Dataset #	M1	M2 [4]	M3 [7]			M4 (Proposed)
			ΔI_ε	ΔI_E	$\Delta I_{1/2}$	
109	(12 000,24 000)/ (12 000,24 000)	(19 000,2000)/ (13 875,750)	(11 250,1500)/ (11 250,1500)	(12 000,24 000)/ (6000,12 000)	(12 000,24 000)/ (11 500,1000)	(7875,750)/ (7875,750)
110	(12 000,24 000)/ (12 000,24 000)	(19 000,2000)/ (18 500,1000)	(11 250,1500)/ (11 250,1500)	(6000,12 000)/ (6000,12 000)	(6000,12 000)/ (11 500,1000)	(4125,750)/ (15 375,750)
111	(12 000,24 000)/ (12 000,24 000)	(19 000,2000)/ (18 500,1000)	(11 250,1500)/ (12 500,1000)	(6000,12 000)/ (1500,1000)	(6000,12 000)/ (11 500,1000)	(375,750)/ (4125,750)
112	(12 000,24 000)/ (12 000,24 000)	(6000,12 000)/ (18 500,1000)	(23 625,750)/ (23 625,750)	(6000,12 000)/ (6000,12 000)	(6000,12 000)/ (23 625,750)	(4125,750)/ (14 500,1000)
174	(12 000,24 000)/ (12 000,24 000)	(6000,12 000)/ (21 375,750)	(2625,750)/ (16 875,750)	(6000,12 000)/ (22 000,4000)	(6000,12 000)/ (16 500,3000)	(8500,1000)/ (14 500,1000)
175	(12 000,24 000)/ (12 000,24 000)	(6000,12 000)/ (23 000,2000)	(5625,750)/ (5625,750)	(6000,12 000)/ (9375,750)	(6000,12 000)/ (5500,1000)	(15 375,750)/ (17 000,2000)
176	(12 000,24 000)/ (12 000,24 000)	(12 000,24 000)/ (23 000,2000)	(5625,750)/ (5625,750)	(12 000,24 000)/ (375,750)	(12 000,24 000)/ (9375,750)	(8500,1000)/ (17 000,2000)
177	(12 000,24 000)/ (12 000,24 000)	(19 500,1000)/ (13 125,750)	(15 750,1500)/ (14 500,1000)	(6000,12 000)/ (18 000,12 000)	(12 000,8000)/ (20 000,8000)	(375,750)/ (17 000,2000)

Notes: (f_c, Bw): f_c = center frequency and Bw = bandwidth.

proposed method, where a relatively narrow frequency band is selected and highlighted by a red rectangle. Based on their selected frequency bands, squared envelope spectrums are generated and are shown in Figs. 7 and 8, respectively. According to Fig. 7, the fault characteristic frequency f_{BPFI} shows up in the spectrums, but it is not a dominant component. Therefore, their diagnosis results are all in the "P" category in Table 3. If the fault characteristic frequency is not dominant, it is not convincing that the indicators is changed due to the associated faults, although the frequency can be observed. This is why we define this case probably diagnosable. On the other hand, the component f_{BPFI} and its harmonics not only shows up but also are dominant components that have the largest amplitude in the overall spectrum in Fig. 8. Based on our definitions, this diagnosis category should be "Y" for the proposed method. It indicates that the frequency band determined by the proposed method is the right one intrinsically affected by the inner race fault. However, the dominant component is the shaft rotating frequency f_n (about 28.77 Hz for the dataset #177) and its harmonics in Fig. 7 for other three methods. It means that the component f_n could show impulsive and cyclostationary characteristics as the fault transient and may further mislead a correct FBS for rotating machinery fault diagnosis.

We also test five indicators (excess kurtosis proposed in M2; NE_{SE}, NE_{SES}, and $(NE_{SE} + NE_{SES})/2$ proposed in M3; accuracy proposed in M4) mentioned in this paper for the selected frequency bands in the healthy and test conditions. Results are summarized in Table 5. From this table, it clearly shows that the indicators employed by the methods 2–4 reach their maximums in the test condition, respectively. For example, spectral kurtosis adopts excess kurtosis as the indicator to select the most sensitive frequency band, and thus its value of excess kurtosis (107.92) is the maximum one among that in all frequency band candidates. A very important observation from this table is that the indicators in M2 and M3 sometimes has a larger value in the healthy condition than that in the test condition. These cases are highlighted by underlines in Table 5. It somehow explains the importance of considering a health reference in the task of FBS, because a larger absolute value may not indicate a fault-related frequency band. The proposed method actually uses a relative value by considering both the healthy and test conditions. It should partially be the key point of success of the proposed method.

3.2. Experimental gearbox dataset

In this section, the proposed method is further validated on an experimental dataset reported in [25]. The dataset is regarding to a tooth missing fault of a sun gear in a planetary gearbox test rig shown in Fig. 9. The experimental text rig consists of a spur gearbox and a one-stage planetary gearbox driven by a 2.24 Kw three phase electrical motor with a motor speed controller. The physical parameters of the one-stage planetary gearbox are

Table 5
Indicator results of selected frequency bands for the dataset #177 (FE).

	M1	M2 [4]	M3 [7]			M4 (Proposed)
			ΔI_ε	ΔI_E	$\Delta I_{1/2}$	
Excess kurtosis	(0.89, −0.01)	(**107.92**, 0.06)	(102.06, 3.38)	(25.56, −0.63)	(53.10, 0.28)	(12.17, 0.04)
NE_{SE}	(1.77, 1.13)	(5.14, 1.24)	(**7.02**, 1.18)	(4.83, 0.54)	(4.99, 1.42)	(6.03, 1.41)
NE_{SES}	(2.76, 3.47)	(0.92, 1.31)	(0.77, 0.50)	(**3.86**, 4.95)	(3.74, 1.33)	(2.28, 1.26)
$(NE_{SE} + NE_{SES})/2$	(2.27, 2.30)	(3.03, 1.28)	(3.90, 0.84)	(4.35, 2.75)	(**4.37**, 1.38)	(4.16, 1.34)
Accuracy	89.53%	88.40%	98.40%	66.53%	85.20%	**99.48%**

Notes: (A, B): A = indicator value under the test condition and B = indicator value under the healthy condition.

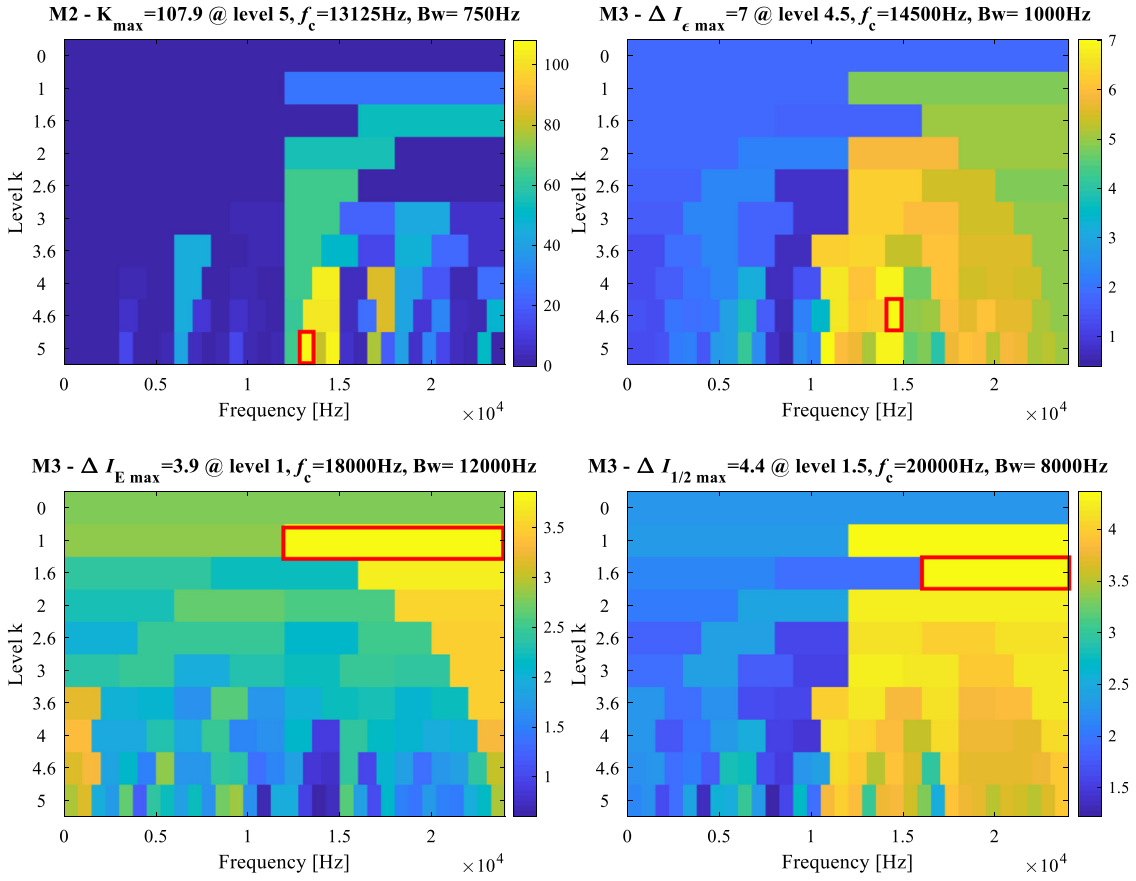

Fig. 5. Kurtogram and infogram for the dataset #177 (FE).

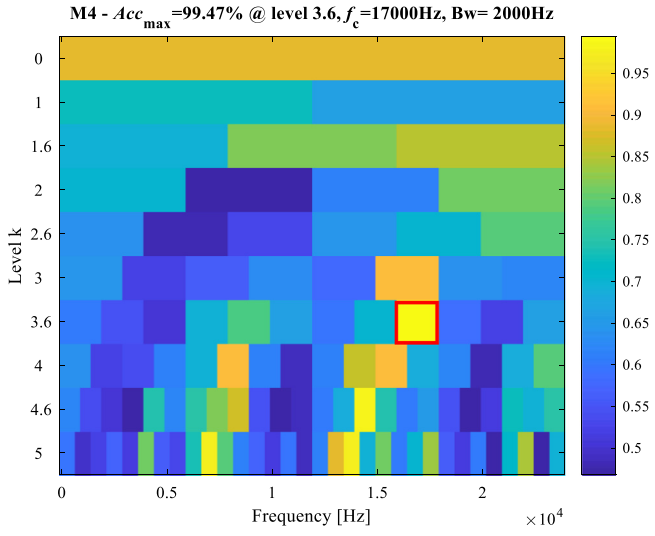

Fig. 6. Accugram for the dataset #177 (FE).

as follow: sun gear, 28 teeth; planet gear, 36 teeth; ring gear, 100 teeth; number of planet gears, 4. Therefore, it can validate generalization performance to gear fault diagnosis other than bearing fault diagnosis.

Two experiments have been conducted: one is for the health reference; another is for the faulty condition. The load level is 0 and the shaft rotating frequency is 50 Hz for both experiments. For each experiment, vibration data have been collected for about 6.4 s on a vertical accelerometer settling on the case of the gearbox with a sampling frequency of 30 720 Hz. Similar to bearing fault diagnosis, if a tooth missing fault occurs on a sun gear, it usually excites an amplitude increasing of the sun gear rotational frequency (f_{sun}) that is equal to $3.125 \times f_n$ [25]. In the following, we use only this characteristic frequency to show a potential application of the proposed method, even though there may exist other promising indicators.

The selected frequency bands for M2 and M3 are (8960, 2560), (11 520, 7680), (240, 480), and (11 520, 7680) Hz, respectively. Fig. 10 shows squared envelope spectrums after filtering their selected bands. From this figure, the characteristic frequencies and

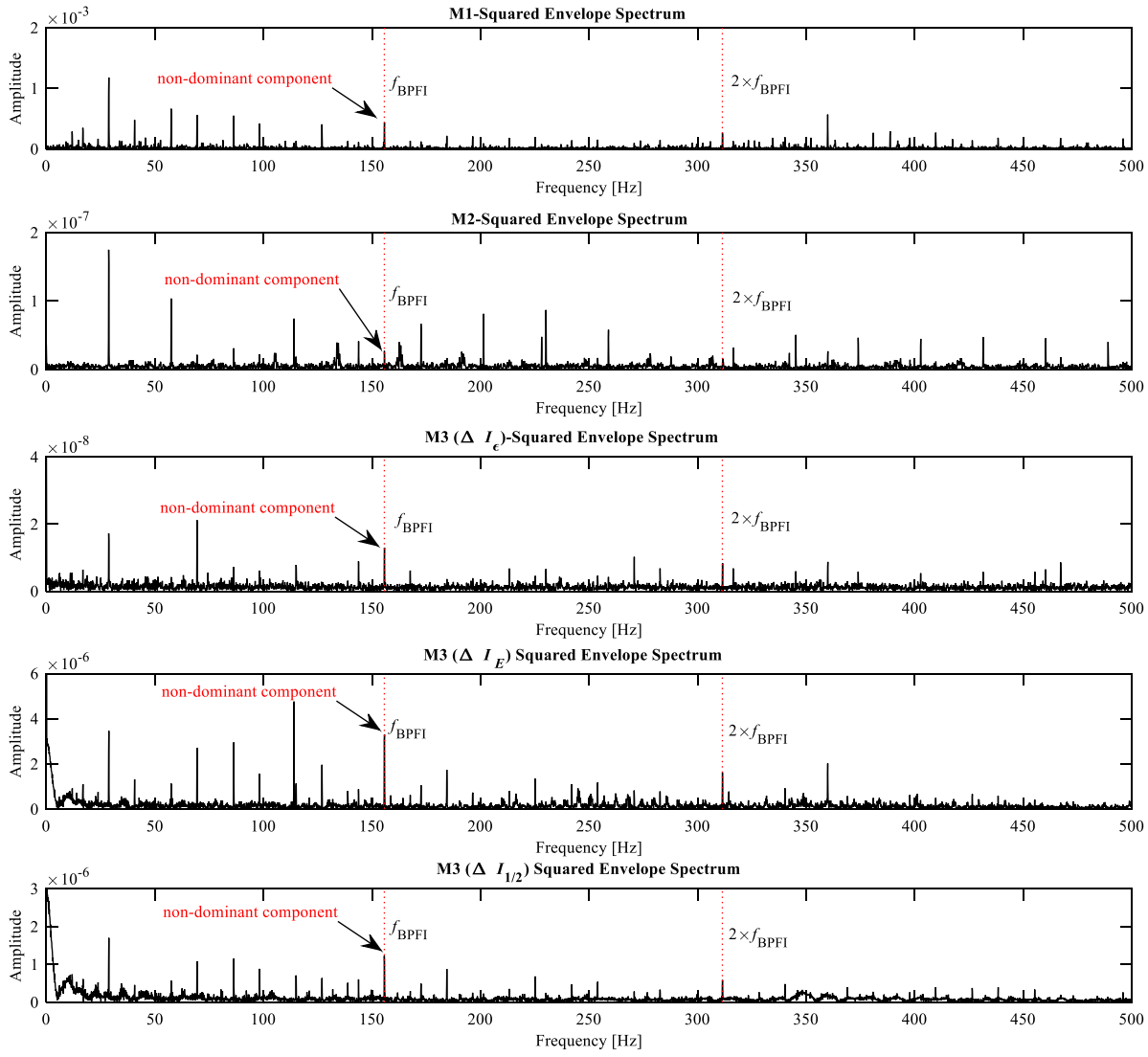

Fig. 7. Squared envelope spectrum with M1, M2, and M3 for the benchmark bearing dataset.

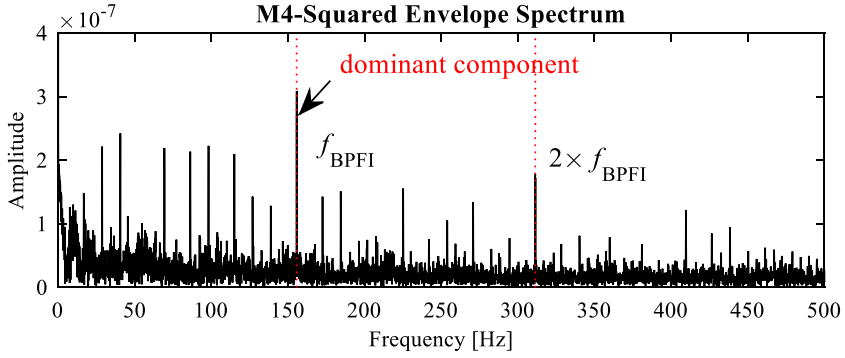

Fig. 8. Squared envelope spectrum with the proposed method.

its second-order harmonics clearly show up as a dominant component in all spectrums except the one from M3 (ΔI_E). Following definitions in Table 2, it means that the diagnosis categories are "Y" for M1, M2, M3 (ΔI_ε), and M3 ($\Delta I_{1/2}$), and it should be "N" for M3 (ΔI_E). M3 (ΔI_E) can select the smallest bandwidth, but its squared envelope spectrum is dominated by the shaft rotating frequency (50 Hz) and its harmonics that also have cyclostationary characteristic. This fact shows that M3 (ΔI_E) sometimes fail to recognize the right frequency band related to gear fault diagnosis.

From the accugram for the gearbox dataset, it can select the smallest bandwidth of 480 Hz with a center frequency of 4560 Hz among all four methods. After filtering the test signals, the squared envelope spectrums are presented in Fig. 11,

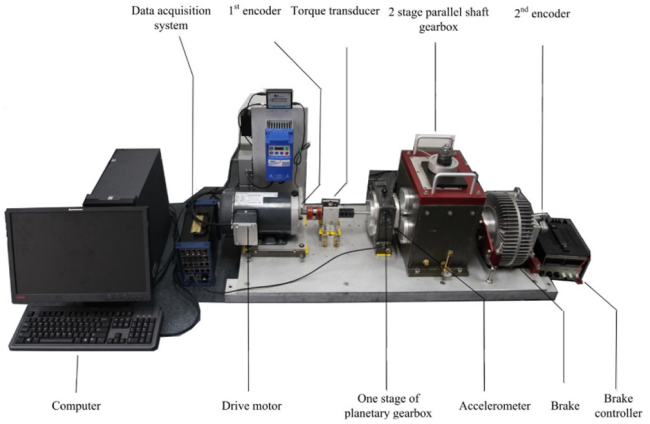

Fig. 9. The planetary gearbox test rig.

where the characteristic frequency dominates the spectrum and increases significantly in amplitude comparing to its health reference. Thus, the proposed method can also make accurate gear fault diagnosis based on our observation that the fault characteristic frequency also dominates the envelope spectrum. Table 6 provides five indicators proposed in M2, M3, and M4 for the selected frequency bands. From this table, we can observe that some indicators in the healthy condition is even larger than that in the fault condition. These cases are highlighted by underlines in this table. For example, NE_{SES} in M3 (ΔI_{ε}) is maximized at 2.89, but this maximum value is still smaller than 2.91 in the healthy condition. It explains why this method fails to locate the fault transient as pointed in the above paragraph. This kind of weird phenomenon is observed in Section 3.1 as well. On the contrary, the five indicators in the proposed method follow a reasonable relationship of numerical value between the healthy and faulty data. In addition, the proposed method gives an accuracy of 62.56% for the selected frequency band by M3 (ΔI_{ε}), which exactly reflects its fault diagnosis category "N". It again demonstrates that the proposed method can eliminate effect from the shaft rotating frequency by introducing the healthy data. From the above analysis, we can see that the conclusions from gearbox fault diagnosis is consistent with that made from bearing fault diagnosis in Section 3.1.

4. Conclusion and remarks

This paper proposes a novel method, coined by accugram, to find the most sensitive frequency band for rotating machinery

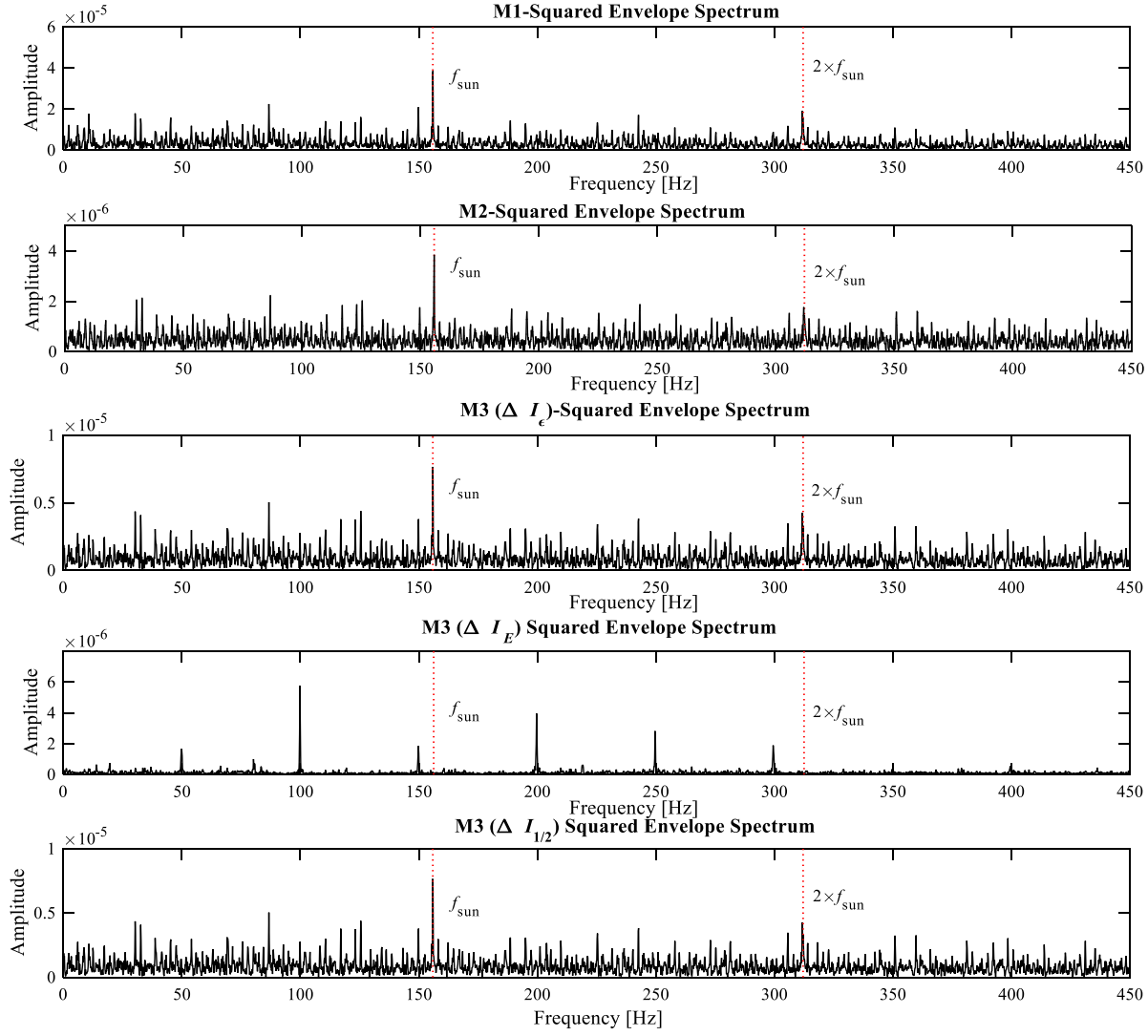

Fig. 10. Squared envelope spectrum with M1, M2, and M3 for the experimental gearbox dataset.

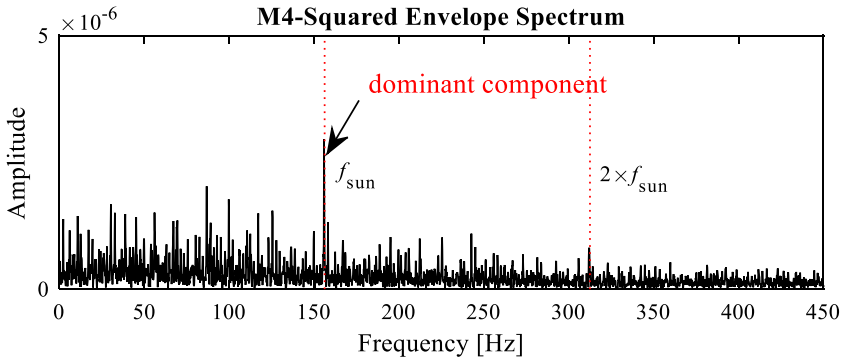

Fig. 11. Squared envelope spectrum with the proposed method.

Table 6
Indicator results of selected frequency bands for the gearbox dataset.

	M1	M2 [4]	M3 [7]			M4 (Proposed)
			ΔI_E	ΔI_E	$\Delta I_{1/2}$	
Excess kurtosis	(30.07 0.65)	(**172.5**, 7.93)	(92.4, 2.36)	(−0.24, −0.14)	(92.4, 2.36)	(38.0, 1.34)
NE_{SE}	(6.23, 1.57)	(6.53, 2.10)	(**7.03**, 2.48)	(0.61, 0.58)	(7.03, 2.48)	(3.76, 1.89)
NE_{SES}	(1.04, 3.25)	(0.83, 1.07)	(1.11, 1.27)	(**2.89**, 2.91)	(1.11, 1.27)	(1.30, 0.93)
$(NE_{SE}+NE_{SES})/2$	(3.63, 2.41)	(3.68, 1.59)	(4.07, 1.87)	(1.75, 1.75)	(**4.07**, 1.87)	(2.53, 1.41)
Accuracy	97.31%	75.38%	77.05%	62.56%	77.05%	**99.48%**

Notes: (A, B): A = indicator value under the test condition and B = indicator value under the healthy condition.

fault diagnosis. MATLAB source codes of the accuracy can be found at [26]. This paper has three contributions: (1) health reference is extensively investigated and recognized as a critical aspect in frequency band selection other than impulsiveness and cyclostationarity; (2) a novel approach based on classification is proposed to integrate health reference, impulsiveness and cyclostationarity for frequency band selection; and (3) classification accuracy is developed as a novel indicator to select the most sensitive frequency band for rotating machinery fault diagnosis. The proposed method has two new features: (1) it can make use of healthy data and their associated labels; and (2) it can fuse multiple indicators for FBS. Extensive validation on 16 bearing datasets and one gearbox dataset shows that the accugram outperforms the conventional envelope analysis, the kurtogram, and the infogram in terms of a more accurate fault diagnosis and a smaller selected frequency band. We are currently continuing to work on development of the accugram towards a flexible frequency band partition. More results will be presented later.

Remarks about the current work are provided as follows.

(1) ***Health reference is critical for FBS***. Impulsiveness and cyclostationarity are two recognized existing evidences of transients. They definitely should be well considered in FBS. However, the two evidences can be affected by other component, e.g. shaft rotating frequency, and thus sometimes may fail to have a successful recognition of fault transients. As shown in Section 3.2, the indicator of M3 (ΔI_E) is contributed by the shaft rotating frequency rather than the fault characteristic frequency. As a result, M3 (ΔI_E) fails to identify the gear fault. This example strongly demonstrates importance of health reference that has been ignored for a long time in FBS. By introducing a health reference, the two characteristics are more likely the existing evidences of fault transients. Therefore, we conclude that health reference is very important and should attract equal attention as the first two evidences considering in FBS, particularly for the purpose of rotating machinery fault diagnosis.

(2) ***Accugram is a semi-supervised FBS method***. Because the accugram needs a health reference, we need to know labels of the reference data in advance from perspective of machine learning. It is usually not a problem to obtain the health reference in practice. Real labels of test data are not necessary to know before applying the accugram. In fact, the accugram considers the test data as an arbitrary label but different from the health reference one, which is already justified in Section 2. It uses a classification method to incorporate label information and learn from the reference data in FBS. Due to involving partial label information, the accugram can be grouped into semi-supervised methods. For the similar reason, FBS methods like spectral kurtosis and infogram can be grouped into unsupervised methods. Obviously, the unsupervised methods cannot make use of and benefit from health reference data and their labels though they do not require known data with the healthy label.

(3) ***Accugram provides a flexible framework for FBS***. In our implementation of the accugram, we depict impulsive and cyclostationary characteristics by using negentropy of the squared envelope and of the squared envelope spectrum, respectively. It could exist another promising indicator or use more than one indicators to represent the same characteristic of a transient. Moreover, beside of the two recognized characteristics, additional characteristics could exist and can be quantified by using some indicators. For the above two cases, the accugram can easily handle them by simply renewing the feature extraction process. From this perspective, classification accuracy is an integrated evidence of transients and its integration level depends on how many relevant indicators involving in FBS. Therefore, the accugram is a flexibly designed framework for FBS.

Acknowledgment

This work was supported in part by The National Key Research and Development Program of China (2018YFB1702400), National Natural Science Foundation of China (61833002), in part by State Key Laboratory of Traction Power, Southwest Jiaotong University (TPL1608). Comments and suggestions from the reviewers, the editors, and Prof. Jerome Antoni are very much appreciated.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have

impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.isatra.2019.05.007.

References

[1] Obuchowski J, Wyłomańska A, Zimroz R. Selection of informative frequency band in local damage detection in rotating machinery. Mech Syst Signal Process 2014;48:138–52.

[2] Barszcz T, Jabłoński A. A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram. Mech Syst Signal Process 2011;25:431–51.

[3] Cui LL, Huang JF, Zhang FB, Chu FL. HVSRMS Localization formula and localization law: Localization diagnosis of a ball bearing outer ring fault. Mech Syst Signal Process 2019;120:608–29.

[4] Antoni J, Randall RB. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. Mech Syst Signal Process 2006;20:308–31.

[5] Antoni J. The spectral kurtosis: a useful tool for characterising non-stationary signals. Mech Syst Signal Process 2006;20:282–307.

[6] Antoni J. Fast computation of the kurtogram for the detection of transient faults. Mech Syst Signal Process 2007;21:108–24.

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

[8] Liu ZL, Jin YQ, Zuo M, Feng ZP. Time-frequency representation based on robust local mean decomposition for multicomponent AM-FM signal analysis. Mech Syst Signal Process 2017;95:468–87.

[9] Lei YG, Lin J, He ZJ. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. Mech Syst Signal Process 2011;25:1738–49.

[10] Wang D, Tse PW, Tsui KL. An enhanced Kurtogram method for fault diagnosis of rolling element bearings. Mech Syst Signal Process 2013;35:176–99.

[11] Wang YX, Liang M. An adaptive SK technique and its application for fault detection of rolling element bearings. Mech Syst Signal Process 2011;25:1750–64.

[12] Xu XQ, Zhao M, Lin J, Lei YG. Periodicity based kurtogram for random impulse resistance. Meas Sci Technol 2015;26:26.

[13] Zhang YX, Randall RB. Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram. Mech Syst Signal Process 2009;23:1509–17.

[14] Wang TY, Han QK, Chu FL, Feng ZP. A new skrgram based demodulation technique for planet bearing fault detection. J Sound Vib 2016;385:330–49.

[15] Liu ZL, Zhao XM, Zuo M, Xu HB. Feature selection for fault level diagnosis of planetary gearboxes. Adv Data Anal Classif 2014;8:377–401.

[16] Wang D. An extension of the infograms to novel Bayesian inference for bearing fault feature identification. Mech Syst Signal Process 2016;80:19–30.

[17] Li C, Cabrera D, Oliveira JV, Sanchez RV, Cerrada M, Zurita G. Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram. Mech Syst Signal Process 2016;76:157–73.

[18] Feng ZP, Ma H, Zuo M. Spectral negentropy based sidebands and demodulation analysis for planet bearing fault diagnosis. J Sound Vib 2017;410:124–50.

[19] Japkowicz N, Stephen S. The class imbalance problem: A systematic study. J Intell Data Anal 2002;6(5):429–49.

[20] Qian YG, Zhou WQ, Yan JL, Li WF, Han LJ. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sens 2015;7:153–68.

[21] Thanh Noi P, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors 2018;18:18.

[22] Liu Y, Sun F. A fast differential evolution algorithm using k-nearest neighbour predictor. Expert Syst Appl 2011;38:4254–8.

[23] Case Western Reserve University Bearing Data Center Website. http://csegroups.case.edu/bearingdatacenter/home.

[24] Smith WA, Randall RB. Rolling element bearing diagnostics using the Case western reserve university data: A benchmark study. Mech Syst Signal Process 2015;64:100–31.

[25] Feng ZP, Zuo MJ. Vibration signal models for fault diagnosis of planetary gearboxes. J Sound Vib 2012;331:4919–39.

[26] Liu ZL. (2019). accugram (https://www.mathworks.com/matlabcentral/fileexchange/70179-accugram), MATLAB Central File Exchange. Retrieved 03.02.19.