Automation of low-speed bearing fault diagnosis based on autocorrelation of time domain features

Authors: Riku-Pekka Nikula[a]*, Konsta Karioja[b], Mika Pylvänäinen[b], Kauko Leiviskä[a]

[a]Control Engineering, Environmental and Chemical Engineering, Faculty of Technology, University of Oulu, P.O. Box 4300, 90014 Oulu, Finland

[b]Intelligent Machines and Systems, Faculty of Technology, University of Oulu, P.O. Box 4200, 90014 Oulu, Finland

*Corresponding author. Tel.: +358 294 482475. E-mail address: riku-pekka.nikula@oulu.fi

ABSTRACT

This study is focused on the application of automated techniques on low-speed bearing diagnostics. The diagnosis in low-speed conditions is hampered by the long periods between defect-related impulses and the high level of noise relative to the magnitude of the impulses. To detect a localised defect in such conditions, a new approach that uses vibration signals and information on the bearing defect frequencies is proposed. At first, the vibration signal is filtered in a specific frequency range to enable the detection of the impulses hidden in the signal. The filtered signal is then segmented into short time windows, the length of which are selected based on the bearing defect frequencies. Statistical time domain features are calculated from these windows to amplify and compress the impulses inflicted by the defect. Then, a criterion based on the autocorrelation values of specific time lags is calculated. An exhaustive search procedure is used to determine the frequency band for signal filtering and to select the statistical feature, which together maximises the proposed criterion. The highest value of the criterion is finally compared with the corresponding value from the baseline condition to detect the localised defect. The proposed technique is demonstrated on simulated signals, and validated based on the vibration signals from laboratory tests with undamaged, slightly damaged and severely damaged rolling elements in a rolling element bearing. Different conditions with shaft speeds from 20 to 80 rpm were studied in the laboratory tests. The proposed technique was compared with automated envelope spectrum diagnosis approaches based on the peak ratio and peak-to-median indicators and the fast kurtogram. The results reveal that the criterion based on autocorrelation gave defect indications associated with the correct type of defect in various circumstances while the tested envelope spectrum approaches were prone to induce an incorrect conclusion. Moreover, the results indicate that the approach could be used successfully on signals with a length that includes relatively few defect periods or impulses. The approach requires a high sampling rate relative to the defect frequencies, which may limit its suitability for the higher shaft speeds.

Keywords: Autocorrelation, Bearing fault, Diagnosis, Low speed, Vibration

1. Introduction

Faults in rolling element bearings (REBs) are a common reason for maintenance work on industrial applications. REBs are used in a large variety of equipment and mechanical devices from roll necks in rolling mills to the main shafts in wind turbines, for example. Abrupt faults in such equipment can be disastrous, cause increased maintenance costs and a decreased rate of productivity. Therefore, the fault prognosis and diagnosis methodologies for REBs have been studied for decades [1]. The localised defects, where a piece of a contact surface in the bearing is dislodged during operation,

are the common fault type studied in the literature [2]. These defects presumably generate a quasi-periodic impulse train in the measured signals [3]. These signals can also be considered cyclostationary [2]. The defects are commonly analysed based on vibration signals by focusing on the bearing defect frequencies [4]. However, in the case of slowly rotating bearings that are typical in industrial applications, the defect-related signal features tend to become weak due to various factors, such as the skidding of the rolling elements, the vibration of other machine parts and machines, noise and disturbances [5]. The conventional diagnostic techniques often turn out insufficient under such circumstances [6].

Shaft speeds in the range from 20 to 600 rpm are often referred to as low speed [7] and significantly lower speeds are sometimes classified separately [8]. However, there is no generally accepted criterion for speed classification. Although there are studies where the slowly rotating bearings have been diagnosed based on vibration signals with successful results [5,6,7,9,10,11,12], the use of vibration appears generally unfavourable. A major part of recent research is focused on the Acoustic Emission (AE) techniques [8,13,14,15,16], but other techniques, such as the Shock Pulse Method (SPM) [17], stress waves [18], and the ultrasound technique [19], have been applied as well. AE techniques have limitations, such as unestablished methods for sensor calibration [20,21] and high sampling rate which may cause computational challenges [15]. The AE waveform also requires a direct transmission path through the solid material to avoid an intermittent signal [20], which is not guaranteed in REBs. The well-established vibration measurements can be considered more practical with respect to such limitations. Therefore, the research on their application on the diagnosis of slowly rotating bearings is also important.

The bearing diagnosis in the low-speed conditions has specific challenges. The impulse strength caused by the localised defects is relative to the shaft speed and it decreases if the shaft speed decreases [5,22]. This may cause difficulties in the diagnosis, if the noise level is high [6]. The industrial surroundings have background noise originating from different machine components and other machines. Consequently, the noise does not decrease although the bearing rotates with a lower speed, which results in a low signal-to-noise ratio in the measured vibration signal. Moreover, the periods between the shocks caused by the bearing defects are relatively long in the low-speed domain. Long samples are required to ensure an adequate number of defect-related impulses in the signals analysed. On the other hand, the spectral smearing due to speed fluctuations and skidding of rolling elements may hinder the diagnosis significantly in such long samples [7,23]. Therefore, a method that is usable on a relatively short signal that includes only a few of such weak defect-related impulses can generally be considered a fitted approach for the low-speed conditions.

The signal processing methods for bearing diagnostics are commonly divided into three categories, which are time, frequency and time-frequency analysis [24]. The time domain techniques traditionally use various statistical features, such as generalised norms [25] and features derived from them [26]. One application for such techniques is the detection of shock-like phenomena in the signal, often based on threshold values. However, it is difficult to diagnose the precise nature of the defect based on the time domain information alone. The frequency domain techniques are used to identify the defect-related spectral components from the signal. The most common method in this category is the high frequency resonant technique (envelope analysis) [27]. Finally, the time-frequency analysis shows how the frequency content of the signal changes with time. This field of analysis is rich with methods, such as short-time Fourier transform [28], wavelet transform [29] and ensemble empirical mode decomposition [30] among others [31]. Such methods appear appropriate for nonstationary signals under time-varying operational conditions [31] such as the start-up or rundown of a machine. In addition, the family of techniques in the cyclostationary analysis [32] has captured a broad interest of the scientific community recently [33].

Furthermore, the methods for the automation of diagnosis draw increasing attention. The automated solutions for diagnosis are studied from the viewpoints of partial and complete automation. As a

partial solution, the automation of envelope spectrum diagnosis, for example, has been studied from many viewpoints. Shiroishi et al. [34] used the peak ratio indicator and Ericsson et al. [24] used the peak-to-median indicator to extract defect-related features in the envelope spectrum diagnosis. Additionally, several methods have been proposed for the frequency band selection in the band-pass filtering stage of the method. These include the spectral kurtosis [28], the fast kurtogram based on multirate filter-bank [35], the combination of the fast kurtogram and genetic algorithms [36], the kurtosis of envelope spectra amplitudes [37], and adaptive time-frequency analysis based on window superposition [38], to mention only a few.

The methods which aim at the fully automated solutions often use machine learning approaches, which classify different bearing conditions based on the supervised learning techniques [39,40]. These approaches commonly utilise an exhaustive feature set to separate the fault signature [41]. Samanta and Al-Balushi [39], for example, used time domain features, Widodo et al. [15] used features from envelope spectra together with time domain features and Rauber et al. [41] extended the feature set even more by incorporating time-frequency features. The inclusion of dimension reduction [42] or feature selection [43] methods is commonly required to improve the viability of the machine learning approach. In addition, the machine learning in industrial applications requires highly careful data pre-processing in order to remove the effects of disturbances unrelated to the operation of the monitored system [44]. A recent addition to the machine learning approach is to generate the training data based on bearing simulations [40].

The research question of this study is focused on the automated detection of localised defects in slowly rotating rolling element bearings based on vibration measurements. The main contribution is the introduction of the criterion that is based on the autocorrelation of statistical features calculated from band-pass filtered signals. The features are extracted in short time windows, which have lengths defined based on the fundamental defect frequencies of the bearing. These statistical features are selected based on their sensitivity to the shocks inflicted by the defect. The short time windows compensate for the minor speed variations and the fact that the defect-related shocks are weak and do not appear in exactly periodic time instances in the raw signal. The criterion is proposed especially for the low-speed bearing diagnosis cases, where the sampling rate is markedly high relative to the presumed defect frequencies and the defect-related impulses are weak in the raw signal. The method is different from the supervised learning approaches in that it does not require training based on the previous operation. On the contrary, the method automatically defines the appropriate diagnostic parameters based on the current operation.

The proposed method is compared with automated envelope spectrum diagnosis approaches, where the defects are identified based on the peak ratio and peak-to-median indicators. Two different approaches are tested for the frequency band selection, because it is the focal processing stage but does not have a single established approach that would be appropriate for each situation. Recently, it has become popular to select the optimal frequency band for demodulation by identifying the band, which shows the strongest response to impulses without considering the origin of the impulses [45]. This approach tends to give fallacious results due to the presence of interfering vibration content in the signals in practical applications [2,37,46]. Therefore, in the first tested approach, the selection is done by maximising the peak ratio and peak-to-median indicators. This approach is based on the idea that the impulses related to the bearing defect frequencies, which are associated with the indicators, explicitly determine the appropriate frequency band. The fast kurtogram was selected as the second approach due to its computational efficiency and the popularity in the current mechanical fault diagnosis literature [45,47,48]. In this approach, the origin of the impulses is not considered.

The paper is organised as follows. The proposed autocorrelation-based criterion and the approaches used for the envelope spectrum diagnosis are introduced in Section 2. Section 3 describes the test rig, the settings and the measurements applied in the rolling element bearing diagnosis experiment under the slow rotational speed circumstances. Section 4 provides the results and discussion based

on the tests on simulated signals and experimental data from the test rig. Finally, Section 5 concludes the paper.

2. Methodology

The criterion based on autocorrelation is introduced in Section 2.1. Section 2.2 introduces the envelope spectrum diagnosis approach based on the peak ratio and peak-to-median indicators. The application of the fast kurtogram in the frequency band selection based on the kurtosis maximisation principle is described in [35].

2.1. Criterion based on autocorrelation

The proposed approach can be divided into separate parts, which include signal filtering, segmentation, feature extraction, and the calculation of autocorrelation and the criterion. A flowchart that illustrates how the parts are linked with each other is presented in Fig. 1. Firstly, the vibration signal is filtered in the frequency domain, which is explained in Section 2.1.1. Then, the filtered signal is segmented into short time windows as shown in Section 2.1.2. These short time windows are compressed into single data points by the extraction of statistical features, which is introduced in Section 2.1.3. Autocorrelation is calculated from each feature, as discussed in Section 2.1.4, and then the criterion based on autocorrelation is calculated, which is introduced in Section 2.1.5. The whole procedure is repeated several times by using different frequency bands in filtering. When all the frequency bands are checked, the optimal settings are selected based on the proposed criterion. The optimal settings include the statistical feature and the frequency band, which together result in the highest value of the proposed criterion with the applied time window size. A demonstration of the signal processing stages in the calculation of the criterion is provided in Section 2.1.6. Finally, Section 2.1.7 illustrates the significance of sampling rate and defect frequency on the resolution of the practical segmentation procedure in the proposed approach.

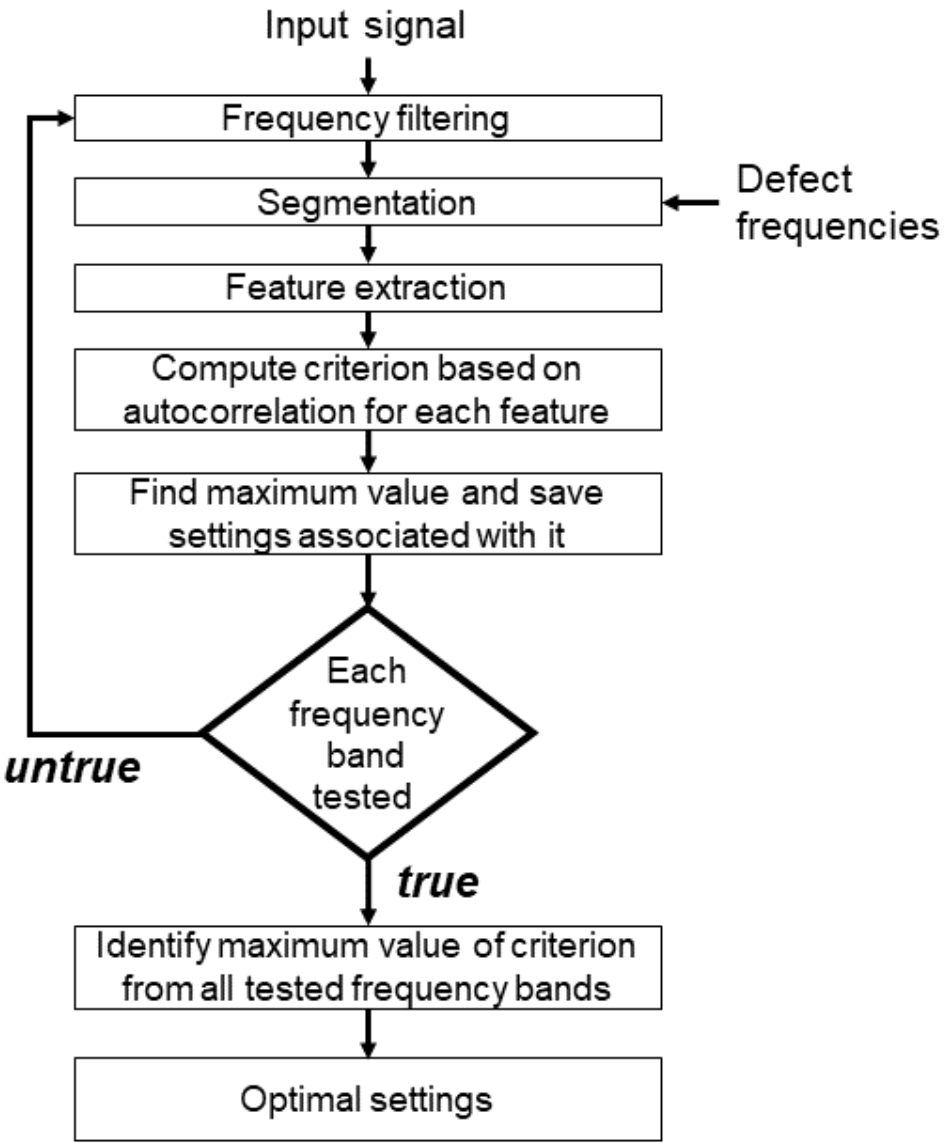

Fig. 1. Flowchart demonstrating the stages in the calculation of the criterion. The optimal settings include the statistical feature and the frequency band, which together maximise the proposed criterion with the applied time window size.

The highest value of the criterion gives insight into the current bearing condition. This value is compared with the corresponding reference value from the baseline condition. The baseline data have been acquired in an earlier stage from the same bearing in a good condition. The operational situations in both bearing conditions are preferably the same including the same rotational speed and load on the bearing. The reference value is computed from the baseline data by using the optimal settings defined from the currently monitored bearing condition. Therefore, the settings applied on the processing of the baseline signal and the monitored signal include the same frequency band, the same segment size (in per cent), and the same statistical feature. The monitored signal is processed first and the baseline signal is processed thereafter using the settings defined based on the monitored signal.

If the value from the current bearing condition is higher than the reference value, a change in the condition can be assumed. Therefore, the fully automated diagnosis requires the definition of a

threshold value for the difference between the current and the baseline conditions. The definition of an appropriate threshold value is done in a case-by-case basis and it is a compromise between the sensitivity to change and false alarms. Additionally, the criterion value from the monitored signal can be used to evaluate the periodically repeating patterns that are associated with the defect frequencies, if the baseline data are not available.

2.1.1. Filtering

Before filtering the acceleration signal in the frequency domain, it is essential to multiply the time domain signal by a window function, such as the Tukey window in this study. If the windowing is not applied, ripples could be created in the time domain signal in the filtering procedure. After the windowing stage, the fast Fourier transform (FFT) is used to transform the vibration signal into the frequency domain. The unwanted frequency components are then removed by multiplying the corresponding complex numbers by zero. This procedure enables the selection of certain frequency range and the removal of all the frequency components outside the specified range. After this procedure, the signal is transformed back into the time domain by using the inverse FFT.

The acceleration signal is filtered repeatedly by selecting different frequency bands, as indicated by the loop in the flowchart in Fig. 1. In this study, a specific procedure for frequency band segmentation was applied. The applicable frequency band was divided into sub-bands with different sizes, as illustrated in Fig. 2. The frequency bands with sizes 2.5%, 3.3%, 5%, 10%, 20%, and 50% of the applicable range were tested. The number of frequency bands with these sizes were 40, 30, 20, 10, 5, and 2, respectively. The applicable frequency range could be from 0 Hz to Nyquist frequency or it could cover the linear frequency response range of an accelerometer.

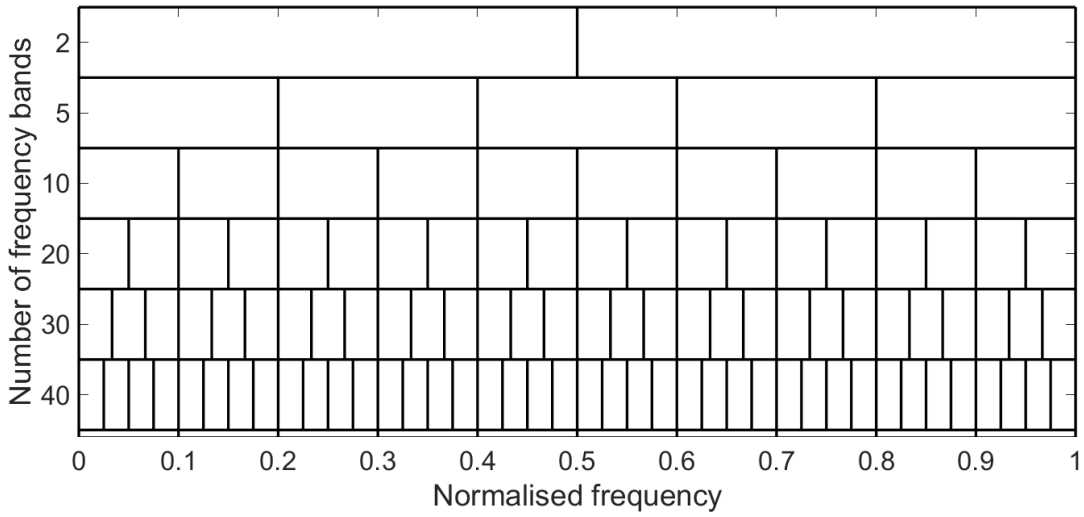

Fig. 2. Frequency bands used in the selection of the optimal frequency range for signal filtering. In this graph, the applicable frequency range is normalised into range 0–1.

2.1.2. Segmentation

The low-speed bearing signal requires a long recording time, because the repetitive pattern of the localised defect has a slow rate. The processing of long signals results in major computational load if some computational procedure is applied on each data point individually. To solve this problem,

the signal can be compressed by chopping it in short consecutive segments (or time windows), from which the features of choice are then calculated. However, it is important to select the proper segment size carefully. Quite often the signal is chopped into segments that have a size unrelated to the frequency of the fault pattern [39,41]. In such a case, the periodic repetitive pattern of the fault becomes masked in the segments and in some cases the fault-related shocks may appear unevenly spread in the segments. The solution proposed in this study is to chop the signal based on the bearing defect frequencies. This means that the segmentation is done in various lengths, because several types of defects are monitored in the automated bearing diagnosis applications. By using this approach, the shocks generated by a particular defect are located on the segments at specific intervals and become thus detectable. Under the assumption that there is no skidding between the rolling elements and races, the defect frequencies are calculated as follows

$$BPFO = \frac{nf_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\}, \quad (1)$$

$$BPFI = \frac{nf_r}{2}\left\{1 + \frac{d}{D}\cos\phi\right\}, \quad (2)$$

$$FTF = \frac{f_r}{2}\left\{1 - \frac{d}{D}\cos\phi\right\}, \quad (3)$$

$$BSF = \frac{Df_r}{2d}\left\{1 - \left(\frac{d}{D}\cos\phi\right)^2\right\}, \quad (4)$$

where BPFO is the ballpass frequency for the outer race, BPFI is the ballpass frequency for the inner race, FTF is the fundamental train frequency (cage speed), and BSF is the rolling element (roller or ball) spin frequency. Parameter f_r is the shaft rotational frequency, n is the number of rolling elements, d is the rolling element diameter, D is the pitch diameter of the bearing, and ϕ is the angle of the load from the radial plane.

In order to chop the signals into segments, the defect periods are calculated by taking the inverse of bearing defect frequencies. The length of a segment is set to be 1–2% of the defect period, because this length is proportional to the typical amount of variation due to skidding [2]. If the segment length is considerably increased, the impulses unrelated to the defect possibly begin to dominate in the segments and the risk of misdiagnosing increases.

The segmentation approach requires consideration of the practical sampling rate used for the vibration measurements. Above all, the sampling rate must enable the separation of different defect frequencies. Therefore, the sampling rate is selected in such a way that the rounded number of data points in a segment is a different positive integer for different defect frequencies. In addition, the ratio of the sampling rate to the defect frequency must be high enough to ensure a sufficient resolution in the segmentation approach, which is demonstrated in Section 2.1.7. These issues are likely to be more challenging in the higher shaft speeds than in the low-speed conditions.

2.1.3. Feature extraction

The localised faults in rolling element bearings presumably produce series of shocks, which are manifested in the acceleration signal. Therefore, the computational features have to be sensitive to shock-like phenomena. However, it is unclear which feature is the most sensitive to the weak characteristics of the defect in the low-speed case. The generation of a relatively large set of candidate features gives a well-founded basis for the solution. In this study, twenty statistical features were computed. These features have minor requirements for the computing capacity and therefore their use in practical applications is reasonable. Most of the features are based on the l_p norm which is also named as the generalised norm [49]. It is defined by

$$\left\|x^{(\alpha)}\right\|_p = \left(\frac{1}{N}\sum_{i=1}^{N}\left|x_i{}^{(\alpha)}\right|^p\right)^{\frac{1}{p}}, \quad (5)$$

where the real number α represents the order of derivative, x is displacement, N is the number of data points in the time window, and the real number p is the order of the norm. The l_p norm has the same form as the generalised mean, also known as Hölder mean or power mean [50]. In this study, only the acceleration signals ($x^{(2)}$) were used.

Several l_p norms and ratios of l_p norms were calculated in this study. As demonstrated for example in [44], the l_p norm with low order (e.g. p = 0.5 or p = 2) is sensitive to changes in the signal amplitude level, whereas the norm with high order (e.g. p = 10) is more sensitive to individual peaks in the signal. The ratios of high-order and low-order l_p norms give indications about the ratio between the peaks and the general amplitude level. Additionally, indicators, which describe the shape of probability distribution, such as the kurtosis (6) and skewness (7) of signal values, and 95th percentile of absolute signal values, were calculated. Kurtosis and skewness are defined by

$$Kurtosis = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4, \quad (6)$$

$$Skewness = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^3. \quad (7)$$

The kurtosis of a univariate normal distribution is 3. In this study, the excess kurtosis was used to describe the distribution of the signals as well. In that case, the kurtosis of a univariate normal distribution is zero. The set of twenty candidate features is presented in Table 1.

Table 1 Definition of the candidate features.

Features	Definition
$l_{0.5}$, l_2, l_4, l_{10}, l_{20}	Generalised norms, p = {0.5, 2, 4, 10, 20}
l_{20} / l_{10}	Ratio of two high-order generalised norms, p = {10, 20}
l_{20} / l_2, l_{10} / l_2, l_{10} / l_1, l_{20} / $l_{0.5}$, l_{10} / $l_{0.5}$	Ratios of high-order generalised norms, p = {10, 20}, to low-order generalised norms, p = {0.5, 1, 2}
l_4 / $l_{0.5}$, l_4 / l_1, l_4 / l_2	Ratios of a relatively high-order generalised norm, p = 4, to low-order generalised norms, p = {0.5, 1, 2}
l_∞	Maximum of absolute signal values (peak)
l_∞ / l_2	Crest factor, ratio of peak to root-mean-square
l_∞ / $l_{0.5}$	Margin factor, ratio of peak to low-order generalised norm, p = 0.5
$P_{95}(\lvert x^{(2)}\rvert)$	95th percentile of absolute signal values
Kurtosis	Kurtosis describes the tails of probability distribution
Skewness	Skewness describes the asymmetry of probability distribution

2.1.4. Autocorrelation

Autocorrelation is a tool for finding periodically repeating patterns in a signal. It provides information about the structure of the signal or its behaviour in time domain. The autocorrelation for lag τ can be defined by

$$R_{xx}(\tau) = \frac{c_\tau}{c_0}, \quad (8)$$

$$c_\tau = \frac{1}{N-1}\sum_{i=1}^{N-\tau}(x_i - \bar{x})\,(x_{i+\tau} - \bar{x}), \quad (9)$$

where c_τ is the estimate of the autocovariance and c_0 is the sample variance of the time series [51]. This is a measure of correlation between the signal and its time-shifted variant. In rolling element bearing diagnostics, autocorrelation is commonly applied on time series that have the same sampling rate as the raw signal [11,29,46,47]. In such a case, it is assumed that the defect is shown by the peaks that repeat at nearly fixed intervals. The defect period, however, typically deviates 1–2% from the calculated value [2], which corresponds to a large number of data points in low-speed applications, which have accelerometers as sensors. In addition, the shocks generated by the defect tend to be weak and their detection may require derived values, such as the features introduced in the previous section. Therefore, the segmentation approach introduced in Section 2.1.2 is used to compensate for the small frequency variation and the features given in Section 2.1.3 are used to magnify shocks.

After segmentation, the 50th time lag can be considered defect-related, when the segment length is 2% of the defect period. Similarly, the 100th time lag is associated with the defect frequency, when the segment length is 1% of the defect period. With this approach, the autocorrelation function can be used to identify the shocks associated with the defect in a straightforward manner.

2.1.5. Calculation of the criterion based on autocorrelation

The proposed criterion is used to extract the most separable peak at the time lag corresponding to the bearing defect frequency (50th or 100th time lag). This extraction is based on the selection of the statistical feature and the frequency band, which maximise the value of the criterion, i.e. produce the most separable peak. To determine the best setting for this purpose, the autocorrelation of the time lag, which corresponds to the defect frequency, is compared with the autocorrelation values of other time lags nearby. The criterion can be defined as

$$Criterion = max(R_{xx}(\tau_1)) - P_k(R_{xx}(\tau_2)), \quad (10)$$

where $\tau_1 = \{ \tau_{\text{defect}} - 1, \tau_{\text{defect}}, \tau_{\text{defect}} + 1\}$ denotes the search range around the time lag of defect (τ_{defect}). This narrow search range (1–2% in both directions) is applied to find the maximum value close to the theoretical defect time lag. The term $P_k(R_{xx}(\tau_2))$ denotes the k^{th} percentile of autocorrelation values for time lags $\tau_2 = \{1, \ldots, \tau_{\text{limit}}\}$, $\tau_1 \notin \tau_2$. The upper limit (τ_{limit}) for the considered time lags is defined, because only the time lags close to τ_{defect} are of interest. In this study, $\tau_{\text{limit}} = 2.5 \cdot \tau_{\text{defect}}$.

Criterion values above zero indicate there is a discernible peak at the time lag of the defect. The second term $P_k(R_{xx}(\tau_2))$ can be negative, which may result in a high criterion value even when the autocorrelation value of the time lag of the defect is low. Therefore, the k^{th} percentile in the second term of (10) should be a high percentile. It is well-known that the autocorrelation values of signals composed with noise and sporadic impulses tend to be close to zero, as demonstrated in [46]. In such a case, the median (P_{50}) for example could be an unfounded choice for the second term because it would be close to zero, $P_{50}(R_{xx}(\tau_2)) \approx 0$. Then the second term would practically invalidate itself. In addition to noise and sporadic impulses, it is highly probable that the signal includes frequency components unrelated to the defect. These components may cause high autocorrelation values at the time lag of the defect and at other time lags nearby. In order to distinguish the time lag of the defect from other time lags in such a case, it is important to compare it with high values nearby. However, the second term in (10) does not have to be the maximum value ($P_{100}R_{xx}(\tau_2)$) to make a successful comparison. A practical approach is to select the percentile P_k so that $50 << k < 100$. In this study, $k = 95$.

2.1.6. Demonstration of signal processing stages

The stages for the calculation of the criterion are illustrated in Fig. 3. A 30-second signal from the experiment introduced in Section 3 is used in this demonstration. The raw acceleration signal is shown in plot A. In plot B, the signal is band-pass filtered. Distinct peaks are seen in the filtered signal, but their relation to the defect is unclear.

The tapered parts formed in the filtering procedure contained 10% of the total signal length, as shown in plot B. To remove the tapered parts, 1.5 seconds from both ends of the signal were removed for the next processing stage, which is illustrated in plot C. The filtered signal was segmented into time windows, which have a length related to the bearing defect frequency. In this case, it is 2% of the defect period. Each point in plot C corresponds to 0.056 seconds, which equals nearly 2890 points in the original signal presented in plot A. In plot C, the peaks that repeat with the interval of 50 points (2.8 seconds) are associated with the defect frequency, which is FTF in this case. The peaks extracted by the feature (the ratio of I_{10} to I_1) are clearly magnified in plot C when compared with the filtered signal in plot B.

Finally, plot D gives an illustration for the terms in the proposed criterion (10). The first term ($\max(R_{xx}(\tau_1))$ is the maximum autocorrelation value for τ_1. In this case, it is the value for $\tau_{\text{defect}} - 1$. The second term ($P_{95}R_{xx}(\tau_2)$) is the 95$^{\text{th}}$ percentile of autocorrelation values in the range τ_2.

In this particular signal, the second multiple of the time lag of defect (5.6 seconds) has also relatively high autocorrelation. The multiples of defect time lags can be used for diagnosis purposes in some cases. Considering an automated routine, the use of multiples as defect indications however increases the risk of misdiagnosing, because such components are not undoubtedly related only with the defect.

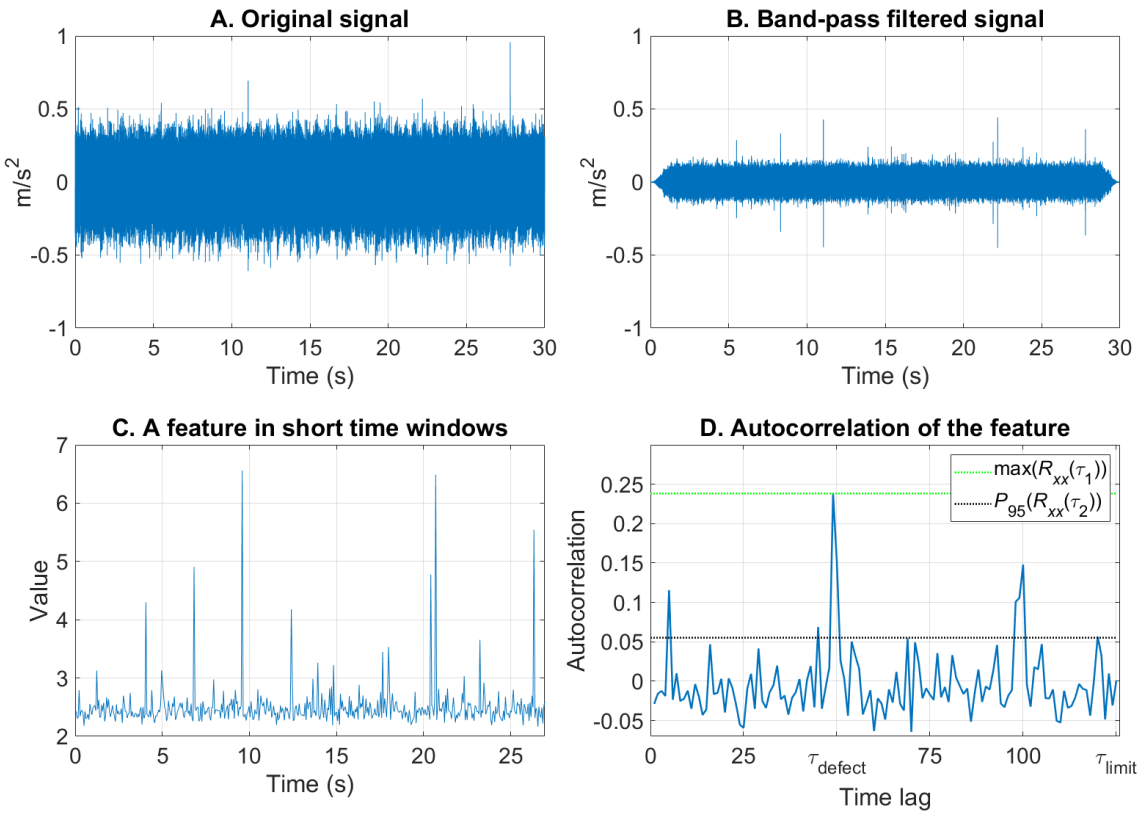

Fig. 3. Demonstration of the stages of the proposed method.

2.1.7. Significance of sampling rate and defect frequency for resolution in segmentation

The sampling rate and the monitored defect frequency together affect the accuracy of the defect period approximation as a result of the segmentation approach presented in Section 2.1.2. The rounding procedure applied in the generation of segments weakens the resolution of the proposed criterion in the practical solution. If the number of data points in a single segment is relatively low, a high error is induced. In other words, the actual defect period (aka inverse of defect frequency) and the approximated defect period differ considerably. The approximated defect period is the sum of fixed-sized segments, i.e. the sum of fifty time windows, when the 2% window size is used.

In order to study the appropriate ratio between the sampling rate and the defect frequency, the following equation was used

$$Number\ of\ data\ points = ceil(|fs \cdot var_1 - var_2 \cdot var_3| - var_4 \cdot fs \cdot var_1), \quad (11)$$

where *fs* is the sampling rate (Hz); var_1 is the inverse of the defect frequency (s); var_2 is the number of segments in a defect period (50 or 100 segments); var_3 is the length of a time window in data points after rounding (positive integer); and var_4 is the percentual window size in decimals (0.02 or 0.01). The resulting number shows the distance of the approximated defect period from the defined window (2% or 1%) around the actual defect frequency. If the number is positive, the approximated defect period is outside the window, which indicates a poor resolution. A negative number indicates that the resolution is adequate.

Plot A in Fig. 4 shows the ratio of sampling rate to defect frequency. The sampling rates 20 and 51.2 kHz were tested and the defect frequencies up to 50 Hz are shown with 0.1 Hz step size. Plots B and C show the results of (11) with 1% and 2% window sizes, respectively. The points where (11) gave higher value than zero for the first time, while the defect frequency increased, are marked in plot A. The ratios associated with these points give indications about the appropriate ratio between the sampling rate and the defect frequency.

Plot A indicates that a ratio larger than 1225 should be used with 51.2 kHz sampling rate in the case of 2% window size. The use of the 1% window requires a higher ratio (>4452). The appropriate ratios are relatively similar in the case of 20 kHz sampling rate, but significantly lower defect frequencies should be diagnosed. It should be mentioned that lower ratios are adequate in some cases, because such ratios may also result in negative values in (11), as shown in plots B and C. However, the resolution is not always adequate then. A poor resolution can be compensated in some cases by using an appropriate search range around the time lag associated with the defect period, as shown in Section 2.1.5.

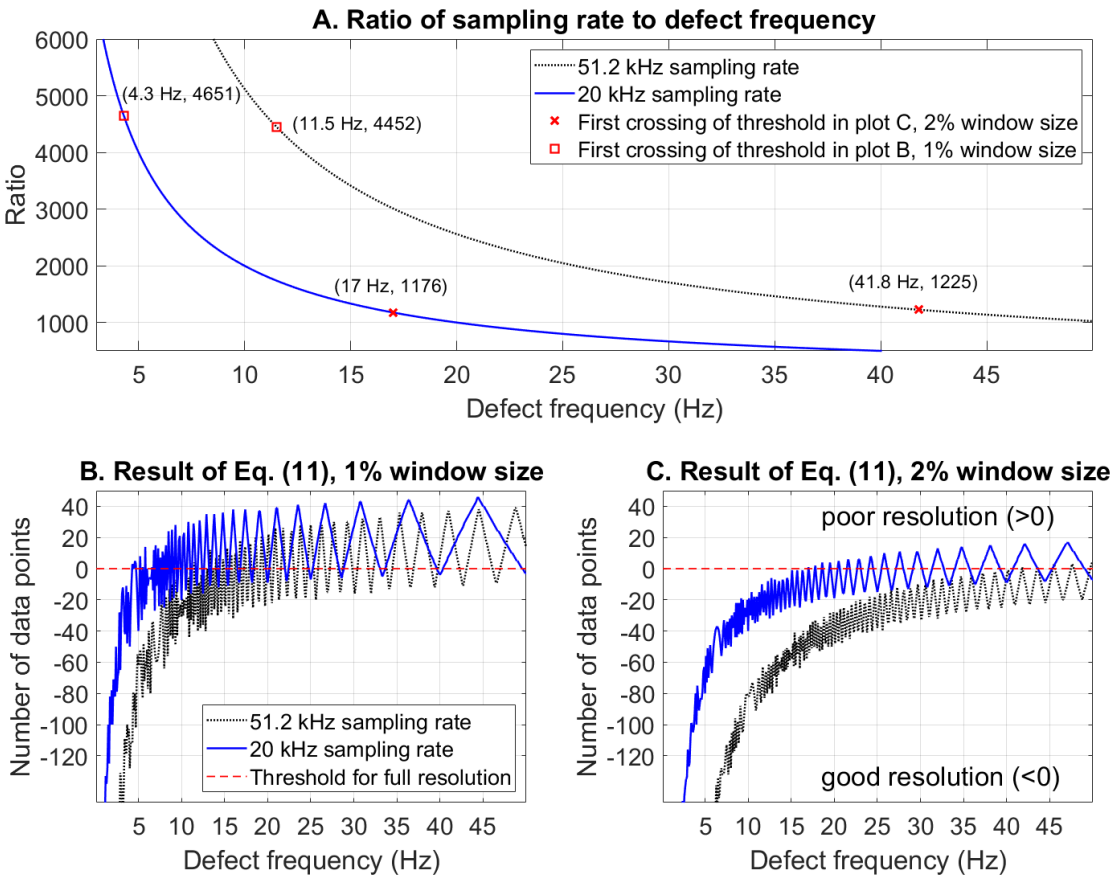

Fig. 4. Effects of ratio between sampling rate and defect frequency on resolution in segmentation. Plot A shows the ratio. Plots B and C show the corresponding effects on the resolution based on Eq. (11).

2.2. Peak ratio and peak-to-median indicators

The following procedure is used for the calculation of envelope spectra in this study. The time domain signal is windowed with a Tukey window using 10% taper. The signal is filtered by the approach explained in Section 2.1.1. The filtered signal is rectified by calculating the absolute values and then the average is subtracted from the rectified signal. The average is subtracted to ensure that the non-zero average does not cause high amplitudes at the low frequencies in the envelope spectrum. Five percent of data points are removed from each end of the filtered signal to remove the tapered parts. Finally, the filtered signal is windowed with a Tukey window using a 10% taper and FFT is used to generate the envelope spectrum.

The peak ratio is a dimensionless ratio applied on the envelope spectrum amplitudes [34]. The peak ratio is defined as the sum of the peak values of the defect frequency and its harmonics over the average value of the spectrum as follows

$$Peak\ ratio = \frac{N \cdot \sum_{j=1}^{n} P_j}{\sum_{k=1}^{N} A_k}, \tag{12}$$

where P_j is the amplitude value of the peak located at the defect frequency or its multiple, A_k is the amplitude of frequency k, and N is the number of points used from the spectrum. The number of peaks in the selected spectrum area is n.

Due to the practical considerations, slight modifications were made on the peak ratio presented in [34]. Firstly, the bearing frequencies can only be estimated in a search range limited by some maximum error ε due to the measurement precision, skidding and speed variations. The peaks are selected from the search ranges $I_j = [j \cdot f(1-\varepsilon), j \cdot f(1+\varepsilon)]$, where f is the bearing defect frequency. In this study, the peak value is the maximum value inside such a search range limited by $\varepsilon = 0.02$, which equals 2% of the considered frequency. Parameter j is restricted to be one or two due to the fact that a large number of harmonics increases the risk of misdiagnosing in an automated approach. The search range used for peak detection increases as the number of harmonics increases, which consequently increases the risk that peaks unrelated to the defect become representative features of the defect. The number of points (N) in the spectrum is restricted to be $(n+1) \cdot n_f$, where n_f is the running number of the frequency component corresponding to bearing defect frequency f.

The peak-to-median ratio is calculated by dividing the maximum peak P_j by the median value of the amplitudes A_k at any frequency except the points inside the search ranges for peaks (I_j). A slightly different form of the indicator was presented in [24].

In the first tested approach, the optimal frequency band for band-pass filtering is selected based on the maximum value given by the peak ratio or peak-to-median indicator. The frequency bands introduced in Fig. 2 are applied in this study. In the second tested approach, the frequency band is selected based on the fast kurtogram. In diagnosis, the condition under analysis is compared with the baseline condition using the same band for filtering.

3. Experimental setup

Section 3.1 introduces the test rig, the rolling element bearing and the defects, which were studied in the laboratory tests. Section 3.2 describes the instrumentation and the measurements including a statistical description of the measured acceleration signals.

3.1. Test rig

The measurements were done in the RWTH Aachen test laboratory. The basic layout of the test rig is sketched in Fig. 5 [12]. The studied bearing is located in between the test drive shafts. In addition, the test rig includes an asynchronous motor rotating the power train comprised of straight shafts, a loading mechanism (retaining device) and a reverse gear with the gear ratio 5:4. The rotational speed, the load of the system and the angle of the test drive shafts can be adjusted in the test rig. Twelve test rig states were used in the study of three bearing conditions, which included the baseline condition, initial defect and worsened defect. The settings for each run are shown in Table 2. Three different speed settings were applied in the asynchronous motor. In half of the test runs, the test drive shafts were straight and in the other half, they had 5° angle. Considering the bearing monitoring, it is assumed that the angle of the load in the radial plane is fixed in Eqs. (1)–(4) in the test runs with different shaft angles. Half of the test runs had a small load applied in the system by using the loading mechanism, which is shown in photo A in Fig. 6. Unfortunately, the system had no measurement showing the absolute value of the load. However, it was made sure that the system had a fixed load in the loaded states by using a torque wrench. The maximum torque that can be applied in the test rig is 10 kNm. The load applied in the tests had a small effect on the shaft speed, which was slightly lower than in the unloaded tests. This is demonstrated in Table 3, which shows the mean and variation (two standard deviations) of the drive shaft speed.

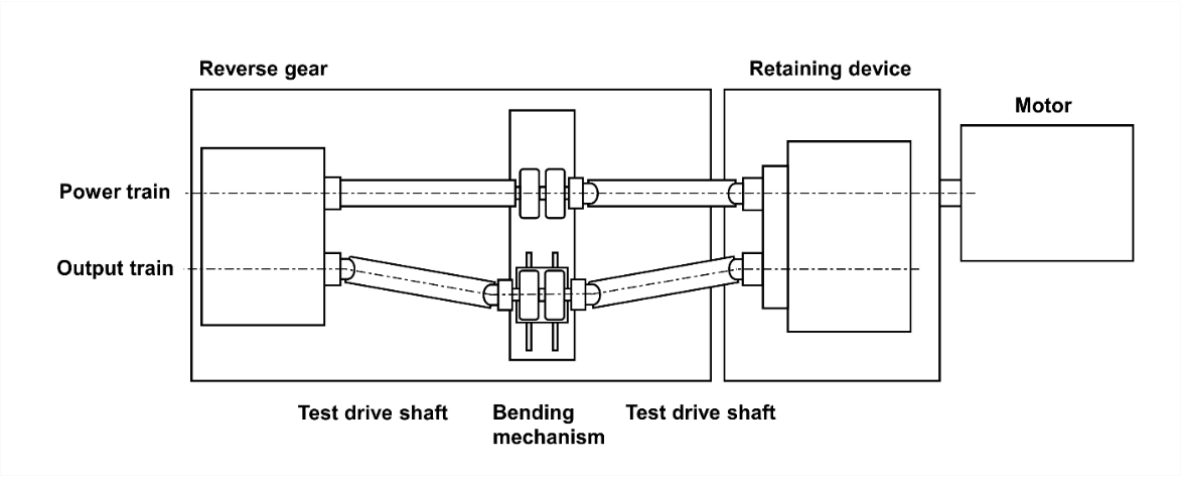

Fig. 5. Drawing of drive shaft test rig (modified from [12]).

Table 2 Test rig settings in the test runs.

Test rig state	Motor speed setting (rpm)	Shaft angle (degrees)	Load
1	30	0	no
2	60	0	no
3	90	0	no
4	30	5	no
5	60	5	no
6	90	5	no
7	30	0	yes
8	60	0	yes
9	90	0	yes
10	30	5	yes
11	60	5	yes
12	90	5	yes

Table 3 Measured rotational speed ($\mu \pm 2\sigma$) of the drive shaft next to the bearing based on 60-second samples (in rpm).

Test rig state	Baseline	Initial defect	Worsened defect
1	24.21 ± 0.54	23.83 ± 0.46	24.23 ± 0.44
2	48.42 ± 0.64	48.01 ± 0.38	48.45 ± 0.47
3	72.62 ± 0.93	72.20 ± 0.47	72.63 ± 0.66
4	23.97 ± 0.65	23.93 ± 0.65	23.70 ± 0.74
5	48.15 ± 0.63	48.09 ± 0.40	47.85 ± 0.60
6	72.23 ± 0.89	72.26 ± 0.43	71.98 ± 0.73
7	23.57 ± 0.59	23.61 ± 0.53	22.94 ± 0.58
8	47.70 ± 0.61	47.79 ± 0.36	47.16 ± 0.53
9	71.86 ± 0.90	71.92 ± 0.45	71.30 ± 0.68
10	23.79 ± 0.95	23.22 ± 1.09	23.36 ± 0.91
11	47.92 ± 0.70	47.35 ± 0.51	47.51 ± 0.67
12	72.05 ± 0.92	71.45 ± 0.50	71.60 ± 0.78

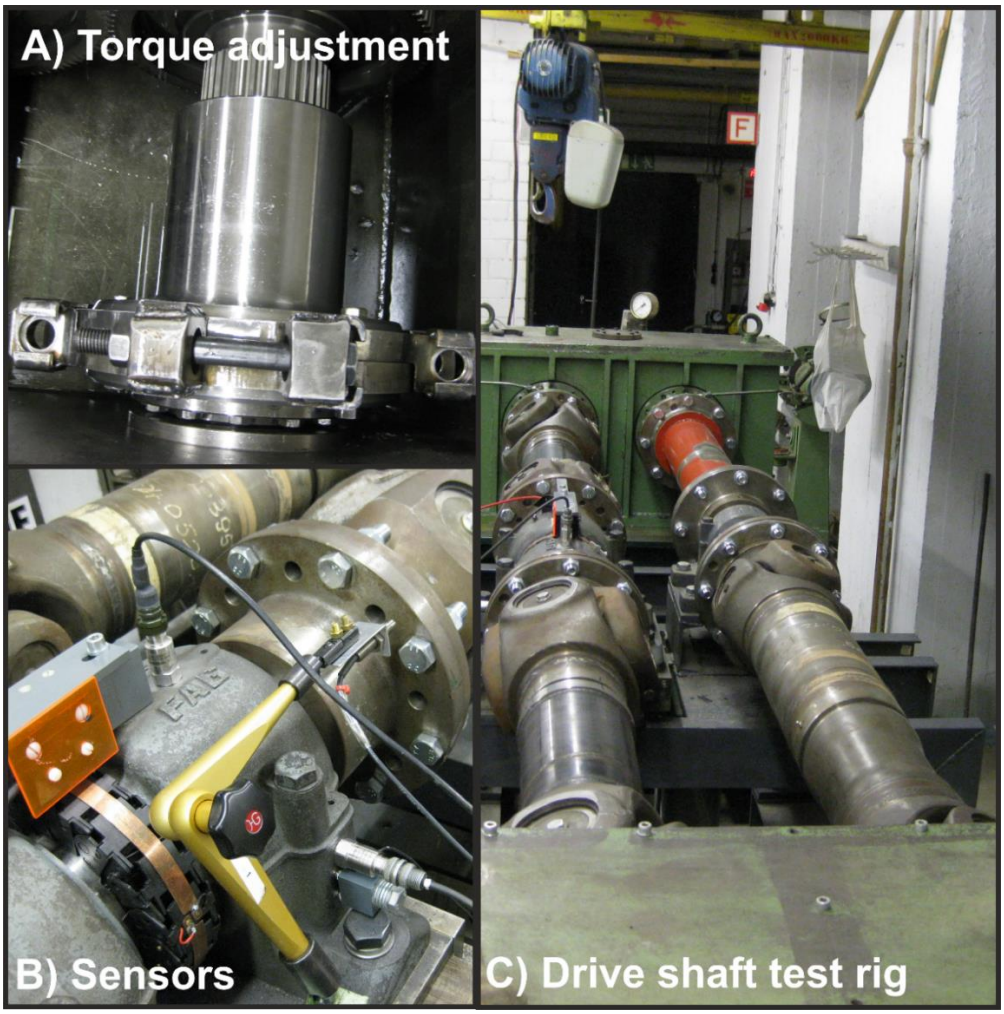

Fig. 6. Photographs of the drive shaft test rig. Photo A shows the loading mechanism of the test rig. The accelerometer used for vibration measurements is mounted on top of bearing housing and the eddy current sensor is located next to the bolts in the drive shaft in photo B. The test rig is depicted from the angle of the motor in photo C.

The bearing type in the test rig was FAG 23028-E1A-K-M + H3028. The defect frequencies related to 60 rpm are 11.1173, 13.8827, 4.4178, and 0.4447 (Hz) for BPFO, BPFI, BSF, and FTF, respectively. To generate the initial defect, one rolling element from both sides in the bearing was slightly damaged by drilling with a multipurpose tool. In the case of worsened defect, the other rolling element was damaged even more. Fig. 7 shows the defects from one side of the bearing.

3.2. Instrumentation and measurements

The accelerometer used for the vibration measurements was IMI Sensors 622A01, which has a frequency response from 0.2 to 10000 Hz with ±3 dB deviation. The measurements were done in the vertical direction and the sensor was stud-mounted on top of the bearing housing, as shown in photo B in Fig. 6. The measurement hardware included NI9234 data acquisition card. The sampling rate was 51.2 kHz and the only filter used at the hardware level was the built-in antialiasing filter of the data acquisition card.

The shaft speed was measured by an eddy current sensor by recording the pulses caused by the bolts in the rotating shaft. The arrangement is shown in photo B in Fig. 6. NI cRIO-9215 data acquisition card was used and the sampling rate was 10 kHz in the shaft speed measurements. LabVIEW software was used for the recording of the signals.

Fig. 8 shows the l_2 (root-mean-square) and excess kurtosis values of the 60-second acceleration signals to give an illustration of the signals from each test rig state. The values of l_2 describe the signal amplitude level, whereas the kurtosis values give an insight into the tails of the distribution. The l_2 values clearly indicate that the amplitude level increased together with the shaft speed, but the increase was smaller when the test drive shafts were bent. According to the excess kurtosis, the baseline condition at the lowest shaft speed had the most leptokurtic distribution. Otherwise, the differences between the signals in different bearing conditions were generally small based on these statistical features.

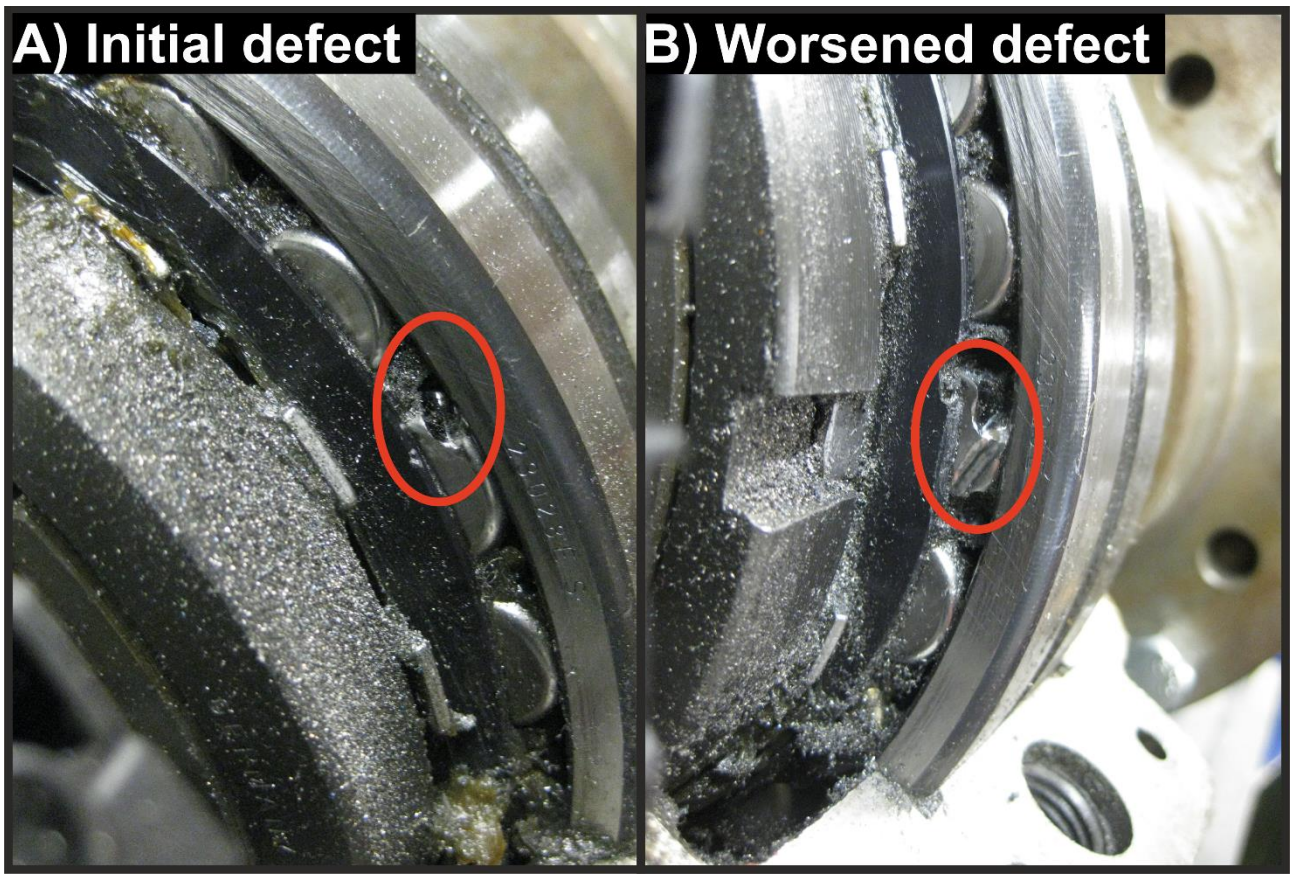

Fig. 7. Defects in the rolling element of the bearing. The initial defect in the roller element is shown on the left and the worsened defect in the same roller element is shown on the right.

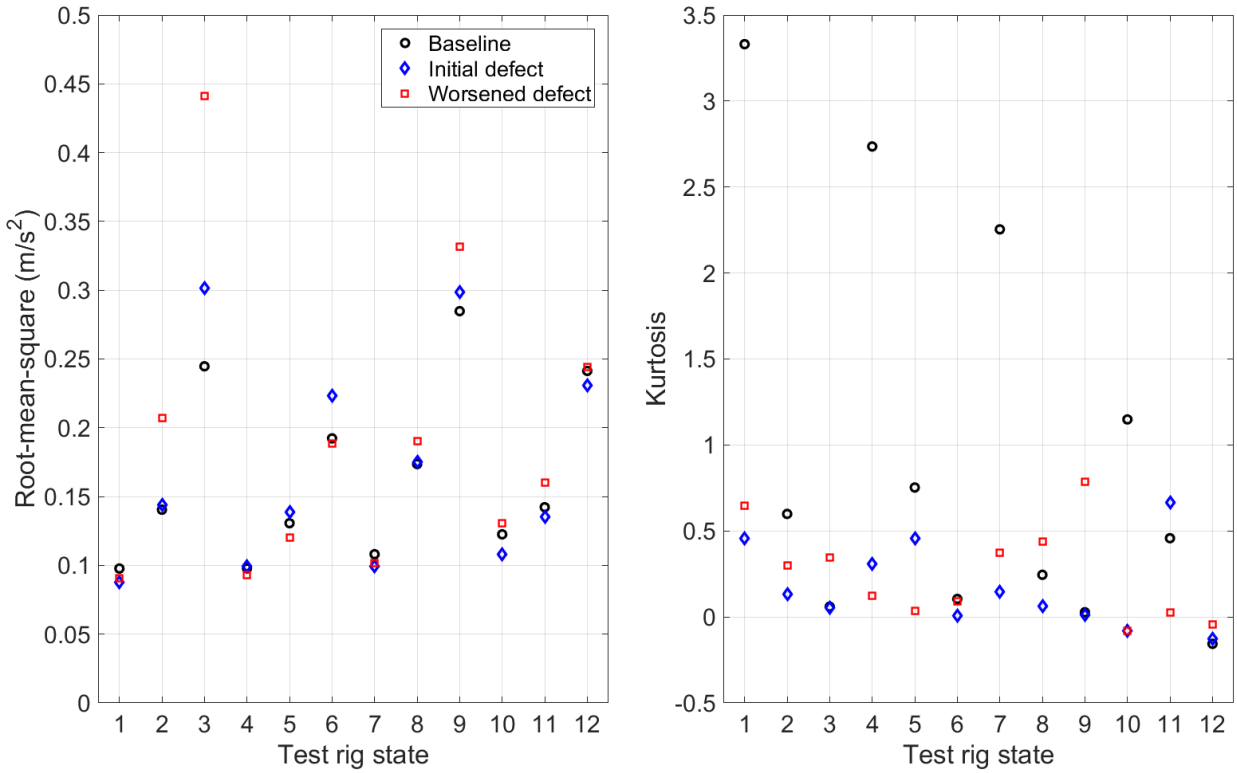

Fig. 8. Root-mean-square (on the left) and excess kurtosis (on the right) computed from 60-second samples of the acceleration signals.

4. Results and discussion

The performance of the methods was evaluated based on the bearing fault simulations and laboratory test data. Section 4.1 presents the results from the simulations. The proposed method is first compared with the reference approaches based on the bearing fault simulations. Thereafter, a sensitivity study is made for the proposed method and the computation time of the method is studied. Results from the laboratory tests are presented in Section 4.2. Finally, Section 4.3 discusses the results presented in the previous sections and outlines the applicability of the tested methods in automated bearing diagnosis.

The full bandwidth and the 107 frequency bands shown in Fig. 2 were used in the frequency band selection stage of the criterion based on the autocorrelation and the peak ratio and peak-to-median based selections. Tukey window with 10% taper was used before the application of FFT on these methods. In the frequency band selection based on the fast kurtogram, six decomposition levels were used which stands for 220 frequency bands altogether [33].

In the criterion based on autocorrelation, both window sizes i.e. 1% and 2% of the defect period, were tested. Unless stated otherwise, the window size that gave the highest criterion value was used in the diagnosis.

4.1. Simulation study

4.1.1. Comparison of the bearing diagnosis methods

An inner race defect was studied on a simulated system in Matlab®. A gear was included in the system to produce harmonic effects that could be included also in a practical application and to see if these simplified gear signals influence the performance of the tested diagnosis approaches. The BPFI was calculated using the formula (2). The number of rolling elements in the bearing was n = 8 with the diameter d = 0.02 (m). The pitch diameter was D = 0.12 (m) and the contact angle was ϕ = 15 (º). Each defect-related shock was modelled as a 3 kHz sinusoid that had amplitude 1. The sinusoid was windowed by a Kaiser window with the length L = 100 points and *beta* = 40 in the Kaiser function in Matlab®. The sampling rate in the simulations was 20 kHz, and therefore, the length of the simulated shock was 5 milliseconds. The shock was made periodic by convolving it with a comb function that had value 0.33 at the periods corresponding to the BPFI and zero elsewhere.

The defect was studied with input shaft speeds 20, 40, 60, and 300 rpm, which correspond to f_r = {⅓, ⅔, 1, 5} Hz. The corresponding defect frequencies were 1.548 Hz, 3.096 Hz, 4.644 Hz, and 23.220 Hz, respectively. In addition to the shocks, the signal included sinusoids corresponding to the gear mesh frequency, input shaft speed and output shaft speed with the amplitudes 1, 0.4, and 0.2, respectively. The gear mesh frequency was $f_{mesh} = f_r \cdot N_p$ and the output shaft frequency was $f_{gear} = f_r \cdot N_p / N_g$. The number of teeth on the pinion (input) was N_p = 13 and the number of teeth on the gear was N_g = 35. Normally distributed noise with variance σ^2 = 0.04 was added on each signal. Similar signals without the shocks caused by the defect were created as the baseline signals for comparison. All the signals had the length of 20 seconds.

Fig. 9 presents the amplitude spectra of each signal with the defect. The defect is visible around 3 kHz frequency on the 60 and 300 rpm cases and masked behind the noise in the lower speed cases. Table 4 shows the frequency bands that were selected by each of the tested methods in the BPFI diagnosis. The deterministic gear frequencies apparently had no effect on the diagnosis results, because they were filtered out by each method during the frequency band selection stage. However, more complicated gear effects are likely to happen in real applications.

The criterion based on autocorrelation identified a frequency band around 3 kHz in each case. Interestingly, this frequency band matches with the frequency of the sinusoids modelling the shocks. The peak ratio and peak-to-median based approach resulted in a relatively high frequency band in the 20 rpm shaft speed. The selection of the band was presumably influenced by the small but indiscernible variations of the random noise in the signal (see Fig. 9). The frequency bands presented in Table 4 are based on the peak ratio indicator with two harmonics, because it had the highest value in each case. Furthermore, the largest differences between the defect and baseline conditions were obtained with this indicator in each rotational speed.

The fast kurtogram resulted in a relatively high frequency band with the shaft speeds 20 and 40 rpm, which indicates that these bands had the highest kurtosis in lowest shaft speeds. The higher shaft speeds resulted in the selection of frequency bands close to 3 kHz. The maximum excess kurtosis values of the complex envelopes [35] based on the fast kurtogram were 0.188, 0.183, 0.261, and 0.770 from the slowest to the highest shaft speed, respectively.

In the criterion based on autocorrelation, the selected statistical features were I_{10}, I_4, $P_{95}(|x^{(2)}|)$, and I_4 on the shaft speeds from the lowest to the highest, respectively. This result indicates that a single feature, which would be optimal for all conditions, did not exist. The selected segment size was 2% of the defect period in each case.

The values of the indicators used in the diagnosis are given in Fig. 10. The upper part of Fig. 10 shows the peak ratio values, when the optimal frequency bands were selected based on the peak

ratio and the fast kurtogram. The lower part shows the values of autocorrelation at the defect time lags and the values obtained by the criterion based on autocorrelation. Fig. 10 demonstrates that the peak ratio shows the defect clearly only on the two highest shaft speeds when the value from the defect is compared with the value from the baseline condition. The slightly higher values in the defect condition in 20 and 40 rpm speeds do not give a clear basis for unambiguous defect detection. In the case of the autocorrelation-based criterion, the difference between the defect and baseline conditions is more obvious. However, the values of the criterion are low. The autocorrelation of the defect time lag and the proposed criterion reveal the defect equally clearly in these simulated cases.

The autocorrelation values of the selected time domain features (I_{10}, I_4, $P_{95}(|x^{(2)}|)$, and I_4) for each rotational speed are shown in Fig. 11 in order to illustrate how successfully the defect time lag is separated from other time lags. A distinguishable peak can be seen at the time lag associated with the defect period (e.g. the inverse of the defect frequency) in each case in the defect condition. Figs. 12 and 13 show the envelope spectra, which had the filtering bands selected based on the peak ratio and the fast kurtogram, respectively. These plots confirm that the defect became diagnosable only in 60 and 300 rpm shaft speeds. Distinct peaks emerged at the defect frequency and its multiples in those cases.

Table 4 Selected frequency bands (Hz) for filtering based on the 20-second signals.

Shaft speed (rpm)	20	40	60	300
Crit. based on autocorrelation	2000–4000	2000–4000	2000–4000	2000–4000
Peak ratio (two harmonics)	7659–7992	2997–3330	2000–4000	2000–4000
Fast kurtogram	6667–10000	6667–10000	2500–3125	2917–3333

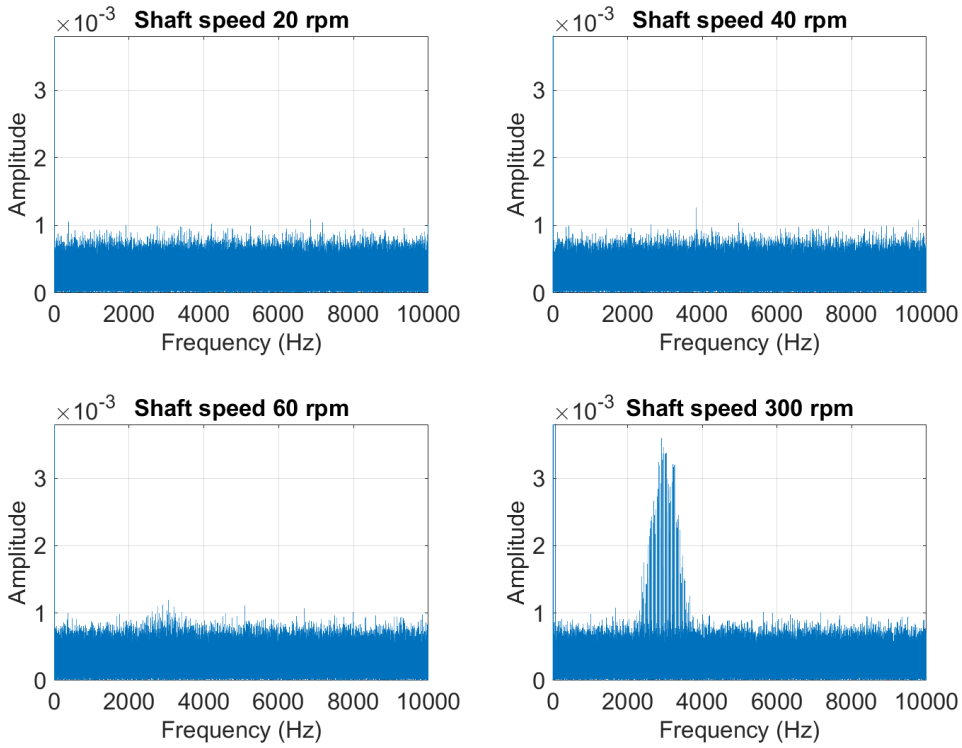

Fig. 9. Amplitude spectra of the simulated 20-second signals with an inner race defect. The effect of the defect is visible around the 3 kHz region in the shaft speeds 60 and 300 rpm.

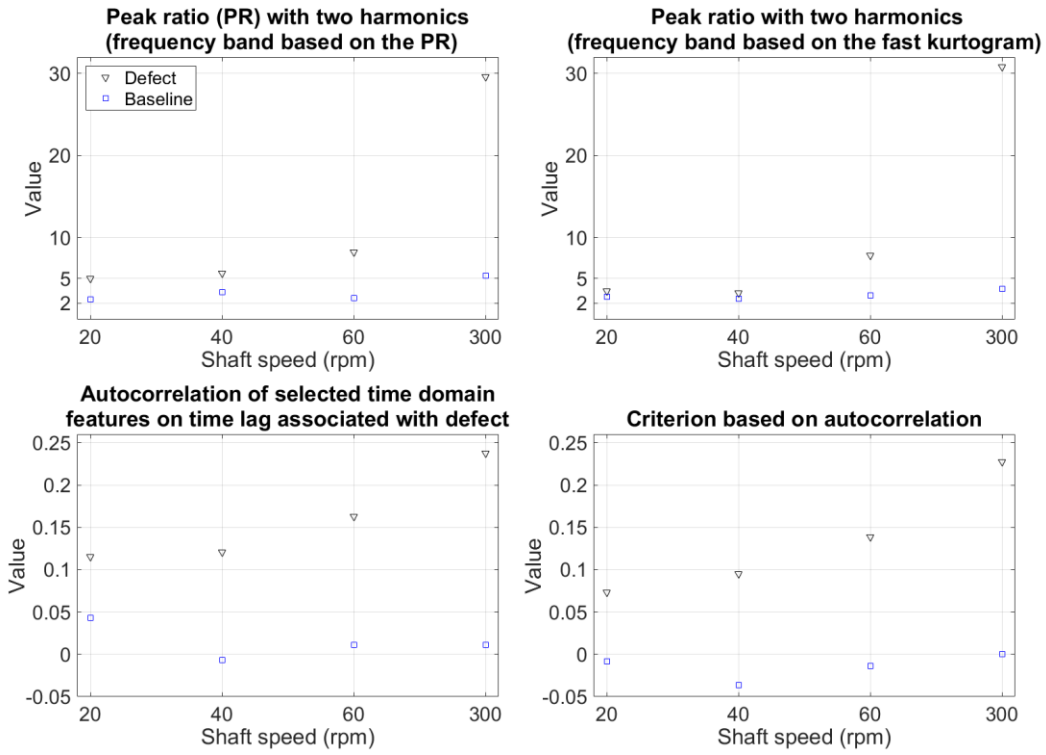

Fig. 10. Values of indicators in defect and baseline conditions.

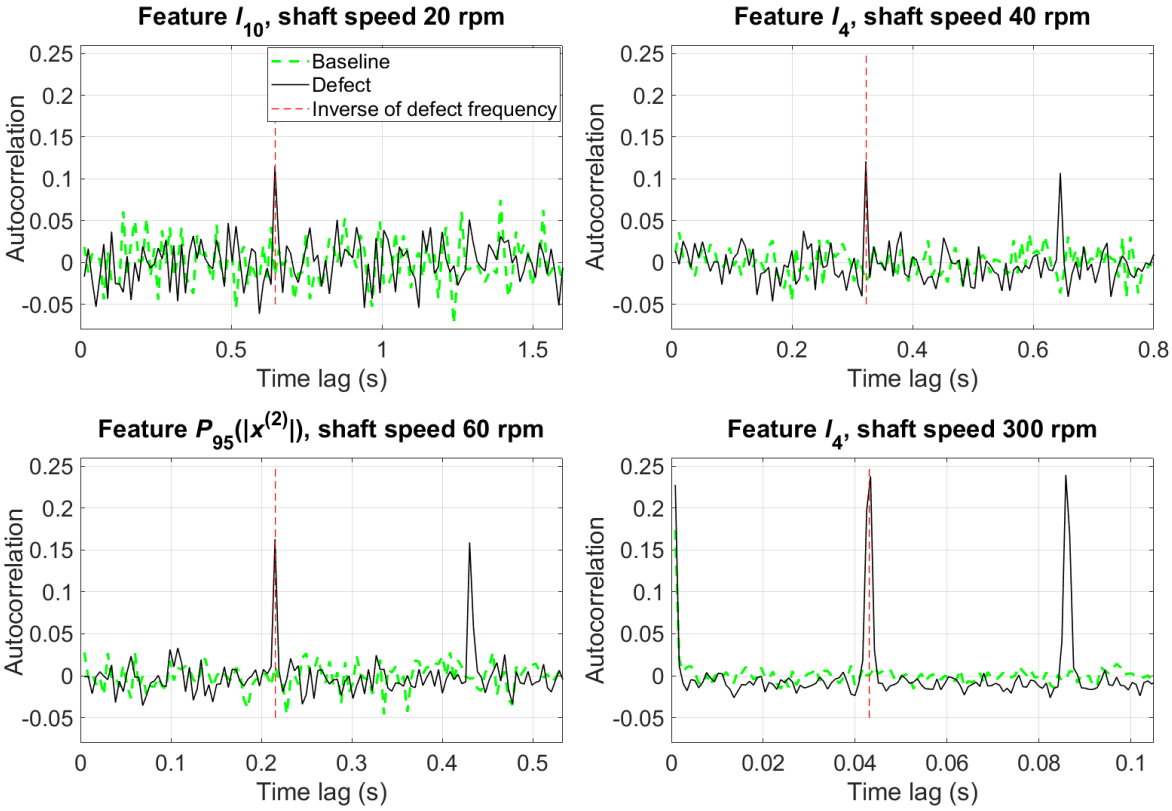

Fig. 11. Autocorrelation of the selected time domain features from the signals simulating different shaft speeds.

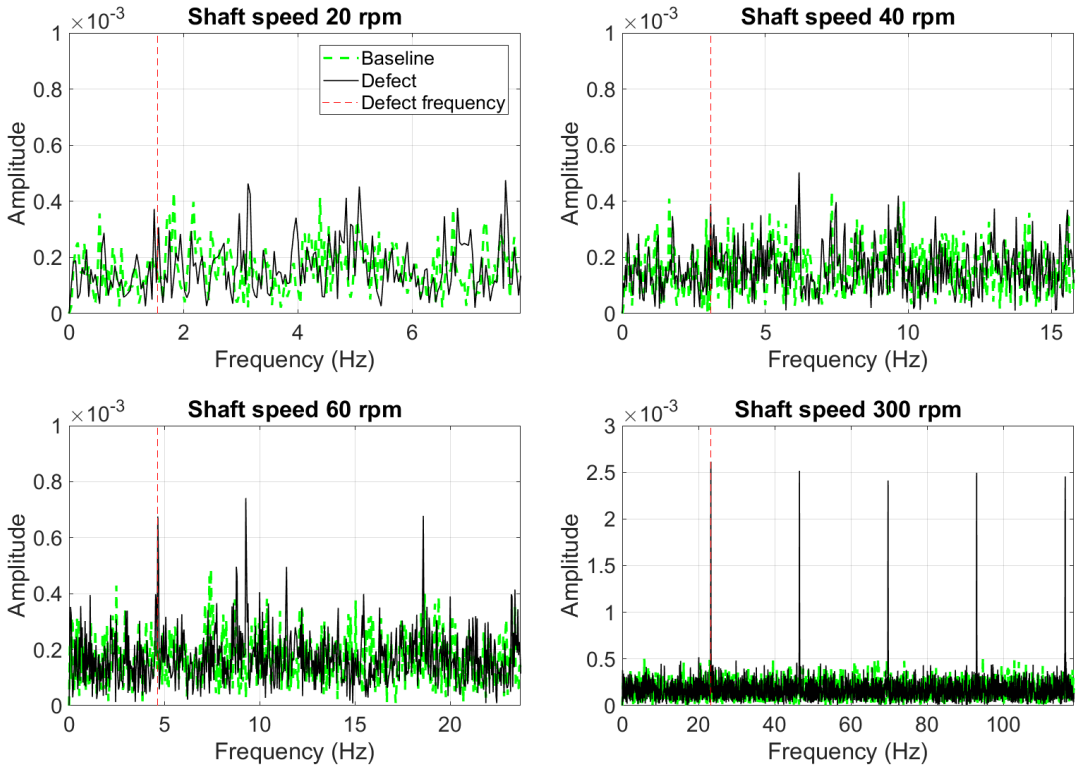

Fig. 12. Envelope spectra from the signals simulating different shaft speeds using the frequency bands selected based on the peak ratio indicator.

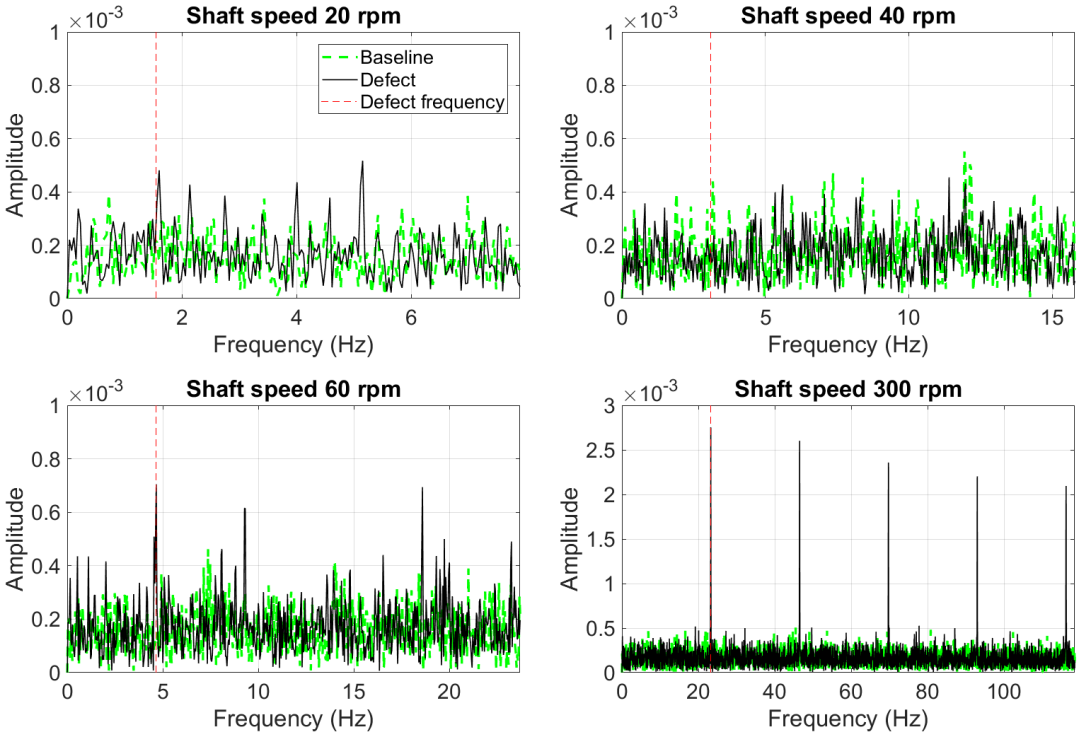

Fig. 13. Envelope spectra from the signals simulating different shaft speeds using the frequency bands selected based on the fast kurtogram.

In order to demonstrate the influence of signal length on the autocorrelation and envelope spectrum, an additional test was done on 300-second signals simulating the 20 rpm case. In this case, both the proposed criterion and the peak ratio with two harmonics resulted in the selection of the 2-4 kHz frequency band. The proposed criterion value was 0.10 and the corresponding baseline value was -0.02. The value of peak ratio with two harmonics was 9.84 and the corresponding baseline value was 4.12. The results indicate that the performance of the approach based on envelope spectrum improved when the signal length was significantly increased (15×) from the 20-second length. The results of the criterion based on autocorrelation did not have such a substantial change. The autocorrelation of the selected feature with 2% segment size and the envelope spectrum are illustrated in Fig. 14.

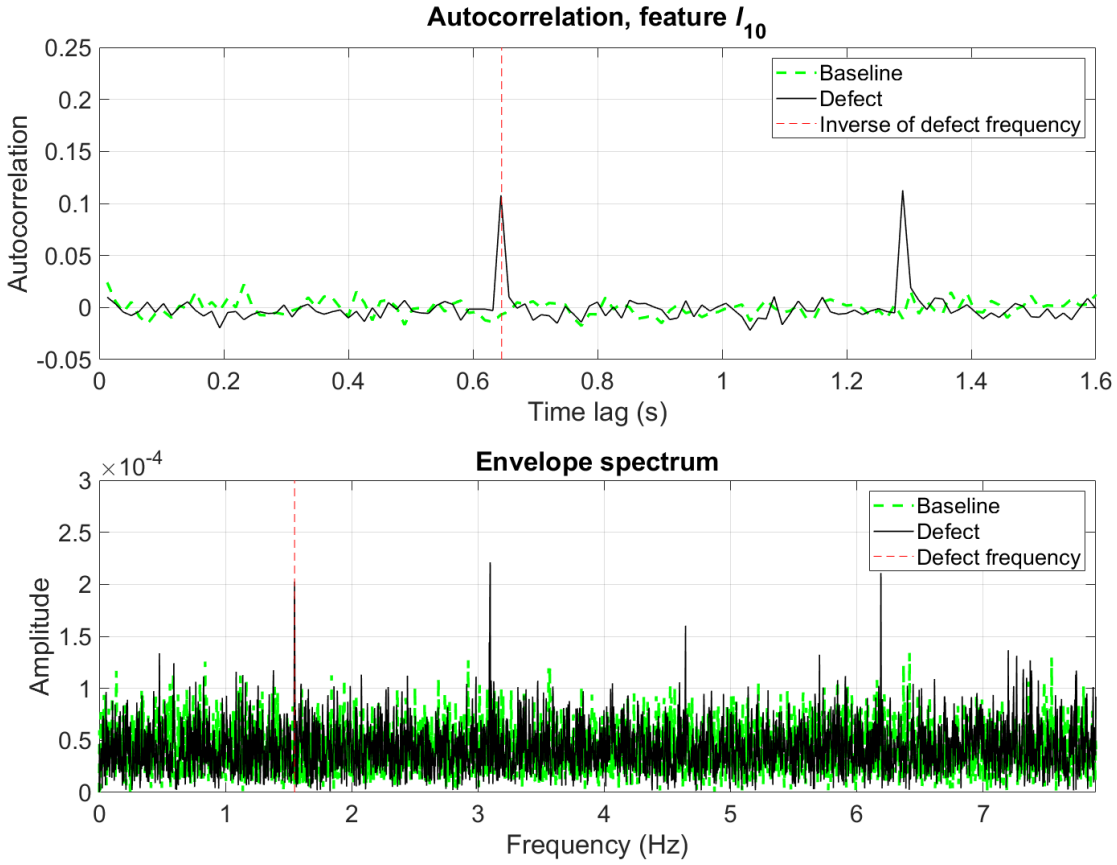

Fig. 14. Autocorrelation of feature I_{10} (above) and the envelope spectrum (below) using 300-second signals simulating the 20 rpm shaft speed case.

4.1.2. Sensitivity study of the criterion based on autocorrelation

Simulations were done to study the sensitivity of the proposed criterion on the number of shocks and their magnitudes. Moreover, both window sizes (1% and 2%) were used separately to study the effect of the window size. The simulation settings were otherwise the same as in the simulations in Section 4.1.1 with the 20-second signals, but in this case, the comb function amplitude varied from 0.1 to 1 with the step size 0.1 to simulate varying shock magnitudes. Two shaft speeds, 20 and 60 rpm, were tested to illustrate the effect of the number of shocks on the criterion value. The lower

speed case included 30 full defect periods while the defect frequency was 1.548 Hz. Similarly, the higher speed case included 92 full defect periods.

Fig. 15 shows the criterion values with the studied settings. The results reveal that the criterion value increased together with the defect amplitude. The 2% window size resulted in a higher criterion value than the smaller window size each time.

The number of shocks in the monitored period had only a small influence on the results, which is shown by the small deviation between the values calculated from the cases with different shaft speeds with a comparable window size. The higher shaft speed, i.e. more shocks, resulted in a slightly higher value of the criterion in most cases. However, the lower shaft speed resulted in higher values in some cases with the 2% window size, including the cases with defect amplitudes 0.1, 0.2, 0.5, and 0.8. Based on Fig. 15, it can be concluded that the defect amplitude had a larger influence on the criterion value than the signal length.

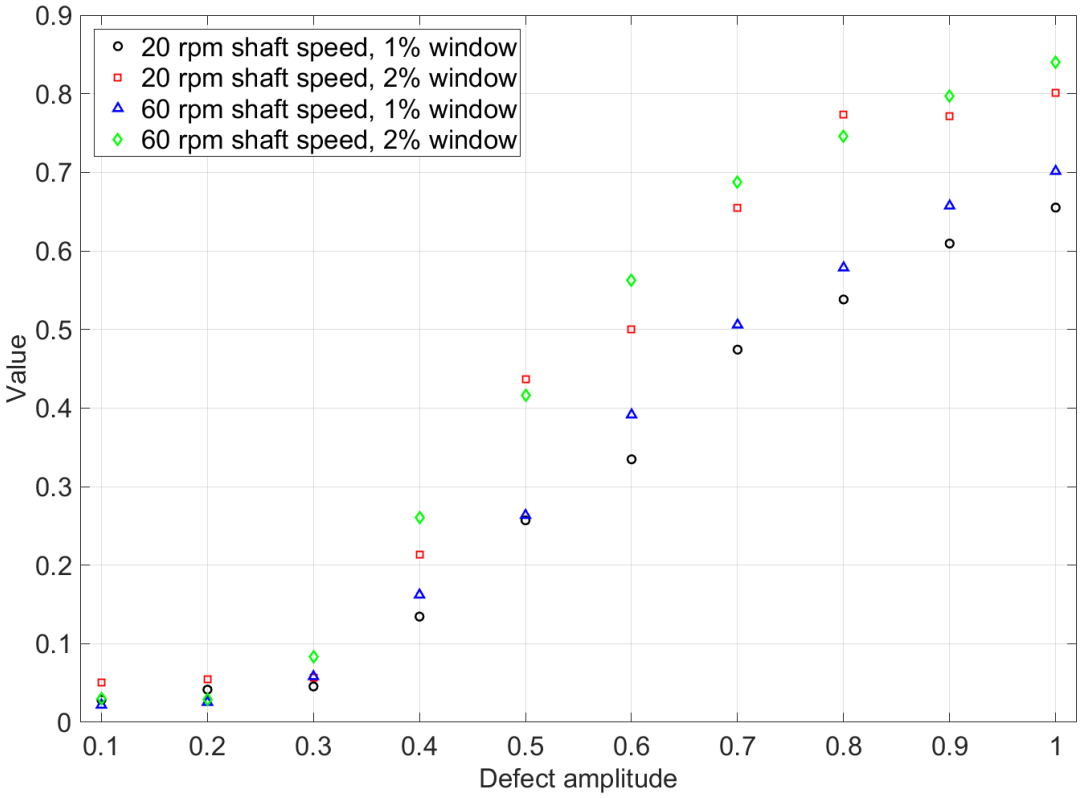

Fig. 15. The criterion values from simulations with two shaft speeds using 1% and 2% window sizes. The defect amplitude on the horizontal axis is equal to the amplitude of the comb function used in the inner race defect simulation.

4.1.3. Computation time of the criterion based on autocorrelation

Simulations were done to evaluate the computation time used by the proposed criterion. The computation time is directly related to the length of a signal, and therefore, each simulation was done on a signal that had the length of ten defect periods. Four defect frequencies {0.1, 0.2, 1, 5} Hz from the low-speed circumstances were studied. The corresponding signal lengths were therefore {100,

50, 10, 2} seconds, respectively. Two sampling rates, 20 and 51.2 kHz, were applied to illustrate the influence of the number of data points on the computation time.

Two servers administered by the University of Oulu, which are named here as servers A and B, were applied in the simulations to increase the objectivity of the results. Server A includes four 2.13 GHz dual-thread 8-core E7-4830 Xeon processors with 256 GB memory. Server B includes two 2.8 GHz 10-core E5-2680v2 Xeon processors with 512 GB memory. The computations were done with Matlab® R2019a by using a single processor and by using 12 workers for parallel computing.

The computation times given in Fig. 16 are the mean values of ten simulations. The results reveal that computation was faster with the larger window size (2%), which can be explained by the lower number of segments in a signal. In addition, the computation was faster with the lower sampling rate (20 kHz) due to the smaller number of data points in a time window compared with the case with the higher sampling rate (51.2 kHz). The results show that the computation becomes faster as the defect frequency increases. This can be explained by the fact that the number of data points in a time window decreases when the frequency increases. In addition, the results clearly show that the type of computer and the parallelism of computation both have a major influence on the computation time in general.

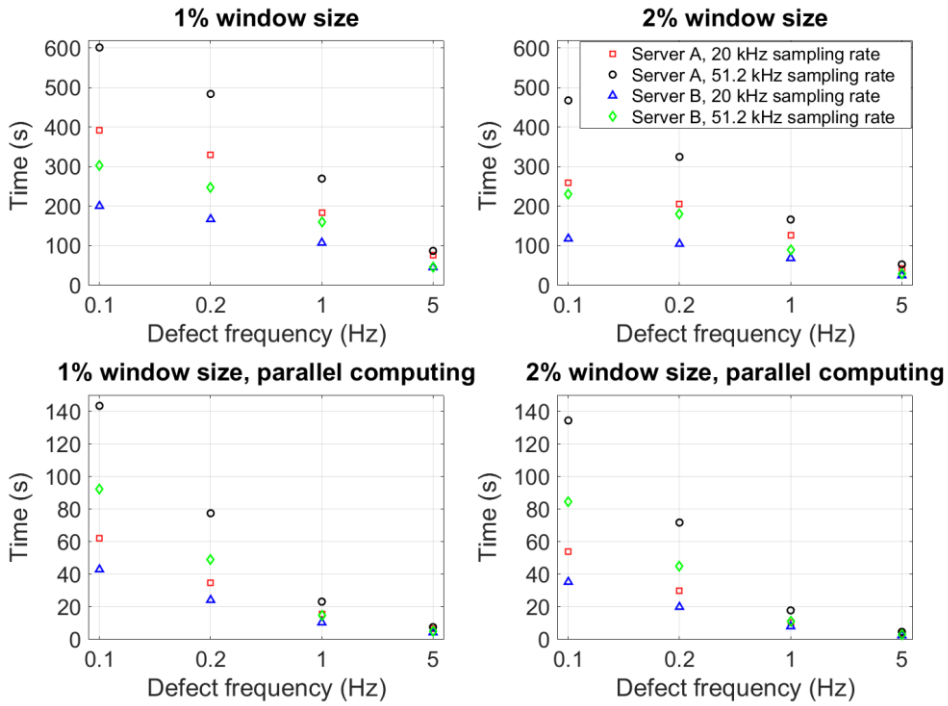

Fig. 16. Average computation time spent in the computation of one criterion value. The upper plots show the computation time with a single processor, whereas the lower plots show the computation time when 12 workers were applied in parallel computing.

4.2. Validation on experimental data

All the bearing defect frequencies (1) – (4) were analysed, because in practice the nature of the developing defect would be unknown in automated diagnosis. The defect frequencies were calculated based on the mean drive shaft speeds shown in Table 3 and the reference frequencies given in Section 3.1 for the 60 rpm speed. Section 4.2.1 reports the analysis results for each defect

frequency by focusing on the difference of diagnostic values between the damaged and undamaged bearing. The emphasis is on the detection of the correct type of bearing defect. Only the rolling elements were damaged and therefore the defect should be presumably shown on the BSF or FTF frequencies [48].

Section 4.2.2 provides detailed results from the diagnosis of one bearing defect frequency (FTF) by using the criterion based on autocorrelation. The reference results for the FTF diagnosis obtained by the automated envelope spectrum diagnosis approaches are given in Section 4.2.3. Finally, Section 4.2.4 compares the frequency band selections done by the applied methods in the FTF diagnosis. These three sections focus on the parameter selection behind the automated diagnosis of one defect frequency.

The bearing condition was studied based on 30-second and 60-second signals separately in order to evaluate the effect of signal length on the indicator values in a practical application. The upper limit for the applicable frequency range in signal filtering was set 9 kHz, which is 1 kHz below the upper limit of the ±3 dB frequency response range specified by the manufacturer of the accelerometer (see Section 3.2). To be on the safe side, only the frequencies below this limit were accepted in the frequency band selection. In addition, the full bandwidth signals were tested.

The frequency bands for filtering and the statistical features were selected based on the damaged conditions. The selections were done for each test rig state, both signal lengths and both damaged conditions separately. The same settings were then used in the baseline condition to obtain the reference values for comparison. The same approach was used for all the tested methods.

4.2.1. Diagnosis of all defect frequencies

Fig. 17 presents the values of the autocorrelation-based criterion in the initial defect condition (left), baseline condition (middle), and the difference of the values (right). The highest value was obtained on the fundamental train frequency (FTF) in each test rig state in the initial defect condition. The corresponding values were lower in the baseline condition. However, the baseline values were relatively high in test rig states 3 and 9 as well. The values of other defect frequencies were lower than 0.1 in the damaged condition except for test rig state 3 where BPFO was above this value. The differences between the criterion values from the 30-second and 60-second signals are mainly a result of different parameter configurations, which are omitted from this section. Considering the difference between the defect and baseline conditions, FTF had a larger difference than the other defect frequencies in most test rig states. However, the difference values should not be examined alone because the baseline signals often had negative values as shown in Fig. 17.

Fig. 18 shows the values of the autocorrelation-based criterion in the worsened defect and baseline conditions and their differences. Like in the initial defect condition, the highest value was obtained by FTF in each test rig state in the case of worsened defect. However, the deviation from the values of BPFI, BPFO, and BSF was small in the test rig states with 5° angle in the drive shafts (test rig states 4–6 and 10–12). The baseline value was generally near zero or negative, but in two cases it was above 0.1. Apart from the test rig state 5 with the 60-second signal, an increased value for FTF was obtained in each case in the damaged condition when compared with the baseline condition.

Based on Figs. 17 and 18, the defect frequencies BPFI, BPFO, and BSF gave weak signs or no signs of the defect. BPFO had an increased value (>0.1) in the test rig state 3 in the damaged conditions, but the value was lower than the value for FTF. In the case of worsened defect, BSF was slightly increased (≈0.13) when compared with the baseline values (<0) in test rig state 8. However, BSF did not generally have substantially increased values even though the rolling element was damaged.

The signal length had some influence on the results. Figs. 17 and 18 clearly indicate that the largest effects were seen on FTF. However, the signal length, which would generally give the clearest damage indication cannot be concluded based on these results.

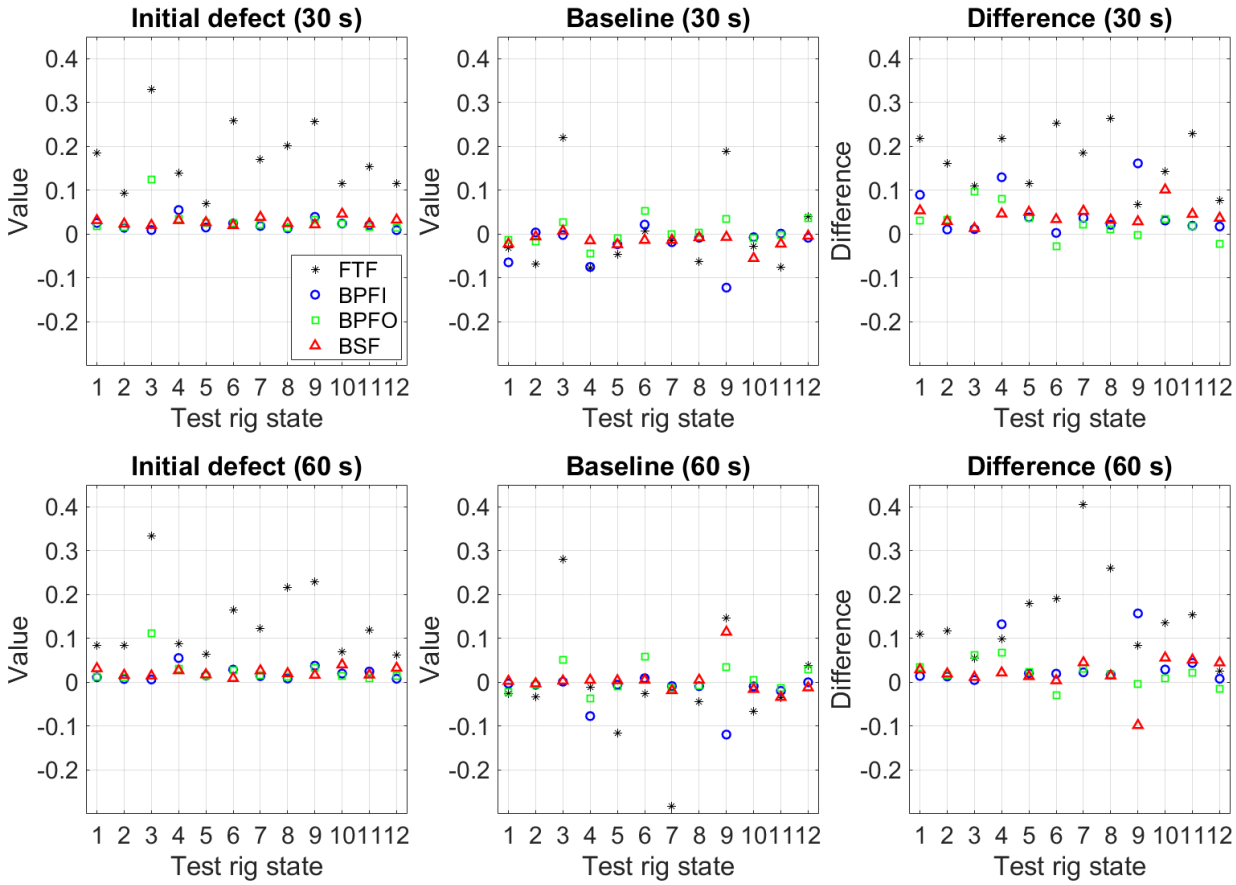

Fig. 17. Values of criterion based on autocorrelation in twelve test rig states. The initial defect is shown on the left, the corresponding baseline condition is shown in the middle and their difference is shown on the right.

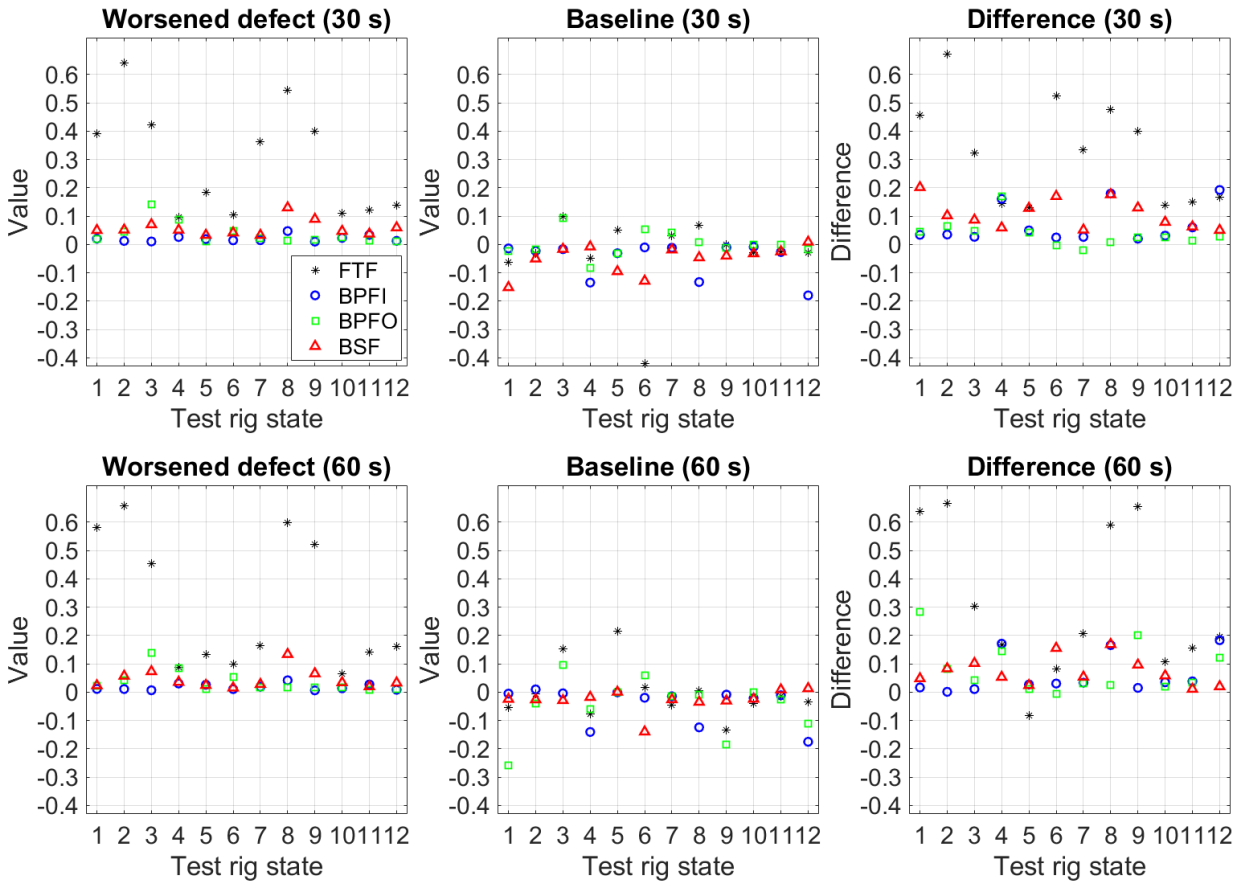

Fig. 18. Values of criterion based on autocorrelation in twelve test rig states. The worsened defect is shown on the left, the corresponding baseline condition is shown in the middle and their difference is shown on the right.

Fig. 19 shows the differences between the defect and baseline conditions based on the peak-to-median with two harmonics. The frequency band was selected based on the maximum value of the same indicator in the defect condition. The positive values in Fig. 19 indicate that the indicator had a higher value in the defect condition than in the baseline condition. The negative values indicate an opposite result.

The highest value was obtained by different defect frequencies depending on the test rig state in consideration. In many test rig states, the highest difference was obtained based on defect frequencies, such as BPFO and BPFI, which are unrelated to the actual damage. However, the defect-related frequencies BSF and FTF had the highest values in certain test rig states as well. In conclusion, the relatively high values on defect frequencies that are unrelated to the actual damage increase the risk of misdiagnosing. A similar risk of misdiagnosing is present in the peak-to-median with one harmonic and peak ratio indicators. Those results are omitted in this section due to their low additional evidential value.

Like in Fig. 19, the differences between the bearing conditions based on the peak-to-median indicator are shown in Fig. 20, but the frequency band for band-pass filtering was selected by using the fast kurtogram. The differences between the defect and baseline conditions are generally lower in Fig. 20 when compared with Fig. 19. This indicates that the bearing defect was not shown clearly in most of the test rig states in Fig. 20. Moreover, BPFI or BPFO had the highest difference in many cases, which increases the risk of misdiagnosing. The results indicate that sound diagnosis

conclusions cannot be done based on this data set by using the fast kurtogram for the selection of the frequency band. However, in some cases, such as the test rig states 8 and 9, the worsened defect was correctly indicated by the relatively high values on FTF.

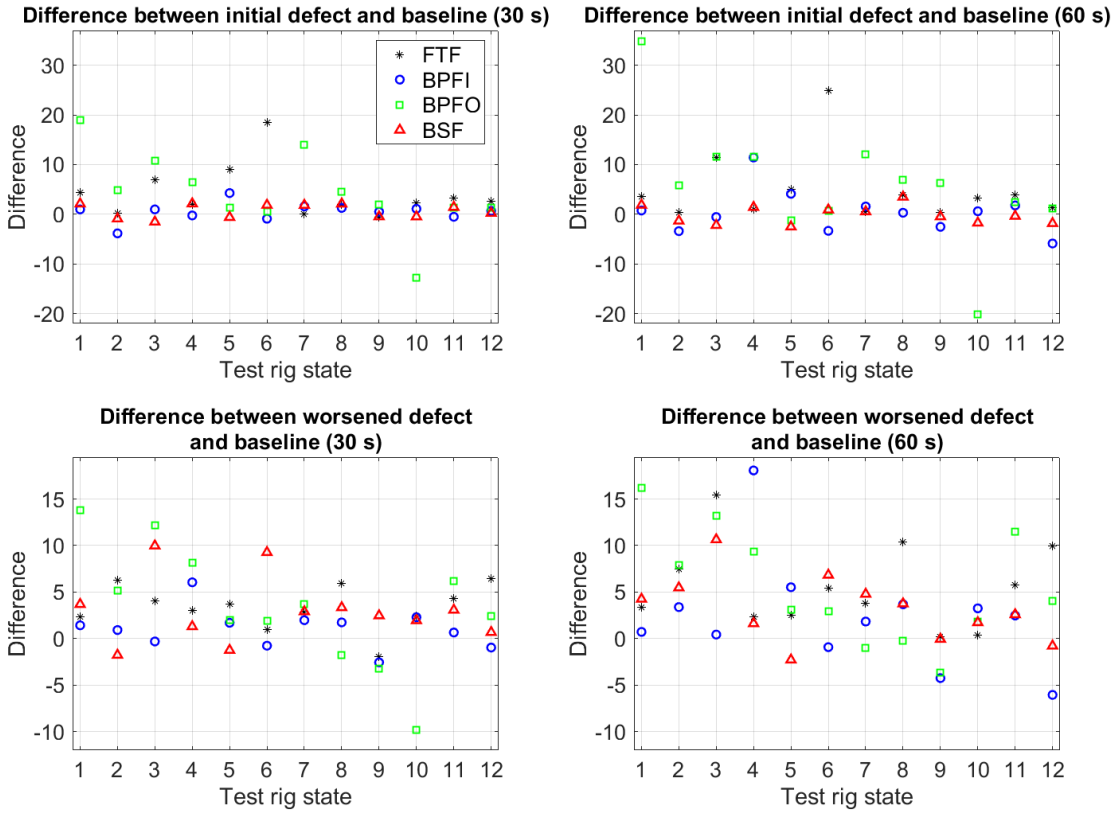

Fig. 19. Differences between the defect and baseline conditions based on peak-to-median indicator with two harmonics. The frequency bands for band-pass filtering were selected based on the peak-to-median indicator.

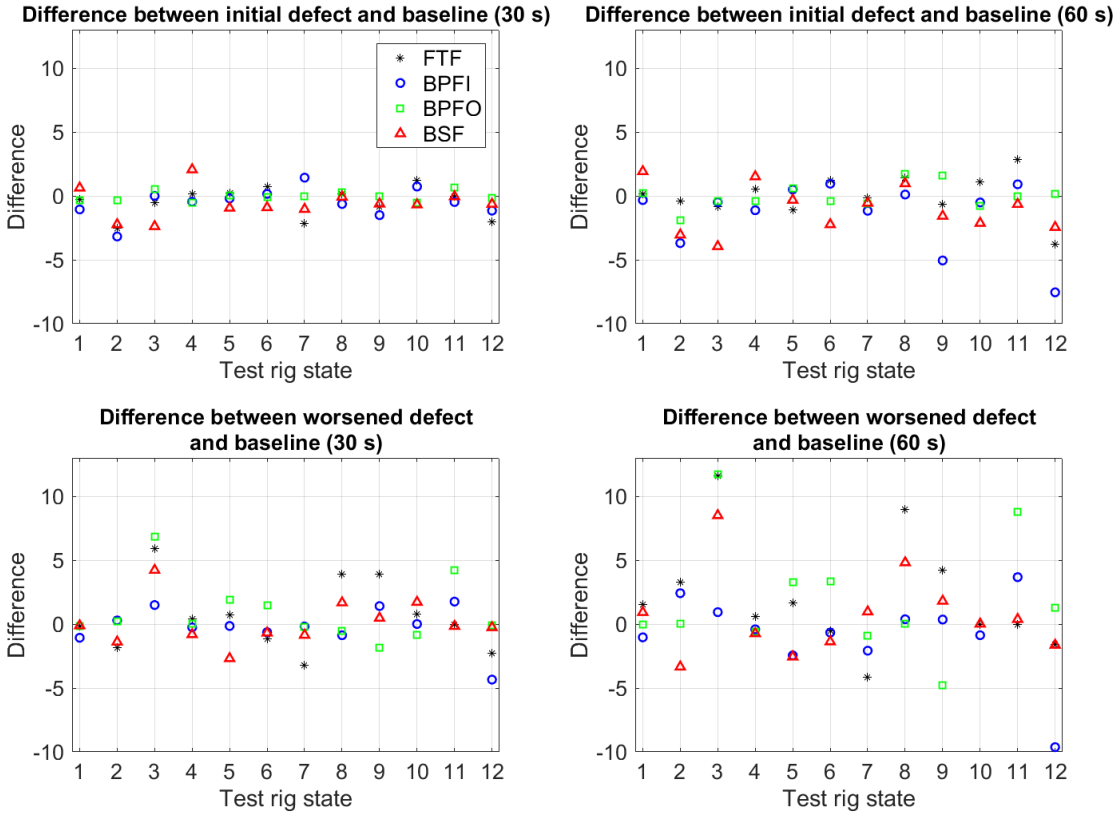

Fig. 20. Differences between the defect and baseline conditions based on peak-to-median indicator with two harmonics. The frequency bands for band-pass filtering were selected based on the fast kurtogram.

Finally, the computation time used by the proposed criterion is shown in Fig. 21. The total computation time for individual signals is shown on the left and the computation time for each defect frequency is shown on the right. The computations were done by using parallel computing on server B, introduced in Section 4.1.3. The presented results are the mean of 10 tests. Both segment sizes (1% and 2%) were included in the computations.

The results are shown from test rig states 1–3 with the worsened defect using 30-second signals. The other test rig states (4–12) with nearly the same defect frequencies had very similar computation times to the results presented. The computation times on the 60-second signals were approximately twofold. The results show that the computation time increased together with the shaft speed and the defect frequency. This is explained by the increasing number of segments in the fixed length signals.

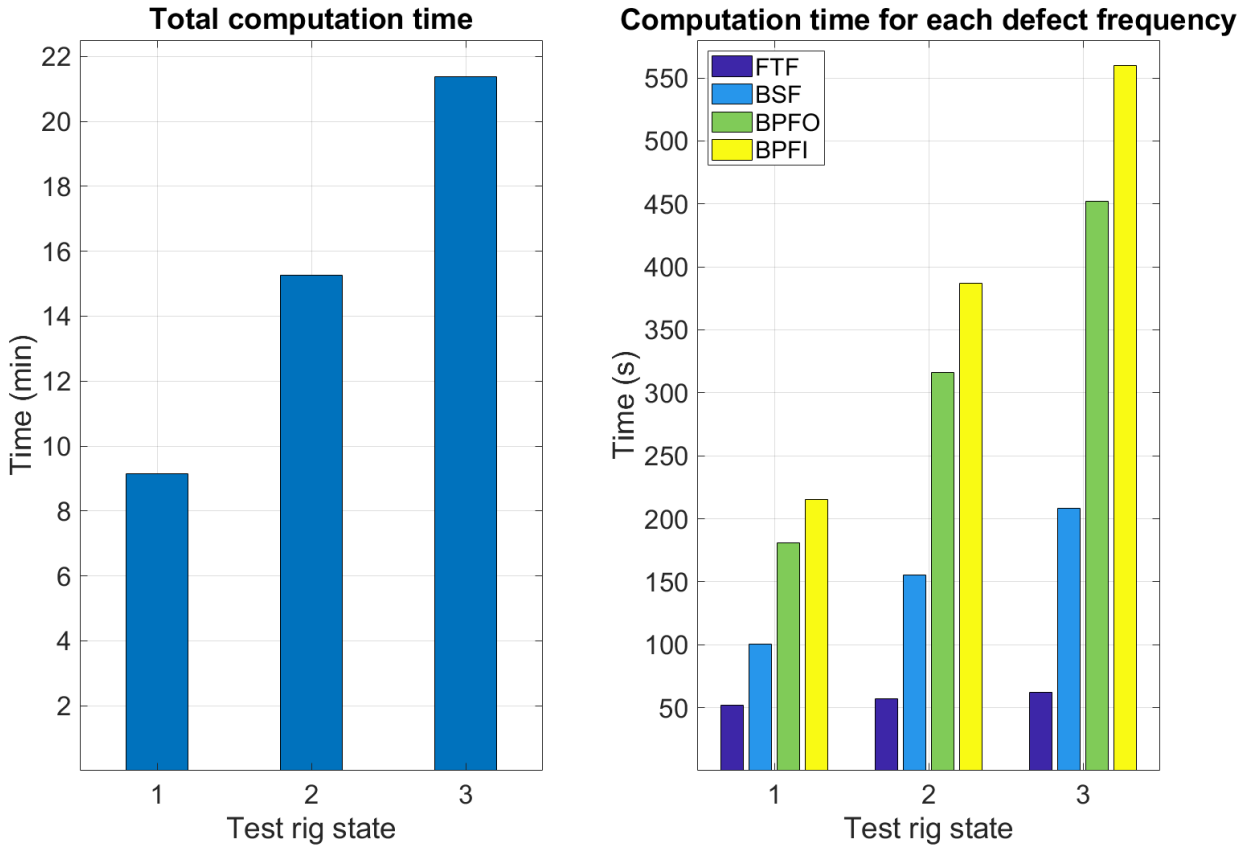

Fig. 21. Total computation time for individual signals (left) and computation time for each defect frequency (right) based on the proposed criterion using 30-second signals from the worsened defect case on test rig states 1–3.

4.2.2. Diagnosis of the fundamental train frequency based on the proposed criterion

Fig. 22 illustrates the values of the proposed criterion based on autocorrelation in each test rig state as bars. The values of initial and worsened defects are shown in the upper part and the corresponding values from the baseline condition are shown in the middle part in Fig. 22. The differences between the defect and baseline conditions are shown on the lowest part. At least a small change can be observed in all the test rig states by comparing the damaged and baseline conditions. In general, the defects had higher values than the corresponding baseline condition.

Test rig states 1–3 and 7–9, where the drive shafts were straight, show increasing damage the most explicitly. The worsened defect resulted in higher values than the initial defect in these test rig states. Otherwise, the signs of increasing damage are mainly indistinct. The relatively low values on the defect conditions in the test rig states with five degrees angle in the drive shafts give unclear evidence of the damage. A potential reason for the low values is the skidding of the rolling elements inside the bearing in these test rig states.

In some test rig states, such as the states 3 and 9, the baseline values were relatively high when the parameter settings of the initial defect were used. The criterion values of the initial defect were still slightly higher, but the indication of damage is unclear in those cases. In the test rig state 5 with 60-second signals, the criterion value was actually higher (=0.22) in the baseline condition than in the worsened defect condition (=0.13). In the case of the 30-second signal, the result was the opposite;

worsened defect had value 0.18 and the baseline had the value 0.05. This result clearly indicates that it is necessary to compare the defect and the baseline conditions with each other in order to make sound diagnostic conclusions.

In the test rig states 6 and 7, the baseline condition had relatively high negative values. A high negative value indicates that the time lags around the defect time lag had higher autocorrelation values than the defect time lag. This indicates that the parameters identified based on the monitored condition may distort the signal from the baseline condition in some cases.

Table 5 presents the statistical features that were selected by the proposed criterion in each test rig state. Fourteen different features from the set of twenty features described in Section 2.1.3 were selected. This result indicates that the characteristics of the damage in the vibration signals were varying in different conditions and a large set of candidate features became therefore justifiable. In most cases, the segment length that was 2% of the defect period gave clearer defect indications than the shorter alternative. The longer time window apparently had a better tolerance for the variation of the period between the shocks inflicted by the damaged bearing. Interestingly, a different feature was selected in many test rig states from the 30-second and 60-second signals. The initial and worsened defect resulted in the selection of different features in the most of the test rig states as well.

The selected features were mostly the ones, which are generally responsive to shocks. These include kurtosis, skewness, 95^{th} percentile, generalised norms with relatively high order, and the ratios of generalised norms. In addition, l_2 (root-mean-square) became selected in many test rig states (3, 7, 8, 9, 12) although it is generally not the most sensitive feature to the shocks. It is probable that the criterion based on autocorrelation yields relatively similar values on different statistical features in certain situations. Only the feature, which resulted in the highest value of the criterion, had significance in this study. The selected statistical feature was considered only as a part of the optimal settings to be used in the automated approach.

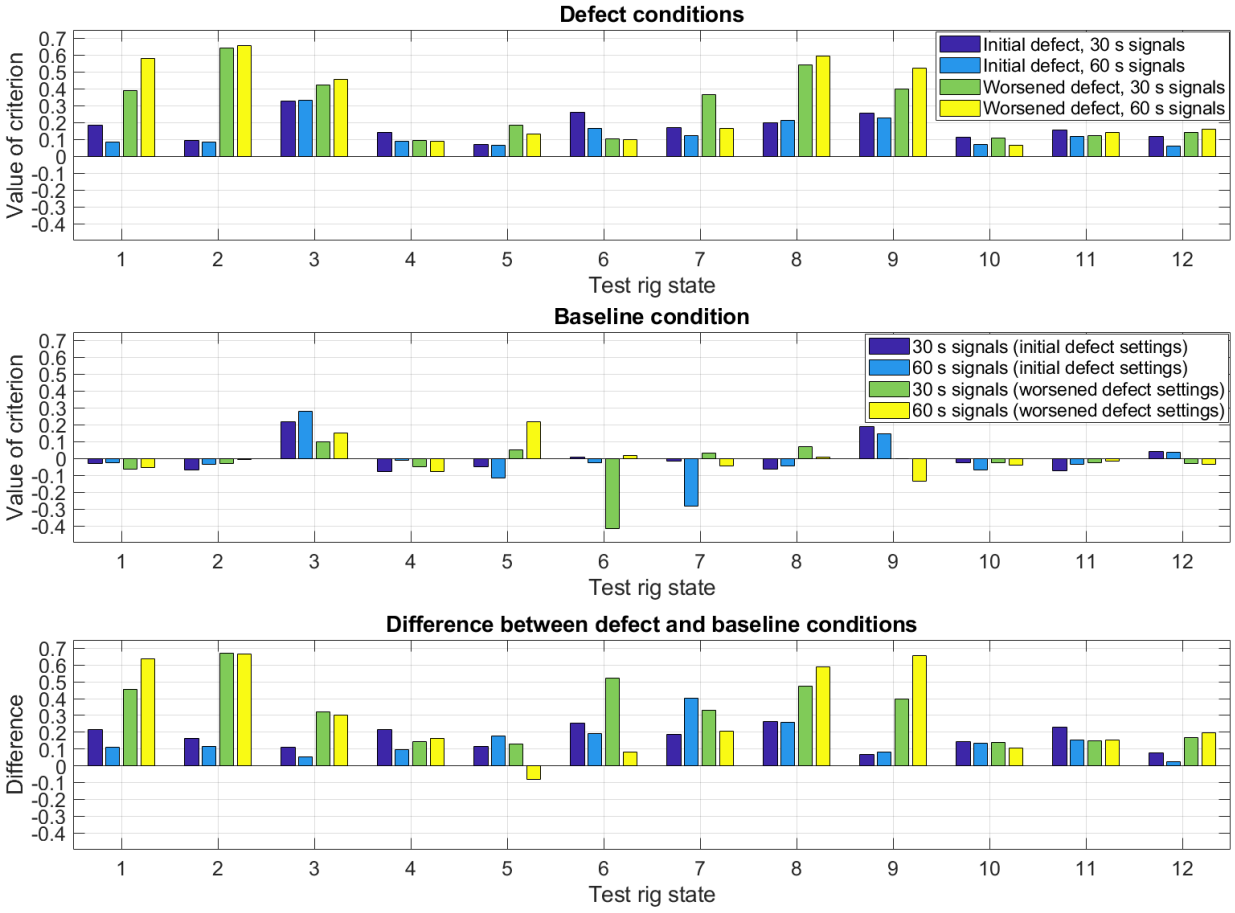

Fig. 22. The values of the criterion based on autocorrelation in FTF diagnosis. Baseline values were computed using the settings defined based on the defect conditions.

Table 5 The statistical features and segment lengths (percentage from the defect period) for each test rig state based on the proposed autocorrelation-based criterion.

	Initial defect				**Worsened defect**			
	30-second signal		*60-second signal*		*30-second signal*		*60-second signal*	
Test rig state	feature	length	feature	length	feature	length	feature	length
1	*skewness*	1	*skewness*	2	*kurtosis*	2	*kurtosis*	2
2	I_4	2	I_4	2	*kurtosis*	2	*kurtosis*	2
3	I_2	2	I_2	2	I_4	2	I_2	2
4	*skewness*	2	skewness	1	I_{20} / I_{10}	2	I_{10} / I_2	2
5	*peak*	2	$P_{95}(\lvert x^{(2)}\rvert)$	2	I_{10} / I_1	2	*kurtosis*	2
6	*kurtosis*	2	*kurtosis*	2	I_{20}	2	$P_{95}(\lvert x^{(2)}\rvert)$	1
7	*skewness*	2	I_2	1	*kurtosis*	2	I_{20} / $I_{0.5}$	2
8	I_2	2	I_2	2	*kurtosis*	2	*kurtosis*	2
9	I_2	2	I_4	2	*kurtosis*	2	I_{10}	2
10	I_4 / $I_{0.5}$	2	I_4 / $I_{0.5}$	2	*skewness*	1	*crest factor*	2
11	*kurtosis*	2	I_{10}	2	I_4	2	I_{10}	2
12	I_4	2	I_2	2	I_4	1	I_4	1

4.2.3. Diagnosis of the fundamental train frequency based on the peak ratio and peak-to-median indicators

The maximum differences between the defect and baseline conditions based on the peak ratio and peak-to-median indicators are shown in Fig. 23. In this case, the same indicators were used as the method to select the frequency band. Fig. 23 reveals that higher values were obtained in the defect conditions than in the baseline condition in each test rig state. The difference was generally the smallest in the lowest shaft speed; in the test rig states 1, 4, 7, and 10. In the test rig states 5, 6, and 10, the difference was larger with the initial defect, but the worsened defect resulted in a larger difference from the baseline condition more commonly.

Fig. 24 shows the corresponding differences by using the fast kurtogram as the approach to frequency band selection. In many cases, the difference values between the defect and baseline conditions were low (<2) or even negative which suggests that the bearing condition had not degraded. The clearest defect indications with both signal lengths were given in the test rig states 3, 8, and 9 in the case of the worsened defect. The depicted differences in Fig. 24 are lower than in Fig. 23, which indicates that the frequency bands selected by the fast kurtogram were generally not as appropriate as the ones selected based on the peak ratio and peak-to-median indicators.

The peak ratio and peak-to-median indicators, which yielded the results shown in Fig. 23 are presented in Table 6 together with their values in the defect conditions in each test rig state. Table 6 indicates that the values of the indicators increased together with the shaft speed in the defect conditions with some exceptions. Moreover, the results indicate that different indicators with a varying number of harmonics became selected in different conditions. The corresponding indicators and their values that were computed by using the frequency bands selected by the fast kurtogram are given in Table 7. The comparison between the results in Tables 6 and 7 reveals that the indicator values were mostly higher in Table 6 in the corresponding test rig states and with the same signal length.

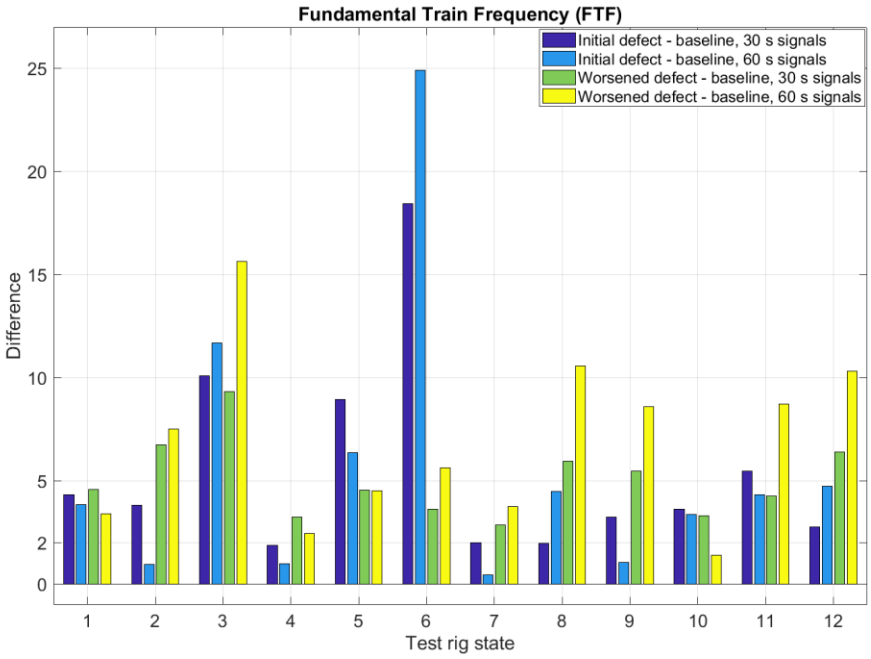

Fig. 23. Maximum differences between the defect and baseline conditions based on the peak ratio and peak-to-median indicators.

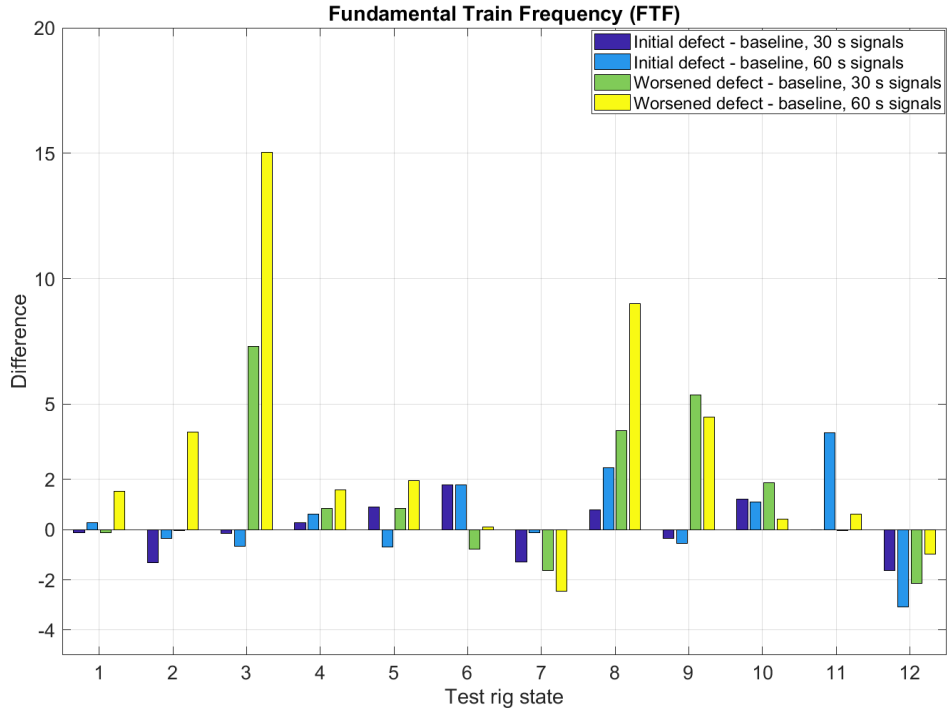

Fig. 24. Maximum differences between defect and baseline conditions based on the peak ratio and peak-to-median indicators. The frequency bands for band-pass filtering were selected based on the fast kurtogram.

Table 6 Maximum peak ratio and peak-to-median values in defect conditions. The term 'PTM' stands for peak-to-median, 'PR' stands for peak ratio, and the following number stands for the number of FTF harmonics.

	Initial defect				**Worsened defect**			
	30-second signal		*60-second signal*		*30-second signal*		*60-second signal*	
Test rig state	indicator	value	indicator	value	indicator	value	indicator	value
1	PTM 2	6.50	PTM 1	6.86	PTM 1	4.90	PR 2	6.62
2	PR 2	6.01	PTM 1	10.21	PTM 1	8.31	PR 2	10.98
3	PTM 1	18.78	PTM 1	22.70	PTM 1	11.01	PR 2	20.70
4	PTM 2	3.18	PTM 2	2.30	PTM 1	4.34	PR 2	4.23
5	PTM 2	10.22	PTM 1	6.44	PTM 1	5.29	PR 2	11.00
6	PTM 2	21.16	PTM 2	28.27	PR 2	7.30	PTM 1	11.11
7	PR 2	3.64	PTM 2	5.78	PTM 2	4.69	PTM 2	4.95
8	PTM 2	7.15	PTM 1	7.92	PTM 2	7.54	PTM 1	12.85
9	PTM 1	12.09	PR 1	8.85	PR 2	10.12	PR 2	16.08
10	PTM 1	4.62	PTM 1	4.97	PTM 1	4.06	PR 2	5.15
11	PTM 1	8.10	PTM 1	7.37	PTM 2	5.37	PR 2	10.71
12	PTM 1	6.80	PR 2	7.54	PTM 2	9.70	PTM 1	13.51

Table 7 Maximum peak ratio and peak-to-median values in defect conditions. The frequency bands were selected based on the fast kurtogram.

	Initial defect				**Worsened Defect**			
	30-second signal		*60-second signal*		*30-second signal*		*60-second signal*	
Test rig state	indicator	value	indicator	value	indicator	value	indicator	value
1	PTM 1	0.67	PTM 1	1.42	PTM 2	1.43	PTM 2	2.64
2	PR 1	1.73	PR 1	0.58	PR 2	3.45	PR 2	5.41
3	PR 1	1.62	PR 1	2.07	PR 2	10.46	PR 2	19.38
4	PR 2	1.36	PR 2	1.47	PR 2	1.93	PR 2	2.44
5	PR 2	2.63	PR 1	1.84	PR 2	3.37	PR 2	3.47
6	PTM 1	2.30	PTM 1	3.79	PR 1	0.39	PR 2	2.31
7	PR 2	2.49	PTM 2	5.55	PR 2	2.16	PR 1	1.24
8	PR 2	2.70	PR 2	4.49	PTM 2	7.66	PTM 2	13.56
9	PR 2	1.62	PR 1	1.77	PR 2	8.91	PTM 1	8.28
10	PTM 2	1.80	PTM 2	2.39	PTM 1	2.83	PR 2	2.28
11	PTM 2	2.42	PR 2	6.06	PTM 2	1.42	PR 2	3.84
12	PR 1	1.49	PR 1	1.11	PR 1	0.81	PR 2	3.85

4.2.4. Analysis of the selected frequency bands

This section discusses the frequency band selections made by each of the tested methods in the FTF diagnosis. The selections for the initial defect are presented on the left and the selections for the worsened defect are shown on the right in Fig. 25. The optimal frequency band varied significantly based on the method used, the length of a signal, the state of the test rig, and the severity of the damage. The fast kurtogram was an exception, because the most commonly selected frequency band was 0–2133 Hz. However, this band was a poor alternative considering the defect detection, as shown in Fig. 24. A different frequency band was selected in the cases, from which an indication about the defect can be concluded.

The same frequency band was selected based on both signal lengths by the fast kurtogram apart from the test rig state 11 with the initial defect and the test rig state 10 with the worsened defect. On the contrary, the approach based on the peak ratio and peak-to-median indicators resulted in the same band for both signal lengths only in the test rig states 3 and 11 with the initial defect and in the test rig states 1, 3, 9, and 10 with the worsened defect. The criterion based on autocorrelation yielded the same band for different signal lengths in the test rig states 3, 6, 8, 9, 11, and 12 with the initial defect and in the test rig states 2, 3, 8, 11, and 12 with the worsened defect. These results indicate that the signal length had the biggest influence on the peak ratio and peak-to-median based frequency band selection.

The width of the selected frequency bands varied considerably. In the case of initial defect, the narrowest band (225 Hz) became selected on three test rig states by the criterion based on autocorrelation. The peak ratio and peak-to-median based approach had the 225 Hz band in three test rig states with the initial defect and in four test rig states with the worsened defect. The narrowest band selected based on the fast kurtogram was 400 Hz.

The full bandwidth was selected on some occasions as well. The criterion based on autocorrelation selected the full bandwidth in the test rig states 8 and 9 with the worsened defect. The approach based on the peak-to-median and peak ratio indicators selected the full bandwidth in the test rig state 9 with the worsened defect. On the other hand, the use of the fast kurtogram interestingly resulted in 533 Hz and 400 Hz bands in the corresponding cases. In the test rig state 3 with the

worsened defect, almost the same frequency bands became selected based on the fast kurtogram (400–800 Hz) and the peak ratio and peak-to-median (450–900 Hz) based approach. The selection by the proposed criterion resulted in a higher and wider frequency band (1800–3600 Hz). Obviously, the characteristics of the worsened defect were present in a wide frequency range in the test rig states 3, 8, and 9, which can be concluded based on the results in Figs. 22, 23, and 24. In those cases, the defect was diagnosable with multiple parameter settings and all the methods. Fig. 25 clearly indicates that the varying operating conditions had a notable influence on the automated selection of the optimal frequency band.

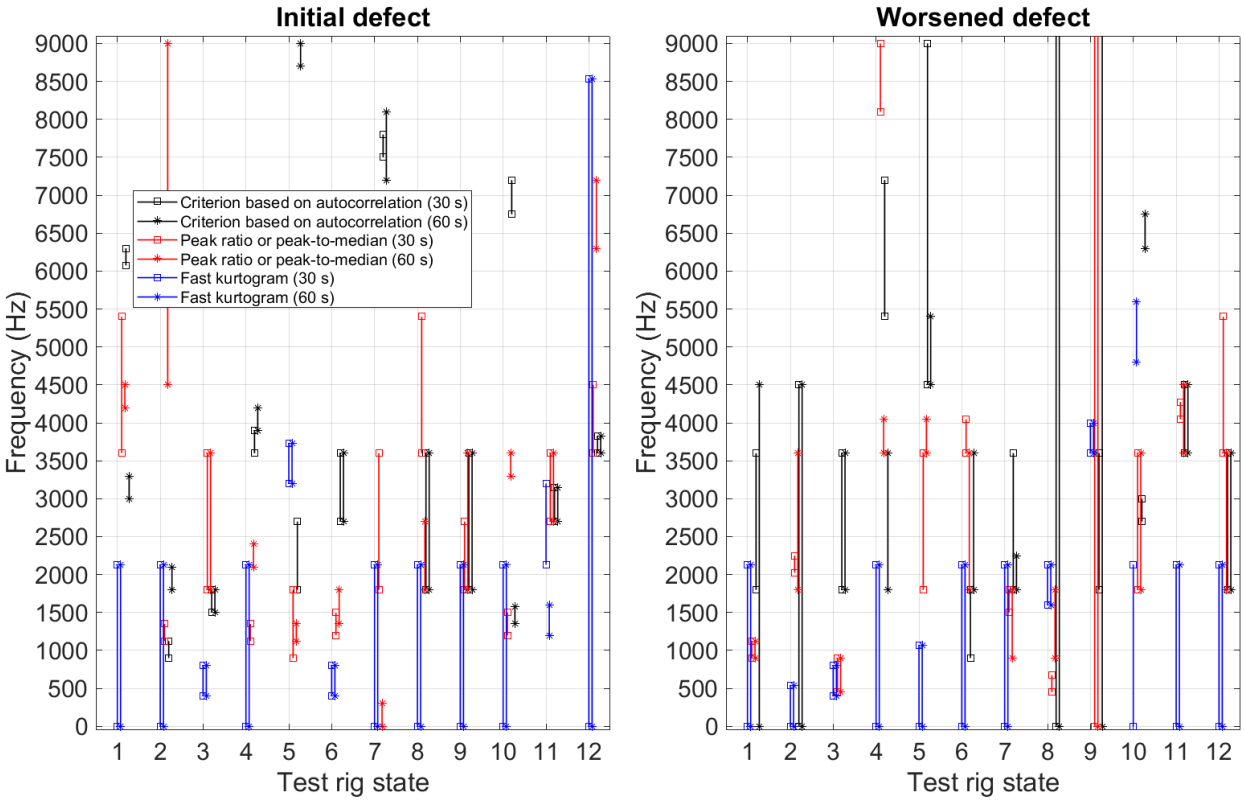

Fig. 25. Selected frequency bands. The lines continuing above 9 kHz indicate full bandwidth. The lines portraying the selected frequency bands by different methods are non-overlapping for the clarification.

Finally, Appendix A shows the autocorrelation values and the envelope spectra, of which frequency bands for band-pass filtering were selected based on the peak ratio or peak-to-median indicators and the fast kurtogram. These plots can be used to confirm the diagnostic conclusions in the case of 30-second signals.

4.3. Discussion

Based on the simulations and the laboratory tests, the proposed criterion based on autocorrelation was able to indicate defects also on the lowest shaft speeds, whereas the envelope analysis gave weak or obscured indications. The FTF diagnosis on test rig states 1 and 7 with the worsened defect

and the simulation study with the lowest rotational speeds gave the strongest proof of that. However, the performance of the envelope analysis improved in the simulation study, when the signal was lengthened substantially. On the higher rotational speeds, the envelope analysis gave indications about the defects in many cases, if the frequency band was selected based on the peak ratio and peak-to-median indicators. Frequency band selection based on the fast kurtogram was mostly unsuccessful apart from the simulation study with the highest rotational speeds and the FTF diagnosis on test rig states 3, 8, and 9 with the worsened defect.

The results reveal that the frequency band selection based on the defect frequencies gave superior results compared with the selection by the fast kurtogram. The reason behind the weaker performance of the latter approach could be the low magnitude of the shocks inflicted by the defects in the speed regime under consideration. In addition, shocks unrelated to the defect may distractingly increase the values of kurtosis in certain frequency bands.

In some cases, such as the FTF diagnosis on test rig states 3 and 5 with the initial defect, the envelope spectra, which had frequency bands selected based on the peak ratio and peak-to-median based approach, gave the clearest indication about the defects. However, a drawback of this method was given in Section 4.2.1, which revealed that the method yields high values for the defect frequencies unrelated to the actual damage, such as BPFI and BPFO. The latter is a harmonic of FTF, which can be concluded based on (1) and (3), but a high value for BPFI indicates a different type of defect. Such performance restricts the applicability of this approach on automated fault diagnosis. A supplementary study, which is not presented here, indicated that the use of squared envelope spectrum does not remove this problem on the studied data set either.

The peak ratio and peak-to-median indicators are ratios and therefore sensitive to the low values of the frequency amplitudes around the peaks. If the mean or median of amplitudes is close to zero, then even low peaks may result in high ratios. Moreover, such indicators are sensitive to frequency components with high amplitudes inside the search range used for the peaks. This could be a problem especially on the higher frequencies with relatively wide search range, such as the search range presented in Section 2.2.

The signal length is an essential factor for successful diagnosis in the slow rotational speed. It should be adjusted based on the shaft speed. However, the signal length optimisation is challenging in practice because the rolling elements of the bearing may skid, and then the shocks disappear completely or have a variable period. A signal length that could theoretically contain e.g. 2–3 shocks based on the presumed bearing defect frequency would be too short. In this study, the presumed period between the defect related shocks was around 5.9 seconds in the slowest case, which was test rig state 7 with the worsened defect. This means that six of such shocks at the most could be expected in a 30-second signal. The proposed criterion was able to give a clear indication of the defect contrary to the other tested methods. Moreover, the results presented in Figs. 15 and 22 suggest that lengthening the signal (i.e. increasing the number of defect periods) does not give explicit advantage to the diagnosis. This indicates that the proposed criterion could be used successfully with relatively short signals, whereas the envelope analysis requires a higher number of defect-related impulses to show the defects clearly, as illustrated in Figs. 12 and 14.

The automated data-driven approach to the frequency band selection excludes the possibility of poor selection by the vibration analyst. However, further research is recommended on the optimisation of the frequency band selection. The proposed method is optimal on the grounds that all the defined frequency bands become tested and the optimal one becomes selected based on the considered criterion. This is not the optimal solution considering the speed of computation, as indicated by the results in Sections 4.1.3 and 4.2.1. Therefore, some rules of thumb for the appropriate frequency bands and their sizes should be defined to reduce the search space. However, parallel computing accelerates the computation significantly, which was demonstrated in Section 4.1.3.

Moreover, the automated selection is vulnerable to disturbances and machine vibrations that have the same frequency as the bearing defect. In such a case, high criterion values may be generated by factors unrelated to the damage. The difference between the currently monitored and baseline conditions may reveal change in the bearing condition, if the disturbance unrelated to bearing defect stays constant in both conditions. However, in certain applications some additional method for filtering out sinusoids from other sources may be required.

The selection of the optimal frequency bands and computational features by the techniques under discussion requires two stages. At the first stage, the optimal settings are defined and the values of condition indicators are calculated in the monitored system state. At the second stage, the defined settings are applied on the baseline signal from a corresponding undamaged system state. The second stage is essential for diagnosis, because without comparison to normal situation the discussed condition indicators do not contain sufficient diagnostic information, as indicated by the results in Section 4.2. The selections in the laboratory tests, such as the frequency bands shown in Fig. 25 and the features in Table 5, reveal that different conditions in the monitored system resulted also in different parameter selections. This indicates that raw baseline data should be saved from different operating conditions for an automated diagnostic application. Exactly the same parameters can then be used for the comparison of the currently monitored and baseline conditions. Pre-determined signal processing settings would potentially yield sub-optimal results in varying conditions. This should be considered in the development of autonomous and intelligent sensors, which process acquired signals before sending the processed information forward.

Considering an automated application, it is important to define suitable thresholds, which indicate defects or potential defects. In the case of damage, the criterion based on autocorrelation should have a positive value, which is higher than the corresponding value in the baseline state. Based on Figs. 10 and 22, the criterion value around 0.1 and the difference from the baseline around 0.1 could together indicate a small change in the condition. With such thresholds, the defect at the lowest shaft speed case would have remained undetected in the simulation study (Fig. 10). On the other hand, too low thresholds result in repeated alarms, which is a problem in many automated solutions. In the cases, where the reference value from the baseline condition is negative, it may be advantageous to use zero as the reference value. The thresholds obviously require case-specific consideration. If abundant data are available from the monitored system, the methods of Statistical Process Control (SPC) [52] could provide appropriate tools for the threshold definition.

Considering the peak ratio and peak-to-median indicators, the threshold adjustment is even more difficult, because these indicators are ratios. Even a difference around value 4 between the damaged and baseline conditions was not large enough to give an unambiguous indication of damage in some cases. This can be inferred based on the results from test rig state 1 with the worsened defect, as shown in Fig. 23 and Appendix A (Fig. A4), for example. However, the peak ratio and peak-to-median indicators could be used together with the envelope spectrum to support the analysis in a semi-automated approach. One solution for an automated approach would be to combine diagnosis methods. If the criterion based on autocorrelation indicates a change in the bearing condition, other methods such as the envelope spectrum could be used to verify the observation.

Finally, the roller element defect was shown only by the frequency component associated with FTF in the laboratory tests. This means that a certain differentiation between the roller element and cage defects was not achievable based on the acquired data and the methods tested. Because only the roller element defects were diagnosed in the laboratory tests, further research on the diagnosis of different bearing defects and multiple simultaneous defects by using the proposed method would be advantageous.

5. Conclusions

A criterion based on autocorrelation was proposed to the localised defect diagnosis in slow rotating rolling element bearings. The steps in the criterion calculation included filtering, signal segmentation, statistical feature selection and autocorrelation computation. In order to enhance the defect manifestation, the autocorrelation of the time lag corresponding to the defect was compared with other time lags. The signal processing settings were selected so that the defect time lag became the most discernible based on the criterion. The approach was compared with automated envelope spectrum diagnosis approaches based on a simulation study and experiments on a drive shaft test rig.

The results revealed that the proposed criterion yielded correct defect indications at the lowest shaft speeds, while the envelope spectrum approaches gave obscured results or no defect indications at all. On the experimental data, the proposed criterion indicated increased values on the fundamental train frequency, which could be associated with the rolling element damage. However, the lack of defect signs on the roller spin frequency indicated that a certain differentiation between the roller defect and cage defect was not achievable in this case. The automated envelope spectrum diagnosis provided defect indications unrelated to the actual damage, which increases the risk of misdiagnosing. Nevertheless, the worsened defect in the rolling element was found by each approach based on the fundamental train frequency in some cases. In addition, the results indicate that the proposed criterion could be used successfully on relatively short signals, which include only a few defect periods.

The results also suggest that the signal processing settings should be identified based on the prevailing condition, because pre-determined signal processing settings produce non-optimal results in systems with varying conditions. In addition, the results suggest that the automated fault diagnosis benefits from the acquisition of unprocessed baseline signals in different conditions as the reference information. Further research is recommended on the use of the proposed criterion in the real-world cases with different types of defects and multiple simultaneous defects. Moreover, the combination of different methods should be studied to advance the development of robust automated diagnostic techniques.

Acknowledgements

The laboratory experiments were done during the SIMP (System Integrated Metals Processing) programme coordinated by DIMECC Oy. R.-P. Nikula and K. Karioja would like to thank the staff in the Institute for Advanced Mining Technologies (AMT) at RWTH Aachen University for the collaboration during the laboratory experiments. The authors would like to thank Jérôme Antoni for sharing his Matlab® scripts for the fast kurtogram.

Appendix A. Autocorrelations and envelope spectra in laboratory test runs

Appendix A illustrates the autocorrelation values and the envelope spectra calculated from the 30-second signals in the laboratory experiments focusing on the fundamental train frequency diagnosis. Fig. A1 shows the autocorrelation values of the initial defect and the baseline conditions. Similarly, Fig. A2 shows the autocorrelation values of the worsened defect and the baseline conditions. The configurations used in the computations include the statistical features presented in Table 5 and the frequency bands given in Fig. 25. Figs. A3 and A4 illustrate the corresponding envelope spectra from the cases, where the frequency bands were selected based on the peak ratio and peak-to-median indicators. Finally, Figs. A5 and A6 show the envelope spectra from the cases, which had the frequency bands selected by the fast kurtogram.

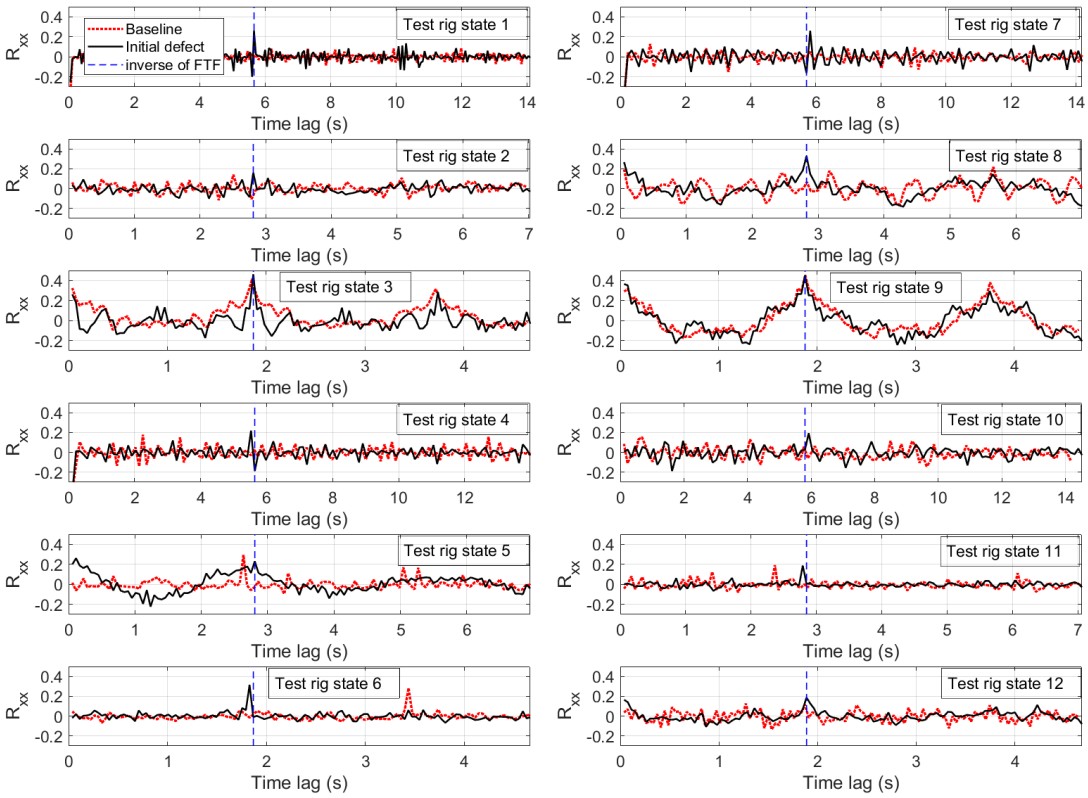

Fig. A1. Autocorrelation values from the initial defect and baseline conditions based on 30-second signals.

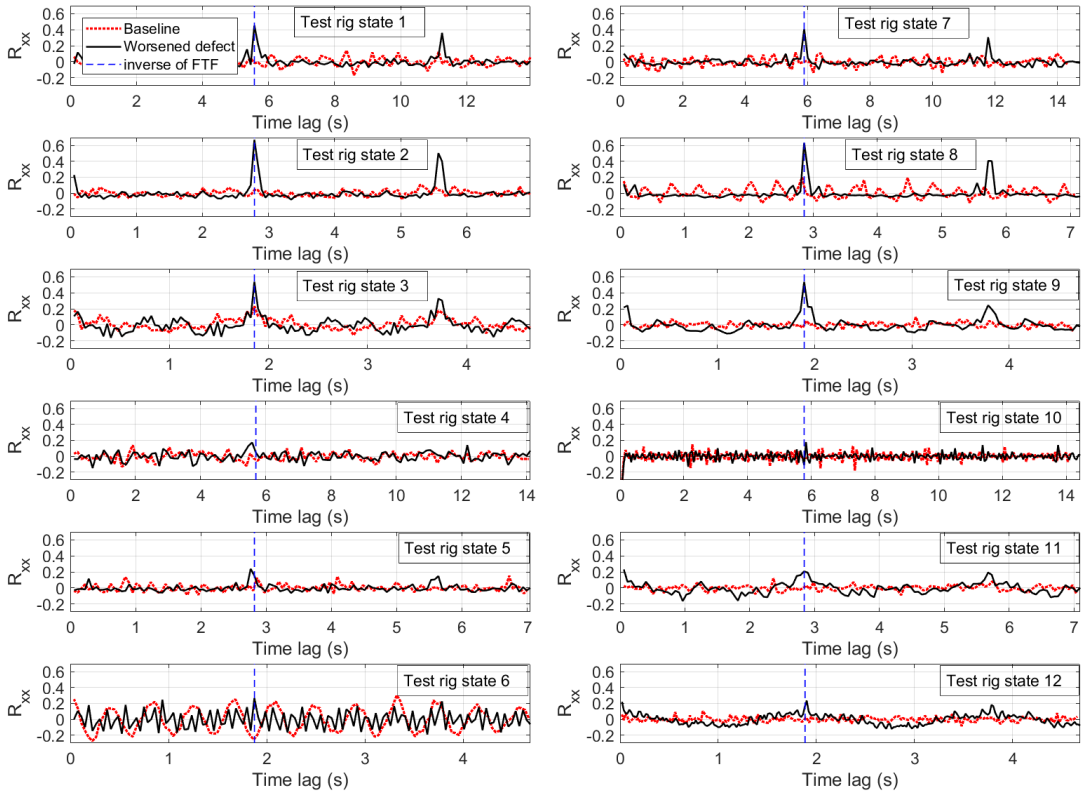

Fig. A2. Autocorrelation values from the worsened defect and baseline conditions based on 30-second signals.

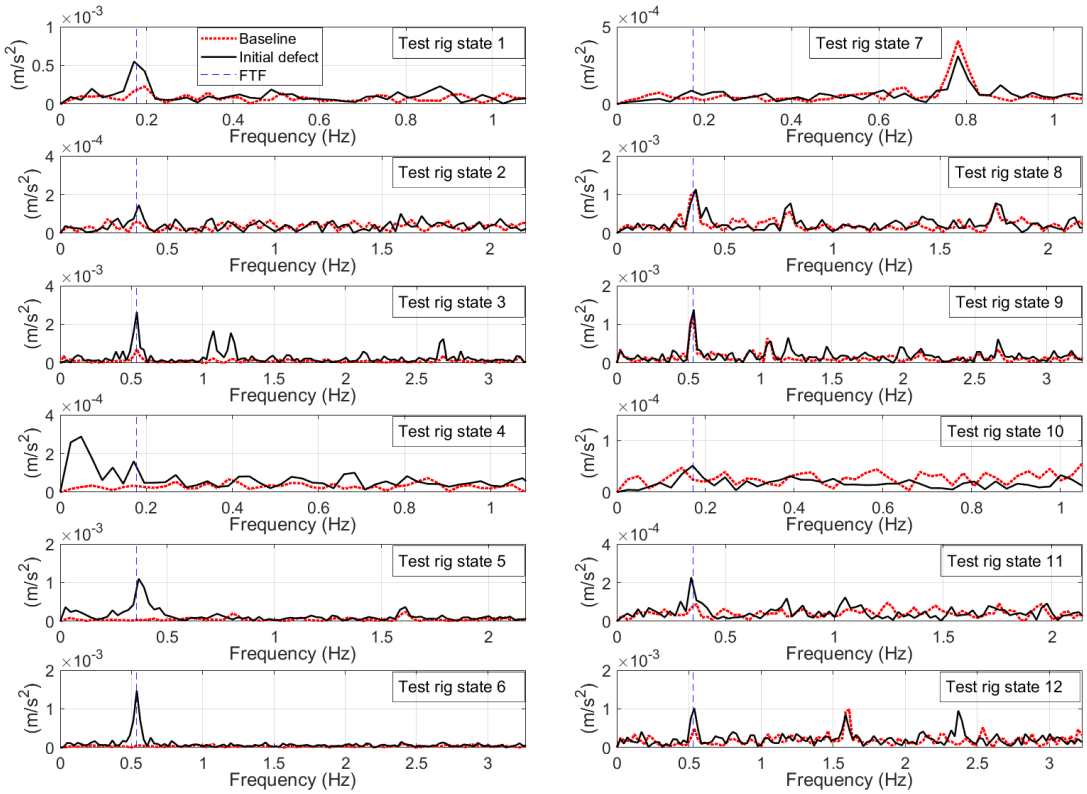

Fig. A3. Envelope spectra from the initial defect and baseline conditions based on 30-second signals. The frequency bands were selected based on the peak ratio and peak-to-median indicators.

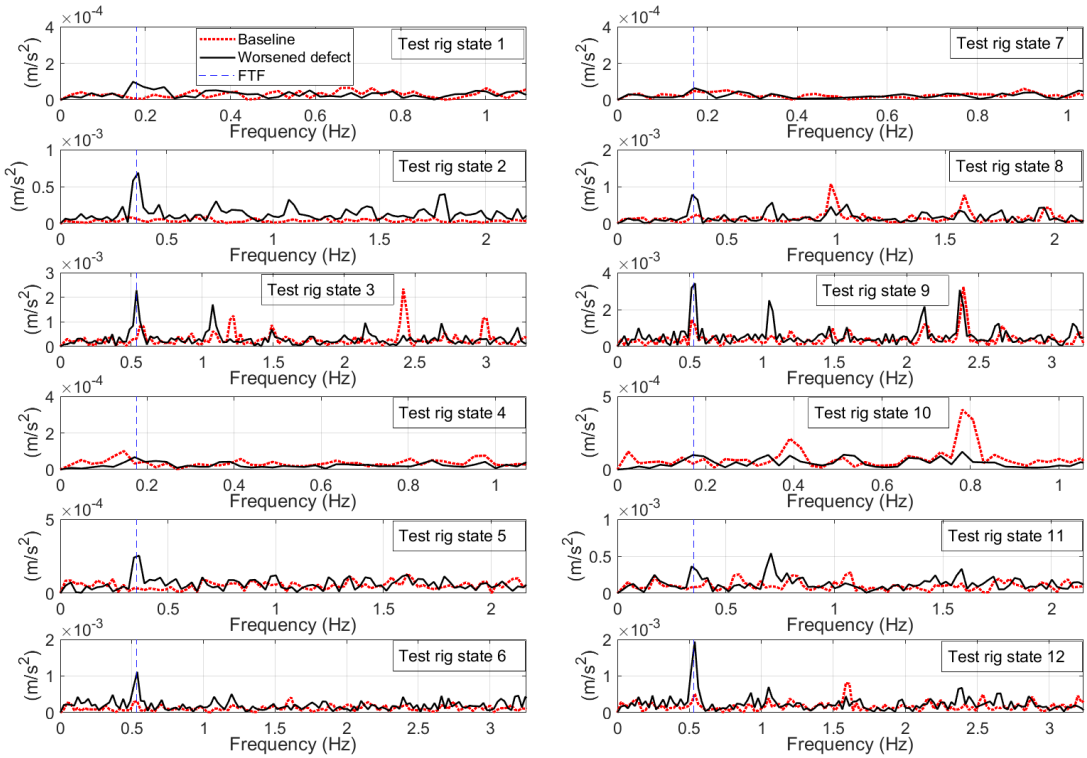

Fig. A4. Envelope spectra from the worsened defect and baseline conditions based on 30-second signals. The frequency bands were selected based on the peak ratio and peak-to-median indicators.

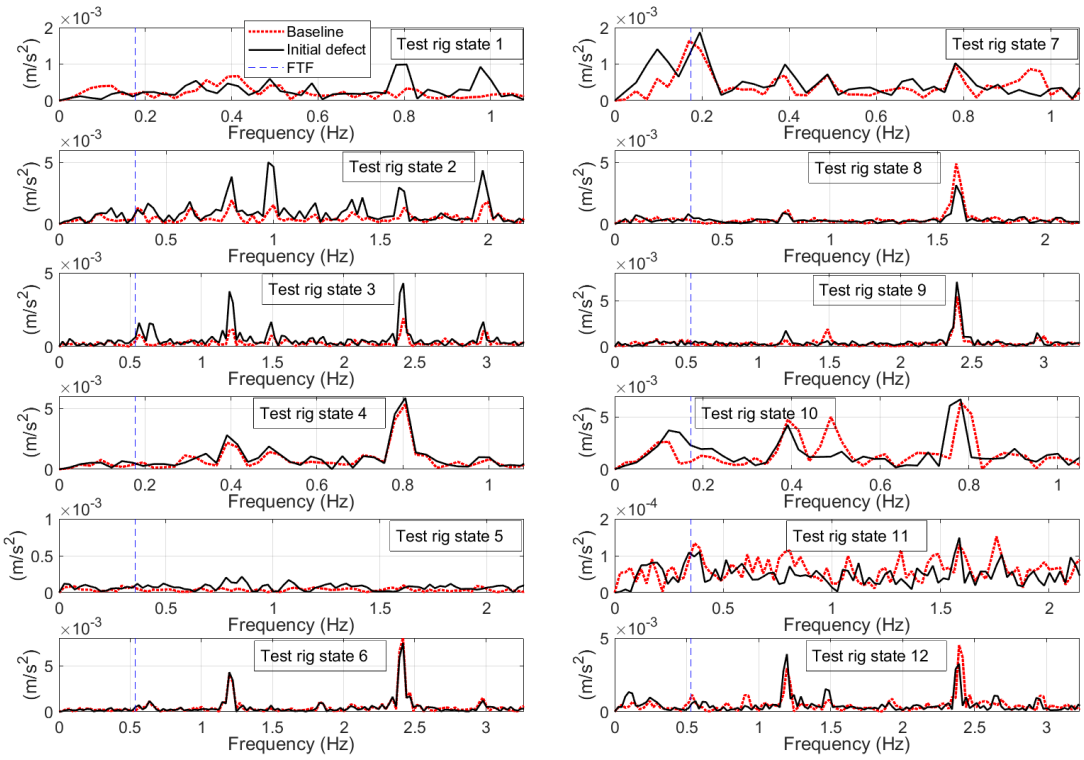

Fig. A5. Envelope spectra from the initial defect and baseline conditions based on 30-second signals. The frequency bands were selected based on the fast kurtogram.

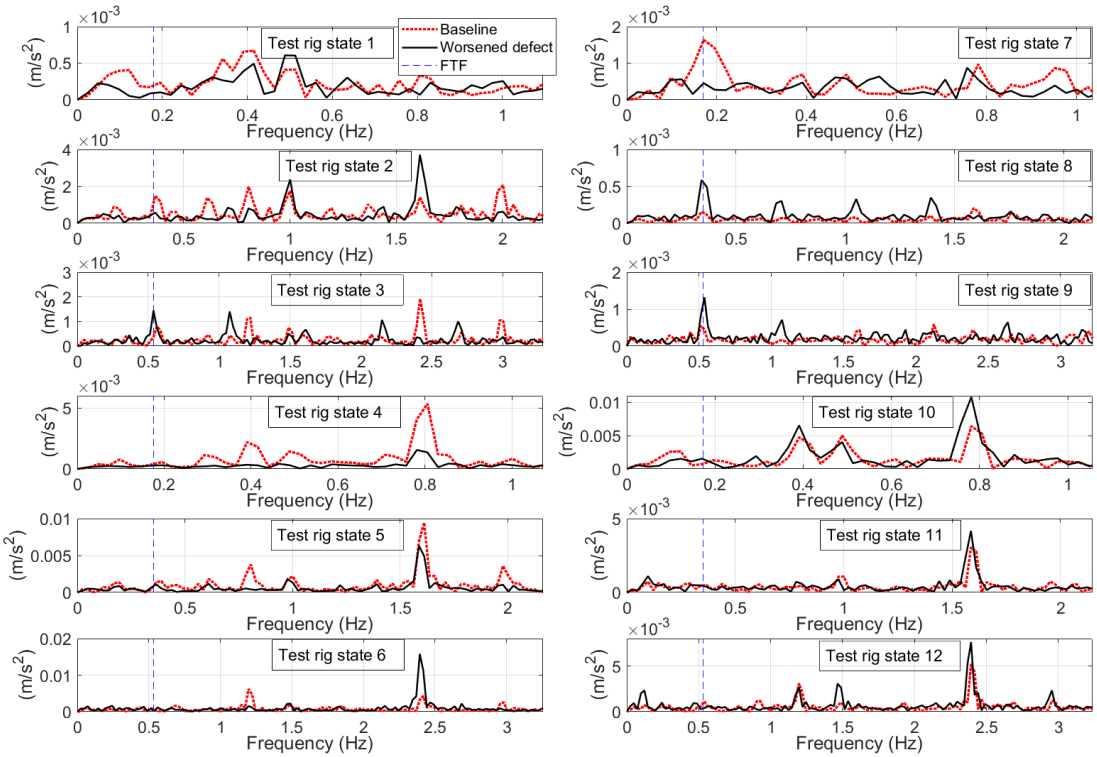

Fig. A6. Envelope spectra from the worsened defect and baseline conditions based on 30-second signals. The frequency bands were selected based on the fast kurtogram.

References

[1] A. Rai, S.H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribol. Int. 96 (2016) 289–306. https://doi.org/10.1016/j.triboint.2015.12.037.

[2] R.B. Randall, J. Antoni, Rolling element bearing diagnostics – A tutorial, Mech. Syst. Signal Process. 25(2) (2011) 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017.

[3] D. Brie, Modelling of the spalled rolling element bearing vibration signal: An overview and some new results, Mech. Syst. Signal Process. 14(3) (2000) 353–369. https://doi.org/10.1006/mssp.1999.1237.

[4] C.M. Harris, A.G. Piersol, Harris' Shock and Vibration Handbook, fifth ed., McGraw-Hill, New York, 2002.

[5] C. Mishra, A.K. Samantaray, G. Chakraborty, Rolling element bearing fault diagnosis under slow speed operation using wavelet de-noising, Measurement 103 (2017) 77–86. https://doi.org/10.1016/j.measurement.2017.02.033.

[6] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis, Mech. Syst. Signal Process. 15(5) (2001) 963–977. https://doi.org/10.1006/mssp.2001.1410.

[7] R.G. Canada, J.C. Robinson, Vibration measurements on slow speed machinery, in: Predictive Maintenance Technology National Conference (P/PM Technology) 8(6), Indianapolis, Indiana, 1995, pp. 33–37.

[8] B. Van Hecke, J. Yoon, D. He, Low speed bearing fault diagnosis using acoustic emission sensors, Appl. Acoust. 105 (2016) 35–44. https://doi.org/10.1016/j.apacoust.2015.10.028.

[9] C.F. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearing. Part I: the use of parametric spectra, Mech. Syst. Signal Process. 6(4) (1992) 297–307. https://doi.org/10.1016/0888-3270(92)90032-E.

[10] J.C. Robinson, R.G. Canada, R.G. Piety, Vibration monitoring on slow speed machinery: new methodologies covering machinery from 0.5 to 600 rpm, in: Proceedings of the Fifth International Conference on Profitable Condition Monitoring – Fluids and Machinery Performance Monitoring, bHr Group Ltd., Publication 22, Harrogate, UK, 1996, pp. 169–182.

[11] Y.-F. Wang, P.J. Kootsookos, Modeling of low shaft speed bearing faults for condition monitoring, Mech. Syst. Signal Process. 12(3) (1998) 415–426. https://doi.org/10.1006/mssp.1997.0149.

[12] K. Karioja, R.-P. Nikula, C. Niedringhaus, C. Küch, R. Baltes, Roller bearing fault detection using real order derivatives and cyclostationarity, in: 13th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Paris, 2016, pp. 152–162.

[13] T. R. Lin, E. Kim, A.C.C. Tan, A practical signal processing approach for condition monitoring of low speed machinery using Peak-Hold-Down-Sample algorithm, Mech. Syst. Signal Process. 36(2) (2013), 256–270. https://doi.org/10.1016/j.ymssp.2012.11.003.

[14] M. Elforjani, D. Mba, Accelerated natural fault diagnosis in slow speed bearings with Acoustic Emission, Eng. Fract. Mech. 77(1) (2010) 112–127. https://doi.org/10.1016/j.engfracmech.2009.09.016.

[15] A. Widodo, E.Y. Kim, J.-D. Son, B.-S. Yang, A.C.C. Tan, D.-S. Gu, B.-K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Syst. Appl. 36(3) Part 2 (2009) 7252–7261. https://doi.org/10.1016/j.eswa.2008.09.033.

[16] N. Jamaludin, D. Mba, Monitoring extremely slow rolling element bearings: part I, NDT&E Int. 35(6) (2002), 349–358. https://doi.org/10.1016/S0963-8695(02)00005-1.

[17] A.G. Ray, Monitoring rolling contact bearings under adverse conditions, in: IMechE Conference on Vibrations in Rotating Machinery, Cambridge, 1980, pp. 187–194.

[18] N. Jamaludin, D. Mba, R.H. Bannister. Condition monitoring of slow-speed rolling element bearings using stress waves, Proc. Inst. Mech. Eng., E: J. Process. Mech. Eng. 215(4) (2001) 245–271.

[19] Y.-H. Kim, A.C.C. Tan, J. Mathew, B.-S. Yang, Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements, in: J. Mathew, J. Kennedy, L. Ma, A. Tan, D. Anderson (Eds.), Engineering Asset Management, Springer, London, England, 2006, pp. 182–191.

[20] J.Z. Sikorska, D. Mba, Challenges and obstacles in the application of acoustic emission to process machinery, Proc. Inst. Mech. Eng., E: J. Process. Mech. Eng. 222(1) (2008) 1–19.

[21] H. Lan, L. Yan, D. Xiao, Z. Jiao, I.-M. Chen, Surface-to-surface calibration of acoustic emission sensors, Sensor. Actuat. A-Phys., 174 (2012) 16–23. https://doi.org/10.1016/j.sna.2011.11.014.

[22] A. Klausen, K.G. Robbersmyr, Cross-correlation of whitened vibration signals for low-speed bearing diagnostics, Mech. Syst. Signal Process. 118 (2019) 226–244. https://doi.org/10.1016/j.ymssp.2018.08.048.

[23] J. Mathew, A. Szczepanik, B.T. Kuhnell, J.S. Stecki, Incipient damage detection in low speed bearings using the demodulated resonance technique, in: Proceedings of International Tribology Conference, Melbourne, 1987, pp. 366-369.

[24] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjöberg, J.-O. Strömberg, Towards automatic detection of local bearing defects in rotating machines, Mech. Syst. Signal Process. 19(3) (2005) 509–535. https://doi.org/10.1016/j.ymssp.2003.12.004.

[25] S. Lahdelma, E. Juuso, Signal processing and feature extraction by using real order derivatives and generalised norms. Part 2: Applications, International Journal of Condition Monitoring 1(2) (2011) 54–66. https://doi.org/10.1784/204764211798303814.

[26] R.B.W. Heng, M.J.M. Nor, Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition, Appl. Acoust. 53(1–3) (1998) 211–226. https://doi.org/10.1016/S0003-682X(97)00018-2.

[27] P.D. McFadden, J.D. Smith, Vibration monitoring of rolling element bearings by the high-frequency resonance technique – a review, Tribol. Int. 17(1) (1984) 3–10. https://doi.org/10.1016/0301-679X(84)90076-8.

[28] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20(2) (2006) 308–331. https://doi.org/10.1016/j.ymssp.2004.09.002.

[29] W. Su, F. Wang, H. Zhu, Z. Zhang, Z. Guo, Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement, Mech. Syst. Signal Process. 24(5) (2010) 1458–1472. https://doi.org/10.1016/j.ymssp.2009.11.011.

[30] X. Zhang, J. Zhou, Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines, Mech. Syst. Signal Process. 41(1–2) (2013) 127–140. https://doi.org/10.1016/j.ymssp.2013.07.006.

[31] Z. Feng, M. Liang, F. Chu, Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples, Mech. Syst. Signal Process. 38(1) (2013) 165–205. https://doi.org/10.1016/j.ymssp.2013.01.017.

[32] I. Antoniadis, G. Glossiotis, Cyclostationary analysis of rolling-element bearing vibration signals, J. Sound Vib. 248(5) (2001) 829–845. https://doi.org/10.1006/jsvi.2001.3815.

[33] D. Abboud, M. Elbadaoui, W.A. Smith, R.B. Randall, Advanced bearing diagnostics: A comparative study of two powerful approaches, Mech. Syst. Signal Process. 114 (2019) 604–627. https://doi.org/10.1016/j.ymssp.2018.05.011.

[34] J. Shiroishi, Y. Li, S. Liang, T. Kurfess, S. Danyluk, Bearing condition diagnostics via vibration and acoustic emission measurements, Mech. Syst. Signal Process. 11(5) (1997) 693–705. https://doi.org/10.1006/mssp.1997.0113.

[35] J. Antoni, Fast computation of the kurtogram for the detection of transient faults, Mech. Syst. Signal Process. 21(1) (2007) 108–124. https://doi.org/10.1016/j.ymssp.2005.12.002.

[36] Y. Zhang, R.B. Randall, Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram, Mech. Syst. Signal Process. 23(5) (2009) 1509–1517. https://doi.org/10.1016/j.ymssp.2009.02.003.

[37] T. Barszcz, A. Jablonski, A novel method for the optimal band selection for vibration signal demodulation and comparison with the Kurtogram, Mech. Syst. Signal Process. 25(1) (2011) 431–451. https://doi.org/10.1016/j.ymssp.2010.05.018.

[38] Y. Wang, M. Liang, An adaptive SK technique and its application for fault detection of rolling element bearings, Mech. Syst. Signal Process. 25(5) (2011) 1750–1764. https://doi.org/10.1016/j.ymssp.2010.12.008.

[39] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mech. Syst. Signal Process. 17(2) (2003) 317–328. https://doi.org/10.1006/mssp.2001.1462.

[40] C. Sobie, C. Freitas, M. Nicolai, Simulation-driven machine learning: Bearing fault classification, Mech. Syst. Signal Process. 99 (2018) 403–419. https://doi.org/10.1016/j.ymssp.2017.06.025.

[41] T.W. Rauber, F. de Assis Boldt, F.M. Varejao, Heterogenous feature models and feature selection applied to bearing fault diagnosis, IEEE T. Ind. Electron. 62(1) (2015) 637–646. https://doi.org/10.1109/TIE.2014.2327589.

[42] M. Zvokelj, S. Zupan, I. Prebil, Multivariate and multiscale monitoring of large-size low-speed bearings using Ensemple Empirical Mode Decomposition method combined with Principal Component Analysis, Mech. Syst. Signal Process. 24(4) (2010) 1049–1067. https://doi.org/10.1016/j.ymssp.2009.09.002.

[43] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.

[44] R.-P. Nikula, K. Karioja, K. Leiviskä, E. Juuso, Prediction of mechanical stress in roller leveler based on vibration measurements and steel strip properties, J. Intell. Manuf. 30(4) (2019) 1563–1579. https://doi.org/10.1007/s10845-017-1341-3.

[45] Y. Wang, J. Xiang, R. Markert, M. Liang, Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications, Mech. Syst. Signal Process. 66–67 (2016) 679–698. https://doi.org/10.1016/j.ymssp.2015.04.039.

[46] J. Singh, A.K. Darpe, S.P. Singh, Rolling element bearing fault diagnosis based on over-complete rational dilation wavelet transform and auto-correlation of analytic energy operator, Mech. Syst. Signal Process. 100 (2018) 662–693. https://doi.org/10.1016/j.ymssp.2017.06.040.

[47] A. Moshrefzadeh, A. Fasana, The Autogram: An effective approach for selecting the optimal demodulation band in rolling element bearing diagnosis, Mech. Syst. Signal Process. 105 (2018) 294–318. https://doi.org/10.1016/j.ymssp.2017.12.009.

[48] W.A. Smith, R.B. Randall, Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study, Mech. Syst. Signal Process. 64–65 (2015) 100–131. https://doi.org/10.1016/j.ymssp.2015.04.021.

[49] S. Lahdelma, E.K. Juuso, Signal processing in vibration analysis, in: The Fifth International Conference on Condition Monitoring and Machine Failure Prevention Technologies, Edinburgh, 2008, pp. 867–878.

[50] P.S. Bullen, Handbook of Means and Their Inequalities, second ed., Kluwer Academic Publishers, Dordrecht, 2003.

[51] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, revised ed., Holden-Day, Oakland, California, 1976.

[52] J.S. Oakland, R.F. Followell, Statistical Process Control, a Practical Guide, second ed., Heinemann Newnes, Oxford, 1990.