ScuDo

Scuola di Dottorato ~ Doctoral School

WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation

Doctoral Program in Mechanical Engineering (30^{th}cycle)

Fault Detection in Rotating Machinery

vibration analysis and numerical modeling

By

Ali Moshrefzadeh

******

Supervisor(s):

Prof. Alessandro Fasana

Doctoral Examination Committee:

Prof. Jerome Antoni, Referee, University of Lyon

Prof. Keith Worden, Referee, University of Sheffield

Prof. Paolo Pennacchi, Politecnico di Milano

Prof. Riccardo Rubini, University of Modena and Reggio Emilia

Prof. Stefano Marchesiello, Politecnico di Torino

Prof. Stefano Mauro, Politecnico di Torino

Politecnico di Torino

2018

Abstract

This thesis investigates vibration based machine condition monitoring and consists of two parts: bearing fault diagnosis and planetary gearbox modeling.

In the first part, a new rolling element bearing diagnosis technique is introduced. Envelope analysis is one of the most advantageous methods for rolling element bearing diagnostics but finding the suitable frequency band for demodulation has been a substantial challenge for a long time. Introduction of the Spectral Kurtosis (SK) and Kurtogram mostly solved this problem but in situations where signal to noise ratio is very low or in presence of non-Gaussian noise these methods will fail. This major drawback may noticeably decrease their effectiveness and goal of this thesis is to overcome this problem. Vibration signals from rolling element bearings exhibit high levels of 2nd order cyclostationarity, especially in the presence of localized faults. A second-order cyclostationary signal is one whose autocovariance function is a periodic function of time: the proposed method, named Autogram by the authors, takes advantage of this property to enhance the conventional Kurtogram. The method computes the kurtosis of the unbiased autocorrelation (AC) of the squared envelope of the demodulated and undecimated signal, rather than the kurtosis of the filtered time signal. Moreover, to take advantage of unique features of the lower and upper portions of the AC, two modified forms of kurtosis are introduced and the resulting colormaps are called Upper and Lower Autogram. In addition, a new thresholding method is also proposed to enhance the quality of the frequency spectrum analysis. Finally, the proposed method is tested on experimental data and compared with literature results so to assess its performances in rolling element bearing diagnostics.

Moreover, a second novel method for diagnosis of rolling element bearings is developed. This approach is a generalized version of the cepstrum pre-whitening (CPW) which is a simple and effective technique for bearing diagnosis. The superior

performance of the proposed method has been shown on two real case data. For the first case, the method successfully extracts bearing characteristic frequencies related to two defected bearings from the acquired signal. Moreover, the defect frequency was highlighted in case two, even in presence of strong electromagnetic interference (EMI).

The second part presents a newly developed lumped parameter model (LPM) of a planetary gear. Planets bearings of planetary gear sets exhibit high rate of failure; detection of these faults which may result in catastrophic breakdowns have always been challenging. Another objective of this thesis is to investigate the planetary gears vibration properties in healthy and faulty conditions. To seek this goal a previously proposed lumped parameter model (LPM) of planetary gear trains is integrated with a more comprehensive bearing model. This modified LPM includes time varying gear mesh and bearing stiffness and also nonlinear bearing stiffness due to the assumption of Hertzian contact between the rollers/balls and races. The proposed model is completely general and accepts any inner/outer race bearing defect location and profile in addition to its original capacity of modelling cracks and spalls of gears; therefore, various combinations of gears and bearing defects are also applicable. The model is exploited to attain the dynamic response of the system in order to identify and analyze localized faults signatures for inner and outer races as well as rolling elements of planets bearings. Moreover, bearing defect frequencies of inner/outer race and ball/roller and also their sidebands are discussed thoroughly. Finally, frequency response of the system for different sizes of planets bearing faults are compared and statistical diagnostic algorithms are tested to investigate faults presence and growth.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The key motivation for this PhD project is to improve the trustworthiness of rolling element bearings diagnosis. Moreover, increasing the understanding of the interactions between bearings and gears in a planetary gear set is another main motivation of this thesis.

Rolling element bearings (REBs) are one of the most used elements in rotating machinery and their failure is the most important cause of machinery breakdowns. Thus, correctly detecting and diagnosing bearing faults at stages prior to their complete failure is of vital importance. It avoids potential catastrophic damage not only to the apparatus but also to the personnel. Also, it reduces the machinery downtime which results in increasing the productivity.

Planetary gears, also recognized as epicyclic gears, are extensively used power transmission elements in numerous fields such as automotive, aerospace, wind turbines and marine applications. They have several benefits including compactness, high torque to weight ratio, high efficiency, multiple gear ratios and reduced noise in comparison with fixed-shaft gearboxes. Therefore, investigating planetary gear noise and vibration in healthy and faulty conditions is crucial to keep them functional and also to avoid any machinery breakdown as a result of a partial failure.

1.2 Research objectives

This thesis consists of the following two topics: diagnosis of bearing faults under complex environmental conditions is the first one, the other one is planetary gearbox modeling.

The first objective is to develop a new method to find the proper frequency band of demodulation for bearing faults diagnosis. It is an attempt to overcome the major drawback of the benchmark method in this field, the Fast Kurtogram, i.e. it is vulnerable to impulsive noise which makes it more sensitive to individual impulses than to series of transients. Enhancing the selection of correct frequency band for demodulation and further frequency analysis is advantageous since it limits the non-periodic impulses and noise from the raw time data, which are not related to any actual defects of bearings. In addition, it boosts the quality of the frequency analysis and increases the chance of successful fault diagnosis.

The second objective is to develop a new approach based on the cepstrum pre-whitening (CPW) method which has shown great potential for REBs diagnosis, despite its simplicity.

The third objective of this research is to develop an analytical model, which is capable of modeling the interaction between gears and bearings of a planetary gear set in the presence of bearing and gear faults. Mathematical modeling is an advantageous approach to scrutinize characteristics of mechanical systems. It gives a good understanding of structure dynamic characteristics; it is reasonably accurate and suitable for evaluations during design stages. In this regard, mathematical models such as lumped parameter models (LPMs) have been vastly used to study modal properties and also vibration signals of planetary gear trains. Moreover, the model should be able to simultaneously simulates the response of the system for different sizes, locations and profiles of bearings and gears defects.

1.3 Thesis Structure

The layout of this thesis is as follows:

Chapter 2: the literature of the vibration-based condition monitoring, with emphasis on rolling element bearings (REBs) diagnostics, will be reviewed.

Chapter 3: a new method for optimal demodulation band selection for non-stationary signals containing repetitive transients, e.g. bearing with localized defects, will be established.

Chapter 4: another new method for diagnosis of rolling element bearings based on the cepstrum pre-whitening idea will be developed. It takes a completely different approach than the method proposed in chapter 3.

Chapter 5: the performance of two proposed methods in chapter 3 and chapter 4 will be compared by conventional approaches for REBs diagnosis which will be reviewed in chapter 2.

Chapter 6: the literature of planetary gearbox simulation will be reviewed.

Chapter 7: a lumped parameter model with a comprehensive bearing model will be developed to investigate the gears and bearings interaction of a planetary gear train in presence of faults.

Chapter 8: the conclusion and the further work will be discussed.

Part I

ROLLING ELEMENT BEARINGS FAULT DIAGNOSIS

Chapter 2

Review of Rolling Element Bearing Diagnosis

2.1 Introduction

Rolling element bearings (REBs) are one of the most used elements in rotating machinery and their failure is the most important cause of machinery breakdowns. Thus, correctly detecting and diagnosing bearing faults at stages prior to their complete failure is of vital importance. It avoids potential catastrophic damage not only to the apparatus but also to the personnel.

In this chapter, the literature of the vibration-based condition monitoring, with emphasis on rolling element bearings (REBs) diagnostics, will be reviewed. Oil debris analysis, acoustic emissions and temperature monitoring are some other methods for bearing diagnostics but they are not discussed in this chapter as these are out of the scope of this thesis.

In section 2.2 the basic characteristics of vibrations induced by various bearings localized defects and defect frequencies related to each fault will be discussed. Proposed mathematical models to simulate the signal generated by localized faults are presented in section 2.3. The classic methods used in bearing fault diagnosis and enhancement of the bearing signals are presented in section 2.4 and section 2.6 respectively. Section 2.5 provides an overview of the techniques for separation of random and deterministic part of the signals.

2.2 Vibration-based bearing fault diagnosis

The outer race, inner race, cage and rolling elements are the key components of a bearing. The schematic structure of a rolling element bearing is presented in Figure 2.1.

As a localized defect develops either on an inner race, an outer race or a roller part of a bearing, an impact is generated each time the defect is engaged and consequently the bearing and machine structure are excited, in particular at their resonance frequencies. The corresponding vibration signal will comprise all the harmonics of this impact, which repeats almost periodically at a rate dependent on bearing geometry. Investigation of the generated vibrations is indispensable to detect the faults and many methods have been developed to extract the bearing characteristic frequencies (rate of generation of the impulses) from the measured vibrations.

Ref. [1] gives a clear and detailed explanation of the kinematics of bearings and general equations for the various frequencies shown in Figure 2.2 are given as follow:

Ballpass frequency, inner race (BPFI):

$$f_{bpfi} = \frac{N_b}{2}(f_o - f_i)(1 + \frac{D_b}{D_p}\cos\xi) \tag{2.1}$$

Ballpass frequency, outer race (BPFO):

$$f_{bpfo} = \frac{N_b}{2}(f_o - f_i)(1 - \frac{D_b}{D_p}\cos\xi) \tag{2.2}$$

Ball (roller) spin frequency (BSF):

$$f_{bsf} = \frac{f_o - f_i}{2}\frac{D_p}{D_b}(1 - (\frac{D_b}{D_p}\cos\xi)^2) \tag{2.3}$$

Fundamental train frequency/cage speed (FTF):

$$f_{cage} = \frac{f_i}{2}(1 - \frac{D_b}{D_p}\cos\xi\ ) + \frac{f_o}{2}(1 + \frac{D_b}{D_p}\cos\xi\ ) \tag{2.4}$$

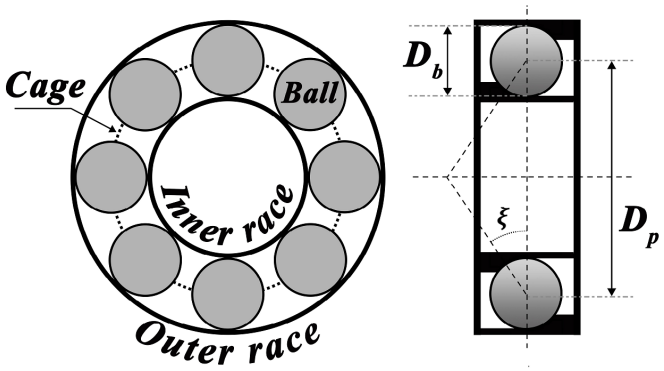

Fig. 2.1 Bearing schematic

where f_i is the inner race frequency (shaft rotating frequency), f_o is the outer race frequency, N_b is the number of rolling elements, D_b and D_p are the rolling elements and pitch diameters and ξ is the operating contact angle (see Figure 2.1).

Refs. [2–4] explain that there are two main sources of amplitude modulation of these series of broadband bursts

1. The strength of the bursts: the load carried and transported by the rolling elements can be modulated by the rate at which the fault is passing through the load zone
2. Rotation of the faults: the transmission path varies between the fix transducers and the rotating defects

Outer race is usually stationary and its faults mostly occur in the load zone. Therefore, the defect impulses will not be modulated as shown in Figure 2.2. In contrast, the defect impulses for inner race faults and rolling elements are modulated as they pass through the load zone at shaft frequency and FTF (see Figure 2.2), i.e. sidebands are spaced at shaft frequency and FTF around BPFI and BSF respectively. Moreover, even harmonics of the BSF are often dominant as the fault on the rolling element is engaged with the inner race and outer race once per each revolution.

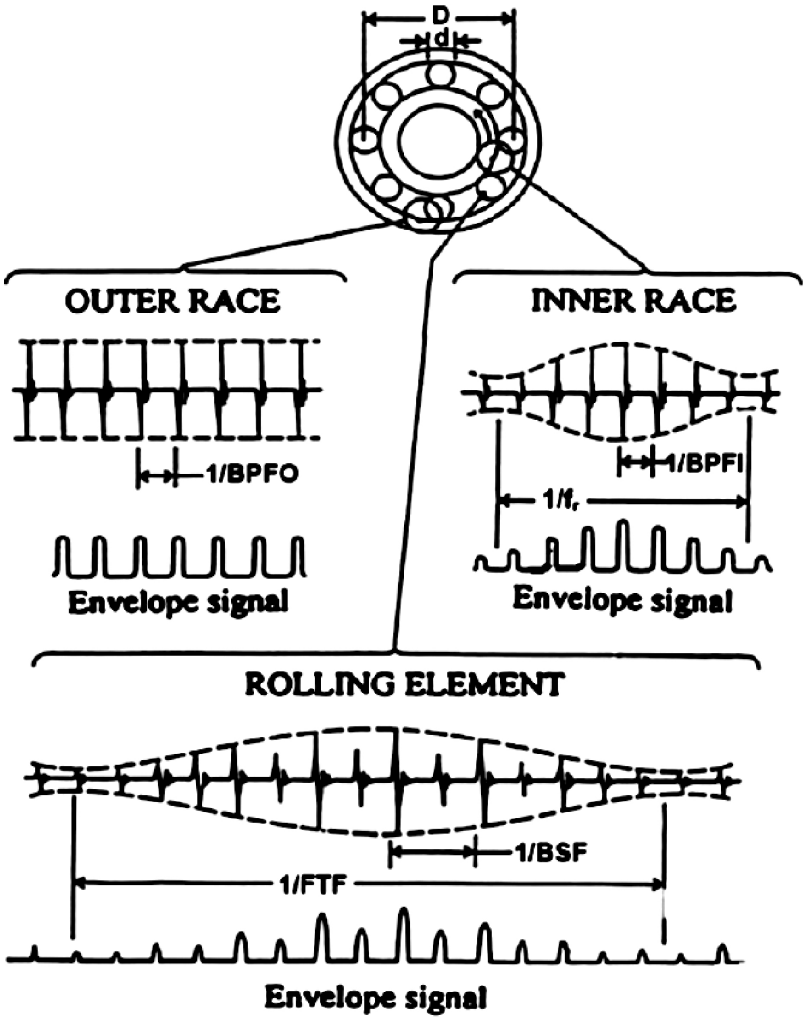

Fig. 2.2 Rolling element bearing vibration signal characteristics due to local faults [2]

2.3 Modeling the vibration of bearing localized defects

Several mathematical models are proposed to simulate the signal generated by localized faults in rolling element bearings. McFadden and Smith [3] in their classic paper developed a model to describe the vibration produced by a single point defect on the inner race of a rolling element bearing under constant radial load. The vibration produced by the defect is modelled as a series of impulses spaced with constant period and amplitude modulated by different sources. Acquired vibration $v(t)$ by the transducer is presented by the follow equation:

$$v(t) = [d(t)\, q(t)\, a(t)] \times e(t)$$

This equation is graphically demonstrated in Figure 2.3 where $d(t)$ is a series of impulses generated by a localized defect with constant amplitude d_0, $q(t)$ presents the load distribution on the circumference of a bearing under a radial load, $a(t)$ considers the effect of the changing transmission path between the location of the defect and the transducer on the machine casing and also changing in angle of the applied impulse and $e(t)$ represent the exponential decay of vibration.

The frequencies and relative amplitudes of the spectral lines produced by this model agree with those found in the measured spectrum. McFadden and Smith [4] further extended this model to describe the vibration produced by multiple point defects. Moreover, they studied the pattern of the lines in the spectrum as in cases with multiple defects the varying phase, related to the different position of the defects, angles give cancellation and reinforcement of the defect frequency and sidebands. The model was latter refined by Ho and Randall [5] by adding slight random variations in the time between defect pulses so as to produce more realistic vibration signals. Though this variation is small, it is sufficient to exclude the resulting signal from the class of periodic processes.

2.3.1 Cyclostationarity

Vibration signals are categorized as deterministic, random or a combination of both. Deterministic signals are further considered as periodic and non-periodic, and random

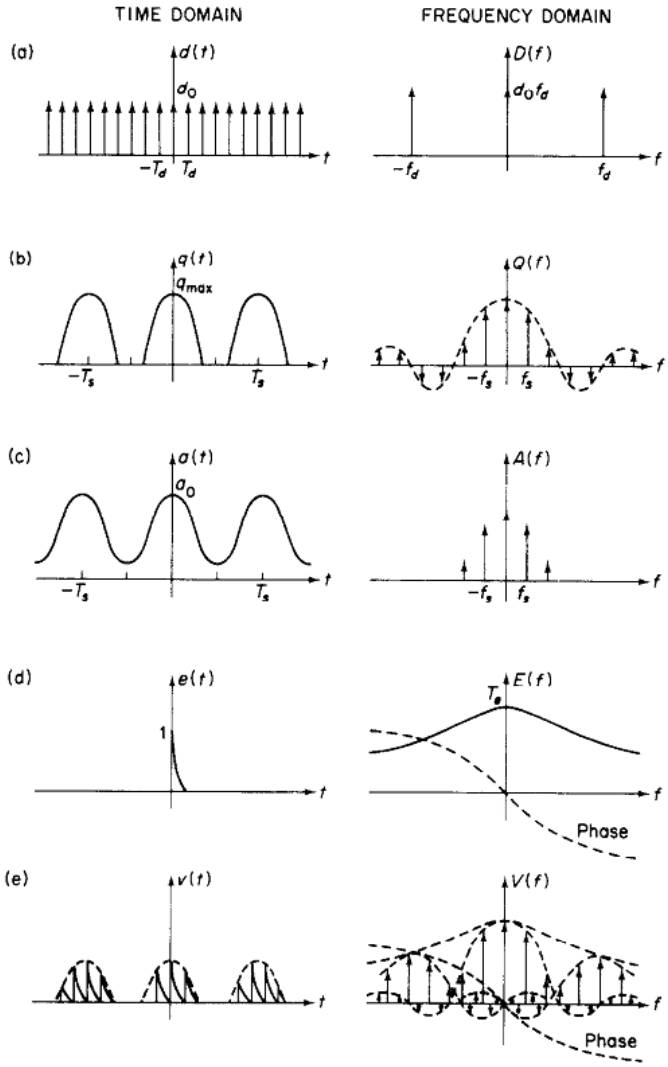

Fig. 2.3 Model of the signal generated by a bearing with a single defect [4]

signals as stationary and non-stationary. Some processes are not generally periodic functions of time yet are fundamentally generated by a hidden periodic mechanism. Many processes in mechanics emerge form periodic phenomena such as rotation and reciprocation of gears, bearing, belts, chains, shafts, propellers, pistons, and so on. Signals of this kind are called cyclostationary, their statistical properties change cyclically with time and the capacity of traditional signal processing techniques can be extended to take advantage of these characteristics. A signal is assumed to be cyclostationary of order n when its nth order statistics is periodic, i.e. after passing the signal through any nonlinear transformation including nth power the resultant signal is periodic; therefore, peaks can be spotted in its representation in frequency domain [6].

Considerable amount of research has been published on cyclostationarity. A concise survey of the literature can be found in Ref. [7]. Moreover, an introduction to cyclostationarity from first principles, with a special emphasis on intuition rather than on mathematics, to design and implement simple cyclostationary-based estimators is provided by Ref. [8]. Refs. [9, 10] investigated the bearing signal characteristic in presence of localized faults. The inner race, outer race and rolling elements defects are the most frequent bearing faults. As was discussed, the vibration produced by the localized faults are a sequence of impulses dominated by the high resonance frequencies of the structure. Slippage of the rolling elements and cage introduce some level of randomness in spacing of the impacts, and although it is not larger than a few percent (typically on the order of 1-2%) of the rate of oscillating bursts repetition and the impacts have a period almost equal to the ball-pass period, the resulting signal cannot be categorized as a periodic process. Ref. [9] pointed out that the signal of a bearing with localized fault may be modelled as a 2nd order cyclostationary process. A second order cyclostationarity determines processes with a periodic autocovariance function in time.

$$R_{xx}(t,\tau) = \mathbf{E}\{x(t-\tau/2)x(t+\tau/2)\} \tag{2.5a}$$

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) \tag{2.5b}$$

where $\mathbf{E}\{\,.\}$ is the expected value operator and T is the period of autocovariance. Autocovariance function is a function of not only the time lag τ but also the instantaneous time t and it should not be confused with the autocorrelation function of a stationary process, calculated as the time average.

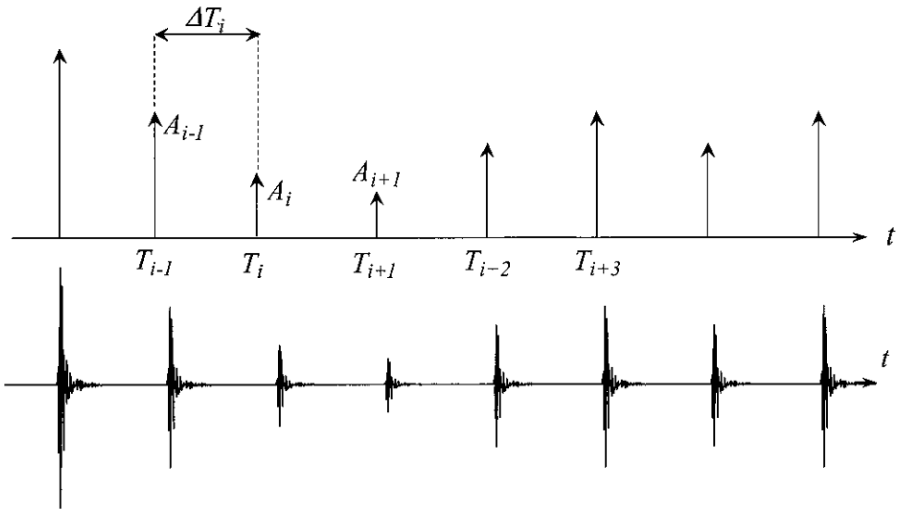

Fig. 2.4 Expected bearing vibration signal [10]

Figure 2.4 shows an example of a bearing defect impulses. For a second order cyclostationary process, the times of occurrence are periodic with an average time between two impacts T, but with some random jitter around each node so that

$$T_i = iT + \delta T_i$$

where i is the number of the impulse and δT_i is random uncertainty (typically a few percent of T). Effect of a very slight random slip on the frequency domain representation for a train of Dirac pulses (impact forces) is shown in Figure 2.5 where the higher harmonics vanish and turn into a continuous baseline.

Ref. [10] denotes the cyclostationarity assumption may not be so accurate because an impact should occur T seconds later than the actual time of occurrence of the previous one and as a result the uncertainty increases as the time of prediction increases. Therefore, the signals from a bearing localized fault is strictly speaking a quasi-cyclostationary process rather than a cyclostationary process. When the randomness is not high, Ref. [10] denotes these processes as pseudo-cyclostationary, as they seem to be cyclostationary but in fact are not. However, they can be treated as cyclostationary in a first approximation as the departure from cyclostationarity may essentially be really slow.

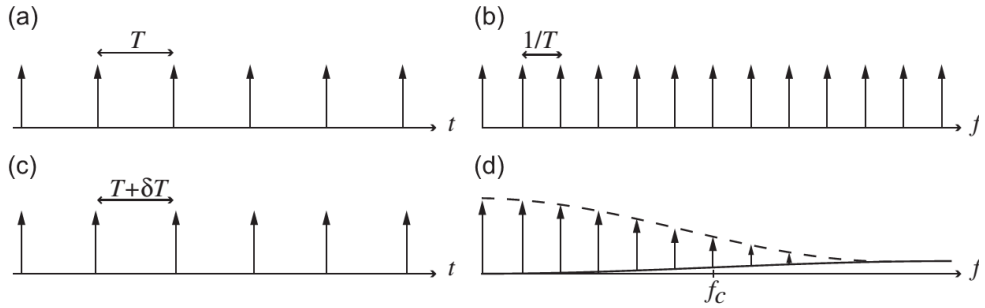

Fig. 2.5 (a) An ideal train of Dirac pulses in the time-domain and its (b) frequency domain representation (c) the same train of Dirac pulses plus random fluctuations and its (d) frequency domain representation [11]

2.4 Vibration analysis techniques for diagnosis of rolling element bearings

Various damage detection techniques based on the vibration analysis have been developed over the years for diagnosis of rotary machinery, e.g. rolling element bearings and gearboxes, as fault development deviates the vibration signature of a machine form its standard condition. These changes can be related to the fault and be utilized for diagnosis of the system [12]. In general, these techniques can be divided in five categories:

- Time domain techniques
- Frequency domain techniques
- Time-frequency techniques
- Cyclostationary analysis
- Cepstral analysis

2.4.1 Time domain analysis

Initially, the statistical characteristics of the signal in the time domain collected from transducers, typically accelerometers, were the main focus of study [13]. To distinguish between faulty and healthy conditions these methods employ certain

statistical parameters of the waveform. Statistical parameters such as peak value, root mean square (RMS), crest factor, skewness and kurtosis are among the mostly used indicators.

$$\begin{aligned}
\text{RMS}_x &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} \\
\text{Crest factor}_x &= \frac{\text{maximum peak value}}{\text{RMS}} \\
\text{Kurtosis}_x &= \frac{\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2} \\
\text{Skewness}_x &= \frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^3}
\end{aligned} \tag{2.6}$$

where x_i is the sample at time t_i, $\bar{x}$ and N are the mean value and length of the data set.

Moreover, different indicators have been developed over the years for different diagnosis purposes, e.g. more sensitive parameter to peaks, continuing growth of the fault, damage progresses from localized to distributed, etc. [14].

Comparing the vibration levels in machines with the standard criteria for vibration severity is the simplest but not the most reliable technique for fault detection. Figure 2.6 demonstrates an example of peak, RMS and crest factor evaluation for typical spall development in a bearing [12].

2.4.2 Frequency domain analysis

As it was discussed in section 2.2 defects in rolling element bearings produce pulses of very short duration. The natural frequencies of bearing elements and housing structures are excited by these pulses which increase the vibrational energy at these high frequencies. The main idea of frequency domain or spectral analysis methods analysis is to either investigate the whole spectrum or certain frequency components of interest and thus obtain features from the signal and then extracting the bearing fault feature [15, 16]. In this technique, mainly the power spectrum of the time domain signal in logarithmic or decibel scale is generated by means of FFT as a tool for comparison (see Figure 2.7). For instance, Ref. [12] mentions that based on the

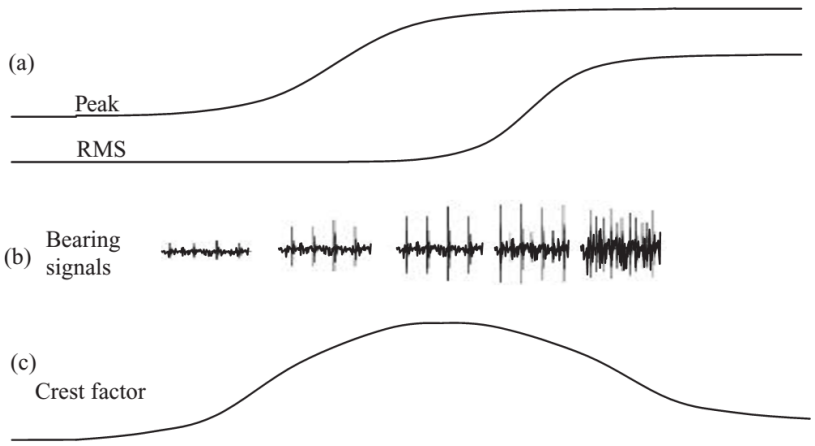

Fig. 2.6 Typical trend of time domain indicators with fault development [12]

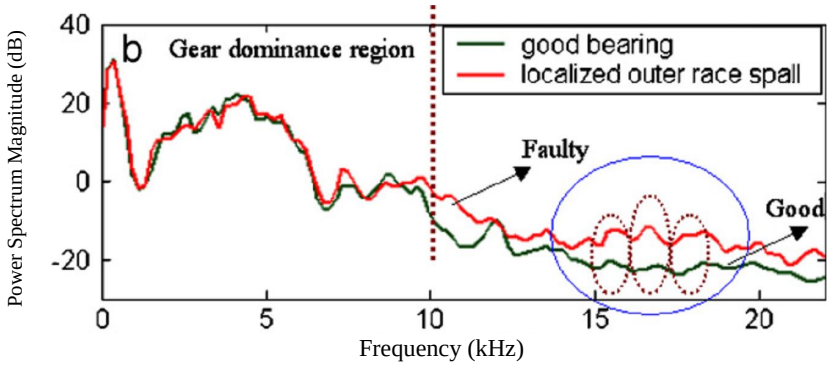

Fig. 2.7 Power spectrum comparison for good and outer race defect bearing [17]

well-established standard a change in vibration level of 6–8 dB is significant and a change of 20 dB should be considered as a serious condition.

2.4.3 Time-frequency analysis

Frequency domain analysis are limited to stationary signals and are not able to examine nonstationary waveforms. Under these circumstances time-frequency analysis, which processes the vibration signal simultaneously in time and frequency domains, increases the chances of detecting faults in rotating machinery as nonstationary techniques provide information about the local time-domain properties of a signal.

Refs. [16, 18] reviewed the class of time–frequency distributions. Short-time Fourier transform (STFT) or spectrogram (the power of STFT), wavelet analysis (WA) and Wigner–Ville distribution (WVD) are the most widespread techniques.

The idea of STFT which is a time variable version of the traditional FT, is to slice the whole waveform signal into (overlapped) segments by using a time window moving along the time axis to divide the signal and then calculate the local FT for each segment. The spectrogram is the squared magnitude of the STFT and represents the energy density spectrum of the signal as a function of time. Window length, which is responsible for the time resolution of the analysis, should be set in advance but it should be considered that it is not possible to improve the time and frequency resolution at the same time since the product of them is a constant.

The WVD is defined as the FT of autocorrelation function with respect to the time-lag variable. This is the basis of almost all the bilinear time-frequency distributions. As the method is not based on signal segmentation it has the highest time-frequency resolution, but it suffers from possibility of negative energy levels and inevitable cross-term (non-physical) interferences which make it difficult to interpret. To take advantage of the high time-frequency resolution and also to overcome these drawbacks, enhanced versions such as affine class distributions, Cohen class distributions, adaptive optimal kernel and reassignment method have been proposed [18].

WA is a well-known and powerful tool for time-frequency analysis of waveforms and has been quickly developed in the past decade and has wide application in various area, e.g. fault diagnostics of gears, bearings and other mechanical systems [19–22]. In contrast to FT which employs sinusoidal functions as the basis it uses wavelets which can be orthogonal or non-orthogonal, and continuous or discrete. As the wavelets can independently be dilated and shifted as a function of time this method is effective for time–frequency localization, and is suited to transient signal analysis. An example of short time Fourier transform, Wigner–Ville distribution and continuous wavelet transform spectrogram of a synthetic frequency modulated (FM) signal (Figure 2.8a) is shown in Figure 2.8b-d.

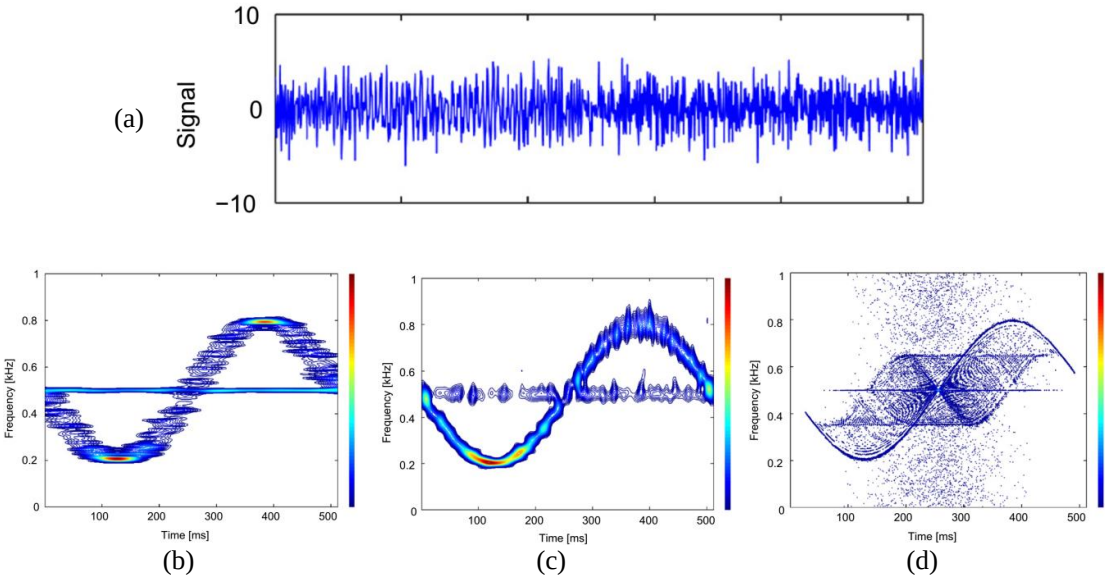

Fig. 2.8 (a) Signal composed of a sinusoidal wave and a sinusoidal frequency modulated (FM) wave (b) short time Fourier transform (c) Wavelet transform (d) Wigner–Ville distribution [18]

2.4.4 Cyclostationary analysis

Separation of deterministic and random signals will be discussed in section 2.5, but it is also possible to further categorize the random signals based on their cyclostationary properties. Second order cyclostationary analysis for bearing diagnosis has become popular in the field of machine diagnostics and a series of digital signal processing techniques have been developed to study their signature. In this section, we shortly review some of the most widespread techniques due to their high computational efficiency and/or simple implementation.

2.4.4.1 Envelope analysis

Envelope analysis is the best-known and the most important bearing diagnostic algorithm. It was first proposed by Ref. [23] and named High-frequency resonance technique (HFRT) by author but now it is mainly known with the former name. Spectrum of raw signal often does not contain sufficient diagnostic information related to bearing faults therefore envelope analysis has become the benchmark method for bearing diagnostics.

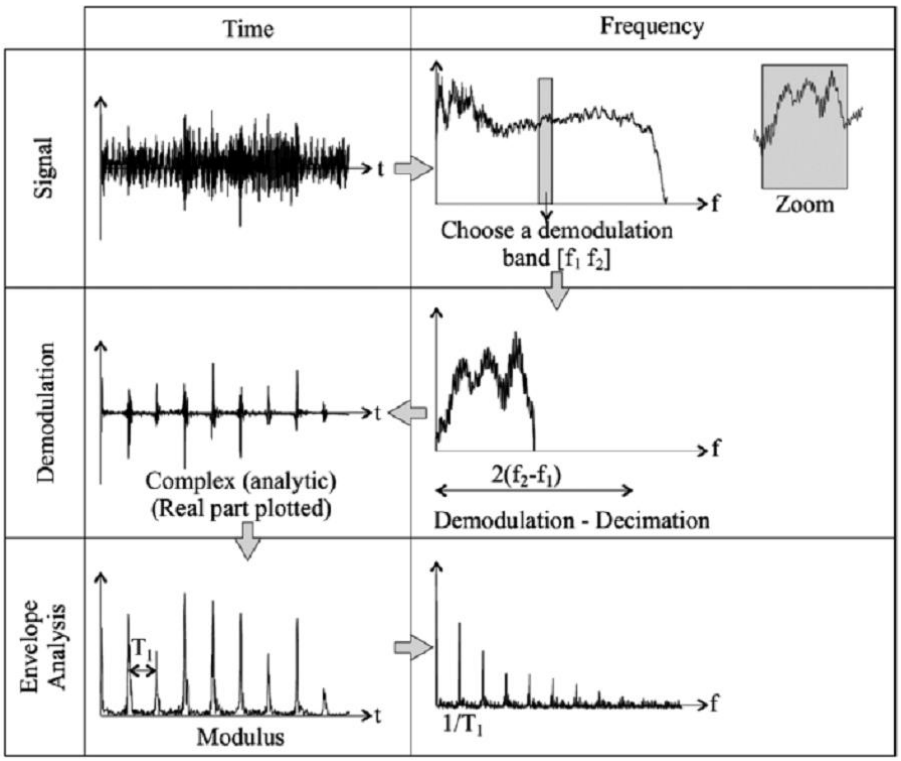

Fig. 2.9 Procedure for envelope analysis [12]

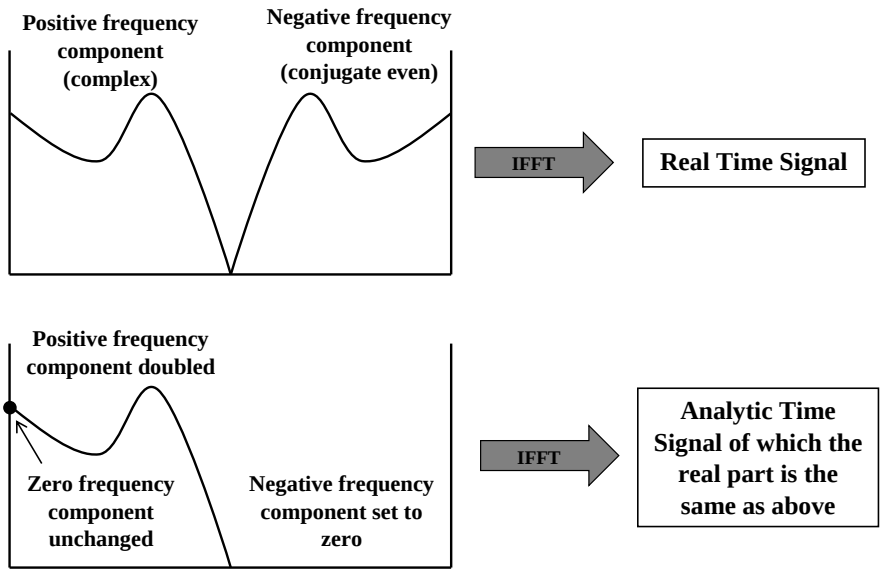

Fig. 2.10 Manipulation of the positive frequency spectrum to obtain a real time signal (IFFT = inverse Fourier transform) [12]

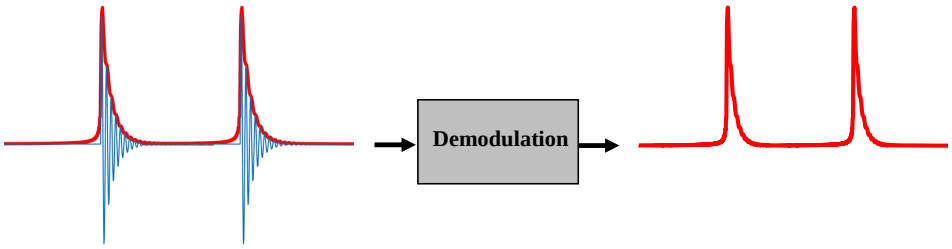

Fig. 2.11 Amplitude demodulation principle

As it was discussed in section 2.2, resonances of the rolling element bearing and the surrounding structure will be excited by the impacts produced by localized defects. This technique band-pass filters a bearing signal for these resonance frequencies to remove low-frequency mechanical noise and interference from other sources. Afterward, envelope of the filtered signal is generated by means of Hilbert transform or a squaring operation followed by a low-pass filtering. Then the periodicity of the envelope signal is computed and compared with the defect frequencies (see section 2.2) to investigate the presence of the defects [24, 12].

This process is depicted in Figure 2.9 in which the envelope of the signal is formed by calculating the square of the complex analytical signal which is produced by taking the inverse FT of the zero-padded and shifted one-sided frequency band of interest for demodulation [6]. The schematic for computing the analytic time signal by using this approach is shown in Figure 2.10.

Amplitude demodulation is the process by which the signal envelope (modulation signal) is extracted (see Figure 2.11). But, selecting the proper frequency band which carries the cyclostationary components for demodulation of the bearing signal is crucial for a successful diagnosis. In subsection 2.6.2, some of the most well-known methods proposed for this purpose will be discussed.

Based on the log-transformation of the signal envelope a new 2nd order cyclostationary indicator namely the “log-envelope” has been proposed by Ref. [25] to eliminate the interaction between different 2nd order cyclostationary components. Furthermore, Ref. [26] investigates the superiority of the log-envelope in dealing with signals containing high impulsive noise either as a result of harsh operating situations or electric noise generated by power electronics and captured by the sensor.

2.4.4.2 Cyclic Spectral Analysis (CSA)

As it was discussed in subsection 2.3.1, cyclostationary is a subclass of nonstationary signals which exhibit some cyclical behavior and the term *cyclic frequency* indicates the periodicity in the magnitude (power, variance, envelope) of the random signal. In case of defected bearing this periodicity is analogous to the bearing characteristic frequencies (see section 2.2)

Cyclic Spectral Analysis (CSA) is a powerful approach for condition monitoring of rotary machinery which separates and describes different 2nd order cyclostationary components in terms of the spectral (carrier) frequency and the cyclic (modulation) frequency variables.

CSA was brought to the field of diagnostics of mechanical systems by Antoni [27, 28] but it did not get the attention it deserves, perhaps due to its advanced theory of stochastic processes and computational cost [29]. Spectral Correlation (SC) which is a bi-spectral (modulation frequency (α) and carrier frequency (f)) map (see Figure 2.12), is one of the main tool for the CSA of machine signals. The SC is defined in Equation 2.7 as the two-dimensional FT of the instantaneous autocovariance function (Equation 2.5). Also, it is noticeable that the integral of the spectral correlation over all frequency f is effectively the squared envelope spectrum [6].

$$S_x(\alpha, f) = \lim_{N\to\infty} \frac{1}{(2N+1)f_s} \sum_{n=-N}^{N} \sum_{m=-\infty}^{\infty} R_{xx}(t_n, \tau_m) e^{-j2\pi n \frac{\alpha}{f_s}} e^{-j2\pi m \frac{f}{f_s}} \qquad t_n = \frac{n}{f_s}, \quad \tau_n = \frac{m}{f_s} \tag{2.7}$$

where f_s is the sampling frequency.

Various methods have been proposed to estimate the SC [31]. Among them Averaged Cyclic Periodogram (ACP) or time-smoothed cyclic periodogram is popular estimator which can be computed as follows [29]:

$$S_x^{ACP}(\alpha, f) = \frac{1}{K\|w\|^2 f_s} \sum_{i=1}^{K-1} X_w(i, f) X_w(i, f-\alpha)^* \tag{2.8}$$

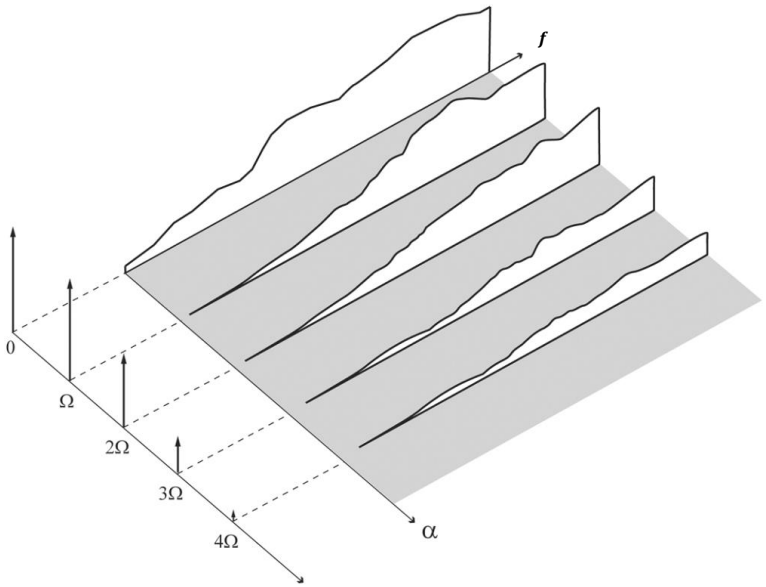

Fig. 2.12 Schematic spectral correlation density showing a typical cyclostationary signature, continuous in the spectral frequency (f) and discrete in the cyclic frequency (α) [30]

$$\|w\|^2 = \sum_{n=0}^{L_w-1} |w_n|^2$$

where $X_w(f)$ is SFTF of the signal $x(t)$ with length of N, w_n is the window function, $K = \frac{N-L_w}{R} + 1$ is the total number of averaged segments (total number of windows with length of L_w shifted by R samples).

The Spectral Coherence is defined as:

$$\gamma_x(\alpha, f) = \frac{S_x(\alpha, f)}{\sqrt{S_x(f)S_x(\alpha, f)}} \tag{2.9}$$

where $S_x(f)$ is a magnitude normalized form of the SC within 0 and 1. The Spectral Coherence intensify the effect of weak cyclostationary signals which may have very small amplitude in the SC.

ACP is an ideal tool in condition monitoring due to its high capability in decomposition of the signal with respect to the modulation and the carrier frequencies which makes the identification of fault frequencies easy, i.e. ACP provides a high-resolution version of the envelope spectrum. But being computationally intensive makes it unsatisfactory for industrial application.

Cyclic Modulation Spectrum (CMS) has been proposed as a fast alternative to the SC [8, 30]. It is actually a cascade of envelope spectra in all possible frequency bands, i.e. FT of the spectrogram or the squared envelope at the output of a filter bank. Ref. [32] gives a new approach to include CMS in the unified form of cyclostationary estimators as an estimator of the SC. However, CMS suffers from a major drawback because cyclic frequencies larger than the frequency resolution of the STFT cannot be detected.

Recently Ref. [29] introduced a fast algorithm to compute the SC which substantially increases the computational efficiency. This is mainly achieved by calculating the auto/cross correlation function by means of the FFT algorithm instead of a loop on time-lags. In contrast to biased estimators such as the CMS, Fast-SC generates similar statistical performance as the ACP. They also proposed a new spectral quantity of "Enhanced Envelope Spectrum" (EES) which is calculated based on the Spectral Coherence for a given frequency band $[f_1\ f_2]$ as follows:

$$\mathrm{EES}(\alpha) = \int_{f_1}^{f_2} |\gamma_x(\alpha, f)| \mathrm{d}f \tag{2.10}$$

2.4.5 Cepstral analysis

Cepstral analysis is advantageous in detecting spectrum periodicities by reducing a whole family of harmonics into a single cepstral line. In other words it can be considered as 'spectrum of a spectrum'. The definition of the complex cepstrum is:

$$C = \mathrm{IFT}\left\{\log\left(\mathrm{FT}\left(x\right)\right)\right\} = \mathrm{IFT}\{\log(A(f)) + j\phi(f)\} \tag{2.11}$$

where FT and IFT refer to Fourier transform and inverse Fourier transform respectively and

$$\mathrm{FT}(x) = A(f)e^{j\phi(f)} \tag{2.12}$$

The real cepstrum is calculated by setting the phase to zero as follows:

$$C = \mathrm{IFT}\left\{\log\left|\mathrm{FT}\left(x\right)\right|\right\} \tag{2.13}$$

The word cepstrum is generated by manipulating the word spectrum. In similar manner, quefrency, rahmonics and lifter are also warped versions of the correspond-

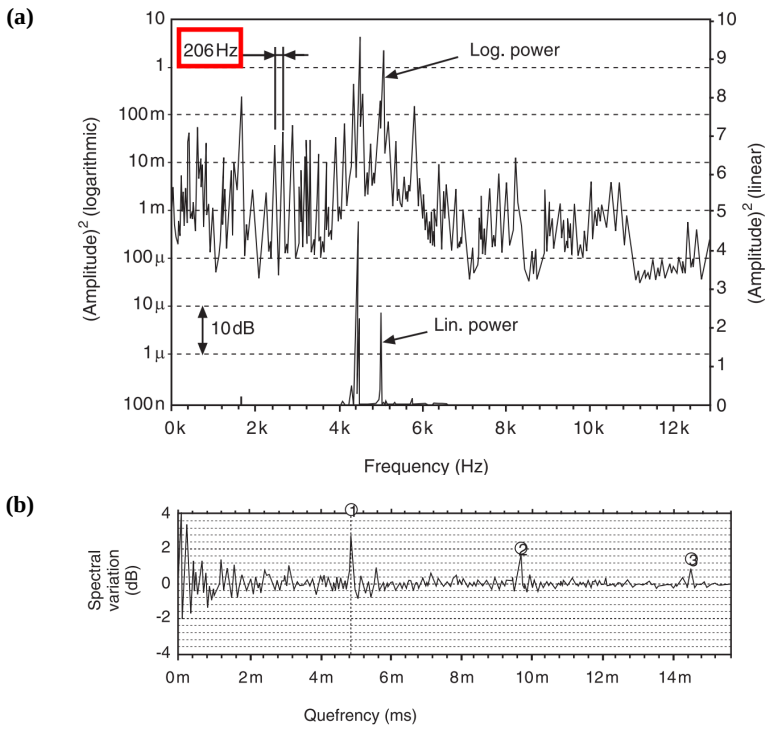

Fig. 2.13 (a) linear vs logarithmic amplitude scales power spectrum (b) cepstrum [24]

ing spectrum terms of frequency, harmonics and filter. The x-axis of the cepstrum is quefrency and has the unit of time. Rahmonics are a series of uniformly spaced peaks in the cepstrum and liftering is achieved by windowing the cepstrum.

The major distinction of cepstrum roots in the logarithmic conversion. It has two important characteristics. First, level of periodic components such as families of harmonics and sidebands are boosted by log-spectrum in comparison to the linear spectrum. Therefore, the inverse Fourier transform will manifest the periodicity effectively. Second, the multiplicative effect of transmission path is changed into an additive one and the two effects can be separated in this manner.

To exemplify the first feature, spectrum and log-spectrum of a signal related to a bearing with a defected outer race are depicted in Figure 2.13a. For this case the BPFO is equal to 206 Hz. By comparing these two plots, peaks spaced at BPFO are only noticeable in the log-spectrum. The cepstrum for the same signal is shown in Figure 2.13b. The peak at quefrency 4.85 ms ($\frac{1}{\text{BPFO}} = \frac{1}{206\ \text{Hz}}$) which indicates the presence of the fault is shown by circle number 1. Also, 2nd and 3rd rahmonics of this quefrency can be detected in the cepstrum by circle number 2 and 3.

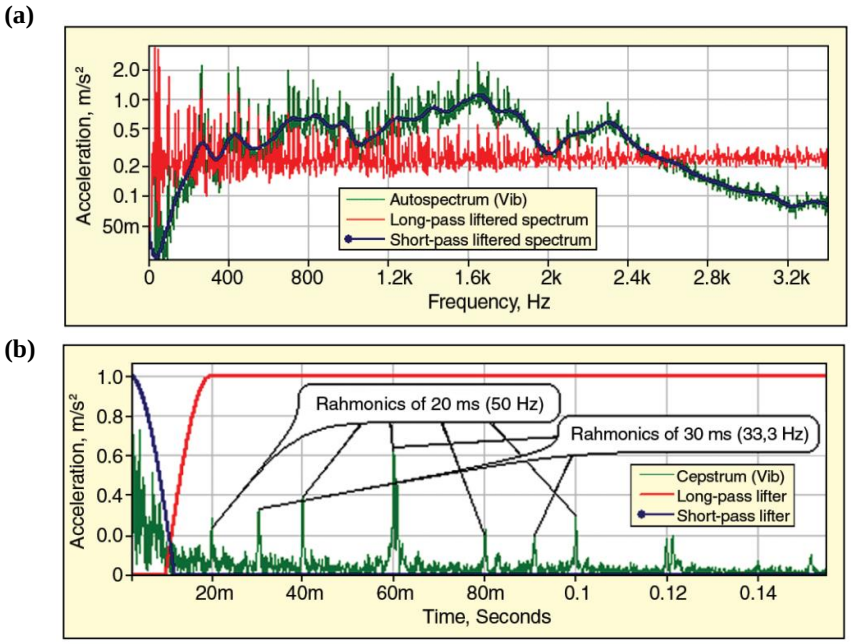

Fig. 2.14 Cepstral analysis (a) original and liftered vibration spectra (b) cepstrum and lifters [33]

The second important characteristic is illustrated with the following example. The green line in Figure 2.14a shows the spectrum of a signal collected from a faulty gearbox. The cesptrum is depicted by green line in Figure 2.14b. Ref. [33] claims that the unknown transfer path manifests itself at short quefrencies in the cepstrum. In contrast, the defect quefrencies have long quefrencies. Therefore, by employing short-pass lifter and long-pass lifter (see Figure 2.14b) these two components can be efficiently separated.

Although, it should be mentioned cepstral analysis is not a popular approach despite some of its interesting features. One reason might be that the computation and interpretation is not as easy as the envelope analysis which is an everlasting tool in condition monitoring.

2.5 Random and deterministic signal separation

Vibration analysis of machines gains much in efficiency if periodic vibrations can be separated out from non-deterministic ones. Discrete frequency “noise” from

various sources such as gears and fans are usually strong and mask the bearing signal. Therefore, it is beneficial to separate the signals of different types before diagnosis of bearings. The random slippage in bearings, though small, causes to signals be random. It makes it possible to effectively separate the random bearing signal from the deterministic signals. A number of techniques have been developed for this purpose and in the following the main methods will be discussed.

2.5.1 Autoregressive model (AR)

Autoregressive model (AR) is a widespread method to establish the deterministic components in a data sequence. This method tries to build a linear model to forecast the current value of a signal based on its previous values. An AR model of order p can be written as [34]

$$x(n) = \sum_{i=1}^{p} a(i)\, x(n-i) + e(n)$$

where the $e(n)$ is the error function, i.e. the residual or unpredictable part of the signal. The discrete frequency "noise" which is deterministic and predictable can be established by the model. On the other hand, the bearing signals are random and cannot be predicted. Therefore, the error function $e(n)$ will include the stationary white noise and the desired bearing signal. As spectrum of the residual part is white this process is called pre-whitening.

2.5.2 Time Synchronous Averaging (TSA)

Time synchronous averaging (TSA) is the oldest and one of the most effective signal processing tools applied to rotating machinery for extraction of the periodic parts from a composite signal with minimum disruption. It was first proposed by Ref. [35] to extract the repetitive waveforms buried in noise, and takes advantage of the redundant information provided by repetition. Time domain averaging is achieved by averaging the data segments for each period as follows [34]:

$$y(t) = \frac{1}{N} \sum_{\mathrm{i}=0}^{N-1} x(t + i\, T)$$

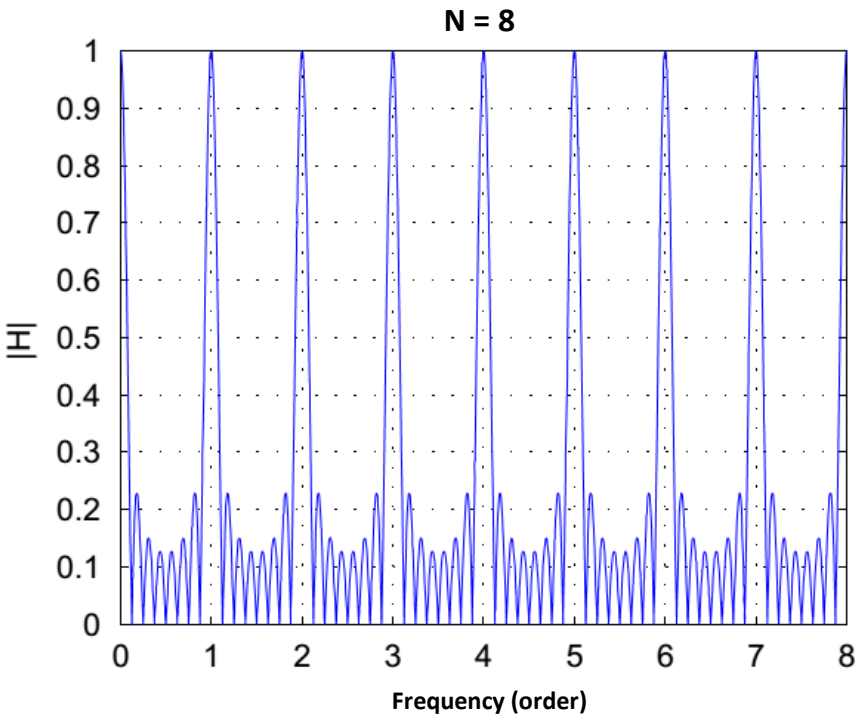

Fig. 2.15 Comb filter for an average of eight synchronous [37]

where $y(t)$ is the synchronizing signal, T is the chosen period and N is the number of section averaged. Periodic components related to the period T are thus preserved in $y(t)$, where other component, e.g. signal corresponding to different periods or noise, will be attenuated and converge to zero. In the frequency domain it is equivalent to the multiplication of the Fourier transform of the signal by a comb filter (Figure 2.15), therefore only those components with the fundamental and harmonic frequencies of the desired signal are passing through the filter [36]. Increasing the N narrows the teeth of the comb and reduces the level of the side lobes between the teeth, consequently improving the estimate of the desired signal [36].

Mcfadden [36] revised the comb filter presented by Ref. [35] which requires an infinite number of samples of the signal and produces a result which is not exactly periodic. It was achieved by applying a rectangular window to the noisy signal in the time domain and sampling the Fourier transform of the signal in the frequency domain. A more general approach of synchronous average can be found in Ref. [37] which discusses adaptation of specific comb filters for different type of signals.

One drawback of TSA is that it needs prior knowledge of a system period and it must separately be applied for each period T, e.g. in the case of gearboxes with multiple shafts. Moreover, in real cases speed of rotation is not completely constant, thus 'order tracking' or 'angular resampling' should be performed before the averaging process to remove the effect of speed fluctuation. Therefore, a tachometer or key phasor signal is also needed to implement this method. This process could be tedious as requires separate angular sampling for each harmonic family.

2.5.3 Self-Adaptive Noise Cancellation (SANC)

One usual method of estimating a signal corrupted by additive noise is to pass it through a filter that suppresses the noise while not disrupting the signal of interest.

A schematic of an adaptive noise cancelation (ANC) is depicted in Figure 2.16. It requires little or no prior knowledge of the signal and noise characteristics [38]. In this method, a reference input signal is passing through a filter and then the resultant is subtracted from the primary input containing both the signal and the noise. Consequently, the noisy part of the signal is attenuated or eliminated and the outcome of the process should be the denoised signal.

Adaptive filters are able to filter out part of the signal in the primary input (combination of signal and noise) which has correlation with the reference input (noise) by automatically adjusting its impulse response. It is not necessary that the reference signal be equal to the noisy part of the primary input, which is desired to be removed. The filtering process will be successful if they can be related by a linear transfer function [34].

Self-adaptive noise cancellation (SANC) is a modified version of the ANC in which the reference signal is replaced by a delayed version of the primary signal. Therefore, there is no need of an acquired reference signal which is not available in many cases.

When a signal is combination of random and deterministic parts, the adaptive filter finds the transfer function between the signal and its delayed version. As there is no correlation between the random part of the signal and its delayed version, if the delay is larger than the correlation-time of the random part, the transfer function only represents the relationship between the deterministic part of the signal and the delayed version of it, i.e. the transfer function is a delay function. A schematic

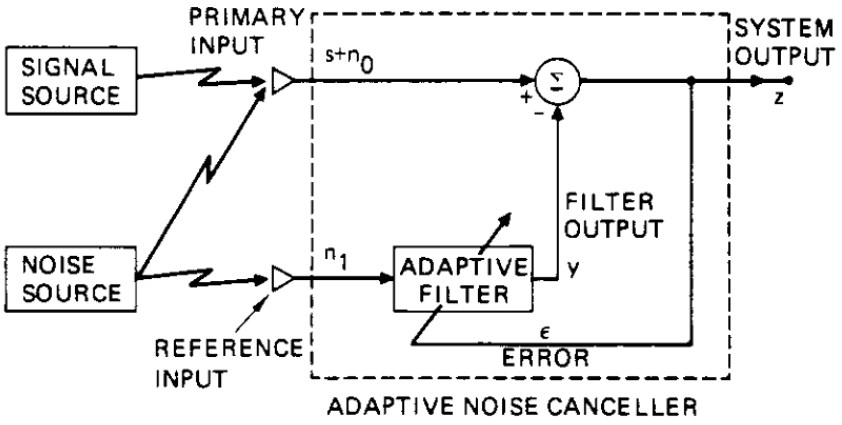

Fig. 2.16 The adaptive noise cancelation (ANC) concept [38]

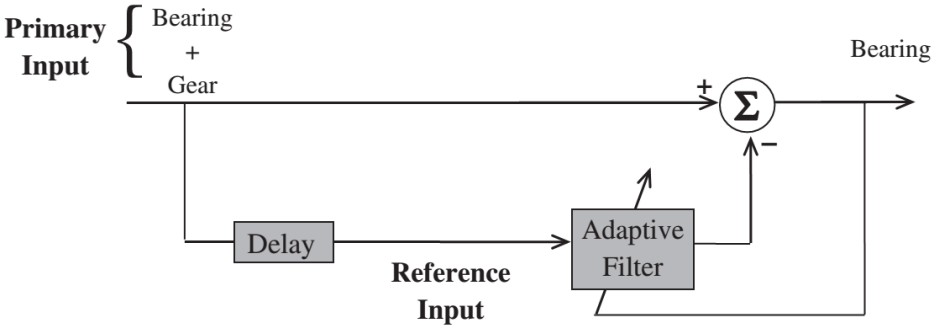

Fig. 2.17 Schematic diagram of SANC for separating bearing and gear signals [6]

diagram of SANC used for removing periodic interference in a gearbox is depicted in Figure 2.17.

Setting the filter parameters in advance is one of the major drawbacks of the SANC. Ref. [39] provides some guidelines for selecting the time delay and length of the filter parameters of the algorithm when applied to vibration signals. Very long filter length is another disadvantage of this method when SNR for a signal with several harmonics is high. This means that the algorithm may slowly converge or even fail if other parameters are not tuned properly [39].

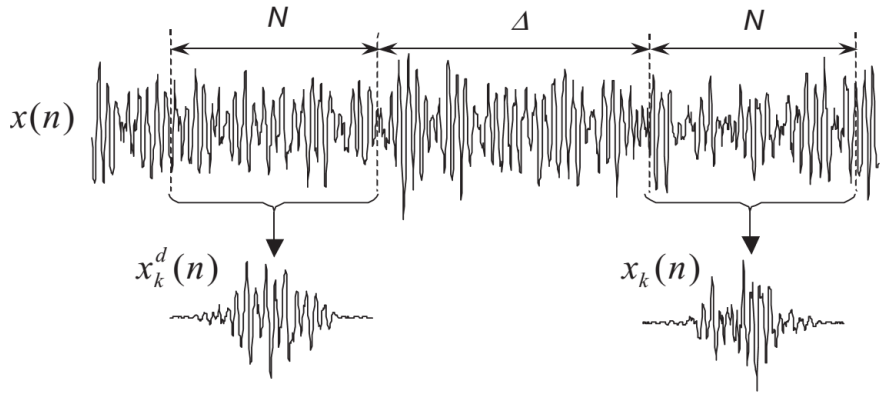

Fig. 2.18 Short-time sequences used in the DRS [40]

2.5.4 Discrete Random Separation (DRS)

This method is proposed by Ref. [40] and similar to the SANC, it finds the coherent relationship between the signal and a delayed version of itself but, as the solution is calculated in frequency domain, is faster and simpler than adaptive algorithms.

In this method, the data are divided in different blocks and then the average of their frequency representations is computed to estimate the optimal Wiener frequency response (see Figure 2.18). This approach is more beneficial in comparison to the SANC because it takes advantage of efficient fast Fourier transformation (FFT) algorithm. Also, the optimal solution is computed in one run.

The optimal frequency response function $H(f)$ between a signal and its delayed version, is given by [40]

$$H(f)=\frac{S_{x_k^d\,x_k}(f)}{S_{x_k^d\,x_k^d}(f)}\approx\frac{\sum_{k=1}^{M}X_{k,c}(f)\,X_{k,c}^{d}(f)}{\sum_{k=1}^{M}\left|X_{k,c}(f)\right|^2}$$

where S is the cross-power spectrum, $X(f)$ is the Fourier transform of $x(n)$, M is the number of sequences, $*$ denotates the complex conjugate and superscripts d and c refers to delayed signal by factor Δ and c-long ($c\geq N$ due to possible zero-padding) discrete Fourier transforms respectively.

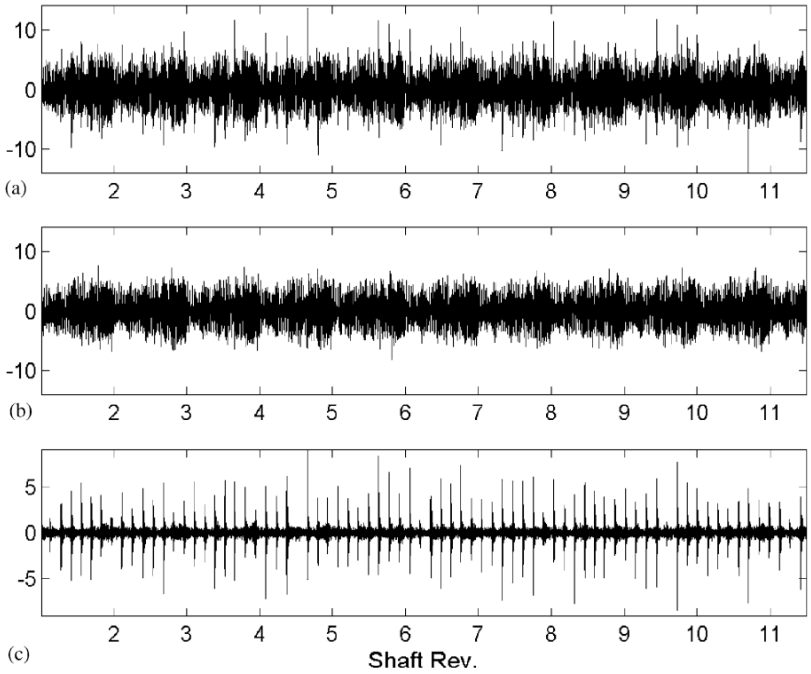

Fig. 2.19 Example of a faulty bearing in a gearbox (a) measured vibration signal (b) extracted periodic part (c) extracted non-deterministic part [40]

An example of the DRS filter performance is shown in Figure 2.19 where a real case vibration signal of a helicopter gearbox with a defected bearing is divided to its deterministic (gear signal, Figure 2.19b) and random (bearing signal, Figure 2.19c) parts.

It should be noted that in contrast to SANC which can cope with some speed variation, DRS requires order tracking to suppress speed variation and makes the signal deterministic part perfectly periodic.

2.5.5 Cepstrum editing and pre-whitening

Cepstral analysis has been investigated in subsection 2.4.5. Randall and Swahili [41] employed cepstrum editing procedure to separate periodic and random components in a signal. The method is shown schematically in Figure 2.20. This is based on editing the real cepstrum which is calculated in Equation 2.13. First the selected discrete quefrency components, corresponding to the families of harmonics and sidebands, in the real cepstrum are removed and then the time signal of the random

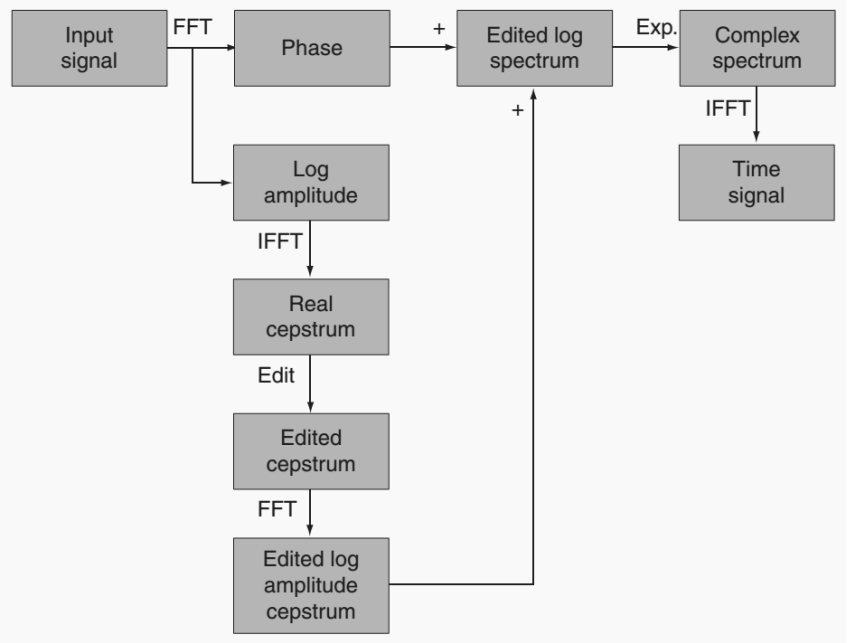

Fig. 2.20 Schematic diagram of cepstrum editing [41]

residual is generated by combining the edited log amplitude cepstrum with the original phase spectrum. This process does not necessary need order tracking, for limited speed variation, and can remove selected discrete frequency components just in one operation while desired periodic components can also be preserved.

A simpler approach is achieved just by dividing the Fourier transformation of a signal by its norm and then transforming back to the time domain:

$$x_{cpw} = \mathrm{IFT}\left\{\frac{\mathrm{FT}(x)}{|\mathrm{FT}(x)|}\right\} \tag{2.14}$$

where x_{cpw} is the pre-whitened signal. This operation, which is called cepstrum pre-whitening (CPW) by Ref. [42], is equivalent to setting to zero the values of the entire real cepstrum. Although, this technique has been well-known for long time as the phase-only reconstruction technique [43], Borghesani et al. [42] were the first that used this approach for REBs diagnosis. They also showed that using CPW after performing TSA is beneficial in cases with variable speeds as the application of TSA is not able to completely remove the effect of deterministic components, such as gear coupling and misalignments. Since this approach is fast, easy to implement and can be performed without any additional input parameters, it is suited for practical

applications. Although, removing all the desired periodic components and reduction of SNR are the major drawbacks of CPW.

Peeters et al. [44] proposed automated cepstrum editing procedure (ACEP) in which certain peaks in the real cepstrum are selected automatically and instead of the full real cepstrum only these peaks are set to zero. Moreover, they compare the performance of the ACEP and CPW methods and assess the potential advantages/disadvantages of each method.

2.6 Enhancement of bearing signals

Even after removing the discrete frequencies from other sources unrelated to bearings defects, by using the methods explained in section 2.5, the defect frequencies may not be achievable. For this reason, the bearing signal should be enhanced to increase its impulsivity and SNR. In what follows, some of the most well-known methods for this purpose will be outlined.

2.6.1 Minimum Entropy Deconvolution (MED)

Deconvolution approaches design a finite impulse response (FIR) filter to maximize a measure (norm) of interest in a signal. MED was first developed by Wiggins [45] to extract the detailed reflectivity information from amplitude anomalies in reflection seismic recordings. This method iteratively finds a filter which minimizes the entropy/maximizes the kurtosis of the filtered output signal. Although MED was successfully used for rotary machinery diagnosis [46–48] it suffers from some major drawbacks. First, MED results is not necessarily optimal. Second, the kurtosis value is maximized with only one spike. Third, it is an iterative method which may be time consuming.

An example of the MED technique performance is shown in Figure 2.21. The MED filter is applied to a time domain signal acquired from a gearbox (Figure 2.21a) with a faulty bearing, after removing the gear deterministic signal by utilizing an AR filter (Figure 2.21b). Consequently, the impulsiveness of the signal is enhanced as the kurtosis value is increased from 1.25 for the AR residual signal to 38.55 (Figure 2.21c).

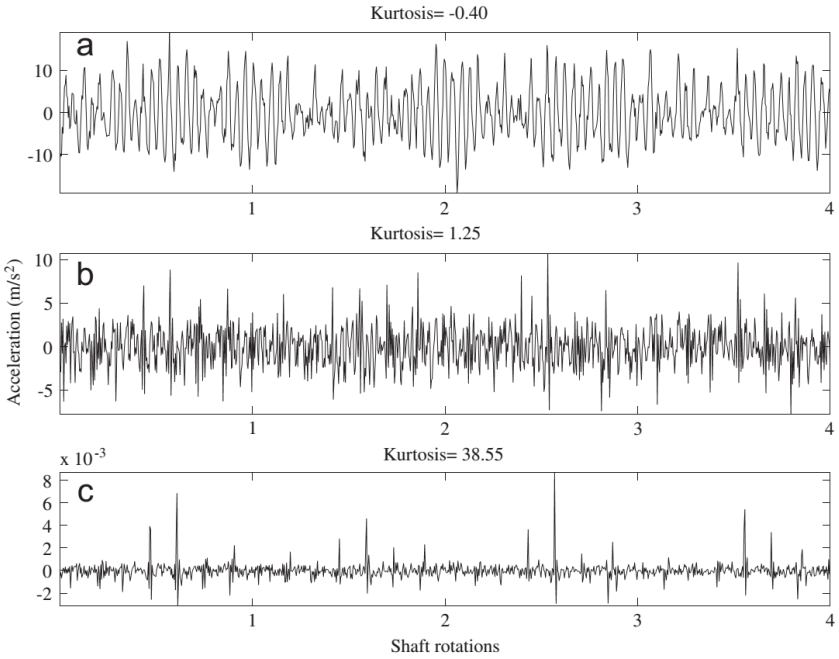

Fig. 2.21 (a) Raw signal (b) residual of the AR linear prediction filter (c) signal of Figure 2.21b filtered using MED [46]

McDonald et al. [49] proposed a new deconvolution formulation called Maximum Correlated Kurtosis Deconvolution (MCKD) to deconvolve periodic impulses separated by a known period T. The FIR filter is sought to maximize the newly defined Correlated Kurtosis (CK) instead of the kurtosis used in MED.

M-shifted CK of signal x with length L is defined as:

$$CK_M(T) = \frac{\sum_{n=1}^{N}(\prod_{m=0}^{M} x_{n-mT})^2}{(\sum_{n=1}^{N} x_n^2)^{M+1}} \tag{2.15}$$

where T (period of interest) should be set in advance.

Prior knowledge of the fault period and high computational cost are the two major drawbacks of the method.

Cabrelli [50] proposed an exact optimal solution for the inverse filter in MED by using a norm called the D-Norm (Equation 2.16) so there is no need for an iterative process. But it is not suitable for rotary machinery diagnosis because, as it was mentioned, the optimal solution is a single impulse rather than the desired periodic fault impulses.

$$\text{D-Norm} = D(\overrightarrow{x}) = \max_{k=1,\dots,N} \frac{|x_k|}{\|\overrightarrow{x}\|} \tag{2.16}$$

McDonald and Zhao [51] further proposed Multipoint Optimal Minimum Entropy Deconvolution (MOMED) for multi D-Norm objective function. Similar to Cabrelli [50] the result is optimal but based on their proposed objective function the periodicity of the impulses are considered too. Since the solution for the filter is calculated non-iteratively, the process is not time consuming but similar to MCKD prior knowledge of the fault period is required.

2.6.2 Narrowband amplitude demodulation techniques

Signals related to local damages usually contain the resonance frequencies of the structure. These frequencies are excited by the impacts which are generated once the defect is engaged. Therefore, this frequency band can be used to extract the signal associated to the defect and filter out irrelevant and extraneous components.

Many techniques have been proposed to automatically find the most suitable center frequency and the bandwidth for demodulation [20, 52–59]. Spectral Kurtosis, Protrugram, enhanced and optimized Kurtogram try to determine the frequency bands with maximum impulsivity by using kurtosis indicator while Infogram uses negentropy as the indicator. These well-known methods will be shortly reviewed here.

Furthermore, Ref. [60] studied some other statistics criteria such as Jarque–Bera, Kolmogorov–Smirnov, Cramer–von Mises, Anderson–Darling, quantile–quantile plot and a method based on the local maxima approach to investigate their abilities to select informative frequency band (IFB).

2.6.2.1 Spectral Kurtosis (SK) and Kurtogram

Kurtosis is a statistical indicator which measures the peakedness of a data set, therefore it can be used to detect faults impulsiveness in signals related to rotary machinery. Kurtosis is defined in Equation 2.6. The SK is an extension of kurtosis notation to frequency domain where a band is found to demodulate the signal in order to extract its impulsive and non-stationary component. It can be represented

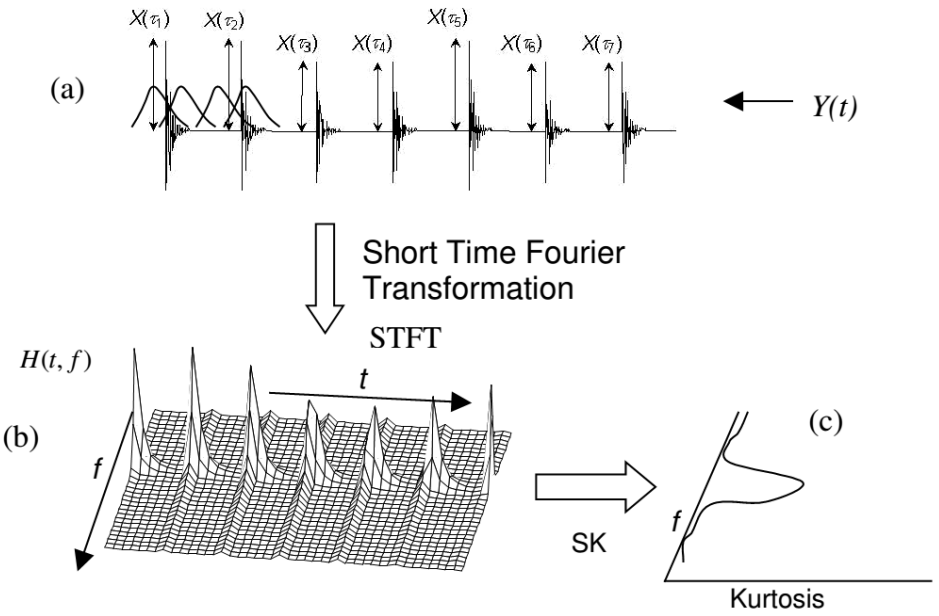

Fig. 2.22 Calculation of spectral kurtosis (SK) for a simulated bearing fault signal (a) time signal and moving time windows (b) amplitude of STFT (c) SK [54]

by the fourth-order normalized cumulant [53]:

$$K_x(f) = \frac{\left\langle |Y(t_i,f)|^4 \right\rangle}{\left\langle |Y(t_i,f)|^2 \right\rangle^2}, \tag{2.17}$$

where $\langle\rangle$ is the time-averaging operator. $Y(t_i,f)$ is the short-time Fourier transform (STFT) of a signal $x(t_k)$ obtained at time $t_i = i/f_s$,f_s is the sampling frequency, by moving a constant length window (N_w) along the signal. Therefore, SK is a function of frequency and STFT window length (frequency resolution). The SK for a rolling element bearing signal modelled as a periodic series of impulse responses is depicted in Figure 2.22.

The principle of SK is similar to power spectral density (PSD) except for a time signal it represents fourth-order statistics (kurtosis) vs. frequency instead of second order statistics (power) used in PSD. Moreover, in contrast to PSD, SK strongly depends on the window length (frequency resolution) of STFT. For different length of windows PSD (Welch estimation) and SK of a signal from a rolling element bearing with a localized fault are shown in Figure 2.23.

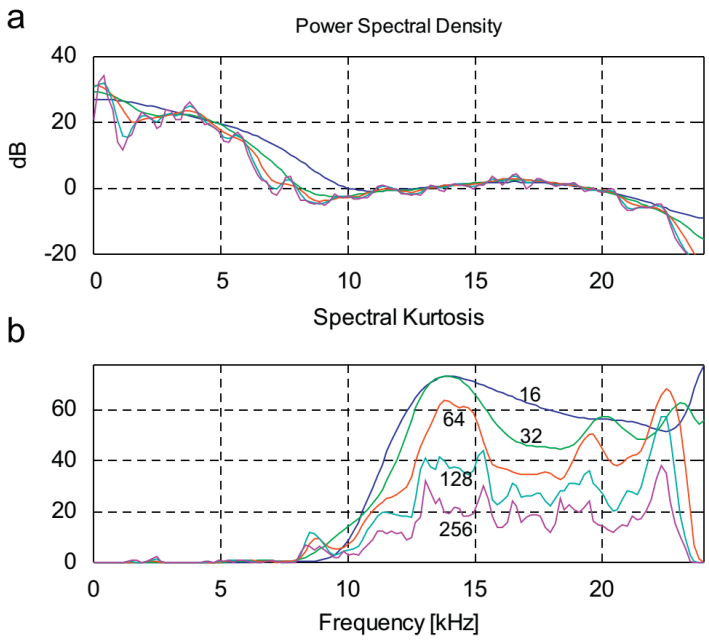

Fig. 2.23 (a) PSD and (b) SK computed for different frequency resolution/window length (N_w = 16; 32; 64; 128; 256) [54]

Full Kurtogram [53] was introduced as the representation of SK for all possible combinations of frequency and N_w. The tree-dimensional plot which is actually a cascade of SK for various window length is shown in Figure 2.24. Consequently, the optimal parameters, window length and frequency band, which maximize the SK value can be selected to design the band-pass filter.

An example of a real case bearing signal [54] can be seen in Figure 2.25 in which the optimal band-pass filter (Figure 2.25b) as obtained from the maximum of the Kurtogram in Figure 2.24 is utilized to filter the raw signal for the bearing with defected outer race shown in Figure 2.25a. The outcome of the filtering process is illustrated in Figure 2.25c where it can be seen that the kurtosis value of the filtered signal has significantly increased as compared to the raw signal.

This process is very time consuming and it is not suitable for online industrial applications. Fast Kurtogram (FK) [55] was proposed to overcome this difficulty. It subdivides the bandwidths into rational ratios that allow the use of fast multirate processing, and then investigates the kurtosis value of the complex filtered signal for each selected bandwidth. The schematic of this procure is shown in Figure 2.26 in which low-pass/high-pass decomposition is iterated in a pyramidal manner to

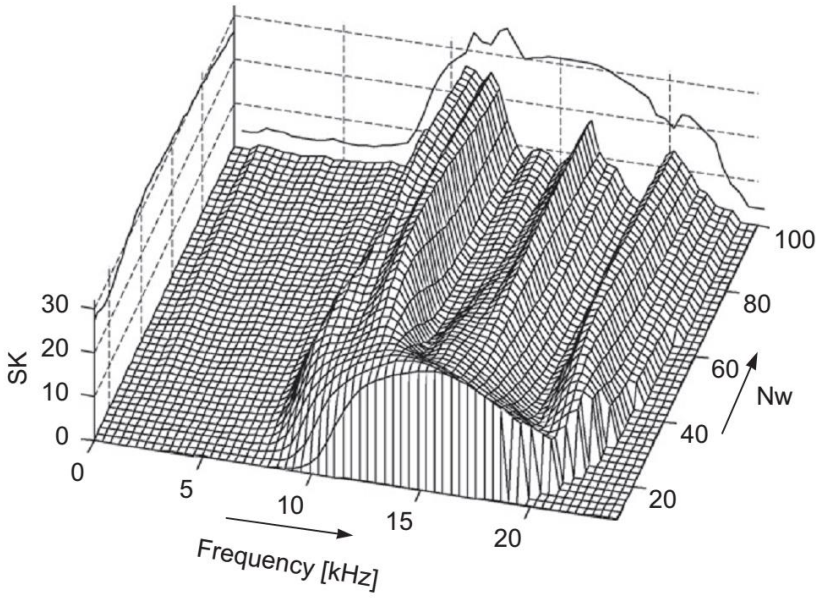

Fig. 2.24 Kurtogram of a rolling element bearing signal with an outer race fault [54]

produce a tree of filter-banks (Figure 2.26b). Finally, Kurtogram is estimated by computing kurtosis of all sequences shown in Figure 2.26b.

Dyadic (Figure 2.27a) and 1/3-binary (Figure 2.27b) tree structures are frequently used for designing these band-pass filters. At the kth level, the bandwidth for each node is equal to $\frac{f_s}{2^{k+1}}$, i.e. the halved-band is divided into 2^k bands. For Dyadic and 1/3-binary tree the value of k is equal to $k = 0,1,2,3,\ldots$ and $k = 0,1,1.6,2,2.6,\ldots$ respectively. SK is optimal in comparison to Fast Kurtogram due to its finer resolution, however two methods generate almost similar results.

2.6.2.2 Protrugram, Enhanced and Optimized Kurtogram

Fast Kurtogram has been successfully used in many cases [6], including in presence of strong Gaussian noise, but it has been found to be ineffective in several conditions, e.g. in presence of relatively strong, non-Gaussian noise containing large peaks, high repetition rate of fault impulses and low signal-to-noise ratio. To overcome the above-mentioned drawbacks a novel method, called Protrugram, for optimal band selection to demodulate vibration signals was proposed by Barszcz and Jabonski [56]. There are two main differences between Protrugram and Fast Kurtogram. First, in contrast to Fast Kurtogram that tries to find simultaneously the optimal bandwidth

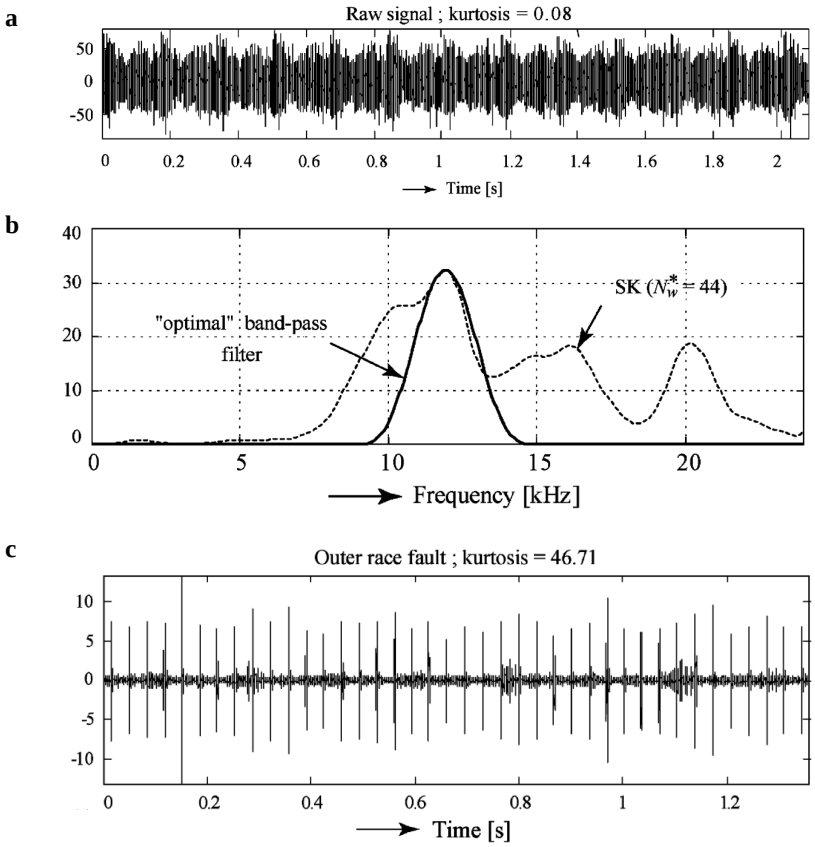

Fig. 2.25 (a) Raw signal for a bearing with defected outer race (b) Optimal band-pass filters (thick line) as obtained from the maximum of the Kurtogram Figure 2.24 (c) Filtered signals with the designed optimal band-pass filters [54]

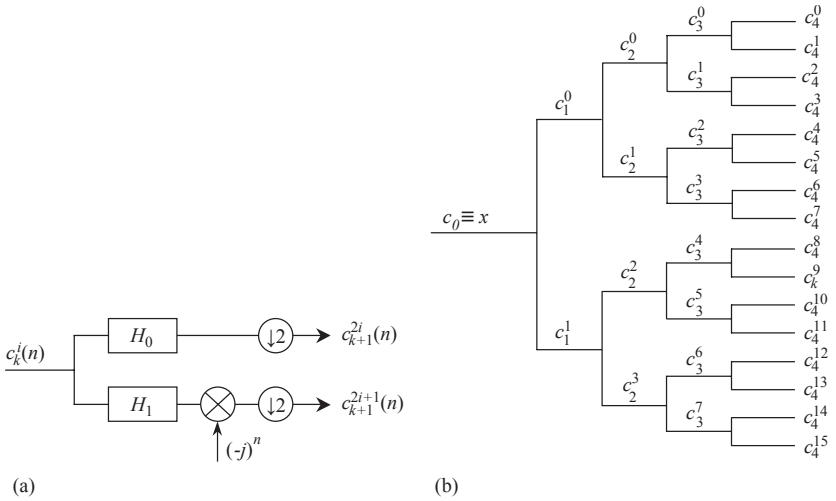

Fig. 2.26 (a) Low-pass and high-pass decomposition (b) Fast computation of Kurtogram by means of an arborescent filterbank structure [55]

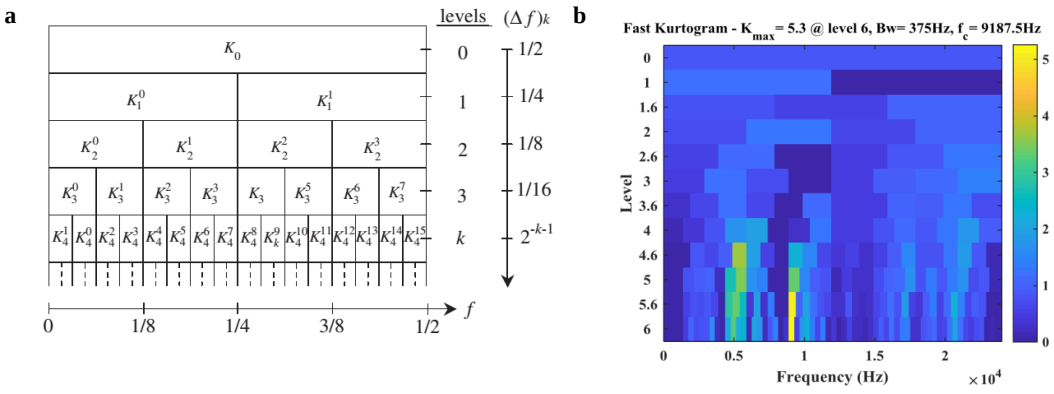

Fig. 2.27 Fast Kurtogram with (a) Dyadic [6] (b) 1/3-binary tree structure

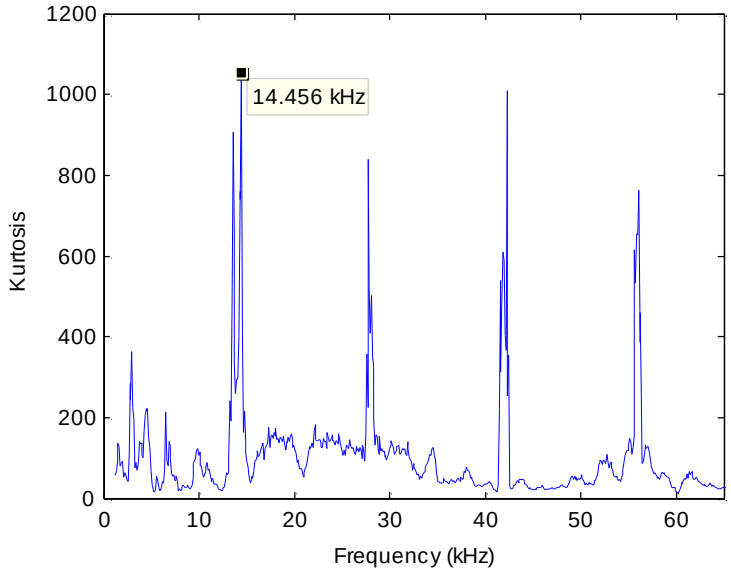

Fig. 2.28 Protrugram result for a real case bearing signal [61]

and center frequency for demodulation, in this method the former is set in advance based on the mechanical characteristics of the system under investigation and then the optimal center frequency is sought. They proposed to select a short frequency band to reduce the negative effect of the extraneous signals but long enough to include 1 to 5 harmonics of the bearing characteristic frequencies. Also, as prior knowledge about the system is required this method is not blind. Second, Protrugram computes kurtosis of the envelope spectrum of the demodulated signal rather than the filtered time signal which makes the method less vulnerable to non-Gaussian noise, also more sensitive to high repetition rate of impulses. After setting the constant frequency band, Protrugram tries to find the center frequency. For each step the center frequency is shifted by a predefined value, seeking for the highest kurtosis for the envelope spectrum.

An example of Protrugram for a real case bearing signal is depicted in Figure 2.28. The value of kurtosis gradually reaches its local maximum as the center frequency is shifted and get closer to the bearing defect vibration frequency, afterwards it decreases.

Ref. [57] proposed enhanced Kurtogram by employing the fault indicator used in Protrugram, i.e. similar to Kurtogram the optimal frequency band and bandwidth

are sought. Moreover, the filter bank was replaced by wavelet packet transform due to its effectiveness in extracting the transient impulsive signals [22].

Ref. [61] borrowed the idea of fix bandwidth from Protrugram and proposed optimized Kurtogram to overcome the negative effect of impulsive wideband interference in cases of electromagnetic interference (EMI) from a variable frequency drive (VFD). Similar to Protrugram it is not blind since certain knowledge of the system is needed.

2.6.2.3 Infogram

In the field of thermodynamics, disorder in a system is measured by entropy. Also, transients are considered as departures from a state of equilibrium. Motivated by this idea Ref. [58] proposed Infogram. In this method, kurtosis as an indicator of defects repetitive transients in Kurtogram and enhanced Kurtogram is substituted by negentropy - a contraction for negative entropy.

By employing this statistical indicator, spectral negentropy in time and frequency domains are calculated respectively as follows:

$$\Delta I_{\varepsilon}(f;\Delta f) = \left\langle \frac{\varepsilon_x(n;f,\Delta f)^2}{\left\langle \varepsilon_x(n;f,\Delta f)^2 \right\rangle} \ln\left(\frac{\varepsilon_x(n;f,\Delta f)^2}{\varepsilon_x(n;f,\Delta f)^2} \right) \right\rangle \tag{2.18}$$

$$\Delta I_{E}(f;\Delta f) = \left\langle \frac{|E_x(\alpha;f,\Delta f)|^2}{\left\langle |E_x(\alpha;f,\Delta f)|^2 \right\rangle} \ln\left(\frac{|E_x(\alpha;f,\Delta f)|^2}{\left\langle |E_x(\alpha;f,\Delta f)|^2 \right\rangle} \right) \right\rangle \tag{2.19}$$

where $\langle\rangle$ is time-averaging operator. ε_x and E_x are squared envelope (SE) and squared envelope spectrum (SES) respectively. Moreover, x_n is a discrete-time signal for $n = 0, \ldots, L-1$, L is the signal's length and $x(n;f,\Delta f)$ is defined as complex envelope in a frequency band $[f - \Delta f/2 \;\; f + \Delta f/2]$.

Two resultant colormaps, called SE Infogram and SES Infogram are respectively produced. Flowchart of the algorithm for computing Infogram is illustrated in Figure 2.29.

Ref. [58] argues that when signals are corrupted with impulsive contents or in cases with high repetition rate of transients it is advantageous to jointly consider the Infograms. Moreover, as SE and SES Infograms units ("nat", a unit of information) are the same it is possible to combine them together. Although, the averaging

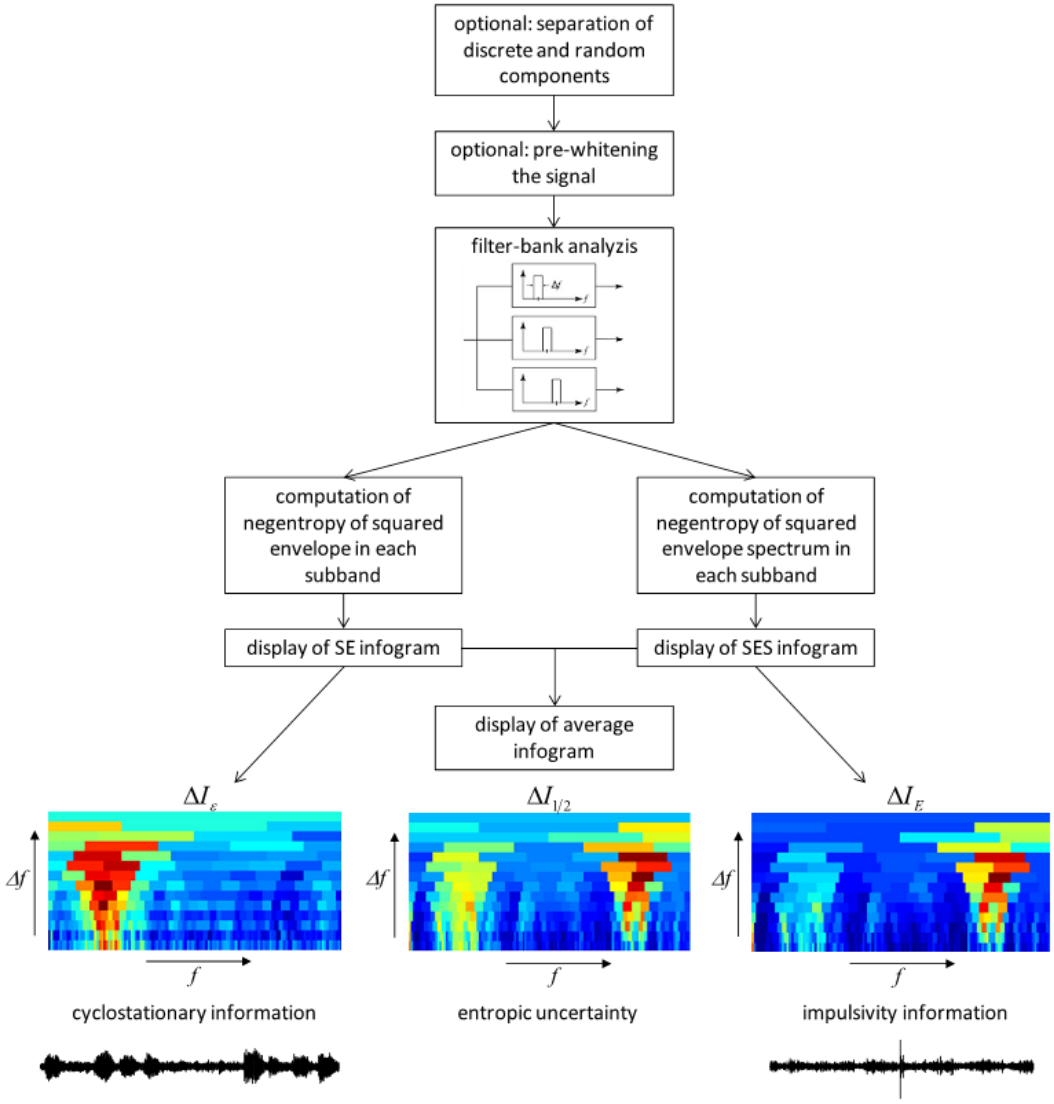

Fig. 2.29 Flowchart of the algorithm for computing Infogram [58]

combination of the Infograms which is called by author “average Infogram” has not been found useful.

Inspired by the Infogram idea and as an attempt to find a more beneficial combination algorithm, Ref. [62] proposed a multiscale clustering grey Infogram (MCGI) which combines both negentropies in a grey fashion using multiscale clustering.

2.7 Conclusion

In this chapter the most important and widespread techniques for diagnosis of rolling element bearings were investigated. In section 2.2 the bearing characteristics frequencies were discussed. Once identification of localized defects is required, computing theses frequencies is the first step toward a successful diagnosis. Mathematical models which try to generate signals that mimic the defected bearings signals were discussed in section 2.3. Although, these models help to have better understanding of the features of signals they are not comprehensive, especially when sources of modulation are considered.

Different methods for diagnosis of bearings by using vibration signal were discussed in section 2.4. Time domain and frequency domain approaches analyze data statistically, so it is necessary to have a reference signal from a healthy situation to be compared with the damaged condition. Therefore, these techniques are more suitable in combination with machine learning algorithms. In such methods, first features are extracted from time and/or frequency domains then a model is trained. Afterwards, it is used for classification and detection of anomaly in signals. In contrast, for time-frequency and cyclostationary analysis, a signal from undamaged condition is not needed and presence of the bearing characteristic frequencies are considered as the signature of faults. The effectivity of time-frequency analysis might deteriorate as the level of slippage increases and additional problems may occur with high level of noise. Under this circumstance, the bearing defect frequencies that have lower amplitude than resonance frequency of the structure may completely be masked, therefore they cannot be detected.

Cyclostationary analysis has been used for long time and it is the mostly used and effective method for diagnostic of bearings. The aim is to extract periodicity of statistical quantity of a process, i.e. envelope analysis examines the periodicity of

a signal envelope. Cyclic Spectral Analysis (CSA) provides a different family of bearings diagnosis techniques which represent a signal in a bi-spectral map, with carrier and modulation frequencies. It is not as straight forward as envelope analysis as different approaches have been developed for its estimation.

For envelope analysis, it is necessary to find the frequency band in which the structure is excited by the impacts generated by the defects. This technique band-pass filters the signal for this frequency band to remove mechanical noise and interference from other sources. This process enhances the SNR ratio for the bearing signal and increase the chance of detecting the defects. The main challenge has always been finding the most suitable frequency band for demodulation.

cepstral analysis tries to highlight the signature of defects in magnitude spectrum by combining a defect frequency and its harmonics into a single cepstral line. Consequently, this approach is not beneficial for cyclostationary signals with relatively higher slippage rate and/or lower SNR which may not produce any detectable discrete frequency components, as the peaks might be entirely masked by the noise because of their weak energy level [11].

Different techniques used for separation of discrete and random components and pre-whitening are presented in section 2.5. SANC method tries to find the transfer function between the signal and its delayed version. As it is achieved adaptively by minimizing the power of residual signal, the process is time consuming. DRS borrowed the idea of seeking the relationship between the signal and its delayed version but the method is implemented in the frequency domain. In addition, the optimal solution is calculated in one step, so it is faster than SANC. In general, both methods are able to remove all deterministic components. For situation with small speed variation, SANC is more robust than DRS. Time Synchronous Averaging (TSA) is achieved by averaging the data segments for a specific period.

Among these three methods for separation of deterministic and random components, TSA provides the best results with minimum disturbance of the signal, but it has three drawbacks. First, prior knowledge of a system periods is essential, and the deterministic parts must be removed for each period separately. Second, order tracking is necessary to remove the effect of speed fluctuation which means extra device for collecting the rotation speed is needed. Third, modulation sidebands removable by SANC and DRS cannot be eliminated. AR and cepstrum pre-whitening

are effective pre-whitening methods. Pre-whitening does not achieve a separation, but it only reduces the effects of the deterministic components.

In section 2.6 the most effective and widespread approaches for enhancement of bearing signals are reviewed. Fast Kurtogram has become the benchmark method and it has been a significant step to unravel this problem. It is a method which effectively detects the sequence of impulses in a signal and can be used to determine the proper demodulation frequency band in which a signal has the maximum impulsivity.

Enhanced Kurtogram is more robust than Fast Kurtogram as it is less vulnerable to non-Gaussian noise such as large impulses and low SNR. But its effectiveness is limited to synthetic signal where the signal is only combination of bearing signal and noise. This does not often happen in real cases as signal from other sources interferes with bearing signals.

The idea behind Protrugram and enhanced Kurtogram is similar. The only difference is related to the bandwidth of the filter which is constant for Protrugram and variable for enhanced Kurtogram. As this constant frequency band for filtering process is chosen based on the bearing characteristics, the method is more suited for industrial application than enhanced Kurtogram. However, the method is not blind and prior knowledge about the defect frequencies is needed. This drawback vastly reduced its usage in practical applications.

Nonetheless, both Protrugram and enhanced Kurtogram may fail when defect frequencies are not dominant in the spectrum in comparison to discrete components generated by other sources such as shaft frequency. The weak performance of Protrugram in cases with intense EMI is also shown in Ref. [61].

Infogram was an attempt to connect the concepts of Kurtogram and Protrugram to avoid the drawbacks of these methods and produce a more robust approach and capture the signature of repetitive transients in both domains. However, it has not received as much attention as Fast Kurtogram possibly because the two colormap generated by Infogram were not so different from Fast Kurtogram and enhanced Kurtogram. Second, in practice the joint consideration of the Infograms did not provide promising results. This even made the interpretation of the resultant colormap more difficult.

Overall, based on the literature reviewed the three following approaches are the mostly used techniques for rolling element bearings:

1. Fast Kurtogram to find the demodulation frequency band and then calculation of the envelope spectrum
2. Spectral correlation
3. Cepstrum pre-whitening and then calculation of the envelope spectrum

In the following chapters of this part, this thesis mainly focusses on methods 1 and 3. As was explained, method 1, despite its potential and vastly usage, suffers from some major drawbacks. The aim is to introduce a new method which is an improved version of Kurtogram and benefits from advantages of other methods such as enhance Kurtogram and Fast Kurtogram. Also, another objective is to present a comprehensive and general method to be suitable for industrial application.

In contrast to method 1 in which first the signal is band-pass filtered and then the envelope spectrum is computed, spectral correlation immediately returns the envelope spectrum in all frequency bands. It might raise the question why an analysis by spectral correlation (method 2) should not be preferred to method 1. In the following chapter, there will be an attempt to answer this concern too.

Furthermore, method 3, despite its simplicity, provides valuable diagnostic information. The idea of manipulating the spectrum magnitude has a great potential which has not yet received the attention it deserves. Therefore, it would be appropriate to examine this idea more deeply which may open a new perspective for rolling element diagnosis.

Chapter 3

The Autogram: A novel method for selecting the optimal demodulation band [1]

3.1 Introduction

This chapter is organized as follows. Theoretical background is discussed in section 3.2. Section 3.3 establishes a new method for optimal demodulation band selection for non-stationary signals containing repetitive transients, e.g. bearing with localized defects. A comprehensive explanation of the proposed method is presented in section 3.4. In addition, experimental validation is carried out and the outcome is compared with the FK [64], Fast-SC [29] and literature results [65] to examine the performance of the new proposed method.

[1] "Part of the work described in this chapter has been previously published in Ref. [63]"

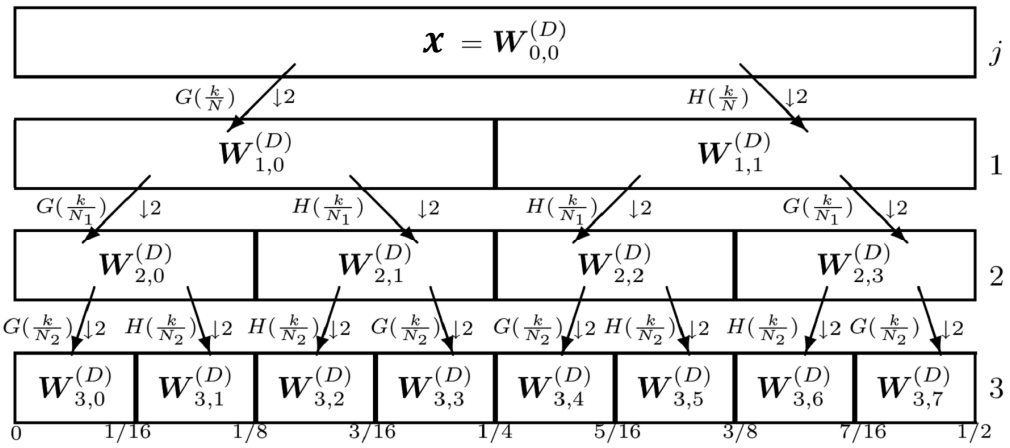

Fig. 3.1 Wavelet packet table showing the DWPT of x [66]

3.2 Theoretical Background

3.2.1 Maximal Overlap (Undecimated) Discrete Wavelet Packet Transform [66, 67]

In this section we first discuss the concept behind discrete wavelet packet transform (DWPT) which is a generalization of pyramid algorithm for the discrete wavelet transform (DWT). In contrast to DWT, DWPT provides subband filtering of an input signal into progressively better equal width frequency intervals. The bandpass accuracy depends on frequency-localization of the filters. At each decomposition level j the frequency axis $[0,\ F_s/2]$ is partitioned into 2^j equal subbands which are called nodes, where F_s is the Nyquist frequency.

Column vector x is assumed to be a time series data with length N, $\{x_i : i = 0, \ldots, N-1\}$. At the first step of the DWPT algorithm, to calculate the sequence of coefficients at the first level $(j = 1)$, the original signal $W_{0,0}^{(D)} \equiv x$ is circularly filtered by the scaling (low-pass) $\{g_l : l = 0,...,L-1\}$ and wavelet (high-pass) $\{h_l : l = 0,...,L-1\}$ filters respectively. Then the resulting time series are downsampled by factor 2, $W_{1,0}^{(D)} = \{W_{1,0,t}^{D}, t = 0, \ldots, (N/2) - 1\}$ and $W_{1,1}^{D} = \{W_{1,1,t}^{(D)}, t = 0, \ldots, (N/2) - 1\}$.

Filtering of coefficients can be repeated for each level as shown in Figure 3.1. The $(j-1)$th level coefficients can be further decomposed by the circular filter with

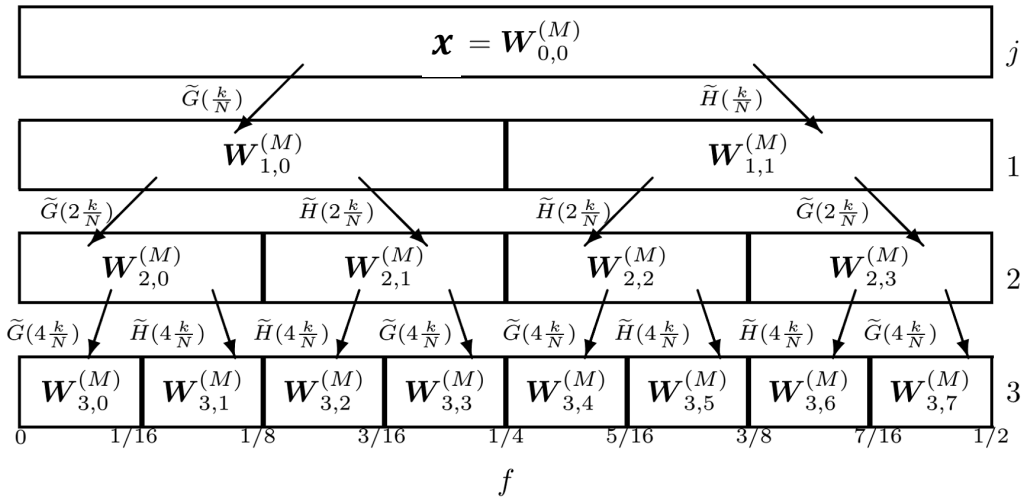

Fig. 3.2 Undecimated wavelet packet table showing the MODWPT of x [66]

discrete Fourier transform $\{H(\frac{k}{N_j-1})\}$ or $\{G(\frac{k}{N_j-1})\}$, $N_j{=}\frac{N}{2^j}$, and then downsampled to attain the jth level coefficients. At each level the number of filtered sequences is increased by a factor of 2 and length of the coefficients are half of the prior level as a result of downsampling. $\{W_{j,n,t}^{(D)}\}$ which are the DWPT coefficients for levels j are calculated by using downsampled wavelet packet coefficients of the previous step as follows:

$$W_{j,n,t}^{(D)} \equiv \sum_{l=0}^{L-1} \left(r_{n,l}\, W_{j-1,\lfloor n/2 \rfloor,(2t+1-l) \bmod N_{j-1}}^{(D)} \right), \qquad t = 0, \ldots, N_j - 1$$

where

$$r_{n,l} = \begin{cases} g_l, & \text{if } n \bmod 4 = 0 \text{ or } 3 \\ h_l, & \text{if } n \bmod 4 = 1 \text{ or } 2 \end{cases}$$

$\lfloor \bullet \rfloor$ denotes 'the integer part' operator. This DWPT produces a sequency-ordered wavelet packet tree. Level j and band n is nominally associated with frequencies in the interval $(\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$ where $n = 0, 1, ..., 2^j - 1$.

As a result of downsampling operation, in higher levels the lengths of the coefficient series are much shorter than the original data which will cause larger estimation error and limit the possibility to investigate the coefficients. Furthermore, DWPT can be sensitive to the selected starting point, as wavelet and scaling coefficients are not circularly shift-equivariant, i.e. a change in the starting point for a time series can produce rather different outcomes. These drawbacks can be overcome by

using maximal overlap (undecimated) discrete wavelet packet transform (MODWPT) which removes the downsampling step in DWPT.

MODWPT filters $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$ are defined as the rescaled terms of their DWPT counterparts, namely $\tilde{g}_l = g_l/\sqrt{2}$ and $\tilde{h}_l = h_l/\sqrt{2}$. For levels j of the transform $2^{j-1}-1$ zeros, $j \geq 1$, are inserted between each of the L elements of the filters $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$. Therefore, as $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$ have transfer functions of $\tilde{G}(f) = \sum_{l=0}^{L-1} \tilde{g}_l\, e^{-i2\pi fl}$ and $\tilde{H}(f) = \sum_{l=0}^{L-1} \tilde{h}_l\, e^{-i2\pi fl}$ the upsampled filters would have transfer functions given by $\tilde{G}\left(2^{j-1}f\right)$ and $\tilde{H}\left(2^{j-1}f\right)$ respectively.

By using these filters, level jth coefficients $\{W^{(M)}_{j,\ n,\ t}\}$ can be defined in a recursive scheme for which it is assumed that level $(j-1)$th coefficients are already known $\{W^{(M)}_{j-1,\ \lfloor n/2\rfloor\ ,\ t}\}$ (the process initiates with the time series itself, $W^{(M)}_{0,0} = x$, for level $j = 0$):

$$W^{(M)}_{j,n,t} \equiv \sum_{l=0}^{L-1} \left(r_{n,l}\, W^{(M)}_{j-1,\lfloor n/2\rfloor\ ,\left(t-2^{j-1}l\right) \bmod N} \right), \qquad t = 0,\ldots,\ N-1$$

where

$$r_{n,l} = \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3 \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2 \end{cases}$$

Since the results are not downsampled, all the coefficients have the same length (N) as the original time data. The diagram illustrating the transformation of signal x into MODWPT coefficients $W_{j,n}$ for levels $j = 1,2,3$ is shown in Figure 3.2. The main differences between MODWPT and DWPT are renormalization of filters and elimination of downsampling. Since essentially the same filters are used to generate DWPT and MODWPT coefficients, MODWPT coefficients are nominally related to the same frequency interval as DWPT coefficients, $(\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$ where $n = 0,1,...,2^j-1$.

3.3 Proposed Method

3.3.1 Autogram

The usage of FK to determine the most impulsive frequency band, followed by envelope analysis of the bandpass filtered signal, has become the benchmark method

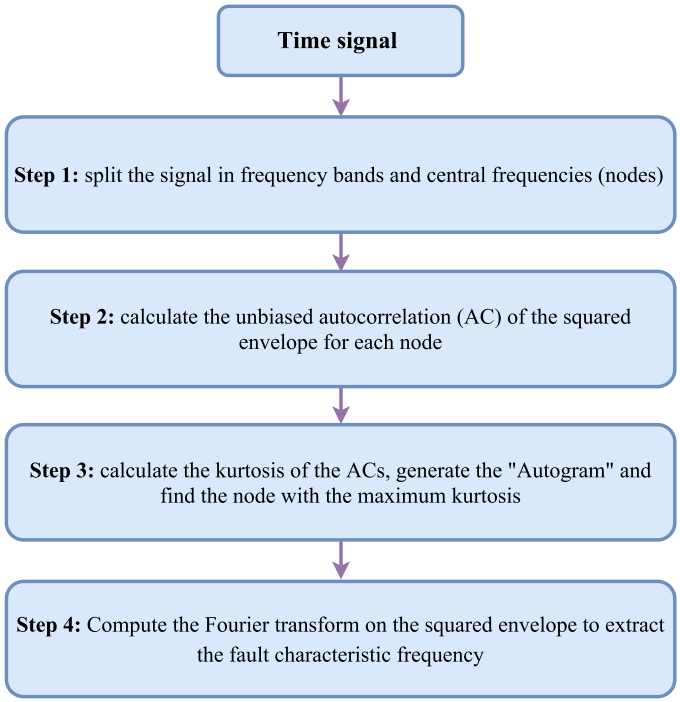

Fig. 3.3 Flowchart of the proposed method

for bearing diagnostics for years and has accomplished significant results [65]. FK is numerically very efficient and largely diffused so that also in the present it work has been selected as a reference method. Kurtogram is commonly capable of detecting localized hidden non-stationarities, even in presence of strong Gaussian noise, but its performance is limited in several conditions, i.e. low signal to noise ratio (SNR) or strong non-Gaussian noise such as randomly distributed impulses [65]. In these cases, FK was found to be ineffective in seeking the transient signal. These circumstances are common in industrial applications as multiple devices such as gearboxes and bearings work alongside in a complex machine. Also, the acquired signal in harsh environment can be extremely affected by external sources.

This chapter proposes a new procedure based on unbiased AutoCorrelation (AC) to overcome the restrictions imposed by heavy Gaussian and also non-Gaussian background noise. The proposed method is thought to be sufficiently general for detection of rotary machinery faults with impulsive signal, e.g. REBs and gears. Similar to FK, the new method is blind and no prior knowledge of signals is required.

The flowchart of the proposed method is shown in Figure 3.3 and the details of each step are described as follows.

Step 1: In this first step, time domain data are divided in frequency bands, according to a dyadic tree structure, by means of the wavelet transform (WT). Wavelets have very good local properties in both time and frequency domains, and WT can be used as an effective filter to split a signal in different frequency bands and central frequencies [22].

Basically, MODWPT is applied as a filter to the investigated time history and a series of signals is consequently produced at each level of decomposition. The filtered signals, each corresponding to a frequency band and central frequency (node), are the inputs for the following Step 2.

This division in frequency bands is similar to the decimated filterbanks used by FK [55], although MODWPT is computationally more expensive as it does not make use of multirate processing. Unquestionably, the procedure is not as fast as FK and a comparison of the required CPU (Central Processing Unit) times is presented in subsection 3.4.8. MODWPT has been chosen because, besides overcoming the DWPT drawbacks, it preserves full-time resolution [66] which is essential for the proposed method to calculate the AC. But other filters may be used, provided that the length of the resulting time history remains unchanged. It should be noted the comparison of MODWPT filter with other filters, e.g. STFT or the filternabks used in Ref. [55], is out of the scope of this research.

Step 2: The fundamental concept which motivates this work is to take advantage of periodicity of the autocovariance function which characterizes the 2nd order cyclostationarity of bearing vibration signals (subsection 2.3.1). Therefore, the unbiased AC of the (periodic) instantaneous autocovariance of the signal $R_{xx}(t_i, 0)$ is calculated, where x is the signal filtered by MODWPT at Step 1. It can be seen from Equation 2.5a that the instantaneous autocovariance is computed by the ensemble average operator. Unfortunately, in many situations the expected value cannot be obtained since only a single record of data rather than a set of records is available. Once cyclostationarity is assumed, other features of the signal, e.g. the envelope function, provide similar information on periodicity as the instantaneous autocovariance [68].

A simple example is provided by the white noise with periodical amplitude modulation, shown in Figure 3.4a. Figure 3.4b and Figure 3.4c (left columns)

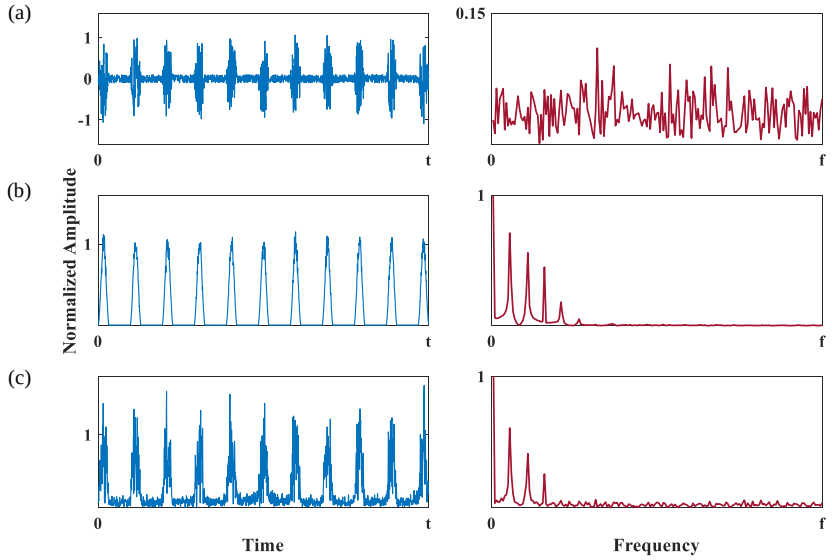

Fig. 3.4 Example of (a) amplitude modulated white noise and its spectrum (b) instantaneous autocovariance $R_{xx}(t,0)$ and its spectrum (c) envelope signal and its spectrum. Generic engineering units may be associated to the data

illustrate the instantaneous autocovariance $R_{xx}(t_i,0)$ and the envelope of the signal. The frequency representation of the signal (Figure 3.4a, right) does not provide any useful information. On the other hand, the modulation frequency can be clearly detected in the very similar Figure 3.4b and Figure 3.4c (right). Note that the ensemble average in Figure 3.4b is produced by using 200 realizations, which would be impossible in practical situations. on the contrary, the computation of envelope signal in Figure 3.4c only requires a single realization.

In this step, unbiased AC is computed on the squared envelope of the signal as follows:

$$\hat{R}_{XX}(\tau) = \frac{1}{N-q}\sum_{i=1}^{N-q} X(t_i)X(t_i+\tau) \tag{3.1}$$

where X is the squared envelope of the filtered signal, $\tau = q/f_s$ is the delay factor and $q = 0,\ldots,N-1$.

The AC has the benefit of removing the uncorrelated components of the signal, i.e. noise and random impulsive contents, both unrelated to any specific bearing fault. Furthermore, the periodic part of the signal (directly related to the defects) is enhanced, showing an additional virtue of this process. This is even more advan-

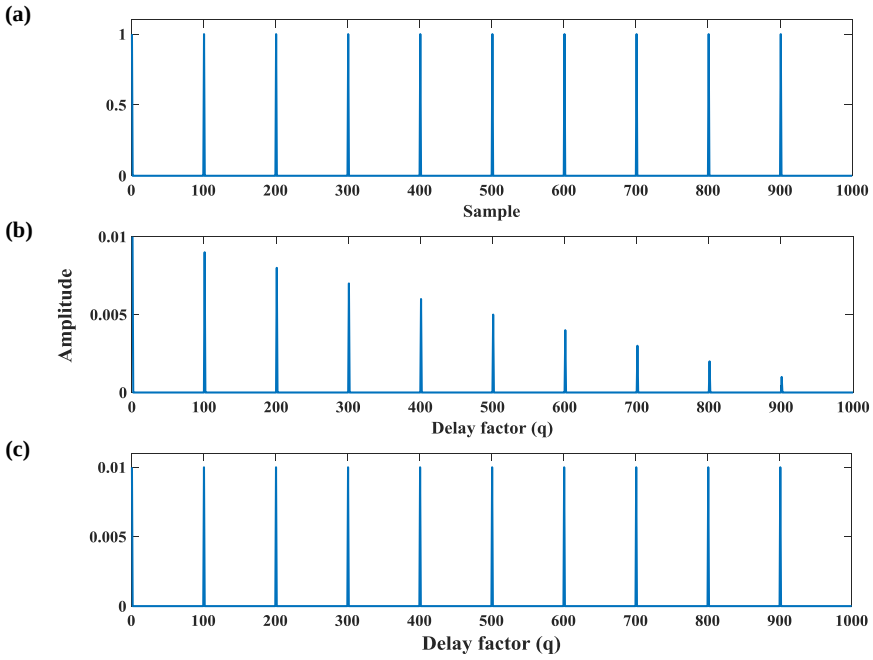

Fig. 3.5 Example of (a) a series of impulses spaced each 100 samples (b) biased auto-correlation (c) unbiased auto-correlation

tageous since it is done for each node separately rather than on the complete raw signal, so that SNR for each demodulated band signal is increased.

A simple example is provided here to show the reason behind selection of unbiased AC and its difference with biased one. Figure 3.5 shows 10 impulses spaced at every 100 samples. The biased and unbiased AC are depicted in Figure 3.5 b and c respectively. In Figure 3.5b the amplitude of impulses decreases for higher value of q (see Equation 3.1) but in Figure 3.5c the amplitude of impulses are constant regardless of q. This difference roots in normalization factor which is N (total length of the signal) and $N-q$ for these two cases.

In addition, with τ increasing, the number of data samples for computing the AC will decrease (see Equation 3.1), and therefore the resulting part will not have an adequate estimation variance. As a result, only a part of the computed AC is chosen for further investigation: unless stated otherwise, throughout this chapter the first half of the AC is selected. Moreover, the first coefficients of filtered signal, that are affected by the transients of filters, should not be included in calculation of the AC [55].

The output of this step is then the AC which leads to a more accurate diagnostic process than possible with the original outputs of MODWPT. In fact, impulsive noise, which ineffectively assigns very high kurtosis to a signal, can largely be removed.

Step 3: The objective of this step is to find the most suitable frequency band for demodulation. This is substantial to have a successful diagnosis of bearings faults, since fault information cannot be extracted from the demodulated signal if the appropriate frequency band and central frequency are not selected. In this regard, FK and Protrugram are two widely known methods which compute kurtosis of the filtered time signal and spectral lines of the envelope spectrum respectively. In this chapter, an alternative approach is introduced to achieve an optimal frequency band of demodulation. The proposed method differs from both mentioned techniques because the kurtosis is evaluated for the signals resulting from step 2, i.e. the unbiased AC of the squared envelope, for each level and frequency band (nodes). Subsequently, kurtosis values of all nodes, similar to FK, are presented in a colormap, whose color scale is proportional to kurtosis value and the vertical and horizontal axis represent level of MODWPT decomposition and frequency respectively. Since the concept is analogous to Kurtogram, and this proposal is based on autocorrelation, the name of "Autogram" is suggested for this newly developed approach.

To quantify the impulsivity of the AC for each node, three equations which are modified versions of the kurtosis (Equation 2.6) are proposed as follows:

$$\text{Kurtosis}(X) = \frac{\sum_{i=1}^{\frac{N}{2}} \left[\hat{R}_{XX}(i) - \min(\hat{R}_{XX}(\tau))\right]^4}{\left[\sum_{i=1}^{\frac{N}{2}} \left[\hat{R}_{XX}(i) - \min(\hat{R}_{XX}(\tau))\right]^2\right]^2}, \tag{3.2a}$$

$$\text{Kurtosis}_u(X) = \frac{\sum_{i=1}^{\frac{N}{2}} |\hat{R}_{XX}(i) - \overline{X}_T(i)|_+^{\,4}}{\left[\sum_{i=1}^{\frac{N}{2}} |\hat{R}_{XX}(i) - \overline{X}_T(i)|_+^{\,2}\right]^2}, \tag{3.2b}$$

$$\text{Kurtosis}_l(X) = \frac{\sum_{i=1}^{\frac{N}{2}} |\hat{R}_{XX}(i) - \overline{X}_T(i)|_-^{\,4}}{\left[\sum_{i=1}^{\frac{N}{2}} |\hat{R}_{XX}(i) - \overline{X}_T(i)|_-^{\,2}\right]^2}, \tag{3.2c}$$

where N is length of the original signal and operators $|\blacksquare|_+$ and $|\blacksquare|_-$ mean that only positive or negative values are accepted and values of other data points are set to zero. Also, $\overline{X}_T$ is the threshold level and it is defined as the moving mean value of

the AC:

$$\overline{X}_T(i) = \frac{1}{k} \sum_{j=i}^{i+k-1} \hat{R}_{XX}(j) \tag{3.3}$$

k is length of the windowed signal to be averaged, typically a very small fraction of the total number of samples N.

Equation 3.2a is analogous to the standard definition, except that the minimum is used instead of the mean value. The AC of the squared envelope is in fact a positive function, whose minimum value is arguably different from zero.

Equation 3.2 b and c have been suggested by empirical observations on the results of some damaged bearings. Basically, they should be used when high levels of impulsive noise or low SNR occur in recorded data. Examples will be presented in the following section.

Colormap presentation of the results based on the Equation 3.2 a-c will be called Autogram, Upper Autogram and Lower Autogram correspondingly. The advantages and characteristics of each proposed Autogram and the condition in which using Upper/Lower Autogram is beneficial will be discussed in detail within the first example of section 3.4.

Ultimately the signal associated with the node with the highest kurtosis is considered for further investigation.

Step 4: The Fourier transform is finally applied to the squared envelope of the signal associated to the node selected in Step 3. The fault characteristic frequencies are extracted and a diagnosis of the bearing is performed.

3.3.2 Lower/Upper threshold and Squared Envelope Spectrum

As a consequence of step 2 of the proposed method, the level of uncorrelated signals such as noise and random impulses is reduced and the distinction between two parts of the signal (noise and defect impulses) has become more clear. Periodic impulses - corresponding to defective bearing - are in fact enhanced at step 2 and can be separated from noise more effectively. This gives the opportunity to separate the two parts without losing any useful information for diagnosis of bearings, as the main interest is in repetition frequency of the peaks corresponding to bearings defects, which is conserved after this process. To achieve this objective, in this step

a thresholding procedure is introduced and performed on the resulting coefficient of step 2 (AC of the envelope signal), for the node with the highest kurtosis selected in step 3.

In this chapter, the non-constant $\overline{X}_T$, defined in Equation 3.3, is used as the threshold level. The Lower/Upper thresholding process extracts the significant information by setting to threshold level the AC values with higher/lower than the threshold level.

This process directly affects the quality of the frequency analysis as it controls which coefficients will be retained and which will be discarded. The benefit of this step is not merely limited to suppression of noise, since both lower and upper part of the squared envelope signal after performing AC has some unique and virtuous features which will be discussed with the help of the first example in the section 3.4.

3.3.3 Combined Squared Envelope Spectrum (CSES)

A major problem related to all methods which try to find the best frequency band and center frequency for demodulation, such as Autogram or Kurtogram, is that other nodes which may contain useful information are neglected. It is more problematic especially in cases with multiple defects on different bearings as they may have different carrier frequencies, or when a defect on a bearing excites different resonance frequencies. To overcome this difficulty, an approach is proposed to combine the outcomes of the all nodes with valuable diagnostic information. The steps of this procedure can be explained as follow:

1. For each level of decomposition, the nodes with valuable diagnostic information are selected. This is achieved by selecting the node with highest kurtosis in each level, in addition to the nodes with kurtosis larger than a certain threshold. In this chapter, the threshold for each level is half of the maximum kurtosis of that level.

2. For each level of decomposition, the squared envelope spectrum (SES) is computed for all nodes selected in previous step. These spectra are normalized within 0 and 1, then they are combined. Hereafter, this spectral quantity is

called "Combined Squared Envelope Spectrum" (CSES):

$$\mathrm{CSES}(level) = \frac{1}{n_{level}} \sum_{i=1}^{n_{level}} \mathrm{SES}(i, level) \tag{3.4}$$

where n_{level} is the number of the nodes selected in the previous step for each level.

3.4 Results and Discussion

The data sets provided by the Case Western Reserve University (CWRU) bearing data center [69] has become a standard reference for diagnosis of bearings. For example, a detailed benchmark study has been provided by Smith and Randall [65] in which three diagnostic methods such as envelope analysis of the raw signal, cepstrum pre-whitening and discrete/random separation (DRS) followed by SK and bandpass filtering were applied to all the data sets. Therefore, these data will be used to examine the performance of the proposed method. The results will be compared with the benchmark study, FK and Fast-SC, whose codes have been provided by [64, 29]. The comparison makes it possible to properly evaluate the proposed method.

The benchmark study characterized the result of its diagnosis in the same six categories employed throughout the following section - Table 3.1.

The basic experimental setup is shown in Figure 3.6. The test rig includes a 2 hp electric motor, a dynamometer which provides the load and a torque transducer/encoder. Two test bearings, namely drive-end and fan-end bearings, support the motor shaft and a single localized fault is introduced on inner race, outer race or rolling element. The vibration data were acquired for about 10 seconds with sampling frequencies of 12 and 48 kHz for each case. The data are labeled by DE, FE and BA which indicate that the signals are acquired by accelerometers mounted on the housing of the drive-end (DE) bearing, fan-end (FE) bearing and the motor supporting base plate (BA). More details on bearings and faults specification, can be found in Refs. [69, 65].

Note that signal pre-whitening and/or separation of deterministic and random signals is usually conducted on the raw signal before the diagnosis. The aim is to reduce the effects of the deterministic components (discrete frequency “noise”) such as gears signals and therefore increase the bearings SNR. To this purpose Refs.

Table 3.1 Categorisation of diagnosis outcomes Diagnosis [65]

Diagnosis category	Diagnosis success	Explanation
Y1	Yes	Data clearly diagnosable and showing classic characteristics for the given bearing fault in both the time and frequency domains
Y2	Yes	Data clearly diagnosable but showing non-classic characteristics in either or both of the time and frequency domains
P1	Partial	Data probably diagnosable; e.g., envelope spectrum shows discrete components at the expected fault frequencies but they are not dominant in the spectrum
P2	Partial	Data potentially diagnosable; e.g., envelope spectrum shows smeared components that appear to coincide with the expected fault frequencies
N1	No	Data not diagnosable for the specified bearing fault, but with other identifiable problems (e.g., looseness)
N2	No	Data not diagnosable and virtually indistinguishable from noise, with the possible exception of shaft harmonics in the envelope spectrum

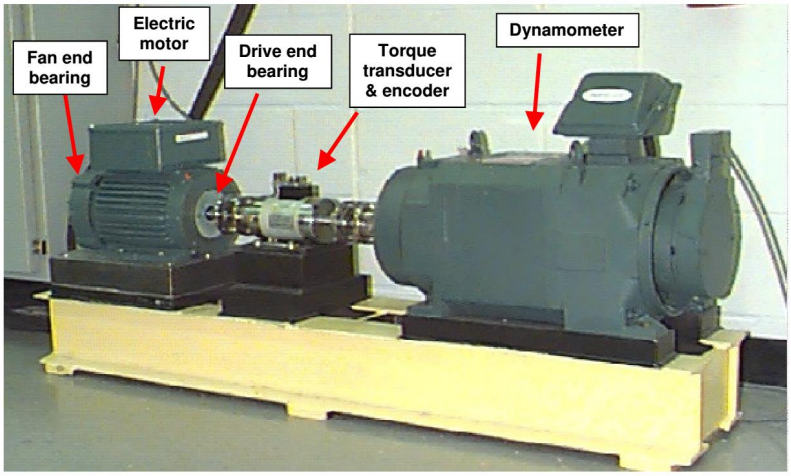

Fig. 3.6 CWRU bearing test apparatus [65, 69]

[65, 57] recommend discrete/random separation (DRS) and autoregressive (AR) model methods respectively, while some other techniques are discussed in Ref. [34]. Nevertheless, in this chapter neither pre-whitening nor separation of deterministic and random parts is performed prior to analysis because there are no components such as gear in the system which may produce high energy deterministic signals and mask the bearing signal. In this regard, the benchmark study employed DRS filter but the author couldn't find any positive effect of this technique on the diagnosis of CWRU data set, i.e. the results are the same and using DRS makes no difference. Moreover, if the parameters of the filters (e.g. DRS and AR), which should be set in advance, are not correctly selected, their performances would be improper.

Daubechies wavelets (db12) are employed to decompose the signal in Step 1. Moreover, to have proper minimum frequency band of demodulation for data with sampling frequencies of 12 and 48 kHz, 5 and 6 levels of decomposition are performed respectively, for both FK and Autogram.

Note that in the envelope spectra throughout this section, green dash-dot lines cursors are depicted at the nominal shaft frequency, red dashed lines at harmonics of the expected fault frequency, and red dotted lines show the first order modulation sidebands around the fault frequency and its harmonics. These sidebands are spaced at shaft and cage frequencies for defects on the inner race and rolling element respectively. For outer race faults there will be no sidebands because the outer race is fix and as a result the transmission path between the defect and the signal acquisition point does not vary.

This section is divided in two parts. The purpose of the first part is to explain in deeper details the steps described in subsection 3.3.1. Additionally, the performance of the Upper, Lower and original Autogram and the effects of different proposed thresholding will be illustrated. In the second part the CSES results will be presented and a comparison with Fast-SC will be shown.

3.4.1 Synthetic signal

In this section the properties of the method are explored and illustrated using a simple cyclostationary synthetic data.

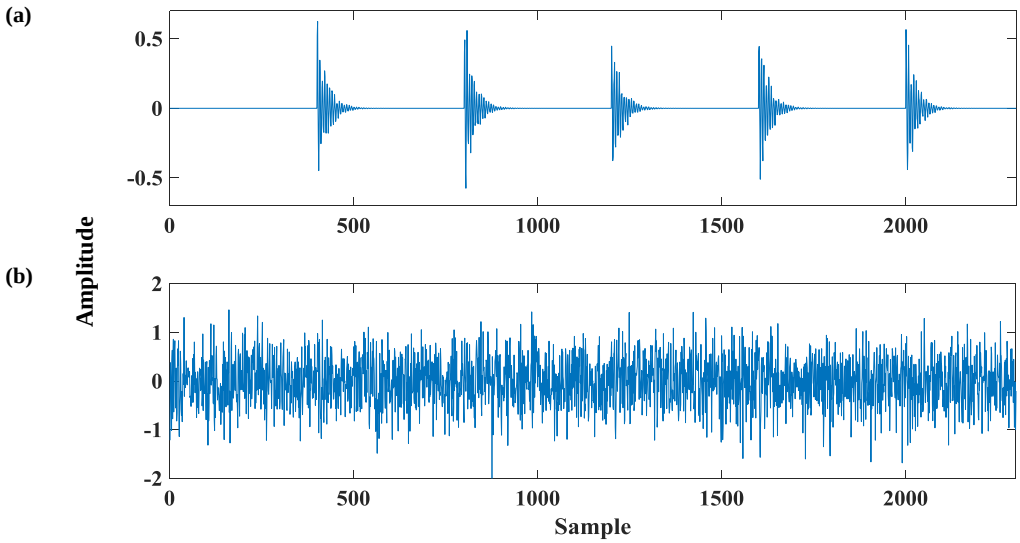

Fig. 3.7 (a) simulated bearing defect signal (b) combined simulated signal with SNR = -17 db

3.4.1.1 Model of the bearing defect signal

The simulated signal $x(t)$ with one resonance frequency is generated as follows:

$$x(t) = \sum_i A_i \delta(t - T_i) \tag{3.5a}$$

$$A_i = \sum_j C_j e^{-\beta j} \sin(2\pi f_r j) \tag{3.5b}$$

$$T_i = iT + \delta T_i \tag{3.5c}$$

where in Equation 3.5a, A is the response of the system when the defect is engaged and $\delta(t)$ is the Dirac delta function. In Equation 3.5b, β is the decay factor, C is amplitude and f_r is the resonance frequency of the structure. Finally, in Equation 3.5c, T is a fixed period of the bearing defect impulses occurrence (inverse of the bearing defect frequency) and δT_i is the random uncertainty around it. The actual signal under analysis is the combination of the simulated bearing defect signal and a certain amount of white noise.

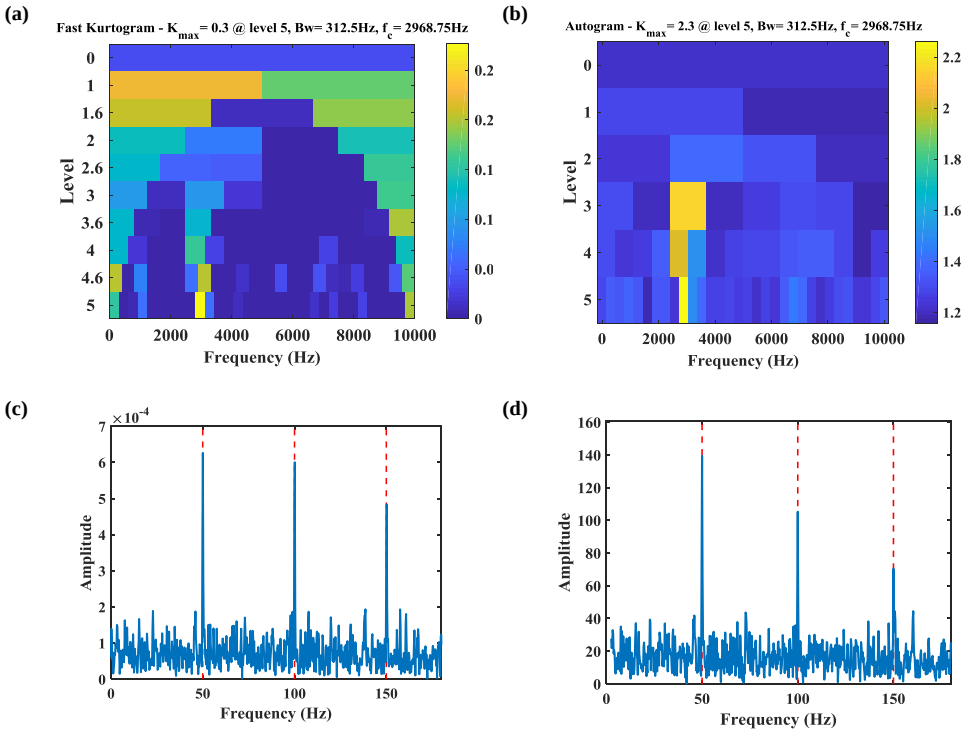

Fig. 3.8 First simulated case (a) Fast Kurtogram (b) Autogram, squared envelope spectrum (SES) of the signal related to node with highest kurtosis in (c) FK (d) Autogram

3.4.1.2 Autogram

The selected values of these variables for the first case are given in Table 3.2. To have more realistic signal, C and f_r are subjected to discrete uniform distributions which are a few percent of their nominal values.

The sampling frequency is equal to 20 kHz for the following signals and 80000 samples are the total length of each signal. The simulated defect signal and the combined signal are depicted in Figure 3.7 a and b (for the sake of clarity only part of the signals are shown). For this case the SNR is equal to -17 dB. Kurtogram and Autogram are shown in Figure 3.8 a-b. Both methods are able to detect the proper frequency bands for demodulation and the defect frequency and its harmonics are present in the SESs of the selected nodes in Figure 3.8 c and d. However, the result provided by Autogram clearly shows better outcome as the proper band in each level is identified by Autogram. In contrast to Kurtogram, the nodes containing the defect

Table 3.2 Parameters of the simulated signal

β	C	f_r (Hz)	T (sec)	δT
800	0.5±50%	3000±2.5%	0.02	0.03T

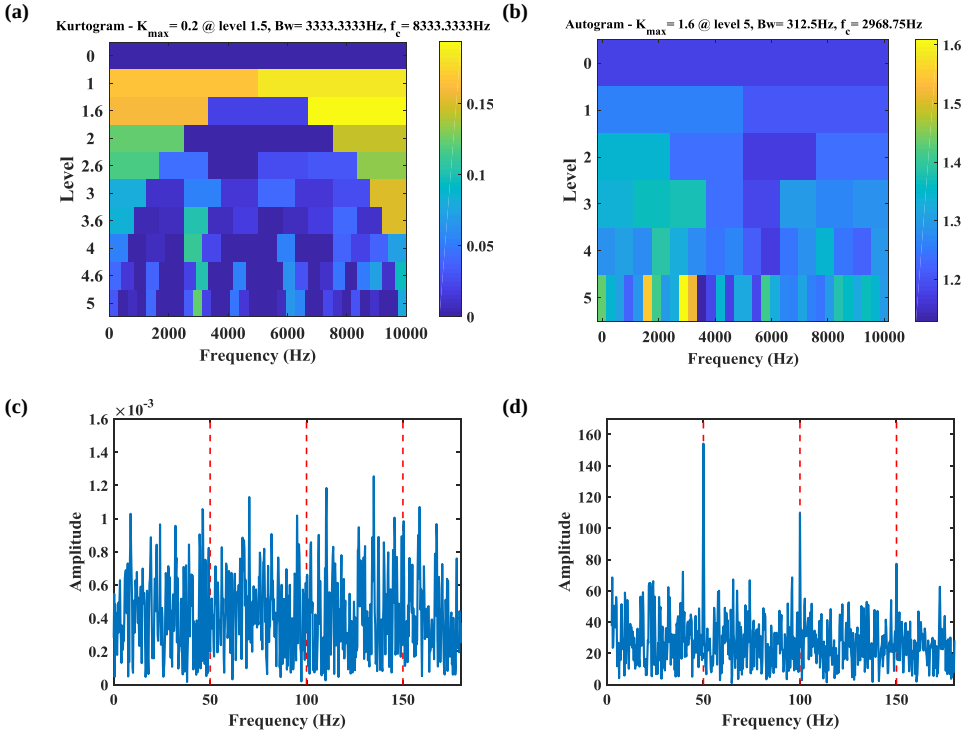

Fig. 3.9 Second simulated case (a) Fast Kurtogram (b) Autogram, squared envelope spectrum (SES) of the signal related to node with highest kurtosis in (c) FK (d) Autogram

signal are more clearly differentiated from the nodes without any useful diagnostic data.

For the second case, the SNR is decreased to -19.8 dB. Kurtogram, depicted in Figure 3.9a, fails and the suitable node is missed. Consequently, the defect frequency cannot be spotted in the SES of the selected node in Figure 3.9c. On the contrary, the node is detected by Autogram (Figure 3.9b) and the SES in Figure 3.9d provides successful diagnostic information.

The synthetic signal of the first case corresponds to the minimum SNR (-17 dB) below which Kurtogram can not detect the effective frequency band for demodulation. But, Autogram can be used for signals with SNR as low as -19.8 dB as shown

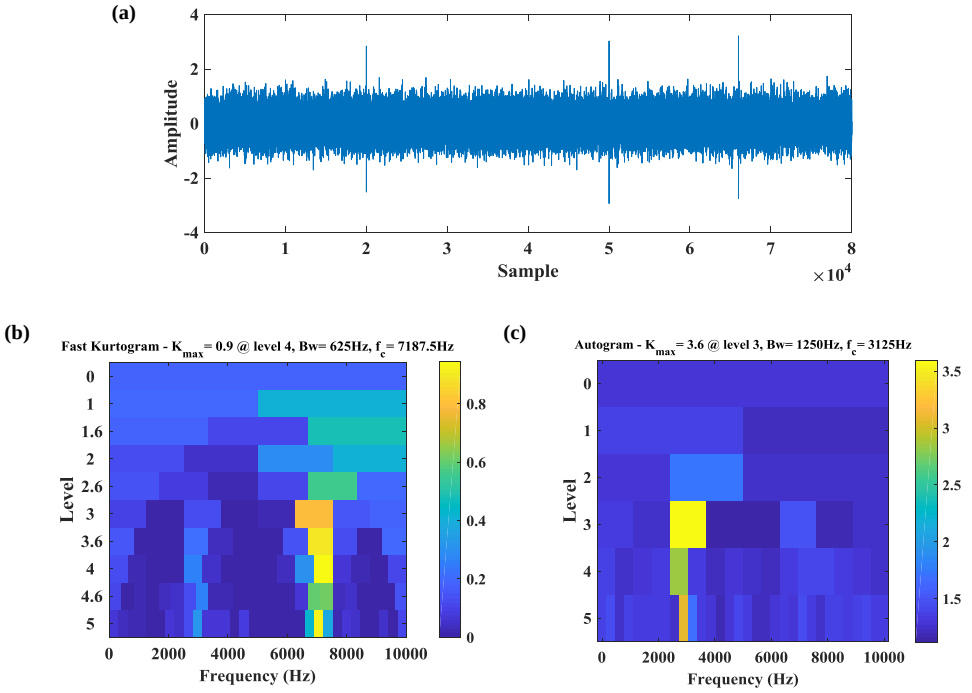

Fig. 3.10 Third simulated case (a) simulated signal with three large impulses (b) Fast Kurtogram (c) Autogram

in the second case. Overall, these two examples reveal superior performance of the Autogram over Kurtogram in dealing with greater levels of noise. As it was discussed, this is one the main disadvantages of the Kurtogram which is improved by the proposed method.

For the third case the SNR is set to -16 dB. In addition, three impulses (noise) are added to the simulated signal (see Figure 3.10a). These non-periodic impulses have the resonance frequency equal to 7000 Hz and are distributed randomly. FK is depicted in Figure 3.10b in which the frequency band including the impulses signature has the highest kurtosis. In contrast, as the impulses are not periodic, their effect is reduced after performing the AC, therefore the node related to the defect signal has the highest kurtosis in the Autogram in Figure 3.10c which also lead to a successful diagnostic. This example shows that Autogram has the ability to overcome another drawback of FK, its vulnerability to impulsive contents.

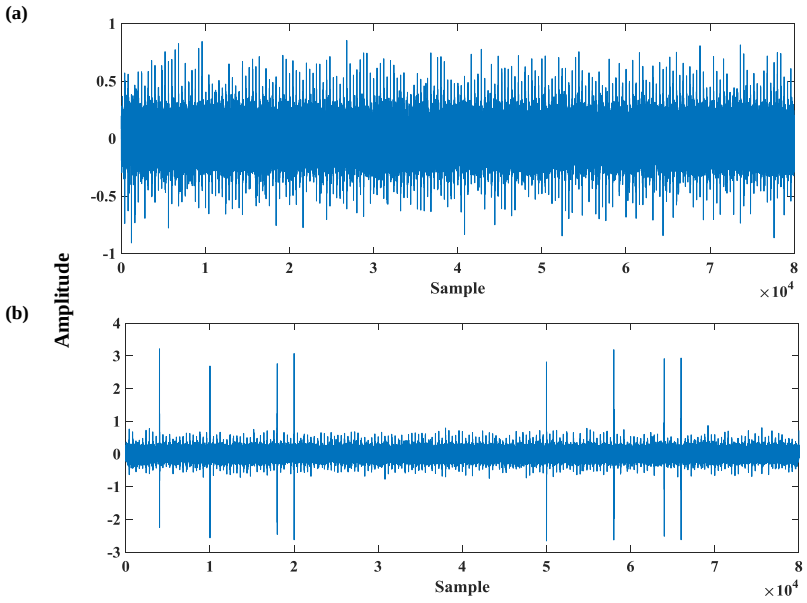

Fig. 3.11 Simulated signals (a) without impulsive noise (b) with multiple large impulses

3.4.1.3 Lower Autogram

Although, Autogram can tolerate presence of non-periodic impulses to some extent, for more severe cases, more impulsive contents with higher amplitude, it also fails similar to FK. Under this circumstance, it is advantageous to employ Lower Autogram. Figure 3.11a shows a simulated signal with the same feature as mentioned in Table 3.2. Multiple impulses with large amplitude are added to this signal and the outcome is depicted in Figure 3.11b. The ACs of these two signals are displayed in Figure 3.12 a and b respectively and the yellow lines are the threshold levels calculated by Equation 3.3.

An interesting observation is that for each maximum (peak) in the AC there is a corresponding minimum (valley) and, in contrast to maxima, minima are not affected by the modulation sources and presence of impulses. To exploit this property, Lower Autogram is proposed in which a lower threshold (yellow line) is applied on the AC of the envelope signal and then Autogram is computed for the modified signal, i.e. lower Autogram is constructed by using Equation 3.2c in which the impulsivity of the lower part of envelope AC is quantified.

The presence of impulses has dramatically increased the kurtosis value. Kurtosis value for the two ACs are equal to 13 and 80. In contrast, the kurtosis of lower

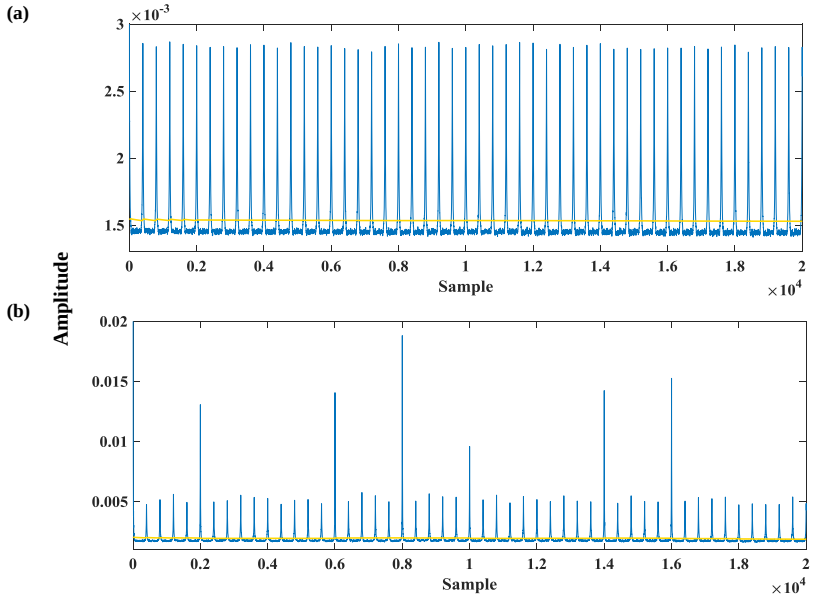

Fig. 3.12 ACs of the simulated signals (a) without impulsive noise (b) with multiple large impulsive noise (yellow lines are the threshold levels)

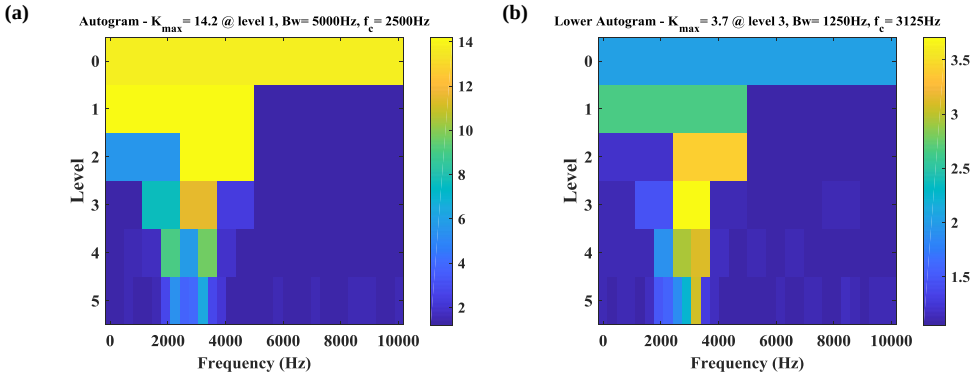

Fig. 3.13 Simulated signal without impulsive noise (a) Autogram (b) lower Autogram

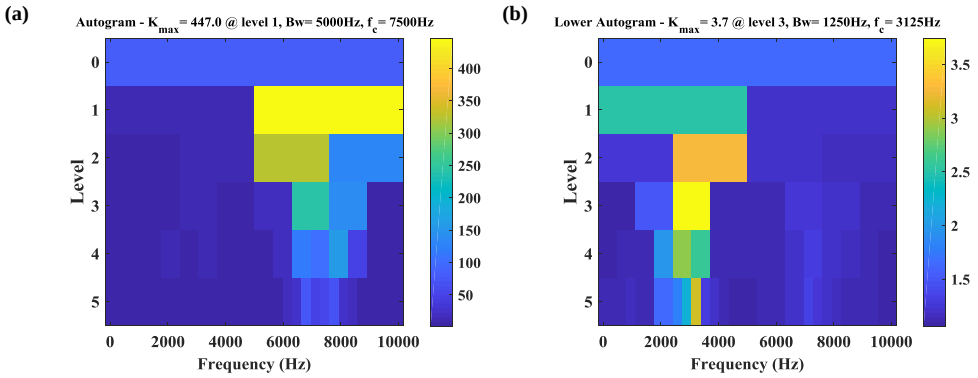

Fig. 3.14 Simulated signal with multiple large impulsive noise (a) Autogram (b) lower Autogram

parts in Figure 3.12 a and b are computed as 1.7 and 1.9. Therefore, also the second case with impulsive noise is comparable to the case without impulsive noise. It is advantageous to employ Lower Autogram when large non-periodic impulses in the signal cause high kurtosis in both FK and Autogram, even after performing AC.

Autogram and Lower Autogram for the first signal (Figure 3.11a) are shown in Figure 3.13 a and b. In the present case, the signal does not contain impulsive noise and the Lower Autogram is almost comparable to the originally proposed Autogram. For the second data set (Figure 3.11b) in which multiple large peaks unrelated to a bearing defect also exist, Autogram (Figure 3.14a) is not able to find the effective frequency band for demodulation. The selected band is related to the impulse and it has very high kurtosis. The lower Autogram is depicted in Figure 3.14b. As it is less vulnerable to existence of the very large non-periodic peaks, the frequency bands attributed to the defect signal are distinguishable in most levels.

3.4.1.4 Upper Autogram

Upper Autogram is introduced to deal with signals with lower SNR. But this feature cannot be investigated by using synthetic data as the proposed models are not able to present comprehensively characteristics of a real case data, especially the interaction between signals generated by different sources. For this reason, upper Autogram properties will be studied in the following example by using a typical experimental data.

3.4.2 Case 1: investigation of different Autograms and thresholding process by using a real data

3.4.2.1 Autogram

In this case, record 176 FE with an inner race defect is examined. The vibration signal of the record is plotted in Figure 3.15a. Defect repetitive transients are visible in the time waveform and the fault is categorized as P1, Y2 and P2 by methods 1, 2 and 3 of the benchmark study [65] respectively.

FK is depicted in Figure 3.15b and two frequency bands, which are both related to the defect, can be detected in the plot. The center frequency 9187.5 Hz with bandwidth 375 Hz has the highest kurtosis. The SES of the filtered signal, selected by the FK, is displayed in Figure 3.15d. Although it is categorized as P2 in the benchmark, the strongest component in the spectrum is indeed the ballpass frequency inner race (BPFI). This misclassification might be a result of the DRS filter or insufficient maximum level of decomposition, since the latter is not mentioned in the paper.

The proposed method is also applied to the same signal. The Autogram is shown in Figure 3.15c and the maximum value is found with center frequency 6750 Hz and bandwidth 1500 Hz, at node (4,5). The SES of the signal related to node (4,5) is presented in Figure 3.15e. An extra defect frequency band, in addition to the couple in the FK, is present in the Autogram with very high frequency.

Figure 3.16a illustrates the squared envelope of the filtered signal associated to the mentioned node (4, 5), while its AC is depicted in Figure 3.16b (to have a closer look, only short portions are presented). The filtered signal has high SNR and the uncorrelated part, Gaussian noise in this case, is canceled in the AC. As a result, almost all the defect pulses are evidently distinguishable. One of the main reason behind proposition of the Autogram is to utilize this very beneficial feature. This is the reason why the Kurtogram is not able to detect the high frequency band as clearly as the Autogram does (see Figure 3.15b and c). The defect pulses are mixed and masked by high frequency noise, and therefore the filtered signal has low kurtosis.

Bearing signals are usually modulated by other sources such as shaft frequency, and in some cases more than one defect exists in a bearing. Since the carrier frequencies, e.g. bearing defect frequencies, are not usually integer multiple of the

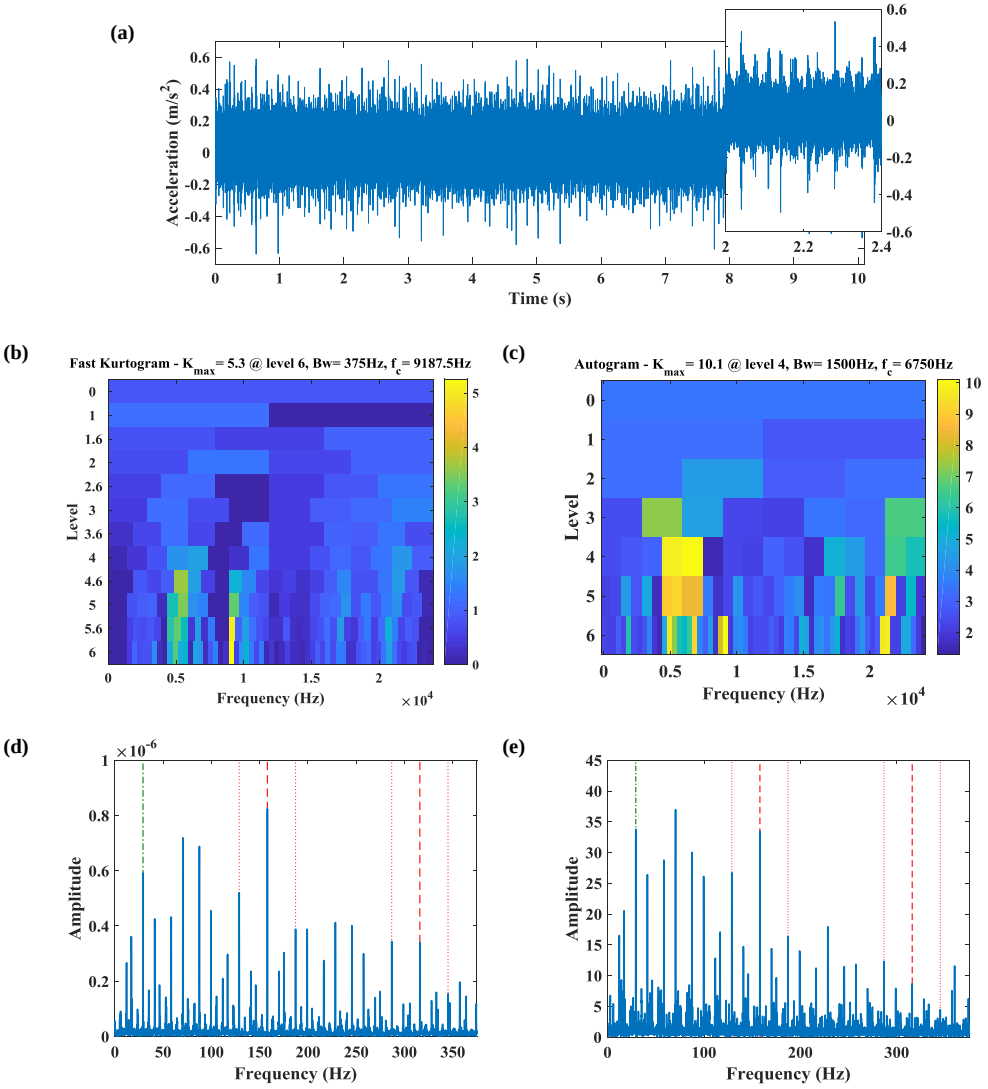

Fig. 3.15 Case 1 (a) Time domain signal: 176 FE (b) FK (c) Autogram, squared envelope spectrum (SES) of the signal related to node with highest kurtosis in (d) FK (e) Autogram (Green dash-dot line: nominal shaft frequency, red dashed lines: first two harmonics of the BPFI, red dotted lines: first order modulation sidebands at shaft speed around the BPFI and its harmonics)

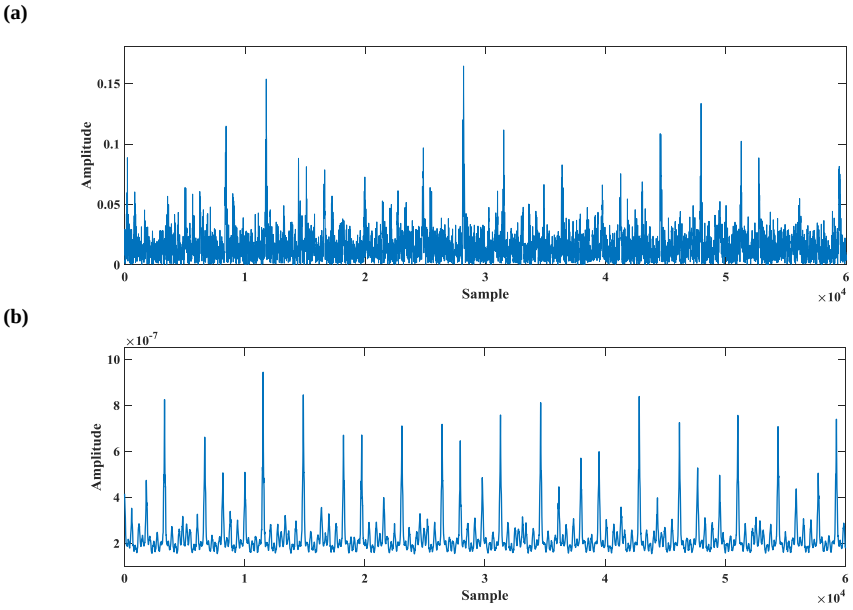

Fig. 3.16 Case 1 (a) squared envelope of the filtered signal associated to the node selected by Autogram and (b) its AC

modulation frequency, e.g. shaft frequency (f_s), the signal (and mainly its envelope) will also have, in addition to shaft and defect periodicities, a fundamental period (FP) which is different from the carrier and modulation periodicities.

FP is the *least common multiple* (LCM) of all individual periods of a signal, i.e. the fundamental frequency (FF) is the *greatest common divisor* (GCD) of all frequency components of a signal. When maxima for both modulation and carrier signals occur at the same time, the corresponding peaks, multiplication of these two values, have the largest amplitudes. Their periodicity will be equal to the FP, i.e. the peaks with the same amplitude will have the period equal to the FP. From now on, to distinguish between theses peaks and peaks spaced by the defect period we will refer to them as the ”second kind peaks”. Similarly, peaks amplitude in the signal’s envelope and its AC will manifest the same pattern.

As the distinction between peaks is more clear in the AC than in the envelope signal, it is possible to effectively alleviate the effect of the shaft frequency modulation. This is achieved by using the threshold introduced with equation Equation 3.3. The AC and the threshold level (yellow line) are depicted in Figure 3.17a. Figure 3.17b shows the spectrum of AC signal after implementing "lower threshold" (only samples below the yellow line are selected) where the amplitude of shaft frequency and its

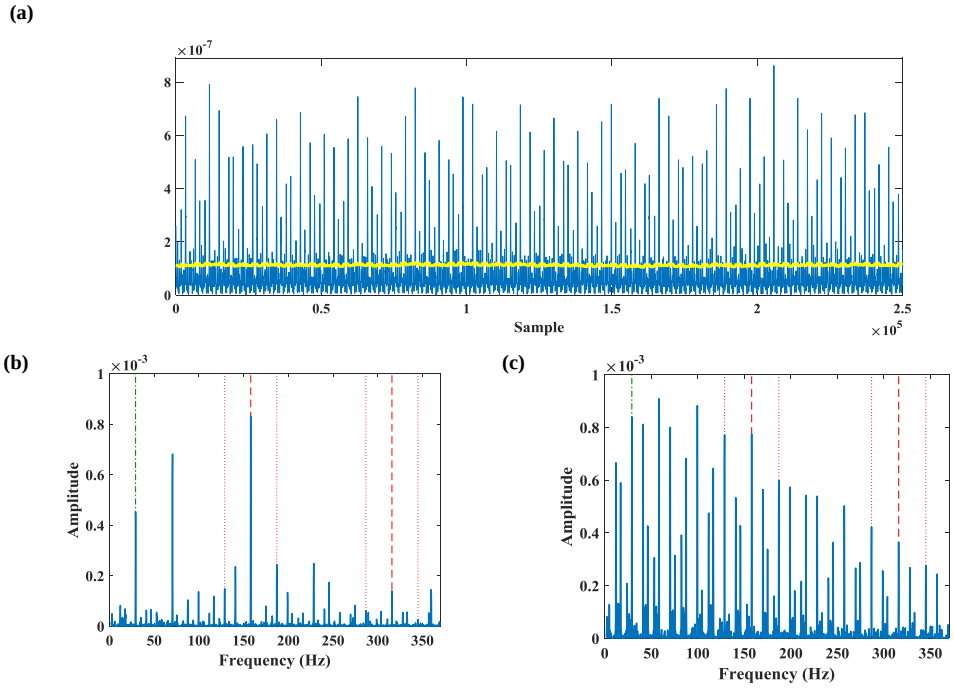

Fig. 3.17 Case 1 (a) AC of the envelope signal (blue line) and the threshold level (yellow line), spectrum of the AC signal after performing (b) lower threshold (c) upper threshold

harmonic are decreased. On the contrary, the amplitude of shaft frequency and its harmonic as well as the corresponding sidebands around defect frequency are increased after the "upper threshold" (Figure 3.17c) as the resultant signal mostly contains the "second kind peaks". These characteristics of the lower and upper parts of the AC are one of the main reasons behind the proposed Lower and Upper Autogram which are discussed in the following.

3.4.2.2 Upper Autogram

As it was discussed earlier, the largest peaks of the AC are spaced by the FP (second kind peaks). Therefore, when the SNR is low for a filtered signal, many defect peaks are masked by noise and mainly the second kind peaks can be extracted. However, since the level of noise is high and only a few peaks are present, the kurtosis value of the signal is very low and a proper frequency band cannot be detected for demodulation. Upper Autogram is introduced in subsection 3.3.1 to overcome this difficulty, in which an upper threshold is applied to the AC of the envelope signal to remove the noisy part and then Autogram, which is called Upper Autogram, is

constructed for the modified signal. Consequently, the value of kurtosis for each frequency band which mainly includes second kind peaks will rise drastically as the lower part is removed.

The Upper Autogram for record 176 FE is shown in Figure 3.18a and the maximum value is assigned to the node (5, 24), with center frequency 17625 Hz and bandwidth 750 Hz. Even though this frequency band comprises valuable diagnostic information it cannot be detected by neither FK nor Autogram.

The squared envelope of the filtered signal is displayed in Figure 3.18b and its AC and the threshold level (yellow line) are depicted in Figure 3.18c. After performing AC, the level of noise is reduced, the defect pulses are enhanced and now these peaks are more noticeable. These large second kind peaks are spaced at the envelope's period and other peaks related to defects with considerably lower amplitude can also be spotted. The spectrum of the envelope signal is shown in Figure 3.18d while Figure 3.18e presents the AC's spectrum after performing upper threshold, where it can be realized that the level of noise in the spectrum is reduced. Moreover, sidebands are more dominant in this spectrum because the lower part, which mainly contains the unmodulated part of the signal, is removed. Furthermore, since mainly second kind peaks are present after upper thresholding, the spectrum will mainly encompass the fundamental frequency and its harmonics. This phenomenon is greatly advantageous in finding bearing defect frequencies when the SNR is low because even in presence of a few peaks, the existence of the defect frequency could be revealed in the spectrum.

The spectrum of the lower part is plotted in the Figure 3.18f where the defect frequency dominates the spectrum and the shaft frequency and sidebands are almost negligible.

It must be noted Lower and Upper Autograms should be considered as complementary approaches to the original Autogram, taking into account that they highlight some specific features of signals.

3.4.3 Case 2: impulsive noise

In this case, record 275 DE - Figure 3.19a - with a defect on the inner race is examined. A number of transients are visible in the time waveform and the signal shows some level of non-stationarity. The bearing fault is diagnoseable (P1) by methods 1 and 2

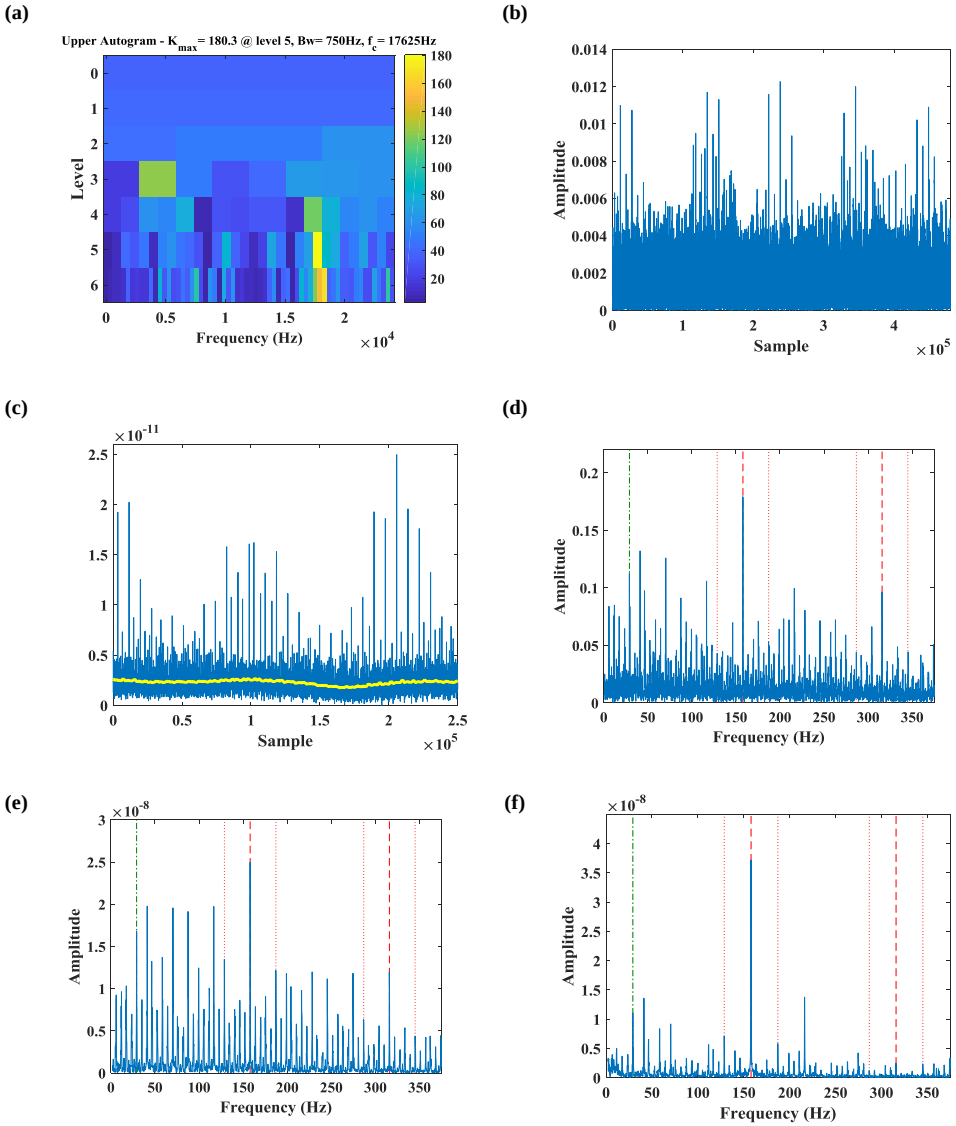

Fig. 3.18 Case 1 (a) Upper Autogram (b) envelope of the filtered signal associate to the node selected by Upper Autogram (c) AC of the filtered signal's envelope and the threshold level (yellow line) (d) SES of the filtered signal (e) spectrum of the upper part (f) spectrum of the lower part

of the benchmark study but method 3 does not determine the presence of the defect (N1) [65]. The reason is that FK is vulnerabile to impulsive noise, i.e. it tends to highlight the presence of individual impulses rather than sequences of transients. Figure 3.19b illustrates the FK in which the highest kurtosis, for the node with center frequency 3625 Hz and bandwidth 250 Hz, is related to impulsive noise. The SES is plotted in Figure 3.19d but does not provide valuable diagnostic information.

Figure 3.20a shows the filtered signal for this node which mainly contains impulses belonging to non-stationarity portions of the raw signal. The non-stationarity around 3.8 seconds causes the high value of kurtosis and it is independent from the specified bearing fault.

The Autogram is shown in Figure 3.19c and the maximum value of the kurtosis (Equation 3.2) is assigned to node (4, 4), with center frequency 1312.5 Hz and bandwidth 375 Hz. The frequency band selected by FK is not noticeable, which indicates elimination of the impulsive noise in the AC. The SES of the signal for node (4, 4) is displayed in Figure 3.19e where the harmonics of BPFI can be detected. The filtered signal and its zoomed part are shown in Figure 3.20b and, in contrast to the filtered signal related to the node selected by FK, a series of transients is present which, based on the spectrum, is indeed related to the bearing fault.

Furthermore, another interesting frequency band can be identified in the right side of the Autogram. In this branch, node (4,12) with center frequency 4312.5 Hz has the highest kurtosis and its spectrum, not reported here for brevity, provides an additional successful diagnosis.

Other data sets provided by CWRU bearing data center, such as 275 BA, 276 DE with inner race defects and 284 BA with rolling element defect (see Ref. [70]), have a similar behaviour. As a consequence, they are not diagnosable with method 3 (N1) [65] but are correctly classified by Autogram. Moreover, the defect frequency (ball spin frequency - BSF) could be detected after applying lower threshold for data 292 FE, which is not diagnosable with any of the applied methods in the benchmark study.

These results show the capability of Autogram in dealing with signals containing impulsive (non-Gaussian) noise.

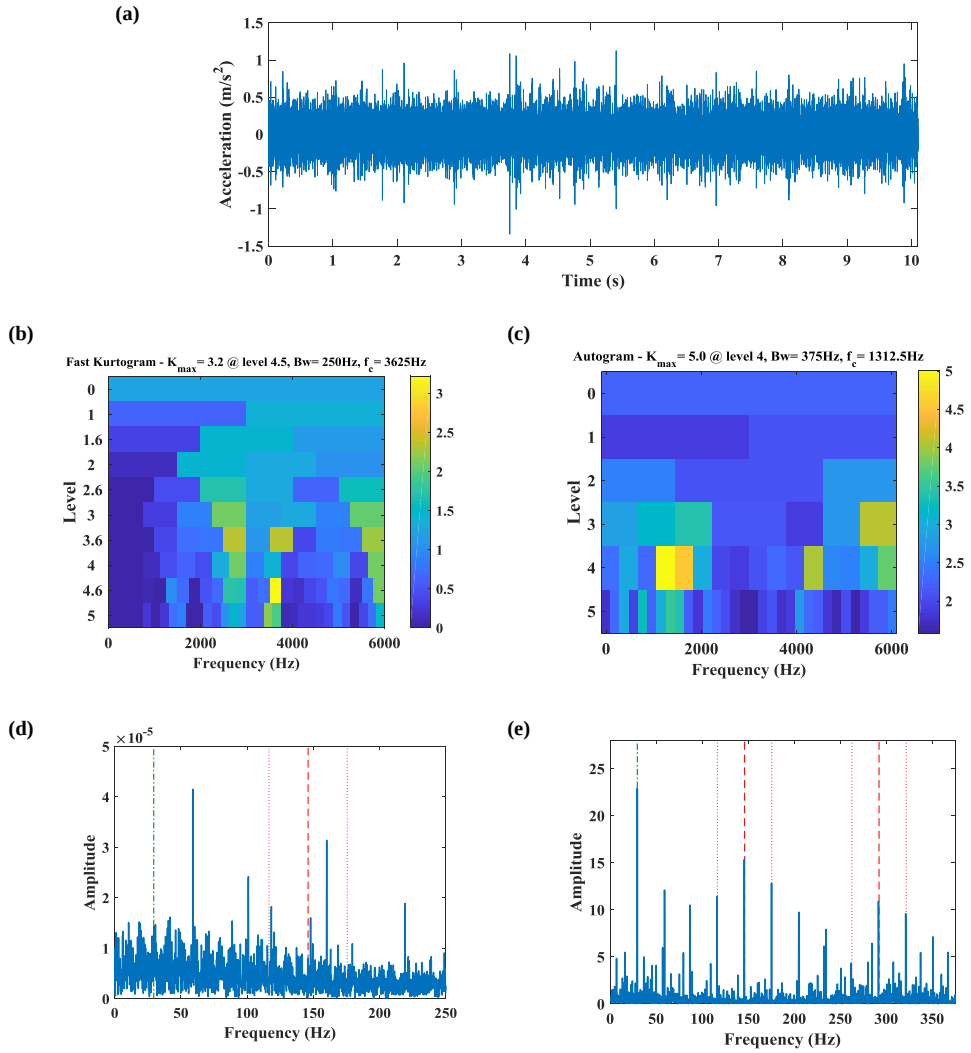

Fig. 3.19 Case 2 (a) Time domain signal: 275 DE (b) FK (c) Autogram, SES of the node selected by (d) FK (e) Autogram

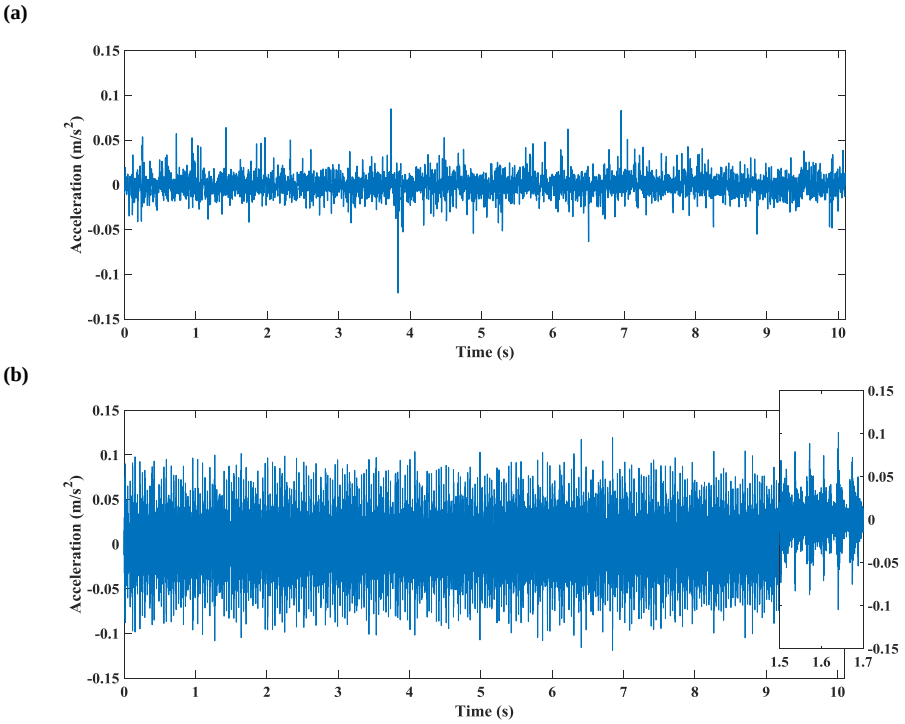

Fig. 3.20 Case 2: Filtered signal associated with the selected node by (a) FK (b) Autogram

3.4.4 Case 3: corrupted signal

In this case, record 177 FE with an inner race defect is studied and the vibration data are plotted in Figure 3.21a. From the time waveform, it can be seen that the record is corrupted with patches of electrical noise around the 0.3-0.7 seconds. Many defect repetitive transients are visible in the time waveform and the fault diagnosis result is categorized as P1 and Y2 by methods 1 and 2 of the benchmark study but method 3 failed to find the defect (N1) [65]. Figure 3.21b indicates that the FK is vulnerable to this high frequency electrical noise and the highest kurtosis is detected at center frequency 23250 Hz with bandwidth 500 Hz.

Although Autogram has a high potential to automatically deal with corrupted signals, a more guided approach is suggested to even increase this capability. When the delay (τ in Equation 3.1) is very low, the correlation between the currupted part and its delayed version generates high values in the AC. On the contrary, when the delay exceeds the correlation length of the corrupted part with itself, only the relationship between the defect signals will be recognised in the AC. Since AC of the envelope signal is employed to calculate the kurtosis, this provides a flexible tool

to select the appropriate part of the AC without losing the defect information. This feature is used to manage the corrupted signal of record 177 FE, disregarding the first 10% of the AC.

The Autogram is shown in Figure 3.21c and the maximum value is assigned to node (6, 15), with center frequency 9187.5 Hz and bandwidth 375 Hz. The SESs of the filtered signals, selected by the FK and Autogram, are displayed in Figure 3.21d and Figure 3.21e respectively. In Figure 3.21e inner race ballpass frequency (BPFI) and its second harmonic, together with their sidebands at the shaft frequency, can clearly be spotted.

As just discussed, Autogram has the potential to find the proper frequency band of demodulation even when part of a signal is corrupted. Similarly, successful diagnosis can be achieved for data sets 177 DE with inner race defect and 158 BA with outer race defect.

3.4.5 Case 4: multiple defects

In this case, data set 222 DE is studied. It presumably has a defect on a rolling element but there is often indication of inner/outer race fault for several records [65]. Data are plotted in Figure 3.22a and reveals a highly non-stationary vibration signal with many large impulses. In regard to the ball defect, the fault diagnosis is categorized as P1, Y2 and Y2 by methods 1, 2 and 3 but neither method 2 nor method 3 are able to diagnose the unintentional inner race fault [65].

The FK is plotted in Figure 3.22b and the node with center frequency 5750 Hz and bandwidth 500 Hz has the highest kurtosis. The SES of the signal related to this node is depicted in Figure 3.23a where smeared peaks can be detected at the first and second harmonics of the BSF, particularly at 2×BSF. High slippage is usually responsible for smeared components in frequency domain but sources of these smeared peaks are ascribed to random and impulsive amplitude modulation of the signal rather than ball slip [65].

The Autogram is plotted in Figure 3.22c, where an optimal band with center frequency 3843.75 Hz and bandwidth 187.5 Hz is found. In Figure 3.23b, the SES of the filtered signal provides ball fault signatures for bearing fault diagnosis. It shows smeared components that correspond to BSF and its harmonics, specifically even harmonics (2×BSF and 4×BSF). Being BSF the frequency at which the fault

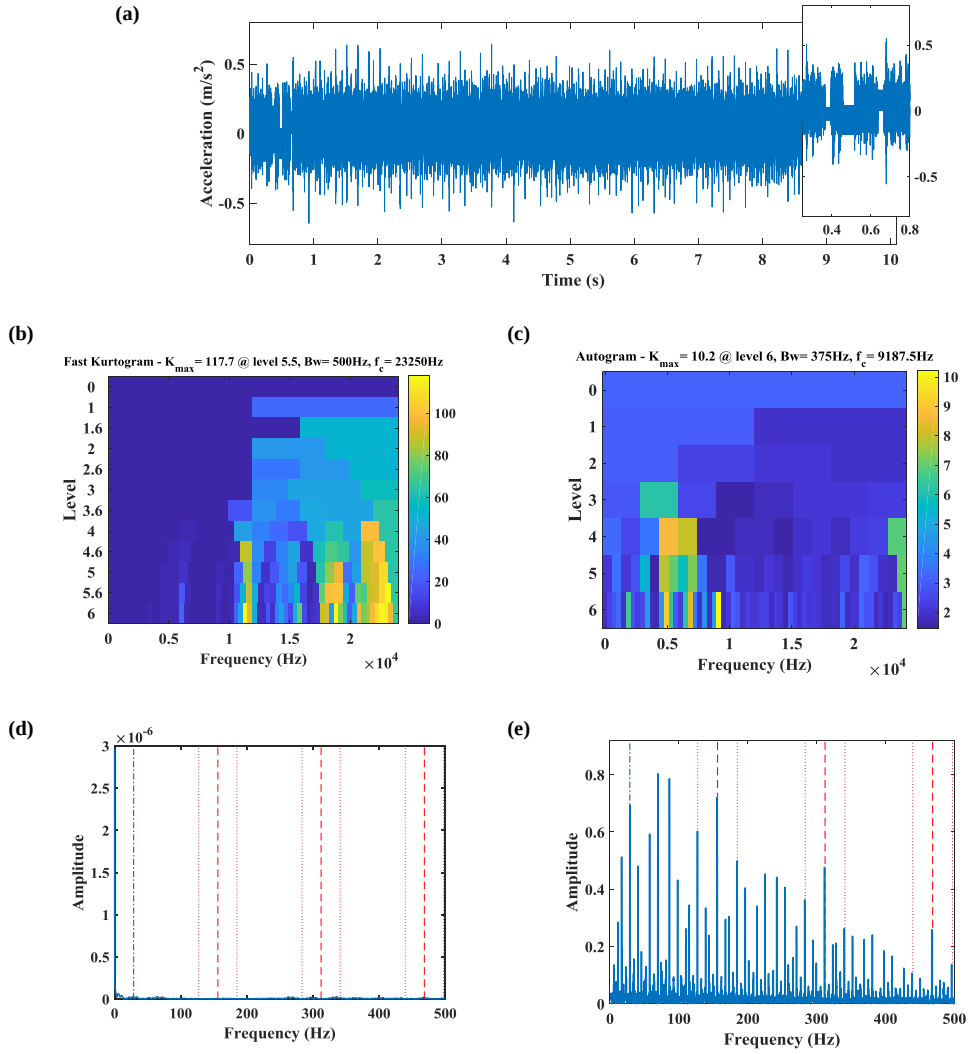

Fig. 3.21 Case 3 (a) Time domain signal: 177 FE (b) FK (c) Autogram, SES of the node selected by (d) FK (e) Autogram

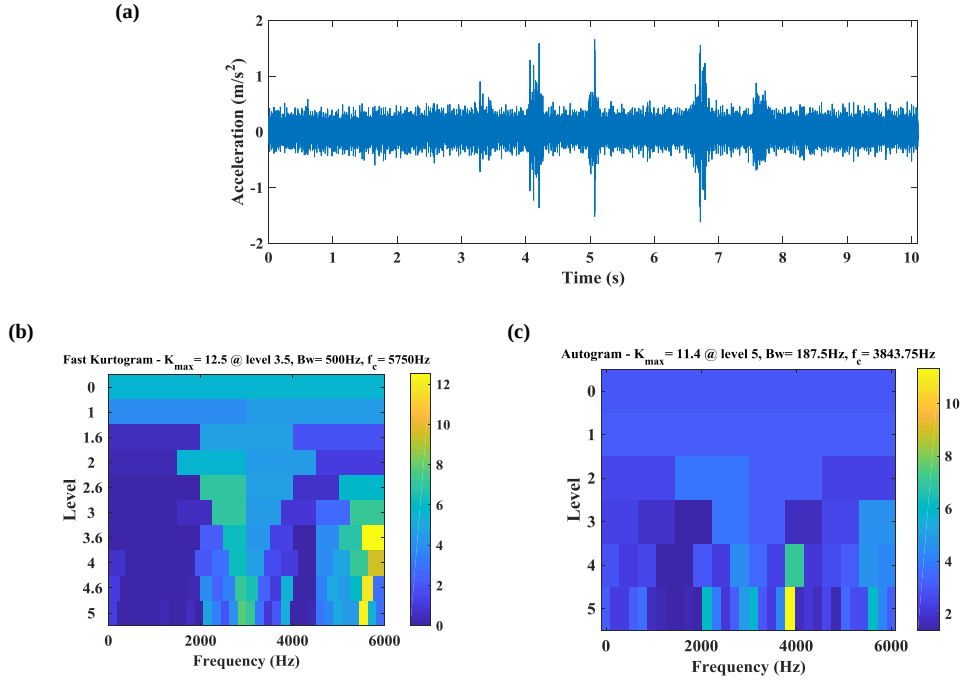

Fig. 3.22 Case 4 (a) Time domain signal: 222 DE (b) FK (c) Autogram

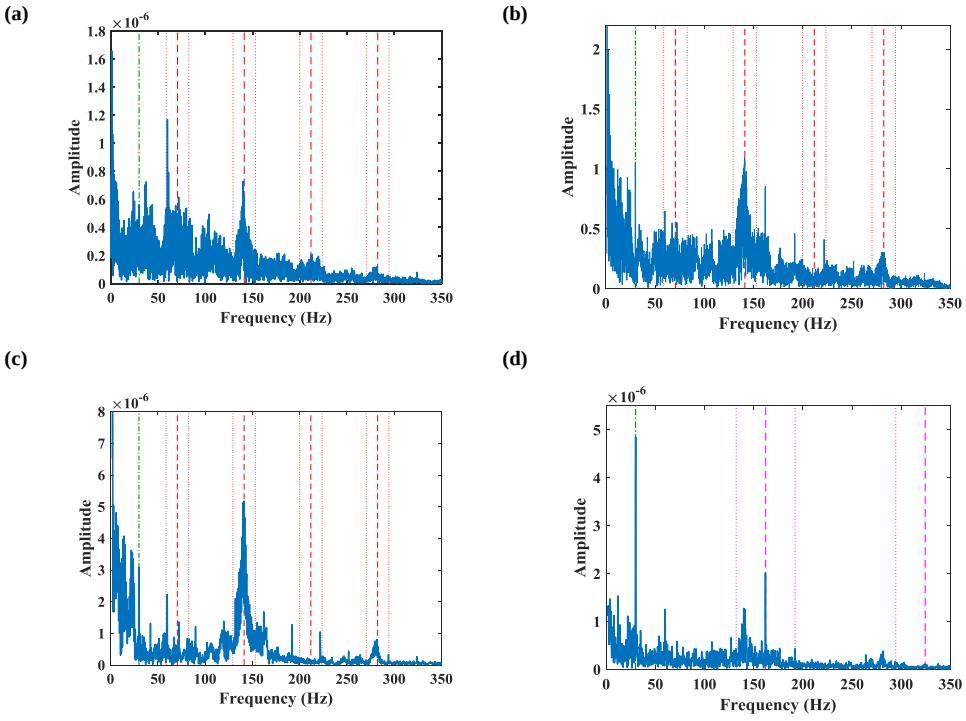

Fig. 3.23 Case 4, SES of the node selected by (a) FK (b) Autogram (c) upper part spectrum (d) lower part spectrum

is engaged with the same race (outer or inner race), the even harmonics of BSF are often dominant in the envelope spectrum.

The large impulses in time are attributed to the ball defect and FK enhances only the largest pulses related to BSF, which causes high kurtosis. Conversely, Autogram tries to find the frequency band which contains a series of impulses and, as a result, detects a different frequency band, with more dominant even harmonics of the BSF. An interesting observation is that a discrete frequency, the BPFI at 162.2 Hz, is present in Figure 3.23b but does not exist in Figure 3.23a. This indicates better performance of Autogram over FK in presence of multiple defects, especially when they are not subjected to the same random and impulsive amplitude modulation.

The SESs after applying upper threshold and lower threshold are depicted in Figure 3.23c and d respectively and the level of noise is reduced drastically. Moreover, as the Upper Autogram mainly contains the highly amplitude modulated impulses associated to the ball defect, the BSF is more evident in (Figure 3.23c) than in (Figure 3.23b). On the contrary, since the lower part is less affected by modulation, in Figure 3.23d the dominancy of the BSF is decreased and now the shaft frequency and BPFI have the largest amplitude in the spectrum. In Figure 3.23d the pink dashed line cursor is tuned at the first two harmonics of the expected BPFI, and dotted lines demonstrate the first order shaft modulation sidebands around the fault frequency and its harmonics.

3.4.6 Case 5: several non-periodic impulses

In this case, record 291 FE which has a defect on its rolling element is examined. The vibration signal of the record is plotted in Figure 3.24a, showing high level of non-stationarity due to several large but non-periodic impulses. This data is part of the fan end measurements. The important characteristic of this data sets is that the bearing characteristic frequencies are almost integer multiples of the shaft frequency which, in many cases, makes it difficult to differentiate between the defect frequency and harmonics of the shaft frequency. The bearing fault is diagnoseable (P2) by methods 1 and 3 of the benchmark study but method 2 does not determine the defect (N1) [65].

Figure 3.24b illustrates that the FK selects the frequency band with center frequency 3000 Hz and bandwidth 6000 Hz which is associated with the raw signal.

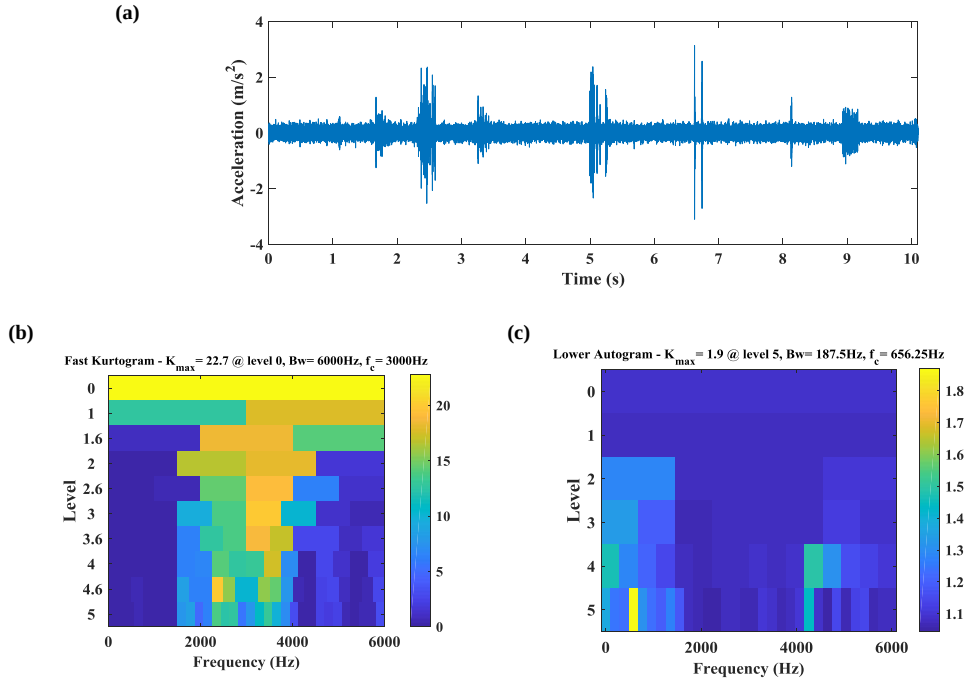

Fig. 3.24 Case 5 (a) Time domain signal: 291 FE (b) FK (c) Autogram

The SES of the raw signal is shown in Figure 3.25a. Though it is categorized as P2 there is no evidence of the BSF, possibly because contrary to Ref. [65], no DRS filter has been applied.

Very large impulses in the signal cause high kurtosis and the effectivity of the FK and to some extent Autogram is reduced. This drawback can nonetheless be overcome by taking advantage of Lower Autogram.

The Lower Autogram is shown in Figure 3.24c and the maximum value is assigned to the node (5, 4). In the present case, the defect frequency is almost twice the shaft frequency. Figure 3.25b displays the envelope spectrum of the filtered signal which clearly provides the ball fault signature, as the BSF has larger amplitude than the shaft frequency. As it was discussed, the effect of modulation could be alleviated by computing the spectrum of the lower part of the envelope's AC. The outcome is depicted in Figure 3.25c in which the defect frequency is even more relevant than in Figure 3.25b. Also, the third harmonic of the shaft frequency is completely canceled which means it was a sideband of the defect frequency rather than a harmonic of the shaft frequency.

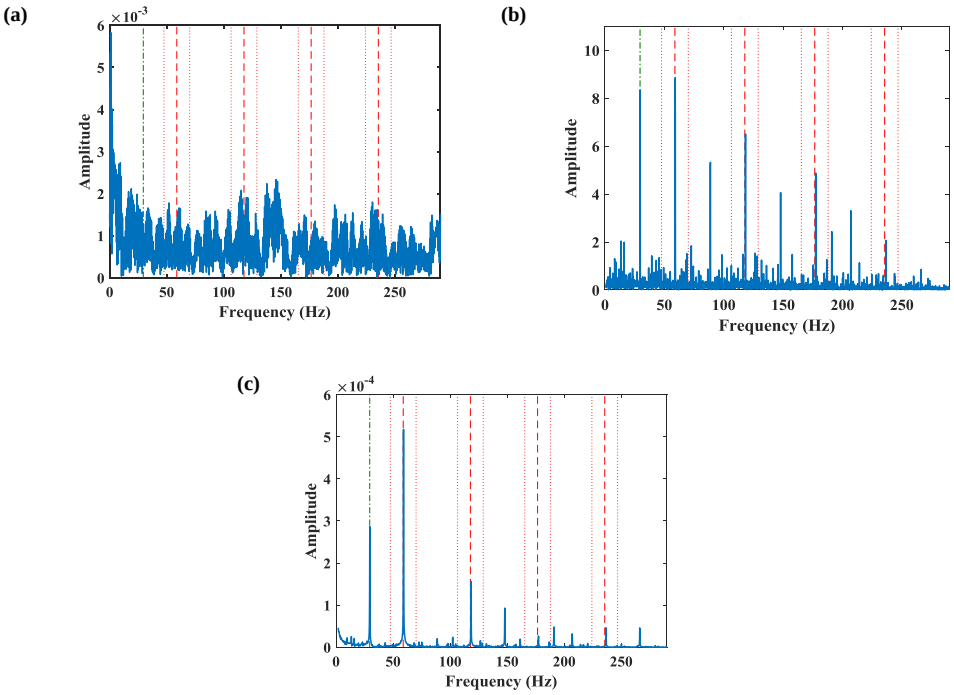

Fig. 3.25 Case 5, SES of the node selected by (a) FK (b) Autogram (c) lower part spectrum

By using Lower Autogram other data sets such as 291 BA, 283 DE, 284 DE/FE, 285 DE can be diagnosed similarly.

3.4.7 Case 6: low signal to noise ratio

As the last case of the first part, record 204 FE with a defect on its outer race is studied and Figure 3.26a presents the vibration data in the time domain. The SNR is very low and the data set is not diagnosable with any of the applied methods of the benchmark study (N1/N1/N2) [65].

The FK is depicted in Figure 3.26b, and Figure 3.26d shows the SES for the indicated optimal band, which does not contain useful diagnostic information.

As the SNR is very low, the Upper Autogram is preferable (see subsubsection 3.4.2.2) and it is plotted in Figure 3.26c. The SES of the node selected by Upper Autogram reveals fault related frequencies (ballpass frequency outer race - BPFO) in Figure 3.26e, although they are not the most dominant components.

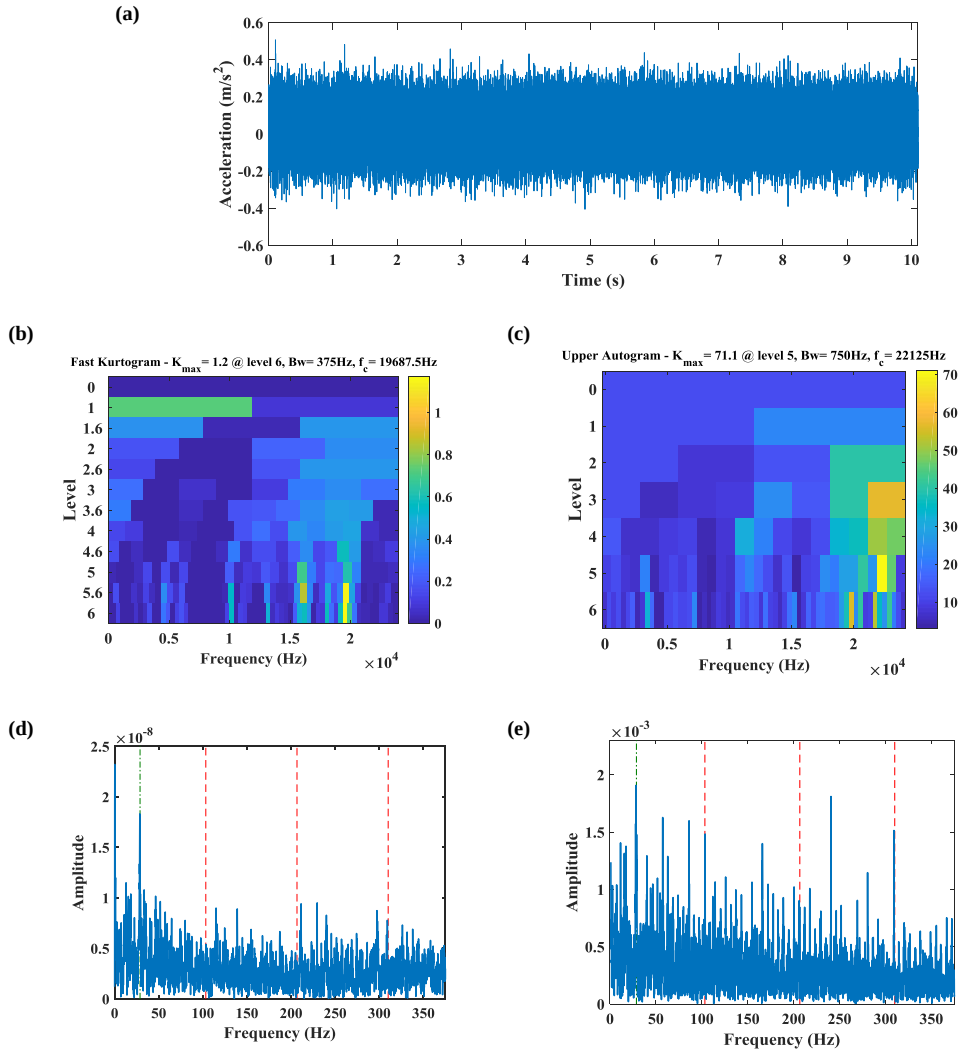

Fig. 3.26 Case 6 (a) Time domain signal: 204 FE (b) FK (c) Autogram, SES of the node selected by (d) FK (e) Autogram

Table 3.3 CPU time in seconds required to compute Autogram for 6 cases studied in this section

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Length of Signal (sample)		481e3	121e3	482e3	121e3	121e3	480e3
Level of Decomposition		6	5	6	5	5	6
CPU Time (sec)	Autogram	32	4.5	39	4.5	4.3	32

Records 199 DE, 203 DE and 204 DE are other cases which can be diagnosed successfully.

This part demonstrates that the proposed method not only can provide comprehensive information regarding probably diagnosable (P) data sets but also successfully diagnoses data sets in the N categories.

3.4.8 Computational time

Before ending the first part, it is worth mentioning the computational time of the Autogram for these 6 cases. Cases 1, 3 and 6 consist of roughly 480000 samples and a 6 levels decomposition is performed. Cases 2, 4 and 5 include roughly 120000 samples and the signals are decomposed in 5 levels. Table 3.3 reports the required CPU time on a laptop computer (Intel Core i7-4810HQ Processor 2.50 GHz). The computational time mainly consists the time for filtering the signal, calculation the ACs and their envelopes. Since all the filtered signals share the same length of the original data, each level of decomposition requires twice the computational time of the preceding one.

On the other hand, less than 1 second is needed to compute the FK as it exploits multirate filtering process and only kurtosis values are calculated for each level and frequency band (but not the corresponding time series).

Although Autogram is not as fast as FK, based on the reported computational time in Table 3.3, it could still be considered for online condition monitoring of industrial systems.

In the following two cases, the performance of Autogram and CSES will be shown and for the sake of comparison, the Fast-SC and the Enhanced Envelope Spectrum (EES) (Ref. [29]) are also presented.

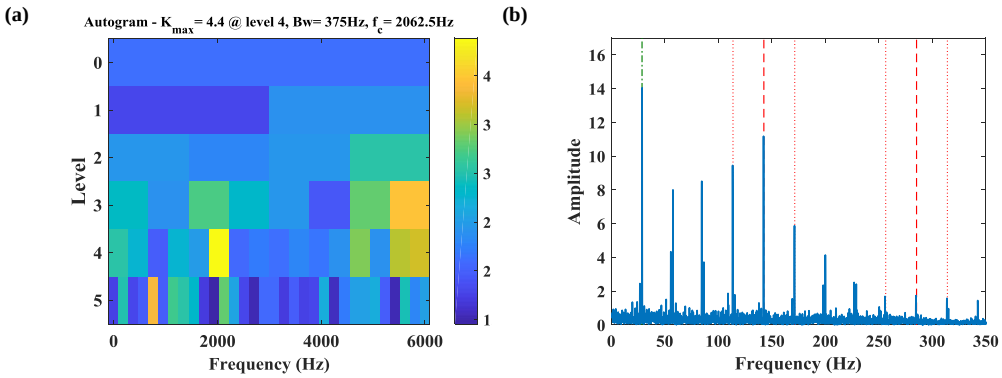

Fig. 3.27 Case 7 (a) Autogram (d) SES of the node selected by Autogram

It should be noted that the EES and CSES produce different quantities and cannot then be compared directly. The aim of these last examples is to show the capability of the proposed method in providing diagnosis information similar to Fast-SC and EES.

3.4.9 Case 7: two faulty bearings

In this case, record 277 DE is examined. The bearing located on the fan end is reported to be defective on inner race but the signal is collected on the drive end. As it is illustrated in Figure 3.27a, the Autogram selects the frequency band with center frequency 2062 Hz and bandwidth 375 Hz. The SES of the signal related to this node is shown in Figure 3.27b where the modulation or cyclic frequency associated with the BPFI of the fan end bearing can be clearly detected. Other nodes have high kurtosis in the Autogram so that, to investigate their frequency components, CSES (subsection 3.3.3) is calculated. The CSES result is depicted in Figure 3.28a and Figure 3.28b represents the average of the CSES for these six levels. The BPFI of the drive end bearing (156 Hz), that was not highlighted in the SES of the selected node by Autogram, can now be detected (pink dashed line). This example shows the improved performance of the Autogram by utilizing the CSES instead of the SES.

For the sake of comparison, the Fast-SC and EES in full band are shown in Figure 3.29a and Figure 3.29b respectively. In this case, the EES and CSES generate almost comparable results and both defects frequencies can be detected in their spectra.

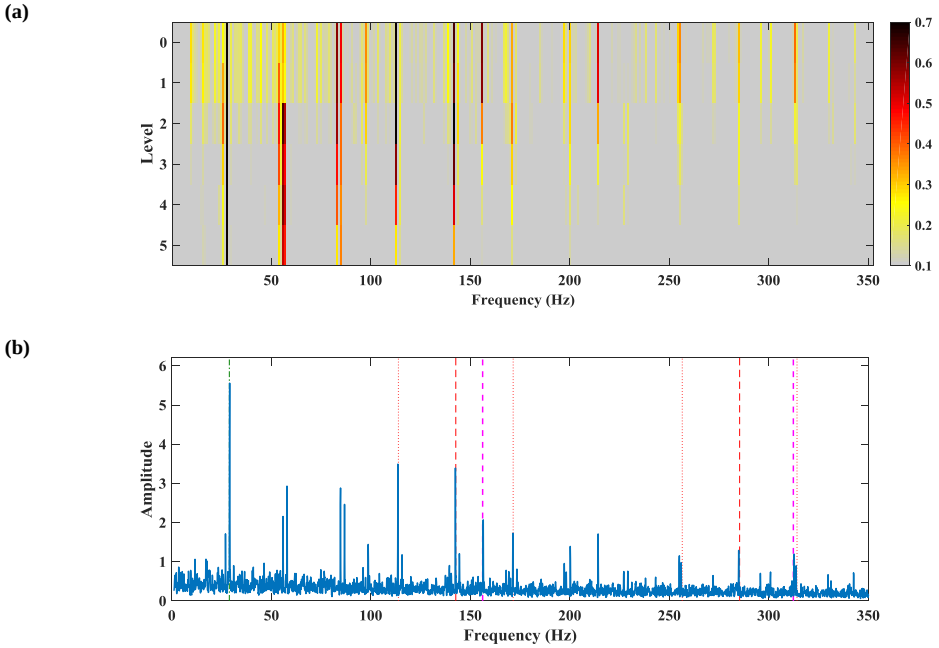

Fig. 3.28 Case 7 (a) Combined Squared Envelope Spectrum (CSES) (b) average Combined Squared Envelope Spectrum (CSES) for all levels

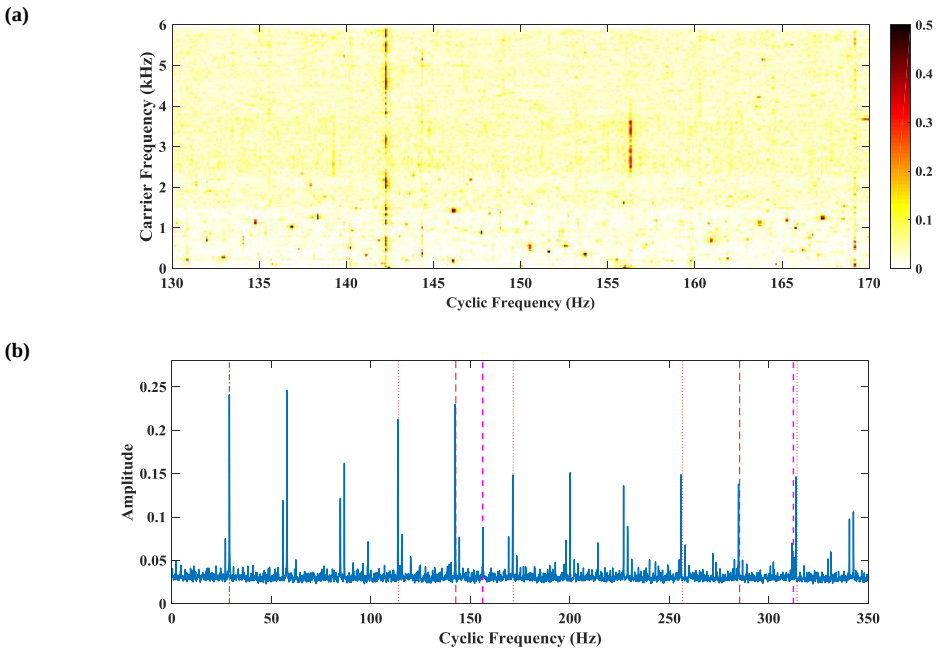

Fig. 3.29 Case 7 (a) Fast Spectral Correlation (b) full band Enhanced Envelope Spectrum (EES)

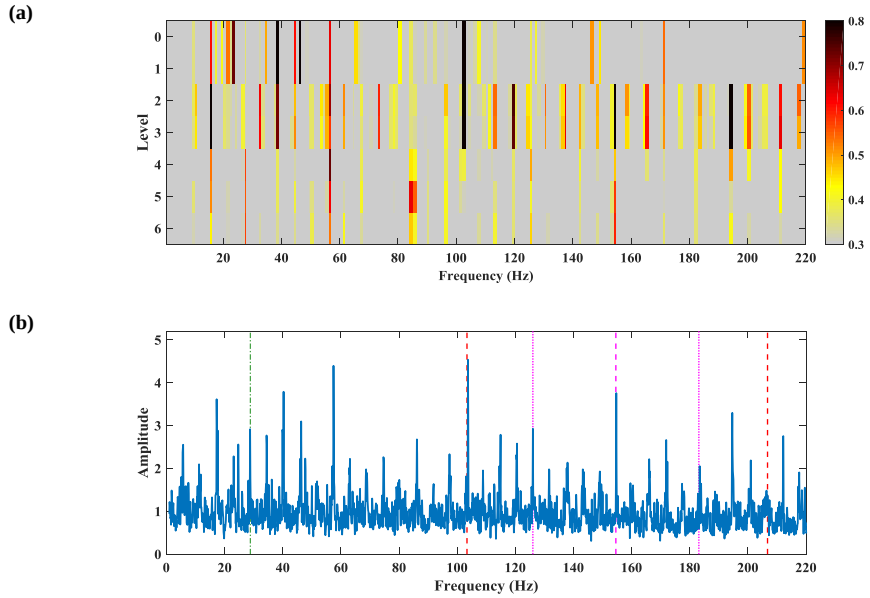

Fig. 3.30 Case 8 (a) Combined Squared Envelope Spectrum (CSES) (b) average Combined Squared Envelope Spectrum (CSES) for all levels

3.4.10 Case 8: bearing with two defected races

Record 204 DE, with an outer race fault on the drive end bearing, is studied as the last case. The CSES is depicted in Figure 3.30a for the Upper Autogram and the BPFO of the drive end bearing (103.4 Hz) is the dominant frequency through levels 0 to 3. The average CSES for these 7 levels of decomposition is shown in Figure 3.30b. Figure 3.31a illustrates the Fast-SC and its zoomed portion around the BPFO, and the EES in full band is shown in Figure 3.31b. Although no defect on the inner race is reported, the BPFI (156.1 Hz) can also be observed in the CSES (Figure 3.30) and EES (Figure 3.31b). The EES computed in the frequency band [2000 4000] Hz, in which the BPFO of the drive end bearing shows high values, is shown in Figure 3.32a and comparing to full band EES Figure 3.31b, the defect frequency is more dominant. Finally, Figure 3.32b displays the average CSES for levels 0 to 3 where the BPFO has high values.

The SC has the benefit of automatically showing the carrier frequency related to each modulation frequency. Moreover, computing EES for selected bands can enhance its diagnosis capabilities, because the spectrum can reveal more harmonics of the defect frequencies as well as their sidebands. Also the proposed method

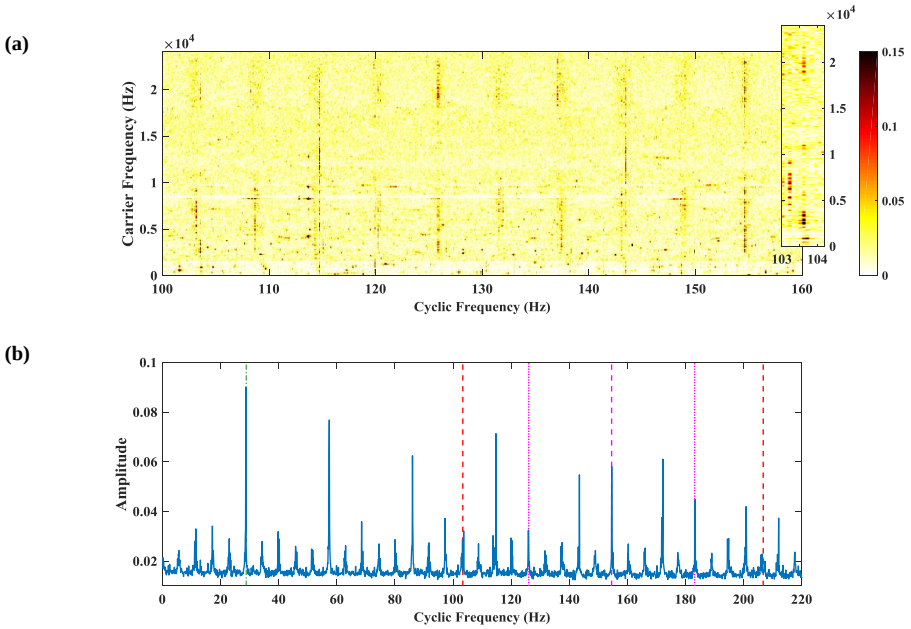

Fig. 3.31 Case 8 (a) Fast Spectral Correlation (b) full band Enhanced Envelope Spectrum (EES)

(Autogram + CSES) is able to automatically select the carrier frequencies containing useful information, even if the signal processing approach is different. According to the so far presented examples, Autogram can indeed achieve valuable results, at least comparable to other state-of-the art procedures.

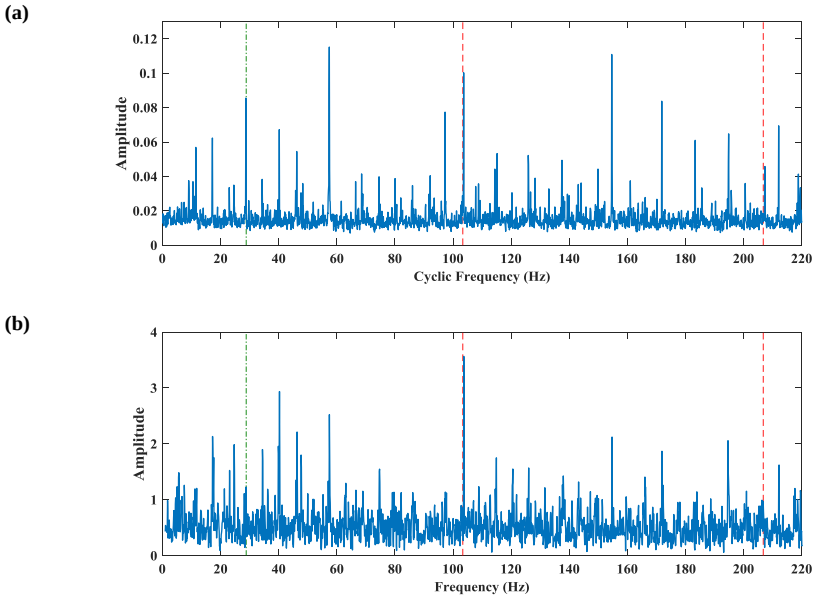

Fig. 3.32 Case 8 (a) Enhanced Envelope Spectrum (EES) in selected frequency band [2 4] kHz (b) average Combined Squared Envelope Spectrum (CSES) for levels 0 to 3

Chapter 4

A new method for diagnosis of rolling element bearings

4.1 Introduction

Cepstrum editing [41] and pre-whitening (CPW) [42] has been discussed in chapter 2. in this chapter, a new approach which is a generalized version of the CPW is developed. Although CPW is an effective approach for diagnosis of bearings, it suffers from two major drawbacks. First, as the entire real cepstrum is set to zero or, in other words, the whole magnitude of the signal in frequency domain is set to one, the SNR for the reconstructed signal is decreased because the frequency components of noise will have the same magnitude as the carrier frequencies linked to the defect signal. Second, considering bearing signal only as cyclostationary of the second order leads to the conclusion that the peaks related to the bearing damage in the real cepstrum have not considerable amplitude [42]. But this is not always acceptable since in many cases the random slippage is not high and the peaks can be seen both in frequency and quefrency domains. This explains why for some cases of the benchmark study [65] the defect frequencies are present in the squared envelope spectrums (SES) of the raw signals but they cannot be detected after performing CPW. Moreover, even by assuming second order cyclostationarity of the bearing defect signals, or pseudo-cyclostationarity as a more realistic process [10], for a typical percentage of slippage in REBs, peaks (discrete frequencies) are generated in the magnitude spectra. While the level of background noise increases, this is likely

to masks the peaks completely. Though it is worth mentioning certain peaks are amplified by resonance frequencies of structures and even under this circumstance are detectable in the spectrum.

4.2 Proposed method

Time signals are transformed to the frequency domain by Fourier transform (FT) in terms of amplitude ($A(f)$) and phase ($\phi(f)$) as follows:

$$X(f) = \mathrm{FT}[x(t)] = A(f)\, e^{j\phi(f)} \tag{4.1}$$

The real cepstrum is calculated as follows:

$$C = \mathrm{IFT}\{\log[A(f)]\} \tag{4.2}$$

In the cepstrum editing method proposed by Ref. [41], the real cepstrum is modified to remove selected harmonics and/or sidebands from the spectrum. As the process was shown in Figure 2.20 the residual signal is constructed by combining the edited real cepstrum and the original phase of the signal. As it was mentioned, in the CPW method the whole real cepstrum is set to zero to reduce the effect of deterministic parts. By using Equation 2.14 this approach is directly applicable in the frequency domain and there is no need for transformation to cepstral domain. This equation can be rewritten in the following form:

$$x_{cpw}(t) = \mathrm{IFT}\{A(f)^0 e^{j\phi(f)}\} \tag{4.3}$$

The phase relatively contains more information of a signal than the magnitude and to completely reconstruct a signal, within a scale factor, the information preserved in phase is adequate [43], consequently many of the significant features of a signal, in this case valuable diagnostic information, are preserved after this process. Both methods introduce some disturbance as the phases related to frequencies in which the amplitudes are edited are incorrect but this drawback is often insignificant. The proposed method generalizes this equation in form of:

$$x_m(t,n) = \mathrm{IFT}[A(f)^n\, e^{j\phi(f)}] \tag{4.4}$$

where x_m is called modified signal. Power of zero of magnitude is replaced by n in which $a < n < b$, a and b are two arbitrary numbers. Hereafter, the variable n will be called Magnitude Order (MO).

Similar to cepstrum editing and CPW, disturbance of signal is inevitable as the magnitude spectrum is altered. Nonetheless, the main interest for diagnosis of REBs is the repetition rate of the impulses generated by defects, therefore, if the upper and lower limit of MO (a and b) are selected appropriately, the effects of these disturbances are not problematic as the repetition rate of the impulses will remain unaffected. Based on the values of MO, the effect of the proposed method on the modified time signal could be explained in three categories.

First, for $n > 1$ the effects of frequencies with higher magnitude are amplified. Also, as the value of n increases, part of the time signal represented by frequencies with lower magnitude will be gradually masked by other parts containing frequencies with higher level of energy.

These values of MO have also denoising effect as the distinction between resonance frequencies of the structure excited by defect impacts and other frequencies associated with the noise will be better highlighted. Although, it should be mentioned that very high values of b should be avoided since the effect of only few large peaks will dominate the magnitude spectrum. it is suggested not to use values larger than 1.5 to evade introduction of any unacceptable disturbance to the signal.

For $0 < n < 1$, as the value of n varies from 1 to 0 the dominancy of frequencies with larger amplitude will gradually reduce in comparison to frequencies with lower amplitude. In contrast to CPW, all frequency components are not set to the same magnitude. As a result, the effect of both high and low energy parts, as well as cyclostationary and periodic components will be preserved to some extent. For instance, in cases that the bearing defect signal is weak and masked by signals from other sources such as gears meshing, for these values of MO the impact of "discrete frequency noise" will be lessened, meanwhile the role of bearing defect signal will be improved in the modified signal.

For $n < 0$, this process has a very different and unique outcome. The effect of frequencies magnitude will be reversed. Amplitudes of more dominant frequency components are now negligible which is analogous to filtering out these frequencies. On the contrary, the influence of previously insignificant frequencies will be boosted. It is worth mentioning that using very low negative values for a may be problematic

since frequencies with very low amplitudes are usually related to noise. Therefore, when they are enhanced there is a possibility that unacceptable peaks be introduced in the SES, so it is recommended not to utilize values smaller then -0.5.

Negative values of n are advantageous especially for diagnosis of bearings with multiple defects and/or when a rotating machinery has more than one faulty bearing and there is a considerable difference among their energy levels. For these MO, when the level of noise is not significant to dominate the modified signal, the defect with very low level of energy could be extracted more effectively.

Implementation of the proposed method is simple and fast and no prior knowledge of signals is needed. Moreover, it is not required to set any parameter in advance or to find a frequency band for demodulation. Since it automatically decomposes a signal for different resonance frequencies and discrete frequencies based on their amplitude.

After computing the modified signal for a range of MO, SESs of each signal is calculated. Then these spectra are normalized between 0 and 1 and are demonstrated by a 3-dimensional plot, where x, y, z axis represent the MO, modulation frequency and normalized amplitude respectively. As a matter of fact, the SESs of signals after CPW and raw signals happen to be two particular cases of the proposed method with MO equivalent to 0 and 1 respectively.

4.3 Results and discussion

This section exemplifies the usage of the new proposed method for diagnostics of rolling element bearings on two real case data. As the previous chapter, the experimental data provided by CWRU bearings data center [69] are used.

4.3.1 Case 1

The purpose of the first case is to explain and show the performance of the proposed method with the help of a real case data. For this example, record 275 DE is utilized. This data and the data 277 DE share similar characteristics as the vibration signals are acquired for a same defected bearing, only the shaft speed of rotation is different, and both cases were individually investigated in chapter 3. The defected bearing with

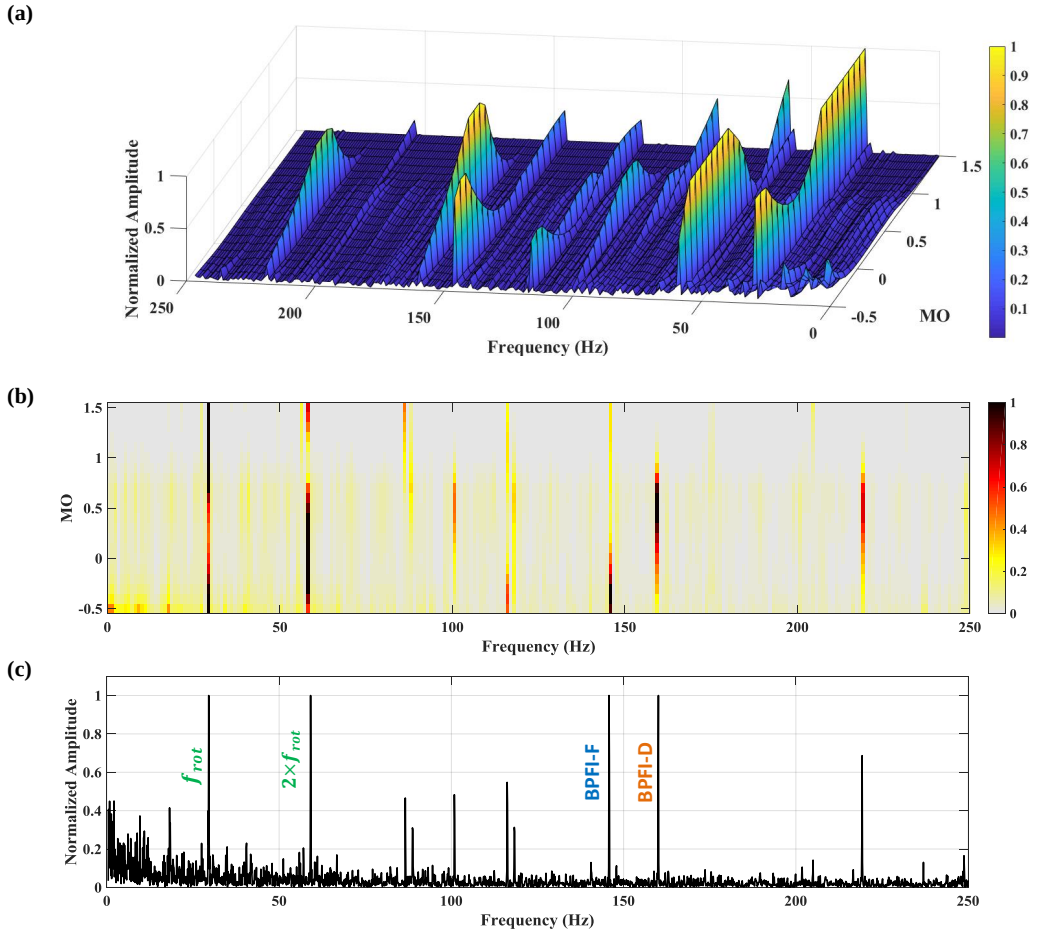

Fig. 4.1 Case 1: the results obtained by proposed method for record 275 DE (a) 3D plot (b) above view (c) view along the x-axis (MO)

seeded inner race fault is located on the fan-end of the test rig. On the other hand, the transducer acquires the acceleration signal from the drive end so the transmitted signal through the apparatus attenuates, which makes it more difficult to investigate the defect signature. Moreover, as it was shown in the previous chapter, the ball pass frequency of inner race (BPFI) for drive-end in addition to its sidebands are also present in the SES. Because the accelerometer is located close to this bearing, it might be related to the misalignment of the drive-end bearing and not an evidence of a defect bearing [29].

All things considered, this record, based on those explained features, could be considered as an interesting case to evaluate the performance of any novel bearing diagnostic technique such as the proposed method.

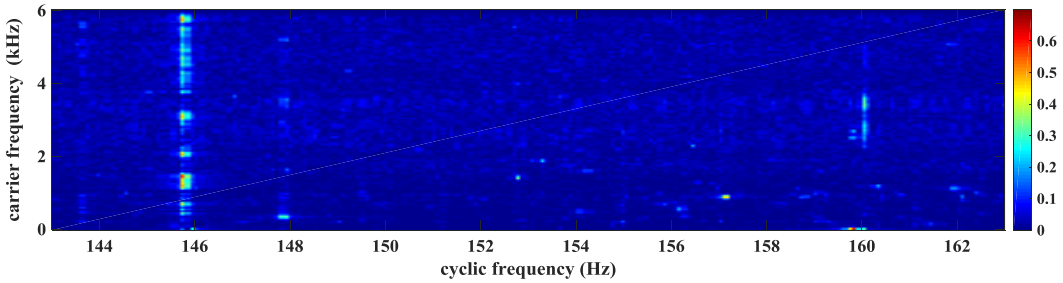

Fig. 4.2 Case 1: Fast-SC

The outcome of the proposed method is illustrated with a three-dimensional graph in Figure 4.1a where the x, y and z axis represent Magnitude Order (MO), cyclic frequency and normalized amplitude respectively, i.e. cascades of normalized SESs for modified signals generated by series of MO (n). As it was explained, to prevent unacceptable disturbance of the time signal, the values of MO ranges from -0.5 to 1.5 and the increment of 0.1 is utilized in this chapter.

Figure 4.1b displays the 2D view of Figure 4.1a in which the colors are proportional to the normalized amplitude values. The variation of each cyclic frequency intensity based on the MO values can be followed more clearly in this figure. Moreover, the quantity of MO in which a specific cyclic frequency has its highest value could be considered as an indicator of energy intensity of carrier frequencies associated to that cyclic frequency.

Figure 4.1c depicts the view along the x-axis (MO) of Figure 4.1a in a two-dimensional graph ($y-z$ plane). In fact, this plot shows the maximum normalized amplitude value of each cyclic frequency for different MO.

For this case, the shaft frequency (f_{rot}), BPFI of fan-end (BPFI-F) and BPFI of drive-end (BPFI-D) bearings are equal to 29.5 Hz, 146.1 Hz and 159.9 Hz respectively. Both damage frequencies in addition to shaft frequency and its harmonics can be noticed in Figure 4.1a-c and in fact, all the cyclic frequencies related to the signal are revealed. This is a great advantage of this method over other approaches which try to find the best frequency band for demodulation of signal.

Based on the result presented in Figure 4.1, the BPFI-F has high values for MO from -0.5 to 0 and its highest value is achieved by -0.4. On the other hand, the BPFI-D is the dominant frequency in the SESs for MO of 0.5 and 0.6 also it has high values throughout -0.5 to 1.1. As the maximum value of BPFI-F in achieved

for lower MO comparing to the BPFI-D, it indicates that the signal from drive-end bearing has higher energy level than the fan-end bearing signal. This deduction is in good agreement with the fact that the transducer is positioned near the drive-end bearing and far from the fan-end bearing.

In order to have a better understanding of carrier frequencies related to each cyclic frequency, particularly the BPFI-F and BPFI-D, the outcome of spectral correlation for this signal by using Fast-SC algorithm [29] is accessible in Figure 4.2. Several carrier frequencies are responsible for the BPFI-F which makes it difficult to extract all parts of this defect signal, i.e. even knowing the exact carrier frequencies this is corresponding to a series of filtering processes for multiple frequency bands and central frequencies. The carrier frequencies related to the BPFI-D are mainly concentrated in the mid-range. Although, this band is not completely continuous and in this range some carrier frequencies are still linked to the BPFI-F.

To investigate in more details the performance of the method and its diagnosability, the SESs of the modified signals for MOs equal to 1, 0, -0.4, 0.5 and 1.5 are depicted in Figure 4.3a-e respectively. In addition, the magnitude spectra of the modified signals are illustrated in Figure 4.4a-e.

The SES of the raw signal and its magnitude spectrum are shown in Figure 4.3a and Figure 4.4a. The BPFI-F and BPFI-D and their harmonics are present in the SES but they are not the most dominant frequencies and the maximum amplitude in the SES is corresponding to the shaft frequency. It was discussed that the resonance frequencies of BPFI-D have a high level of energy and can be spotted around 3200 Hz in Figure 4.4a. Two resonance frequency bands around 500 Hz and 1100 Hz have also sizable magnitudes but amplitudes of other frequencies are very small and negligible. Another observation is that the discrete frequencies which are amplified by the resonances are present in the magnitude spectrum.

The SES and spectrum of the acceleration signal after performing CPW are depicted in Figure 4.3b and Figure 4.4b respectively. As expected, the magnitude of whole frequency range is one. Consequently, the dominance of shaft frequency is reduced and the two ball pass frequencies as well as their sidebands are now more noticeable, comparing to the SES of the raw signal.

It should be noted that sidebands related to the BPFI-F are spaced at the shaft frequency. But the signal related to the drive-end defect is modulated by frequency twice the rotation speed thus sidebands of BPFI-D are spaced at two times of the

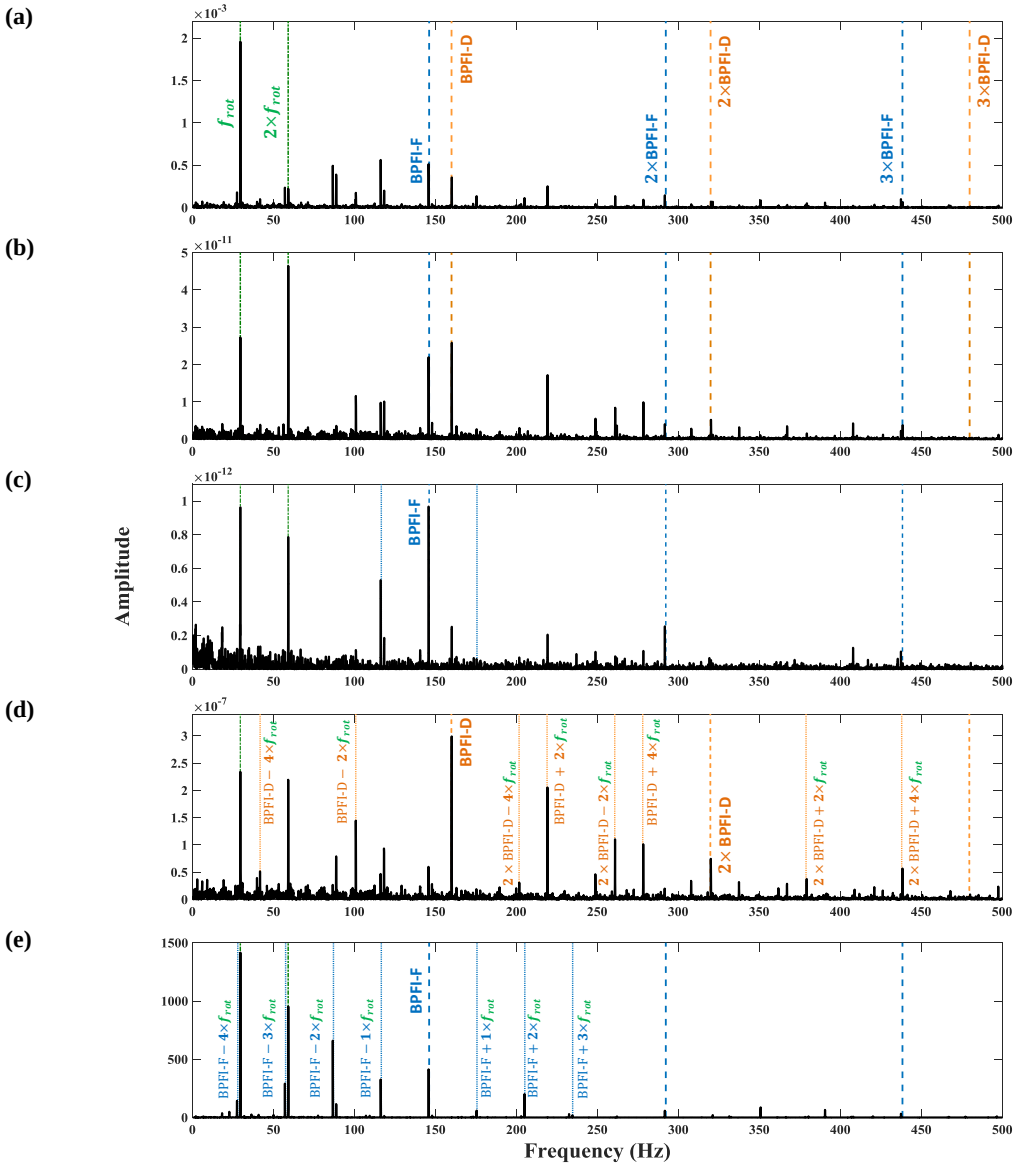

Fig. 4.3 Case 1: SESs of modified signals for different value of MO **(a)** SES of raw signal (MO = 1) **(b)** SES of the modified signal after application of CPW (MO = 0) **(c)** MO = -0.4 **(d)** MO = 0.5 **(e)** MO = 1.5

shaft frequency. This might be the reason behind the large magnitude of the first harmonic of the shaft frequency in the SES.

The SESs and magnitude spectra for MO equal to -0.4 and 0.5 are shown in Figure 4.3c-d and Figure 4.4c-d. As it was discussed, for these two values the BPFI-F and BPFI-D are the most dominant frequencies in the SESs of their modified signals.

Figure 4.3c clearly reveals the BPFI-F and its three harmonics as well as the shaft frequency and the resultant sidebands. Since the value of MO is negative the dominant resonance frequencies which are related to the BPFI-D and shaft frequency are canceled (see Figure 4.3a and Figure 4.3c). As a result, amplitude of the BPFI-D is decreased in the SES. Moreover, the influence of other carrier frequencies with negligible amplitude, mainly include signature of the fan-end bearing defect signal (see Figure 4.2 and Figure 4.4c) are enhanced.

For MO equal to 0.6, the SES is displayed in Figure 4.3d. The BPFI-D accompanied by two harmonics and their sidebands, clearly spaced at twice and forth of the shaft frequency, are highlighted. As the positive value of MO is employed, the overall behavior of the magnitude spectrum depicted in Figure 4.4d is comparable to the magnitude spectrum of the raw signal in Figure 4.4a. The value of MO is lower than one hence the effect of the frequencies with lower amplitude, including resonance frequencies around 3100 Hz which are linked to the BPFI-D (see Figure 4.2) are improved and dominates the time signal. Furthermore, the effect of large discrete frequencies is weakened which result in reducing the amplitudes of the shaft frequency and BPFI-F in the SES. It can be noticed that in Figure 4.1a and b, in addition to the absolute maximum of the BPFI-F for MO = -0.4 the value of BPFI-F has another relative maximum around MO = 1.5, i.e. it starts to increase again for MO greater than 1 and reach its maximum at MO = 1.5. The SES of the modified signal for this value of MO is shown in Figure 4.3e. Although the shaft frequency has a larger amplitude than the BPFI-F, the presence of defect frequency in addition to its three harmonics provide valuable diagnostic information. The corresponding time signal is highly modulated by the rotation of the shaft, therefore, the shaft frequency and subsequent sidebands around BPFI-F and its harmonics can be identified evidently.

The magnitude spectrum of the modified signal is exhibited in Figure 4.4e. Since the MO is greater than 1 the influence of frequencies with larger amplitude dominates the spectrum. These frequencies mainly consist of discrete frequencies enhanced

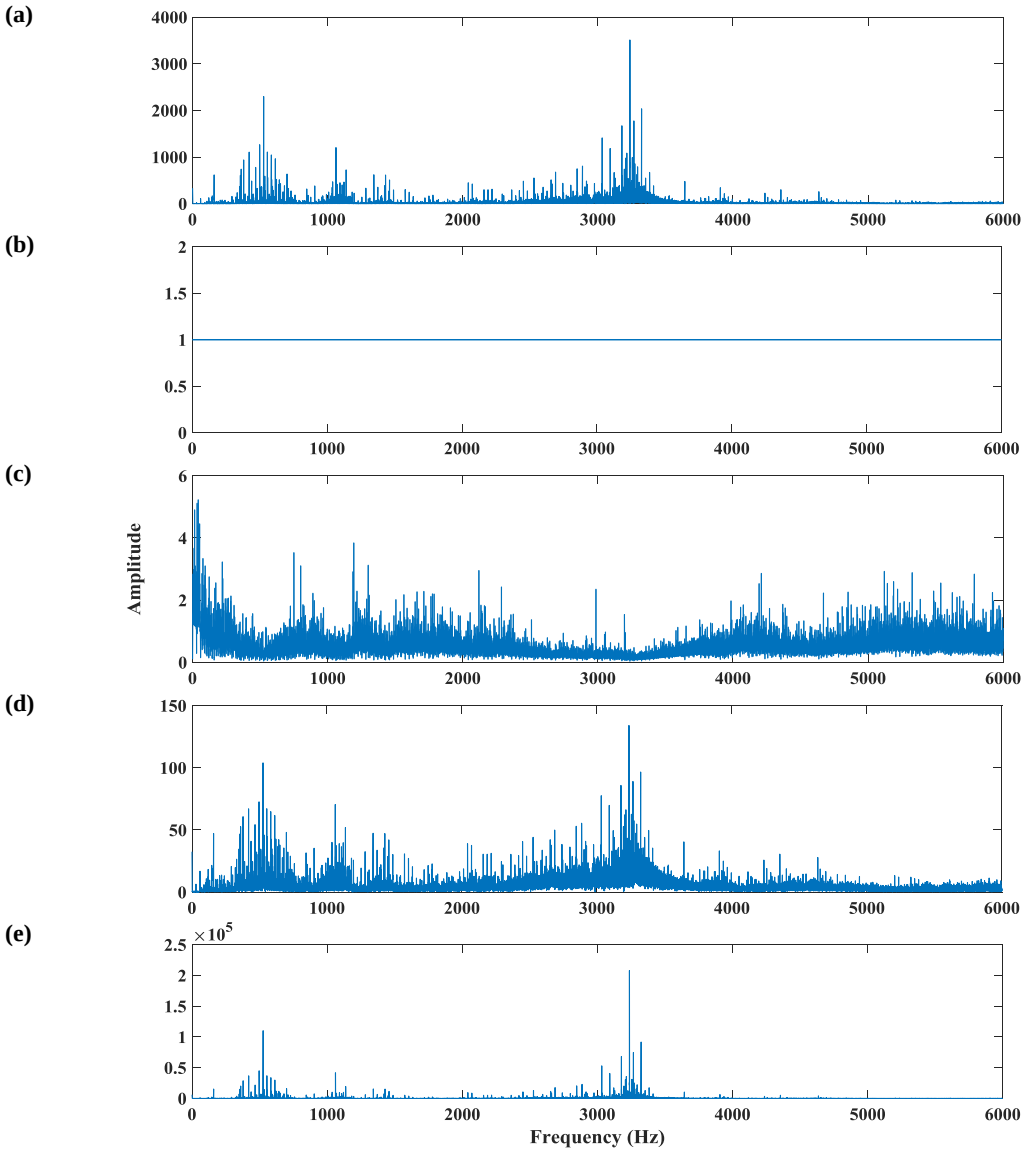

Fig. 4.4 Case 1: magnitude spectra of modified signals for different value of MO **(a)** spectrum magnitude of raw signal (MO = 1) **(b)** spectrum magnitude of the modified signal after application of CPW (MO = 0) **(c)** MO = -0.4 **(d)** MO = 0.5 **(e)** MO = 1.5

by the two resonance frequency bands adjacent to 500 Hz and 3200 Hz. As it was explained, these discrete frequencies are mostly related to the shaft frequency but also it contains symptoms of BPFI-D (see Figure 4.2).

Comparison of SESs depicted in Figure 4.3c and e discloses two dissimilar patterns for shaft and defect frequencies. This is attributed to the contribution of two phenomena with different characteristics in the defect signal. In other words, the signal associated with the fan-end bearing's defect includes two main parts, cyclostationary and periodic parts. The cyclostationary part, presented by multiple flows of energy in the spectrum (see Figure 4.2 and Figure 4.4c), has been extracted by setting the value of MO to -0.4. On the other hand, the periodic part, with multiple discrete frequencies in the spectrum, has been obtained by setting the value of MO to 1.5.

As it was mentioned, there is an overlap between carrier frequencies related to the BPFI-D and BPFI-F in the resonance frequency around 3100 Hz. However, this frequency band is still selected for MO = 1.5 but there is no sign of BPFI-F and only BPFI-D exists in the SES. It is due to the fact that the proposed method separate components of the signal mostly based on the energy levels of their carrier frequencies, therefore, the spectral lines linked to the BPFI-D, which have greater amplitudes, are effectively separated from carrier frequencies related to the BPFI-F in this frequency range.

4.3.2 Case 2

The second case will investigate the data 318 BA. The fan-end bearing, in this case, has a defect on its outer race. The acceleration signal and zoomed portion of it are depicted in Figure 4.5a. The signal is not impulsive and the defect signature neither can be spotted in the time signal nor the SES of the raw signal (Figure 4.5b), the BPFO is equal to 88.2 Hz but it is not present in the SES. The dominant frequency is associated to 114.1 Hz and also the shaft frequency ($f_{rot} = 28.8$ Hz) is detectable. The magnitude spectrum of the raw signal is illustrated in Figure 4.6a. The spectrum is contaminated with some discrete frequencies "noise" which have very high amplitude. The source of this noise is not completely evident, however, it might be because of strong electromagnetic interference (EMI) from variable frequency drives (VFD) [61]. EMI could be explained as "any unwanted signal that is either radiated or

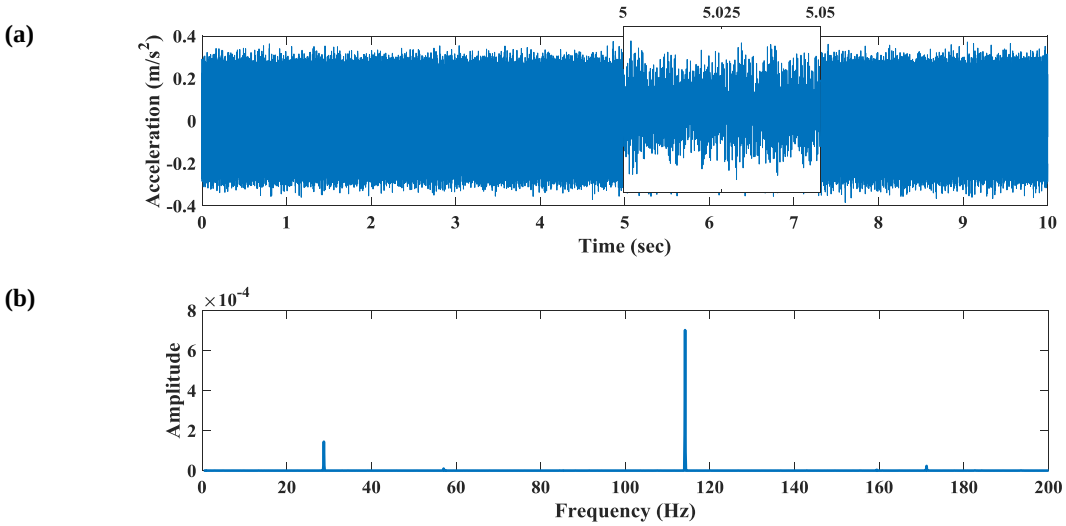

Fig. 4.5 Case 2 (a) acceleration signal of record 318 BA (b) SES of raw signal

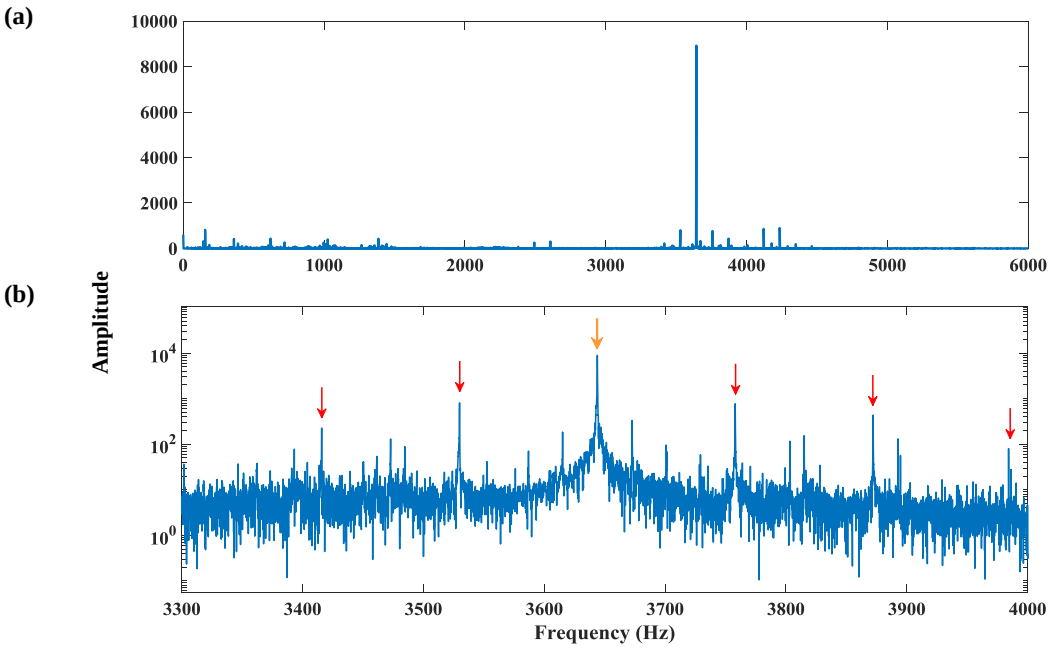

Fig. 4.6 Case 2 (a) magnitude spectrum of raw signal (b) zoomed magnitude spectrum in logarithmic scale. Carrier (switching) frequency of VFD (orange arrow) and sidebands spaced at pseudo line frequency of 114.1 Hz (red arrows)

conducted to electronic equipment and negatively affects the performance of the equipment" [61].

To investigate this assumption a zoomed portion of Figure 4.6a around the resonance frequency 3644 Hz is depicted in Figure 4.6b in logarithmic scale. Ref. [61] explains in detail the frequency representation of EMI interferences from a VFD with pulse-width modulated (PWM) output. PWM waveforms contain switching frequency of the VFD and its harmonics, modulation frequency which are also called 'pseudo line frequency' and sidebands around the carrier harmonics spaced at pseudo line frequency [61]. The carrier (switching) frequency of the VFD (3644 Hz) and sidebands spaced at pseudo line frequency of 114.1 Hz (f_{EMI}) around the carrier frequency can be spotted clearly.

This carrier frequency is the absolute maximum of the spectrum and has considerable amplitude comparing to other spectral lines. As a result, all the other frequencies of the signal, in particular the frequency band comprising the bearing defect signal, are masked by this interference.

The 3D result of the proposed method is shown in Figure 4.7a and Figure 4.7b and c demonstrate this plot from two different views. The view along in the x-axis (MO) in Figure 4.7c could be considered as a useful diagnostic tool in which the maximum value of each cyclic frequency for various MOs is accessible, i.e. one cyclic frequency might be the largest peak in the SES for a certain MO, conversely does not exist for some other values of MO but this plot displays the entire cyclic frequencies of a signal just in a single figure.

Moreover, it can be noticed from Figure 4.7a and b that throughout the -0.5 to 0.4 values of MO the BPFI is the dominant cyclic frequency. The SES of the modified signal and its magnitude spectrum for MO equal to 0.4 are displayed in Figure 4.8a and b correspondingly. Since the value of MO is smaller than 1 the effect of EMI is lessened. Correspondingly, other carrier frequencies containing the defect signal are enhanced (see the difference between Figure 4.6a and Figure 4.8b) and now the presence of BPFO and its harmonic combined with the shaft frequency provide a successful diagnosis of the bearing, though the frequency band linked to the EMI is still part of the spectrum. It should be noted that CPW as a specific case of the proposed method for MO = 0 will also detect the presence of the defect but the result is presented for MO = 0.4 to attain higher SNR.

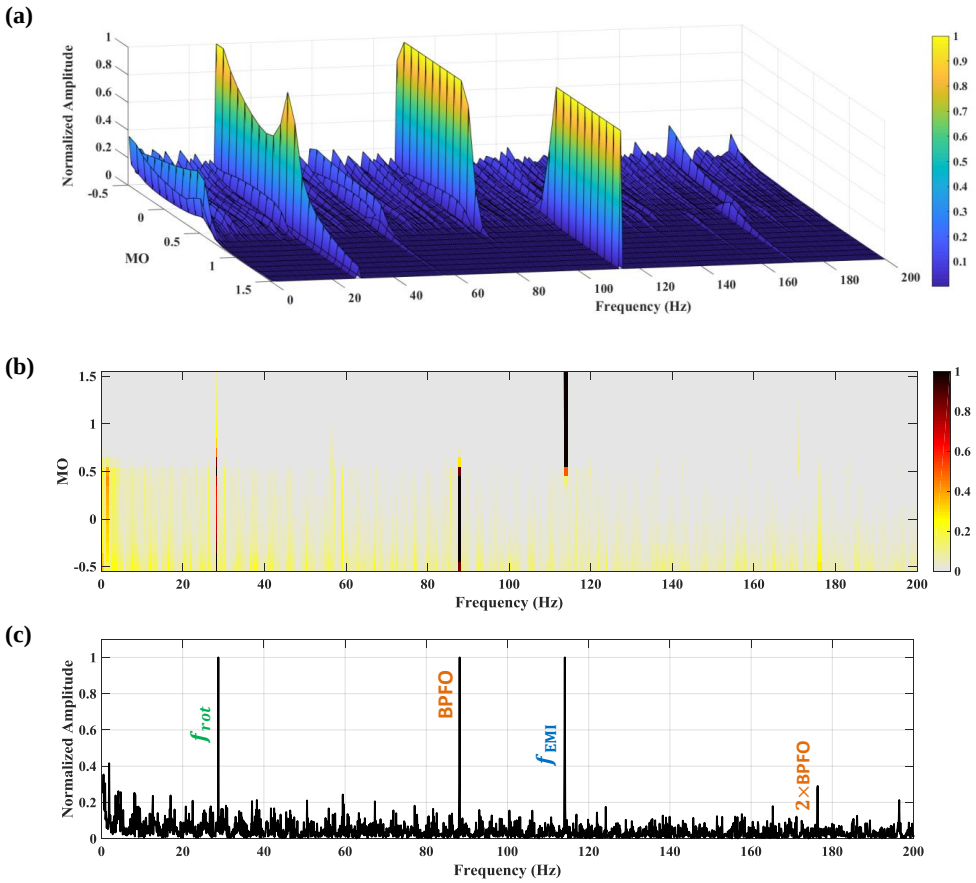

Fig. 4.7 Case 2: the results obtained by proposed method (a) 3D plot (b) above view (c) view along the x-axis (MO)

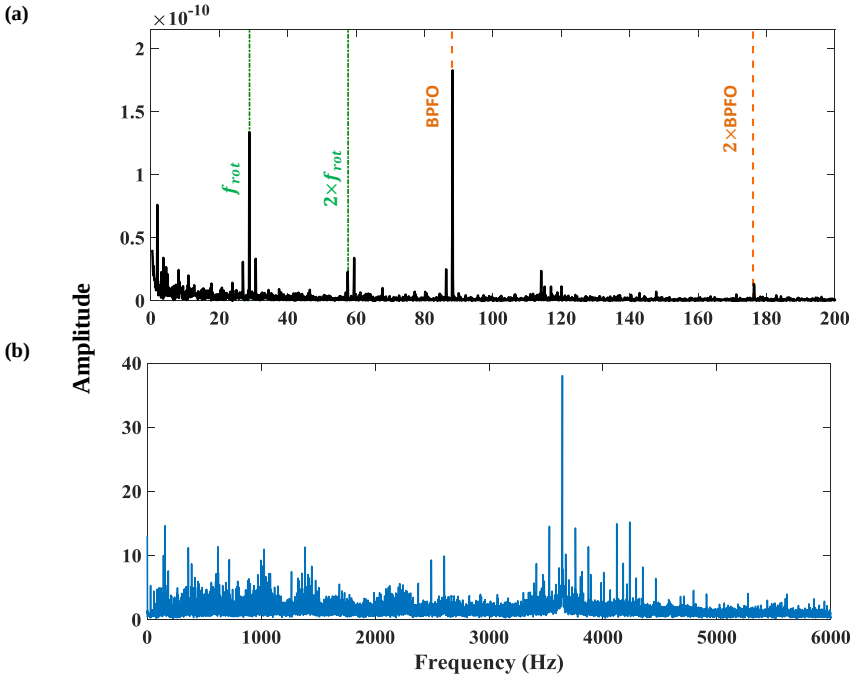

Fig. 4.8 Case 2 (a) SES of modified signal for MO = 0.5 (b) magnitude spectrum of modified signal for MO = 0.5

Instead, for all other quantities of MO larger than 0.5, the EMI effect will mask the defect signal. As a consequence, the pseudo line frequency (modulation frequency) of 114.1 Hz which is produced by the VFD will dominate the SESs and completely eradicate the signature of the defect frequency.

Chapter 5

Comparison of various diagnosis tools

5.1 Introduction

In this chapter, the performance of the newly proposed methods in chapter 3 and chapter 4 will be compared to the mostly used methods discussed in chapter 2. These approaches are applied to the same data sets; therefore, it is possible to draw conclusions about their relative merits. Moreover, to obtain a reliable assessment, data from two different test rigs are utilized, the test rig assembled in the labs of the Dynamics and Identification Research Group (DIRG) in the Mechanical and Aerospace Engineering Department, Politecnico di Torino (PoliTo) and Center for Intelligent Maintenance Systems (IMS), University of Cincinnati [71, 72].

Five vibration analysis techniques for diagnosis of rolling element bearings were presented in section 2.4. But, the focus of this chapter is on the methods for bearing signals enhancement (section 2.6) and cyclostationary analysis techniques (subsection 2.4.4) and other presented approaches will not be considered.

Time domain (subsection 2.4.1) and frequency domain (subsection 2.4.2) analyses are comparative tools rather than diagnostic approaches and per se are not useful. In addition, cepstral analysis tries to highlight the signature of defects in magnitude spectrum by combining a defect frequency and its harmonics into a single cepstral line. Consequently, this approach is not beneficial for cyclostationary signals with relatively higher slippage rate and/or lower SNR which may not produce any

detectable discrete frequency components, as the peaks might be entirely masked by the noise because of their weak energy level [11].

Time-frequency approaches are also not considered as they only map the signal and the information about cyclic (modulation) frequency (the time interval between bursts of energy produced by impacts) should be calculated separately. This could be down manually or by using methods such as SC or CMS.

Autogram has been introduced as a technique to find the effective frequency band for demodulation of signals; therefore, its performance will be compared to the previously proposed approaches for the same purpose, such as Fast Kurtogram (FK), Enhanced Kurtogram (EK) and Protrugram.

As the new approach presented in chapter 4 performs on the full band and no filtering is performed, its results will be presented besides Squared Envelope Spectrum (SES) of the raw signals, signals filtered by MED and cepstrum pre-whited signals which have similar characteristic.

Other alternatives to MED, discussed in subsection 2.6.1, will not be considered in this chapter as they suffer from a major drawback. The period of impulses should be set in advance which makes the methods impractical.

As it was mentioned in chapter 3, when the defect signal is weak, using upper Autogram is more appropriate as it is very sensitive to repetitive transients even in presence of noise. Nonetheless, it should be mentioned that upper Autogram may also be vulnerable to impulsive noise. Thus, the best option is a combination of CSES and upper Autogram.

Upper Autogram detects the frequency bands with valuable data more efficiently because the lower part which mainly contains noise and not the peaks related to defects is removed; as a result, the kurtosis values are increased and differentiation between nodes with and without valuable information is clearer. In addition, CSES and average CSES combine SESs and provide more comprehensive diagnostic information. Consequently, the outcome will be compared with the combination of Spectral Correlation (SC) and the Enhanced Envelope Spectrum (EES).

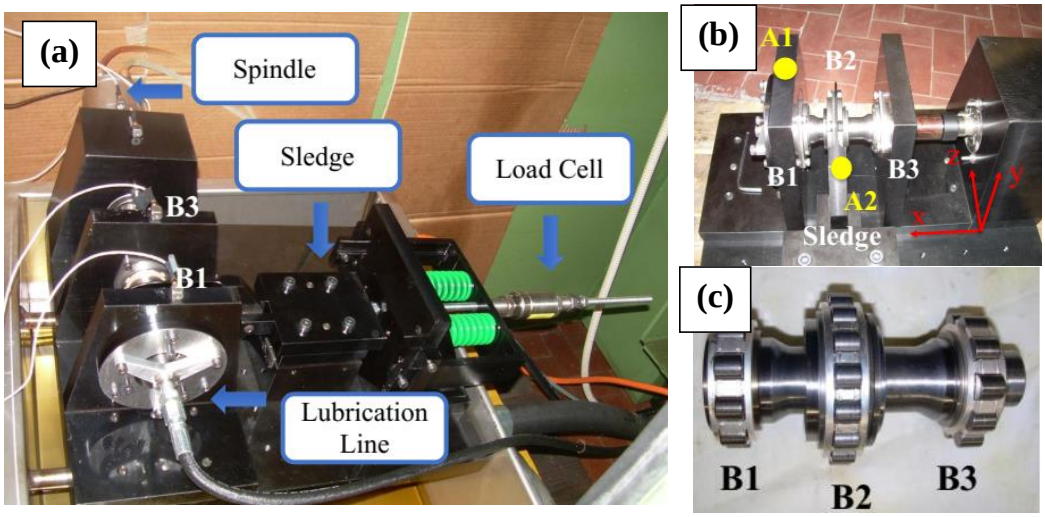

Fig. 5.1 **(a)** Test rig assembled at Politecnico di Torino (PoliTo) **(b)** positions of the two accelerometers (A1 and A2) **(c)** three bearings and the shaft (B1, B2 and B3)

5.2 Test rigs description

The experimental setup assembled in the lab of the DIRG in PoliTo is illustrated in Figure 5.1a and b. It is equipped with an electro-spindle, its power supply, three bearings B1, B2 and B3 (Figure 5.1c) and their supports, a load applying mechanism, a load cell and accelerometers to measure vibration signals. The electro-spindle has been selected so that high rotating speed up to 30000 rpm can be achieved. Rockwell tool has been utilized to create conical indentation on the inner ring or on one of the rollers. The size of the resultant localized faults is given by the diameter of the cones base (150, 250, 450 μm).

Two accelerometers A1 and A2 are mounted on the structure (Figure 5.1b). A1 is located on the support of the bearing B1 with damaged outer race or rolling element and A2 is located on the support of the larger bearing B2. Bearing B2 is dedicated to the application of the external load. The acceleration signals were acquired for about 10 second with sampling frequency of 51.2 kHz.

The IMS test rig is shown in Figure 5.2. The shaft is supported by four double row bearings and it is driven by an AC motor and coupled by rub belts. The speed of shaft rotation is retained constant to 2000 rpm and 6000 lbs load is applied to the system via bearings 2 and 3 by using a spring mechanism. Accelerometers are mounted on the housing of the bearings and record the signal with sampling frequency of 20.48

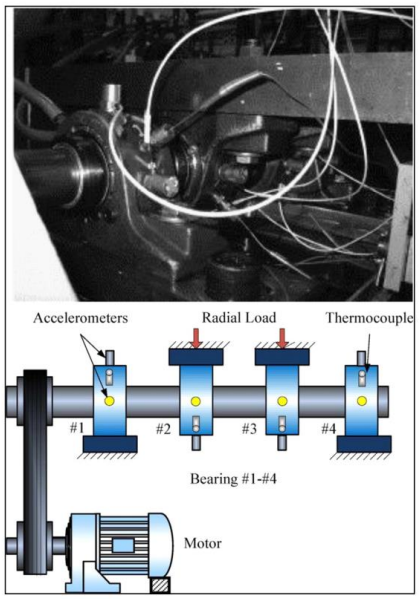

Fig. 5.2 IMS test rig [73]

kHz. Data are acquired during an endurance test and the experiment is stopped when the accumulated debris exceeds a certain level. A more detailed description of the experimental activity can be found in Ref. [71].

5.3 Results

5.3.1 Case 1

The data used for this case belongs to the IMS data set [72]. The test rig is shown in Figure 5.2. In this case the inner race defect occurred in bearing 3 and the signal is related to the accelerometer mounted on this bearing (first data set, channel 5 of data file 2003.11.25.15.47.32, the file's name indicates the date and hour it has been collected). The raw time signal is depicted in Figure 5.3 in which impulsive noise with relatively large amplitude can be detected.

The outcome of Fast Kurtogram is shown in Figure 5.4a and node with center frequency 8640 Hz and bandwidth 640 Hz at level 4 is selected as the filtered signal with highest kurtosis. The SES of the signal associated to this node is displayed in

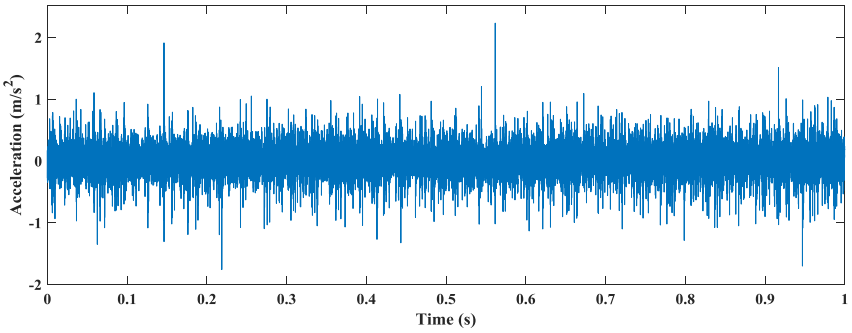

Fig. 5.3 Case 1, IMS data: acceleration signal (first data set, channel 5 of data file 2003.11.25.15.47.32)

Figure 5.4b but the inner race ball pass frequency (BPFI) is not detectable and the bearing is not diagnosable by using this method.

The Autogram and the SES of the selected node can be seen in Figure 5.4c and e. The selected node center frequency is 4800 Hz and bandwidth 640 Hz at level 4. The shaft frequency, BPFI and its sidebands spaced at shaft frequency can be spotted clearly which result in a successful diagnosis.

Figure 5.4e shows the Protrugram. As it was mentioned in subsubsection 2.6.2.2 this method is not blind and the bandwidth and step should be set in advance. For this case, the bandwidth and the step are equal to 4$\times$BPFI and 50 Hz and the highest kurtosis is achieved for 3600 Hz. The SES of the filtered signal for this center frequency and the selected bandwidth is depicted in Figure 5.4f.

The Enhanced Kurtogram is displayed in Figure 5.4g. The SES of the filtered signal related to the node with center frequency 4160 Hz and bandwidth 640 Hz at level 4 has the highest kurtosis. The SES in Figure 5.4h includes two large peaks which resulted in high kurtosis. The first one is related to the twice the shaft frequency and the second peak has the same frequency as the sideband of the BPFI. But, the BPFI has negligible amplitude and the diagnosis result is negative.

MED is another method whose performance is studied in this chapter. The signal filtered by MED and its SES are shown in Figure 5.5 a and b respectively and Figure 5.5c shows the SES of the raw signal. Although, a small peak at the BPFI can be detected in Figure 5.5b, by comparing the two SESs it can be concluded that the SES of the signal filtered by MED includes less diagnostic information than the raw signal.

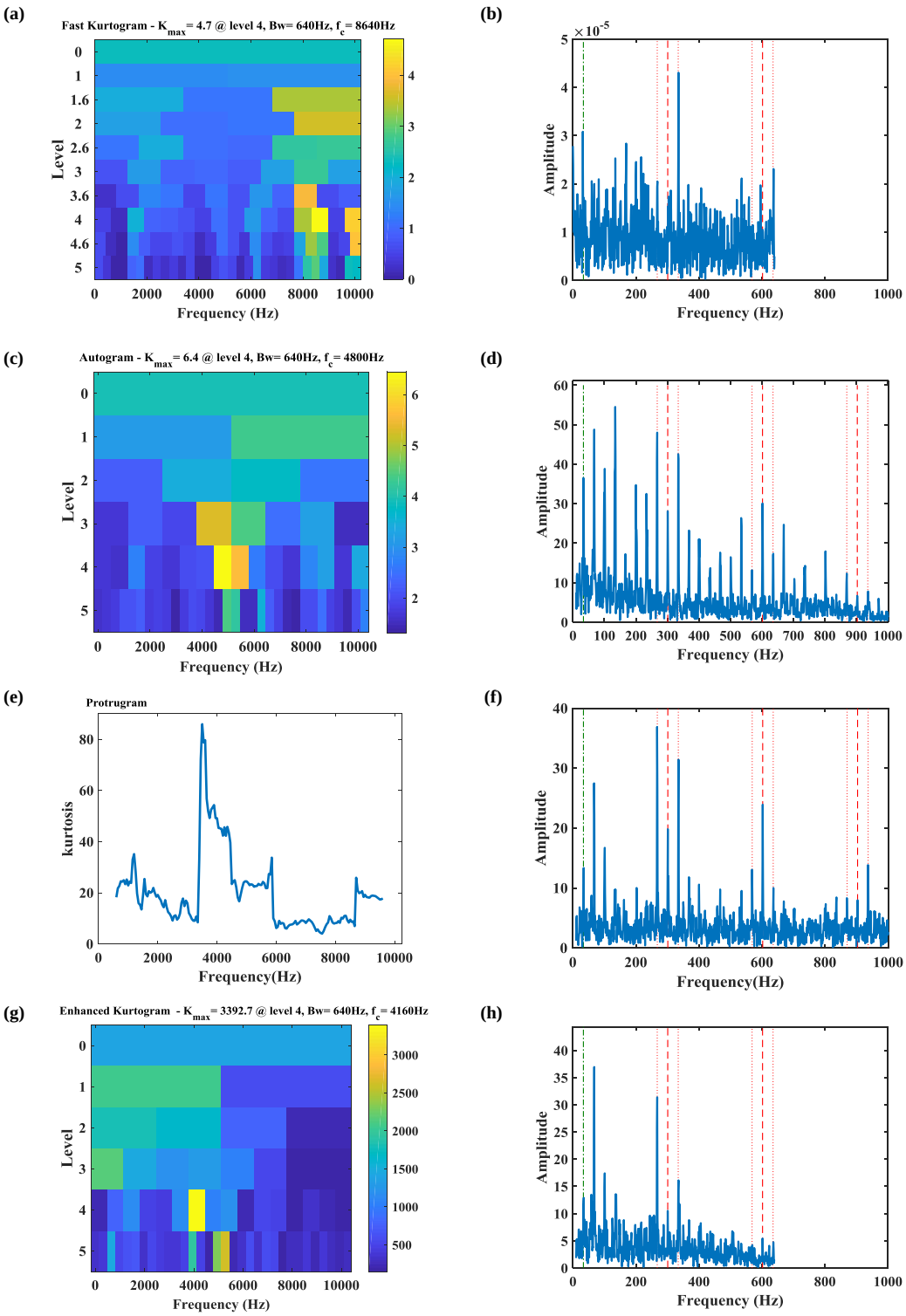

Fig. 5.4 Case 1, IMS data (a) Fast Kurtogram (b) Autogram (c) Protrugram (d) Enhanced Kurtogram; SES of the signal related to node with highest kurtosis in (e) Fast Kurtogram (f) Autogram (g) Protrugram (h) Enhanced Kurtogram

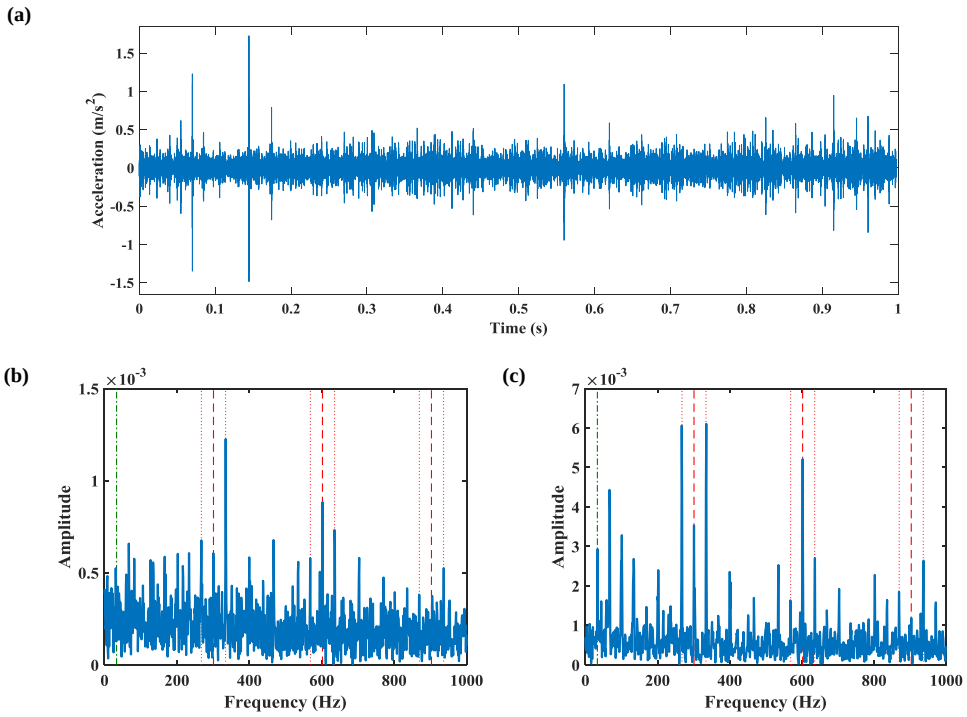

Fig. 5.5 Case 1, IMS data (a) Acceleration signal after applying MED filtering (b) SES of the signal filtered by MED (c) SES of the raw time signal

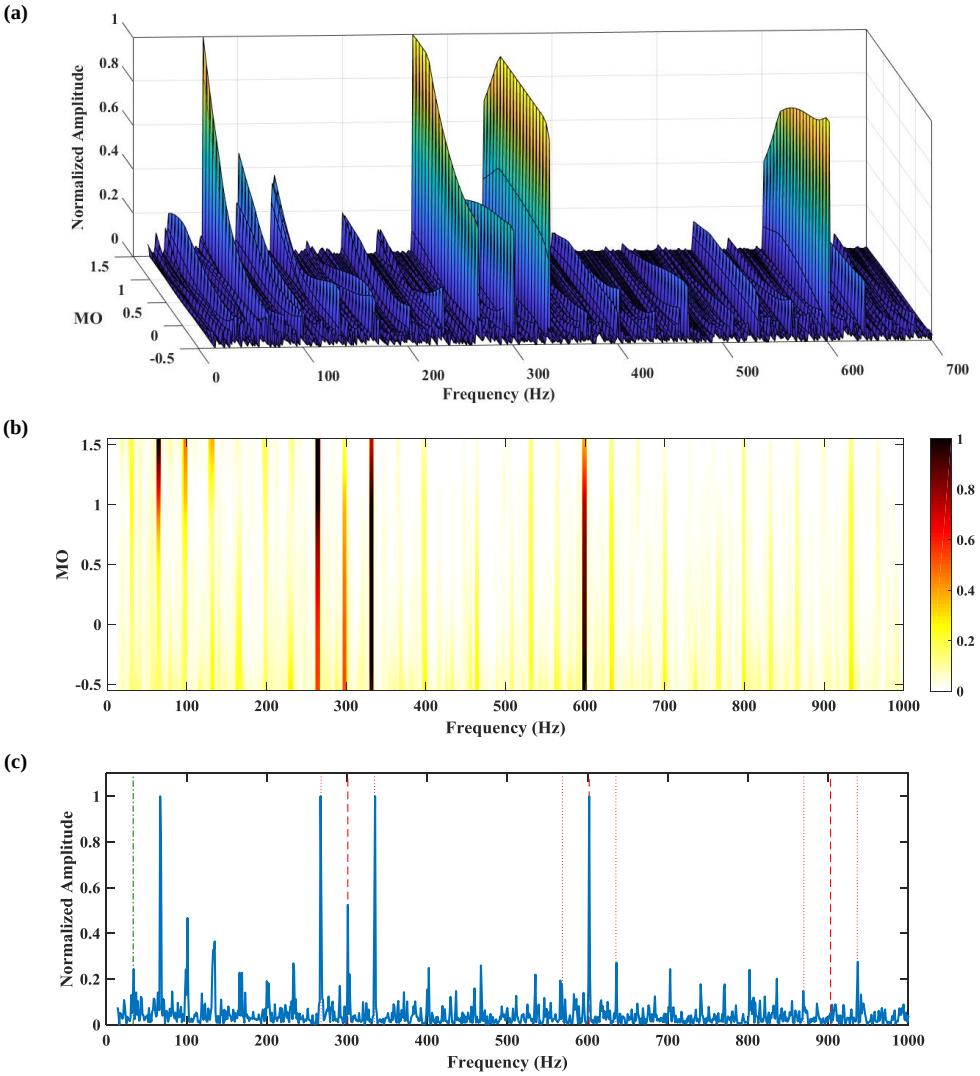

Fig. 5.6 Case 1, IMS data: the results obtained by proposed method in chapter 4 (a) 3D plot (b) above view (c) view along the x-axis (MO)

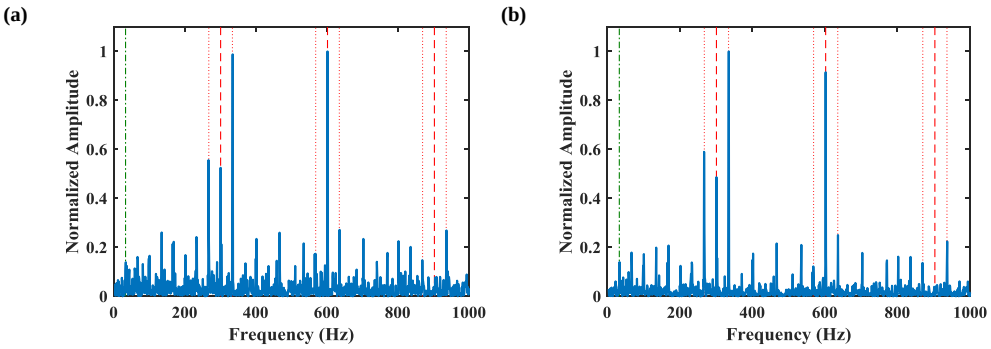

Fig. 5.7 Case 1, IMS data: SES of the modified signal for (a) MO = -0.4 (b) MO = 0 (cepstrum pre-whited signal)

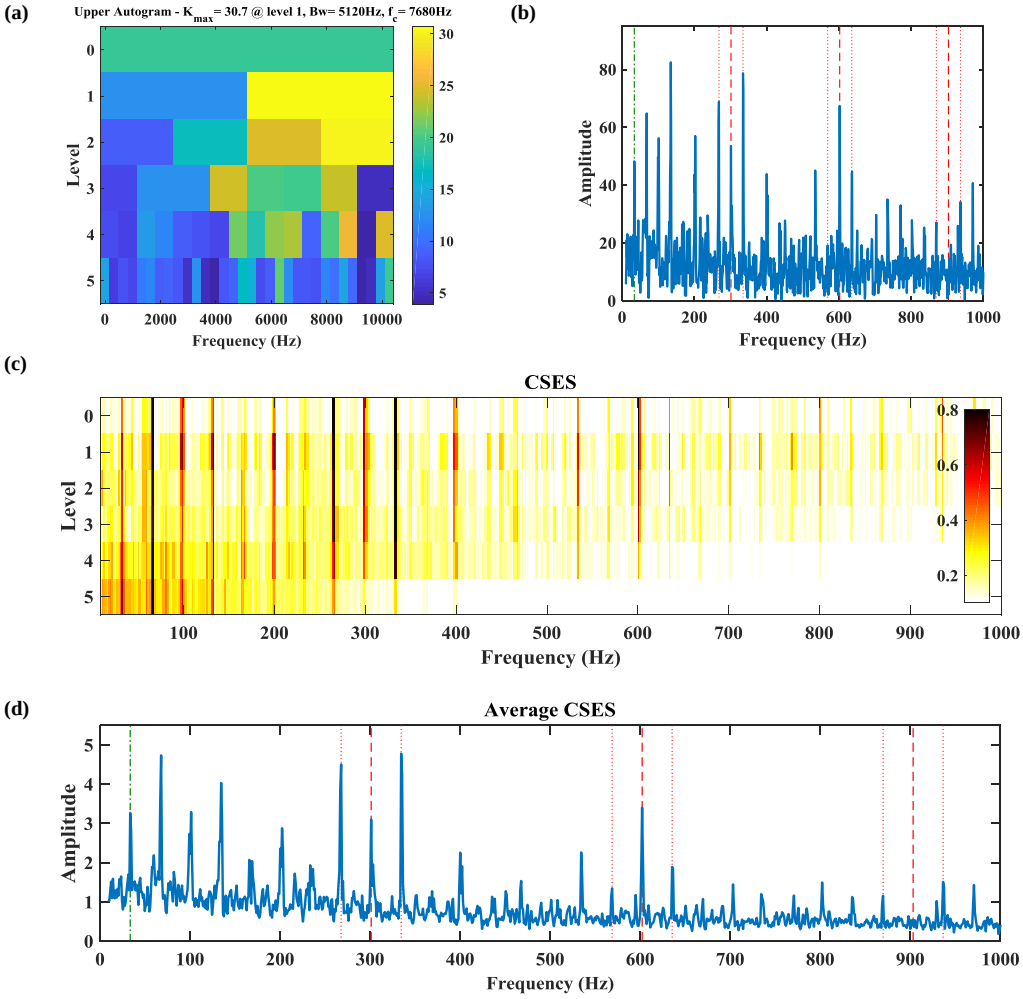

Fig. 5.8 Case 1, IMS data (a) upper Autogram (b) SES for node with the highest kurtosis (c) CSES (d) average CSES

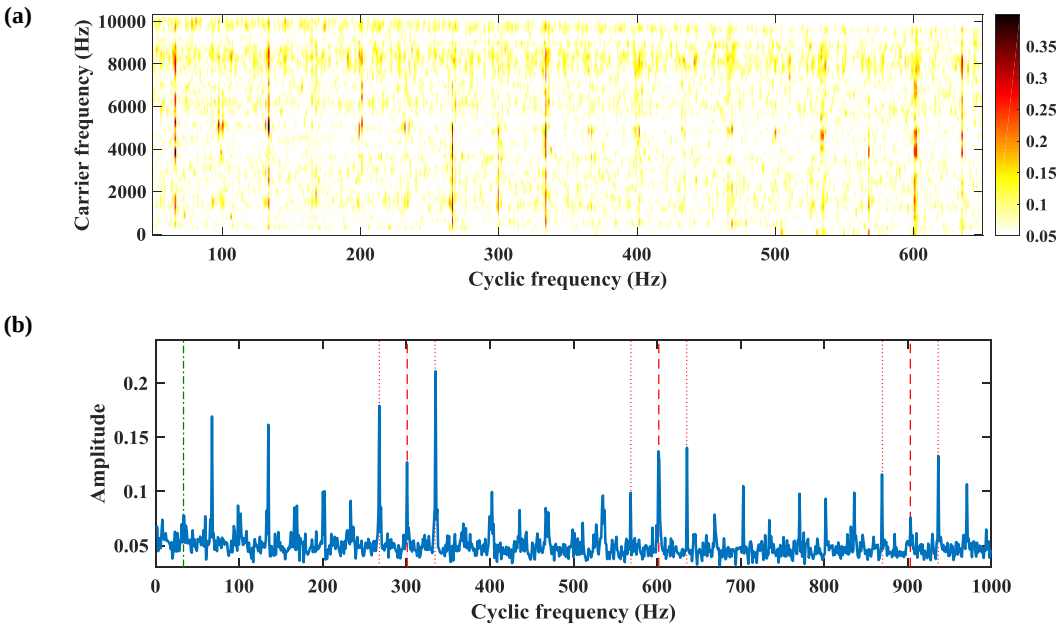

Fig. 5.9 Case 1, IMS data (a) fast spectral correlation (b) full band EES

The new method proposed in chapter 4 is also applied to this signal. The three-dimensional outcome is illustrated in Figure 5.6a where the x, y and z axis represent Magnitude Order (MO), cyclic frequency and normalized amplitude respectively. The views from above and along the MO axis of this plot are depicted in Figure 5.6 b and c respectively. The presence of the BPFI can clearly be detected for MO from -0.5 to 1 and its highest value is achieved by -0.4. The SES of the modified signals for MOs equal to -0.4 and 0 (CPW signal) are depicted in Figure 5.7a and b. Both SESs successfully provide diagnostic information as the BPFI besides its harmonic and sidebands can be spotted evidently.

Combination of upper Autogram and CSES is the next method will be investigated. The upper Autogram and the SES of its selected node are depicted in Figure 5.8 a and b. The defect frequency is present in the SES of the node with the highest kurtosis. Moreover, the resultant CSES and average CSES are shown in Figure 5.8 a and b and the CSES detects the BPFI throughout level 0 to 5.

SC is the last method implemented on this data. The result of Fast SC algorithm and full band EES are depicted in Figure 5.9 a and b. The BPFI and its harmonics and sidebands can also be detected in the EES.

It is noted that the second harmonic of the defect frequency has larger amplitude than the first one. From Figure 5.9a it can be inferred that the second harmonic consist some carrier frequency which do not exist for the first harmonic. In should be mentioned that the source of this phenomenon is unknown.

5.3.2 Case 2

Second case data is recorded from the test rig in DIRG lab. A bearing with a localized defect on its inner race, size of 250 µm, is placed at position B1 (see Figure 5.1) and the signal is acquired by the sensor mounted on the casing of the same bearing (sensor A1). The time signal is shown in Figure 5.10.

All the methods utilized previously are applied to this signal as well. For this case the bearing defect signal is weaker and it is also masked by signals from other sources.

Fast Kurtogram, Autogram, Protrugram and Enhanced Kurtogram are depicted in Figure 5.11 a-d respectively and the corresponding SESs are also shown in Figure 5.11 e-h. All approaches fail to find the effective demodulation band and none of the SESs include the bearing defect frequency. Fast Kurtogram and Autogram provide comparable results and the shaft frequency and its harmonics dominate the SESs. Protrugram and Enhanced Kurtogram highlight two different frequency bands which also do not comprise the defect signals.

To employ MED, the length of the FIR filter should be set in advance. The SES of the signal filtered by using a FIR filter with length of 30 is depicted in Figure 5.12a. SES of the raw signal is presented in Figure 5.12b with no signature of the BPFI. Consequently, it can be concluded that MED has effectively extracted the signal related to the defect as the defect frequency is clearly distinguishable. However, for a filter with length of 40 the outcome is completely different and the SES of the filtered signal in Figure 5.12c does not include the bearing characteristic frequency.

The outcome of the second novel approach explained in chapter 4 is displayed in Figure 5.13 a-c. Presence of the BPFI is noticeable for MOs ranging from -0.5 to 0.5 and its highest value is contributed to MO = -0.1. For this MO, SES of the modified signal is exhibited in Figure 5.13d and the promising diagnostic information reveals the potential of the proposed method to take out the defect signal from the raw signal.

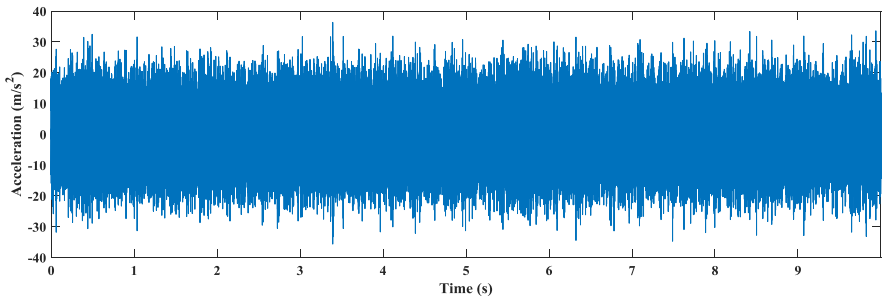

Fig. 5.10 Case 2, PoliTo data: acceleration signal

The upper Autogram and the SES of the node with the highest kurtosis are plotted in Figure 5.14 a and b respectively and the defect frequency is present in the SES. Moreover, the combination of upper Autogram and CSES clearly demonstrate the presence of the fault throughout the level 1 to 4 (Figure 5.14 c and d). Level 0 (raw signal) does not provide the defect signature and the shaft frequency prevails in the SES. Similarly, in level 5 selected nodes with high kurtosis only contain the shaft frequency and not the defect frequency.

The Fast SC and the full band EES are shown in Figure 5.15 a and b respectively. The presence of defect can be noticed from these figures but the BPFI has relatively small amplitude. The EES for selected band of carrier frequency [14 20.4] kHz where BPFI has higher amplitude (see Figure 5.15 a) is depicted in Figure 5.15c. In comparison to the full band EES, the EES for the selected band is more suitable as the BPFI is more evident.

5.4 Discussion

It should be mentioned that the case 1 is not difficult to diagnose as even from the SES of the raw signal (Figure 5.5c) the BPFI can be detected clearly.

Fast Kurtogram is generally a powerful method to extract the defect signal from background gaussian noise but its vulnerability to impulsive noise, which is common in industrial application, deteriorates its performance. This disadvantage causes the negative diagnostic of this case. In contrast, the Autogram successfully eliminates the effect of impulses by performing the ACs and provides positive diagnostic information.

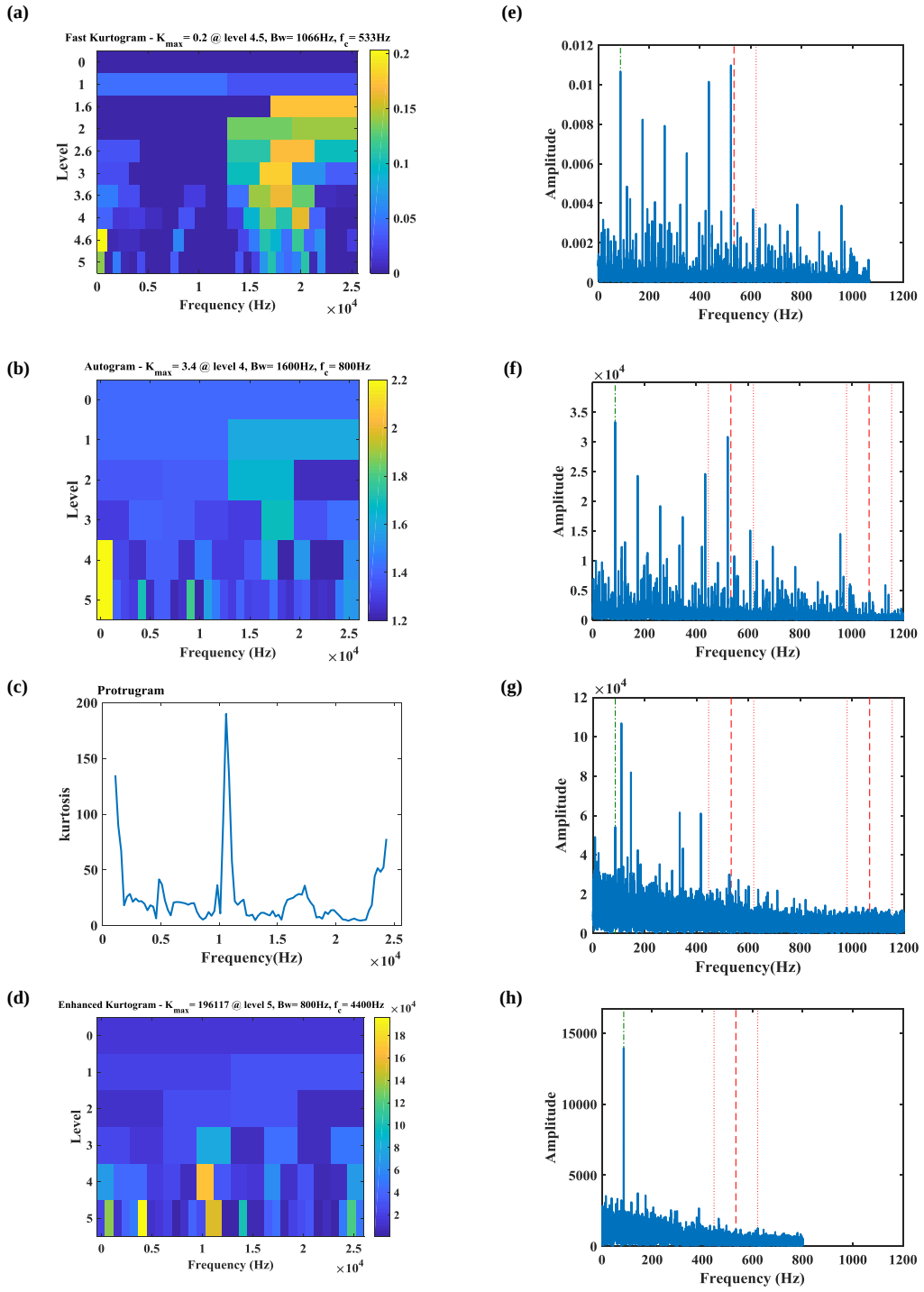

Fig. 5.11 Case 2, PoliTo data (a) Fast Kurtogram (b) Autogram (c) Protrugram (d) Enhanced Kurtogram; SES of the signal related to node with highest kurtosis in (e) Fast Kurtogram (f) Autogram (g) Protrugram (h) Enhanced Kurtogram

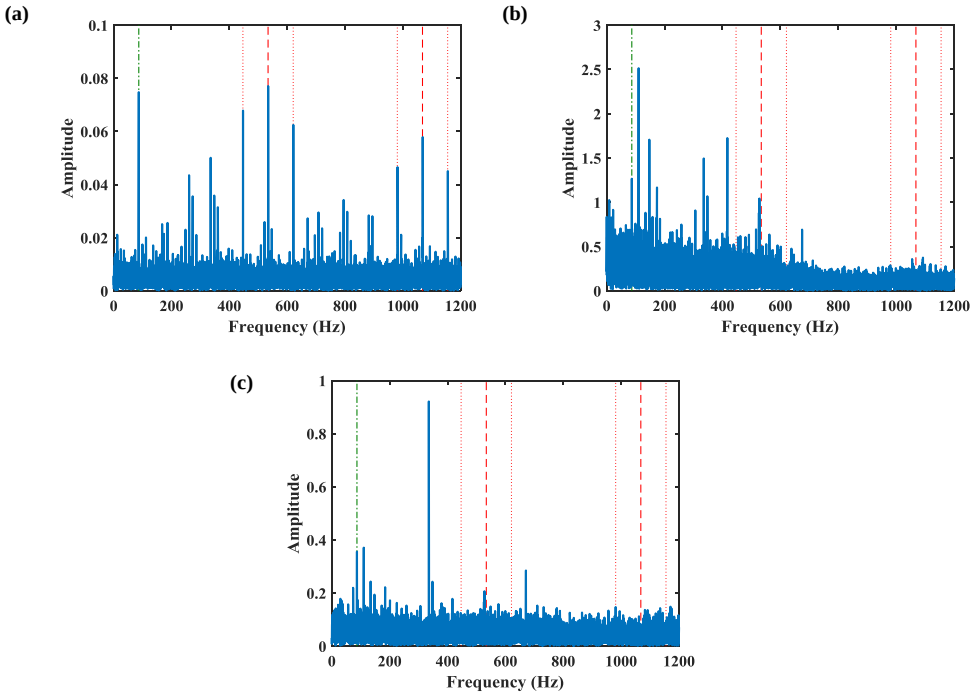

Fig. 5.12 Case 2, PoliTo data (a) SES of the signal filtered by MED by using a FIR filter of length 30 (c) SES of the raw time signal (a) SES of the signal filtered by MED by using a FIR filter of length 40

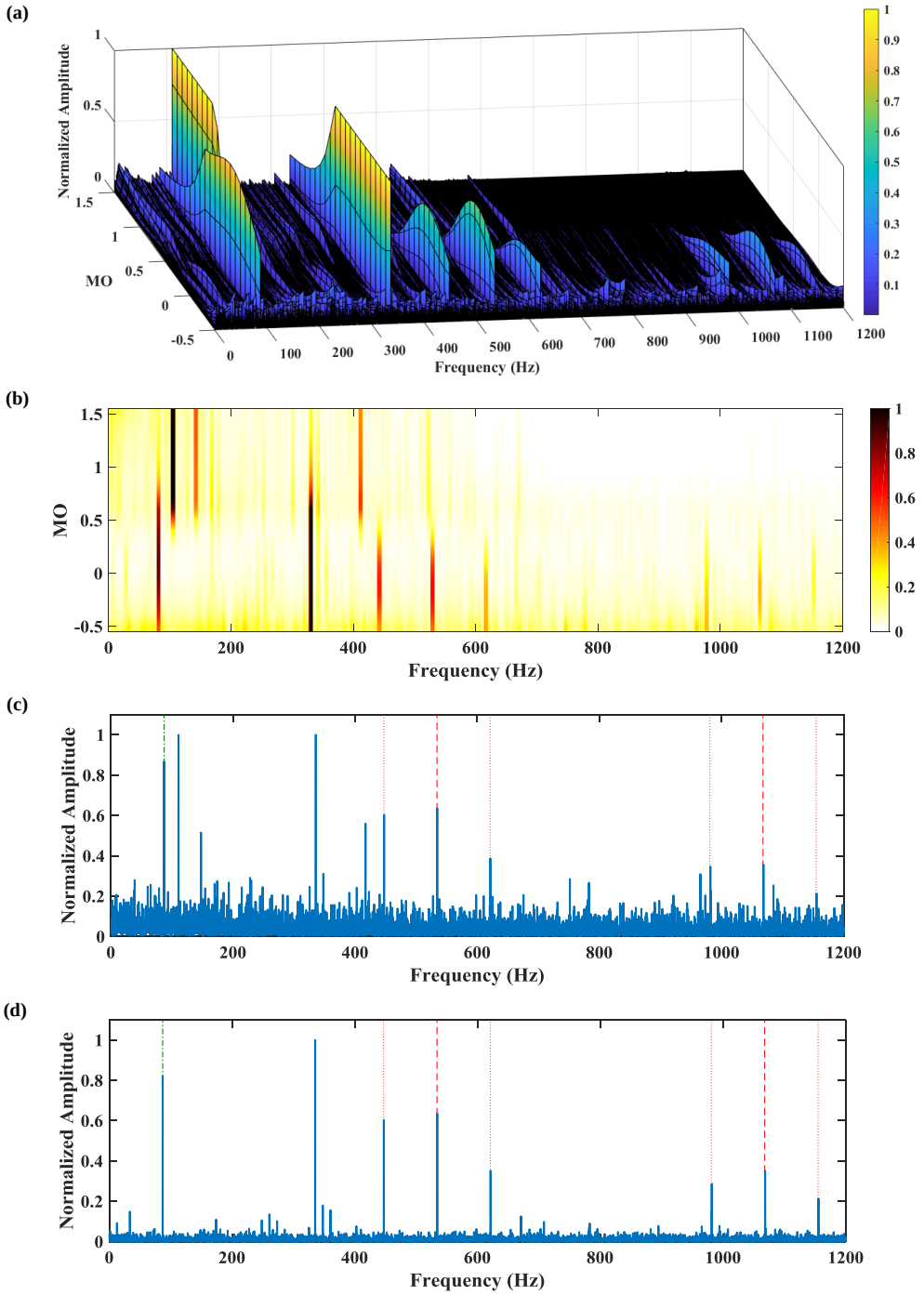

Fig. 5.13 Case 2, PoliTo data: the results obtained by proposed method in chapter 4 (a) 3D plot (b) above view (c) view along the x-axis (MO) (d) SES of the modified signal for MO = -0.1

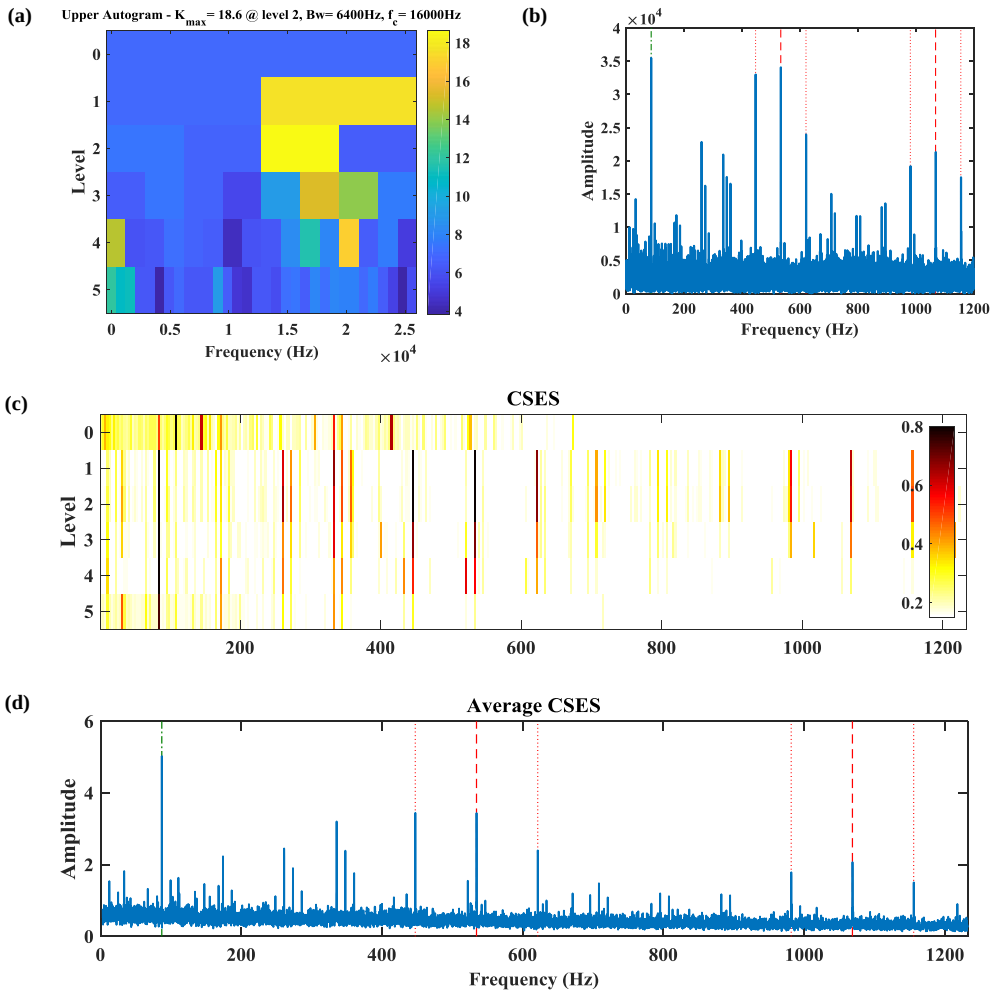

Fig. 5.14 Case 2, PoliTo data (a) upper Autogram (b) SES for node with the highest kurtosis (c) CSES (d) average CSES

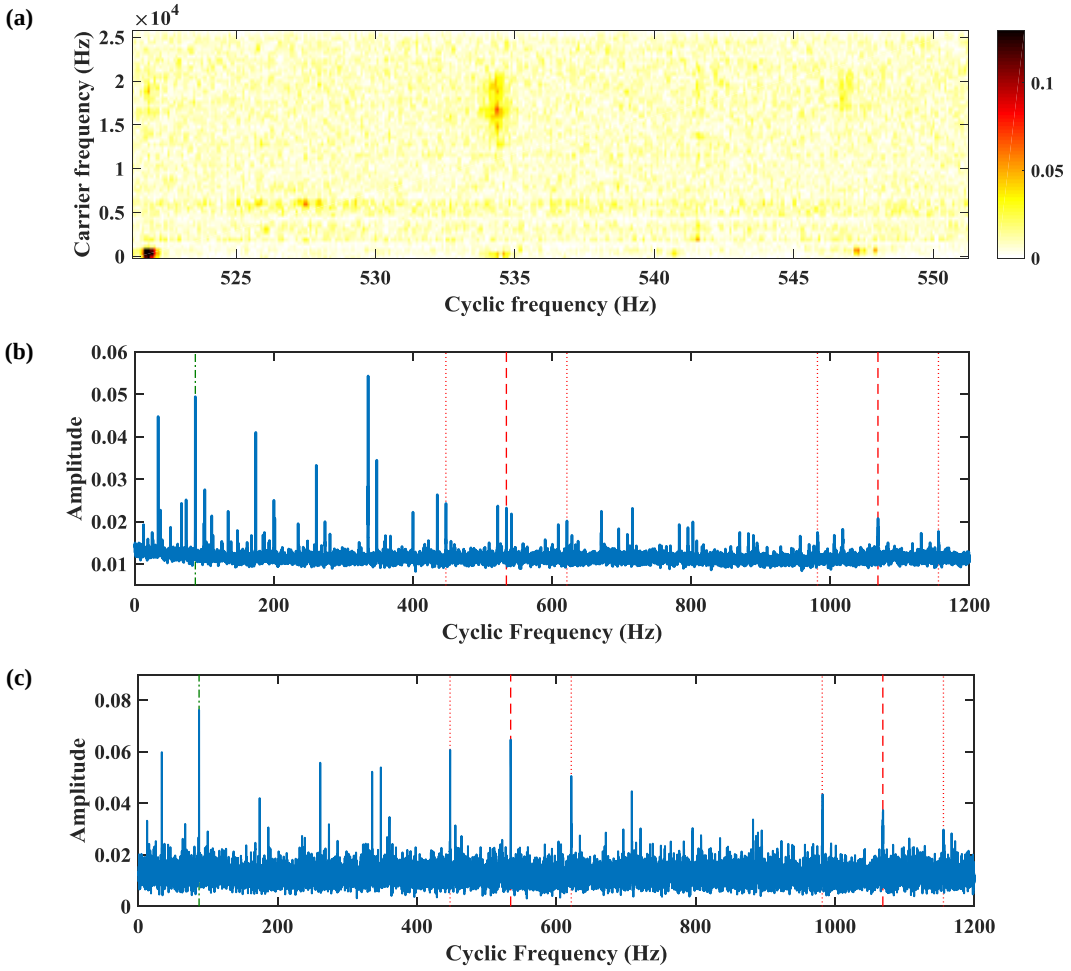

Fig. 5.15 Case 2, PoliTo data (a) fast spectral correlation (b) full band EES (c) EES for selected frequency band [14 20.4] kHz

Also, Protrugram is successful to find the proper frequency band for demodulation but EK fails. This can be attributed to the fix length bandwidth in Protrugram which also requires certain knowledge of the system. In this regard, Protrugram is not as universally applicable as FK and Autogram, which would have the best chance of success in blind applications. Another reason of EK failure is related to calculation of the kurtosis of the power spectrum of envelope signals (SES^2) instead of kurtosis of SES which is utilized in Protrugram. Additional drawbacks of Protrugram and EK is that they are not necessarily sensitive to impulsivity of signals in time domain as the kurtosis is computed in frequency domain which could also contain large peaks as a result of periodic signals.

As it was discussed, Protrugram is not blind and the parameters should be set in advance but there is no comprehensive guideline for selection of bandwidth and shift frequency. Nonetheless, practicality of Protrugram and EK are limited once defect frequencies are not dominant in SESs in comparison to discrete components generated by other sources such as shaft frequency.

For the second case none of these methods is able to select the frequency band related to the defect. These nodes (second half of the frequency range) have relatively high kurtosis in the FK, Autogram and EK but their kurtosis is not the highest. It is worth mentioning, upper Autogram effectively finds the frequency bands associated to the bearing defect signals for both cases.

Another point which can be mentioned here is related to the frequency range of the SES. A down-sampling process is performed at each level of FK and EK and consequently the frequency range of SESs is not constant and varies with the levels. For instance, the selected nodes by FK and EK for case 1 belong to the level 4, thus the whole frequency range is divided in 16 (2^4) sections. The maximum achievable frequency range for the SES is equal to 640 Hz (f_s/2/16) and it is only possible to detect the first harmonic of the BPFI. On the other hand, Autogram does not suffer from this drawback since down-sampling is neglected in the filtering process. The explained phenomenon is a disadvantage since the number of harmonics can be used as a fault severity indicator. This is more significant when diagnosis of bearings for a high-speed shaft is required, since in such cases the bearings characteristic frequencies are very high and might be out of the SES frequency range. By utilizing higher sampling frequency this problem can be limited to some extent but more storage space and in some situation new equipment are needed.

MED performance also has been studied for these two cases. For case 1, by comparing the SES of the raw signal and SES of the signal filtered by MED it was noticed that MED has canceled the bearing defect signals. As it was discussed, the data for case 1 is contaminated with impulsive noise. Additionally, MED tries to maximize the kurtosis and then it is inclined to highlight a single impulse over a series of periodic impulses. Consequently, those impulses with larger amplitude are mainly enhanced by MED and the bearing defect signal is deteriorated.

Another important consideration in performance of MED is linked to its sensitivity to filter length. As the solution of MED is achieved iteratively, it is not necessarily optimal; therefore, it may converge to different solutions for different preset parameters. The results for case 2 revealed this significant effect, for two different filter length of 30 and 40, the resultant signals are not the same.

So far, all the implemented methods assume that the bearings defect signals are the only (impulsive) component in a signal. But the presented results demonstrate that this assumption is troublesome while dealing with real case data which are combination of heterogeneous vibration sources such as periodic signals, stationary signals, periodically-modulated signals, repetitive transients, etc. In addition, even after performing random/deterministic separation techniques, deterministic components can be still present. For instance, Ref. [42] stated that even TSA, which is the most effective technique for this purpose, is not able to completely remove the deterministic parts. Also, Ref. [61] reported that their filtering efforts were not successful to remove the EMI, which has similar characteristics as bearing defect signals (second order cyclostationary), from their measured signals. Hence, a more reliable approach would be to extract all modulation frequencies included in a signal to eliminate the chance of neglecting desirable ones. As it was explained, this is particularly essential for practical applications since signals from various known and unknown sources and with different carrier frequencies are present. Moreover, relatively low energy of bearing faults makes them inferior to more powerful signals such EMI signals.

Two approaches proposed in this thesis (combination of Autogram and CSES and the second new method discussed in chapter 4) and combination of SC and EES try to generate more comprehensive diagnostic information by emphasizing all the cyclic (modulation) frequencies in a signal.

For both cases the second proposed method effectively reveals the bearing characteristic frequencies. For these two specific cases, SESs of CPW signals also contains proper diagnostic information. But the new approach provides more comprehensive outcome which contains almost all modulation frequencies even if some of them are not present in SESs of CPW signals.

For the second case all the methods with the purpose of finding the proper frequency band fail to detect the defect frequency. FK, Autogram and EK are able to highlight the suitable nodes but their kurtosis are not the highest. This was the primary incentive to propose CSES which combine SESs of all nodes with valuable information. Moreover, as it can be seen in FK and Autogram for case 2, for different bandwidth (levels) different frequency bands are emphasized; therefore, by using CSES all the modulation frequencies are revealed without the need to define any parameter in advance.

In addition, despite failure of other methods, upper Autogram has effectively detected the nodes with bearings defect signals. This shows the greater potential of this method to find nodes with impulsive contents in comparison to the original Autogram. As a result, CSESs are generated by employing upper Autogram. CSESs for case 1 and case 2 also suggests the upper Autogram has successfully extracted the frequency bands with valuable information for each level.

As it was discussed, CSES combine SESs for various bandwidth. It is advantageous since incipient faults may have very localized frequency bands; thus, a finer frequency resolution is required to detect them. On the other hand, larger defects are present in wider frequency range and their signature are better captured for coarser frequency bands.

Last but not least, SC and EES were presented. SC is a very powerful method for mapping a signal based on its carrier and cyclic frequencies. However, it may be more difficult to interpret than the SES. This problem has been solved by using EES which is an improved version of the SES.

For the case 1 the CSES and the full band EES provide comparable results. But when carrier frequencies for a specific cyclic frequency such as a bearing defect frequency is localized, first the signal should be filtered for that frequency band, otherwise diagnosis information of full band EESs would not be satisfactory. This could be noticed for case 2 where the BPFI is more noticeable in the EES of selected frequency than the full band EES. Also, it should be noted that the CSES is more

similar to the EES for selected frequency band. The better performance of the CSES over the full band EES is as result of selecting the nodes with valuable information and discarding others.

Overall, CSES can provide comparable and in some cases even better results than full band EES. On the other hand, EES for a selected frequency band where a specific cyclic frequency, such as bearing defect frequency, is present, may offers notable results even for difficult cases. The major disadvantage is that this frequency band should be found and set manually.

Part II

SIMULATION OF PLANETARY GEARBOX

Chapter 6

Review of planetary gearbox modeling

6.1 Introduction

Planetary gears, also recognized as epicyclic gears, are extensively used power transmission elements in numerous fields such as automotive, aerospace, wind turbines and marine applications. They have several benefits including compactness, high torque to weight ratio, high efficiency, multiple gear ratios and reduced noise in comparison to fixed-shaft gearboxes. Therefore, investigating planetary gear noise and vibration in healthy and faulty conditions is crucial to keep them functional, as well as to avoid any machinery breakdown as a result of a partial failure. A schematic of a planetary gear is shown in Figure 6.1.

Mathematical modeling is an advantageous approach to scrutinize characteristics of mechanical systems. It gives a good understanding of structure dynamic characteristics; it is reasonably accurate and suitable for evaluations during design stages. In this regard, mathematical models such as lumped parameter models (LPMs) have been vastly used to study modal properties and also vibration signals of planetary gear trains.

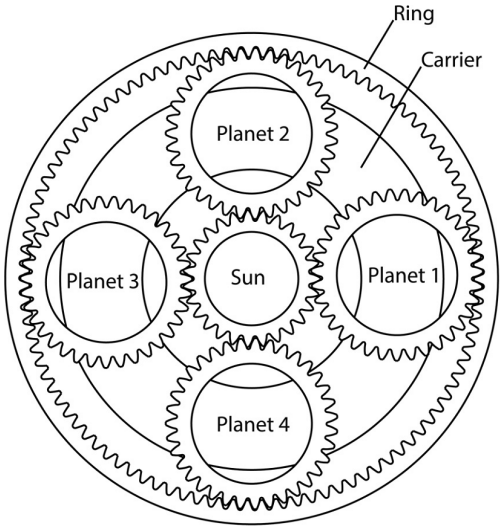

Fig. 6.1 Schematic of a planetary gear with four planets [74]

6.2 Modeling of planetary gear sets

Cunliffe et al. [75] developed a two dimensional mathematical model for a planetary gearbox with 13 degrees of freedom and studied the natural frequencies and vibration modes of the system. Modes were grouped into low frequency "bearing modes" and high frequency "tooth modes," but they did not identify the vibration structure (see Figure 6.2).

Botman [76] analyzed the effect of planet-bearing stiffness and rotation of the carrier on the natural frequencies of in-plane vibration of a single stage spur planetary gearbox with eighteen degrees of freedom. He categorizes the vibration modes as "axisymmetric" and "nonaxisymmetric". In the axisymmetric or rotational modes, all planets perform the same motion with respect to the sun and the other components have only rotational vibration and nonaxisymmetric modes are due to lateral motion of other components of the gearbox, e.g. ring or sun gear.

Frater et al. [77] extended Botman's model to take into account the unequal mesh stiffness between the sun-planet and ring-planet meshes. Also, stiffness alternation due to variation of number of teeth in contact can be considered in the model. The effect of unequal and constant mesh stiffness on natural frequencies and vibration modes of planetary gears is investigated and good agreement between the resonant frequencies of the model and the dynamic system was achieved.

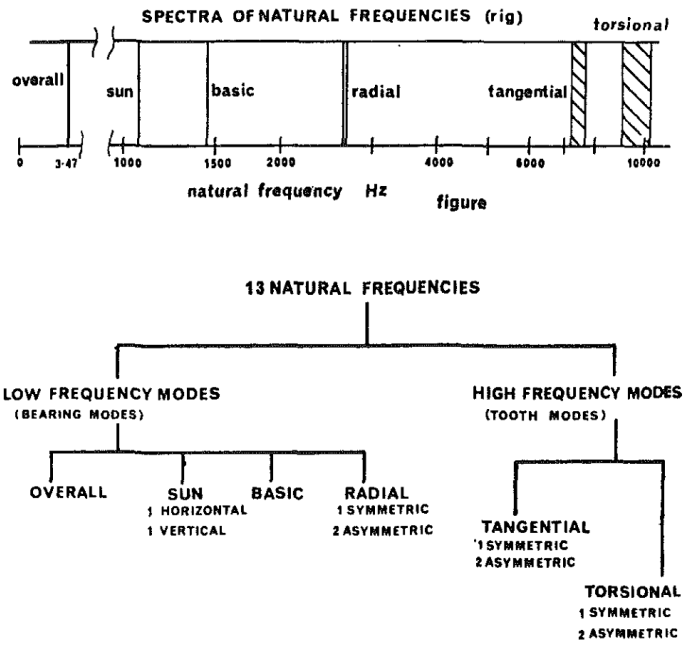

Fig. 6.2 System natural frequencies [75]

The tooth mesh stiffness, planet bearing stiffness, and ring support stiffness influence on the natural frequencies of planetary gears was further studied by Ref. [78] by using finite element method.

Kahraman derived closed form expressions for natural modes of a planetary gear by using a purely torsional model [79] and concluded that the torsional model, given its simplicity, provides accurate result and therefore can be utilized to study complicated phenomena in planetary gears.

Lin and Parker [80] developed an analytical model by including radial and tangential planet deflections for each component, and rigorously investigate main properties of natural frequencies and vibration modes of a general planetary gear system with equally spaced planets which is applicable to various gears configurations. They showed that planetary gears, independent of the number of planet gears and the system parameters, have three types of vibration modes- planet, rotational, and translational modes. Their 2D planar lumped-parameter planetary gear model is shown in Figure 6.3.

They later studied the free vibration of this set with unequally spaced planets [81] and it was shown that the modal properties of the systems are comparable to the equally spaced planet systems.

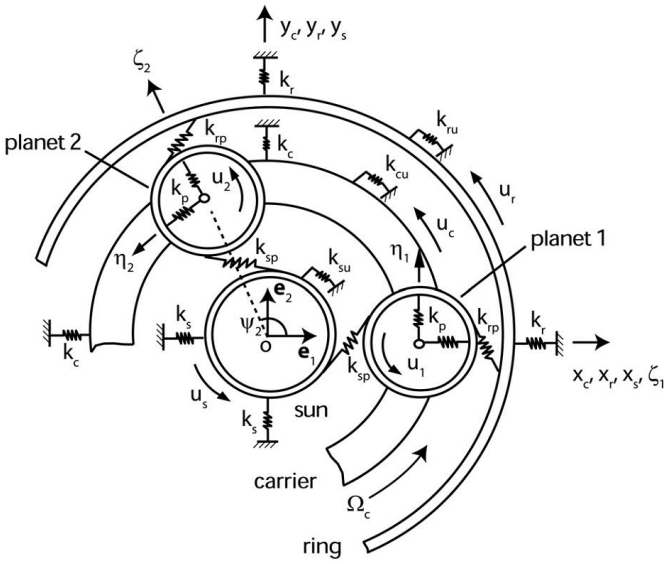

Fig. 6.3 Lumped-parameter planetary gear model from Ref. [80]

Varying contact conditions at the gear tooth mesh is the main excitation source in gear sets. For instance, excitation of the system close to its resonances causes large vibration which leads to contact lose, i.e. mesh stiffness vanishes suddenly and strong nonlinearity is introduced to the system. Therefore, to have more realistic model it is necessary to model the gear mesh accurately. August and Kasuba [82] analyzed the nonlinearity due to gear mesh stiffness variation. Also the effect of the floating sun gear (zero translational stiffness) on the general performance of the system was studied.

Kahraman [83] developed a nonlinear time-varying dynamic model of a planetary gear set with an arbitrary number of pinions which includes tooth separations and mesh stiffness fluctuations and investigated the issue of dynamic load sharing among the arbitrarily spaced pinions.

Sun and Hu [84] developed a lateral-torsional coupled nonlinear dynamic model for a planetary gear system with multiple clearances by using the harmonic balance method. They also discussed the effects of the variation of mesh stiffness and static transmission errors on the nonlinear dynamics behavior of the system.

Ambarisha and Parker [85] used both lumped parameter model and finite element model to examine tooth separation nonlinear dynamic behavior of spur planetary gears. They conclude nonlinear jumps, chaotic motions, and period-doubling bifur-

cations occur when the mesh frequency or any of its higher harmonics are near a natural frequency of the system.

When a gear tooth is in contact on the drive-side and back-side simultaneously, tooth wedging or tight mesh occurs. Guo and Parker [86] introduced tooth wedging, tooth contact loss and bearing clearance into a lumped parameter model and investigated the interplay between tooth wedging and bearing clearance. It was shown that tooth wedging is a possible source of bearing failure as it elevates planet bearing forces considerably and destroys load sharing among the planets. They also [87] studied nonlinear behavior, bifurcations and chaos caused by bearing clearance as well as interaction between tooth separation and bearing clearance by applying harmonic balance method with arc length continuation to lumped parameter model.

The effect of gravity on the dynamic response of wind turbines was studied by Guo et al. [88]. It has significant effect on the dynamics of wind turbines, i.e. dynamic response is affected by the stiffening effects which is caused by tooth wedging and bearing-raceway contacts at vibratory resonances.

6.2.1 Modeling of gear faults

Simulation has also been used to model gear faults due to its inherent advantages such as deeper insight into the signatures of different faults on the system vibration signals, the reduction of number of experiments and testing of fault diagnostic and prognostic techniques.

Chaari et al. [89] studied the influence of tooth defects (pitting and crack) on the response of the system by modifying the gear mesh stiffness. These two defects were modeled by a phase shift and amplitude reduction of gear mesh stiffness (see Figure 6.4).

Chen and Shao [90–92] presented a series of papers on the effects of gears cracks on planetary gear dynamics. The mesh stiffness model, derived based on the potential energy principle, is incorporated into the dynamic model of a planetary gear set with 21 DOF to investigated the effect of the rigid ring gear tooth crack and the crack size on the dynamic response of the planetary gear [90]. Time and frequency response of a system with cracks with different sizes and inclination angles on sun and planet gears teeth was further studied by Ref. [91] and growth in sidebands around the mesh frequency and its harmonics in frequency spectrum of the simulated results

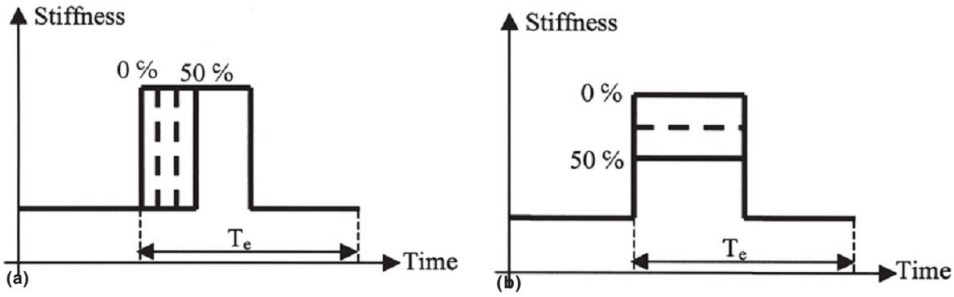

Fig. 6.4 Stiffness changes in gear mesh stiffness model (a) Tooth flank pitting (b) Tooth cracking [89]

was observed. Moreover, Chen et al. [92] studied the influence of flexible ring gear rim and cracked tooth on mesh stiffness and dynamic features of a planetary gearbox. The mesh stiffness was computed by an analytical model, developed based on the potential energy principle and uniformly curved Timoshenko beam theory.

Bahk and Parker [93] investigated the influence of tooth profile modification on nonlinear dynamics of spur planetary gears by employing Perturbation analysis which yields a closed-form approximation of vibration response and an analytical model for the tooth profile modification. Also, impact of the system parameters, such as mesh stiffness amplitude fluctuation and variation of relative mesh phase between sun–planet and ring–planet meshes, on the tooth profile modification and dynamic response of the set was determined.

6.2.2 Deformable models

Assumption of rigid gear bodies may not provide accurate results under many situations such as high-speed gears therefore the elastic vibration of gear bodies should be considered [74]. Wu and Parker [94] expanded the lumped parameter planetary gear model from Ref. [80] and derived the characteristic modal properties of planetary gears having equally spaced planets and an elastic continuum ring gear by using thin ring theory (Figure 6.5a). According to the closed-form results achieved by Perturbation method, all the vibration modes based on their unique characteristics are classified into four different types: planet, rotational, translational, and purely ring modes.

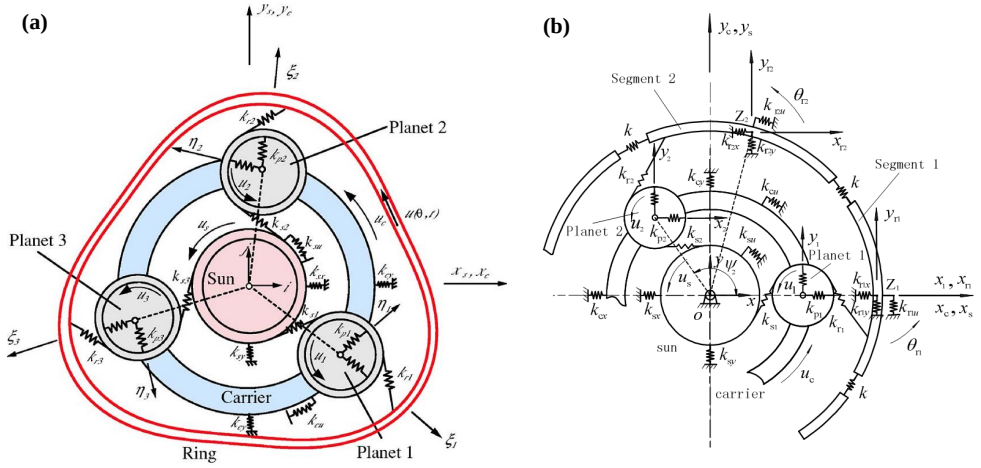

Fig. 6.5 (a) Elastic-discrete model of a planetary gear and corresponding system coordinates (The distributed springs around the ring circumference are not shown) [94] (b) discrete model of a planetary gear train with flexible ring [95]

Zhang et al. [95] calculated natural frequencies and corresponding vibration modes of a planetary gearbox with flexible ring by dividing the continuum ring into finite rigid sectors (Figure 6.5b). The equations of motion of each individual component are assembled to produce the governing equations of planetary gear system. Finally, the eigenvalue problem is solved to obtain the natural frequencies and corresponding vibration modes. To determine the vibration signatures of localized planet-bearing faults, Jian [96] developed an analytical model including flexible ring gear.

6.2.3 Transmission path effect

In a planetary gearbox there are multiple vibration sources and, due to rotation of carrier, the transmission path of vibration signals varies. This causes amplitude modulation (AM) of vibration signals which results in modulation sidebands in the frequency domain.

Inalpolat and Kahraman [97] proposed an analytical framework to predict these sidebands due to AM of signals from different sources for five configuration of gears. Number of planets, planet position angles and planet phasing relationships were considered in the model. They also validated their proposed analytical model experimentally.

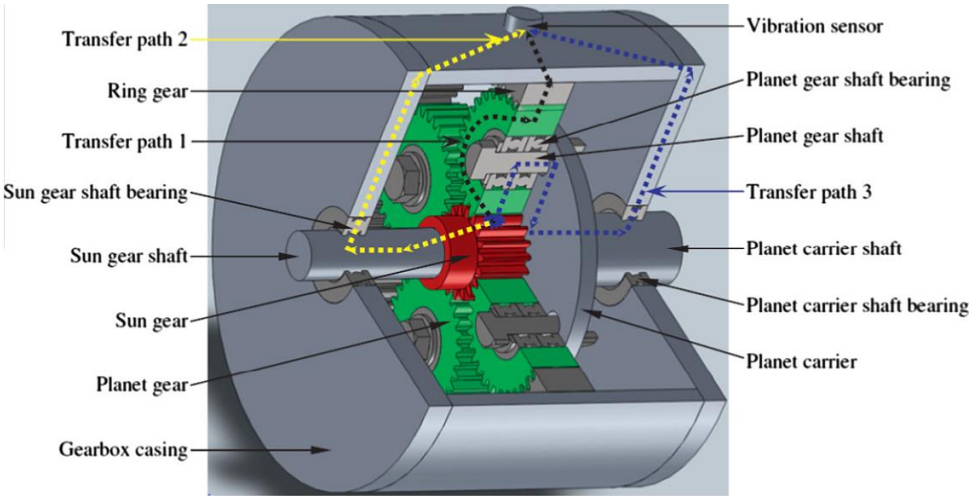

Fig. 6.6 Possible transmission paths in a planetary gearbox [100]

Inalpolat and Kahraman further [98] proposed a nonlinear time-varying dynamic model and studied the frequency modulation (FM) of the vibration signals due to gear manufacturing errors in the forms of run-out or eccentricity and variation of the instantaneous gear mesh locations.

Feng and Zuo [99] modeled the vibration signals of planetary gear sets, considering both the AM and FM effects as result of the either local or distributed gear damage and periodically time-variant operational condition (speed or load), in addition to the effect of vibration transmission path.

Liu et al. [100] proposed a model to investigate the effects of the transmission path on the acquired signal on the casing of a planetary gearbox. The transmission path includes two parts, first the transmission path inside the gearbox from each gear to the casing, second the transmission path along the casing to the transducer position (see Figure 6.6). Furthermore, they investigated influence of different transmission paths and properties of the resultant vibration signals and validated their proposed model with experimental data.

Chapter 7

Simulation of Planetary Gearbox with Localised Bearings and Gears Faults [1]

7.1 Introduction

This chapter develops a lumped parameter model to investigate the gears and bearings interaction of a planetary gear train in presence of faults. A 18-DOFs model of planetary gearbox, based on [102], is combined with a comprehensive bearing model. The resulting set of equations has the capability of simulating the behaviour of the system for different sizes, locations and profiles of defects, both on gears and on bearings, and also includes nonlinear effects due to the Hertzian contact assumption. The purpose is to build a numerical tool where faults, different in type, size and position can be implemented to produce time domain signals. The aim is not to exactly reproduce the behaviour of a real gearbox, or even of a test rig, but to produce numerical simulations of faults which are an important step in testing and developing diagnostic techniques. Numerically generated signals, with specific characteristics directly related to particular defects, can provide an efficient and very economical way to verify and enhance the capabilities of data processing methods. In real life, the signature of the defect can in fact be masked by external noise, making it impossible

[1] "Part of the work described in this chapter has been previously published in Ref. [101]"

to figure out what the real contribution of the defect on the measured output is; and even more difficult is to have at disposal a test rig with known and controlled defects.

Furthermore, this chapter comprehensively discusses the frequency components of signals associated to planet-bearing defects for inner/outer race and rolling element and sources of sidebands around bearing damage frequencies. Finally, frequency analysis and statistical features are conducted on dynamic response of the planetary gear to study presence and growth of the bearings faults.

7.2 Mathematical Model

7.2.1 Lumped Parameter Model for Planetary Gear Sets

The primary aim of this simulation is to calculate vibration signals of each gear of a planetary gearbox in presence of defects on bearings and gears. The two-dimensional lumped parameter models by Lin and Parker [80] and Liang et al. [102] are the basis for the present discussion. There are some distinctions between these two models: first, configurations of planets coordinates are different; second, damping is not considered in the Lin and Parker [80] model. Whether planets deflections are described cartesian coordinates [102] or in polar coordinates [80], the dynamic behavior of this gearbox will not change. The coordinates in this chapter are based on Ref. [102] which simplify the final expression of the equations of motion.

A 2D lumped parameter model for a spur planetary gearbox is illustrated in Fig. 7.1. Each gear (sun, ring and planets) and carrier are considered as rigid bodies with three degrees of freedom (DOFs) - one rotational (θ) and two transverse motions in the x and y directions. The resulting Lumped Parameter Model (LPM) has $3(N+3)$ DOFs where N is number of planets. The degrees of freedom of this system are: $[x_c,\ y_c,\ \theta_c,\ x_r,\ y_r,\ \theta_r,\ x_s,\ y_s,\ \theta_s, x_{pj},\ y_{pj},\ \theta_{pj}]$ where c, r, s and $pj,\ j = 1,\ \ldots,\ N$ are assigned to the carrier, ring, sun and planet gears, respectively. The flexible gear teeth contacts are modelled by springs and dampers acting along the gear line of action and tooth contact loss is assumed not to occur. The gears translational displacements are calculated with respect to a rotating frame of reference Oxy with origin O, the center of the planetary gear set. This frame is attached to the carrier and rotates with the same angular speed as carrier. The angular rotation θ is defined in OXY reference system which is fixed and is not rotating. Furthermore, all gears are

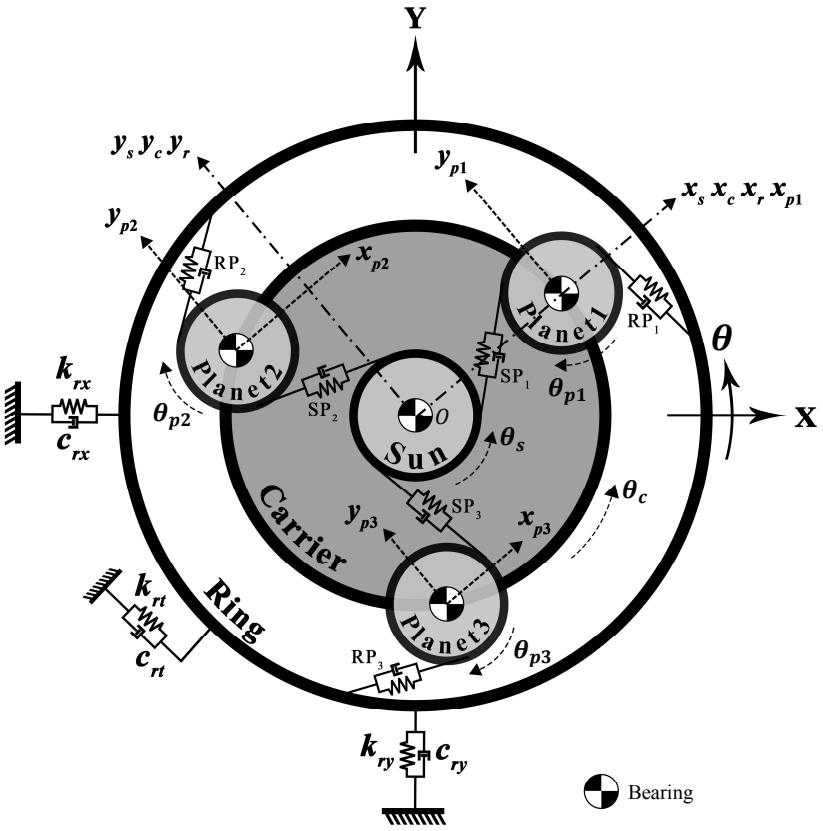

Fig. 7.1 Lumped parameter model of a single stage planetary gear set and its corresponding system coordinates. $\mathrm{SP}_i = [k_{spi},\ c_{spi}]$ and $\mathrm{RP}_i = [k_{rpi},\ c_{rpi}]$ for $i = 1,2,3$ indicates the flexible contacts between sun-planet and ring-planet respectively

assumed perfect without manufacturing and mounting errors. Gears and carriers are also considered free of eccentricities and roundness errors. According to Jian [96], the flexibility of the ring gear influences the relative amplitudes of the sidebands in a fault signature and higher-order sidebands disappear when ring thickness increases. The ring gear herewith described is not deformable but a development of the model is planned so to connect the ring gear, the bearing of the sun gear and the bearing of the carrier to a flexible gearbox casing, to simulate an elastic support and the path from the defect to the measurement point.

7.2.2 Time-Varying Mesh Stiffness

The condition in which gear teeth are in contact varies as they rotate. For a contact ratio lower than two, the number of teeth pairs in contact periodically changes from one to two and this causes a time variation of mesh stiffness, which is the main source of vibration in gearboxes. Mesh stiffness also changes with the contact positions of gear teeth. Teeth pairs enter and exit the mesh constantly and therefore the assumption of equivalent characteristics for every tooth leads to periodic time varying mesh stiffness.

Yang and Lin [103] proposed a potential energy method to calculate the effective mesh stiffness and showed that the energy can be divided in three parts: Hertzian, bending and axial compressive energy. These energies can then be used for the calculations of Hertzian contact stiffness, bending stiffness and axial compressive stiffness, respectively. Tian [104] refined this model by adding the shear energy as the fourth part of the potential energy. This analytically obtained time-varying mesh stiffness can represent the effect of changes in the number of teeth pairs in contact and the contact positions between teeth of engaged gears. He also proposed a method to determine the mesh stiffness for different sizes of the crack on the root of a gear tooth. Similar effects on the mesh stiffness can be produced by tooth profile modifications, as highlighted by Chen and Shao [105]. Finally, Iglesias et al. [106] proposed an advanced model for calculation of internal and external gears meshing forces in spur gear planetary. It is then possible to introduce in the model various forms of defects, simply by a proper definition of the mesh stiffness.

The meshing frequency in a gearbox is the frequency at which gear teeth mate together. When the ring gear of a planetary gearbox is fixed, for every complete revolution of the carrier a planet tooth meshes N_r times with the ring gear teeth (N_r is the ring's number of teeth). Therefore, the meshing frequency can be calculated as follows:

$$f_m(\text{Hz}) = \frac{N_r \omega_c}{60} \tag{7.1}$$

where ω_c denotes the carrier angular velocity (rpm). Furthermore, the angular displacement of the planet gear in one mesh period θ_m can be calculated by the following equation

$$\theta_m = \omega_p T_m = \frac{\omega_p}{f_m} = \frac{2\pi(N_r - N_p)}{N_r N_p} \tag{7.2}$$

where T_m is the meshing period.

In a planetary gearbox several gears with different number of teeth are in contact at the same time (see Fig. 7.1). All planets are assumed to be identical, thus the behavior and periodicity of each ring-planet or sun-planet meshing are similar although it should be considered these are in different phases with each other and it is also true that each planet's contacts with the ring and sun gears are dissimilar in phase [107]. The contact stiffness related to the generic n^{th} planet is

$$k_{upn}(t) = k_{up1}(t - \gamma_{un} T_m) \qquad u = r,s \quad \text{and} \quad n = 1,\ldots,N \tag{7.3}$$

where γ_{rn} and γ_{sn} are the relative phases between the nth ring-planet and sun-planet meshes, respectively. In addition, γ_{rs} represents relative phase between the ring-planet and sun-planet meshes which is identical regardless of which planet is considered [107]. An example of mesh stiffness is given in subsection 7.5.1.

7.2.3 Bearings Model

The outer race, inner race, cage and rolling elements are the key components of a bearing. Fig. 7.2 presents a sketch of the multi-body nonlinear dynamic model that will be used to simulate the vibration response of the planetary gearbox presented in subsection 7.2.1. This model was originally developed by Refs. [108, 109] which consider two degrees of freedom for the inner race with a fixed outer race. However, in the present model (Fig. 7.2) two extra degrees of freedom for outer race are also assumed; as a result, four degrees of freedom of the model comprise the inner raceway displacements x_i and y_i and the outer raceway displacements x_o and y_o. Sawalhi and Randall [17] introduce another DOF and they tune its parameters so to represent an high frequency behaviour of the bearings, in accordance with their experimental results. This DOF would add a new equation for each bearing but would not alter the structure of the equations of motion given in section 7.4. The particular aspect of the flexibility of the bearings, which generates a well separated resonance in the spectra as reported in [17], is not the focus of the present study and has been neglected.

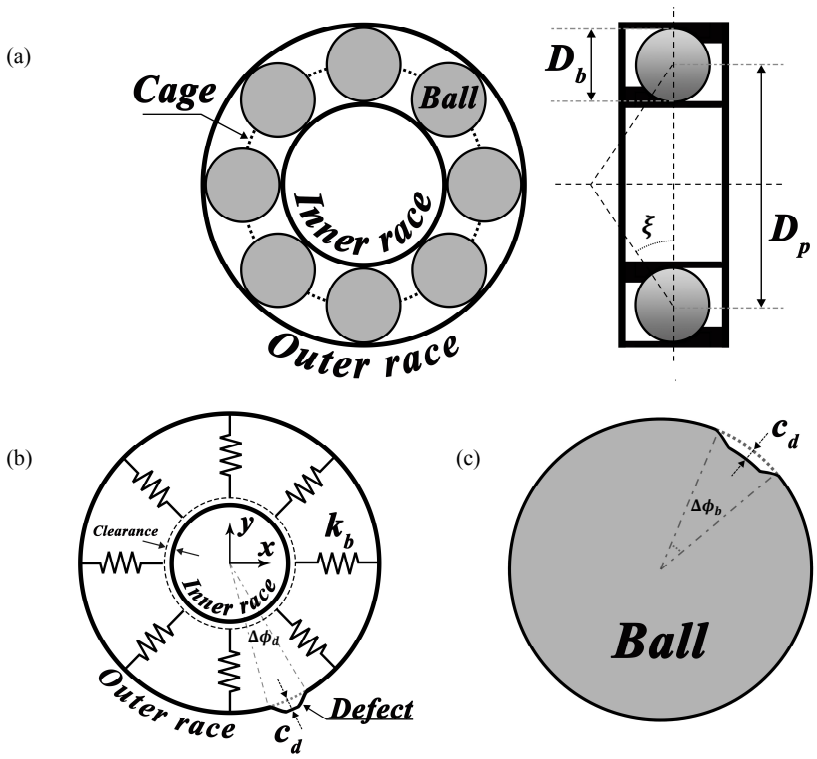

Fig. 7.2 Bearing schematic: (a) bearing components; (b) lumped spring-mass model of bearing and defect model on the outer race; (c) rolling element defect

The rolling element, the element diameter D_b, the pitch diameter D_p and the constant operating contact angle ξ are depicted in Fig. 7.2a.

In this presented model the rolling elements are supposed massless; thus, the centrifugal forces acting on the balls/rollers are negligible. The flexibility of balls/rollers is modelled by circumferentially distributed radial springs with stiffness k_b (Fig. 7.2b) and the radial clearance is considered as well.

The inner and outer race contact deformations can be combined to calculate the overall contact deformation for the jth rolling element δ_j as follow:

$$\delta_j = (x_i - x_o)\cos\phi_j + (y_i - y_o)\sin\phi_j - c \qquad j = 1,2,\ldots,N_b \tag{7.4}$$

where N_b is the number of rolling elements, c is the radial clearance and ϕ_j is the time variant angular position of the center of the rotating elements. Assuming no slippage or sliding between the components of the bearing, the angular position of each ball, ϕ_j, may be calculated based on the races angular rotation and the initial angular position of the first element with respect to x-axis, ϕ_0.

$$\phi_j = \frac{2\pi(j-1)}{N_b} + \theta_{cage} + \phi_0 \tag{7.5}$$

The cage angular speed can be calculated by using the inner race angular speed, ω_i, the outer race angular speed, ω_o, and the geomertry of the bearing [1].

$$\omega_{cage} = \frac{\omega_i}{2}(1 - \frac{D_b}{D_p}\cos\xi\ ) + \frac{\omega_o}{2}(1 + \frac{D_b}{D_p}\cos\xi\ ) \tag{7.6}$$

Due to existence of Hertizian contact between balls/rollers and inner and outer races the reaction force of a roller in position ϕ_j is nonlinear and is given by the following expression:

$$F_j = k_b\delta_j^n \tag{7.7}$$

Harris [110] suggested exponent $n = 1.5$ for ball bearings and 1.1 for roller bearings.

Springs (balls/rollers) forces will exist only if relative motion between inner and outer race causes moving elements to be compressed. Depending on the sign and amount of clearance it is imaginable that in some cases all balls/rollers will not be

in contact simultaneously; therefore, the load zone changes according to the races relative displacements. The contact condition is determined by the positions of the rolling elements. The overall bearing force and stiffness varies with respect to the angular position of balls/rollers and then the total forces in x and y directions can be described as summations of each ball/roller force.

$$F_x = k_b \sum_{j=1}^{N_b} \gamma_j \delta_j^n \cos\phi_j \tag{7.8}$$

$$F_y = k_b \sum_{j=1}^{N_b} \gamma_j \delta_j^n \sin\phi_j \tag{7.9}$$

where γ_j determines whether the jth ball/roller is in contact, according to

$$\gamma_j = \begin{cases} 1 & \delta_j > 0 \\ 0 & \delta_j \leq 0 \end{cases} \tag{7.10}$$

7.2.4 Localized Faults Model

When localized faults in inner/outer races or rolling elements are introduced in a bearing model, equation (7.4) should be written in the following form [17]:

$$\delta_j = (x_i - x_o)\cos\phi_j + (y_i - y_o)\sin\phi_j - \beta_j C_d - c \qquad j = 1, 2, \ldots, N_b \tag{7.11}$$

For the inner/outer race defect case, C_d in this formulation represents the depth of spalls which is considered to be a function of ball's location inside the defected area. β_j determines whether the jth ball/roller is inside defected zone ($\beta_j = 1$) or not ($\beta_j = 0$). A rectangular profile is often used for the shape the of spalls but its sharp borders produce large impulsive forces which cause the vibration response of the system to increase too abruptly. To avoid these impulses, a more realistic profile can be used do define the spalls, as pointed out by Liu et al. [111]; in particular, in this chapter a Tukey window is generated to model C_d (Fig. 7.2b). The maximum height of this window is chosen equal to the maximum of defect depth on the bearing's inner or outer races. When the ball's diameter is larger than the width of the spall,

the maximum depth h theoretically reached by the ball may be calculated as follow (it is an average value for both races):

$$h = \frac{D_p}{2}(1 - \cos\frac{\Delta\phi_d}{2}) \tag{7.12}$$

The angular extent of the spall ($\Delta\phi_d$) and position of the spall on the races are also needed in this model to thoroughly define the spall. β_j equal to one in equation (7.11) indicates that a rolling element is inside the defected region, thus the compression of the ball and the corresponding force will reduce. As it was mentioned earlier, this variation of force will be revealed by a sudden modification in the acceleration response of the system that can eventually reveal defects. In case of spalls with high depth the ball can even totally lose its contact although this event doesn't happen so often in practice.

For the roller/ball defect case, the spall angular speed is identical to the spin speed of a rolling element which can be calculated as follows:

$$\omega_{spin} = \frac{\omega_o - \omega_i}{2}\frac{D_p}{D_b}(1 - (\frac{D_b}{D_p}\cos\xi)^2) \tag{7.13}$$

In this case also a Tukey window is chosen for the spall profile C_d (Fig. 7.2c) and its depth varies as the rolling elements spins. $\beta_j = 1$ indicates contact between the defect on a roller/ball and the inner or outer race and when $\beta_j = 0$ there is no contact. For a full rotation of a defected roller/ball, the fault will be in touch with both outer and inner races. Note that in this case the contact duration between the spall and the inner race is longer than with the outer race due to the difference in curvature of the two races [17]. Depending on the race, the total angular contacts between the fault and outer and inner races are:

$$\begin{aligned} \Delta\phi_{d,i} &= \frac{D_b\,\Delta\phi_b}{D_p - D_b} \\ \Delta\phi_{d,o} &= \frac{D_b\,\Delta\phi_b}{D_p + D_b} \end{aligned} \tag{7.14}$$

where $\Delta\phi_b$ is the angular width of the defect on the roller/ball (Fig. 7.2c).

The so far discussed bearing model is employed for sun, planets and carrier bearings. The sun shaft is assumed to be the input, the output is the carrier shaft:

the rotational speed of the inner races of the sun and carrier bearings are then equal to their shafts speed. Also, the outer races of sun gear and carrier bearings are considered fixed and have no displacements. Additionally, planets outer raceways have the same displacements as the planets and their inner races are assumed be attached to the carrier. It should also be stressed that the axial movements of the inner and outer races may have an influence on the contact deformation of the rollers and the races. The issue has not been addressed in this chapter because of the assumed 2D model (see Fig. 7.1).

7.3 Defect Frequency

7.3.1 Bearing Defect Frequency

Bearing defect frequency is the frequency at which rolling elements pass an imperfection on the inner/outer race or the rolling elements defect approches the raceways. General equations given by Howard [1] are used:

$$f_{bpfi} = \frac{N_b}{2}(f_o - f_i)(1 + \frac{D_b}{D_p}\cos\xi) \tag{7.15}$$

$$f_{bpfo} = \frac{N_b}{2}(f_o - f_i)(1 - \frac{D_b}{D_p}\cos\xi) \tag{7.16}$$

$$f_{bsf} = \frac{f_o - f_i}{2}\frac{D_p}{D_b}(1 - (\frac{D_b}{D_p}\cos\xi)^2) \tag{7.17}$$

where f_o is the outer race frequency, f_i is the inner race frequency, f_{bpfi} is the inner race defect frequency (BPFI), f_{bpfo} is the outer race defect frequency (BPFO) and f_{bsf} is ball or roller spin frequency (BSF).

7.3.2 Gear Defect Frequency

Every time a defected tooth of sun gear meshes with a planet tooth a sudden variation will be introduced into the system vibration signal. During one revolution of the sun gear, this faulty tooth will engage with all the planets and therefore the characteristic

frequency of sun gear with local fault on a single tooth is calculated as follows:

$$f_s = N\frac{f_m}{N_s} \tag{7.18}$$

where N_s is the number of teeth of the sun.

When the planet gear has a tooth with a localized fault every time it meshes with ring or sun, a sudden variation will be introduced to the system vibration signal. During one revolution of the planet gear this faulty tooth will engage with ring or sun so this signal modulation occurs once. During one full rotation of the planet the defected tooth meshes twice (with the sun and ring gear) but only one of these contacts may be considered perfect, depending on the damaged tooth side. Therefore, the characteristic frequency of planet gear with local fault on a single tooth is calculated as follows:

$$f_p = \frac{f_m}{N_p} \tag{7.19}$$

7.4 Equations of Motion

The equations of motion may be written based on the mentioned lumped parameter model of a planetary gearbox and the bearings model. It is worth noting that gyroscopic and centrifugal forces are also considered in this dynamic model.

The sun equations of motion are:

$$m_s\ddot{x_s} + F_{sbx} + \sum_{n=1}^{N} F_{spn}\cos\Psi_{sn} = m_s x_s \dot{\theta_c}^2 + 2m_s\dot{y}_s\dot{\theta}_c + m_s y_s\ddot{\theta}_c \tag{7.20}$$

$$m_s\ddot{y_s} + F_{sby} + \sum_{n=1}^{N} F_{spn}\sin\Psi_{sn} = m_s y_s \dot{\theta_c}^2 - 2m_s\dot{x}_s\dot{\theta}_c - m_s x_s\ddot{\theta}_c \tag{7.21}$$

$$\left(\frac{J_s}{r_s}\right)\ddot{\theta}_c + \sum_{n=1}^{N} F_{spn} = \frac{T_i}{r_s} \tag{7.22}$$

where T_i is the input torque of the system and F_{spn} represents the gear mesh force between the n-th planet and sun gears

$$\begin{aligned} F_{spn} &= k_{spn}\delta_{spn} + c_{spn}\dot{\delta}_{spn} \\ \delta_{spn} &= (x_s - x_{pn})\cos\Psi_{sn} + (y_s - y_{pn})\sin\Psi_{sn} + r_s\theta_s + r_p\theta_{pn} - r_c\theta_c\cos a \\ \Psi_{sn} &= \frac{\pi}{2} - a + \Psi_n \\ \Psi_n &= 2(n-1)\frac{\pi}{N}, \qquad n = 1, \dots N \end{aligned} \tag{7.23}$$

where a is the pressure angle of gears and N is the number of planets. The planets are assumed equally spaced and Ψ_n represents the angular distance between the planets. F_{sbx} and F_{sby} represent the sun bearing force in x_s and y_s directions based on Eqs. (7.8) and (7.9) and ϕ_{sj}, the angular position of sun bearing balls, is calculated with eq. (7.5).

$$F_{sbx} = c_{sx}\dot{x}_s + k_b \sum_{j=1}^{N_b} \gamma_j \left[x_s\cos\phi_{sj} + y_s\sin\phi_{sj} - \beta_j C_d - c\right]^{1.5}\cos\phi_{sj} \tag{7.24}$$

$$F_{sby} = c_{sy}\dot{y}_s + k_b \sum_{j=1}^{N_b} \gamma_j \left[x_s\cos\phi_{sj} + y_s\sin\phi_{sj} - \beta_j C_d - c\right]^{1.5}\sin\phi_{sj} \tag{7.25}$$

$$\phi_{sj} = \frac{2\pi(j-1)}{N_b} + \frac{\theta_s}{2}\left(1 - \frac{D_b}{D_p}\cos\xi\right) + \phi_0 - \theta_c \tag{7.26}$$

The ring equations of motion are:

$$m_r\ddot{x}_r + c_{rx}\dot{x}_r + k_{rx}x_r + \sum_{n=1}^{N} F_{rpn}\cos\Psi_{rn} = m_r x_r\dot{\theta}_c^2 + 2m_r\dot{y}_r\dot{\theta}_c + m_r y_r\ddot{\theta}_c \tag{7.27}$$

$$m_r\ddot{y}_r + c_{ry}\dot{y}_r + k_{ry}y_r + \sum_{n=1}^{N} F_{rpn}\sin\Psi_{rn} = m_r y_r\dot{\theta}_c^2 - 2m_r\dot{x}_r\dot{\theta}_c - m_r x_r\ddot{\theta}_c \tag{7.28}$$

$$\left(\frac{J_r}{r_r}\right)\ddot{\theta}_r + \frac{c_{rt}}{r_r}\theta_r + \frac{k_{rt}}{r_r}\theta_r + \sum_{n=1}^{N} F_{rpn} = 0 \tag{7.29}$$

where F_{rpn} represents the gear mesh force between the n-th planet and ring gears

$$\begin{aligned} F_{rpn} &= k_{rpn}\delta_{rpn} + c_{rpn}\dot{\delta}_{rpn} \\ \delta_{rpn} &= (x_r - x_{pn})\cos\Psi_{rn} + (y_r - y_{pn})\sin\Psi_{rn} + r_r\theta_r - r_p\theta_{pn} - r_c\theta_c\cos a \\ \Psi_{rn} &= \frac{\pi}{2} + a + \Psi_n \end{aligned} \tag{7.30}$$

The planets equations of motion ($n = 1, \ldots, N$) are:

$$\begin{aligned} m_p\ddot{x}_{pn} + F_{cpxn} - F_{spn}\cos\Psi_{sn} - F_{rpn}\cos\Psi_{rn} \\ = m_p x_{pn}\dot{\theta}_c^2 + 2m_p\dot{y}_{pn}\dot{\theta}_c + m_p y_{pn}\ddot{\theta}_c + m_p r_c\dot{\theta}_c^2\cos\Psi_n \end{aligned} \tag{7.31}$$

$$\begin{aligned} m_p\ddot{y}_{pn} + F_{cpyn} - F_{spn}\sin\Psi_{sn} - F_{rpn}\sin\Psi_{rn} \\ = m_p y_{pn}\dot{\theta}_c^2 - 2m_p\dot{x}_{pn}\dot{\theta}_c - m_p x_{pn}\ddot{\theta}_c + m_p r_c\dot{\theta}_c^2\sin\Psi_n \end{aligned} \tag{7.32}$$

$$\left(\frac{J_p}{r_p}\right)\ddot{\theta}_{pn} + F_{spn} - F_{rpn} = 0 \tag{7.33}$$

where F_{cpxn} and F_{cpyn} represent the planets bearing force in x_{pn} and y_{pn} directions based on Eqs. (7.8) and (7.9) and ϕ_{pj}, the angular position of the planet-bearing balls, is calculated with eq. (7.5).

$$\begin{aligned} F_{cpxn} = & c_{pnx}(\dot{x}_{pn} - \dot{x}_c) + \\ & k_b\sum_{j=1}^{N_b}\gamma_j\left[(x_{pn} - x_c)\cos\phi_{pj} + (y_{pn} - y_c)\sin\phi_{pj} + \beta_j C_d + c\right]^{1.5}\cos\phi_{pj} \end{aligned} \tag{7.34}$$

$$\begin{aligned} F_{cpyn} = & c_{pny}(\dot{y}_{pn} - \dot{y}_c) + \\ & k_b\sum_{j=1}^{N_b}\gamma_j\left[(x_{pn} - x_c)\cos\phi_{pj} + (y_{pn} - y_c)\sin\phi_{pj} + \beta_j C_d + c\right]^{1.5}\sin\phi_{pj} \end{aligned} \tag{7.35}$$

$$\phi_{pj} = \frac{2\pi(j-1)}{N_b} + \frac{\theta_{pn} - \theta_c}{2}\left(1 + \frac{D_b}{D_p}\cos\xi\right) + \phi_0 \tag{7.36}$$

The carrier equations of motion are:

$$m_c\ddot{x}_c + F_{cbx} - \sum_{n=1}^{N} F_{cpxn} = m_c x_c \dot{\theta}_c^2 + 2m_c \dot{y}_c \dot{\theta}_c + m_c y_c \ddot{\theta}_c \tag{7.37}$$

$$m_c\ddot{y}_c + F_{cby} - \sum_{n=1}^{N} F_{cpyn} = m_c y_c \dot{\theta}_c^2 - 2m_c \dot{x}_c \dot{\theta}_c - m_c x_c \ddot{\theta}_c \tag{7.38}$$

$$\frac{J_c}{r_c}\ddot{\theta}_c + \sum_{n=1}^{N} F_{cpxn}\sin\Psi_n - \sum_{n=1}^{N} F_{cpyn}\cos\Psi_n = \frac{T_o}{r_c} \tag{7.39}$$

where T_o is the output torque of the system. F_{cbx} and F_{cby} represent the carrier bearing force in x_s and y_s directions based on Eqs. (7.8) and (7.9) and ϕ_{cj}, the angular position of the carrier bearing balls, is calculated with eq. (7.5).

$$F_{cbx} = c_{cx}\dot{x}_c + k_b \sum_{j=1}^{N_b} \gamma_j \left[x_c\cos\phi_{cj} + y_c\sin\phi_{cj} - \beta_j C_d - c\right]^{1.5}\cos\phi_{cj} \tag{7.40}$$

$$F_{cby} = c_{cy}\dot{y}_c + k_b \sum_{j=1}^{N_b} \gamma_j \left[x_c\cos\phi_{cj} + y_c\sin\phi_{cj} - \beta_j C_d - c\right]^{1.5}\sin\phi_{cj} \tag{7.41}$$

$$\phi_{cj} = \frac{2\pi(j-1)}{N_b} + \frac{\theta_c}{2}\left(1 - \frac{D_b}{D_p}\cos\xi\right) + \phi_0 - \theta_c \tag{7.42}$$

It is stressed that the equations of motion are written in the rotating frame of reference Oxy and their results are therefore almost impossible to measure in an experimental rig. In practice, vibration signals of planetary gearboxes can in fact be collected via accelerometers mounted on the casings, being in most cases the internal parts not accessible. Once the previous equations have been solved, it is necessary to project the results in the fixed frame of reference OXY (see Fig. 7.1). In particular, the results discussed in the following section arise from the accelerations of a fixed point on the ring gear, to simulate an actual condition. Gravity can be added to the equations of motion but it has been numerically verified that the presented results would not be modified. In fact, with the parameters set in Table 7.1, the effect of the weight force is negligible in comparison to gear meshing forces, as a result, does not influence the computed time histories, i.e. the accelerations of the output point.

To the best of my knowledge no published work, except [96], has been found in the literature which try to simulate the effect of bearing localised faults in a planetary

gearbox. The present model has indeed some differences with may be worth noting. First, Jain [96], takes into account linear and time-invariant mesh stiffness between the gears, which allows him to define time-invariant mass and stiffness matrices. Consequently, he is able to define and solve an eigenvalue problem and to eventually determine the frequency response functions (FRFs) of the system in terms of modal properties. On the contrary, our proposed model uses time-variant forces and also considers mesh phasing, which can have a substantial effect on the result [107]. A constant mesh stiffness also makes it impossible to introduce any defect on gears, since they are modelled by variation on mesh stiffness (see Fig. 7.3). Moreover the nonlinearity of bearings contacts in addition to fluctuation of the bearing total stiffness, as a result of the cage rotation, are considered. The achieved equations are then nonlinear and time-variant and can't but numerically be solved in the time domain. It also may be observed that computing a FRF in terms of receptance is not too realistic because it is impossible to isolate a single force, as required by the definition of the FRFs. Second, the signature of bearing localised faults is analytically modelled in [96] as a modification of the bearing force, which limits the ability to model bearing faults with different shapes. This approach is here replaced by the model in the subsection 7.2.4, which describes the damage in terms of variation of the contact pattern. Third, the size of the bearing load zone is considered constant in [96] which is not the case here, since it differs for bearings with various number of rolling elements. Finally, the flexibility of the ring gear is modelled in [96] by using a modal expansion, which has here been ignored.

7.5 Results and Discussion

In this section, the set of nonlinear and time-variant equations derived in subsection 7.4 is solved to obtain vibration signals of each gear. In this regards, a code in the Wolfram Mathematica software environment has been developed to numerically solve this system of ordinary differential equations (ODE) by means of the built-in NDSolve function. The gears and bearings parameters are included in Tables 7.1 and 7.2 respectively.

Table 7.1 Parameters of the planetary gearbox

	Sun	Planet	Ring	Carrier
Mass (Kg)	2.1	1	6.7	15.3
Number of Teeth	26	19	64	
Base Circle radius (mm)	42	30.7	103.5	
Root Circle radius (mm)	40	29.2	111.5	
Module (mm)	3.55			
Face width (mm)	60			
Poisson Ratio	0.3			
Young's Modulus (Pa)	2.068×10^{11}			
Pressure Angle, α	22.5			
Input Torque (N.m), T_i	500			
Output Shaft Speed (rpm)				150
Number of Planets, N	3			
Reduction Ratio	3.46			

Table 7.2 Parameters of the bearings

Ball stiffness, k_b (N/m$^{1.5}$)	3.3×10^{11}	D_b (mm)	3
Number of balls, N_b	8	D_p (mm)	13
Contact angle, ξ	0		
$c_{sx} = c_{sy} = c_{rx} = c_{ry} = c_{cx} = c_{cy} = c_{pnx} = c_{pny} = c_{rt} = 1.8 \times 10^3$			

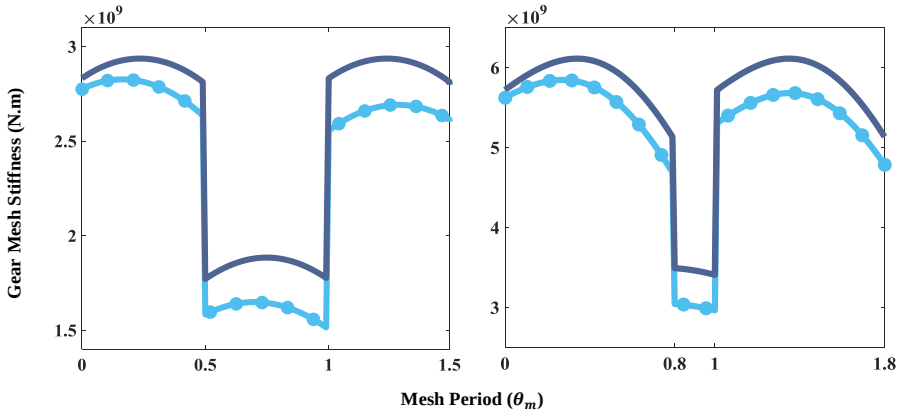

Fig. 7.3 Gear mesh stiffness for healthy (solid line) and cracked (circle-line) tooth (a) sun-planet (contact ratio = 1.5) (b) ring-planet (contact ratio = 1.8)

7.5.1 Gears with Cracked Teeth

The approach of Ref. [112] is used to analytically evaluate the time-varying mesh stiffness of external-external and external-internal gears of our planetary gear set. Fig. 7.3 shows the mesh stiffness of perfect and cracked (10 percent of tooth root is cracked) teeth in the sun-planet and planet-ring coupling for a meshing duration corresponding to the average number of gear teeth pairs in contact while a tooth comes and goes out of contact (contact ratio). As Fig. 7.3 demonstrates, the amplitude of mesh stiffness waveform decreases for damaged gear teeth due to the thickness reduction of the cracked tooth. In this research γ_{rs} is 0.5 and γ_{sn} and γ_{rn} for $n = 1,2,3$ are 0, 2/3, 1/3 and 0, -1/3, -2/3, respectively (see subsection 7.2.2).

The mathematical description of gears contact damping coefficient is complex. In this study a simplified damping model, Ref. [113], is assumed to determine the effective damping factor as follows:

$$c_{jpn} = 2\zeta \sqrt{k_{jpn} \frac{J_p\, J_j}{J_p r_j^2 + J_j r_p^2}} \qquad j = s,r \quad and \quad n = 1,\ldots,N \tag{7.43}$$

where ζ is the contact damping ratio. The value of ζ is between 0.03 to 0.17 [113] and in this study an average value of 0.10 is selected.

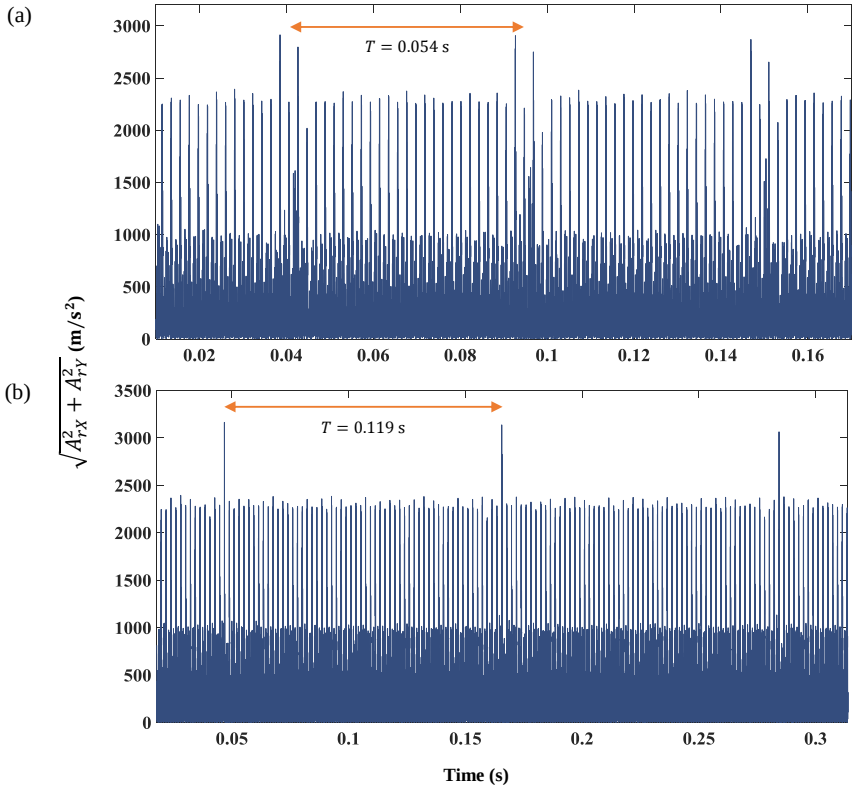

Fig. 7.4 Effect of (a) sun gear or (b) planet gear tooth crack on the ring gear acceleration in X and Y directions (A_{rX},A_{rY})

Fig. 7.4 illustrates the acceleration signal of the ring gear (in the fixed frame OXY) when the crack is seeded in a single tooth of the sun gear (Fig. 7.4a) or planet gear (Fig. 7.4b). When tooth cracks are present, impulsive signals can be observed in time domain. The time duration between two consecutive disturbances in the acceleration signal is equal to the gear defect period. Based on the Eqs. (7.18) and (7.19) sun and planet gear defect frequencies can be calculated as 18.46 Hz (0.054 s) and 8.42 Hz (0.119 s) respectively.

7.5.2 Defected Planet-Bearing

In this section effects of faulty inner race, outer race and rolling element of the planet-bearing will be mainly considered.

In a planetary gearbox forces acting on carrier through planets bearings or on sun gear through the planet-sun gear mesh counterbalance each other and the resultant

forces on carrier/sun bearing is negligible. On the other hand, radial load on planets bearings, which transmit the torque are much higher. As a result, in contrast to sun and carrier, planets bearings exhibit a high failure rate and are considered as one of the most critical components in planetary gearboxes. Therefore, the focus of this section is on defects of planets bearings in a gearbox.

7.5.2.1 Fault on Inner Race

For the healthy gearbox, the acceleration of the ring gear in X direction for zero clearance is shown in Fig. 7.5a. Also, Fig. 7.5b displays the ring acceleration result when a 4-degree defect ($\Delta\phi_d = 4^o$) is seeded in the inner race of the first planet bearing. The localized fault is positioned along the rotating y axis of the planet-bearing and does not move with respect to carrier and bearing load-zone.

For the planets bearings the inner race frequency is equal to carrier frequency ($f_i = f_c$) and the outer race is fixed to the planet ($f_o = f_p = -2.37 f_c$). Therefore, the ball pass frequency of planet inner race (BPFI) when the carrier rotational speed is 150 rpm can be calculated according to Eq. (7.15) as 41.39 Hz. Dashed lines in Fig. 7.5b represents the time duration in which a roller goes through the spall and the spacing between every two successive occurrences is equal to defect period ($T_i = 1/f_{bpfi} = 0.024$ s) which also verifies the validity of the implemented model. As the fault is always inside the load zone its effect is present in every passage of balls over the fault.

Although the bearing fault signature can be presumed inside the dashed box, the signal is mainly dominated by gears meshing components. To attain clearer perception of bearing defect signature, zoomed portions of Fig. 7.5a and Fig. 7.5b are plotted in Fig. 7.5c. Disturbance of the acceleration signal which is due to the defect on the planet-bearing can be seen more evidently.

7.5.2.2 Fault on Outer Race

Planets are responsible to transmit the input torque from sun gear to carrier and therefore forces on their bearings in y direction, which causes the carrier to rotate, are much larger than forces in x direction. When positive or zero clearance exists in planets bearings there will be no sign of defect once the fault is out of the load zone. In this simulation, to represent the effect of outer race spall on output signal

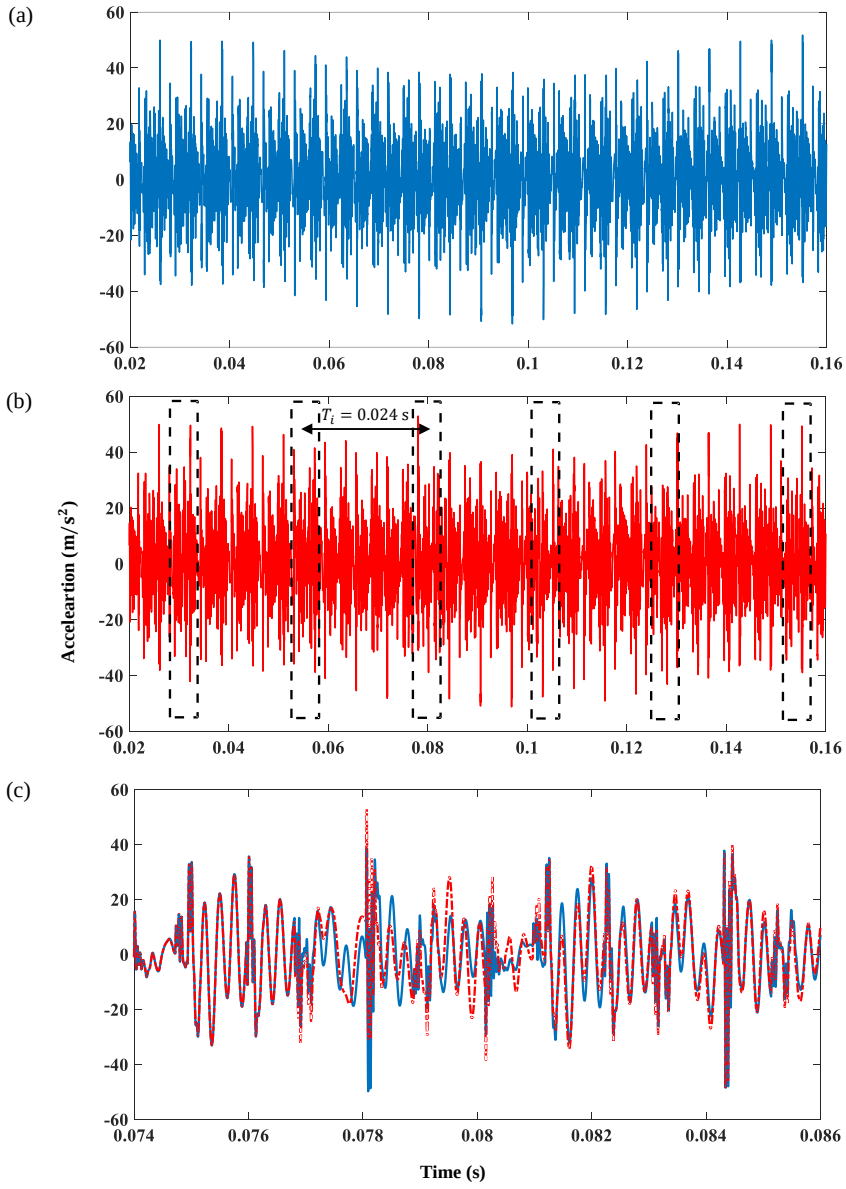

Fig. 7.5 Ring gear acceleration signal (a) healthy planet-bearing (b) defected inner race, $\Delta\phi_d = 4^o$ (c) zoomed portion of healthy and defected signals

more clearly, the clearance is set to an arbitrary value of -5 μm in the planet-bearing model since the negative value of clearance or preload could be generated by elastohydrodynamic lubrication films (EHL) [17].

Ring gear acceleration signal in X direction in case of healthy and faulty planet-bearing, 4-degree rotating spall, are displayed in Fig. 7.6a and Fig. 7.6b respectively. To clarify the influence of outer race spall in the ring gear signal, zoomed portion of Fig. 7.6a and Fig. 7.6b for the period in which one of the balls is inside the defected area are simultaneously represented in Fig. 7.6c. By using Eq. (7.16) defect frequency for the planet outer race (BPFO) is calculated as 25.97 Hz. Defect time period in this case is equal to $T_o = 1/f_{bpfo} = 0.039$ s which is shown in Fig. 7.6b as time difference between two successive defected areas of acceleration signal (dashed box). Similar to the previous case, the bearing defect signal in time domain is vastly masked by gear mesh signals.

For the case of planet-bearing with a defected ball, BSF is calculated by using Eq. (7.17) as 34.87 Hz. Since the balls defects have somehow the same characteristics as the outer race defects, the time domain results are not depicted to not be repetitive.

7.5.2.3 Frequency Analysis of defect Signals

This subsection is devoted to the analysis of the frequency spectrum of signals associated to planet-bearing defects. The objective is to highlight the signal components generated by damage and their sources. The undamaged bearing is considered to generate a reference signal. Hence, the difference between the time responses of damaged and undamaged systems is related to the faults entirely and will be analysed in the following examples.

To determine the frequency spectrum, a Fourier transform is performed on the residual signals. Fig. 7.7 shows the spectra of residual vibration response of the ring gear due to the presence of a single defect on the inner race of the planet-bearing.

Inner race fault frequency, f_{bpfi}, and its harmonics can be seen in Fig. 7.7a. Each of them encompasses a cluster of sidebands separated by the carrier rotation frequency, f_c: a magnified part of Fig. 7.7a is represented in Fig. 7.7b. These sidebands are produced as a result of carrier rotation. In this case, as mentioned earlier, the ring gear is fixed so the centers of planets revolve along with the carrier. This revolution causes the transmission path between the planet-bearing and the

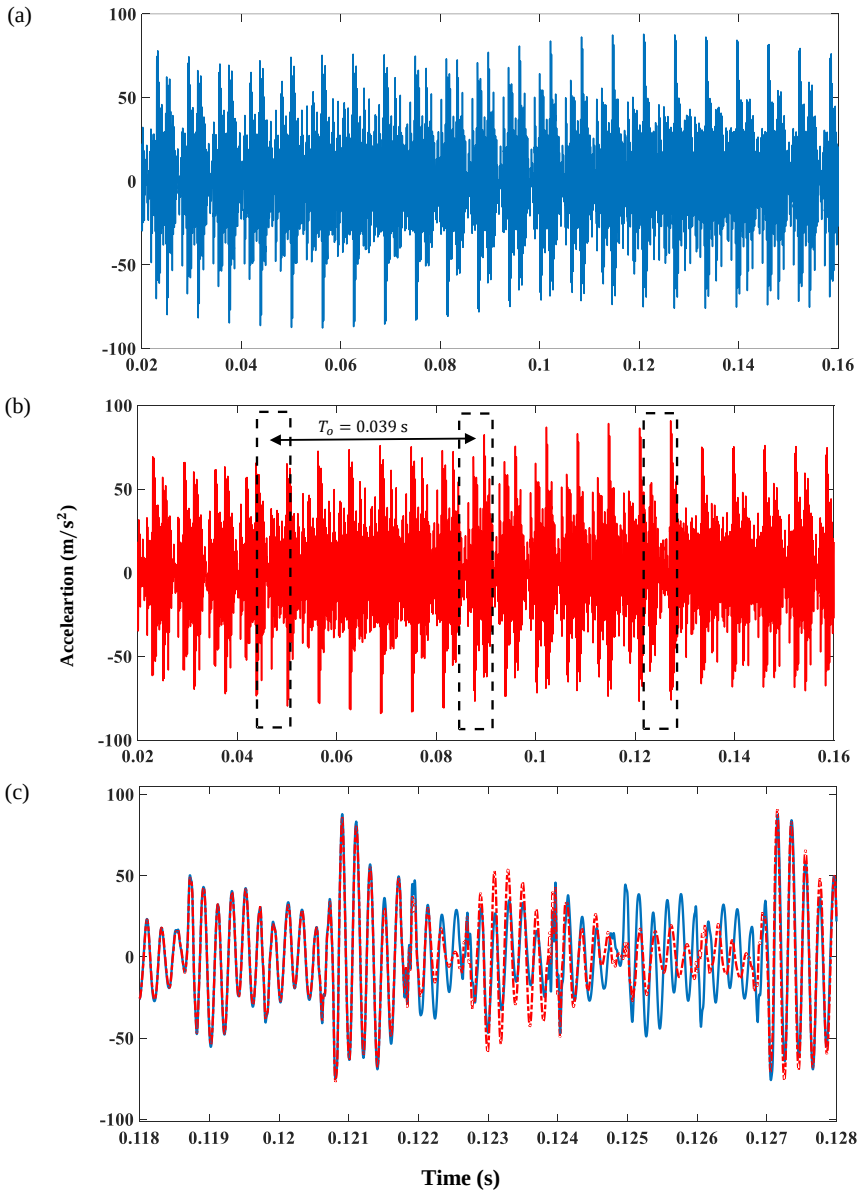

Fig. 7.6 Ring acceleration signal (a) healthy planet-bearing (b) defect outer race, $\Delta\phi_d = 4^o$ (c) zoomed portion of healthy and defected signals

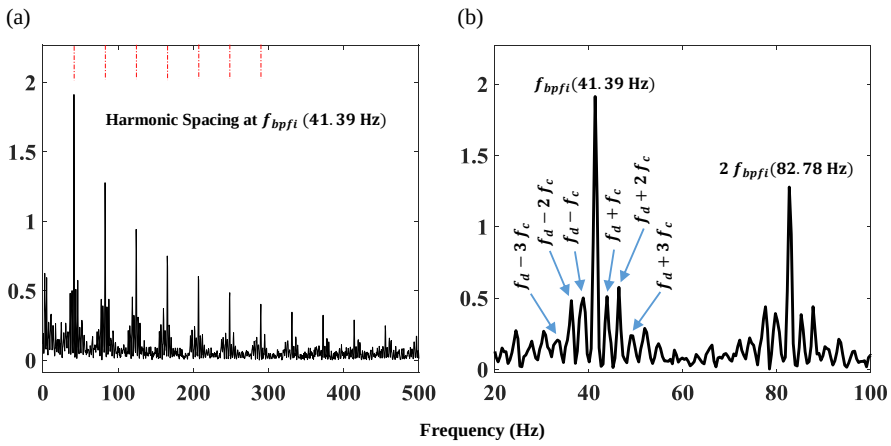

Fig. 7.7 Frequency spectrum of defected inner race planet-bearing signal ($f_d = f_{bpfi}$)

signal acquisition point on the ring gear to vary with time. Due to this variation, the amplitude of planet's bearing vibration signal is modulated and generates the sidebands around the damage frequency and its harmonics. It must be addressed that a spectrum with proper frequency resolution is needed to detect the damage frequencies and their sidebands, so time duration of calculated (or measured) data should be chosen long enough to generate a reliable frequency spectrum.

Frequency spectrum for planet-bearing with defected outer race and ball are shown in Fig. 7.8 and Fig. 7.9 respectively. Similar to the previous case, a cluster of sidebands is present around the outer race damage frequency, f_{bpfo}, and ball damage frequency, f_{bsf}, and their harmonics. But as it is shown in Fig. 7.8b and Fig. 7.9b more sidebands around damage frequencies exist in comparison to the inner race fault since in these cases two frequencies are responsible for creation of these sidebands. As in the former circumstance, the carrier rotation changes the transmission path. This is the first source of planet-bearing signal amplitude modulation so that f_{bpfo} and f_{bsf} are modulated by f_c. When a spall is located on planet-bearing inner race, its position doesn't change relative to rotating frame xy because both are attached to the carrier. However outer race fault is fixed to the planet gear and relative rotation occurs between it and the carrier; moreover, relative rotation between ball and carrier occurs as a result of the cage rotation. Consequently, the following events occur.

First, unlike the inner race fault, the sudden force which is caused by the defective planet-bearing outer race or ball will rotate relative to the xy reference frame. The amount of these relative rotational speeds is equal to ω_{cage} in case of ball defect and

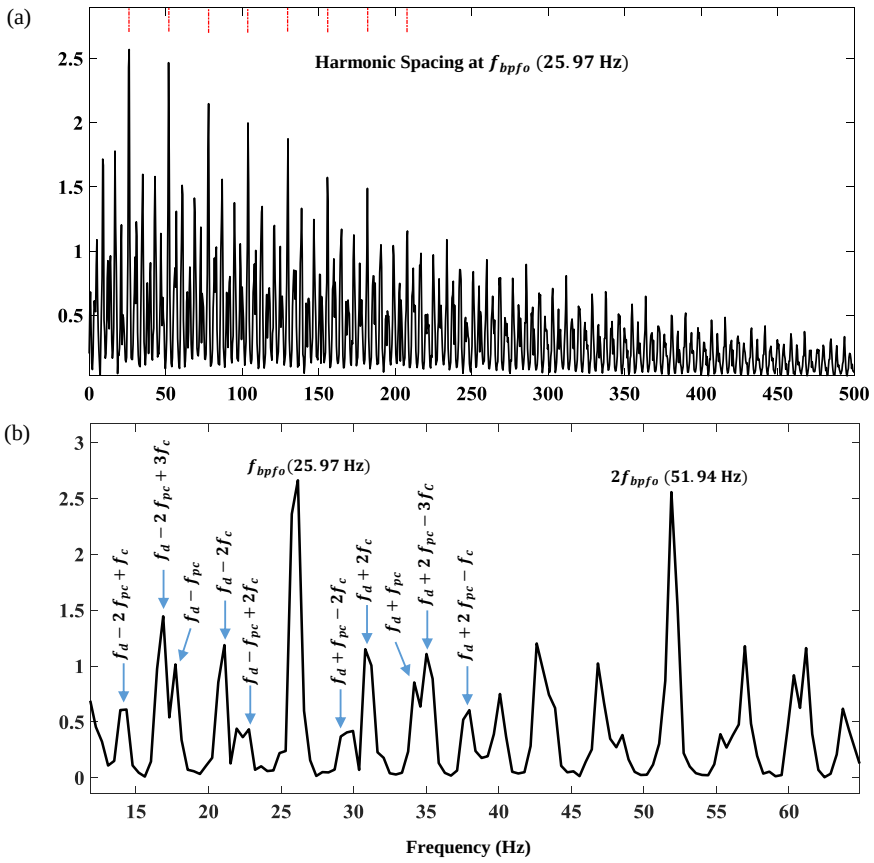

Fig. 7.8 Frequency spectrum of defected outer race planet-bearing signal ($f_d = f_{bpfo}$)

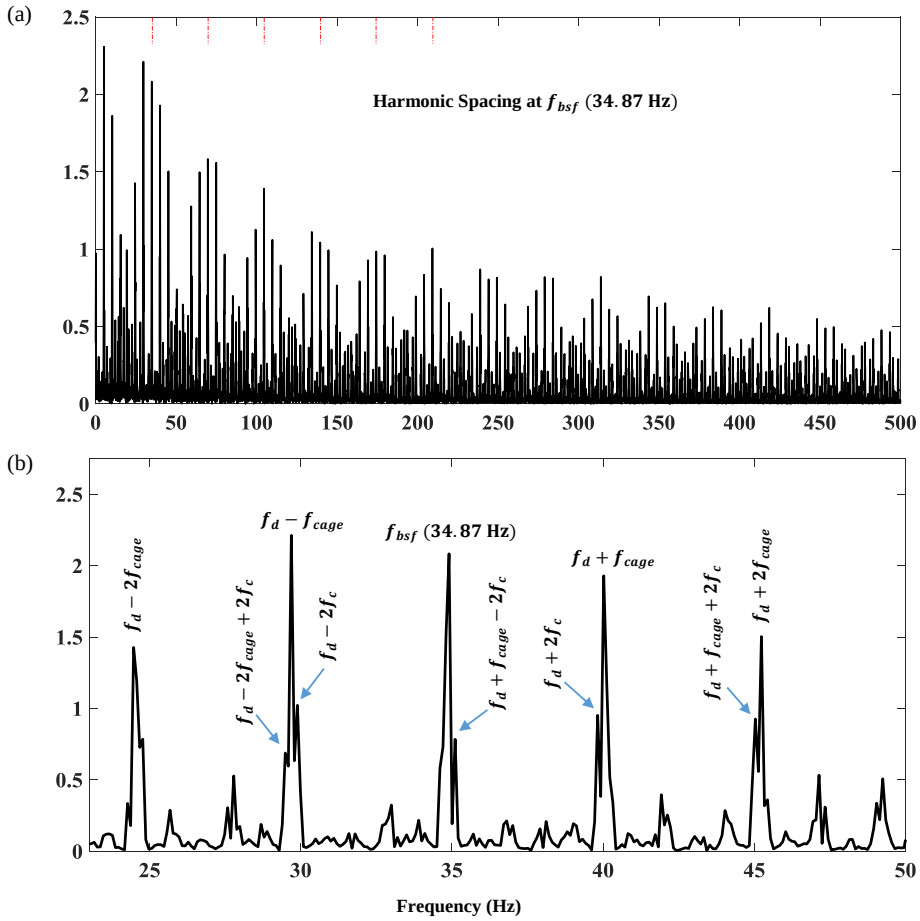

Fig. 7.9 Frequency spectrum of defected ball planet-bearing signal ($f_d = f_{bsf}$)

can be calculated by subtracting the speed of the planet (ω_p) from the speed of the carrier (ω_c) in case of outer race defect.

Second, if the planet-bearing clearance is zero/positive or negative but not sufficient to prevent the ball/roller from losing its contact with the races while it passes through the fault or when the ball defect is in contact with the races, the magnitude of force generated by the defect is different according to the position of the spall relative to the load zone (i.e. amplitude of force will be non-zero inside the load zone and zero when the spall is out of the load zone).

Third, planets vibration signals are transmitted to the ring gear via ring-planet engaging teeth. Due to the spall rotation along with the planet gear or bearing cage the angle between the defect force and ring-planet mesh varies. Maximum force will be transferred to the ring gear when the defect force and the gear meshing spring are in the same direction. On the other hand, perpendicularity minimizes the transmitted force [96].

According to above-mentioned phenomena, the rotation of defect together with the outer race or cage changes magnitude and direction of its force. Consequently, these variations generate a second frequency of modulation which is equal to $f_{pc} = f_p - f_c$ for the outer race spall and f_{cage} for the ball spall. As two modulation frequencies exist for the outer race and ball fault cases, the cluster of peaks around each damage frequency (sidebands), which are shown in Fig. 7.8b and Fig. 7.9b, are combination of f_c and f_{pc} for the outer race fault or f_c and f_{cage} for the ball fault and can be calculated as follow [96]:

$$\begin{aligned} f_{sidebands,o} &= a f_{bpfo} \pm b f_c \pm c f_{pc} \\ f_{sidebands,b} &= a f_{bsf} \pm b f_c \pm c f_{cage} \end{aligned} \qquad a,b,c > 0 \tag{7.44}$$

where $f_{cage} = \omega_{cage}/(2\pi)$ (eq. (7.6)).

It is worth noting that as bearings balls move, their total stiffness varies. In case of planet-bearing, frequency of this variation is equal to inner race ball passage frequency. Thus if the preload and number of balls are not properly selected, even for flawless bearings, vibrations occur at this frequency but increasing the number of balls decreases the amplitudes of bearing vibrations and ball pass frequency (BPF) will reduce accordingly.

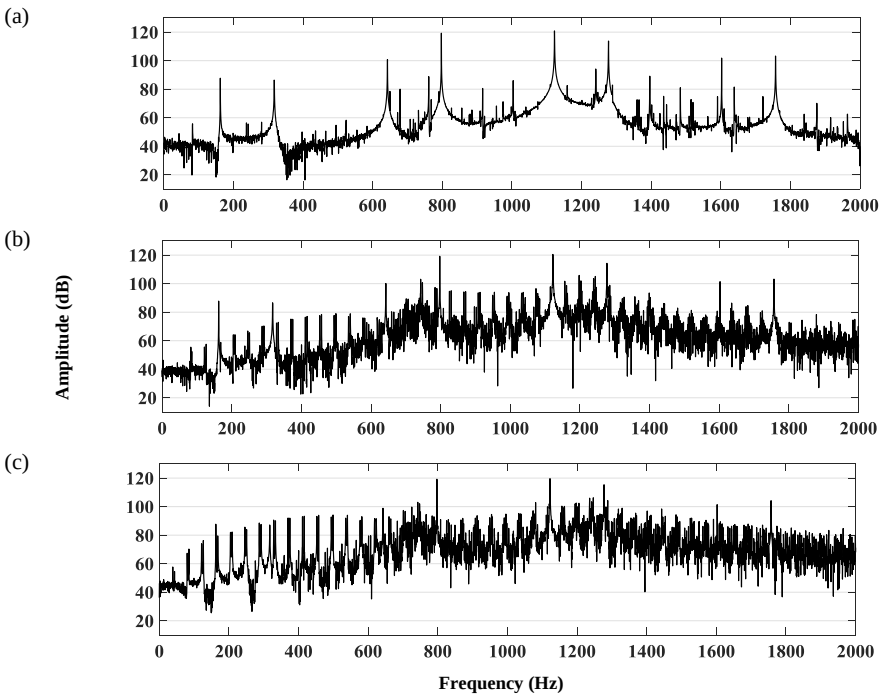

Fig. 7.10 Frequency spectrum (a) healthy gearbox (b) defected planet inner race, $\Delta\phi_d = 2^o$ and (c) 4^o

The damage frequencies and sidebands presented for both planet inner race, outer race and balls faults in this section comply with the theoretical and experimental finding of Ref. [96]. This agreement also verifies the reliability of the bearings and gearbox models as well as accuracy of the simulation results.

7.5.2.4 Frequency spectrum analysis of the ring acceleration signals

Fig. 7.10 shows the frequency spectrum of the ring gear acceleration signal for healthy and defected planet inner race bearings. In Fig. 7.10b and Fig. 7.10c the sizes of defects are 2 and 4 degrees respectively and Fig. 7.10a represent the frequency spectrum of the healthy gearbox. Around the meshing frequency (160 Hz) sidebands emerge when a defect is present. Although the amplitudes of meshing frequency and its harmonics are barely affected by the bearings faults, amplitudes of sidebands increase as the size of spall grows. These collection of frequencies around the meshing frequencies are combination of BPFI and its sidebands as discussed in the previous subsection.

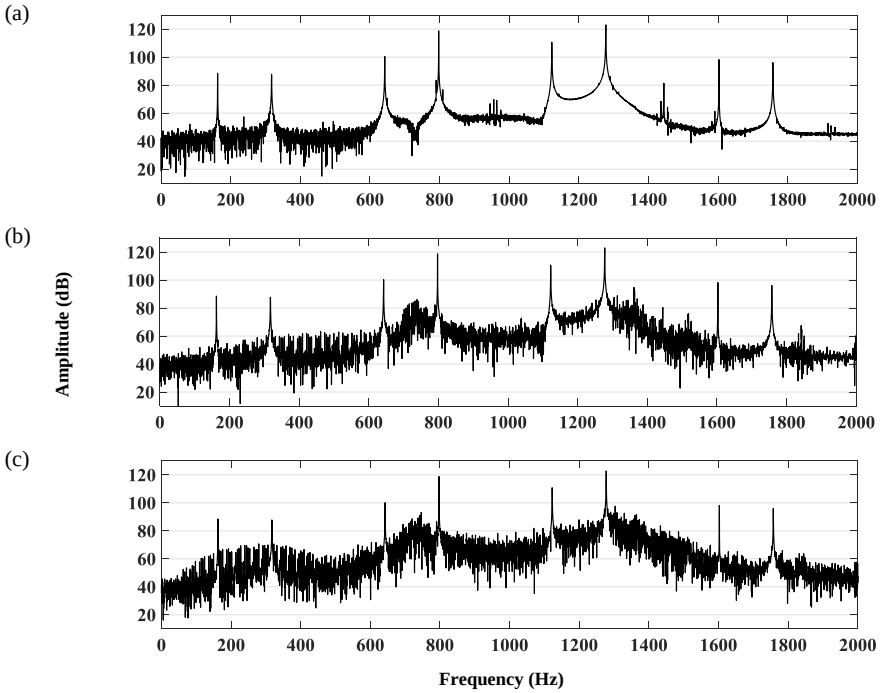

Fig. 7.11 Frequency spectrum (a) healthy gearbox (b) defected planet outer race, $\Delta\phi_d = 2^o$ and (c) 4^o

The spectrum of the gearbox signal with healthy, 2 and 4 degrees faults on the outer race of one planet-bearing are also illustrated in Fig. 7.11. Notice that Fig. 7.10a differs from Fig. 7.11a because of the presence of clearance. In this case, the spectra comprise more sidebands as explained in subsection 7.5.2.3. The BPFO frequency and its sidebands appears in vicinity of the meshing frequencies.

The spectrum for healthy and defected planet-bearing balls are depicted in Fig. 7.12. The maximum balls spalls depths are chosen equal to maximum depths of races defects, h, to have comparable results. The BSF frequency and its sideband are also present in the spectrum. These spectra are similar to those generated by outer race defects case because in both circumstances defects are in rotation relative to the carrier and therefore two sources of modulation exist as explained in the previous subsections.

In all above cases, the amplitudes of meshing frequency and its harmonic are higher than the amplitude of the bearing defect frequency and its harmonics. This is caused by the fact that planetary gearboxes transfer a large amount of torque from

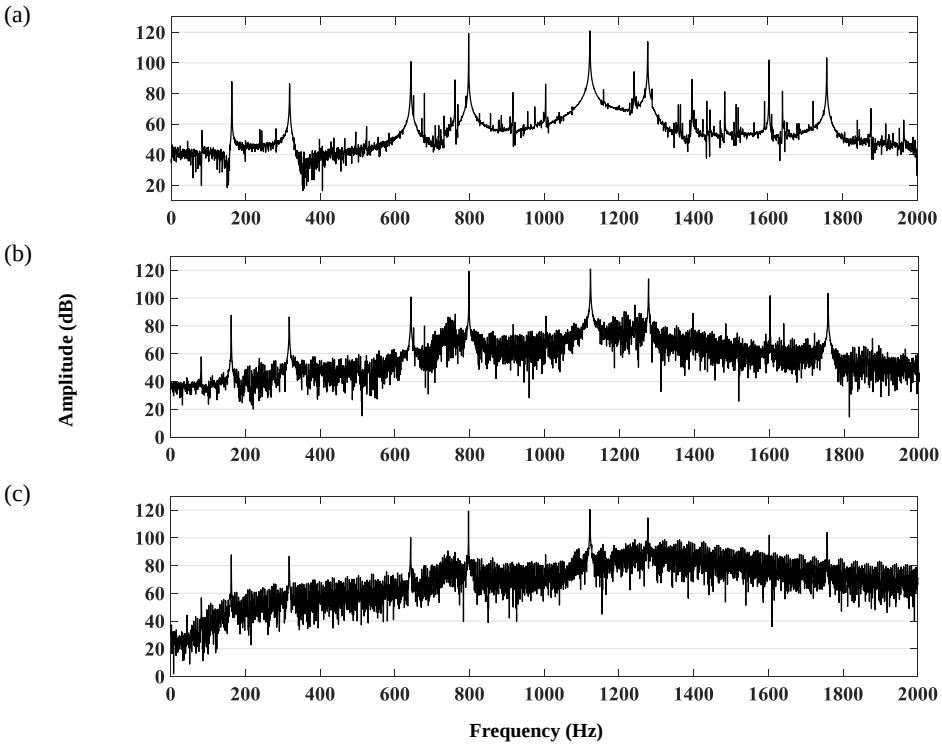

Fig. 7.12 Frequency spectrum (a) healthy gearbox (b) defected planet balls, $\Delta\phi_d = 2^o$ (c) $\Delta\phi_d = 4^o$ ($\Delta\phi_b = 45^o$)

sun gear to carrier through their gear meshes causing rather high energy gear mesh frequencies in comparison to the energy level of signals generated by bearing defects.

7.5.2.5 Condition monitoring of the gearbox

Diagnosis of defects on planets bearings at their early stages has always been problematic. Many statistical features have been developed for condition monitoring of gearboxes and have been frequently used as an indicator of bearings and gears faults presence and growth. Therefore, the vibration signal from the simulation is processed to calculate a set of statistical features. The accelerations for different levels of planet-bearing defects on inner races, outer races and rolling elements will be evaluated by implementing the root mean square (RMS), kurtosis and M8A features to investigate the effectiveness of statistical indicators in detection of planets bearings defects.

Root mean square (RMS) might be one of the most commonly used indicators in vibration monitoring. It is defined as

$$\mathrm{RMS}_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)^2} \tag{7.45}$$

where x_i is the data sample, N and $\bar{x}$ are the length and average of x respectively. The kurtosis is the fourth normalized moment of a signal and provides a measure of its peakedness. It is given by

$$\mathrm{Kurtosis}_x = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2} \tag{7.46}$$

The parameter M8A uses the eighth moment normalized by the variance to the fourth power and is given as

$$\mathrm{M8A}_x = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^8}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^4} \tag{7.47}$$

Many research work has been carried out in the field of statistical fault detection and diagnosis. Typically time domain signals are the inputs of statistical indicators

[114]. In this work these methods are applied on the ring acceleration in time domain but when the sizes of defects are small these indicators are not reliable and cannot detect the presence of the fault because the signal is highly dominated by the gear components.

As it was discussed earlier, frequency sidebands appear about gear mesh frequencies when a bearing fault is introduced and amplitudes of sidebands intensify as the spall grows. These variations are believed to be worthwhile for fault diagnosis of planetary gearboxes. Therefore, the condition indicators on the ring acceleration are exploited in frequency domain to investigate the planet-bearing defect. The procedure is the same as the time domain analysis except that data from Fourier transform of the ring gear time signals are take into account. For every result discussed in this section, white noise (with RMS equal to 15 percent of the acceleration RMS) has been added to the original time sequence to test the performance of these indicators when background noise is not negligible.

The calculated condition indicators for 6 cases, no defect and 5 different sizes of the spalls ($\Delta\phi_d = 1$, 2, 4, 8 and 15 degrees), are calculated. For the defected balls $\Delta\phi_b = 45^o$ and the maximum depth of the ball defect is selected equivalent to maximum depth of the races spall h. For defects on the outer race, inner race and balls of planet-bearing, the difference between the indicators in defected and healthy cases are calculated then these differences are divided by the healthy bearing indicator. The result are plotted in Fig. 7.13.

As it is shown in the Figs. 7.10 - 7.12, when the spall advances not only new peaks appear in the spectrum but also the amplitudes around meshing frequencies ascend generally. Therefore, all the statistical indicators quantities rise with the growth of the planet's defect size (Fig. 7.13). As the average of the data increases the value of RMS gets larger correspondingly, while kurtosis and M8A provide measures of the peakedness of data sequences. A higher positive value indicates a peaked distribution while a lower negative value indicates a flat distribution. When the fault is present, as more peaks appear around the gear mesh frequencies the values of kurtosis and M8A decrease, i.e. the effect of gear mesh frequencies deviation on variance reduces because of sidebands peaks. Therefore, the absolute values of the indicators are presented in Fig. 7.13. The M8A indicator is more sensitive to the presence of the defects and performs better on faults detection, with a potential to diagnose spalls even in their emerging stages. Another observation is that the

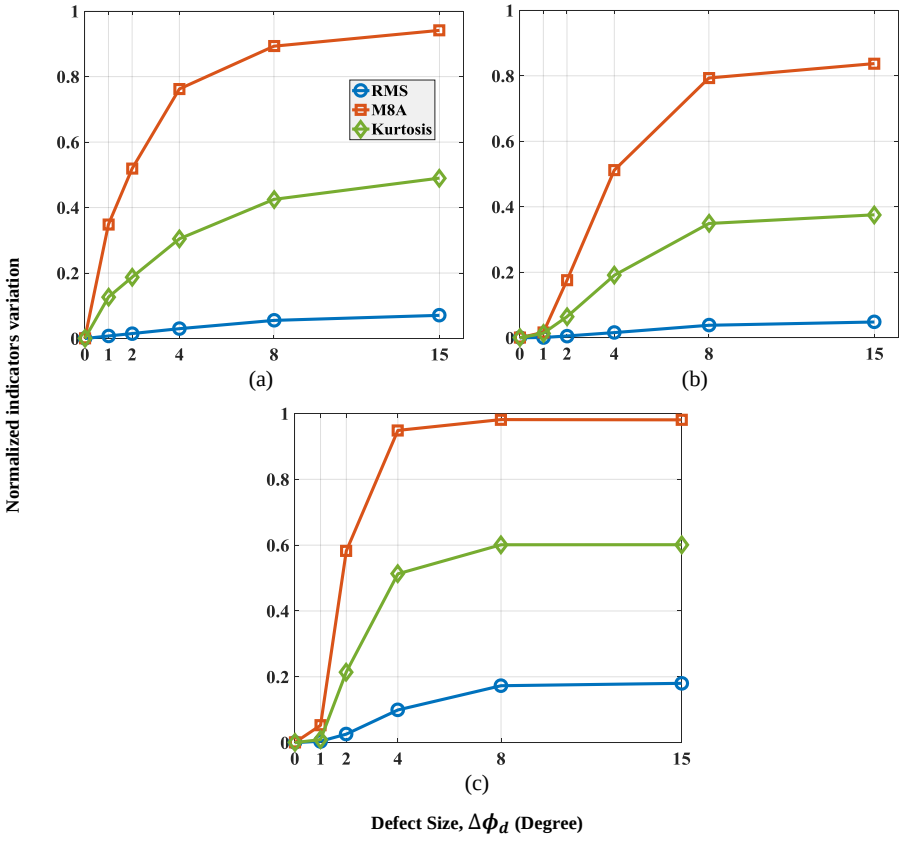

Fig. 7.13 Effects of (a) inner race (b) outer race (c) ball defects sizes on the condition indicators

lower frequencies are more sensitive to the presence of the faults. Lower natural frequencies of the system have larger amplitudes therefore they more effectively amplify the amplitudes of the meshing and defect frequencies. Moreover, as the frequencies amplitudes escalate around the natural frequencies of the system the effect of noise becomes less substantial.

At the end of this section it is worth noting that increased sensitivity may not always be a desired characteristic because an excessively sensitive parameter will yield many false alarms. For this reason, the most useful damage indicator is not definitely the most sensitive one.

Chapter 8

Conclusion and future work

8.1 Summary and conclusions

The research presented in this thesis is summarized in this chapter. Rolling element bearings (REBs) diagnosis and simulation of planetary gearbox are the two main areas investigated in this work.

In the first part the emphasis is mainly on bearings fault detection. In chapter 2, the literature of the vibration-based condition monitoring, with the focus on REBs diagnostics, is reviewed.

Chapter 3 proposes a new method to find the proper frequency band of demodulation for bearing faults diagnosis. At first, undecimated wavelet packet transform (MODWPT) is adopted to split a signal in different frequency bands and central frequencies. Second, unbiased autocorrelation (AC) of the squared envelope of theses signals is calculated to take advantage of the 2nd order cyclostationarity of bearing faults signals, to reduce the level of uncorrelated random noise and enhance the fault related peaks. Third, kurtosis of the ACs is computed and a two dimensional colormap, named Autogram, is presented in order to locate the optimal frequency band for demodulation. Additionally, two modified versions of kurtosis equation are introduced which lead to two special cases of Autogram, namely Lower and Upper Autogram. The advantages and characteristics of each proposed Autogram and the conditions in which using Upper/Lower Autogram is beneficial are also discussed. Finally, the node with the highest kurtosis is chosen and Fourier transform is used to

obtain a frequency domain representation of the envelope signal so to identify the defect frequencies of bearings.

The main advantage of the algorithm is its ability in limiting the influence of non-periodic impulses and noise from raw time data, which are not related to any actual defects of bearings, thus enhancing the selection of a correct frequency band for the ensuing spectrum calculation. This valuable achievement is effectively reached by computing the autocorrelation of the envelope signal. Recognition of faults is based on defect frequencies identification so that it is not necessary to rely on undamaged conditions to diagnose the bearing status. On the other hand, Autogram is specifically designed to improve the detection of periodic impulses and therefore it is not suited to discover other sort of damages like pitting or corrosion.

The procedure can be fully automated to exploit the combined capabilities of Autogram and combined squared envelope spectrum (CSES), in order to generate a single spectrum where defect frequencies are clearly indicated.

The method has been tested on numerical data and experimental data provided by the Case Western Reserve University (CWRU) bearing data center, and compared with both literature results and other standard procedures so to assess its performances in REBs diagnosis. The results are indeed positive and Autogram allows to identify damaged bearings also in case of non-Gaussian noisy data.

In chapter 4 a second novel method for diagnosis of rolling element bearings has been developed. This approach is a generalized version of the cepstrum pre-whitening (CPW) which is a simple and effective technique for bearing diagnosis. CPW sets the amplitude spectrum of a signal to one and then reconstructs the signal only by using its phase. This gives rise to bearings signals which usually have lower energy levels and lower amplitude in the magnitude spectra than deterministic components from other sources. But giving the same value to all frequency lines may cause some problem as well. For instance, the level of noise in the reconstructed signals will be increased; in addition, signals related to resonance frequencies and periodic components generated by bearings defect may be completely masked by the increased level of noise. To overcome this difficulty, the new approach continually modifies the magnitude spectrum of a signal and then reconstructs the modified signals by using the original phase and the altered spectrum. Afterwards, the SES is computed for each of the modified signals and all the SESs are presented in a 3D plot. The superior performance of the proposed method has been shown on two real

case data (section 5.2). For the first case, the method successfully extracts bearing characteristic frequencies related to two defected bearings from the acquired signals. Moreover, the defect frequency is also highlighted in case two, even in presence of strong electromagnetic interference (EMI).

In chapter 5, both proposed methods and some standard bearings diagnostic techniques, reviewed in chapter 2, were implemented on signals from two different test rigs. Comparing all the methods on the same databases made it possible to highlight the pros and cons of them all.

For selection of the effective demodulation band, Autogram had better performance that Fast Kurtogram, Protrugram and Enhanced Kurtogram. But, the assumption behind these methods is that the bearing defect signal is the only component of the acquired signal in addition to white noise. However, for real cases this assumption is not always acceptable since signals from different sources are also present and some of them even have similar characteristics as bearing defect signals, periodicity and impulsivity.

Hence, better answers to REBs diagnostics are achieved by methods which extract all the modulation frequencies in signals. In this way, the chance of neglecting frequencies of interest is virtually eliminated. For this purpose, two methods were proposed in this thesis, combination of Autogram and CSES and new approach presented in chapter 4 and their outcome were compared to SC and the resultant EES. All the methods provide very good results. The second proposed method is the fastest and despite its simplicity has shown great potential for REBs diagnosis. It also was highlighted that EES for a selected frequency band generally offers notable results even for difficult cases, but in such cases the manual selection of the correct frequency band is compulsory which is in contrast to the methods proposed in this thesis. For the first method, this is done by selecting the nodes with valuable information from Autogram. For the second method, the filtering process is avoided and different components of a signal are automatically and effectively extracted based on the level of their magnitude spectrum.

The second part of this thesis chiefly focuses on modeling of planetary gearboxes. In chapter 6, the literature of their mathematical models was reviewed.

In chapter 7 interaction between gears and bearings of a planetary gear set in the presence of faults is investigated by developing a lumped parameter model. The capability of simultaneously simulating the response of the system for different sizes,

locations and profiles of bearing and gears defects is the significant advantage of this proposed model.

The acceleration signals of ring gear for healthy and defected bearings and gears are calculated and compared. Moreover, frequency components of signals associated to planets bearings defects for inner/outer race and rolling elements are discussed. In case of inner race faults, the orbital rotation of planets and in case of outer race/roller faults rotation of spalls relative to carrier as well as planet orbital rotation are sources of amplitude modulation of gearbox components signals and as a result sidebands cluster around the bearing damage frequencies. Furthermore, effects of faults on planets bearings and growth on frequency spectrum of the ring gear acceleration signal are investigated.

Influence of different levels of inner/outer race and rollers planet-bearing defects on dynamic response of the system are evaluated by implementing statistical indicators (RMS, kurtosis and M8A). It is concluded that these indicators are more effective when they are applied to the frequency domain data rather than to time domain signals. The quantities of these three statistical indicators increase as the spalls grow. M8A is the most sensitive feature and seems to have the potential to detects damages in the initial stages.

8.2 Future work

For future work regarding bearing diagnosis, the performance of the proposed methods can be verified on different available data set; in particular, signals collected from gearboxes which also include the interaction between bearings and gears. Moreover, the proposed approaches might be tested and, if necessary, be modified to cope with data acquired under special conditions such as varying shaft rotation speed.

On the topic of Autogram, different indicators such as negentropy used in Infogram or other indicators discussed in Ref. [60] could be used instead of kurtosis to investigate their effectiveness on the selection of the informative frequency band. Moreover, as it was shown by the analysis of various data sets, there is a major difference between SESs of simulated signals and typical signals acquired from real test rigs, especially in terms of the modulation sidebands. This issue can be investi-

gated further for better understanding of bearing defect signals and the interaction among signals from various sources. This may presumably lead to propose a more comprehensive model of bearing defect signals.

The second introduced method has shown great potential for bearings diagnosis but it has been suggested intuitively and it suffers from lack of mathematical background. Therefore, this method could be investigated from a theoretical point of view, which may open a new perspective. This technique has only investigated the effect of the magnitude spectrum manipulation, phase modification is another area which is worth examining.

The performance of EES as an enhanced version of SES can also be studied on the modified signals for different values of MO. Another suggestion could be employing log-envelope spectrum as an alternative to SES which makes the method more robust to cope with highly leptokurtic background noise which is often present in practice when the machine is subjected to highly impulsive phenomena. There are also some evidences showing correlation between the value of MO, when amplitude of a specific modulation frequency such as bearings characteristic frequency is maximized, and the severity of the defect. Although, this matter should be carefully explored.

Furthermore, the planetary gearbox model in chapter 7 can be enhanced to take into account the flexibility of gearbox casings/ring gear, high frequency resonant response of bearings, slippage in bearings and also variable speeds of gears. When the dynamic simulation planetary gearbox model will be realistic enough, its simulated signals could be employed to test the efficiency of diagnosis algorithms in presence of single or multiple defects, possibly with different sizes, of different kinds and located in the various elements of the transmission.

The presented methods are part of a larger problem. For instance, number and position of sensors, fusion of data from different sensors, developing indicators sensitive to evaluation of defects and estimation of remaining useful life of bearings are the areas worth investigating.

References

[1] Ian Howard. A review of rolling element bearing vibration'detection, diagnosis and prognosis'. Technical report, DTIC Document, 1994.

[2] Robert B Randall. State of the art in monitoring rotating machinery-part 1. *Sound and vibration*, 38(3):14–21, 2004.

[3] PD McFadden and JD Smith. Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of sound and vibration*, 96(1):69–82, 1984.

[4] PD McFadden and JD Smith. The vibration produced by multiple point defects in a rolling element bearing. *Journal of sound and vibration*, 98(2):263–273, 1985.

[5] D Ho and RB Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical systems and signal processing*, 14(5):763–788, 2000.

[6] Robert B Randall and Jérôme Antoni. Rolling element bearing diagnostics—a tutorial. *Mechanical systems and signal processing*, 25(2):485–520, 2011.

[7] William A Gardner, Antonio Napolitano, and Luigi Paura. Cyclostationarity: Half a century of research. *Signal processing*, 86(4):639–697, 2006.

[8] Jérôme Antoni. Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009.

[9] Robert B Randall, Jérôme Antoni, and S Chobsaard. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. *Mechanical systems and signal processing*, 15(5):945–962, 2001.

[10] J Antoni and RB Randall. Differential diagnosis of gear and bearing faults. *Transactions-American Society of Mechanical Engineers Journal of Vibration and Acoustics*, 124(2):165–171, 2002.

[11] J Antoni. Cyclic spectral analysis of rolling-element bearing signals: facts and fictions. *Journal of Sound and vibration*, 304(3):497–529, 2007.

[12] Robert Bond Randall. *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. John Wiley & Sons, 2011.

[13] D Dyer and RM Stewart. Detection of rolling element bearing damage by statistical vibration analysis. *Journal of mechanical design*, 100(2):229–235, 1978.

[14] Paul D Samuel and Darryll J Pines. A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of sound and vibration*, 282(1):475–508, 2005.

[15] N Tandon and A Choudhury. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology international*, 32(8):469–480, 1999.

[16] Andrew KS Jardine, Daming Lin, and Dragan Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7):1483–1510, 2006.

[17] Nader Sawalhi and RB Randall. Simulating gear and bearing interactions in the presence of faults: Part i. the combined gear bearing dynamic model and the simulation of localised bearing faults. *Mechanical Systems and Signal Processing*, 22(8):1924–1951, 2008.

[18] Zhipeng Feng, Ming Liang, and Fulei Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mechanical Systems and Signal Processing*, 38(1):165–205, 2013.

[19] C James Li and Jun Ma. Wavelet decomposition of vibrations for detection of bearing-localized defects. *Ndt & E International*, 30(3):143–149, 1997.

[20] Jin Lin and MJ Zuo. Gearbox fault diagnosis using adaptive wavelet filter. *Mechanical systems and signal processing*, 17(6):1259–1269, 2003.

[21] R Rubini and U Meneghetti. Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings. *Mechanical systems and signal processing*, 15(2):287–302, 2001.

[22] Yaguo Lei, Jing Lin, Zhengjia He, and Yanyang Zi. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 25(5):1738–1749, 2011.

[23] Mark S Darlow, Robert H Badgley, and GW Hogg. Application of high-frequency resonance techniques for bearing diagnostics in helicopter gearboxes. Technical report, Mechanical Technology Inc Latham NY, 1974.

[24] David J Ewins, Singiresu S Rao, and Simon G Braun. *Encyclopedia of Vibration, Three-Volume Set.* Academic press, 2002.

[25] Pietro Borghesani and Md Rifat Shahriar. Cyclostationary analysis with logarithmic variance stabilisation. *Mechanical Systems and Signal Processing*, 70:51–72, 2016.

[26] P Borghesani and J Antoni. Cs2 analysis in presence of non-gaussian background noise–effect on traditional estimators and resilience of log-envelope indicators. *Mechanical Systems and Signal Processing*, 90:378–398, 2017.

[27] Jérôme Antoni. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21(2):597–630, 2007.

[28] J Antoni. Cyclic spectral analysis of rolling-element bearing signals: facts and fictions. *Journal of Sound and vibration*, 304(3):497–529, 2007.

[29] Jérôme Antoni, Ge Xin, and Nacer Hamzaoui. Fast computation of the spectral correlation. *Mechanical Systems and Signal Processing*, 92:248–277, 2017.

[30] Jérôme Antoni and David Hanson. Detection of surface ships from interception of cyclostationary signature with the cyclic modulation coherence. *IEEE Journal of Oceanic Engineering*, 37(3):478–493, 2012.

[31] W Gardner. Measurement of spectral correlation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1111–1123, 1986.

[32] P Borghesani. The envelope-based cyclic periodogram. *Mechanical Systems and Signal Processing*, 58:245–270, 2015.

[33] Hans Konstantin-Hansen and Henrik Herlufsen. Envelope and cepstrum analyses for machinery fault identification. *Sound and Vibration*, 44(5):10, 2010.

[34] RB Randall, N Sawalhi, and M Coats. A comparison of methods for separation of deterministic and random signals. *International Journal of Condition Monitoring*, 1(1):11–19, 2011.

[35] Charles R Trimble. What is signal averaging? *Hewlett-Packard Journal*, 19(8):2–7, 1968.

[36] PD McFadden. A revised model for the extraction of periodic waveforms by time domain averaging. *Mechanical systems and signal processing*, 1(1):83–95, 1987.

[37] S Braun. The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing*, 25(4):1087–1102, 2011.

[38] Bernard Widrow, John R Glover, John M McCool, John Kaunitz, Charles S Williams, Robert H Hearn, James R Zeidler, JR Eugene Dong, and Robert C Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

[39] J Antoni and RB Randall. Unsupervised noise cancellation for vibration signals: part i—evaluation of adaptive algorithms. *Mechanical Systems and Signal Processing*, 18(1):89–101, 2004.

[40] J Antoni and RB Randall. Unsupervised noise cancellation for vibration signals: part ii —a novel frequency-domain algorithm. *Mechanical Systems and Signal Processing*, 18(1):103–117, 2004.

[41] Robert B Randall and Nader Sawalhi. A new method for separating discrete components from a signal. *Sound and Vibration*, 45(5):6, 2011.

[42] P Borghesani, P Pennacchi, RB Randall, N Sawalhi, and R Ricci. Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. *Mechanical Systems and Signal Processing*, 36(2):370–384, 2013.

[43] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[44] Cédric Peeters, Patrick Guillaume, and Jan Helsen. A comparison of cepstral editing methods as signal pre-processing techniques for vibration-based bearing fault detection. *Mechanical Systems and Signal Processing*, 91:354–381, 2017.

[45] Ralph A Wiggins. Minimum entropy deconvolution. *Geoexploration*, 16(1-2):21–35, 1978.

[46] N Sawalhi, RB Randall, and H Endo. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems and Signal Processing*, 21(6):2616–2633, 2007.

[47] H Endo and RB Randall. Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter. *Mechanical Systems and Signal Processing*, 21(2):906–919, 2007.

[48] Tomasz Barszcz and Nader Sawalhi. Fault detection enhancement in rolling element bearings using the minimum entropy deconvolution. *Archives of acoustics*, 37(2):131–141, 2012.

[49] Geoff L McDonald, Qing Zhao, and Ming J Zuo. Maximum correlated kurtosis deconvolution and application on gear tooth chip fault detection. *Mechanical Systems and Signal Processing*, 33:237–255, 2012.

[50] Carlos A Cabrelli. Minimum entropy deconvolution and simplicity: A noniterative algorithm. *Geophysics*, 50(3):394–413, 1985.

[51] Geoff L McDonald and Qing Zhao. Multipoint optimal minimum entropy deconvolution and convolution fix: application to vibration fault detection. *Mechanical Systems and Signal Processing*, 82:461–477, 2017.

[52] NG Nikolaou and IA Antoniadis. Demodulation of vibration signals generated by defects in rolling element bearings using complex shifted morlet wavelets. *Mechanical systems and signal processing*, 16(4):677–694, 2002.

[53] Jérôme Antoni. The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006.

[54] Jérôme Antoni and RB Randall. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20(2):308–331, 2006.

[55] Jérôme Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007.

[56] Tomasz Barszcz and Adam JabŁoński. A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram. *Mechanical Systems and Signal Processing*, 25(1):431–451, 2011.

[57] Dong Wang, W Tse Peter, and Kwok Leung Tsui. An enhanced kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 35(1):176–199, 2013.

[58] Jérôme Antoni. The infogram: Entropic evidence of the signature of repetitive transients. *Mechanical Systems and Signal Processing*, 74:73–94, 2016.

[59] Xiaohui Gu, Shaopu Yang, Yongqiang Liu, and Rujiang Hao. Rolling element bearing faults diagnosis based on kurtogram and frequency domain correlated kurtosis. *Measurement Science and Technology*, 27(12):125019, 2016.

[60] Jakub Obuchowski, Agnieszka Wyłomańska, and Radosław Zimroz. Selection of informative frequency band in local damage detection in rotating machinery. *Mechanical Systems and Signal Processing*, 48(1):138–152, 2014.

[61] Wade A Smith, Zhiqi Fan, Zhongxiao Peng, Huaizhong Li, and Robert B Randall. Optimised spectral kurtosis for bearing diagnostics under electromagnetic interference. *Mechanical Systems and Signal Processing*, 75:371–394, 2016.

[62] Chuan Li, Diego Cabrera, José Valente de Oliveira, René-Vinicio Sanchez, Mariela Cerrada, and Grover Zurita. Extracting repetitive transients for rotating machinery diagnosis using multiscale clustered grey infogram. *Mechanical Systems and Signal Processing*, 76:157–173, 2016.

[63] Ali Moshrefzadeh and Alessandro Fasana. The autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis. *Mechanical Systems and Signal Processing*, 2018.

[64] Jérôme Antoni. Fast kurtogram. https://mathworks.com/matlabcentral/fileexchange/48912-fast-kurtogram.

[65] Wade A Smith and Robert B Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mechanical Systems and Signal Processing*, 64:100–131, 2015.

[66] Andrew T Walden. Wavelet analysis of discrete time series. In *European Congress of Mathematics*, pages 627–641. Springer, 2001.

[67] Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge university press, 2006.

[68] Jérôme Antoni. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21(2):597–630, 2007.

[69] Case western reserve university bearing data center website. http://csegroups.case.edu/bearingdatacenter/home.

[70] A Moshrefzadeh, A Fasana, and L Garibaldi. Using unbiased autocorrelation to enhance kurtogram and envelope analysis results for rolling element bearing diagnostics. In *International Conference Surveillance 9, Morocco*, 2017.

[71] Hai Qiu, Jay Lee, Jing Lin, and Gang Yu. Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of sound and vibration*, 289(4-5):1066–1090, 2006.

[72] Nasa ames prognostics data repository.

[73] William Gousseau, Jérôme Antoni, François Girardin, and Julien Griffaton. Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the university of cincinnati. In *The Thirteenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Paris, France*, pages 1–13, 2016.

[74] Christopher G Cooley and Robert G Parker. A review of planetary and epicyclic gear dynamics and vibrations research. *Applied Mechanics Reviews*, 66(4):040804, 2014.

[75] F Cunliffe, JD Smith, and DB Welbourn. Dynamic tooth loads in epicyclic gears. *Journal of Engineering for Industry*, 96(2):578–584, 1974.

[76] M Botman. Epicyclic gear vibrations. *Journal of Engineering for Industry*, 98(3):811–815, 1976.

[77] JL Frater, R August, and FB Oswald. Vibration in planetary gear systems with unequal planet stiffnesses. *NASA Technical Memorandum 83428*, 1982.

[78] A Saada and P Velex. An extended model for the analysis of the dynamic behavior of planetary trains. *TRANSACTIONS-AMERICAN SOCIETY OF MECHANICAL ENGINEERS JOURNAL OF MECHANICAL DESIGN*, 117:241–241, 1995.

[79] A Kahraman. Natural modes of planetary gear trains. *Journal of sound and vibration*, 173(1):125–130, 1994.

[80] Jian Lin and RG Parker. Analytical characterization of the unique properties of planetary gear free vibration. *Journal of vibration and acoustics*, 121(3):316–321, 1999.

[81] J Lin and RG Parker. Structured vibration characteristics of planetary gears with unequally spaced planets. *Journal of Sound and Vibration*, 233(5):921–928, 2000.

[82] R Kasuba and R August. Torsional vibrations and dynamic loads in a basic planetary gear system. *ASME J. Vib., Acoust., Stress, Reliab. Des*, 108:348–353, 1986.

[83] Ahmet Kahraman. Load sharing characteristics of planetary transmissions. *Mechanism and Machine Theory*, 29(8):1151–1165, 1994.

[84] Tao Sun and HaiYan Hu. Nonlinear dynamics of a planetary gear system with multiple clearances. *Mechanism and Machine Theory*, 38(12):1371–1390, 2003.

[85] Vijaya Kumar Ambarisha and Robert G Parker. Nonlinear dynamics of planetary gears using analytical and finite element models. *Journal of sound and vibration*, 302(3):577–595, 2007.

[86] Yi Guo and Robert G Parker. Dynamic modeling and analysis of a spur planetary gear involving tooth wedging and bearing clearance nonlinearity. *European Journal of Mechanics-A/Solids*, 29(6):1022–1033, 2010.

[87] Yi Guo and Robert G Parker. Dynamic analysis of planetary gears with bearing clearance. *Journal of Computational and Nonlinear Dynamics*, 7(4):041002, 2012.

[88] Yi Guo, Jonathan Keller, and Robert G Parker. Nonlinear dynamics and stability of wind turbine planetary gear sets under gravity effects. *European Journal of Mechanics-A/Solids*, 47:45–57, 2014.

[89] F Chaari, T Fakhfakh, and M Haddar. Dynamic analysis of a planetary gear failure caused by tooth pitting and cracking. *Journal of Failure Analysis and Prevention*, 6(2):73–78, 2006.

[90] Zaigang Chen and Yimin Shao. Dynamic simulation of planetary gear with tooth root crack in ring gear. *Engineering Failure Analysis*, 31:8–18, 2013.

[91] Zaigang Chen and Yimin Shao. Dynamic features of a planetary gear system with tooth crack under different sizes and inclination angles. *Journal of Vibration and Acoustics*, 135(3):031004, 2013.

[92] Zaigang Chen, Zhifang Zhu, and Yimin Shao. Fault feature analysis of planetary gear system with tooth root crack and flexible ring gear rim. *Engineering Failure Analysis*, 49:92–103, 2015.

[93] Cheon-Jae Bahk and Robert G Parker. Analytical investigation of tooth profile modification effects on planetary gear dynamics. *Mechanism and machine theory*, 70:298–319, 2013.

[94] Xionghua Wu and Robert G Parker. Modal properties of planetary gears with an elastic continuum ring gear. *Journal of Applied Mechanics*, 75(3):031014, 2008.

[95] Jun Zhang, Yi Min Song, and Jin You Xu. A discrete lumped-parameter dynamic model for a planetary gear set with flexible ring gear. In *Applied Mechanics and Materials*, volume 86, pages 756–761. Trans Tech Publ, 2011.

[96] Sharad Jain. *Skidding and fault detection in the bearings of wind-turbine gearboxes*. PhD thesis, University of Cambridge, 2013.

[97] Murat Inalpolat and A Kahraman. A theoretical and experimental investigation of modulation sidebands of planetary gear sets. *Journal of Sound and Vibration*, 323(3):677–696, 2009.

[98] Murat Inalpolat and Ahmet Kahraman. A dynamic model to predict modulation sidebands of a planetary gear set having manufacturing errors. *Journal of Sound and Vibration*, 329(4):371–393, 2010.

[99] Zhipeng Feng and Ming J Zuo. Vibration signal models for fault diagnosis of planetary gearboxes. *Journal of Sound and Vibration*, 331(22):4919–4939, 2012.

[100] Libin Liu, Xihui Liang, and Ming J Zuo. Vibration signal modeling of a planetary gear set with transmission path effect analysis. *Measurement*, 85:20–31, 2016.

[101] Ali Moshrefzadeh and Alessandro Fasana. Planetary gearbox with localised bearings and gears faults: simulation and time/frequency analysis. *Meccanica*, 52(15):3759–3779, 2017.

[102] Xihui Liang, Ming J Zuo, and Mohammad R Hoseini. Vibration signal modeling of a planetary gear set for tooth crack detection. *Engineering Failure Analysis*, 48:185–200, 2015.

[103] DCH Yang and JY Lin. Hertzian damping, tooth friction and bending elasticity in gear impact dynamics. *Journal of mechanisms, transmissions, and automation in design*, 109(2):189–196, 1987.

[104] Xinhao Tian. *Dynamic simulation for system response of gearbox including localized gear faults*. Library and Archives Canada= Bibliothèque et Archives Canada, 2005.

[105] Zaigang Chen and Yimin Shao. Mesh stiffness calculation of a spur gear pair with tooth profile modification and tooth root crack. *Mechanism and Machine Theory*, 62:63–74, 2013.

[106] M Iglesias, A Fernandez del Rincon, A de Juan, A Diez-Ibarbia, P Garcia, and F Viadero. Advanced model for the calculation of meshing forces in spur gear planetary transmissions. *Meccanica*, 50(7):1869–1894, 2015.

[107] RG Parker and Jian Lin. Mesh phasing relationships in planetary and epicyclic gears. In *ASME 2003 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 525–534. American Society of Mechanical Engineers, 2003.

[108] CS Sunnersjö. Varying compliance vibrations of rolling bearings. *Journal of sound and vibration*, 58(3):363–373, 1978.

[109] Satoru FUKATA, Emil Halim GAD, Takahiro KONDOU, Takashi AYABE, and Hideyuki TAMURA. On the radial vibration of ball bearings: computer simulation. *Bulletin of JSME*, 28(239):899–904, 1985.

[110] Tedric A Harris and Michael N Kotzalas. *Rolling bearing analysis*. CRC/Taylor & Francis,, 2006.

[111] Jing Liu, Yimin Shao, and Teik C Lim. Vibration analysis of ball bearings with a localized defect applying piecewise response function. *Mechanism and Machine Theory*, 56:156–169, 2012.

[112] Xihui Liang, Ming J Zuo, and Mayank Pandey. Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set. *Mechanism and Machine Theory*, 76:20–38, 2014.

[113] Hsiang H Lin and Chuen-Huei Liou. A parametric study of spur gear dynamics. Technical report, DTIC Document, 1998.

[114] David G Lewicki, Kelsen E LaBerge, Ryan T Ehinger, and Jason Fetty. Planetary gearbox fault detection using vibration separation techniques. *NASA/TM—2011-217127*, 2011.