

Bearing fault detection with application to PHM Data Challenge

Pavle Boškosi¹, and Anton Urevc²

¹ *Jožef Stefan Institute, Ljubljana, Slovenia*
pavle.boskoski@ijs.si

² *Centre for Tribology and Technical Diagnostics,*
Faculty of Mechanical Engineering,
University of Ljubljana, Slovenia
anton.urevc@ctd.uni-lj.si

ABSTRACT

Mechanical faults in the items of equipment can result in partial or total breakdown, destruction and even catastrophes. By implementation of an adequate fault detection system the risk of unexpected failures can be reduced. Traditionally, fault detection process is done by comparing the feature sets acquired in the faulty state with the ones acquired in the fault-free state. However, such historical data are rarely available. In such cases, the fault detection process is performed by examining whether a particular pre-modeled fault signature can be matched within the signals acquired from the monitored machine. In this paper we propose a solution to a problem of fault detection without any prior data, presented at PHM'09 Data Challenge. The solution is based on a two step algorithm. The first step, based on the spectral kurtosis method, is used to determine whether a particular experimental run is likely to contain a faulty element. In case of a positive decision, fault isolation procedure is applied as the second step. The fault isolation procedure was based on envelope analysis of band-pass filtered vibration signals. The band-pass filtering of the vibration signals was performed in the frequency band that maximizes the spectral kurtosis. The effectiveness of the proposed approach was evaluated for bearing fault detection, on the vibration data obtained from the PHM'09 Data Challenge.

1. INTRODUCTION

Stable and predictable condition of process equipment, high process availability and reliability are key factors that keep a company competitive. However, wear, material stress and environmental influences can cause mechanical faults which result in equipment breakdowns. Since the emergence of faults

is inevitable, it is of utmost importance to construct an effective fault detection system capable of detecting these faults in their incipient phase. Such early detection will prevent unscheduled interruptions in the machine operation, which effectively will increase the overall performance.

In the research domain there is an impressive body of literature that addresses the issues of fault detection. One group of authors mainly focuses on modeling the vibration signals generated by a specific mechanical element, like gears, bearings etc. In that manner, (Bartelmus, 2001) developed model for vibrations produced by meshing gears. (Endo, Randall, & Gosselin, 2009) developed models for gear vibrations under specific tooth faults. Similarly, (Tandon & Choudhury, 1999) derived the relations for the principle frequencies components in the vibration signals produced by localized bearing faults. Although the fault signatures have been thoroughly examined, the detection of the faults in their incipient phase has proved to be a difficult task. Many authors have addressed this issue by employing variety of signal processing techniques. The envelope spectrum analysis of the vibration signals has been one of the most commonly used methods (Rubini & Meneghetti, 2001; Ho & Randall, 2000). However, several authors have shown that a significant increase in the sensitivity of the envelope analysis can be achieved by calculating the envelope spectrum of a filtered signal. In that manner (Wang, 2001) proposed filtering the vibration signals around the system's resonance frequency, (Staszewski, 1998) used wavelet denoising techniques, (Sawalhi, Randall, & Endo, 2007) used spectral kurtosis method for determining the most suitable frequency band.

The majority of the listed methods have been successfully applied in systems where historical data of the fault-free state have been previously acquired, so that any small deviation from the fault-free case could be detected. Conversely, such historical data of the fault-free system were not available on the set of signals acquired from the PHM-09 test-rig (PHM,

Boškosi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2009). In absence of a base–line determining the fault–free runs, the fault detection procedure was implemented using a two–step approach. The first step is used to decide whether the observed experimental run is likely to belong to a faulty machine or not. Provided the observed run is associated to the faulty condition, the second step is applied. This step consists of a fault diagnosis process, in which the fault origins are determined.

The decision whether a machine state is faulty or fault–free was based on the maximal value of the spectral kurtosis (SK) of the acquired vibration signals for a particular run. Such a choice is based on the property of SK for detection of transients, which is explained in more details in Section 4. Additional results of the SK calculation process are the frequency band parameters, central frequency f_c and bandwidth B_w , where the SK maximum resides. These parameters are used to filter the vibration signals. The filtered signals are then used to calculate the envelope spectrum. The resulting spectrum is afterward used as a starting point for the fault detection procedure. Despite the lack of data from the fault–free motor run, the proposed two–step fault detection approach has proved capable of detecting bearing faults on the PHM'09 test–rig (PHM, 2009).

The presentation of the proposed algorithm in this paper is organized as follows. In Section 2 we present the basics of the bearing fault frequency and the used bearing model. A brief overview of the envelope analysis and spectral kurtosis methods are given in Section 3. A detailed description of the proposed fault detection algorithm, with a proposed simplification procedure are presented in Section 4.

2. BEARING FAULTS

In a presence of localized bearing fault, impacts occur every time bearing's roller element passes over the damaged area. Each of these impacts excites an impulse response of the observed bearing, i.e. exponentially decaying oscillation $s(t)$. Under the assumption that the bearing rotates with constant rotational frequency f_{rot} , these impulses will be periodic with some period T . The period T is directly connected to the type of the localized fault. It can be considered that $T = 1/f_e$, where f_e is one of the principle bearing fault frequencies (Tandon & Choudhury, 1999):

$$\begin{aligned} f_{BPFO} &= \frac{Zf_{rot}}{2} \left(1 - \frac{d}{D} \cos\alpha\right) \\ f_{BPF1} &= \frac{Zf_{rot}}{2} \left(1 + \frac{d}{D} \cos\alpha\right) \\ f_{FTF} &= \frac{f_{rot}}{2} \left(1 - \frac{d}{D} \cos\alpha\right) \\ f_{BSF} &= \frac{Df_{rot}}{2d} \left(1 - \left(\frac{d}{D} \cos\alpha\right)^2\right), \end{aligned} \quad (1)$$

where f_{BPFO} is the ball passing frequency of the outer race, f_{BPF1} is the ball passing frequency of the inner race, f_{FTF} is the fundamental train frequency and f_{BSF} is the ball spin frequency, Z is the number of rolling elements, d is the rolling element diameter, D is the pitch diameter, α is the contact angle and f_{rot} is the inner ring rotational speed.

The vibrations $x(t)$ produced by a localized bearing fault may be written as

$$x(t) = \sum_{i=-\infty}^{+\infty} s(t - iT). \quad (2)$$

Although the rotation frequency f_{rot} may be considered as constant, small speed fluctuations are always present. Additionally the speed of the roller element which enters the load zone slightly differs from the one that is outside the load zone. Such random fluctuations are expressed as a time lag of occurrence of each impact, in particular

$$x(t) = \sum_{i=-\infty}^{+\infty} s(t - iT - \tau_i), \quad (3)$$

where τ_i represents the time lag of occurrence of the i^{th} impact.

In Eq. (3) we considered that all impacts have same amplitude. However, each impact excites an impulse with somewhat different amplitude due to changes in the bearing surface, the way the ball enters the damaged region etc. Such random changes are incorporated by adding a factor A_i , which represents the amplitude of i^{th} impact

$$x(t) = \sum_{i=-\infty}^{+\infty} A_i s(t - iT - \tau_i). \quad (4)$$

Finally, in order to take into account all surrounding non–modeled vibrations as well as any other environmental influence, a purely random component $n(t)$ is added to Eq. (4). The final model of bearing vibrations becomes (Randall, Antoni, & Chobsaard, 2001):

$$x(t) = \sum_{i=-\infty}^{+\infty} A_i s(t - iT - \tau_i) + n(t). \quad (5)$$

3. METHODS OVERVIEW

The vibration signal defined by (5) is an amplitude modulated (AM) signal, where the modulation itself can be considered as a random signal. The information about the present fault is contained in the mean period T of the occurrence of the impacts. Hence all the needed diagnostic information is contained within the signal's envelope. When the amplitudes A_i in (5) are high enough a fairly simple analysis of the envelope spectrum is sufficient for the fault diagnosis process. However, when these amplitudes are small and dominated by the noise $n(t)$, the simple envelope spectrum analysis turns to be

In the case of PHM'09 challenge all bearings were ball bearings

ineffective in extracting information about the fault. A way around this problem is to calculate the envelope of the signal after the vibration signal has been let through the band-pass filter centered at a selected carrier frequency. An improper selection of band-pass filtering parameters can significantly hinder the effectiveness of the fault detection process. As a result of this effect, a variety of different approaches have been developed offering different solutions to the issue of band-pass filter parameter selection. In our case we adopted the spectral kurtosis method, which has proved capable of detection transients buried in noise.

3.1 Spectral Kurtosis

The spectral kurtosis method was firstly introduced by (Dwyer, 1983), as a method that is able to distinguish between transients (impulses and unsteady harmonic components) and stationary sinusoidal signals in background Gaussian noise. Spectral kurtosis takes high values for frequency bands where the vibration signal $x(t)$, defined with Eq.(5), is dominated by the corresponding impulses, and it takes low values for frequency bands where the signal is dominated by the Gaussian noise $n(t)$ or stationary periodic components. If we rewrite the signal from Eq.(5) as

$$x(t) = y(t) + n(t), \quad (6)$$

where

$$y(t) = \sum_i A_i s(t - iT - \tau_i), \quad (7)$$

than the SK values $K_x(f)$ for the signal $x(t)$ contaminated by additive noise $n(t)$ can be calculated as (Antoni & Randall, 2006)

$$K_x(f) = \frac{K_y(f)}{[1 + \rho(f)]^2}, \quad (8)$$

where $K_y(f)$ is the spectral kurtosis of the signal $y(t)$, and $\rho(f)$ is the noise-to-signal ratio for that particular frequency f . The value for $K_y(f)$ can be obtained using the following relation

$$K_y(f) = \frac{S_{4y}(f) - 2S_{2y}^2(f)}{S_{2y}^2(f)}, \quad (9)$$

where $S_{2y}(f)$ and $S_{4y}(f)$ are the second and fourth spectral moments respectively. The maximum of Eq.(8), actually determines the frequency band where the signal-to-noise ratio in the observed signal is the biggest and at the same time the observed signal $x(t)$ is the closest to the original, uncontaminated signal, $y(t)$.

The definition of SK given by the Eq.(9) bears resemblance with the statistical definition of kurtosis. However the actual physical interpretation and its ability for detection of non-stationary transients in signals is not so obvious. One way for clarification is to observe the time-frequency characteristic of the simulated vibration signal $x(t)$, defined by Eq. (5). The simulation was conducted with $T = 333$ Hz, $s(t) =$

$e^{-100t} \sin(2\pi 1500t)$ and $\text{SNR}=1$. The random time fluctuations τ_i and random amplitudes A_i were Gaussian random processes with zero-mean with $\sigma_{\tau_i} = 0.05T$ and $\sigma_{A_i} = 0.5$ respectively (cf. Figure 1).

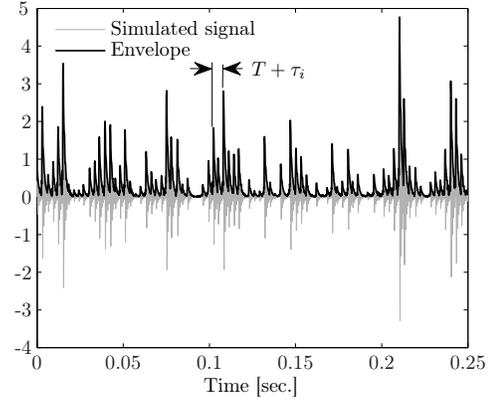


Figure 1. Simulated signal $x(t)$, Eq. (5)

The time-frequency characteristic of the simulated signal is shown in Figure 2. It is noticeable that the highest peaks are around 1.5 kHz, which was the chosen simulated eigenfrequency of the impulse response $s(t)$. The amplitudes of the spectral components around 1.5 kHz vary in time considerably, compared to the ones above and below this frequency band. By calculating the average over time we will obtain the standard power-spectral density (PSD).

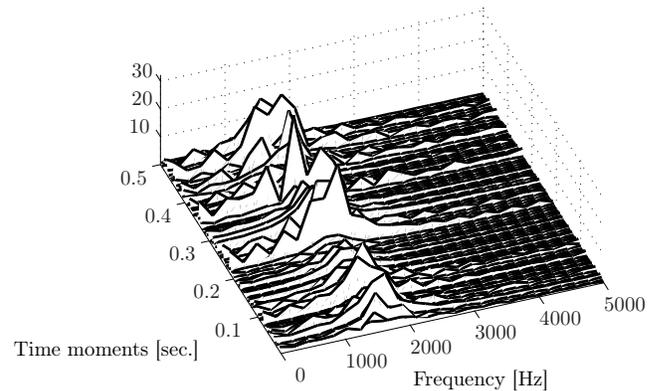


Figure 2. Time-frequency characteristic of the simulated signal $x(t)$ Eq. (5)

If we now consider the changes of the amplitude of particular spectral components as a stochastic process, the spectral kurtosis actually searches for the frequency band where this stochastic process exhibits the highest kurtosis. For our simulated signal Eq. (5), the frequency band in question is around the resonance frequency of 1.5 kHz. The time change of the amplitude of that spectral component compared with its average value is shown in Figure 3.

It can be concluded that for non-stationary processes these

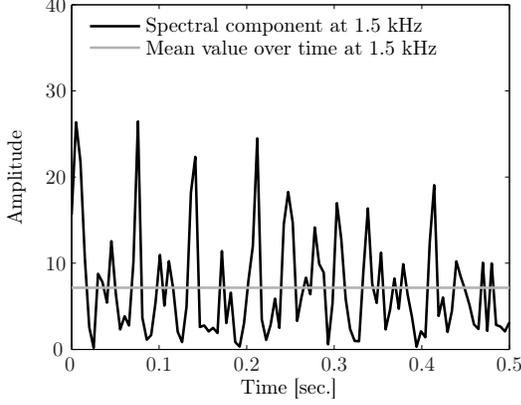


Figure 3. Changes of amplitude of the spectral component at 1.5 kHz over time

discrepancies in amplitude of some spectral components in time will be more expressed then in the cases of stationary processes. Consequently, we can use the SK as an indicator for a frequency band where the signal's non-stationarities are most expressed.

The estimation of the time-frequency characteristics of the simulated signal was based on the Short-time Fourier transform (STFT). Since the resolution in time and in frequency depends on the used window length, the search for the best frequency range is performed by examining the SK value for several window lengths (Antoni, 2007). This procedure produces a diagram called *kurtogram*. The kurtogram diagram for the simulated signal $x(t)$ defined by Eq. (5) is shown in Figure 4.

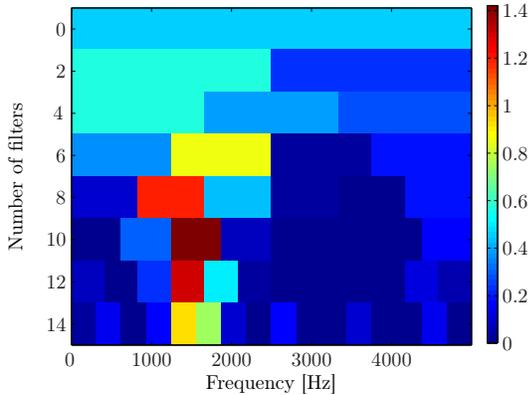


Figure 4. Kurtogram of the simulated signal $x(t)$ defined by Eq. (5)

From the kurtogram shown in Figure 4, we can see that the maximal value of SK is obtained in the frequency band with central frequency $f_c = 1500$ Hz and bandwidth $B_w = 625$ Hz, and the maximum of SK in that frequency band is 1.4.

For a comprehensive derivation and all properties of SK one

should refer the following references (Antoni, 2006; Antoni & Randall, 2006; Sawalhi et al., 2007).

3.2 Envelope analysis

After filtering the signal in the frequency range determined by the SK method, the final step in the fault detection process is done by envelope analysis. The use of envelope analysis is justified, because the information about the fault is in the occurrence period T of the quasi-periodic impact impulses. For the simulated signal $x(t)$ defined by Eq. (5), the envelope is marked with black line in Figure 1.

The signal's envelope is obtained from the Hilbert transform, i.e. by analyzing the amplitude of the analytical signal $x_a(t)$. The analytical signal $x_a(t)$ is a complex signal whose real part is the original signal $x(t)$, and the imaginary part is the Hilbert transform of the original signal $x(t)$

$$x_a(t) = x(t) + i\mathcal{H}[x(t)], \quad (10)$$

where $\mathcal{H}[x(t)]$ is the Hilbert transform of the signal $x(t)$

$$\mathcal{H}[x(t)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau. \quad (11)$$

Fourier transform of the analytical signal $x_a(t)$ is:

$$\begin{aligned} X_a(t)(f) &= X(f) + j\mathcal{H}[X(f)] \\ &= \begin{cases} 2X(f) & \text{for } f > 0, \\ X(f) & \text{for } f = 0, \\ 0 & \text{for } f < 0. \end{cases} \end{aligned} \quad (12)$$

This shows that the analytical signal has spectrum only in the positive frequency range. By calculating the amplitude of the analytical signal (10)

$$a(t) = \sqrt{x^2(t) + \mathcal{H}^2[x(t)]} \quad (13)$$

we obtain the envelope of the signal.

The spectrum of the envelope of the simulated signal $x(t)$ is shown in Figure 5. From that spectrum we can easily identify the spectral component at 333 Hz, which is the period of the occurrence of the simulated impacts.

3.3 Alternative way of estimating signal envelope

Although the calculation of the signals envelope using the Hilbert transform can be performed using Fast Fourier transform (FFT), a good estimation can be done by calculating the spectrum of the absolute value of the signal (Benko, Petrovčić, Musizza, & Juričić, 2008). By doing so we skip one FFT calculation of the whole signal. Namely, calculating envelope spectrum using the Hilbert transform we first calculate FFT of the original signal. Then we obtain the analytical signal $x_a(t)$ by employing the Eq. (12). As the last step we calculate FFT of $x_a(t)$, thus obtaining the envelope spectrum. Using the estimation method for envelope spectrum, we just have to calculate one FFT of the $|x(t)|$.

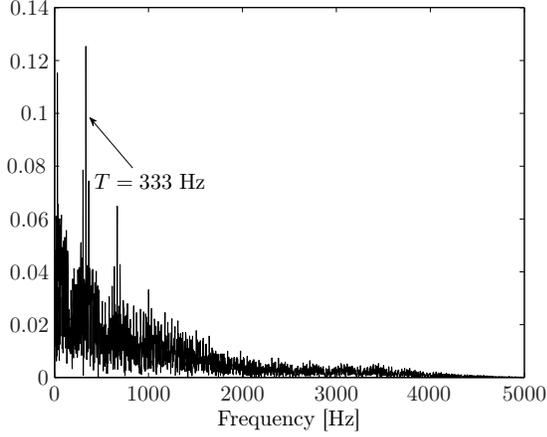


Figure 5. Envelope spectrum of the simulated signal $x(t)$, defined by Eq. (5)

This procedure will be presented by analyzing the envelope of an amplitude modulated (AM) signal $y_{am}(t)$ in form:

$$\begin{aligned} y_{am}(t) &= [C + m(t)]\cos(\omega_c t) \\ &= y_0(t)\cos(\omega_c t), \end{aligned} \quad (14)$$

where $y_0(t) = [C + m(t)]$ is the modulation signal, or the envelope, and $\omega_c = 2\pi/T_c$ is the carrier frequency.

The Fourier transform of the absolute value $|y_{am}(t)|$ can be obtained as

$$\begin{aligned} \mathcal{F}\{|y_{am}(t)|\} &= \sum_p \int_{T_p} y_{am}(t)e^{-j\omega t} dt \\ &\quad - \sum_q \int_{T_q} y_{am}(t)e^{-j\omega t} dt \\ &= \mathcal{F}\{y_{am}(t)\}(\omega) \\ &\quad - 2 \sum_q \int_{T_q} y_{am}(t)e^{-j\omega t} dt, \end{aligned} \quad (15)$$

where T_p is one of the intervals where the signal $y(t)$ is positive, and T_q is one of the intervals where $y(t)$ is negative. In lower frequencies $\omega_0 < \omega_{max}$, the spectrum of the observed signal is zero, i.e. $\mathcal{F}\{y_o(t)\}(\omega) = 0, \forall \omega \geq \omega_{max}$, where $0 < \omega_{max} \ll \omega_c$. Under these conditions the Eq. (15) can be rewritten as

$$\begin{aligned} \mathcal{F}\{|y_{am}(t)|\}(\omega_0) &= \\ &= -2 \sum_q \int_{T_q} y_{am}(t)e^{-j\omega_0 t} dt \\ &= -2 \sum_q \int_{T_q} [C + m(t)]\cos(\omega_c t)e^{-j\omega_0 t} dt. \end{aligned} \quad (16)$$

If we now allow $\omega_c \rightarrow \infty$, the intervals T_q will tend to zero, and within the observed interval the function $[C + m(t)]\cos(\omega_c t)e^{-j\omega_0 t}$ may be considered as constant. The observed value is the value of the function in moment t_q , which

is the middle of the T_q interval. Under these assumptions Eq. (16) becomes

$$\begin{aligned} \lim_{\omega_c \rightarrow \infty} \mathcal{F}\{|y_{am}(t)|\}(\omega_0) &= \\ &= -2 \sum_q [C + m(t_n)]e^{-j\omega_0 t_n} \int_{T_q} \cos(\omega_c t) dt = \\ &= -2 \sum_q [C + m(t_n)]e^{-j\omega_0 t_n} \left(-\frac{2}{\omega_c}\right) = \\ &= 2 \sum_q [C + m(t_n)]e^{-j\omega_0 t_n} \frac{T_c}{\pi}. \end{aligned} \quad (17)$$

Since we have allowed $\omega_c \rightarrow \infty$, consequently the period $T_c \rightarrow 0$. In such case we can change the summation to integration and the Eq. (17) becomes

$$\lim_{T_c \rightarrow 0} \mathcal{F}\{|y_{am}(t)|\}(\omega_0) = \frac{2}{\pi} \int [C + m(t)]e^{-j\omega_0 t} dt. \quad (18)$$

The last part of Eq. (18) is actually the Fourier transform of the signals envelope. Finally we can conclude that when $\omega_0 \ll \omega_c$ then

$$\mathcal{F}\{|y_{am}(t)|\}(\omega_0) \propto \mathcal{F}\{C + m(t)\}(\omega_0). \quad (19)$$

4. FAULT DETECTION PROCEDURE

Traditionally, the presence of fault is determined based on a results of a spectrum comparison with previous data (fault-free data) (Sawalhi & Randall, 2008). Since such set of fault-free data was unavailable, we had to use a different approach in order to detect the faulty experimental runs. We have adopted a two step approach. Firstly we have constructed a set of experimental runs that are most likely to represent a run with faulty element. As a second step, each of these presumably faulty runs was analyzed using envelope spectra of the filtered signal in the frequency band that maximizes the SK value.

Since the PHM'09 Data challenge consisted of 560 test-runs conducted under 5 different running speeds, we have divided the complete data set into 5 batches each comprising experiments conducted under the same rotational speed f_{rot} . Such a pre-processing step was necessary since the principle bearing fault frequencies, defined by Eq. (1), are dependent on rotational speed f_{rot} . Thus we have analyzed 5 different batches consisting of 112 runs each using the same algorithm. The results presented in this section refer to only one group, but the same principle applies to the remaining 4 batches.

The decision, whether a particular experimental run is likely to contain a faulty element or not, was based on the maximum value of the SK for the particular measurement. The idea behind this approach is the property of the SK which states (Antoni & Randall, 2006) that the value of SK $K_x(f)$, defined by Eq. (8), increases with the intensity of the fluctuations in the impulse amplitudes. Consequently the value of the SK can be used as an indication of the severity of the

damage. So SK value was calculated for each measurement separately and the measurements were sorted by decreasing values of SK. Thus, the measurements at the top of the list were more likely to represent faulty runs, and those at the bottom of the list were considered as more likely to be fault-free runs.

In the second step, the fault isolation procedure, started by analyzing measurements from the top of the sorted list. Each signal from the list was filtered in the frequency band in which the value of SK was the highest. After that the envelope spectrum of such filtered signal was calculated by applying the Eq. (19), although one might also opt for the calculation of the signal's envelope based on Hilbert transform and Eq. (13). The fault detection was based on the amplitudes of spectral components at the bearing characteristic frequencies, defined by Eq. (1), obtained from the envelope spectrum. A block diagram of the described approach is shown in Figure 6.

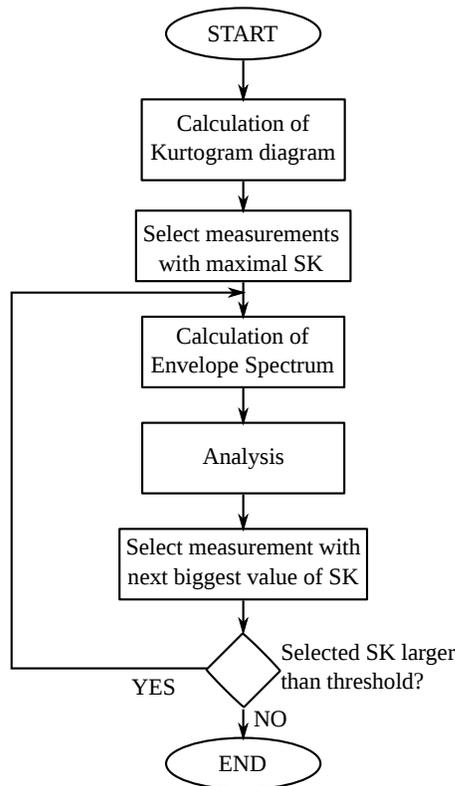


Figure 6. Block diagram of the used fault detection algorithm

After the analysis of all experimental runs from the top of the list, the next step is to determine whether the remaining measurements consist of only fault-free runs. In this is the case the algorithm ends. The decision whether the remaining runs contain experiments with damaged elements is based on the value of SK. The values of SK for a group of 112 measurements for one particular speed is shown in Figure 7. Under assumption that the fault-free runs have the smallest value of

SK, we can determine a threshold that can serve as a boundary between faulty and fault-free runs. The procedure for selection of the threshold takes into account the fact that SK values for multiple measurements with same kind of fault should have similar values. Therefore, if we take a window consisting of several neighboring measurements and calculate the interquartile interval on their SK values, we can use the width of the interquartile interval to determine whether the measurements originate from the same machine configuration or not. If they do originate from the same configuration, the width of the interquartile interval will be smaller. Otherwise it will be significantly larger. Thus, we can use the width of interquartile interval as an indication which neighboring measurements have been done in different machine configuration.

The first change in the machine configuration occurs when some kind of fault was introduced into a fault-free machine. In order to identify this change we analyzed the interquartile intervals of the measurements with lowest value of SK. The values of the interquartile intervals are shown in Figure 8 using a box plot. From the figure we can detect that the measurements with the smallest width of the interquartile interval are found for measurements 18, 19 and 20. This leads to a conclusion that the first change in the machine configuration occurred several measurements before i.e. somewhere between the 14th and the 17th measurement. In this interval the 16th measurement has the widest interquartile interval. Therefore, as a threshold we have chosen the upper interquartile value of the 16th measurement.

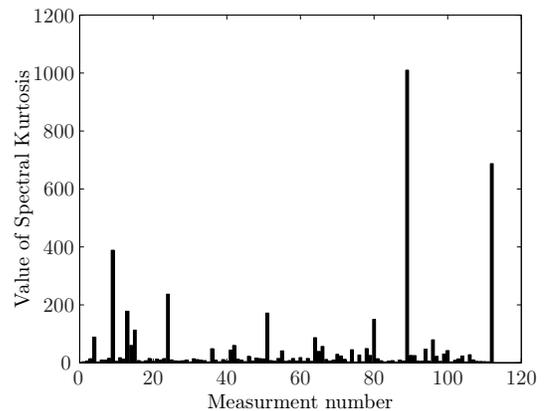


Figure 7. Values of the spectral kurtosis for 112 measurements

Although we have decided that 16 measurements with the lowest value of SK represent fault-free runs, it can be noticed that there are two segments of measurements with smaller interquartile intervals within these 16 measurements, in particular the 3rd, 4th and the 5th measurement, as well as the 10th and the 11th. These two intervals indicate the two different load levels under which the machine was operating during the conducted experiments. However these changes did not influence

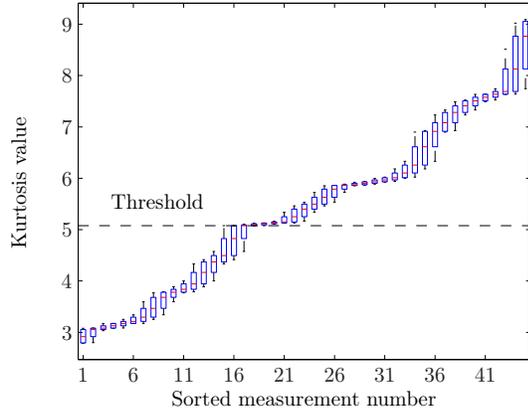


Figure 8. Box plot for the values of SK for determining the threshold (only the lower values of SK are depicted)

the overall fault detection process.

4.1 Results

The values of SK for a batch of 112 measurements with same speed is shown in Figure 7. We can notice that some measurements show high values of SK compared to the rest of the batch. As already explained, the iterative procedure starts with the measurement that has the highest SK.

The kurtogram diagram for the experimental run with the highest kurtosis is shown in Figure 9. In this particular case, the frequency band with the highest SK can be obtained by filtering the original signal with a band-pass filter with central frequency $f_c = 18.7$ kHz and filter bandwidth $B_w = 1.4$ kHz.

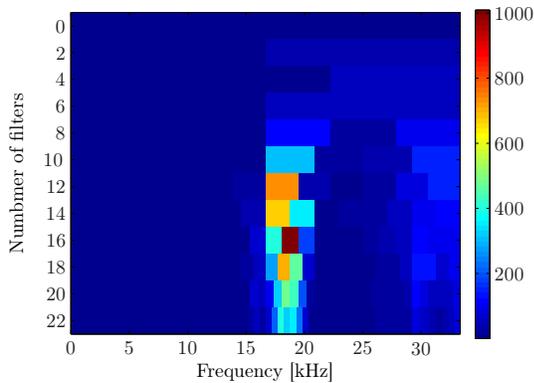


Figure 9. Kurtogram for the measurement from the top of the sorted list

By using this filter parameters we calculated the envelope spectrum of the examined measurement, shown in Figure 10. The envelope spectrum is dominated by a single spectral component at f_{BPFI} .

Since the spectrum is dominated by a single spectral component the fault diagnosis procedure is fairly simple. However,

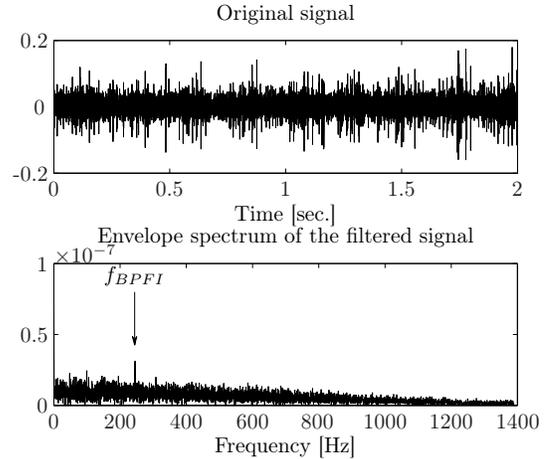


Figure 10. Envelope spectrum of the filtered vibration signal from the output shaft sensor

the final task for the data challenge was the selection which of the 6 bearings was damaged. All three shafts of the gear box rotated with different rotating speeds f_{rot} . Since all 6 bearings were of the same type, each pair of bearings on a particular shaft had its own set of bearing fault frequencies defined by Eq. (1). So, as a first step, we have located the shaft on which the faulty bearing was mounted. Afterwards, the selection which of the two possible bearings was damaged was done by observing the values of the signals acquired at both ends of the gearbox. The hypothesis is that the closer the sensor is to the fault itself the larger are the amplitudes of the corresponding bearing fault frequency. For the examined case of bearing inner race fault, the envelope spectrum of the filtered signal from the sensor mounted on the input shaft is shown in Figure 11. It is visible that the spectrum contains the spectral component at f_{BPFI} , however the amplitude is 50 times smaller than the corresponding one from the sensor mounted on the output shaft, shown in Figure 10. So for this particular experimental run we can conclude that the fault was on the inner race of the bearing on the output side of the gearbox.

The task of fault detection becomes more complicated as the severity of the fault decreases. One such case is bearing with damaged ball. This fault is characterized with increased amplitude of a spectral component at f_{BSF} (cf. Eq. 1). The kurtogram diagram for this fault is shown in Figure 12. Unlike the diagram for bearing inner race fault, the maximum value of the SK for this case is 4.5, compared to the SK value of 1000 observed in the previous case. Another important observation is that the impulses generated by this fault are most visible around $f_c = 7.5$ kHz. Hereupon, we can conclude that different faults may excite different eigenfrequencies of the observed system. So, for each run we have to calculate the maximal value of SK, thus obtaining the proper frequency band parameters valid for that particular run. Calculated enve-

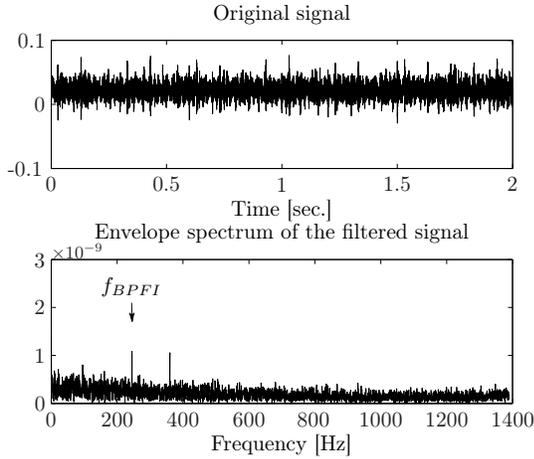


Figure 11. Envelope spectrum of the filtered vibration signal from the input shaft sensor

Envelope spectrum of the filtered signal around the proposed central frequency f_c is shown in Figure 13. Similarly to the first case, the fault detection for the currently observed run is quite straightforward. The most dominant spectral component is centered at f_{BSF} , which unambiguously points towards roller element damage.

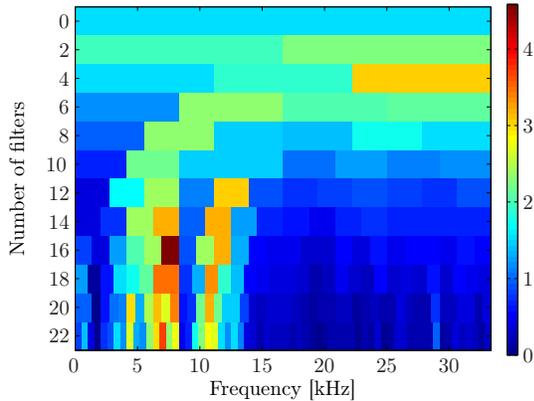


Figure 12. Kurtogram for the measurement from the top of the sorted list

4.2 Procedure simplification for cases with heavy damage

Spectral kurtosis is quite efficient for cases when the transient components in the measured signals have small energy compared to other vibration sources in the vicinity, like meshing gears or any other environmental influence. However, for the cases where the amplitude of impulses produced by the observed fault was sufficiently high, the fault detection procedure can be done by calculating the envelope spectrum of the whole unfiltered signal. Such a spectrum for the case of f_{BPFI} fault is shown in Figure 14. Comparing this spectrum

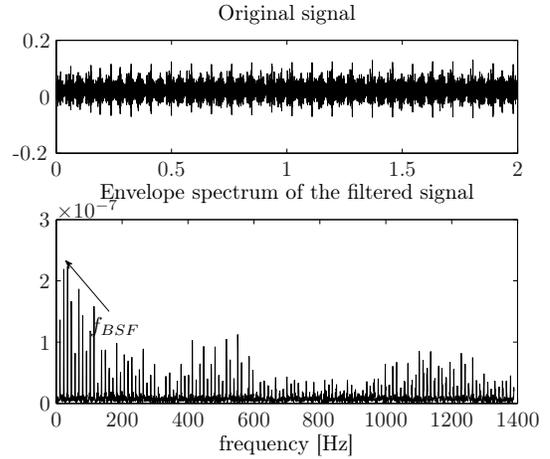


Figure 13. Envelope spectrum of the filtered vibration signal from the input shaft sensor

with the one shown in Figure 10, we can conclude that both spectra are dominated by the same frequency component.

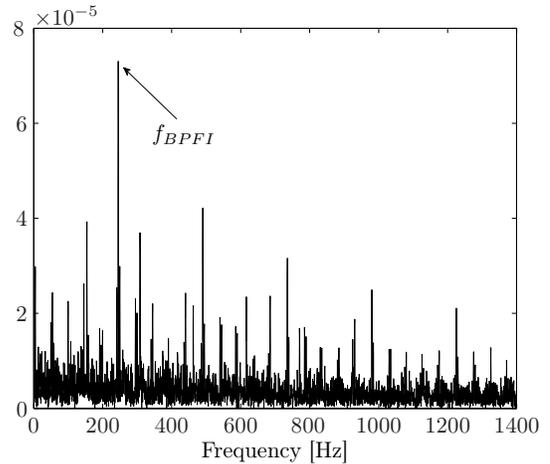


Figure 14. Envelope spectrum of the unfiltered signal from the output shaft sensor

This leads to a conclusion that for cases where the severity of the fault has passed the incipient stage the use of SK, although effective, is unnecessary. Comparably good results could be obtained by using a simple envelope analysis. Taking this finding into consideration the fault detection algorithm can be simplified, by removing the calculation of the SK for each measurement. With the removal of SK, the selection whether a particular experimental run belongs to the group of fault-free runs or not, can be done based on the amplitudes of spectral components of bearing principle frequencies.

However for the cases of mild faults or for cases of smaller load, the analysis of the envelope spectrum of unfiltered signal turns to be inefficient. This can be seen on the second observed case where the maximal SK value was 4. The enve-

lope spectrum of the unfiltered signal is shown in Figure 15. From the figure we can notice that the signal is dominated by two spectral components located at meshing gears frequencies, unlike the spectrum of the filtered signal shown in Figure 13 which is dominated by the spectral component at f_{BSF} .

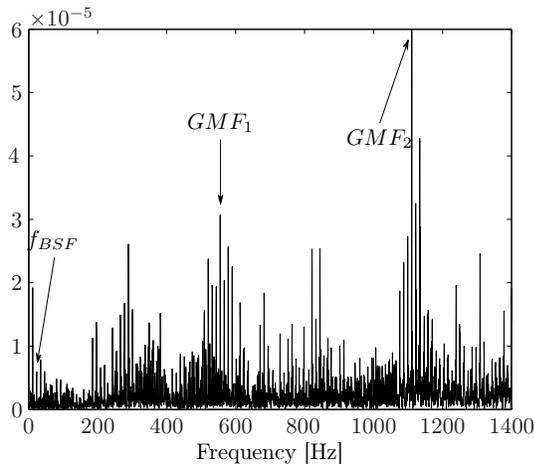


Figure 15. Envelope spectrum of the filtered signal from the input shaft sensor

Therefore we can conclude that for experimental runs conducted with severely damaged bearings the calculation of SK is unnecessary. However for cases of minor damages or cases where the experimental runs were conducted under smaller loads, the calculation of SK and filtering the vibration signals prior to the analysis of the envelope spectrum is an essential step for obtaining the satisfactory results.

5. CONCLUSION

The problem of fault detection in a case when there is no reference fault-free data, like in the case of the PHM Data Challenge 2009, was resolved using a two step approach: combining spectral kurtosis method and analysis of envelope spectrum. In achieving our goal of bearing fault detection, the gained results of the calculation of SK were twofold. Firstly, the maximum value of SK was used as an indication whether the observed experimental run is likely to represent a run with a faulty element. Secondly, the method selects a frequency band where the impulses generated by the localized faults are most visible. This is very important step, since the amplitude of the localized bearing faults are usually significantly lower than the amplitudes of the vibrations produced by meshing gears, which effectively mask their presence. Thus a simple spectral analysis is ineffective in the process of their detection. So by filtering out all other vibration influences, only spectral components originating from the localized bearing faults remain in the envelope spectrum, which additionally simplifies the fault detection procedure.

Despite all listed results, the general approach based solely on the maximal value of SK has its drawbacks. In cases of multiple faults, there may occur several frequency bands with similarly high values for SK. In such cases all of them should be examined and not just the one with the highest value (Combet & Gelman, 2009).

Regarding the vibration measurements obtained from PHM'09 Data Challenge, the effectiveness of SK can not be fully expressed. This is mainly due to the fact that most of the presented faults had already passed the incipient stage. As it was shown, in such cases the standard envelope analysis of the complete unfiltered signal produces satisfactory results. However, the SK pre-processing approach is effective for the cases of smaller machine loads and mild and incipient faults.

ACKNOWLEDGMENTS

The research of the first author was supported by Ad futura Programme of the Slovene Human Resources and Scholarship Fund. We also like to acknowledge the support of the Slovenian Research Agency through Research Programme P2-0001. The authors would like to thank Nader Sawalhi from the School of Mechanical and Manufacturing Engineering, The University of New South Wales, for providing the scripts for Spectral Kurtosis.

REFERENCES

- Antoni, J. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 20, 308-331.
- Antoni, J. (2007). Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21, 108-124.
- Antoni, J., & Randall, R. (2006). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mechanical Systems and Signal Processing*, 20, 282-307.
- Bartelmus, W. (2001). Mathematical modelling and computer simulations as an aid to gearbox diagnostics. *Mechanical Systems and Signal Processing*, 15, 855-871.
- Benko, U., Petrovčič, J., Musizza, B., & Juričič, Đ. (2008). A System for Automated Final Quality Assessment in the Manufacturing of Vacuum Cleaner Motors. In *Proceedings of the 17th World Congress The International Federation of Automatic Control* (p. 7399-7404).
- Combet, F., & Gelman, L. (2009). Optimal filtering of gear signals for early damage detection based on the spectral kurtosis. *Mechanical Systems and Signal Processing*, 23, 652-668.
- Dwyer, R. (1983). Detection of non-Gaussian signals by frequency domain Kurtosis estimation. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, 8, 607-610.

- Endo, H., Randall, R., & Gosselin, C. (2009). Differential diagnosis of spall vs. cracks in the gear tooth fillet region: Experimental validation. *Mechanical Systems and Signal Processing*, 23, 636–651.
- Ho, D., & Randall, R. B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, 14(5), 763-788.
- PHM. (2009). *Prognostics and Health Management Society 2009 Data Challenge*. <http://www.phmsociety.org/competition/09>.
- Randall, R. B., Antoni, J., & Chobsaard, S. (2001). 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, 945 - 962.
- Rubini, R., & Meneghetti, U. (2001). Application of the envelope and wavelet transform and analyses for the diagnosis of incipient faults in ball bearings. *Mechanical Systems and Signal Processing*, 15(2), 287-302.
- Sawalhi, N., & Randall, R. (2008). 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, 1924-1951.
- Sawalhi, N., Randall, R., & Endo, H. (2007). 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, 2616-2633.
- Staszewski, W. J. (1998). Wavelet based compression and feature selection for vibration analysis. *Journal of Sound and Vibration*, 211, 735 - 760.
- Tandon, N., & Choudhury, A. (1999). A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribology International*, 32, 469-480.
- Wang, W. (2001). Early Detection of gear tooth cracking using the resonance demodulation technique. *Mechanical Systems and Signal Processing*, 15, 887-903.