Mécanique & Industries

Spectral Kurtosis against SVM for best frequency selection in bearing diagnostics

Alessandro Fasana[1], Stefano Marchesiello[1], Miriam Pirra[1], Luigi Garibaldi[1,a] and Alessandra Torri[2]

[1] Dipartimento di Meccanica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[2] Avio S.p.A. Strada del Drosso 145, 10135 Torino, Italy

Abstract – Rolling bearing is probably the most widely used component in rotating mechanical equipments and its condition monitoring and fault diagnosis to prevent the occurrence of breakdown is growing in interest since many years. Vibration signal based methods are the most popular and have been adopted in many kinds of condition monitoring systems. Starting in the early 60, an immense range of different methods has been proposed on this basis, to perform diagnosis, fault identification and classification of bearing faults. Among the others, one typical approach consists in deep analysis of the most informative frequency range output of the system under test; the identification of this band is not straightforward because the fundamental task consists in finding out the band which is the most informative in contents which, in turn, might not be corresponding to that one of the maximum response, as claimed by some authors. In this paper, Spectral Kurtosis and Support Vector Machine are analysed and compared and it is shown that they typically reach similar results, in spite of their totally different approach. A brief description of both methods is given and laboratory data are analysed from a lab rig which uses spare parts of a full size power transmission gearbox, designed by AVIO. By taking advantage of these comparisons, the analyses are conducted using classical indicators applied to the specific bands suggested by previous analysis such as the RMS and other statistical quantities. Multi dimensional graphs are reported to show the reliability of the obtained results.

Key words: Spectral Kurtosis (SK) / Support Vector Machine (SVM) / bearing diagnostics / frequency selection

1 Introduction

Since rotating machinery is widely used in various industrial, military, and commercial processes, fault diagnosis and failure prognosis have gained significant attention in the past few years. It is well known that bearings are the best location for measuring machinery vibrations since this is where the basic dynamic loads and forces of machines are applied and they are a critical component of machinery. This means that condition monitoring and fault diagnosis of bearings can represent properly the condition of machine and their failures often result in a critical damage, downtime, and costly repair. Therefore, a proper fault diagnosis and failure prognosis providing a condition based maintenance strategy to either machinery or components is very important to the safety of the system and results in substantial economic benefits.

Vibration based condition maintenance is widely accepted as an appropriate strategy in many industrial situations as the best way to minimise maintenance costs and avoid unscheduled production interruptions. So far, for instance, an important class of bearing fault detection and diagnosis techniques relies on bearing defect frequencies. Typically, the diagnosis approaches are driven in a specific range of frequencies, say the most informative one, where the system response shows higher levels.

In [1] Antoni et al. presented the Spectral Kurtosis (SK) as a powerful analysis tool of non-stationary signals, after demonstrating in [2] its high potential in detecting and characterising non-stationary signals. According to the authors, the SK provides a robust way of detecting incipient faults even in presence of strong masking noise and allows to design optimal filters for cancelling the mechanical signature of faults. In the same context, the kurtogram concept is presented as a useful tool to seek for the optimal band-pass filter for the envelope analysis.

a Corresponding author: luigi.garibaldi@polito.it

On the other side, Widodo et al. describe in [3] a survey of machine condition monitoring and fault diagnosis using support vector machine (SVM), in particular for the diagnosis of rolling element bearings. Among various authors using SVM to detect faults, Jack and Nandi [4] rely on both SVM and artificial neural network (ANN). They analyse vibration data from a lab test rig and simulate the bearing condition with four different faults. Statistical features based on moments and cumulants are defined and calculated and the optimal features are selected using a genetic algorithm (GA) while the classification is led with SVM using Radial Basis Function (RBF) kernel with constant kernel parameter. Samanta et al. [5] apply GA for feature selection too and to search proper RBF kernel parameters in a procedure for diagnosis of bearing condition using the two classifiers ANN and SVM. On the whole, the performance of SVM have been found to be substantially comparable or better than ANN with the entire feature set. Rojas and Nandi [6] propose in their paper a mechanism for selecting adequate training parameters in order to make the classification procedure fast and effective. In general, many authors present SVM as a way to select features and to detect damages in rolling elements; in spite of this, the SVM is not exploited as a method to choose the best informative frequency band, as in the case of this paper.

Here, SK and SVM are shortly described and adopted to choose the best informative frequency band. Whilst the Spectral Kurtosis has been fully analysed and adopted to treat experimented data [2] up to the diagnosis, in this case we just compare the kurtogram concept and the SVM approach to detect the best range of interest on the basis of experimental data from damaged and undamaged bearings.

2 Spectral Kurtosis method

The *Spectral Kurtosis* (SK) is a statistical tool useful to detect the presence of series of transients and their location in the frequency domain. Up to now, it has rarely been used as mean of detecting problems related to vibrating and rotating machines, but we want to show that it can be efficiently used for their surveillance and diagnosis, as proposed in [1]. The general idea is that the presence of faults in rotating elements (such as rolling bearings) generates series of impacts producing transient vibration signals. Unluckily, these signals are often seriously corrupted by strong levels of background noise, so that this problem can be formulated as the detection of transient signals, called $X(t)$, in strong additive noise, called $N(t)$. The measured signal is hence in the form $Y(t) = X(t) + N(t)$ and a model able to describe incipient faults in rolling bearings is the generalised shot noise process, as seen in [1].

A possible way to analyse the characterisation of incipient faults is then to consider statistical indicators sensitive to the peakiness of the signal. The *Kurtosis* is a feasible one, thanks to the high values achieved in presence of the fault signal $X(t)$ and the ideal zero value it

takes in case of background noise $N(t)$. In order to avoid the masking of the signal by the strong vibrations produced from several competing sources which span a large frequency range, the Kurtosis could be locally applied in different frequency bands. That is the case of *Spectral Kurtosis* presented by Antoni in [2].

Furthermore, in [1] the author proposes a SK estimator based on the short-time Fourier transform (STFT) using an analysis window $w(n)$ with duration N_w shorter than the mean between two consecutive impulses. So, the SK of a vibration signal $Y(t)$ with sampling frequency f_s and fault rate of occurrence f_d becomes approximately

$$K_Y(f) \approx \left(\frac{f_s}{f_d} \frac{\gamma_{4w}}{N_w} \left(\kappa_X + 3 \right) - 2 \right) \frac{1}{[1 + \rho(f)]^2}$$

where γ_{4w} is the time–bandwidth product of the square of the analysis window, κ_X is the intensity of the fluctuations in the impulse amplitudes and $\rho(f)$ is the noise-to-signal ratio. It is clear that it increases if the window length N_w of the STFT decreases and this fact suggests that a shorter N_w should be preferred.

It is now necessary to create some robust detection filter with an imposed band-pass structure which needs only two parameters to be identified, namely the central frequency f_c and the bandwidth Δf, or equally the frequency f and the window length. The name "kurtogram" is so coined to indicate the SK map formed as function of f and N_w and its maximum value gives the optimal central frequency and bandwidth of the band-pass filter. A further simplification is proposed in [7] and is named *fast kurtogram*. It involves the computation of the Kurtosis of coefficients at the output of filter-banks of quasi-analytic filters with central frequency f and bandwidth Δf. On the whole, the fast kurtogram gives a strict method to find automatically the best frequency band able to filter the signal. Moreover, it detects in which frequency band transients take place and it returns the complex envelope for each selected band. Once some transients have been identified by the kurtogram, it is possible to compute the PSD of the filtered signal in order to find out spectral lines related to defect frequencies.

3 Support vector machine method

Support vector machine (SVM) is a computational learning method developed by Vapnik in the 80s and based on the statistical learning theory, rather suitable in case of classification and regression of a large set of data [8]. The first one is a common need in machine learning: given some data points which belong to a certain class, the purpose is to understand which class a new data point would be in. An input data $\boldsymbol{x}$ in an n-dimensional space is given and it is made up of a number M of samples, each of these belonging to a class, namely a positive or a negative one. The SVM constructs a hyperplane that should separate the two classes and hence it tries to place a boundary between them. Moreover, this boundary has to be oriented such that its distance from the nearest data
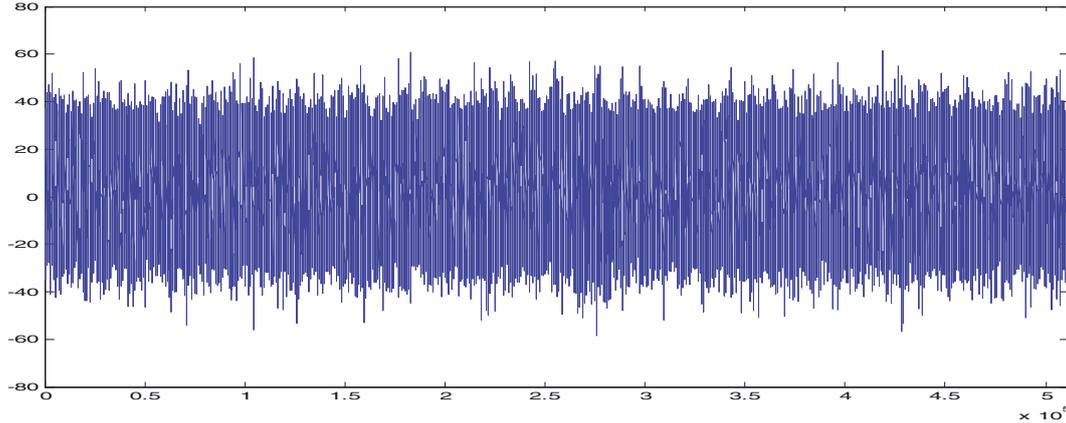
Fig. 1. Signal analysed: acceleration from a test rig with 1.4×10^3 N radial load on the bearing and 24 000 RPM angular speed.

points in each class is maximal. In this way, an optimal separating hyperplane is created, namely the maximum margin. In both classes, the points nearest to this margin are called *support vectors* and, once selected, contain all the necessary information to define the classifier.

To formalise this concept in case of linear data, the hyperplane that separates them is expressed as

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b = 0$$

where the M-dimensional vector $\boldsymbol{w}$ and the scalar b are the parameters that exactly define its position.

The nearest data points lie instead on the planes $f(\boldsymbol{x}) = \pm 1$ and so the margin between them is $2 \left\| w \right\|^{-1}$. The goal is to maximise this one and it can be done through the following optimisation problem:

minimize $\frac{1}{2} \left\| w \right\|^2 + C \sum\limits_{i=1}^{M} \xi_i$

subject to $\begin{cases} y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1 - \xi_i & i = 1, \dots, M \\ \xi_i \geq 0 & i = 1, \dots, M \end{cases}$

where ξ_i (slack variables) measure the distance between the margin and the points $\boldsymbol{x}_i$ lying on the wrong side of the margin itself, C is the error penalty and $y_i \in \{-1, +1\}$ are the negative and positive class labels respectively.

This problem can be converted into the equivalent, and easier to solve, Lagrangian dual problem [9]:

maximize $L(\alpha) = \sum\limits_{i=1}^{M} \alpha_i - \frac{1}{2} \sum\limits_{i,j=0}^{M} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$

subject to $\alpha_i \geq 0, \qquad i = 1, \dots, M, \qquad \sum\limits_{i=1}^{M} \alpha_i y_i = 0$

where α_i are the Lagrangian multipliers, necessary to define the decision function

$$f(\boldsymbol{x}) = \text{sign} \left(\sum\limits_{i=1}^{M} \alpha_i y_i (\boldsymbol{x}_i \boldsymbol{x}_j) + b \right).$$

In case of non-linear classification SVM can also be applied by means of a function $\Phi(\boldsymbol{x})$ that maps the data onto a high-dimensional feature space, where the linear classification is then possible. Furthermore, if a kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\Phi^{\mathrm{T}}(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j))$ is applied, it is not necessary to evaluate $\Phi(\boldsymbol{x})$ explicitly in the feature space. So

Fig. 2. Test rig (on the left) and damage on the rolling element after 50 h of testing (on the right).

that, in the decision function, the term $y_i(\boldsymbol{x}_i \boldsymbol{x}_j)$ could be replaced by $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which is easier to compute. There are various possible choices of the kernel function, such as linear, polynomial or Gaussian RBF. This fact is rather outstanding because it enables SVM to be used in case of very large feature spaces: the dimension of classified vectors does not influence directly the SVM performance.

If the classes the input data could be classified in are more than two, multi-class SVM is necessary. Two different approaches are taken into account: One-against-all (OAA) and One-against-one (OAO). In the first one the ith SVM is trained with all of examples in the jth class with positive labels and all the other examples with negative labels, while in the latter one each classifier is trained on data from two classes.

4 Application to bearing data

The methods presented in the previous sections are here applied to real signals searching for the most informative frequency range of the system output. A signal such as in Figure 1 is analysed, representing 10 seconds of an acceleration time history sampled at 51.2 kHz. The data are collected from a test rig assembled by the Dynamics & Identification Research Group (DIRG) at the Department of Mechanics, which is shown in Figure 2; for the described tests, the radial load on the damaged bearing is settled at 1.4×10^3 N and the rotational speed is 24 000 RPM. Different levels of fault on rolling element

Table 1. Frequency bands for level 3 decomposition.

Band n°	Low freq	Upper freq
1	0	3200
2	3200	6400
3	6400	9600
4	9600	12 800
5	12 800	16 000
6	16 000	19 200
7	19 200	22 400
8	22 400	25 600

are analysed, that is 450, 250 and 150 micron size fault. An example of damage of 450 micron as it appears after 50 h of testing is presented in Figure 2. Faults on the rolling element have been produced by using a Rockwell type indentation hedge, causing a penetration surface area of different sizes. The expected fault frequency

$$f_b = \frac{D}{d}\omega \left(1 - \frac{d^2}{D^2}\cos^2\varphi\right),$$

where D is the pitch diameter, d the ball diameter, ω the rotational speed, φ the thrust contact angle, null in this case, is computed and it is 1.7 kHz.

With reference to [7], the signal is passed through the detection filters and its kurtogram is then defined. In Figure 3 three kurtograms for different levels of fault on one rolling element are represented, with 450, 250 and 150 micron of their largest dimension, respectively. The x-axis indicates the frequencies, while the y-axis the levels into whom the signal is decomposed; the kurtogram is able to point out transient occurrence in the signal and also in which frequency band it is placed. The level refers to the bandwidth though the relation $Bw = f_{\rm s} \times 2^{-({\rm level}+1)}$. The scale in figures adopts brighter colours for higher kurtosis values. It can be noticed that faults of higher entity, such as for cases (a) and (b) in Figure 3, reach higher Kurtosis values and, as a matter of example, for the case of 450 micron fault the optimal central frequency is placed in the n. 5 band (Fig. 3a, level 3, bright band 12.8÷16.0 kHz). This means that filtering the signal with a band-pass filter characterized by these two parameters will enhance the impulse detection although hidden in the signal.

In the SVM case, using the same band division as for the previous analysis, more signals are taken into account contemporary, representing various entity of damage (450, 250 and 150 micron size fault on the rolling element). This is a case of multi-class SVM analysis and we actually adopt the one-against-one approach. Each of these signals is divided into a certain number N of vectors and their PSD is computed. The division in height frequency bands previously defined is then performed (as in Tab. 1) and their RMS are evaluated. These are taken with reference to the 3rd level computed by the SK method. Hence, a matrix made up of $(n \times N) \times 8$ elements is produced. To set up the parameters that characterise the SVM analysis, $N/2$ points are used to train it, while the others are used to test it. That means that we know the "labels" for
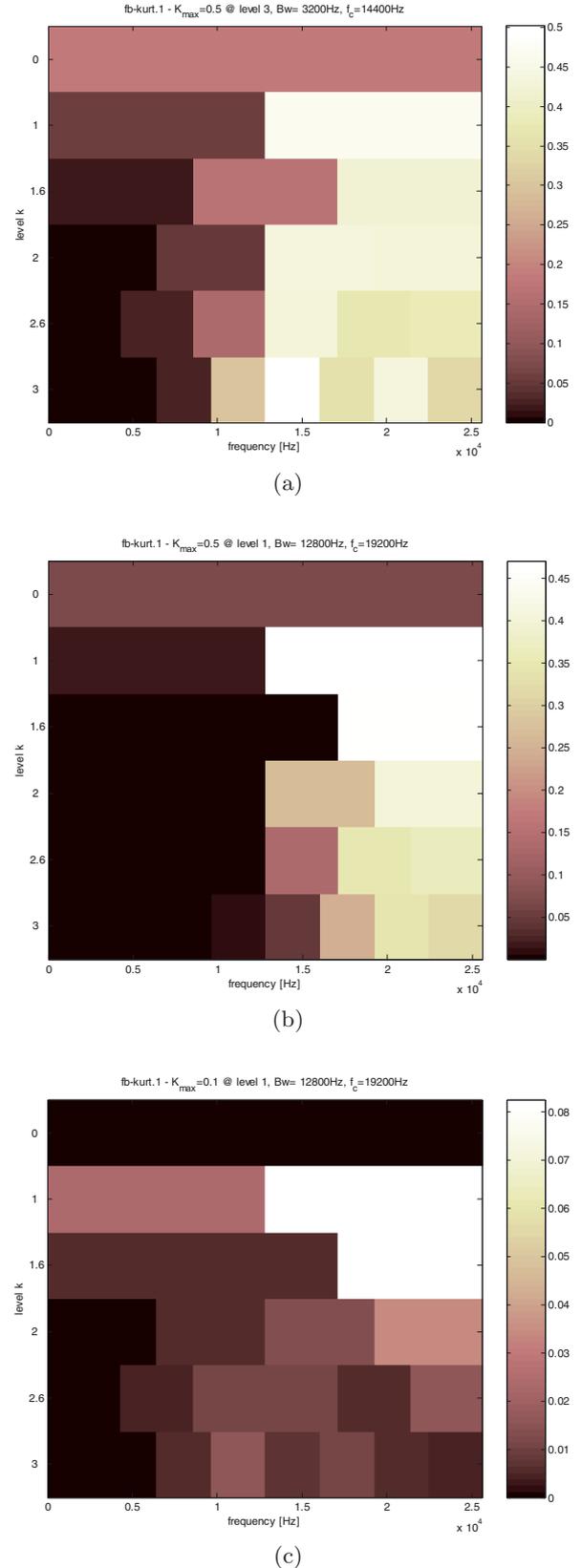

(a)

(b)

(c)

Fig. 3. Kurtogram for different fault levels on a single rolling element: (a) 450 micron (b) 250 micron (c) 150 micron. The brighter colours represent higher values of kurtosis. The titles indicate in which frequency band transients occur.

Table 2. Optimal frequency band intervals and central frequencies for different rotational speeds and same radial loads on three damaged bearings computed for SK and SVM (less than 10% error) methods. For the sake of interpretation: bold numbers indicate that SVM is able to detect fault in all the range indicated (0–3.2 kHz, 16.0–19.2 kHz and 22.4–25.6 kHz), whilst SK seems to be better fitted for the range 21.1–25.6 kHz only.

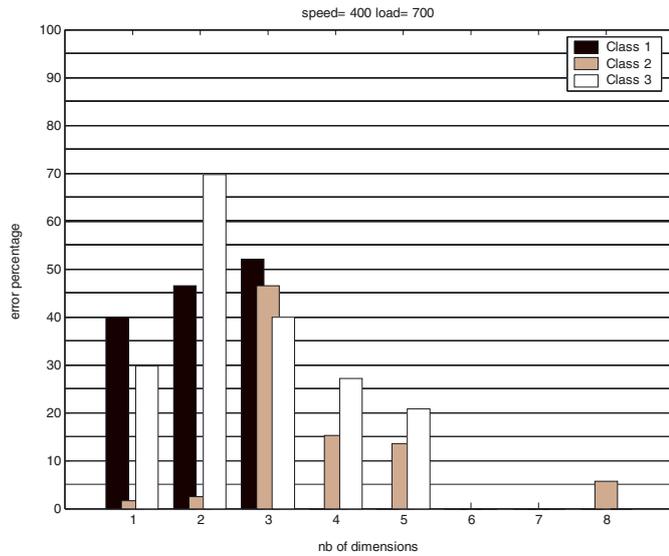| 1.4 kN	12 000 RPM		18 000 RPM		24 000 RPM	
fault [μm]	SK [kHz]	SVM [kHz]	SK [kHz]	SVM [kHz]	SK [kHz]	SVM [kHz]
450	22.4–25.6	9.6–25.6	21.3–25.6	6.4–25.6	12.8–16.0	9.6–25.6
250	**21.3–25.6**	**0–3.2** **16.0–19.2** **22.4–25.6**	12.8–16.0	0–9.6 12.8–19.2 22.4–25.6	12.8–25.6	0–6.4 16.0–25.6
150	0–12.8	22.4–25.6	8.5–12.8	0–3.2 12.8–19.2 22.4–25.6	12.8–25.6	16.0–25.6

Fig. 4. Error percentage in fault detection. Classes 1 are the data for the 450 micron fault on the rolling element, Class 2 for 250 micron and Class 3 for 150 micron. Rotational speed is always 24 000 RPM. Ranges 6 and 7 are clearly the most reliable classes for all the three damage sizes.

each class and we want to see those the SVM would assign. Thus, to find the most informative frequency range the number of errors in each band is considered and those where it is smaller is supposed to be the best one (Fig 4). It is evident from the same figure that, for higher frequencies, the number of errors decreases, so that these ranges are to be taken into account.

5 Comparison between various data

The comparison between the two different methods to extract the most interesting bands is given hereafter. The scope is also to check how the band of highest interest can change depending upon the speed of rotation, the defect size and the load. Data are collected from a number of accelerometers which have been positioned on the test

rig, where bearings with different levels of fault on rolling element are tested (450, 250 and 150 micron size fault). In the first case the radial load is fixed at 1.4×10^3 N and three rotational speeds of 12 000, 18 000 and 24 000 RPM are considered, while in the latter the speed is fixed at 18 000 RPM and three radial loads of 1, 1.4, 1.8×10^3 N are taken into account.

Table 2 shows the optimal central frequencies and respective frequency intervals obtained with the SK and the SVM methods for the first case. It can be noticed that when the fault is bigger, the Kurtosis reaches higher values in shorter intervals. Analysing the optimal central frequency for each fault it can be seen that it grows in the case of the littlest damage, while it decreases when the fault is bigger as a function of speed.

Frequency bands detected with SVM concern those cases with a number of errors lower than 10%. The bands that have less errors seem to be the higher, 6th, 7th and 8th especially, which refer to a frequency band from 16 to 25.6 kHz. It suggests that this is the most informative frequency range that has to be taken into account to analyse this type of data. It can be noticed that when the fault has the smallest size (150 micron) a lower number of bands has proper detection ability while, with the largest damage, the bands that can be considered are almost the same regardless of the speed. For both approaches, SK and SVM, the case of 250 micron damage has a non regular behaviour with respect to the increasing speed.

In Table 3 the optimal central frequencies and respective frequency intervals obtained with the SK and the SVM method for the case of different loads are presented. It is worth to notice that a good agreement between the results from the two methods is confirmed, whilst the dependence of the best frequency choice with respect to the applied load is not evidenced for both the methods.

Once the most informative bands are obtained they are treated by using different philosophies to detect the faults and their relevance: the SK data are passed through an algorithm based on Power Spectrum peaks distribution, the SVM data are passed through other statistical based routines.

It must be observed that the high values of rotational speed are due to the fact that the rig has been conceived

Table 3. Optimal frequency band intervals and central frequencies for a fixed rotational speed (18 000 RPM) and different radial loads: three damaged bearings computed for SK and SVM (less than 10% error) methods.

18 000 RPM	1 kN		1.4 kN		1.8 kN	
fault [μm]	SK [kHz]	SVM [kHz]	SK [kHz]	SVM [kHz]	SK [kHz]	SVM [kHz]
450	21.3–25.6	6.4–25.6	21.3–25.6	6.4–25.6	21.3–25.6	0–3.2 9.6–25.6
250	12.8–16.0	0–3.2 12.8–19.2 22.4–25.6	12.8-16.0	0–9.6 12.8–19.2 22.4–25.6	12.8–16.0	0–3.2 6.4–9.6 12.8–19.2
150	4.3–8.5	12.8–19.2	8.5–12.8	0–3.2 12.8–19.2 22.4–25.6	0–4.3	0–6.4 12.8–19.2

for a specific aeronautical application. Moreover, the short time between impacts might be the cause of misleading interpretations, not being the optimal condition to apply the SK algorithm. Possibly, the relatively low resolution obtained by SK is due to this fact, i.e. the decay of the transient of each impact is partially hidden by the subsequent impact.

6 Conclusions

The paper compares two techniques developed to detect the most informative frequency range from damaged bearing acceleration data. The aim is not to extend the results to assess the best frequency range in a general case but, conversely, to assess the methodology for seeking the best range, for any type of speed, load and damage extension. It is worth to mention that, while the SK approach allows a blind recognition of the most informative analysis bands, this fact does not hold for the case of SVM, which needs a training approach to separate the classes of features.

Under the author experience, however, it is true that complex systems, such as the aeronautical gearbox simulated by the lab rig proposed, the detected signal might come from so many different sources that a pure blind recognition represents a very hard work. Furthermore, just a simple temperature drift might dramatically change the vibration signature, so that important frequency shifting would appear in the output. For this reason, it is the authors feeling that a preventive mapping of the different working spaces is necessary, even in presence of orthogonalisation tool, such as the PCA, frequently adopted to separate the contribute of exogenous inputs.

Under this point of view, since the training phase seems unavoidable, the use of SVM does not overcharge the procedure, being a part of the mentioned mapping procedure.

Regarding the application to real data, only the case of rolling element fault has been treated along this article due to space restriction. Generally speaking, the SVM approach looks a bit more flexible in terms of frequency range choice, but it is true that 10% of errors is tolerated in damage identification, which is too much in real life application. If the percentage is restricted up to 100% of correct answers, than the choice is restricted to certain bands; one example is given for 24 000 RPM, where the $19.2-22.4$ kHz range (6th and 7th bands) show the 100% proper results for all the damage extensions tested.

References

[1] J. Antoni, R.B. Randall, The Spectral Kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20 (2006) 308–331

[2] J. Antoni, The spectral kurtosis: a useful tool for characterising nonstationary signals, Mech. Syst. Signal Process. 20 (2006) 282–307

[3] A. Widodo, B. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (2006) 2560–2574

[4] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural network, augmented by genetic algorithms, Mech. Syst. Signal Process. 16 (2002) 373–390

[5] B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, Artificial neural network and support vector machine with genetic algorithm for bearing fault detection, Eng. Appl. Artif. Intell. 16 (2003) 657–665

[6] A. Rojas, K. Nandi, Detection and classification of rolling-element bearing faults using support vector machines, IEEE Workshop on Machine Learning for Signal Processing 12 (2005) 153–158

[7] J. Antoni, Fast Computation of the Kurtogram for the Detection of Transient Faults, Mech. Syst. Signal Process, in press

[8] V.N. Vapnik, Estimation of dependences based on empirical data, Springer, 1982

[9] C.M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer-Verlag New York, Inc., Secaucus, NJ, 2006