A new methodology for fault detection in rolling element bearings using singular spectrum analysis

H Al-Bugharbee and I Trendafilova

Submitted 26.01.16
Accepted 13.01.17

In this study, a new methodology for fault detection in rolling element bearings is proposed, which is based on singular spectrum analysis (SSA). The main idea of the methodology is to build a baseline space from the feature vectors corresponding to the healthy bearing condition. This baseline space is made from the directions of the first three principal components, which are obtained from the decomposition stage of the singular spectrum analysis. Then, the lagged version of any new signal corresponding to a measured (possibly damaged) condition is projected onto this baseline space in order to assess its similarity to the baseline condition. The Euclidean norms of these projections are used to form three-dimensional feature vectors. The category of a new signal vector is determined on the basis of the Mahalanobis distance (MD) of its feature vector to the baseline ones. The methodology is validated using datasets acquired from two different test-rigs. From the results obtained for the correct classification rate, it is shown that this methodology performs very well. The suggested methodology also has simple steps and is easy to apply.

Keywords: fault detection, singular spectrum analysis, rolling element bearings.

1. Introduction

Fault detection in rolling element bearings (REBs) is commonly carried out using vibration analysis as it is a simple, repeatable and non-destructive strategy.

During recent decades, there have been many studies that have suggested different approaches for the purpose of machinery fault detection. There are several studies that focus on reviewing and comparing the performance of some of the approaches for the health monitoring of rolling element bearings[1-5]. Due to the complexity of the bearing vibration signal resulting from factors such as clearance and friction, some of these approaches are complex in structure but perform well in fault detection. Among these approaches are the wavelet transform, which is a time-frequency analysis method[6,7], and envelope analysis[6,8], to mention a couple. For example, although enveloping is a purely signal analysis-based method, its application for bearing fault detection requires some preliminary information. Enveloping is usually used in combination with other signal analysis techniques to determine the appropriate centre and width of the frequency band of interest in order to develop a bearing fault detection method. To a certain extent, this depends on the *a priori* knowledge of some system characteristics, including bearing assembly resonant frequencies, to identify the centre of the frequency band of interest, and the bearing geometry, to calculate the fundamental fault frequencies of the bearing. Enveloping is usually combined with techniques such as spectral kurtosis[9,10], for example, for a careful windowing of the initial signal. It also requires combining with other signal analysis processes, such as band-pass and low-pass filtering. Thus, research aimed towards a simple and accurate method is still of interest to many researchers.

Most fault detection methods based on bearing vibration signal analysis usually include the extraction of certain features representative of the bearing state, which are used as a means of comparison against the healthy state. There are many features suggested using a number of methods that are based on analysing the vibration signal in the different domains (*ie* time, frequency and time-frequency)[11-13]. Several challenges can accompany the extraction and selection of the features, such as the complexity of the feature extraction method and the sensitivity of the features to the change of a bearing condition.

The methodology proposed in this study is based on singular spectrum analysis (SSA) and is simple in structure and easy to apply. SSA is a time series analysis method that is primarily used to uncover the trend and the periodic components from the original time series. It has two main stages: decomposition and reconstruction. In the decomposition stage, which is the only stage used in this study, signals (in this case a bearing vibration signal) are decomposed into a number of new components called principal components (PCs). These principal components represent the projection of the original signal onto the new space directions, called eigenvectors. The directions defined by the PCs are used to build the reference space. The rationale behind this is that they contain a certain amount of the information contained in the original signal in terms of its variance. Thus, the reference space is expected to preserve a certain amount of the information for the healthy/reference state. From this viewpoint, any new signals with an unknown category are projected onto the reference space in order to assess their similarity to the reference state and hence to the baseline condition.

One of the main contributions of this study is that the eigenvectors corresponding to a healthy bearing condition are used to build a reference state onto which all of the signals are projected. Thus, SSA is only applied to the signals corresponding

Hussein Al-Bugharbee is with the Department of Mechanical & Aerospace Engineering, University of Strathclyde, Glasgow, UK, and the University of Wasit, Iraq, while on leave from the Department of Mechanical and Aerospace Engineering. Email: hussein.al-bugharbee@strath.ac.uk

Irina Trendafilova is with the Department of Mechanical & Aerospace Engineering, University of Strathclyde, Glasgow, UK.

to the baseline condition in order to build the reference state. Once the reference space is built, any signal can be projected onto this space and the norms of its projections can be used to assess its similarity to the baseline condition. For the selection of PCs, there are several guiding criteria mentioned in the literature[14]. One of these criteria is that the percentage of variance portion described in these PCs should be equal to or greater than 75% of the variance of the original signal. Another criteria is that the PCs can be selected at the value where very little or no change in the variance is observed when adding new PCs. In this particular case, the norms of the projections on the first three PCs are used as features. In general, however, the reference state can be built using any number of principal components/directions and using more components may result in a more precise representation of the baseline state. In this case, using the first three principal directions was enough for classification/fault detection purposes.

In general, SSA is a signal analysis method used for climatic and forecasting data analysis[15,16] and biomedical signal analysis[17,18]. It is also used as an anomaly detection method in tool wear health monitoring[19,20] and for damage assessment in wind turbine blades[21], but it is still unpopular for fault detection in rolling element bearings. SSA has the capability to distinguish between different data categories when such are present in the data analysed[22-24]. It does this by increasing the distance between categories and decreasing the distance between signals from the same category[25,26]. The first study of the use of SSA in the fault detection of REBs is presented in[27]. It considers both healthy and faulty inner race conditions. SSA is applied to both healthy and faulty bearing signals and two main time components are obtained, namely trend and residual components. From the trend component of each bearing signal a number of features, such as peak value and standard deviation, are used to form the feature vectors and then presented as input to a neural network classifier. In[14], SSA is used for multi-decomposition analysis. The number of singular values that cumulatively preserve a predetermined variance percentage is used as the indicator for fault presence. In[28], two different variable sets are proposed and used to form two feature vectors, namely the singular values and the energy of the first principal components. Both of these feature vectors are used as input to the back-propagation neural network (BPNN) classifier.

The methodology suggested here is simpler in comparison to the studies mentioned above, since it proposes applying the SSA solely to the signals corresponding to the healthy bearing condition in order to build a reference (*ie* baseline) space. Thus, the only information needed is the information for the healthy bearing state, which can be obtained using a sample of signals measured in the healthy bearing state. These signals are subjected to SSA in order to find their principal directions defined by the first several principal components. So, only the decomposition stage is used and, as mentioned, this is applied solely to the baseline signals measured on the healthy state. It should also be mentioned that the methodology suggested, although simple in its nature and easy to apply, gives very precise results for the case studies presented here.

The norms of the projections on the principal directions are used as features. The features obtained for a new signal with unknown classification are compared to the features obtained for the healthy state. The comparison is carried out on the basis of the Mahalanobis distance (MD), which provides the possibility of comparing each feature vector (FV) to the set of feature vectors corresponding to the healthy condition. Accordingly, the Mahalanobis distance of each FV to the set of baseline FVs is used as a measure of similarity to the baseline (healthy condition). The smaller the MD, the greater the similarity to the healthy condition. The Mahalanobis distance is used in a number of studies for measuring the distance between multidimensional variables. Eventually, a threshold for this distance is established using the baseline FVs and the classification of the signals from different bearing states is carried out using this threshold. If the MD is smaller than the established threshold, the signal and the bearing state are considered healthy and *vice versa*.

As the main goal of the current study is targeting only the detection of faults and, due to the limitation of the manuscript length, the capabilities of the method regarding location identification and fault severity estimation will be investigated in the next installment of this work.

The methodology is validated using two different datasets and it showed excellent performance in terms of correct classification rate. The performance of the methodology is assessed based on the percentages of correct and incorrect classifications.

It is also important to stress that the present method has very good capabilities when applied for fault detection in complex systems. This is because the complex signal will be unfolded and decomposed into a number of PCs by projecting the signal onto the baseline space. Thus, the effect of the presence of a defect will be clearly seen when the projections, more specifically the norms of the projection vectors, deviate from the baseline values. Another important reason for the belief that the present methods will be successful when applied to real complex machines is that the reference space will be created based on the real data of the machine itself (not from a simulated model) when operating in a healthy condition. A threshold will then be set up from this healthy category data and any anomaly will be detected and classified as an observation different to the reference space.

The rest of the paper is organised into the following sections: Section 2 presents the building of a reference space and Section 3 describes the proposed fault detection methodology based on SSA. In Section 4, two case studies for the validation of the suggested method are introduced. In Section 5, the results and discussion are presented. Finally, concluding remarks are given in Section 6.

2. Building a reference space

This section introduces the building of the reference space based on SSA. SSA is a data analysis procedure used for decomposing a time series into a number of interpretable components, such as trend, periodic and structureless noise components. It has two main stages, namely decomposition and reconstruction. As mentioned in Section 1, only the decomposition stage is considered in this study, so the reconstruction stage will not be mentioned. Further information on the reconstruction stage can be found in[29].

First, a trajectory matrix $\mathbf{X}$ is formed from a lagged version of the measured vibration signal $\mathbf{x}$ of length n, $\mathbf{x} = [x(1), x(2), \ldots, x(n)]$ onto a window of length (L):

$$\mathbf{X} = \begin{bmatrix} x(1) & x(2) & x(3) & \ldots & x(K) \\ x(2) & x(3) & x(4) & \ldots & x(K+1) \\ x(3) & x(4) & x(5) & \ldots & x(K+2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(L) & x(L+1) & x(L+2) & \ldots & x(n) \end{bmatrix} \quad \ldots (1)$$

where $K = n - L + 1$.

Next, the covariance matrix of $\mathbf{X}$ is calculated:

$$\mathbf{C}_X = \frac{\mathbf{X}\mathbf{X}'}{L} \quad \ldots (2)$$

The $\mathbf{C}_X$ has a dimension $(K \times K)$ and it defines the covariance between signal realisations. The $\mathbf{C}_X$ is then subjected to eigendecomposition by singular value decomposition (SVD) in order to obtain L eigenvectors $(\mathbf{U}_i, i = 1, 2, \ldots, L)$ and L eigenvalues $(\lambda_i, i = 1, 2, \ldots, L)$ by solving the following expression:

$$\mathbf{C}_X \mathbf{U}_i = \lambda_i \mathbf{U}_i \quad \ldots (3)$$

Each λ_i represents a partial variance proportion of the original signal in the direction of the corresponding $\mathbf{U}_i$. The square roots of each λ_i give the so-called singular value SV_i (*ie* $SV_i = \sqrt{\lambda_i}$). The L eigenvalues (λ_i) are usually arranged in decreasing order, so that the first one is responsible for the largest part of the variance (*ie* in the direction of $(\mathbf{U}_1)$), while the last eigenvalue is responsible for the lowest part of the variance (*ie* in the direction of $(\mathbf{U}_L)$). The plot that displays the trend of λ_i *versus* the index (*ie* number) is called the scree diagram.

All or some of the first L eigenvectors corresponding to the reference state (*ie* in this study the healthy state of the bearing) can be used to define the reference space. Thus, a reference space with dimension L is made of the first L eigenvectors $(\mathbf{U}_i, i = 1, 2, \ldots, L)$: $\mathrm{R}^L = [\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_L]$.

In this study, the first three eigenvectors are used to build the reference space. This is partly for purposes of visualisation but in principle any number of PCs can be used for the baseline/reference state. A larger number of eigenvectors will result in more variance/information being preserved in the reference set of vectors and might presumably lead to improved fault recognition results.

3. Fault detection methodology

In this section, the fault detection process is explained. It has two main parts: feature vector extraction and fault detection.

3.1 *Feature extraction*

This section illustrates how features are made by projecting the lagged version of a signal onto the reference space. First, the signals from the healthy bearing condition are divided into a training sample and a testing sample, while all of the signals from the faulty bearing conditions are used as testing samples. For each signal from the training sample, the trajectory matrix $\mathbf{X}$ is projected onto each eigenvector $\mathbf{U}_i$ and provides the corresponding principal components $\mathbf{PC}_i$:

$$\mathbf{PC}_i = \mathbf{X}'.\mathbf{U}_i / \sqrt{\lambda_i} \quad \ldots (4)$$

The prime denotes the transpose of the $\mathbf{X}$. The trajectory matrix can be decomposed into L elementary matrices $\mathbf{X}_i$:

$$\mathbf{X}_i = \sqrt{\lambda_i}\mathbf{U}_i\mathbf{PC}_i' \quad \ldots (5)$$

so that:

$$\mathbf{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_L \quad \ldots (6)$$

The contribution of the norm of any $\mathbf{X}_i$ to the norm of the original $\mathbf{X}$ follows the same trend of the λ_i[30]. That is, the first $\mathbf{X}_i$ has the highest contribution, while the last one has provided the lowest contribution.

Then, the Euclidean norm of each of the first three PCs of each training signal is calculated according to Equation (7):

$$f_{ij} = \sum_{m=1}^{K} \left(\mathbf{PC}_{ij}(m)\right)^2, \quad j = 1,2,3, \quad i = 1,2,3\ldots k \quad \ldots (7)$$

where f_{ij} is a norm of the jth **PC** obtained from the ith signal, K is the length of a **PC** and equals $(n - L + 1)$, k is the number of signals corresponding to a baseline condition and $\mathbf{PC}_{ij}(m)$ is the mth element of the jth **PC** obtained from the ith signal.

In the current study, the first three PCs are selected, as they have the majority of the original signal variance and to facilitate the graphical representation of the feature space. More PCs can be selected and they might improve the method accuracy for fault detection. However, the latest PCs are not recommended as they usually correspond to the structureless noise[29,31].

The three Euclidean norms corresponding to the first three PCs of a signal i are used to form three-dimensional feature vectors of the baseline condition:

$$\mathbf{fv}_i = [f_{i1} f_{i2} f_{i3}] \quad \ldots (8)$$

Figure 1 presents a 3D visual representation of $\mathbf{fv}_i$. As the Euclidean norm is simply the summation of squared values, all of the features have positive values and when they are projected onto the feature space they are located in the first quadrant.

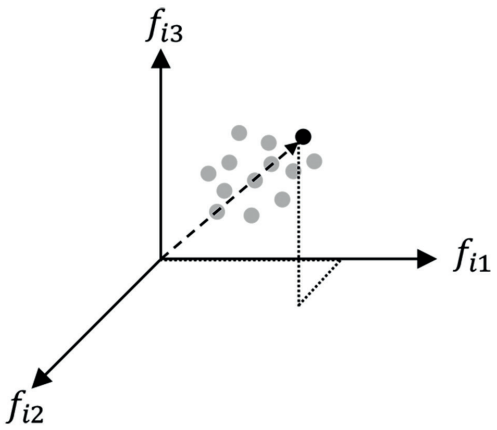

Figure 1. A 3D visualisation of $\mathbf{fv}_i$

A baseline feature matrix ($\mathbf{F}_{baseline}$) is made by arranging the baseline FVs in rows:

$$\mathbf{F}_{baseline} = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ . & . & . \\ . & . & . \\ f_{k1} & f_{k1} & f_{k1} \end{bmatrix} \quad \text{.....(9)}$$

Following the steps described above in Equations (4), (7) and (8), the feature vectors for any other/new signals (*ie* the testing sample) can be obtained.

3.2 *Fault detection*

In this section, the process of fault detection is detailed. The recognition of the healthy/baseline and the non-healthy/non-baseline (faulty) conditions is made on the basis of a threshold. The Mahalanobis distance is used as a measure of similarity between a signal and the baseline condition. Thus, all the FVs where the distance is equal to or less than the threshold are classified as healthy and all of those where the distance is greater than the threshold are considered as faulty.

3.2.1 *Setting a threshold for the baseline condition*

The Mahalanobis distance of each of the baselines, from $\mathbf{fv}_i$ to $\mathbf{F}_{baseline}$, is calculated as shown below:

$$D_i = \left(\mathbf{fv}_i - \mathbf{E}_{baseline}\right).\mathbf{S}^{-1}.\left(\mathbf{fv}_i - \mathbf{E}_{baseline}\right)' \quad \text{.....(10)}$$

where D_i is the Mahalanobis distance of $\mathbf{fv}_i$ to the matrix $\mathbf{F}_{baseline}$ and $\mathbf{S}^{-1}$ is the inverse of the covariance matrix of $\mathbf{F}_{baseline}$. The prime (′) denotes the transpose of the vector $(\mathbf{fv}_i - \mathbf{E}_{baseline})$. $\mathbf{E}_{baseline}$ is the mean of $\mathbf{F}_{baseline}$ rows and is given as follows:

$$\mathbf{E}_{baseline} = \frac{\sum_{i=1}^{k}\left(f_{i1}\ \ f_{i2}\ \ f_{i3}\right)}{k} \quad \text{.....(11)}$$

A threshold for the values of D_i (corresponding to the baseline training FVs) can be made based on the probability distribution of a good fit. In this study, it was found that the lognormal distribution (p) describes well the distribution of the distances D_i. The lognormal distribution has a probability density function that is given by the following equation:

$$p\left(D_i\right)\Big|_{\mu,\sigma} = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\ln(D_i)-\mu\right)^2/\left(2\sigma^2\right)} \quad \text{.....(12)}$$

where $p(D_i)$ is the lognormal probability distribution function value at D_i and μ and σ are the mean and standard deviation of $p(D_i)$, $i = 1, 2, \ldots, k$

A threshold ($Thr_{baseline}$) can be estimated such that there is a probability of 97.5% for D_i from a baseline condition, which is equal to or less than the threshold.

Figure 2 illustrates the threshold (*ie* $Thr_{baseline}$) and a possible distribution of the baseline FVs around it. The *x*-axis represents the number of FVs while the *y*-axis represents the MD described in Equation (10). The horizontal dashed line represents the value of the $Thr_{baseline}$.

Depending on the value of the $Thr_{baseline}$, a certain number of D_i (corresponding to reference FVs) will be below this threshold and all of the rest will exceed it.

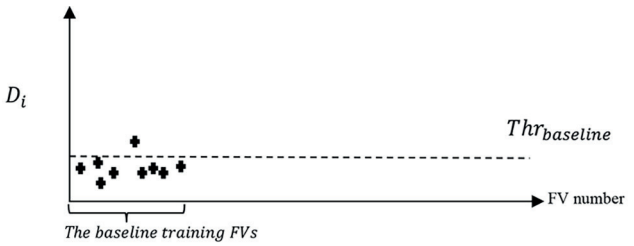

Figure 2. Illustration of the $Thr_{baseline}$ constructed from the baseline training FVs

3.2.2 *Comparison of a new testing D_i to the threshold*

As mentioned at the end of Section 3.1, the FVs of the testing sample can be obtained in the same way as the baseline FVs. For any new signal from the testing sample, its FV is first obtained and the Mahalanobis distance of the FV to the $\mathbf{F}_{baseline}$ is then calculated, as described in Equation (10). Finally, it is compared to the chosen threshold $Thr_{baseline}$. If the Mahalanobis distance of the new FV is less than or equal to the $Thr_{baseline}$, the FV is assigned to the baseline condition and *vice versa*.

$$\left.\begin{array}{l} D_i > Thr_{baseline} \ \ an\ FV_i\ is\ assigned\ to\ faulty\ category \\ D_i \le Thr_{baseline} \ \ an\ FV_i\ is\ assigned\ to\ baseline\ category \end{array}\right\} \quad \text{...(13)}$$

The performance of the methodology in terms of correct assignment of the FVs to their actual category (baseline and non-baseline condition) is evaluated by using the so-called confusion matrix. This matrix is a square ($K_k \times K_k$) matrix, where K_k is two (*ie* healthy (H) and faulty (F)). The columns represent the predicted categories, while the rows represent the actual categories. Thus, the main diagonal represents the correctly classified FVs, while all the other elements represent the miscategorised FVs. Table 1 illustrates the structure of a confusion matrix.

Table 1. The structure of a confusion matrix

Actual class/predicted class	H	F
H	C_{HH}%	C_{HF}%
F	C_{FH}%	C_{FF}%

C_{HH}% and C_{FF}% represent the percentage of FVs assigned to their correct classes, while C_{HF}% and C_{FH}% represent the percentage of miscategorised FVs.

A flowchart showing the steps of the methodology is presented in Figure 3. The flowchart has three main blocks given in boxes with dashed lines:

1. Building a reference space: this includes subjecting a signal from the training sample to SSA and using the first three **U**s as reference space (Section 2).
2. Extraction of FVs: projecting the lagged versions of the signals from the training and testing to obtain the PCs and then calculating their norms.
3. Fault detection: this includes measuring the MD, setting a threshold and comparing the MD of the FVs corresponding to the testing sample with the threshold.

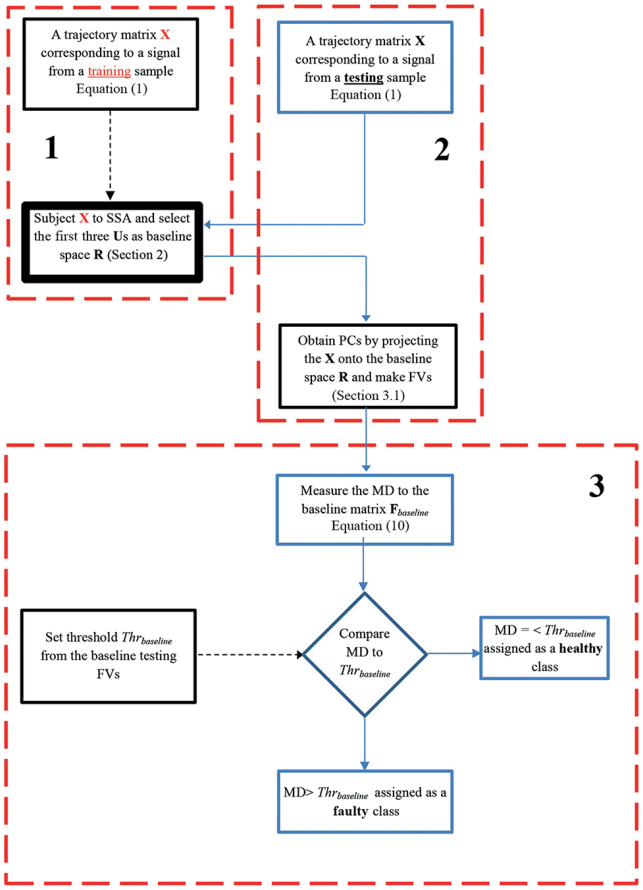

Figure 3. Flowchart showing the steps of the methodology

3.3 *Effect of using the norms of the first three PCs on distinguishing between (baseline and non-baseline) signal classes: visual interpretation*

As mentioned in Section 1, only the first three PCs, which have most of the original signal variance, are used in the fault detection process. In this section, the way in which the use of the norms of the first three PCs, rather than the use of all the elements of these PCs, allows the baseline and non-baseline signal categories to be more easily distinguished will be visually illustrated.

An example from the datasets used in this study is shown to illustrate the improvement of the use of norms of the first three PCs as features to distinguish between two different conditions (*ie* healthy (H) and faulty (F) bearings).

Figure 4 shows the 3D visualisation and its 2D projections using all the elements of the first three PCs. In this Figure, the first three PCs corresponding to 30 signals from the baseline condition and another 30 from the non-baseline condition are plotted. It can be seen that the two conditions are mixed and there is no clear visual separation between them.

Thus, it is suggested in this study that each signal is to be represented by a single feature, which is the Euclidean norm of its first three PCs instead of using all the elements of these PCs. This can help reduce the dimension of FV used to represent a signal and improve the separation between the two signal categories (baseline and non-baseline).

In Figure 5, the 3D visualisation and the three 2D projections of using the norms of the first three PCs are shown. It is clearly seen that the two conditions are well separated.

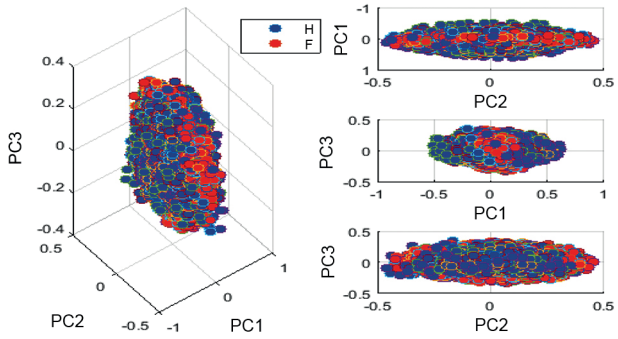

Figure 4. 2D and 3D visualisation for the case using all the elements corresponding to the first three PCs

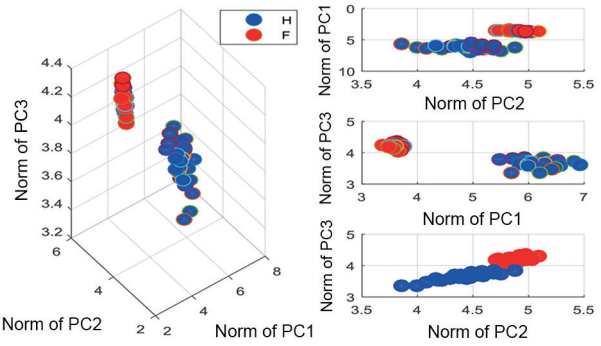

Figure 5. 2D and 3D visualisation for the case using norms corresponding to the first three PCs

4. Method verification: case studies

The proposed methodology is tested and validated using various bearing vibration datasets acquired from different bearing test-rigs.

4.1 *Case study 1*

The experimental datasets used in this analysis are those obtained from the Bearing Data Center of the Case Western Reserve University (CWRU)[32]. The test-rig shown in Figure 6 consists of a dynamometer connected to a three-phase induction motor. The vibration acceleration data were collected from the drive-end SKF 6025 deep groove ball bearing. The datasets were obtained for different bearing conditions, different motor speeds and different bearing fault sizes at a sampling rate of 12 kHz.

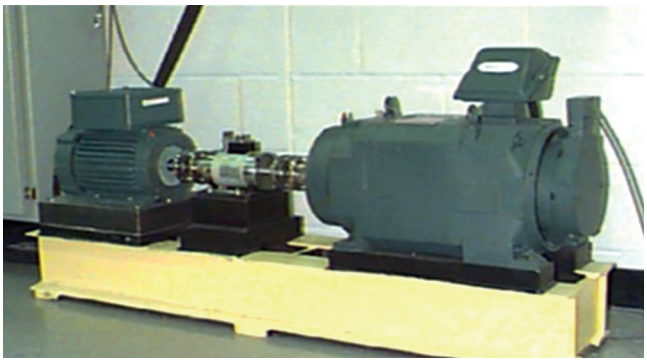

Figure 6. The bearing test-rig of CWRU[21]

4.2 *Case study 2*

The datasets of case study 2 are collected from a bearing test-rig at the Department of Mechanical & Aerospace Engineering, University of Strathclyde, as shown in Figure 7. The test-rig consists of a 1 HP shunt DC motor, bearing assembly and a mechanical loading system. The vibration acceleration data are collected from the housing where the test bearing (in this case a SKF deep groove 6308 ball bearing) is mounted inside. A pinion-toothed belt mechanism is used to transmit the torque from the motor to the bearing assembly. A fault diameter of 0.05 inches was introduced by an electrical discharge machine on the inner race, a ball and the outer race using different bearings. The bearing vibration acceleration datasets were collected at a sampling rate of 12 kHz for different speeds (*ie* 250 r/min, 750 r/min and 1250 r/min) using a magnetic-based accelerometer mounted on the housing of the bearing.

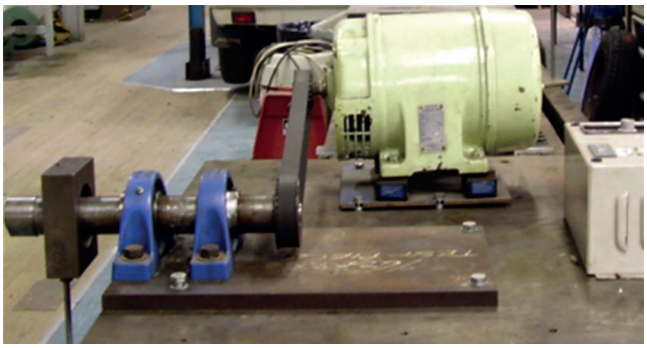

Figure 7. Bearing test-rig of case study 2

5. Results and discussion

5.1 *Case study 1*

The data used in the analysis are shown in Table 2. In Table 2, (H) refers to a healthy bearing condition, IRF refers to a bearing with an inner race fault, BF refers to a bearing with a ball fault and ORF refers to a bearing with an outer race fault. The signals are segmented into 240 equal non-overlapping signals (*ie* 4 signal classes × 60 signals per class, each one containing 2048 data points). Thirty of the healthy signals are used for the training sample and the remaining 210 (*ie* 30 healthy and 180 faulty) signals are used for the testing sample.

Figure 8 presents a signal corresponding to a healthy bearing in the time domain at 1730 r/min. The *x*-axis represents the number of data points while the *y*-axis is the acceleration of the signal vibration.

Figure 9 shows the scree diagram of a healthy bearing condition at a speed of 1730 r/min. The diagram displays the normalised eigenvalues *versus* the number of PCs for a window, L = 10. The *y*-axis represents the normalised eigenvalues (*ie* *normalised* $\lambda_i = \lambda_i / \sum_{i=1}^{L} \lambda_i$), which explain the percentage of variance portion explained by each of the PCs. It can be clearly seen that the first PCs have the highest variance portions when compared to the last ones. In this study, only the first three PCs are considered in forming the FVs. The variance proportion accounted for in the first three PCs is at least 75% of the variance of the original signal, which means that the selection of the threshold percentage for PCs meets with one of the guiding criteria mentioned in[14].

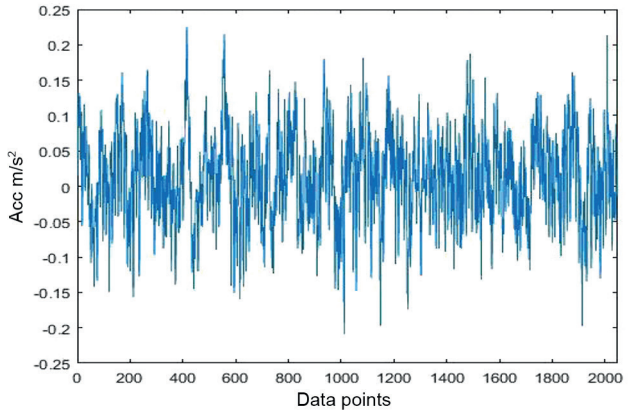

Figure 8. A signal of 2048 data points corresponding to a healthy bearing at 1730 r/min

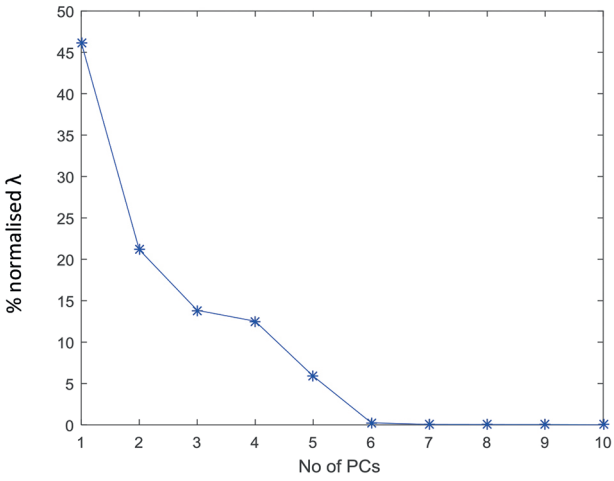

Figure 9. Scree diagram of a healthy signal at 1730 r/min

Figure 10 shows a 3D visualisation of the 30 feature vectors corresponding to the baseline training sample, which are used to form the baseline features. As was described in Section 3.1, all of the FVs are obtained from projecting the lagged version of the signal onto the reference eigenvectors. Since the reference eigenvectors are corresponding to a healthy bearing condition, it is expected that FVs corresponding to non-baseline bearing condition will differ from the baseline FVs.

Table 2. Bearing vibration datasets obtained from case study 1 used in the bearing fault detection

Case no	Motor speed (r/min)	Signal category	No of signals
1	1730	Healthy, (IRF, BF and ORF)$_{\text{small fault 0.007"}}$	240 (4 classes × 60 signals each)
2	1750	Healthy, (IRF, BF and ORF)$_{\text{small fault 0.007"}}$	240 (4 classes × 60 signals each)
3	1772	Healthy, (IRF, BF and ORF)$_{\text{small fault 0.007"}}$	240 (4 classes × 60 signals each)
4	1797	Healthy, (IRF, BF and ORF)$_{\text{small fault 0.007"}}$	240 (4 classes × 60 signals each)

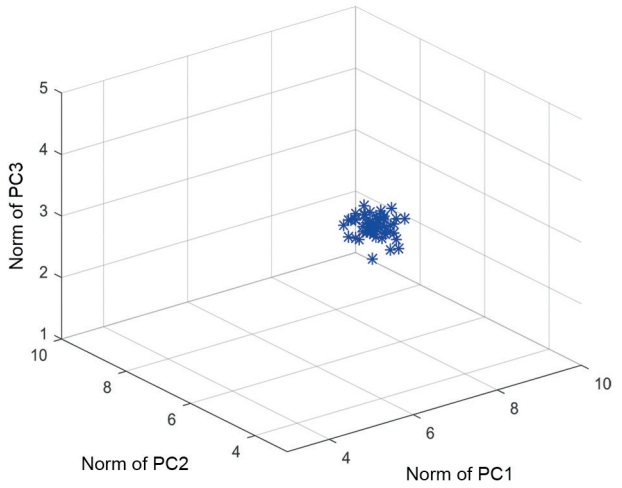

Figure 10. 3D visualisation of FVs obtained from the training sample of the baseline condition at 1730 r/min

Figures 11(a)-(c) show the visualisation of the FVs corresponding to the testing sample (the remaining 30 signals of a healthy bearing condition and the 180 signals of a faulty bearing condition) obtained at 1730 r/min. As mentioned in Section 3, the trajectory matrix of each testing sample signal is projected onto the baseline eigenvectors. Then, the Euclidean norms of the first three PCs are used to form the testing sample FVs. Figure 11(a) shows the 3D visualisation of the FVs corresponding to the baseline bearing condition in blue and the various faulty bearings in red, green and black. It can be seen that FVs corresponding to the faulty classes are separated from the baseline conditions and they are also separated from one another. Figures 11(b)-(d) show the distribution of the FVs from different 2D views. Figure 11(a) shows that selecting three PCs to form the FVs improves the separation not only between the baseline and non-baseline categories but also among the non-baseline categories.

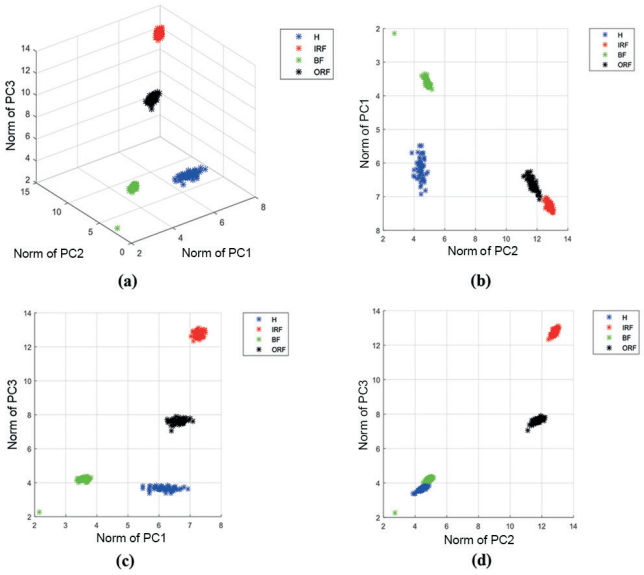

Figure 11. The 2D and 3D visualisation of the FVs corresponding to different bearing conditions obtained at 1730 r/min

Figure 12 presents the Mahalanobis distances of the testing and the training FVs to the reference/baseline FVs, as given in Equation (10). As can be seen from the Figure, there are five regions/parts of MD separated by vertical dashed lines. The leftmost part represents the MD of the training baseline FVs, which represents the MD of each training FV measured to $\mathbf{F}_{baseline}$. This is the part where the threshold is established, see Section 3.2.1. The threshold is represented by the horizontal dashed line in the Figure. The part second from left represents the MD of the testing baseline FVs measured to the reference/baseline FVs. All of the other parts represent the MD of the testing FVs corresponding to different fault locations. It is clear that all the MDs corresponding to training baseline FVs are below the threshold and most of the MDs corresponding to testing healthy samples are also still below (*ie* only two out of 30 are located slightly above the threshold level). However, all the other distances corresponding to the faulty condition are quite far from the threshold and the healthy (for example the *y*-axis in the Figure is shown in a log scale).

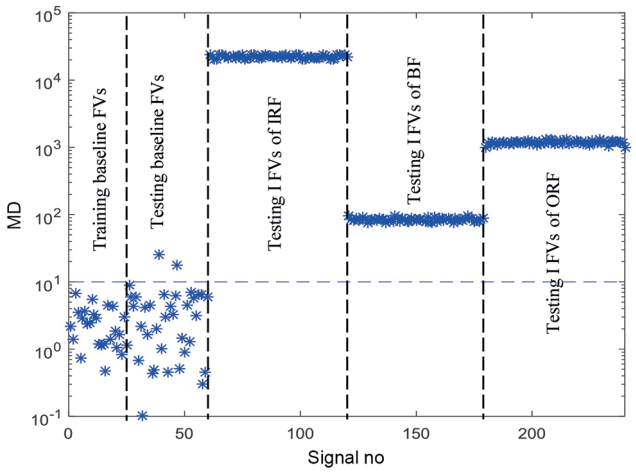

Figure 12. The MDs as indicators for the fault detection in the bearing at 1730 r/min

The confusion matrix corresponding to the classification of testing FVs obtained at 1730 r/min is shown in Table 3. It is clearly shown that the positive false alarm rate (*ie* FVs corresponding to the healthy class but assigned as faulty) is 6.7%, as only two out of 30 are misclassified, while there are no negative false alarms (*ie* none of the FVs corresponding to the faulty bearing category are assigned to the healthy category).

Table 3. Confusion matrix of testing FVs at 1730 r/min

Actual class/predicted classes	H	F
H	93.3%	6.7%
F	0%	100%

For all the datasets considered in this case study (see Table 2), the results regarding the correct classification percentage for both the training and testing samples are given in Table 4. The first column in the Table shows the type of datasets defined in Table 2. The second column shows the percentage of the original variance preserved in the first three PCs corresponding to the baseline

bearing condition. The third, fourth and fifth columns show the correct classification of the training and testing FVs as a percentage. It is clear that all of the 180 testing sample FVs corresponding to the faulty bearing conditions are correctly assigned to the fault class, which means no negative false alarms. The lowest correct classification rate for the healthy training FVs (*ie* $H_{training}$) is 96.7%, where only one out of 30 healthy training FVs were assigned to a faulty class. For the healthy testing FVs, the lowest correct classification rate is seen at the speed of 1772 r/min, where only three out of the 30 FVs were misclassified under the faulty class.

Table 4. Performance of the methodology, PCs (1, 2, 3). Number of FVs: $H_{testing}$ = 30, $H_{training}$ = 30, $F_{testing}$ = 180

Speed (r/min)	% variance contained in the first three PCs	% correctly classified		
		% $H_{training}$	% $H_{testing}$	% $F_{testing}$
1730	81.18	100	93.3	100
1750	79.86	96.7	93.3	100
1772	87.19	100	90	100
1797	94.7	96.7	96.7	100

5.2 *Case study 2*

For case study 2, the datasets used for analysis are described in Table 5. The data corresponding to different motor speeds are considered in the analysis. For each speed, data are obtained for H, IRF, BF and ORF. The number of signals for each case is 320 (4 classes × 80 signals), each of 2048 length. The signals are divided into 40 signals corresponding to healthy bearings as a training sample and the other 280 (40 healthy and 240 faulty) signals as a testing sample.

Table 5. Bearing vibration datasets obtained from case study 2 used in the bearing fault detection

Case no	Motor speed (r/min)	Signal category	No of signals
1	250	Healthy, (IRF, BF and ORF) $_{\text{small fault 0.005"}}$	320 (4 classes × 80 signals each)
2	750	Healthy, (IRF, BF and ORF) $_{\text{small fault 0.005"}}$	320 (4 classes × 80 signals each)
3	1250	Healthy, (IRF, BF and ORF) $_{\text{small fault 0.005"}}$	320 (4 classes × 80 signals each)

Figure 13 shows a vibration acceleration signal corresponding to the baseline training sample collected at 750 r/min.

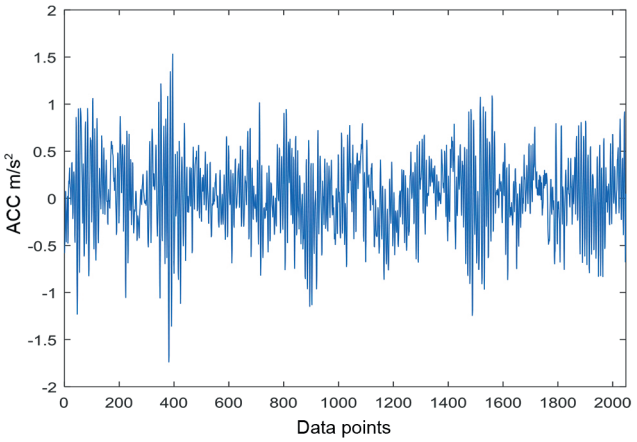

Figure 13. A signal of 2048 data points corresponding to a healthy bearing at 750 r/min

The signal shown in Figure 13 is subject to SSA of window L = 10, to represent the signal in higher dimensional space. Ten PCs are obtained by subjecting the signal to SSA. Figure 14 shows the contribution of the variance of these PCs to the original signal variance. It can clearly be seen that the majority of the original signal variance is included in the first three components, while low contribution to the original signal variance is accounted for in the other PCs.

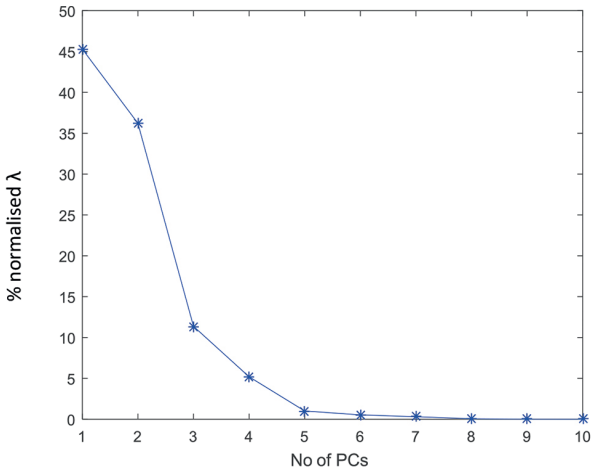

Figure 14. Scree diagram of a healthy signal at 750 r/min (case study 2)

As described in Section 3, the first three eigenvectors corresponding to the baseline training sample are used as the reference vectors onto which all of the lagged versions of testing sample signals are projected. The Euclidean norms of the projection resultants are used to form the FVs. Figures 15(a)-(d) show the 2D and 3D visualisations of the FVs constructed from the training and the testing samples. In these Figures, it can be seen that the separability exists not only between the baseline and non-baseline FVs but also among the faulty categories.

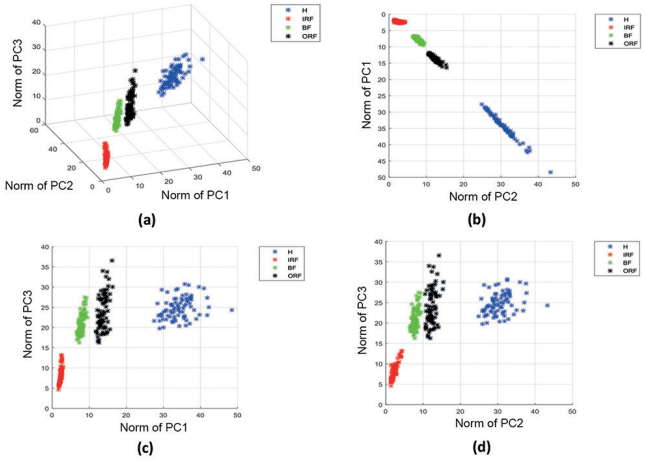

Figure 15. The 2D and 3D visualisations of the FVs corresponding to different bearing conditions obtained at 750 r/min (case study 2)

Figure 16 shows the MDs of both the training and testing FVs measured to the baseline testing FVs category. As the MD is suggested in this study as a fault indicator, it is compared to a predetermined threshold constructed from the MDs corresponding to the training baseline. The separation between the baseline and non-baseline FVs is very clear and it results in accurate detection of the fault.

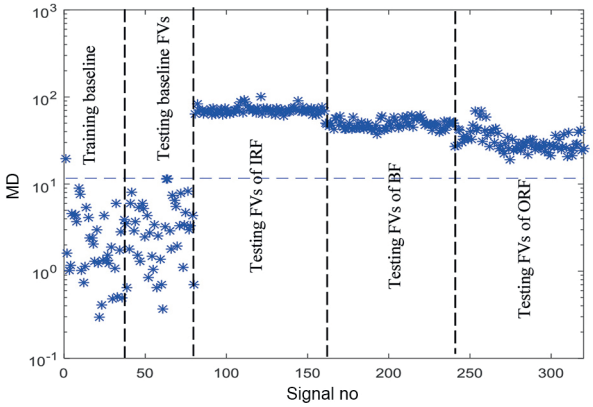

Figure 16. The MDs as indicators for fault detection in the bearing at 750 r/min

The performance of the methodology using the datasets of case study 2 is shown in Table 6. The first column shows the type of datasets as defined in Table 5. The second column shows the percentage of the variance proportion of the original variance preserved in the first three PCs. The third, fourth and fifth columns show the correct classification of the training and testing FVs. It can be seen that only 12 out of the 240 faulty testing FVs were misclassified at a speed of 250 r/min, which gives a correct classification rate of 95%. All of the other FVs corresponding to the testing sample at speeds of 750 r/min and 1250 r/min are correctly classified into the fault class. It can be seen that 95% more of the baseline training FVs are classified to within their actual class, while at least 97.5% of the baseline training FVs are correctly classified (*ie* only one or two out of 40 of the baseline training FVs are misclassified).

Table 6. Performance of the methodology, PCs (1, 2, 3). Number of FVs: $H_{testing}$ = 40, $H_{training}$ = 40, $F_{testing}$ = 240

Speed (r/min)	% variance contained in the first 3 PCs	% correctly classification		
		% $H_{training}$	% $H_{testing}$	% $F_{testing}$
250	94.04	95	100	95
750	92.89	97.5	100	100
1250	75.525	100	97.5	100

6. Conclusions and discussion

The present study suggests a rather simple and easy to apply but accurate method for fault detection in rolling element bearings based on singular spectrum analysis. The suggested method is simple in application as it solely uses the signals from the healthy bearing state to build the reference space. As such, it does not require any previous measurements corresponding to faulty/anomalous conditions. A number of signals measured in the baseline/healthy condition can be used to build the reference space. SSA is used only for the purposes of building this reference space. The new signals, which are to be classified, are not subjected to SSA; their lagged versions are simply projected onto the first principal directions of the reference space. As such, the transformations applied to the measured signals are minimal and very simple. The classification rule, which is based on a threshold of the Mahalanobis distance, is also a robust and simple one. As a result of this, the method holds considerable potential for automatisation and practical implementation.

There are some conclusions that can be drawn from the method itself. The principal components are used to build the reference space and the classification/fault detection is carried out on the basis of projecting the data onto these PCs. As such, it can be argued that the PCs obviously contain information regarding the state of the bearing. In this particular application only three principal directions are used, but potentially more principal directions can be used, which might improve the classification for some applications as it can be argued that more principal directions will contain more information from the measured signals and hence the bearing condition. Thus, further interpretation of the PCs used to define the baseline state is seen as a direction for further research and knowledge enhancement regarding the relationship of the signal information/characteristics to the principal directions obtained using SSA.

It should also be noted that the method is rather general and can be applied to any measured signals regardless of their stationarity. The 3D visualisation shows that not only baseline and non-baseline signal categories can be distinguished but in some cases the method can be used to separate different faulty categories. This remains a subject of further research and as such the interpretation of the PCs/the principal directions will be very helpful.

Based on the results obtained, it has been shown for the case studies considered that SSA is capable of extracting essential information regarding the presence of faults in signals where different fault locations are detected at different motor speeds. As the primary goal of the current study is limited to the detection of faults, further investigations are targeted in terms of expanding the method for fault qualification and quantification purposes and also for online monitoring of the bearing condition.

Acknowledgements

The authors acknowledge the support of the Iraqi Ministry of Higher Education and Scientific Research for making this research possible.

References

1. N Tandon and A Choudhury, 'A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings', Tribology International, Vol 32, No 8, pp 469-480, 1999.
2. C Nataraj and K Kappaganthu, 'Vibration-based diagnostics of rolling element bearings: state of the art and challenges', Proceedings of the 13th World Congress in Mechanism and Machine Science, Guanajuato, Mexico, p 1109, 19-23 June 2011.
3. C Sujatha and C Chandran, 'On various specialised vibration techniques for the detection of bearing faults', Proceedings of the 2002 IMAC – XX: A Conference & Exposition on Structural

Dynamics, LA, California, USA, pp 1306-1312, 4-7 February 2002.
4. N Tandon, G Yadava and K Ramakrishna, 'A comparison of some condition monitoring techniques for the detection of defects in induction motor ball bearings', Mechanical Systems and Signal Processing, Vol 21, No 1, pp 244-256, 2007.
5. J Urbanek, T Barszcz and T Uhl, 'Comparison of advanced signal processing methods for roller bearing fault detection', Metrology and Measurement Systems, Vol 19, No 4, pp 715-726, 2012.
6. P W Tse, Y H Peng and R Yam, 'Wavelet analysis and envelope detection for rolling element bearing fault diagnosis – their effectiveness and flexibilities', Journal of Vibration and Acoustics, Vol 123, No 3, pp 303-310, 2001.
7. N Nikolaou and I Antoniadis, 'Rolling element bearing fault diagnosis using wavelet packets', NDT&E International, Vol 35, No 3, pp 197-205, 2002.
8. R B Randall and J Antoni, 'Rolling element bearing diagnostics – a tutorial', Mechanical Systems and Signal Processing, Vol 25, No 2, pp 485-520, 2011.
9. E Bechhoefer, M Kingsley and P Menon, 'Bearing envelope analysis window selection using spectral kurtosis techniques', Proceedings of the 2011 IEEE Conference on Prognostics and Health Management (PHM), pp 1-6, 20-23 June 2011.
10. N Sawalhi and R B Randall, 'The application of spectral kurtosis to bearing diagnostics', Proceedings of ACOUSTICS 2004, pp 393-398, 3-5 November 2004.
11. H-H Lee, N-T Nguyen and J-M Kwon, 'Bearing fault diagnosis using fuzzy inference optimised by neural network and genetic algorithm', Journal of Electrical Engineering & Technology (JEET), Vol 2, No 3, pp 353-357, 2007.
12. B Sreejith, A Verma and A Srividya, 'Fault diagnosis of rolling element bearings using time-domain features and neural networks', Proceedings of 2008 IEEE Region 10 and the Third International Conference on Industrial and Information Systems (ICIIS 2008), pp 1-6, 8-10 December 2008.
13. M Behsad, A R Bastami and D Mba, 'Rolling bearing fault detection by short-time statistical features', Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, pp 229-237, 2011.
14. B Kilundu, X Chiementin and P Dehombreux, 'Singular spectrum analysis for bearing defect detection', Journal of Vibration and Acoustics, Vol 133, No 5, p 051007, 2011.
15. R Vautard and M Ghil, 'Singular spectrum analysis in non-linear dynamics with applications to paleoclimatic time series', Physica D: Nonlinear Phenomena, Vol 35, No 3, pp 395-424, 1989.
16. R Vautard, P Yiou and M Ghil, 'Singular spectrum analysis: a toolkit for short, noisy, chaotic signals', Physica D: Nonlinear Phenomena, Vol 58, Nos 1-4, pp 95-126, 1992.
17. S Sanei and H Hassani, Singular Spectrum Analysis of Biomedical Signals, CRC Press, Florida, USA, 2015.
18. W C Pereira, S L Bridal, A Coron and P Laugier, 'Singular spectrum analysis applied to backscattered ultrasound signals from *in vitro* human cancellous bone specimens', IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol 51, No 3, pp 302-312, 2004.
19. D Salgado and F Alonso, 'Tool wear detection in turning operations using singular spectrum analysis', Journal of Materials Processing Technology, Vol 171, No 3, pp 451-458, 2006.
20. B Kilundu, P Dehombreux and X Chiementin, 'Tool wear monitoring by machine learning techniques and singular spectrum analysis', Mechanical Systems and Signal Processing, Vol 25, No 1, pp 400-415, 2011.
21. D García, D Tcherniak and I Trendafilova, 'Damage assessment for wind turbine blades based on a multivariate statistical approach', Journal of Physics: Conference Series, Vol 628, p 012086, 2015.
22. J Zabalza, J Ren, Z Wang, S Marshall and J Wang, 'Singular spectrum analysis for effective feature extraction in hyperspectral imaging', IEEE Geoscience and Remote Sensing Letters, Vol 11, No 11, pp 1886-1890, 2014.
23. D Jarchi, L Atallah and G-Z Yang, 'Transition detection and activity classification from wearable sensors using singular spectrum analysis', Proceedings of the 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp 136-141, 9-12 May 2012.
24. S M Mohammadi, S Enshaeifar, M Ghavami and S Sanei, 'Classification of awake, REM and NREM from EEG via singular spectrum analysis', Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp 4769-4772, 25-29 August 2015.
25. I Trendafilova, 'An inverse vibration-based approach towards modelling and damage identification in non-linearly vibrating structures. Application for delamination detection in a composite beam', Journal of Physics: Conference Series, Vol 382, p 012030, 2012.
26. D Garcia, R Palazzetti, I Trendafilova, C Fiorini and A Zucchelli, 'Vibration-based delamination diagnosis and modelling for composite laminate plates', Composite Structures, Vol 130, pp 155-162, 2015.
27. B Muruganatham, M A Sanjith, B Krishna Kumar, S A V Satya Murty and P Swaminathan, 'Inner race bearing fault detection using singular spectrum analysis', Proceedings of the 2010 IEEE International Conference on Communication Control and Computing Technologies (ICCCCT), pp 573-579, 7-9 October 2010.
28. B Muruganatham, M Sanjith, B Krishna Kumar and S A V Satya Murty, 'Roller element bearing fault diagnosis using singular spectrum analysis', Mechanical Systems and Signal Processing, Vol 35, Nos 1-2, pp 150-166, 2013.
29. H Hassani, 'Singular spectrum analysis: methodology and comparison', Journal of Data Science, Vol 5, No 2, pp 239-257, 2007.
30. T Liu, J Chen and G Dong, 'Singular spectrum analysis and continuous hidden Markov model for rolling element bearing fault diagnosis', Journal of Vibration and Control, Vol 21, No 8, pp 1506-1521, 2015.
31. H Al-Bugharbee and I Trendafilova, 'A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling', Journal of Sound and Vibration, Vol 369, pp 246-265, 2016.
32. C.W.R.U.B.D.C. website. Available at: http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website [Accessed: 7 March 2014]