Engineering Applications of Artificial Intelligence 50 (2016) 287–301

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Observer-biased bearing condition monitoring: From fault detection to multi-fault classification

Chuan Li [a,1], José Valente de Oliveira [b,*,1], Mariela Cerrada [c,1], Fannia Pacheco [e], Diego Cabrera [e], Vinicio Sanchez [e], Grover Zurita [d,1]

[a] Research Center of System Health Maintenance, Chongqing Technology and Business University, Chongqing 400067, China
[b] CEOT, Universidade do Algarve, Faro, Portugal
[c] Universidad de Los Andes, Mérida, Venezuela
[d] Universidad Privada Boliviana, Cochabamba, Bolivia
[e] Universidad Politecnica Salesiana, Cuenca, Ecuador

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
21 December 2015
Accepted 5 January 2016

Keywords:
Fuzzy clustering
Observer-biased clustering
Wavelets packed transform
Bearing
Fault detection
Fault diagnosis

ABSTRACT

Bearings are simultaneously a fundamental component and one of the principal causes of failure in rotary machinery. The work focuses on the employment of fuzzy clustering for bearing condition monitoring, i.e., fault detection and classification. The output of a clustering algorithm is a data partition (a set of clusters) which is merely a hypothesis on the structure of the data. This hypothesis requires validation by domain experts. In general, clustering algorithms allow a limited usage of domain knowledge on the cluster formation process. In this study, a novel method allowing for interactive clustering in bearing fault diagnosis is proposed. The method resorts to shrinkage to generalize an otherwise unbiased clustering algorithm into a biased one. In this way, the method provides a natural and intuitive way to control the cluster formation process, allowing for the employment of domain knowledge to guiding it. The domain expert can select a desirable level of granularity ranging from fault detection to classification of a variable number of faults and can select a specific region of the feature space for detailed analysis. Moreover, experimental results under realistic conditions show that the adopted algorithm outperforms the corresponding unbiased algorithm (fuzzy c-means) which is being widely used in this type of problems.

1. Introduction

Bearings are elemental mechanical components in rotary machinery (engines, gearboxes, propellers, turbines, etc.) that have been identified as one of their primary cause of failure, cf. (Yaqub et al., 2012). For example, in induction motors metal bearing faults account up to 40% of all faults (Siyambalapitiya and McLaren, 1990).

Rolling element bearings, such as ball bearings, consist of an inner, an outer race or ring, inside which a set of rolling elements rotate which are all prone to faults. In some models, a cage holds the rolling elements. Bearing faults can have different causes such as excessive load, lubricant failure, or corrosion. Studies exist comparing bearings with different materials and failure mechanisms (Sreenilayam-Raveendran et al., 2013). In general, faults result in abrasion due to steady friction of mechanical parts that, in turn, can have severe consequences for the overall system where the bearing is working in. The healthy condition of the bearings is directly related to the safe and effective operation of mechanical systems (Li et al., 2015). The result of a bearing failure can be catastrophic. This is the case of metal engine bearings supporting a crankshaft. Should this bearing fails the whole engine can disintegrate. Therefore, it is apparent the need for early detection and diagnosis of such faults.

A fault can be classified according to the location where it occurs: at the inner race, outer race, or at the rolling body, cage included. Also, a fault can be classified according to its type: it can be (i) a single point, (ii) localized within a certain region, or (iii) a generalized roughness fault. Often, localized or generalized faults originate from single point faults. The study focuses on metal ball bearings with single point faults. Fig. 1 shows some examples of the considered single-point faults in each one of the main components of a bearing. Different faults can and do occur simultaneously and are considered as well in this work. On the other hand, cage faults are not considered in this study. When present, a cage holds the rolling elements in position and its failure is

* Corresponding author.
[1] Prometeo researcher.

http://dx.doi.org/10.1016/j.engappai.2016.01.038
0952-1976/

Fig. 1. Examples of the three basic types of bearing faults actually studied in this work: (a) ball bearing fault, (b) inner race fault, and (c) outer race fault.

normally secondary as it is due to the failure of the other 3 main bearing components and, as such, cage faults are not normally studied in bearing diagnosis.

Bearing faults leave a trace in the vibration signal captured by accelerometers, the so-called fault signature. Acoustic, electric, thermal, or oil debris signals, e.g., Navarro et al. (2010), Oskouei and Esmaeili (2012), Oh et al. (2012) can also be used for detecting such faults, however vibration analysis is still one of the more widely used method and is also the method used in this work. This is due mainly to the following two aspects: (i) comparatively to electric, thermal, or oil debris signals, the vibration signal is more sensitive to the local defects of rotating components, and (ii) compared to other sensors such as acoustic emission or oil debris sensors, vibration sensors are much cheaper.

Bearing fault diagnosis involves the following general data pipeline: data acquisition and conditioning, feature extraction, feature selection, and classification. Typically, time, frequency, and time-frequency features are extracted from the collected signals. Feature selection is a critical step for optimizing efficiency, accuracy and for mitigating overtraining. Feature selection can be accomplished by experts with or without the help of feature selection methods. These include the employment of genetic algorithms, e.g., Lei et al. (2007), correlation-based methods such principal component analysis (PCA), e.g., Xu et al. (2009), Vijay et al. (2013), Ben Ali et al. (2015), fuzzy measures, e.g., Liu et al. (2008), rough sets (Zhao et al., 2008), orthogonal fuzzy neighborhood discriminant analysis, e.g., Abed et al. (2014), or entropy based criteria like those used for growing decision trees (Robin et al., 2010). The latter computes the information degree contributed by each feature and is the method adopted here. See Section 3.3 for details.

This work focuses on the employment of fuzzy clustering for fault detection and classification. The notions of cluster and clustering can have different meanings. In this paper, we are interested in (partition-based) clustering algorithms that can be viewed as a function mapping patterns (or observations) in a finite, otherwise unlabeled multi-variate date set $\mathbf{X}$ to partitions in $\mathcal{P}$, the set of all $\mathbf{X}$ dimensional compatible partitions. The problem is to partition $\mathbf{X} \subset \mathbb{R}^d$ space into groups (clusters) so that data in one group are similar to each other and are as dissimilar as possible from data in other groups. The (dis)similarities are evaluated through a suitable distance function that satisfies the three properties of a metric: reflexivity, symmetry, and triangle inequality.

Essentially, fuzzy clustering differs from conventional (hard) clustering in the sense that it allows an observation to belong, with different membership degrees, to more than one cluster, cf. (Valente de Oliveira and Pedrycz, 2007). Each membership degree can express how ambiguously or definitely an observation belongs to a given cluster and, under appropriated constraints, can be interpreted as the probability of an observation be a member of a cluster.

Currently there is a wealth of clustering algorithms available. The following focuses only on the algorithms used in studies on bearing fault diagnosis. For a broader perspective on the currently available fuzzy clustering algorithms the interested reader is referred to Valente de Oliveira and Pedrycz (2007).

1.1. Fuzzy clustering algorithms in bearing fault diagnosis

Fuzzy c-means (FCM) is the most popular and widely used fuzzy clustering algorithm in bearing fault diagnosis. Despite being well-known the algorithm is briefly revised here for easy reference. FCM aims at minimizing the objective function J (1) for a specified number of cluster c and a given set of observations $\mathbf{X} = \{\vec{x}_1, \ldots, \vec{x}_j, \ldots, \vec{x}_N\}$

$$J = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^m \| \vec{x}_j - \vec{v}_i \|^2 \tag{1}$$

under the constraints $u_{ij} \in [0,1]$, $\sum_{j=1}^{N} u_{ij} > 0$, and $\sum_{i=1}^{c} u_{ij} = 1$, where u_{ij} represents the membership of observation $\vec{x}_j (j = 1, \ldots, N)$ in the i-th cluster $(i = 1, \ldots, c)$, $\vec{v}_i$ refers to the centroid of the i-th cluster, $\|.\|$ stands for a norm distance in $\mathbb{R}^d$, $m > 1$ being the so-called fuzziness parameter. Increasing m increases the overlapping among the clusters. On the other hand, when $m \to 1$ FCM degenerates into k-means. FCM optimizes J through an iterative process where in each iteration, the centroid of the i-th cluster is updated using:

$$\vec{v}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \vec{x}_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{2}$$

The elements of the partition matrix, u_{ij}, i.e., the membership degrees are computed as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\| \vec{x}_j - \vec{v}_i \|}{\| \vec{x}_j - \vec{v}_k \|} \right)^{\frac{2}{m-1}}} \tag{3}$$

FCM has been extensively studied in bearing fault diagnosis as an exploratory tool (Jia et al., 2005; Guan et al., 2006; Wadhwani et al., 2006; Cui et al., 2008; Pan et al., 2009, 2009; Sui et al., 2010; Fu et al., 2011; Jiang et al., 2011; Ye et al., 2011; Zhang et al., 2011; Cao et al., 2012; Liu and Han, 2012; Xu et al., 2012; Xinbin et al., 2012; Wang et al., 2012b,a, Zanoli and Astolfi, 2012; Liu and Han, 2013; Vijay et al., 2013; Liang et al., 2015; Ou and Yu, 2014; Wang et al., 2014; Meng et al., 2014; Liu et al., 2014; Zhang et al., 2014; Zheng et al., 2015).

Other authors studied specific variants of the algorithm. This is the case of Jiang et al. (2010) in which a specific cost-functional based partitioning clustering algorithm is derived, and the work in Sui et al. (2008), Zhang et al. (2009), Cao et al. (2012) that uses the kernel-based fuzzy c-means.

FCM is one the simplest soft clustering methods, it is well understood, and the experimental evidence reported in the above literature clearly shows its effectiveness in this type of problem, especially when contrasted with more classical methods such as k-means.

Beyond FCM, in Wang et al. (2012) the Gath–Geva clustering algorithm is used, and in Zhang et al. (2014) the Gath–Geva, Gustafson–Kessel, and FCM clustering algorithms are compared. In brief, Gustafson–Kessel algorithm consider ellipsoidal clusters, and Gath–Geva considers ellipsoidal clusters with independent shapes and sizes. Other studies on clustering for bearing faults diagnosis are the work of Chen et al. (2010) where an application of wavelet, PCA, and fuzzy K-nearest neighbor is studied, and the work of Jiang et al. (2011) where an equivalence fuzzy relation based clustering for autoregressive model estimation is proposed. In Zhang et al. (2013) combines clustering with fuzzy support vector machines. No further work on fuzzy clustering based bearing fault diagnosis is found when the scientific index engine Scopus[2] is searched with the string "TITLE-ABS-KEY (fuzzy AND fault bearing) AND PUBYEAR > 1999".

The motivation to use an unsupervised approach such as fuzzy clustering to fault classification is twofold: It is clear that with enough labeled data supervised learning methods are more effective in terms of accuracy. However, when the available data for certain type of faults is scarce or when the type of fault is unknown, unsupervised approaches such as clustering become relevant. Another reason for using clustering is for membership functions elicitation of linguistic variables. This is a common approach to automatically tune fuzzy rule-based, and fuzzy relational systems.

In this study, a novel method allowing for interactive clustering in bearing fault diagnosis is proposed. The method provides a natural and intuitive way to control the cluster formation process, allowing the employment of domain knowledge to guide the clustering process. This allows the user to (i) select a suitable level of granularity, e.g., select between the fault detection case (only two clusters) and fault classification with a variable number of faults, and (ii) to explore in detail specific regions of the feature space. Given these contributions one may legitimately ask what are benefits of doing this, if any? In an attempt to address this question, we start by recalling that finding an optimal partition – independently of the criterion used – is computationally intractable (Fazendeiro et al., 2008; Fazendeiro and Valente de Oliveira, 2015). As a consequence, clustering algorithms explore a small portion of the partition space. The explored region depends on the algorithm and on the underlying assumption on the data distribution, heuristics, or employed model. Exploratory analysis using clustering involves a number of steps including, but not limited to: selection and application of clustering algorithm, (internal) validation according to some cluster validity index, and a final, crucial, interpretation step. This interpretation depends on the application field and relies on the existence of domain experts. Without this interpretation step, results can be completely meaningless. As an illustration, consider a dataset **U** sampled from an uniform distribution. Apply a clustering algorithm over **U** and the algorithm will produced some "clusters" as output. Furthermore, different outputs from different clustering algorithms can be even evaluated against each other, and eventually ranked. All this despite the fact that no structure at all exists in **U**. Thus domain experts are crucial in the identification of *meaningful* clusters, i.e., clusters that actually represents a structure in data. Conventional hierarchical clustering can be viewed as a first attempt to address the above concerns. Hierarchical clustering provides a hierarchy of clusters and let the user decide on their suitable number (level of granularity) cf., e.g., Valente de Oliveira and Pedrycz (2007). Our approach offers two significant advantages over hierarchical clustering: (i) it allows for the detailed exploration of chosen regions of the feature space, and (ii) it allows for the reassignment of observations to clusters when their number changes.

The remaining of the paper is organized as follows. Section 2 presents the proposed clustering algorithm. Section 3 presents the data pipeline required for experimental evaluation of our proposal. It includes short descriptions of data acquisition, signal processing used for feature extraction, and the feature selection process. Section 4 presents both the experimental results and a brief discussion. Section 5 ends the paper with the main conclusions.

2. The observer biased clustering algorithm

The adopted clustering algorithm is inspired by the following metaphor (Fazendeiro and Valente de Oliveira, 2015). In daily live, the visual perception of a group of objects depends on the observer position. The closer an observer is from a set of objects the clearer the set is perceived. Inversely, the farthest the observer is from the objects less detail is visualized. Zooming in and out with an optical lens has exactly this effect. When too close each object is clearly visible, when too far all objects are visualized as a single entity.

In this work, this metaphor is substantiated as follows. The observer position is modeled by a (focal) point based on which a regularization term is defined. Afterwards we use a form of statistical shrinkage to incorporate the regularization term in the objective function of FCM, otherwise viewed as an unbiased clustering algorithm. All this allows one to obtain different *reasonable* clusters, depending on the position of the observer. We use the term *reasonable cluster* in the MacQueen' sense. A reasonable cluster is a cluster that merely belongs to a partition revealing "reasonably good similarity groups" (MacQueen, 1967), as validated by a given internal cluster validity index. This should not be confused with a *meaningful cluster* that actually represents a data structure as recognized by a domain expert.

For formalizing the above let $\mathbf{X} = \{\vec{x}_1, \ldots, \vec{x}_j, \ldots, \vec{x}_N\}$ be a finite set of feature vectors in the $\mathbb{R}^d$ space. Associated with observation $\vec{x}_j$ there is a vector $\vec{u}_j$ such that each one of its elements represents the membership of the observation j in each one of the C classes of observations. Each vector's element, u_{ij}, $i = 1, \ldots, C$, $j = 1, \ldots, N$ takes values in $[0, 1]$, 0 standing for non-membership while 1 corresponds to total membership. Vectors $\vec{u}_j$ can be arrayed as the columns of a $C \times N$ partition matrix, $\mathbf{U} = [u_{ij}] \in \mathbb{R}^{C \times N}$. For the sake of correctness of matrix **U**, the following conditions should be satisfied:

$$u_{ij} \in [0, 1] \text{ for all } i = 1, \ldots, c \quad \text{and} \quad j = 1, \ldots, n \tag{4a}$$

$$\sum_{i=1}^{C} u_{ij} = 1 \quad \text{for all } j = 1, \ldots, N \tag{4b}$$

Furthermore, we assume that the clustering problem is reasonable in the sense that the number of different observations N is much greater than the number of classes, C, i.e., $N \gg C$.

The center of the i cluster is denoted by $\vec{v}_i$ with $\vec{v}_i \in \mathbb{R}^d$. These centers can be arrayed into $\mathbf{V} = [\vec{v}_1, \ldots, \vec{v}_C]$ with $\mathbf{V} \subset \mathbb{R}^{C \times d}$. A regularization term, based on the focal point $\mathbf{P} \in \mathbb{R}^d$ can then be embodied into the cost functional of FCM (1) as follows:

$$Q = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m \| \vec{x}_j - \vec{v}_i \|^2 + \zeta \sum_{i=1}^{C} \| \mathbf{P} - \vec{v}_i \|^2 \tag{5}$$

under constraints (4a) and (4b). The regularization coefficient is $\zeta \geq 0$ and allows one to adjust between the unbiased algorithm (FCM) for $\zeta = 0$ and a biased one. Notice that Q is the sum of two

[2] http://www.scopus.com

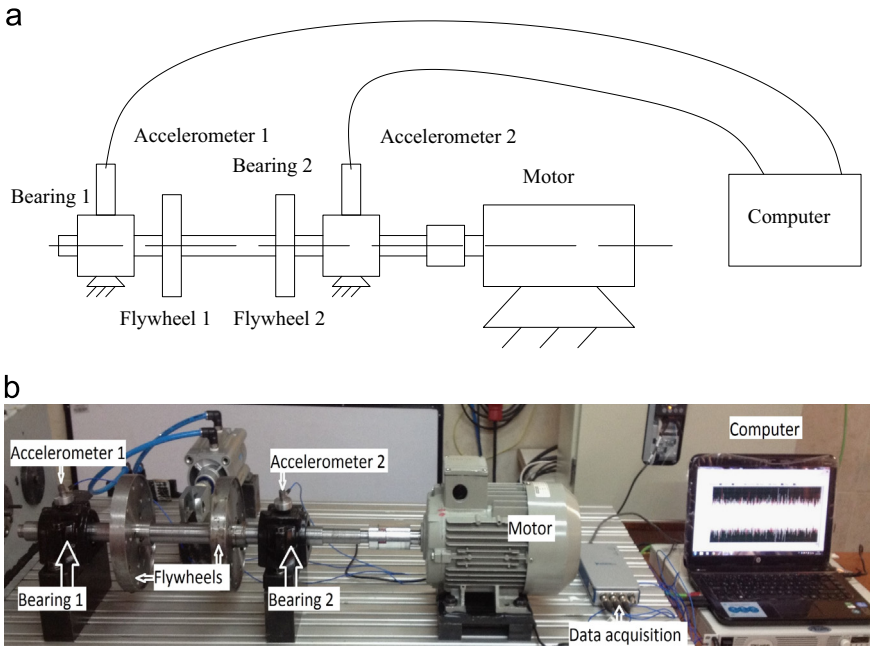

Fig. 2. The experimental apparatus: (a) Schematic test set-up and (b) the actual set-up.

non-negative terms. The second term (i.e., the regularization term) will be zero (minimum) when $\zeta = 0$ or when all prototypes are equal to **P**. If **P** is far enough from the data points, as the *i*th-cluster approaches **P** there is no data belonging to it and thus the membership values u_{ij} will tend to zero. In practice this is equivalent to remove the prototype $\vec{v}_i$.

Consider for instance an unconstrained setup with as many prototypes as data points ($C=N$). We can think of $\zeta = 0$ as the case where the observer is so close to data that each datum is regarded as a cluster; as ζ increases some prototypes are subsumed by **P** resulting in less and less clusters each one of which with more and more elements. One can resort to alternating optimization to minimize (5) under constraints (4a) and (4b). The optimization problem can then be converted into an unconstrained one using Lagrange multipliers, yielding the updating expressions:

$$u_{ij} = \frac{(1/\| \vec{x}_j - \vec{v}_i \|)^{2/(m-1)}}{\sum_{k=1}^{N} (1/\| \vec{x}_k - \vec{v}_i \|)^{2/(m-1)}} \tag{6}$$

$$\vec{v}_i = \frac{\sum_{j=1}^{N} u_{ij}^m \vec{x}_j + \zeta \mathbf{P}}{\sum_{j=1}^{N} u_{ij}^m + \zeta}. \tag{7}$$

The algorithm can be viewed as a sequence of Picard iterations through the necessary conditions (6) and (7) and is given as Algorithm 1 in the Appendix.

The number of clusters C is one of the most important parameter of a partitioning clustering algorithms. Only if C equals the (usually unknown) number of subgroups in the data there is a possibility that the clustering process effectively reveals the existent structure of the data. Often, the merit of selecting a given C is evaluated by a cluster validity analysis. One possibility consists in running the clustering algorithm several times with different initializations, for a sequence of C values. The number C which optimizes the validity measure is elected as the best one.

In order to find a number of reasonable clusters we employ an iterative algorithm which consists of successive runs of Algorithm 1 (in the Appendix) with increasing values of ζ given that **P** is enough distant from data. To ensure this last assumption the focal point **P** is placed in a higher dimensional space $\mathbb{R}^{(d+1)}$. In this way, the algorithm begins by representing both data and the prototypes, initially located in the original input space $\mathbb{R}^d$, into $\mathbb{R}^{(d+1)}$ – a simple way of providing this transformation is by introducing one extra null coordinates both in the data and in the prototypes. Afterwards, the number of clusters is overestimated and denoted by $C_{\max}$. This results in a reduced influence on the overall weights of the partition matrix by those prototypes that are attracted to the focal point. In each iteration of this meta-process the number of candidate clusters is determined and the Xie-Beni validity index is calculated – Eq. (8). Some of the prototypes may have been attracted to the neighborhood of the focal point, in the $\mathbb{R}^{(d+1)}$ space, and can be removed. This meta-process finalizes by producing the best partitions obtained regarding the validity index employed. This iterative process is specified as Algorithm 2 and given in the Appendix.

$$XB = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \| \vec{x}_j - \vec{v}_i \|^2}{n \min_{i \neq j} \| \vec{v}_i - \vec{v}_j \|^2} \tag{8}$$

3. Material and methods

This section presents the data pipeline for bearing fault diagnosis, i.e., it includes short descriptions of the setup used for data acquisition, the signal processing used for feature extraction, and the criterion for feature selection.

3.1. Experimental apparatus

The experimental apparatus used to collect data is displayed in Fig. 2 and consists of the equipment specified in Table 1. In brief, two bearings are installed in a ∅30 mm shaft and mounted in their housings. The shaft is driven by the motor that is controlled by a driver inverter (not identified in the figure). Flywheels are mounted on the shaft when load is required. One accelerometer is installed in each bearing housing for measuring the vibration signals that are collected by the data acquisition card. This test rig allow us to study fault interferences, e.g., the effect of a faulty bearing 1 in the readings of accelerometer 2.

A total of $63 \times 5 = 315$ experiments are performed. Each experiment is characterized by a tuple $\langle speed, load, bs \rangle$ where *speed* is the shaft speed, *load* is the total load on the shaft, and *bs* stands for the bearing status. Given the low variability of the results, each experiment is only repeated 5 times. Three discrete speeds are tested: 8, 10, and 15 Hz. Also, three different types of loads are essayed: zero, one, and two flywheels. The essayed bearing status are described in Table 2. The sampling frequency is 50 kHz being determined considering the following. High frequency band signals, between 1 and 20 kHz, are indicators of faults in bearings. According to the Nyquist–Shannon theorem the sampling frequency should be at least twice the signal highest frequency (20 kHz). Therefore we choose 50 kHz for securely meeting this requirement.

The duration of each sample (measurement time) is 20 s. Fig. 3 shows the first two seconds of two of these signals (samples) corresponding to the same experimental conditions except for bearing health; more concretely, the signals corresponds to the first run of experiments $\langle speed = 8\ \text{Hz}, load = \text{none}, bs = \text{P1} \rangle$ and $\langle speed = 8\ \text{Hz}, load = \text{none}, bs = \text{P3} \rangle$. Clearly the signal corresponding to a faulty case has more energy than the signal corresponding to the healthy case.

Table 1
Specifications of the equipment used in the experimental apparatus.

Component	Specification
Accelerometer 1 & 2	PCB Icp 353c03
Bearing 1 & 2	SKF 1207 Ektn9/C3
Data Acquisition	NI Cdaq-9234
Housing 1 & 2	SKF Snl 507-606
Inverter	Danfoss VLT 1.5kw
Motor	Siemens 1LA7 090-4YA60 2Hp
Shaft diameter	30 mm
Tachometer	Vls5/T/Laser Optical Sensor
Type of load	flywheels

Table 2
Health status of the essayed bearings.

Id	Bearing 1	Bearing 2
P1	healthy	healthy
P2	inner race fault	healthy
P3	outer race fault	healthy
P4	ball fault	healthy
P5	inner race fault	outer race fault
P6	inner race fault	ball fault
P7	outer race fault	ball fault

3.2. Computed features

For each accelerometer 817 features (thus 1634 in total) are computed and are distributed as follows. Seven time domain features, 730 frequency domain, and 80 time-frequency features. In the frequency and the time-frequency domains, signals were divided in 80 bands of 20 KHz each. Afterwards features are computed for each one of these bands. A band is identified by a number between 1 and 80.

Many other features could have been considered, however the below described ones were selected both for their computational efficiency and for our *a priori* knowledge on the their suitability for this application e.g., Li et al. (2012), Li Chuan (2012), Li et al. (2015), Tianyang et al. (2015).

3.2.1. Time domain features

Time domain features include statistics like the mean (μ), standard deviation (σ), variance (σ^2), root mean square (rms),

$$\text{rms}[x] = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{9}$$

kurtosis, k,

$$k[x] = \frac{N\sum_{i=1}^{N}(x_i-\mu)^4}{[\sum_{i=1}^{N}(x_i-\mu)^2]^2} \tag{10}$$

where x is the signal, and N is its duration. Kurtosis is largely used as an indicator of major peaks in a signal and has been shown to be, to a great extend, independent of load and speed variations (Gupta, 1997); Consequently, we have also considered the kurtosis of the speed, acceleration, and of the derivative of the acceleration of the signal. We refer to the former as the *kda* operator which is defined as,

$$\text{kda}[x] = k\left[\frac{d}{dt}\text{Acceleration}(x(t))\right] = k\left[\frac{d^3}{dt^3}x(t)\right] \tag{11}$$

where k is the kurtosis given by (10). Skewness (s), a measure of

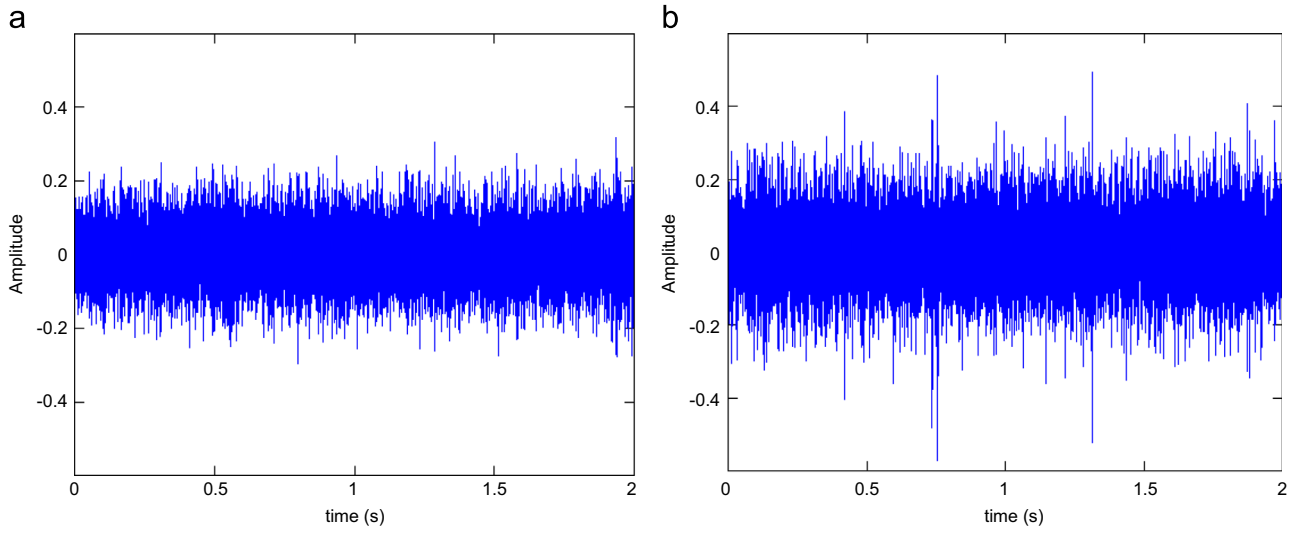

Fig. 3. Signals measured by accelerometer 1 corresponding to experiments (a) $\langle speed = 8\ \text{Hz}, load = \text{none}, bs = \text{P1} \rangle$ and (b) $\langle speed = 8\ \text{Hz}, load = \text{none}, bs = \text{P3} \rangle$. Clearly, the signal (b) corresponding to an outer race fault in bearing 1 has higher energy than the other corresponding to healthy bearings.

Table 3
Used wavelet packet transforms (WPT) and their properties.

WPT	Properties
Biorthogonal (bior6.8)	symmetric, not orthogonal, biorthogonal.
Coiflets (coif4)	near symmetric, orthogonal, biorthogona
Daubechies (db7)	asymmetric, orthogonal, biorthogonal
Symlets (sym3)	near symmetric, orthogonal, biorthogonal
Reverse Biorthogonal (rbio6.8)	symmetric, not orthogonal, biorthogonal

asymmetry of a distribution around its mean,

$$s[x] = \frac{\sum_{i=1}^{N}(x_i-\mu)^3}{N\sigma^3} \tag{12}$$

and the crest factor, *cf*,

$$\mathrm{cf}[x] = \frac{\max[x]}{\mathrm{rms}[x]} \tag{13}$$

were also considered.

3.2.2. Frequency domain features

The vibration time-domain signals are transformed into frequency signals using Fast Fourier Transform (FFT). The results are divided into 80 bands of 20 KHz each. For each band: mean, root mean squares, standard deviation, and kurtosis are computed for both linear amplitude and amplitude in db; for a subtotal of 640 features. In addition, 15 octaves are considered and for each one of them: mean, standard deviation, and kurtosis are computed for both linear and db amplitudes; yielding a subtotal of 90 features. Thus a total of 730 frequency domain features are computed for each accelerometer.

3.2.3. Time-frequency domain features: wavelets packet transforms

For the time-frequency domain, and for each of the 80 bands five wavelets packet transforms are computed, see Table 3.

Wavelet transforms, especially continuous ones, can be viewed as a generalization of Fourier transforms where the trigonometric functions are replaced by wavelet functions. A wavelet function describes a small oscillatory wave whose energy is concentrated in time. Following Ruqiang et al. (2014), the general expression of a continuous wavelet transform, *cwt*, of the time-domain signal $x(t)$ is

$$cwt(S,\tau) = \frac{1}{\sqrt{s}}\int x(t)\Psi^*\left(\frac{t-\tau}{S}\right)dt \tag{14}$$

where S and τ stand for the scale and the translation parameters, respectively, while Ψ^* is the complex conjugate of the wavelet function Ψ.

Wavelet packet transforms are a particular type of discrete wavelet transforms that allows one to assess the detailed information of signals in high frequency bands. Consequently, WPT can be used to explore data features, such as high frequency transients, that might be missed by more traditional methods, such as FFT. Table 3 summarizes the properties of the used WPT.[3]

3.3. Entropy based feature selection

For feature selection we follow the technique proposed in Robin et al. (2010). Although well-known, the main idea is briefly revised here for easy reference. A feature X_j is selected so that it yields the maximum information gain on the data set $\mathbf{X}$, or equivalently that maximizes the reduction in entropy, i.e., that maximizes $I(\mathbf{X},X_j) = H(\mathbf{X}) - H(\mathbf{X},X_j)$ where $H(\mathbf{X})$ is the entropy of the data set before selecting any feature, $H(\mathbf{X}) = -\sum_{i=1}^{C} p_i \log_2 p_i$ with the probability of i-th class, $p_i = n_i/N$ where n_i is the number of samples belonging to class i; $(i = 1,\ldots,C)$ and N is the cardinality of $\mathbf{X}$. Moreover, the conditional entropy (i.e., the entropy after selecting the j-th feature is $H(\mathbf{X},X_j) = -\sum_x p(X_j = x)\sum_y p(X = y|X_j = x)\log_2 p(X = y|X_j = x)$.

This method was found to be much faster and more discriminative than the usually employed genetic algorithm (GA) based feature selection method. In particular, this method selected only 12 relevant features out of 1634 while GA based method selected a number of features one order of magnitude higher typically. This is a direct consequence of the fact that GA are general purpose optimizers whose search is guided by the fitness function only while the above method can be viewed as a specialized method that evaluates the contribution of each feature.

Table 4
Selected features based on maximum information gain.

No.	Domain	ID	Acclr.	Obs
1	time	rms	1	Eq. (9)
2	time	kda	1	Eq. (11)
3	time	cf	1	Eq. (13)
4	freq.	std of FFT band 3	1	
5	freq.	std of FFT band 9	1	
6	freq.	std of FFT band 30	1	
7	time-freq	WPT (db7)–1	1	Energy of node 1 in the db7 decomp.tree
8	time-freq	WPT (db7)–5	1	Energy of node 5
9	time-freq	WPT (sym3)–12	1	Energy of node 12
10	time	cf	2	Eq. (13)
11	time-freq	WPT (db7)–5	2	Energy of node 5
12	time-freq	WPT (coif4)–15	2	Energy of node 15

4. Results and discussion

This section presents some experimental results and a brief discussion. The section starts by presenting and discussing the results of applying the adopted feature selection method to the computed features. The treatment of outliers is also discussed. Afterwards, a presentation of the clustering results obtained is presented for both the fault detection and fault classification cases. This includes some experimental evidence on the merits of FCMFP and its iterated version to providing a natural and intuitive way to control the cluster formation process, allowing the user to iteratively select a suitable level of granularity while searching for meaningful clusters in a given region of the feature space. A comparison of the proposed biased algorithm (FCMFP) with the corresponding unbiased one (FCM) is also given. This comparison measures the quality of the resulting partitions using the external quality Rand and Adjusted Rand indices and check for statistical significant differences between the obtained indices using the non-parametric statistical Wilcoxon signed-rank test.

4.1. Features

When the proposed methodology is applied to the data acquired in the above described experiments 12 out of 1634 features are selected as the most relevant ones. Table 4 presents these features.

Both fault distribution among bearings (see Table 2) and fault interferences dictate that Accelerometer 1 is the most relevant being responsible for capturing 9 of the 12 selected features. Wavelets, i.e., time-frequency domain features corresponds to 5 of

[3] Ready to use implementations of these and other WPT are available from several signal processing software packages.

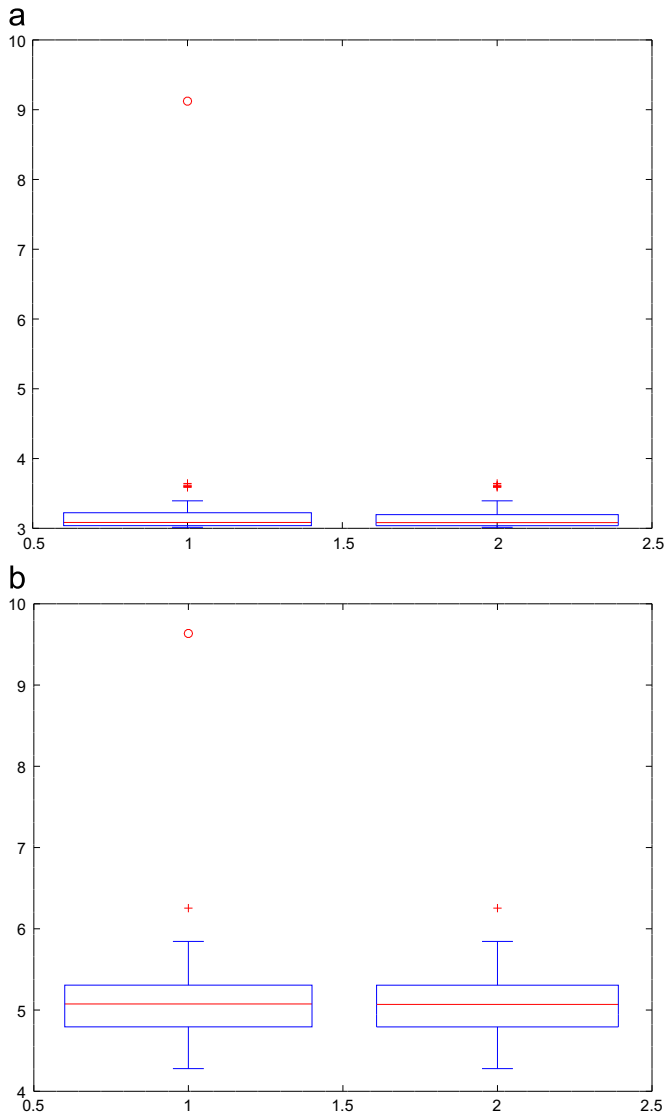

Fig. 4. Boxplots exhibiting the distributions of selected features (1) with and (2) without outliers for healthy bearings: (a) time domain kda (Acceler.1); and (b) time cf (Acceler.1). Each boxplot shows, from bottom up: the minimum, the first quartile, the second quartile (median), the third quartile and the maximum value. The symbol o denotes an outlier that lies outside 3 times the interquartile range while + denotes points between 1.5 and 3 times the interquartile range.

the total number of selected features; in addition, five time domain and three frequency domain features were selected.

When we verify the distribution of the values of the above features, outliers are observed even for healthy bearings. This occurs mainly in time domain features – see Fig. 4 – but also occurs in time-frequency domain, e.g., this is case of the feature corresponding to the energy of the leaf node 1 at the wavelet decomposition tree with Daubechies 7 wavelet (WPT (db7)–1). These outliers correspond to both mechanical and electrical noisy that may degrade the performance of the unsupervised method used and therefore were removed using the Thompson Tau method (Dieck, 2006).

The time-domain noise is even more severe in faulty states. Now all the four time domain features exhibit outliers together with WPT (db7)–1 (Accelerometer 1). Fig. 5 shows one of the more dramatic cases. The figure presents the boxplots of the feature kda (Accelorometer 1) (a) with and (b) without outliers. These outliers, identified by the symbol o, can be many times higher the

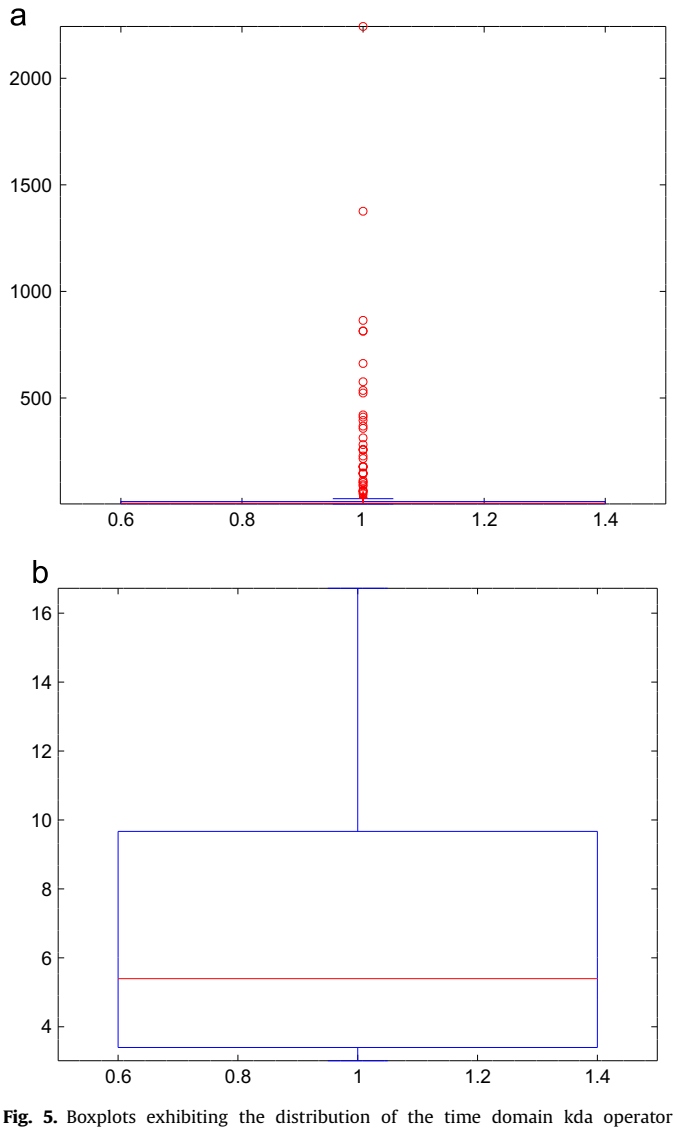

Fig. 5. Boxplots exhibiting the distribution of the time domain kda operator (Accelerometer 1) (a) with and (b) without outliers for faulty bearings.

interquartile range. Even in the faulty condition states, no outliers were observed in the other features.

4.2. Clustering results and discussion

Having applied Algorithm 2 over the above-mentioned 12 features, a wide range of reasonable partitions were revealed. The quality of the different structural alternatives yielded by the algorithm, as assessed by the Xie Beni index (8), is shown in Fig. 6. In this figure, the results correspond to a typical run of the algorithm where the fuzzifier coefficient was set to $m=2$, data were normalized for zero mean, and the focal point was set one order of magnitude above the baricenter of data, and $\Delta\zeta = 0.01$. The number of clusters corresponding to local maxima of the validity index is identified as they correspond to the different reasonable alternatives. These range from partitions with $c=2$ to $c=9$ clusters. Curiously enough $c=5$ was not signaled as a reasonable number of clusters for this problem. For validating the different alternatives suggested by Algorithm 2, internal (i.e., the inverse XB index) and external validation are considered. For external validation both the Rand Index (RI), the Adjusted Rand Index (ARI) (Hubert Lawrence, 1985) and the classical Sammon data dimensionality reduction method for data visualization (Sammon, 1969)

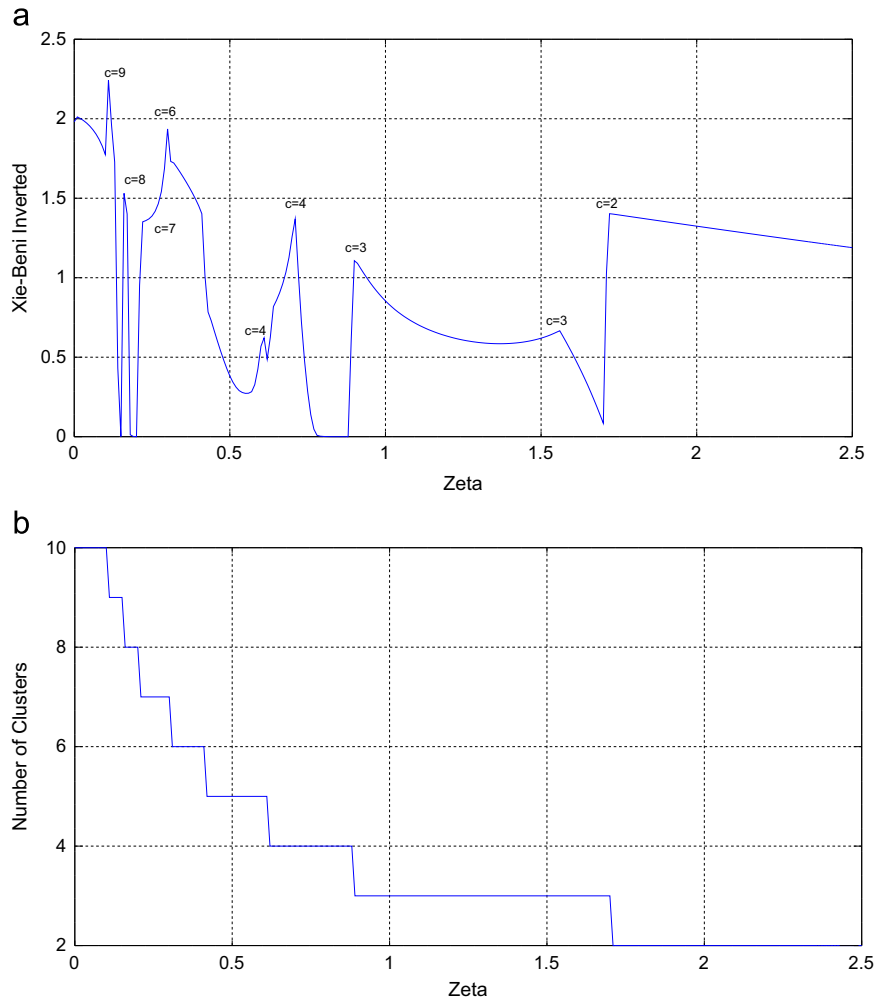

Fig. 6. Evolution of the interval validity index (8)(a) and the corresponding number of clusters (b) as function of ζ for a typical run of Algorithm 2. Peaks of the index reveal reasonable number of clusters. Five clusters were not signaled as a reasonable number.

are employed. Both RI and ARI are two of the most recommendable external validation indexes used to measure the similarity of two partitions. In our case these partitions are respectively the *ground truth* i.e., the real classification of faults, and a hypothetical partition generated by the clustering algorithm. In the Appendix we briefly review both of these indices. The Sammon method provides a nonlinear projection between a high-dimensional space and a lower dimensionality one attempting to preserve the structure of inter-point distances in both spaces.

4.2.1. The fault detection case

We start by inspecting the fault detection case, i.e., the case where we are only interested in verify whether a fault (no mater which) exists. This corresponds to clustering the data with $c=2$ clusters. Fig. 7 shows a Sammom projection of the 12-dimensional feature space into the plane. In this figure, the centers of each cluster are denoted by a red ∗. Around each cluster center there are ten solid line curves with different colors ranging from red to dark blue, each one of them representing a contour of equal membership value; the farthest the curve from the center the lower the membership value (the darker the blue the lower the membership). Each one of the other colored symbols represents to the truth classification of a sample. Samples in the same class are represented by the same color and symbol. More concretely the symbol green • denotes the bearing state identified as P1 (healthy state) in Table 2; symbols magenta ×, blue ⊕, red •, magenta ⊕, cyan ⊙, and black ⊙ corresponding to the states identified as P2, P3, P4, P5, P6, and P7 in the same table, respectively.

From this figure, we can observe that: (i) the total number of samples in the healthy state (marked with green •) are all in the same cluster, i.e., there is no heathy state case that is misclassified which is an asset in this type of applications as nobody wants to stop a machine for replacing a healthy bearing; (ii) however there are some classification errors. These errors are due mainly to samples in state *P*3, an outer race fault in bearing 1 (marked with blue ⊕) and a single sample of *P*4, a ball fault in bearing 1 (marked with red •), see also Table 2. It worths mentioning that from the strict mathematical point of view, this last data point, is completely separated from the other points with same fault.

Relatively to the ground truth, the presented partition has a $RI=0.81533$ and $ARI=0.63076$. Usually $RI>0.7$ corresponds to an acceptable clustering solution in the sense that such values cannot correspond to haphazard agreements between partitions. For further details on *RI*, *ARI*, and their interpretation, the reader is referred to the Appendix and to Hubert Lawrence (1985).

4.2.2. The multi-fault classification case

From the analysis of the Xie Beni index of Fig. 6(a) it follows that different perspectives on the data are possible. In particular, partitions with number of clusters $c=2,3,4,6,7,8$, and 9 provide reasonable

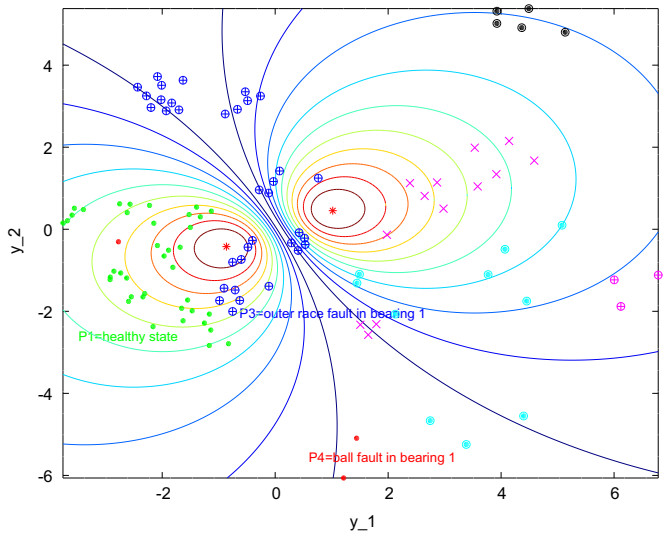

Fig. 7. A Sammom projection of the 12-dimensional feature space into the plane for the fault detection case ($c=2$). The centers of each cluster are denoted by a $*$. Around each cluster center there are ten solid line curves with different colors each one of them representing a contour of equal membership value; the farthest the curve from the center the lower the membership value. Each one of the other colored symbols represent the truth classification of a state. No healthy sample (marked with a green $\bullet$) is misclassified. See text for further details. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

structural alternatives. Fig. 8 shows examples of each one of these different perspectives resorting to Sammom projections, in a sequence that resemble a zooming in process. Starting from Fig. 7 and requiring a more detailed view (passing from 2 to 3 clusters) we observe in Fig. 8(a) that a cluster formed by blue $+$, corresponding to outer race faults in bearing 1 (state $P3$ in Table 2) is formed. From Fig. 8(a) and (b) we observe that the initial cluster corresponding to the healthy state is now subdivided in two clusters: one containing samples in healthy state or in $P3$ state, and other containing samples in healthy state and a sample in $P4$ (ball fault in bearing 1 identified by the red $\bullet$). Following the same process, the current biggest cluster at the right of Fig. 8(b) will be successively subdivided into 2 (Fig. 8 (b)) and 3 subclusters (Fig. 9(c)). At this point, most of faults $P2$ (identified by magenta $\times$) and most of faults $P6$ (identified by cyan $\odot$) have occupied respectively 2 that subclusters. This process of cluster purification continues in the remaining cases. Curiously enough the state $P7$, i.e., outer race fault in bearing 1 and ball fault in bearing 2 (black $\odot$) were always viewed by the algorithm as similar to state $P2$, i.e., inner race fault in bearing 1 and healthy bearing 2 (magenta $\times$), for all the levels of data granularity analyzed. The FCMFP algorithm was able to provide difference data perspectives even for the same number of clusters and the same $\mathbf{P}$. For example, for the run presented in Fig. 6(a) there is a range of ζ values (wider that [1; 1.5]) that yield the same number of clusters $c=3$ and about the same external validation indexes. For instance, for $\zeta=1$ AR=0.56368, RI=0.81274 were obtained. For $\zeta=1.5$ AR=0.60607, RI=0.83247 were obtained, and for $\zeta=1.5$ and AR=0.47217, RI=0.76774 was obtained. It is the prerogative of the domain specialist to choose the view that best fits her needs.

More interestingly, suppose that we are interested in a detailed analysis of a given region of the feature space, say the region where features attains their minimum or their maximum values. This is obtained by positioning the $\mathbf{P}$ within the region of interest. For space saving reasons Fig. 9 shows only three cases ($c=4,6,7$) of obtained results when $\mathbf{P}$ was positioned respectively at the minimum (left side subfigures) and at the maximum (right side subfigures) regions of the feature space. Similar results were obtained for other number of clusters c. Although the same number of clusters is visualized in each pair of subfigures, different levels of detail are observed in each one of the selected regions.

4.2.3. Comparison with the corresponding unbiased algorithm

In this section we temporarily assume that we are *not interested* in the natural and intuitive way to control the cluster formation process provided by FCMFP and its iterated version. We temporarily assumed that we are only interested in the quality of the final results. Under this assumption a reasonable question to ask is how the biased FCMFP compares with its unbiased counterpart (i.e., FCM)? To answer this question we ran both algorithms (FCM and FCMFP) 30 times for each one of the following number of clusters $c=2,3,4,6,7,8,9$ under the *same* initial conditions. Afterwards the quality of the resulting partitions measured by the external quality (Rand and Adjusted Rand) indexes were statistically evaluated.

The non-parametric statistical Wilcoxon signed-rank test (Sheskin, 2011) was used to analyze the obtained results. Non-parametric tests do not require neither normality nor homoscedasticity (equal variance) of the samples in analysis, what is very convenient in the present case. The test is used to answer the question of whether the two samples (quality of partitions in this case) represent two different populations.

The significance level considered was $\alpha=0.05$ corresponding to a confidence interval of 95%, i.e., if p-value <0.05 it is considered that there is a statistical significant difference between the data samples being analyzed (rejection of the null hypothesis). There is no statistical significant difference, otherwise.

The obtained results are summarized in Fig. 10 where a comparison of Adjusted Rand Indexes distributions over 30 independent runs of each algorithm for the different number of clusters can be observed. The figure also presents the corresponding statistical results obtained. From these experiments we conclude that FCMFP outperforms FCM for any number of clusters except for $c=6$ where there is no statistical significant different between the partitions produced by the algorithms. Similar results were obtained for the Rand index and are omitted for brevity. Although this type of improvements was *not* the primary goal beyond FCMFP it is a convenient side effect of the shrinkage technique employed in the design of the algorithm. Shrinkage allow one to balance a high variance estimator (FCM) with a high bias one (the focal point).

Best improvements are observed for small number of clusters and especially for $c=2$, i.e., for the fault detection case. In practical terms this means that, for the former case, most samples in the fault state $P3$ (blue $\oplus$) are now wrongly clustered together with samples in healthy state (marked with a green $\bullet$), see Fig. 11 and compare with Fig. 7.

5. Conclusions

In this paper the bearing fault diagnosis (detection and classification) problem was addressed. As found in the literature, soft clustering is recognisably a promising data analysis tool due to the very nature of the problem.

In this study we further exploit soft clustering techniques for bearing fault diagnosis. We have proposed the employment of the recent observer biased partitioning clustering algorithm named fuzzy C-means with a variable focal point (FCMFP). This algorithm allows the user to iteratively select a suitable level of granularity while searching for meaningful clusters, i.e., clusters that actually correspond to the underlying structure of the data. In addition, the

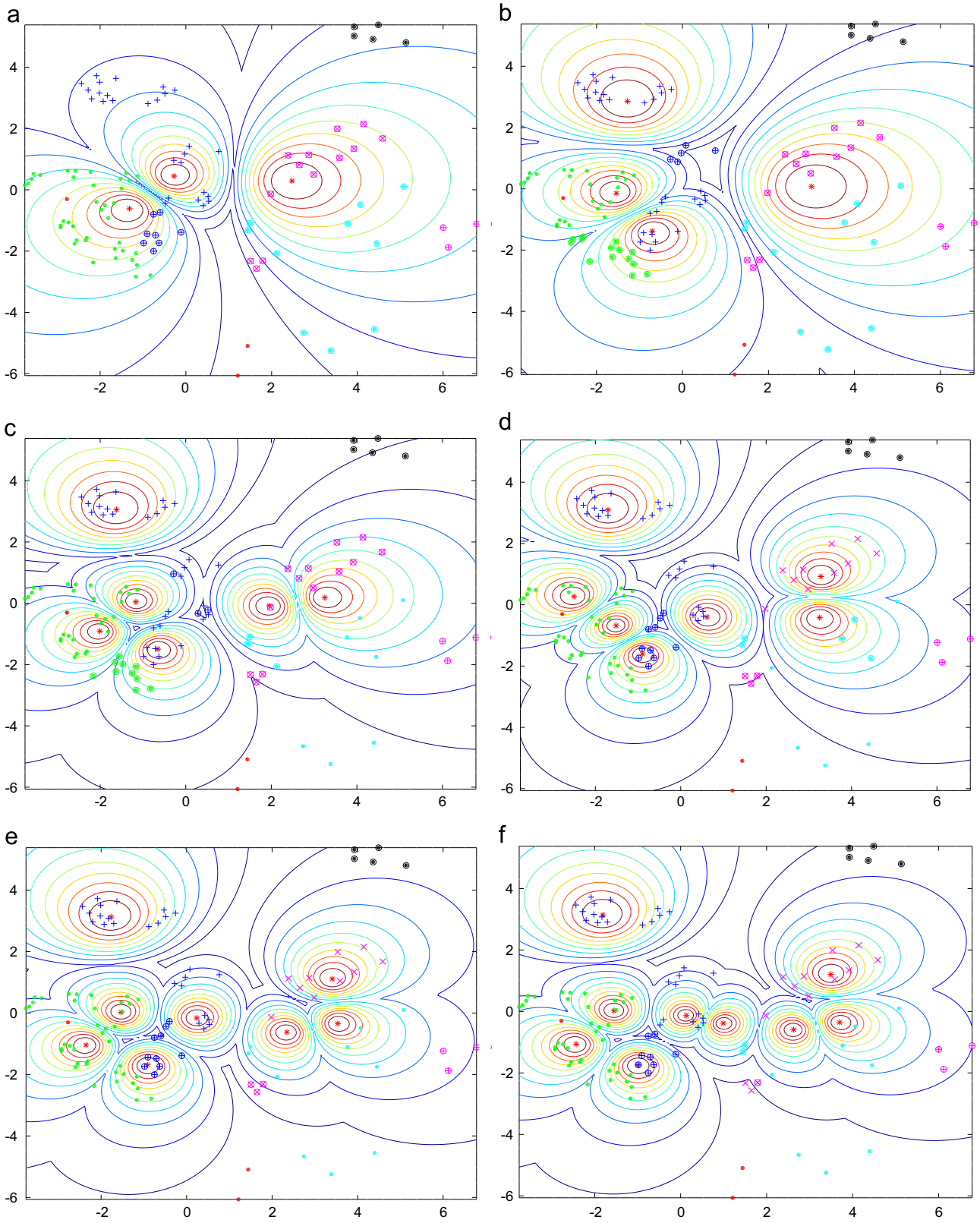

Fig. 8. Sammom projections of the 12-dimensional feature space into the plane for fault classification for: (a) 3, (b) 4, (c) 6, (d) 7, (e) 8, and (f) 9 clusters corresponding to the peak values in Fig. 6(a). The truth classification of the faults are represented by different colors and symbols, the centers (∗) as well as some membership contour curves are also represented. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

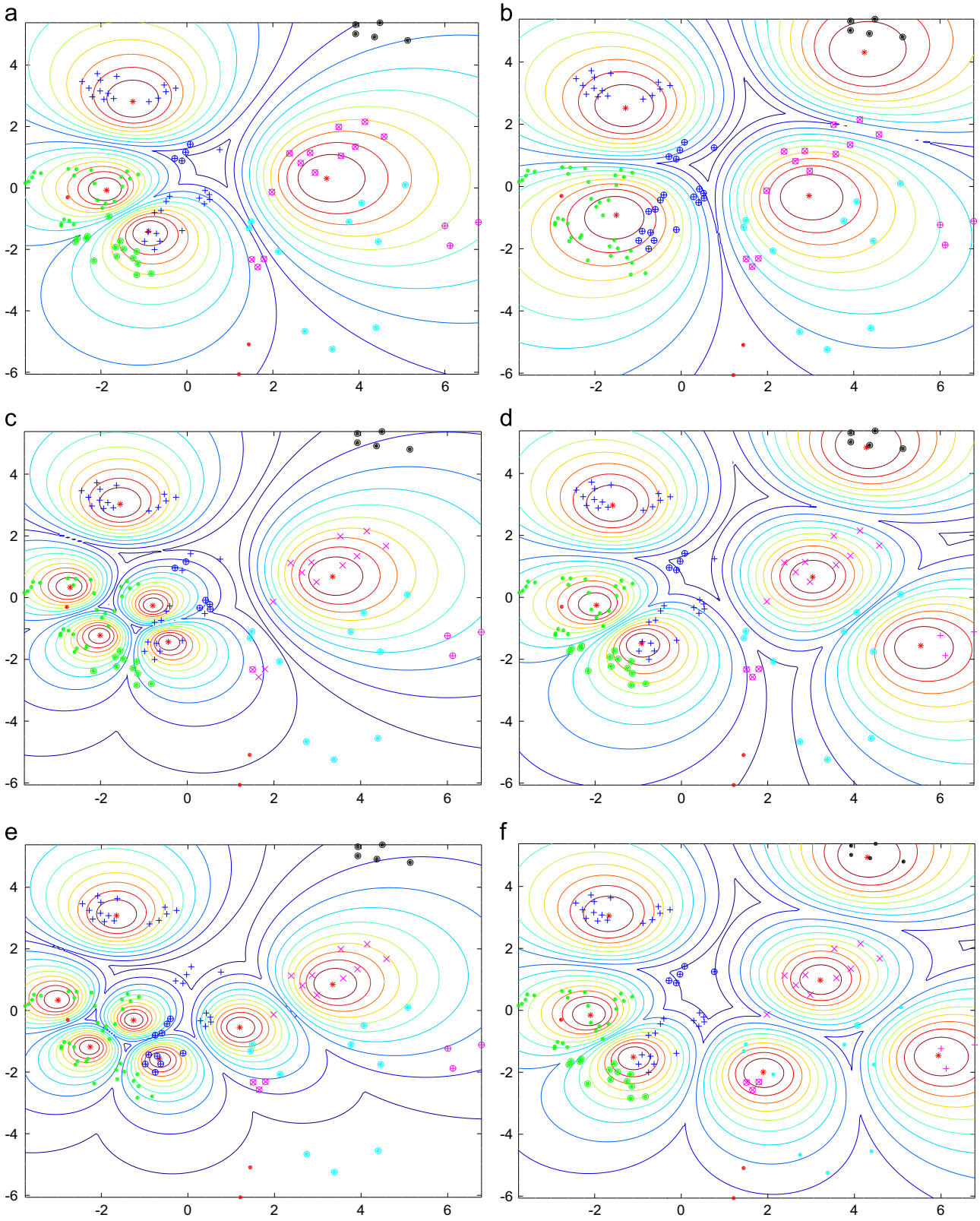

Fig. 9. Sammom projections showing more detailed analysis in different regions of the feature space. Left side subfigures correspond to **P** positioned in the region where features attain their minimum values while right side subfigures correspond to **P** positioned where features attain their max. In each horizontal subfigure pair the same number of clusters is used, i.e., (a) and (b) $c=4$; (c) and (d) $c=6$; and $c=7$ in (e) and (f).

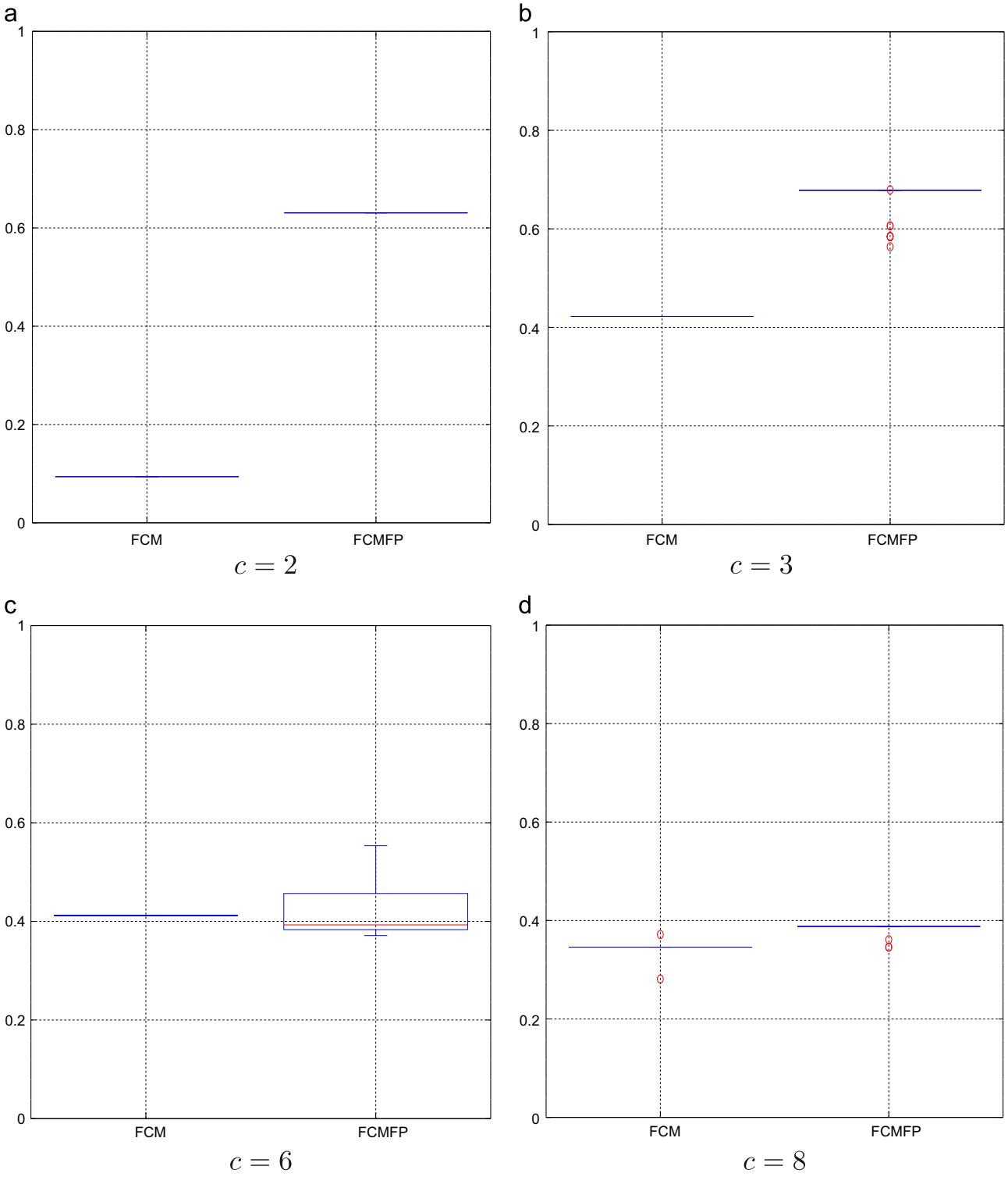

Fig. 10. Boxplots comparing the distributions of Adjusted Rand Indices over 30 independent runs of each algorithm, for different number of clusters c: FCMFP always outperforms FCM (with $p_{\text{Wilcoxon}} = 0.000002$) except for $c=6$ where there is no statistical significant difference between algorithms ($p_{\text{Wilcoxon}} = 0.308615$). Similar results were obtained for the Rand index.

algorithm allows for a detailed analysis of a given region of the feature space.

FCMFP is inspired in the observation that the visual perception of a group of similar objects is (highly) dependent on the observer's position. The position at which the observer is located relatively to a set of objects determines how the observer perceives these objects. If the observer is located at a significant distance from the objects these tend to be undistinguishable, that is, objects tend to be seen as a single cluster. As the observer gets closer to objects differences between them tend to emerge, and the initial single cluster tends to split into more and more clusters.

The included experimental results show that FCMFP is able to provide a number of reasonable partitions at different level of granularity. Reasonable partitions (or hypotheses on data structure) are those successfully evaluated by an internal validation index, the inverted Xie-Beni index in this case. To this end FCMFP

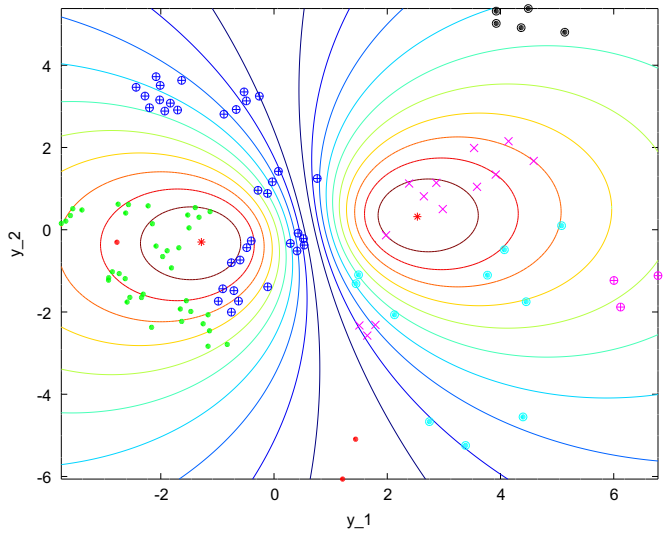

Fig. 11. A Sammom projection of the 12-dimensional feature space into the plane for the fault detection case ($c=2$) when the FCM algorithm is used. Most samples in the fault state $P3$ (blue $\oplus$) are now wrongly clustered together with samples in healthy state (a green $\bullet$). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

shares the advantages of hierarchical clustering while mitigates its limitations in terms of efficiency and flexibility. Moreover, results also show that the biased algorithm (FCMFP) produces partitions exhibiting better external validity index values than the corresponding unbiased algorithm (FCM) for any number of clusters except for $c=6$ where no statistically significant difference was observed between algorithms. Although this type of improvements was *not* the primary goal beyond FCMFP, the improvement is viewed as a convenient side effect of the shrinkage technique employed in the design of the algorithm. From this perspective, one can claim that works currently employing FCM in bearing fault diagnosis would benefit from the use of FCMFP.

Finally, we must say that an effort was made to test the proposed technique under realistic experimental conditions. In industrial applications, high energy noise can corrupt the measurements and suitable vibration signal conditioning (filtering) should be used. In this study, no specific signal conditioning was applied even though there is still noise and interferences resulting from the different components of the test rig such as flywheels, motor, and couplings (see Fig. 2). Hence the proposed technique was tested in the presence of noise corrupted signals as it was illustrated in Fig. 3. Moreover, the vibration signal-to-noise ratio (SNR) is directly related with the size of the smallest detectable fault. The stronger the SNR the smaller the detectable fault. In the noise free theoretical limit, the smallest fault size is determined by the sensitivity of the accelerometer.

Acknowledgments

The work was sponsored in part by the Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation (SENESCYT) of the Republic of Ecuador, by the Chongqing Technology and Business University (CTBU) open Grant number: 1456027, and by the project UID/MULTI/00631/2013 - CEOT.

The experimental work was developed at the GIDTEC research group lab of the Universidad Politécnica Salesiana de Cuenca, Ecuador.

Appendix A

Algorithm 1. Fuzzy C-means with focal point (Fazendeiro and Valente de Oliveira, 2015).

Let $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}\subset\mathbb{R}^d$ be a finite set of unlabeled data.
Initialize the clusters' prototypes $\mathbf{V}\in\mathbb{R}^{d\times c}$.
Set C_{max}, $m>1$, $\mathbf{P}\in\mathbb{R}^w(w\geq d)$ and $\zeta\geq 0$.
Extend $\mathbf{X}$ and $\mathbf{V}$ into $\mathbb{R}^w$ by introducing $(w-d)$ null coordinates per element.
Repeat the following steps until a termination criterion has been met.
Step 1 - For $i=1,2,\ldots,C_{max}$ and $j=1,2,\ldots,n$, update the partition matrix, $\mathbf{U}$, according to (6).
Step 2 - For $i=1,2,\ldots,C_{max}$, update the prototypes, $\mathbf{V}$, according to (7).
Project the prototypes into the original feature space $\mathbb{R}^d$ by computing the intersection of the lines defined by the focal point $\mathbf{P}$ and each $\mathbf{v}_i$ cluster, with the original data space.

Algorithm 2. Iterative fuzzy C-means with focal point (Fazendeiro and Valente de Oliveira, 2015).

Set c', C_{max}, $(1<c'<C_{max})$, $\zeta\geq 0$, and $\Delta\zeta>0$.
Repeat the following steps until the number of candidate clusters is smaller than c'.
Step 1 - Apply the FCMFP algorithm.
Step 2 - Remove neglectable clusters (clusters without any typical datum).
Step 3 - Compute the validity measure for the remaining candidate clusters.
Step 4 - Update $\zeta\leftarrow\zeta+\Delta\zeta$.
Output the partition(s) optimizing the considered validity measure.

A.1. The Rand index and the Adjusted Rand Index

Let a (data) set of n elements be denoted by $X=\{x_1,\ldots,x_n\}$. A disjoint partition of X, denoted as $\mathcal{A}=\Pi(X)$ is s.t. $\mathcal{A}=\{A_i|A_i\subset X, \forall_{i\in I}\}$, and satisfies the following conditions:

(i) $\varnothing\subset A_i\subset X$, i.e., the i-th cluster A_i cannot be empty and is smaller than X;
(ii) $\cup_{i\in I}A_i=X$, i.e., together all clusters fill X;
(iii) $A_i\cap A_j=\varnothing$ for $i\neq j$, i.e., clusters do not overlap.

Given two partition of X, say the ground truth partition, $\mathcal{G}=\Pi(X)$, and a hypothesis partition $\mathcal{H}=\Pi'(X)$ generated by a clustering algorithm, the Rand Index (RI) of these partitions measure how similar $\mathcal{G},\mathcal{H}$ are and is given by Hubert Lawrence (1985):

$$\mathrm{RI}(\mathcal{G},\mathcal{H})=\frac{n_{00}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}=\frac{n_{00}+n_{11}}{n_t} \tag{15}$$

where n_{11} is the number of pairs of elements of X that cluster together in both partitions, n_{00} is the number of pairs that are separated in both partitions, n_{01} number of pairs that cluster together in $\mathcal{G}$ but not in $\mathcal{H}$; and n_{10} is the number of pairs that cluster together in $\mathcal{H}$ but not in $\mathcal{G}$; the total number of pairs n_t is $n_t=\binom{n}{2}=n(n-1)/2$. The index takes values in the unit interval, with 1 meaning that both partitions completely agree. However the RI of two random partitions is not a constant value. This drawback is solved using the Adjusted Rand Index

(*ARI*) (Hubert Lawrence, 1985):

$$\mathrm{ARI}(\mathcal{G},\mathcal{H}) = \frac{\sum_{i,j}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}/\binom{n}{2}\right]}{\frac{1}{2}\left[\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}\right] - \left[\sum_i \binom{n_{i.}}{2}\sum_j \binom{n_{.j}}{2}/\binom{n}{2}\right]} \tag{16}$$

where n is the number of elements of X (the data set), n_{ij} is the number of element that cluster together in G_i and H_j, $n_{i.}$ and $n_{.j}$ are the number of elements in G_i and H_j, respectively. Notice that *ARI* attains also its maximum value at 1 when the two partitions completely agree. In general ARI $\leq$ RI.

References

Abed, W., Sharma, S., Sutton, R., 2014. Diagnosis of bearing fault of brushless DC motor based on dynamic neural network and orthogonal fuzzy neighborhood discriminant analysis. In: 2014 UKACC International Conference on Control (CONTROL), pp. 378–383.

Ben Ali, J., Saidi, L., Mouelhi, A., Chebel-Morello, B., Fnaiech, F., 2015. Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. Eng. Appl. Artif. Intell. 42, 67–81.

Cao, S.-Q., Zuo, X.-M., Tao, A.-X., Wang, J.-M., Chen, X.-Z., 2012. A bearing intelligent fault diagnosis method based on cluster analysis. Appl. Mech. Mater. 152–154, 1628–1633.

Chen, X., Zeng, H., Li, Z., 2010. A multi-fault diagnosis method of rolling bearing based on wavelet-PCA and fuzzy K-nearest neighbor. Appl. Mech. Mater. 29–32, 1602–1607.

Cui, B., Wang, Z., Pan, H., 2008. Application of wavelet analysis-cluster fuzzy to fault diagnosis of roller bearings. Zhendong Ceshi Yu Zhenduan/J. Vibr. Meas. Diagn. 28 (2), 151–154.

Dieck, R.H., 2006. Measurement Uncertainty: Methods and Applications. ISA.

Fazendeiro, P., Valente de Oliveira, J., 2008. Fuzzy clustering as a data-driven development environment for information granules. In: Pedrycz, W., Skowron, A., Kreinovich, V. (Eds.), Handbook of Granular Computing, pp. 153–169.

Fazendeiro, P., Valente de Oliveira, J., 2015. Observer-biased fuzzy clustering. IEEE Trans. Fuzzy Syst. 23 (1), 85–97.

Fu, P., Li, W.L., Cao, W.Q., 2011. Fault diagnosis of bearings based on time-delayed correlation demodulation and fuzzy clustering. Adv. Mater. Res. 211–212, 510–514.

Robin, Genuer, Jean, Poggi, Christine, Tuleau-Malot, 2010. Variable selection using Random Forests. Pattern Recognit. Lett. 14 (31), 2225–2236.

Guan, H., Dong, Y., Sun, J., Liu, K., 2006. Diagnosis system of rolling bearings for freight carriage based on C-means algorithm. Zhendong Ceshi Yu Zhenduan/J. Vibr. Meas. Diagn. 26 (2), 138–141.

Gupta, K.N., 1997. Vibration—A tool for machine diagnostics and condition monitoring. Sadhana 22 (3).

Lawrence, Hubert, Phipps, Arabie, 1985. Comparing partitions. J. Classif. 2 (1), 193–218.

Jia, J., Kong, F., Liu, Y., Wang, J., Liu, W., Chen, J., 2005. Noise diagnosis research based on wavelet packet and fuzzy C-clusters about connecting rod bearing fault. Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach. 36 (6), 87–91.

Jiang, L., Liu, Y., Li, X., Chen, A., 2011. Degradation assessment and fault diagnosis for roller bearing based on AR model and fuzzy cluster analysis. Shock Vibr. 18 (1–2), 127–137.

Jiang, L., Liu, Y., Li, X., Chen, A., 2011. Fault diagnosis of roller bearing based on EMD and fuzzy cluster. Jixie Qiangdu/J. Mech. Strength 33 (5), 650–654.

Jiang, L.L., Liu, Y.L., Li, X.J., Chen, A., 2010. Fault diagnosis of roller bearing based on bispectrum estimation and fuzzy cluster analysis. Appl. Mech. Mater. 36, 129–134.

Lei, Y., He, Z., Zi, Y., Hu, Q., 2007. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. Mech. Syst. Signal Process. 21 (5), 2280–2294.

Chuan, Li, Ming, Liang, 2012. Continuous-scale mathematical morphology-based optimal scale band demodulation of impulsive feature for bearing defect diagnosis. J. Sound Vibr. 331 (26), 5864–5879.

Chuan, Li, Ming, Liang, Tianyang, Wang, 2015. Criterion fusion for spectral segmentation and its application to optimal demodulation of bearing vibration signals. Mech. Syst. Signal Process. 64–65, 132–148.

Chuan, Li, Ming, Liang, Yi, Zhangsnm, Shumin, Hou, 2012. Multi-scale autocorrelation via morphological wavelet slices for rolling element bearing fault diagnosis. Mech. Syst. Signal Process. 31, 428–446.

Chuan, Li, Vinicio, Sanchez, Grover, Zurita, Cerrada, Lozada Mariela, Diego, Cabrera, 2016. Rolling element bearing defect detection using the generalized synchrosqueezing transform guided by time-frequency ridge enhancement. ISA Trans. 60, 274–284.

Liang, Y., Jia, L.M., Cai, G.Q., Liu, J.Z., 2014. A new approach to diagnose rolling bearing faults based on AFD, Lecture Notes in Electrical Engineering 288 LNEE, vol. 2, pp. 573–582.

Liu, C.L., Huang, X.M., Luo, X.J., 2014. Roller bearing fault diagnosis based on ELMD and fuzzy C-means clustering algorithm. Appl. Mech. Mater. 602–605, 1698–1700.

Liu, W.Y., Han, J.G., 2012. A fuzzy clustering-based binary threshold bispectrum estimation approach. Neural Comput. Appl. 21 (1).

Liu, W.Y., Han, J.G., 2013. Rolling element bearing fault recognition approach based on fuzzy clustering bispectrum estimation. Shock Vibr. 20 (2), 213–225.

Liu, X., Ma, L., Zhang, S., Mathew, J., 2008. Feature group optimisation for machinery fault diagnosis based on fuzzy measures. Aust. J. Mech. Eng. 5 (2), 191–197.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Le Cam, L.M., Neyman, J. (Eds.), Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297.

Meng, Z., Wang, Y., Wang, X., 2014. Fault diagnosis of rolling bearings based on DLMD sample entropy and fuzzy clustering (in Chinese). Zhongguo Jixie Gongcheng/China Mech. Eng. 25 (19), 2634–2641.

Navarro, L., Delgado, M., Urresty, J., Cusido, J., Romeral, L., 2010. Condition monitoring system for characterization of electric motor ball bearings with distributed fault using fuzzy inference tools. In: 2010 IEEE of Instrumentation and Measurement Technology Conference (I2MTC), pp. 1159–1163.

Oh, Hyunseok, Shibutani, Tadahiro, Pecht, Michael, 2012. Precursor monitoring approach for reliability assessment of cooling fans. J. Intell. Manuf. 23 (2), 173–178.

Oskouei, A.R., Esmaeili, M., 2012. Neuron-fuzzy-based acoustic emission output parameters prediction for rotary device monitoring. In: Proceedings of ASME International Mechanical Engineering Congress and Exposition (IMECE), vol. 12. pp. 437–442.

Ou, L., Yu, D., 2014. Rolling bearing fault diagnosis based on Laplacian score and fuzzy C-means clustering. Zhongguo Jixie Gongcheng/China Mech. Eng. 10, 1352–1357.

Pan, Y.-N., Chen, J., Li, X.-L., 2009. Fuzzy c-means based equipment performance degradation assessment. Shanghai Jiaotong Daxue Xuebao/J. Shanghai Jiaotong Univ. 43 (11), 1794–1797.

Pan, Y.N., Chen, J., Dong, G.M., 2009. A hybrid model for bearing performance degradation assessment based on support vector data description and fuzzy c-means. Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. 223 (11), 2687–2695.

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Trans. Comput. C-18 (5), 401–409.

Sheskin, D.J., 2011. Handbook of Parametric and Nonparametric Statistical Procedures, fifth ed. Taylor & Francis, Boca Raton.

Siyambalapitiya, D.J.Tilak., McLaren, P.G., 1990. Reliability improvement and economic benefits of online monitoring systems for large induction machines. IEEE Trans. Ind. Appl. 26 (6), 1018–1025.

Sreenilayam-Raveendran, Ranjith-Kumar, Azarian Michael, H., Morillo, C., Pecht Michael, G., Kida, Katsuyuki, Santos, E.C., Honda, Takashi, Koike, Hitonobu, 2013. Comparative evaluation of metal and polymer ball bearings. Wear 302 (1–2), 1499–1505.

Sui W., Lu, C., Zhang, D., 2008. Bearing fault diagnosis based on feature weighted FCM cluster analysis. In: Proceedings of International Conference on Computer Science and Software Engineering (CSSE 2008), vol. 5, pp. 518–521.

Sui, W., Lu, C., Zhang, D., 2010. Bearings fault diagnosis based on feature weighted FCM algorithm. Wuhan Ligong Daxue Xuebao (Jiaotong Kexue Yu Gongcheng Ban)/J. Wuhan Univ. Technol. (Transp. Sci. Eng.) 34 (1), 72–75.

Valente de Oliveira, José, Pedrycz, Witold, 2007. Advances in Fuzzy Clustering and Its Applications. John Wiley & Sons, Inc., New York, NY, USA.

Vijay, G.S., Pai, S.P., Sriram, N.S., Rao, R.B.K.N., 2013. Radial basis function neural network based comparison of dimensionality reduction techniques for effective bearing diagnostics. Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol. 227 (6), 640–653.

Wadhwani, S., Wadhwani, A.K., Gupta, S.P., Kumar, V., 2006. Detection of bearing failure in rotating machine using adaptive neuro-fuzzy inference system. In: 2006 International Conference on Power Electronics, Drives and Energy Systems (PEDES '06).

Wang, C.J., Li, H.Y., Xiang, W., Zhao, D., 2014. A new signal classification method based on EEMD and FCM and its application in bearing fault diagnosis. Appl. Mech. Mater. 602–605, 1803–1806.

Wang, S., Zhang, J., Zhang, S., Liu, Y., 2012. Fault diagnosis of rolling bearings based on Weibull distribution and fuzzy C means clustering analysis. Zhongguo Jixie Gongcheng/China Mech. Eng. 23 (5), 595–599.

Wang, S., Zhang, J., Li, Y., Zhang, S., 2012. Rotating machinery fault diagnosis based on mathematical morphology and fuzzy clustering (in Chinese). Yi Qi Yi Biao Xue Bao/Chin. J. Sci. Instrum. 33 (5), 1055–1061.

Tianyang, Wang, Ming, Liang, Jianyong, Li, Weidong, Cheng, Chuan, Li, 2015. Bearing fault diagnosis under unknown variable speed via gear noise cancellation and rotational order sideband identification. Mech. Syst. Signal Process. 62–63, 30–53.

Xinbin, L., Xi, S., Qiang, C.Y., 2012. Bearing fault diagnosis based on multiple classifiers group of Fuzzy C Means. In: Chinese Control Conference (CCC), pp. 5254–5259.

Xu, H.-B., Chen, G.-H., Wang, X.-H., 2012. Fault identification of bearings based on bispectrum distribution of ARMA model and FCM method in Chinese. Huanan Ligong Daxue Xuebao/J. South China Univ. Technol. (Nat. Sci.) 40 (7), 78–82+89.

Xu, Z., Xuan, J., Shi, T., Wu, B., Hu, Y., 2009. A novel fault diagnosis method using PCA and art-similarity classifier based on Yu's norm. Key Eng. Mater. 413–414, 569–574.
Ruqiang, Yan, Gao Robert, X., Xuefeng, Chen, 2014. Wavelets for fault diagnosis of rotary machines: a review with applications. Signal Process. 96, 1–15.
Yaqub, M.F., Gondal, I., Kamruzzaman, J., 2012. Inchoate fault detection framework: adaptive selection of wavelet nodes and cumulant orders. IEEE Trans. Instrum. Meas. 61 (3), 685–695.
Ye, Z., Guoqiang, C., Yang, J., Zhang, J., Shi, Q., Zeng, H., 2011. Fault diagnosis of railway rolling bearing based on wavelet analysis and FCM. Int. J. Digit. Content Technol. Appl. 5 (3), 47–58.
Zanoli, S.M., Astolfi, G., 2012. Faults diagnosis for a centrifugal machine using the Mahalanobis distance. In: IFAC Proceedings Volumes, vol. 8, PART 1, pp. 444–449.
Zhang, J., Ma, W., Ma, L., He, Z., 2013. Fault diagnosis model based on fuzzy support vector machine combined with weighted fuzzy clustering. Trans. Tianjin Univ. 19 (3), 174–181.
Zhang, L., Li, P., Li, M., Zhang, S., Zhang, Z., 2014. Fault diagnosis of rolling bearing based on ITD fuzzy entropy and GG clustering (in Chinese). Yi Qi Yi Biao Xue Bao/Chin. J. Sci. Instrum. 35 (11), 2624–2632.
Zhang, L.-L., Liao, H.-Y., Cao, Y.-J., Luo, S.-D., Zhao, Y.-G., 2011. Diagnosis on crankshaft bearing fault based on EEMD and fuzzy C mean clustering arithmetic. Neiranji Xuebao/Trans. CSICE (Chin. Soc. Intern. Combust. Engines) 29 (4), 332–336.
Zhang, X.-P., Zhang, X.-Y., Liu, J., 2014. Fuzzy kernel-clustering algorithm based on differential evolution algorithm and its application in fault diagnosis. Dianli Xitong Baohu yu Kongzhi/Power Syst. Prot. Control 42 (17), 102–106.
Zhang, Y., Liu, X.-D., Xie, F.-D., Li, K.-Q., 2009. Fault classifier of rotating machinery based on weighted support vector data description. Expert Syst. Appl. 36 (4), 7928–7932.
Zhao Xinze, Zhao Chunhua, Gao Hongliang, Wu Gang, 2008. Knowledge mining for fault diagnosis based on rough sets theory. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008), pp. 744–749.
Zheng, Z., Jiang, W.-L., Hu, H.-S., Zhu, Y., Li, Y., 2015. Research on rolling bearings fault diagnosis method based on EEMD morphological spectrum and kernel fuzzy C-means clustering. Zhendong Gongcheng Xuebao/J. Vibr. Eng. 28 (2), 324–330.