A Novel Ensemble Clustering for Operational Transients Classification with Application to a Nuclear Power Plant Turbine

Sameer Al-Dahidi, Francesco Di Maio, Piero Baraldi, Enrico Zio, Redouane Seraoui

A NOVEL ENSEMBLE CLUSTERING FOR OPERATIONAL TRANSIENTS CLASSIFICATION WITH APPLICATION TO A NUCLEAR POWER PLANT TURBINE

Sameer Al-Dahidi[1], Francesco Di Maio[1*], Piero Baraldi[1], Enrico Zio[1,2], and Redouane Seraoui[3]

[1] Energy Department, Politecnico di Milano, Milan, 20133, Italy
[2] Chair on Systems Science and the Energetic challenge, Foundation EDF, Centrale Supélec, Paris, 92295, France
[3]EDF-R&D\STEP Simulation et Traitement de l'information pour l'exploitation des systèmes de production, Chatou 78401, France

Abstract. The objective of the present work is to develop a novel approach for combining in an ensemble multiple base clusterings of operational transients of industrial equipment, when the number of clusters in the final consensus clustering is unknown. A measure of pairwise similarity is used to quantify the co-association matrix that describes the similarity among the different base clusterings. Then, a Spectral Clustering technique of literature, embedding the unsupervised *K*-Means algorithm, is applied to the co-association matrix for finding the optimum number of clusters of the final consensus clustering, based on Silhouette validity index calculation. The proposed approach is developed with reference to an artificial case study, properly designed to mimic the signal trend behavior of a Nuclear Power Plant (*NPP*) turbine during shut-down. The results of the artificial case have been compared with those achieved by a state-of-art approach, known as Cluster-based Similarity Partitioning and Serial Graph Partitioning and Fill-reducing Matrix Ordering Algorithms (*CSPA-METIS*). The comparison shows that the proposed approach is able to identify a final consensus clustering that classifies the transients with better accuracy and robustness compared to the *CSPA-METIS* approach. The approach is, then, validated on an industrial case concerning 149 shut-down transients of a *NPP* turbine.

***Keywords*:** Unsupervised Learning, Ensemble Clustering, Final Consensus Clustering, Spectral Clustering, Operational Transients, Nuclear Power Plant (*NPP*) turbine shut-down.

1. Introduction

In industries such as nuclear, oil and gas, automotive and chemical, equipments are subjected to several causes of performance degradation and exposed to faulty conditions, e.g., presence of manufacturing defects, unexpected interactions with the environment, wear and tear (Bolotin & Shipkov, 1998; Muller, Suhner, & Iung, 2008; Baraldi, Di Maio, & Zio, 2012; Baraldi, Di Maio, & Zio, 2013c). Capturing the different operational conditions of these equipments, detecting the onset of abnormal conditions and classifying them in different types can aid the decision maker to decide a proper maintenance intervention policy and, hence, increase equipment reliability and system safety while reducing overall corrective maintenance costs (Jardine, Lin, & Banjevic, 2006; Al-Dahidi, Baraldi, Di Maio, & Zio, 2014).

Measurements of relevant signals are collected during operation. These transient data are representative of different operational conditions of the equipment. For fault diagnosis, these data are manipulated with the objective of partitioning them into dissimilar groups, whose number is "a priori" unknown, such that data belonging to the same group are more similar than those belonging to the other groups, and corresponding to different equipment conditions. In particular, one can distinguish, among the groups, anomalous behaviors of the equipment and relate them to specific root causes (Fred & Jain, 2005; Xiufeng & Changzheng, 2010; Wu & Lee, 2011; Serir, Ramasso, & Zerhouni, 2012; Baraldi, Di Maio, Zio, Rigamonti, & Seraoui, 2013a; Serir, Ramasso, Nectoux, & Zerhouni, 2013).

The problem of grouping the operational transients of an industrial equipment can be formulated as an unsupervised clustering problem aimed at partitioning the transient data into homogeneous clusters so that those data belonging to the same cluster are very similar to each other and dissimilar to those of the other clusters (Salvador, 2002; Bocaniala, Sa Da Costa, & Palade, 2004; Zhou, Zhang, & Wang, 2004; Chaovalit & Zhou, 2005; Wang, Yu, Siegel, & Lee, 2008; Wang, 2010; Baraldi et al. 2013a; Lin, Chen, & Zhou, 2013).

Over the last few decades, several clustering algorithms have been proposed and used in practice, like *K*-Means (Hartigan, 1975; Vlachos, Lin, Eamonn, & Dimitrios, 2003; Siegel & Lee, 2011), Self-Organizing Maps (*SOM*) (Bhavaraju, Kankar, Sharma, & Harsha, 2010; Gonçalves, Bosa, Balen, Lubaszewski, Schneider, & Henriques, 2011; Al-Dahidi, 2014), Fuzzy *C*-Means (*FCM*) (Bezdek, 1981; Leguizamón, Pelgrum, & Azzali, 1996; Baraldi et al. 2012; Di Maio, Hu, Tse, Pecht, Tsui, & Zio, 2012; Baraldi et al. 2013c), Spectral Clustering (Von Luxburg, 2007; Zhao &

Liu, 2007; Baraldi, Di Maio, Zio, Rigamonti, & Seraoui, 2013b), Hierarchical clustering (Johnson, 1967; Van Wijk & Van Selow, 1999; Datta, Mavroidis, & Hosek, 2007), and Hidden Markov Models (*HMMs*) (Baruah & Chinnam, 2005). However, there is no unique clustering algorithm capable of correctly identifying the underlying structure of any kind of dataset. Even the application of different clustering algorithms to the same set of data, or the same algorithm with different parameter settings leads to different clustering results (Fred & Jain, 2005; Fern & Lin, 2008; Vega-Pons & Ruiz-Shulcloper, 2011).

To handle this, ensemble approaches have been proposed that combine multiple base clusterings into a single consolidated clustering, i.e., the final consensus clustering P^* (Strehl & Ghosh, 2002; Topchy, Jain, & Punch, 2004; Topchy, Jain, & Punch, 2005; Chen, 2007; Vega-Pons & Ruiz-Shulcloper, 2011; Iqbal, Moh'd, & Khan, 2012).

A typical ensemble clustering scheme is shown in Figure 1. For a given dataset $\overline{\overline{X}}$, the construction of the ensemble amounts to the aggregation of the results of multiple base clusterings. The base clusterings composing the ensemble can be different because of the different algorithms used and/or because of the different data and features upon which clustering is performed. The outcome of the multiple base clusterings are aggregated into a final consensus clustering P^*, by a given method of aggregation (Strehl & Ghosh, 2002; Topchy et al. 2004; Chen, 2007; Greene & Cunningham, 2007; Vega-Pons & Ruiz-Shulcloper, 2011; Ahuja & Dhanya, 2012).

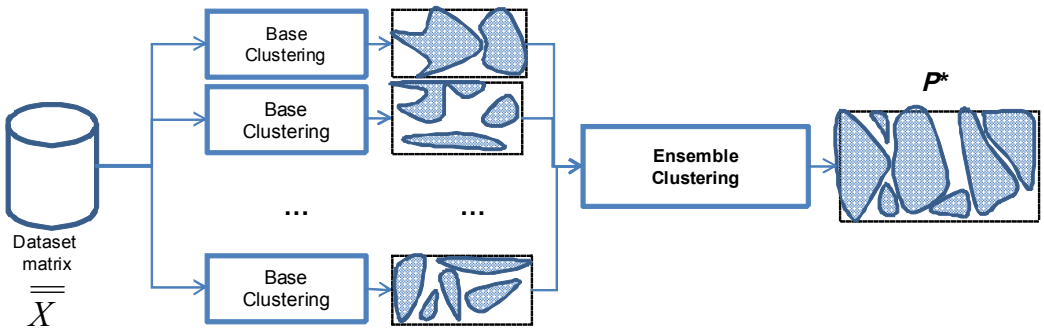

Figure 1: Scheme of ensemble clustering approach.

The main challenges for an effective consensus strategy of aggregation are (Topchy et al. 2004): 1) different base clusterings group data differently and, maybe, in different numbers of clusters, 2) the correspondence between the clusters labels of different base clusterings is unknown, 3) the number of clusters M in the final consensus clustering is "a priori" unknown, 4) some base clusterings might not label some data (missing labels), and 5) for large datasets, large computational times might be needed.

Several methods have been used to obtain the final consensus clustering, for example Relabeling and Voting (Ayad & Kamel, 2010), Co-association Matrix (Vega-Pons & Ruiz-Shulcloper, 2011), Genetic Algorithms (Ghaemi, bin Sulaiman, Ibrahim, & Mustapha, 2011; Chatterjee & Mukhopadhyay, 2013), Finite Mixture Models (Topchy et al. 2004; Topchy et al. 2005), and Graph and Hypergraph partitioning (Karypis, Aggarwal, Kumar, & Shekhar, 1997; Strehl & Ghosh, 2002; Vega-Pons & Ruiz-Shulcloper, 2011). The success of these consensus strategies in addressing the above mentioned challenges is reported in Table 1.

Table 1: Capabilities of ensemble clustering approaches (√ solved, X unsolved).

Ensemble clustering approach	Label correspondence problem	Different number of clusters for each base clustering	"A priori" knowledge of *M*	Missing labels	Computational limitations
Relabeling and Voting	√	X	X	X	No
Co-association matrix	X	√	X	√	Yes
Genetic algorithm	√	√	X	X	Yes
Finite Mixture Models	X	√	X	√	No
Graph and Hypergraph partitioning	X	√	X	√	Yes

The Relabeling and Voting method solves the correspondence between the labels provided by different base clusterings, even for large datasets, by using a simple voting procedure to partition data in clusters (Dimitriadou, Weingessel, & Homik, 2001; Dudoit & Fridlyand, 2003), but it requires the number of clusters in the base clusterings to be the same and known "a priori" (Ghaemi, Sulaiman, Ibrahim, & Mustapha, 2009).

Co-association based methods summarize similarities among base clusterings into a co-association matrix (Strehl & Ghosh, 2002), even for different numbers of clusters for the base clusterings, without any previous knowledge on *M*, but with high computational demands (Fred & Jain, 2005; Vega-Pons & Ruiz-Shulcloper, 2011).

In genetic algorithm-based methods, the search capability of genetic algorithms is used to identify the most stable clusters once the label correspondence problem is solved (Ghaemi et al. 2009). The plus of the method is its ability to identify clusters that are not easily found by other methods, even for different numbers of clusters for each base clustering; on the other hand, its computational burden, and its inability to deal with the missing labels constitute practical limitations (Topchy et al. 2004; Vega-Pons & Ruiz-Shulcloper, 2011).

In Finite Mixture Models, the final consensus clustering is seen as a probability model in the space of the base clusters and is found as a solution to the maximum likelihood problem for a given ensemble clustering (Topchy et al. 2004; Di Maio, Nicola, Zio, & Yu, 2014). The method does not solve the label correspondence problem, it is able to handle missing labels, it deals with different numbers of clusters for each base clustering and does not need any previous knowledge on M (Figueiredo & Jain, 2002), but its computational burden due to the estimation of the covariance matrices, makes the method difficult to apply in practice.

Graph and Hypergraph partitioning algorithms, such as the Cluster-based Similarity Partitioning (*CSPA*), construct a graph from the similarities among the base clusterings, and cluster it using a graphic-based clustering algorithm, such as Serial Graph Partitioning and Fill-reducing Matrix Ordering Algorithm (*METIS*) (Karypis & Kumar, 1995; Karypis & Kumar, 1998; Strehl & Ghosh, 2002; Topchy et al. 2004), for a predetermined value of M (Topchy et al. 2004; Ghaemi et al. 2009). The method does not solve the correspondence between the base clusterings labels, can handle the missing labels and different numbers of clusters for each base clustering, but it suffers computation limitations for large datasets. Despite this, *CSPA* and *METIS* algorithms have been taken as reference for comparison in this paper because *CSPA-METIS* is the simplest and "often" best performing method for consensus aggregation among other Graph and Hypergraph partitioning algorithms, e.g., Meta-CLustering Algorithm (*MCLA*) and HyperGraph-Partitioning Algorithm (*HGPA*) (Strehl & Ghosh, 2002; Chen, 2007), whose pitfall is that the number of final consensus clusters cannot exceed the maximum number of the individual base clusters.

The novelty of the proposed approach is to replace *METIS* algorithm with Spectral Clustering (Von Luxburg, 2007; Baraldi et al. 2013b) and Silhouette validity index (Rousseeuw, 1987), to automatically determine M which by most industrial applications, is not known "a priori" (Chakaravathy & Ghosh, 1996; Strehl & Ghosh, 2002; Li & Chen, 2011). More specifically, the Spectral Clustering technique, embedding the unsupervised K-Means algorithm, is applied to the co-association matrix that describes the similarity among the different base clusterings obtained on a set of diverse sources of data (features) (e.g., vibration, temperature signals), rather than to the similarity values among the data themselves, for mining the clusters that are formed by the most similar data. Then, the optimum number of clusters C^* is selected among several candidates $C_{candidate}$, based on the morphology of the obtained final consensus clusters evaluated by the Silhouette validity index that measures the similarity of the data belonging to the same cluster and the dissimilarity of these in the other clusters (a large Silhouette value indicates that the obtained

clusters of the final consensus clustering are well separated and compacted (Rousseeuw, 1987; Charrad, Lechevallier, Ahmed, & Saporta, 2010).

The proposed approach is developed on an artificial case study properly designed to mimic the signal trend behavior of Nuclear Power Plants (*NPPs*) turbines during shut-down transients. Different sets of features have been simulated and used to obtain different base clusterings, representative of different groupings of the shut-down transients of the turbine. The correct number of clusters, for each base clustering, has been identified by the Davies-Bouldin (*DB*) criterion: the minimum *DB* value is reached for the number of clusters which gives optimal separation and compactness (Davies & Bouldin, 1979). Three controlled datasets containing *M* sparse or overlapping clusters of their base clusterings results have been considered. The results obtained have been compared with those achieved by *CSPA-METIS*. It has been found that the proposed approach is able to identify the final consensus clustering with better accuracy and robustness compared to the *CSPA-METIS* approach.

The approach is, then, applied to a real industrial case concerning 149 shut-down transients of a *NPP* turbine: different base clusterings representative of different groupings of the shut-down transients of the turbine are obtained by using multiple different sources of data (features), i.e., vibration, turbine shaft speed, vacuum, and temperature signals, and a final consensus clustering is obtained that gives the optimal grouping of the shut-down transients of the *NPP* turbine, in terms of groups separation and compactness.

The remainder of the paper is organized as follows. In Section 2, the basics of *CSPA-METIS* ensemble approach are recalled. In Section 3, the proposed ensemble clustering approach is presented. The artificial case study representative of the signal trend behavior of a Nuclear Power Plant (*NPP*) turbine during shut-down transients is introduced in Section 4. Furthermore, the results obtained with the application of the proposed approach to the artificial case and the comparison with *CSPA-METIS*, are discussed. Section 5 verifies the robustness of the proposed approach to clustering overlapping, in identifying the number *M* for three controlled datasets containing sparse or overlapping clusters of their base clusterings results. The real industrial case concerning 149 shut-down transients of a *NPP* turbine is introduced in Section 6 and the results of the application of the proposed approach to the case study are discussed. Finally, Section 7 concludes the paper with some considerations.

2. The *CSPA-METIS* ensemble clustering approach

In this Section, the combination of *CSPA* and *METIS* is described and considered as reference ensemble clustering approach, for the case when the number *M* of clusters in the final consensus clustering is known.

The flowchart for the method is sketched in Figure 2. The algorithm goes along the following two phases: a procedure (i.e., *CSPA*) for establishing a co-association matrix and a procedure (i.e., *METIS*) for partitioning the graph obtained from the co-association matrix to obtain the final consensus clustering *P** (Strehl & Ghosh, 2002; Topchy et al. 2004).

We consider *N* data belonging to the dataset $\overline{\overline{X}}$ that are clustered into *H* base clusterings. For each *j*-th base clustering, *j*=1,…,*H*, each datum is labeled by an integer number ranging in $[1,C_{opt}^{j}]$, where C_{opt}^{j} is the number of clusters for each *j*-th base clustering. The problem of clustering the *N* data is, thus, transformed into an aggregation problem of the base clusterings outcomes $\overline{\overline{Y}}$ of size *N*x*H*.

The algorithm entails three main steps; without loss of generality, these are hereafter described on a simple numerical example where $\overline{\overline{X}}$ contains *N*=5 data, clustered into *H*=3 base clusterings:

Step 1: *Adjacency matrix computation*. In practice, for each *j*-th base clustering (reported in Table 2 for the simple explanatory example), if two data belong to the same cluster they are considered similar, i.e., similarity μ=1, and if not they are dissimilar, i.e., similarity μ=0. Thus, for each *j*-th base clustering, an adjacency binary similarity matrix, $\overline{\overline{A}}^{j}$, of size *N*x*N*, is built (Strehl & Ghosh, 2002) (Figure 3, left, where the different black entries are μ=1 and the white entries are μ=0).

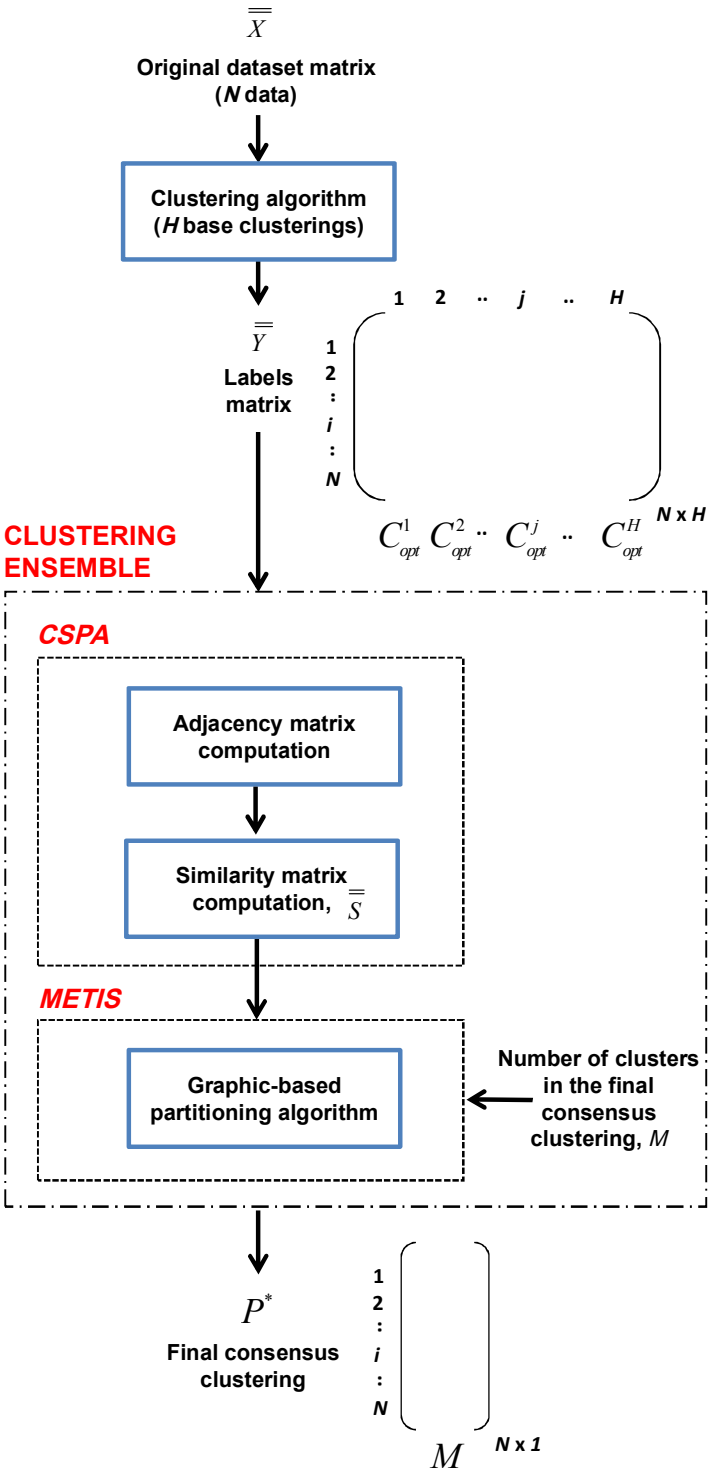

Figure 2: Flowchart of the *CSPA-METIS* approach.

Table 2: The *H*=3 base clusterings results of the *N*=5 data (illustrative example of *CSPA*).

	j=1	*j*=2	*j*=3
X_1	1	2	1
X_2	1	2	2
X_3	2	1	3
X_4	2	1	1
X_5	2	1	2

Step 2: *Similarity matrix computation*. The entry-wise average of the obtained *H* binary similarity matrices leads to obtaining the overall similarity matrix $\overline{\overline{S}} = \frac{1}{H}\overline{\overline{A}}\,\overline{\overline{A}}^T$ (Figure 3, right), of size *N*x*N* (Strehl & Ghosh, 2002). In this way, each entry of the similarity matrix has a value in [0,1], which is proportional to how likely a pair of data is, when grouped together.

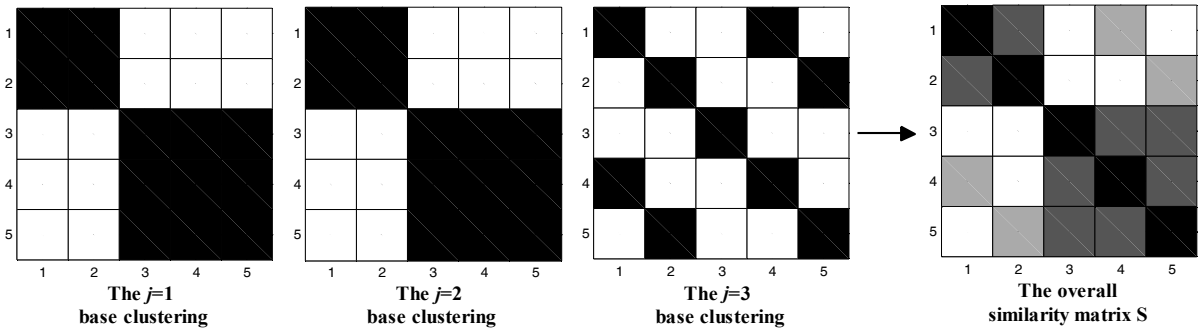

Figure 3: Base clusterings adjacency matrices (left) and the similarity matrix (right) of the numerical example.

Step 3: *Final consensus clustering computation*. To produce a final consensus clustering *P**, the graphic-based clustering algorithm *METIS* is adopted to partition the obtained similarity graph (shown in Figure 3, right) (Strehl & Ghosh, 2002). *METIS* is a multilevel graph partitioning algorithm that entails three main steps (refer to Karypis & Kumar, 1998, for more details):

1. the original graph is collapsed (coarsed) in smaller graphs (where the vertices are the data and the edges are the similarities), by resorting to Random Matching (*RM*) (Bui & Jones, 1993),
2. Spectral Bisection is used for partitioning the coarsened graphs (Barnard & Simon, 1994),
3. The partitions effectiveness is quantified by successively projecting the partitions into the original graph. It has been shown that *METIS* produces a high quality partitioning in a relatively small amount of time. However, the number of partitions to be found and, hence, the number of clusters in the final consensus clustering, has to be known "a priori". One option can be to

assign the number of clusters in the final consensus clustering to be equal to the maximum number of clusters in the H base clusterings, M=max (C_{opt}^{j}), j=1,…,H.

In the following Section, an ensemble approach is proposed to overcome the requirement of an "a priori" knowledge of the number of clusters M in the final consensus clustering.

3. The proposed ensemble clustering approach

In this Section, an ensemble approach is proposed, that evolves from that of Section 2 to avoid the hypothesis on the number of clusters M in the final consensus clustering. The proposed approach is based on the combination of: 1) *CSPA* method to compute the similarity matrix $\overline{\overline{S}}$, 2) Spectral Clustering to transform $\overline{\overline{S}}$ into a normalized laplacian matrix $\overline{\overline{L}}_{rs}$, and then, compute its spectrum information (eigenvectors) (see Appendix A.1), 3) a clustering algorithm, e.g., the K-means algorithm, that is fed with the eigenvectors calculated in the previous step 2), to find the final consensus clustering, and 4) the Silhouette index to quantify the goodness of the obtained clusters (see Appendix A.2).

The flowchart for the method is sketched in Figure 4. The method goes along the following steps:

Step 1: *Adjacency matrix computation*. This Step corresponds to Step 1 of Section 2.

Step 2: *Similarity matrix computation*. This Step corresponds to Step 2 of Section 2.

Step 3: *Spectral Clustering*. Once the overall similarity matrix $\overline{\overline{S}}$ is computed, Spectral Clustering (Appendix A.1) is used to reveal the hidden structure of $\overline{\overline{S}}$. The basic idea of Spectral Clustering is to extract the relevant information of the matrix $\overline{\overline{S}}$, by considering the eigenvectors associated to the ascended eigenvalues $\lambda_1,\lambda_2,...,\lambda_{C_{candidate}},...,\lambda_N$ of the normalized laplacian matrix $\overline{\overline{L}}_{rs}$ of $\overline{\overline{S}}$, to perform dimensionality reduction before clustering in fewer dimensions (see Step 1 in Appendix A.1) (Von Luxburg, 2007; Baraldi et al. 2013c). The eigenvectors $\overline{u}_1,\overline{u}_2,...,\overline{u}_{C_{candidate}},...,\overline{u}_N$ of the eigenvalues $\lambda_1,\lambda_2,...,\lambda_{C_{candidate}},...,\lambda_N$ are calculated and stored in a matrix $\overline{\overline{U}}$ with a size NxN (see Steps 2 and 4 in Appendix A.1), where $C_{candidate}$=[C_{min},C_{max}] and C_{min} and C_{max} are the minimum and maximum numbers of clusters considered for the final consensus clustering, respectively.

Step 4: *Clustering algorithm*. For each candidate number of clusters $C_{candidate}$, the reduced matrix of $\overline{\overline{U}}$ with a size Nx$C_{candidate}$ is partitioned into $C_{candidate}$ clusters by using a single clustering algorithm and the final consensus clustering $P^{*}_{C_{candidate}}$ is obtained. In this work, we resort to the K-means

algorithm, one of the most used clustering methods, to partition $\overline{\overline{U}}$ into $K=C_{candidate}$ clusters (Su & Chou, 2001; Fern & Lin, 2008).

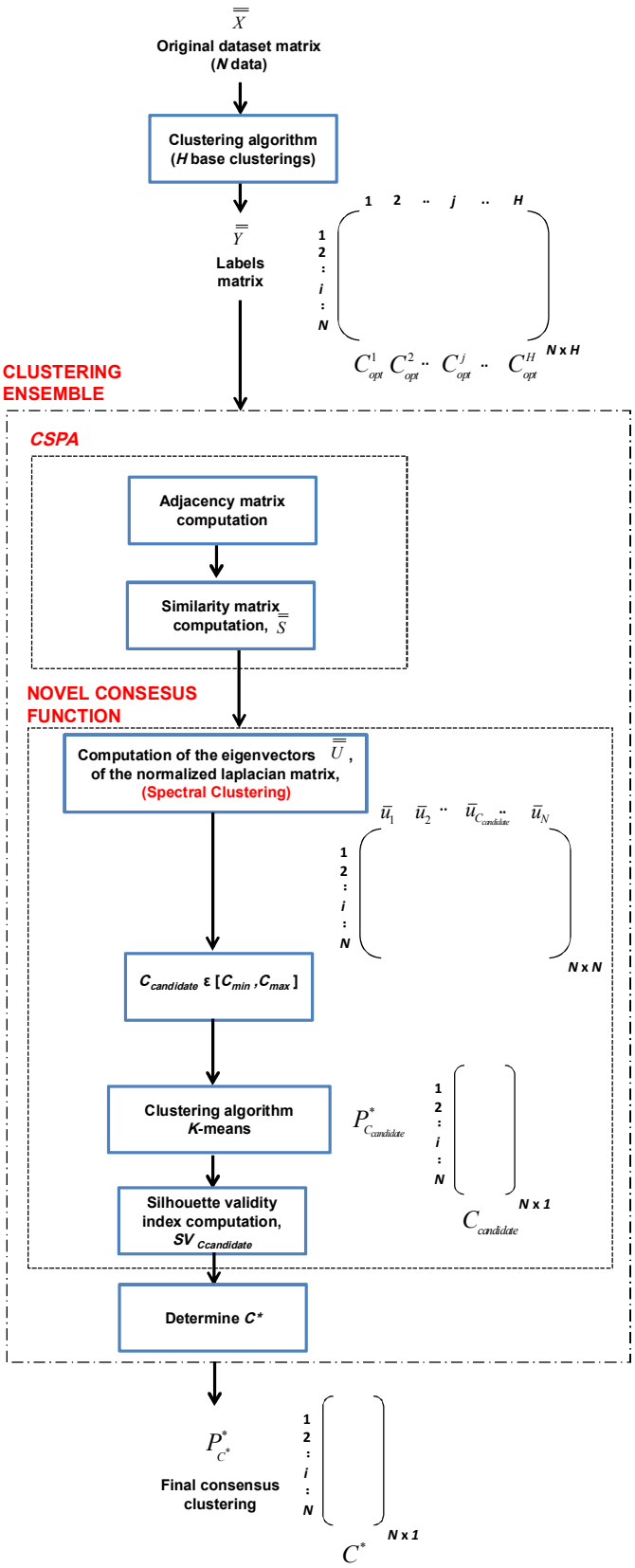

Figure 4: Flowchart of the proposed approach.

Step 5: *Final consensus clustering selection*. For each $C_{candidate}$, the obtained consensus clustering $P^*_{C_{candidate}}$ is evaluated by computing its Silhouette validity index $SV_{C_{candidate}}$ (Rousseeuw, 1987). The most appropriate consensus clustering $P^*_{C^*}$ is the one for which the Silhouette reaches a maximum, for which clusters are well separated and compacted (see also Appendix A.2).

4. Artificial case study

An artificial case study has been designed to generate N=149 data representative of the signals trends behaviors of M=7 different settings of shut-down operations. This is done to mimic the real industrial case of Section 6, concerning N=149 real shut-down transients of a *NPP* turbine. Each datum is described by F=7 features (as for the real case study of Section 6), representative of the turbine condition, e.g., mean value of the vibration signals, and of the environmental and operational conditions that can influence the turbine behavior, e.g., mean values of the vacuum and temperature signals. These data are stored in a matrix $\overline{\overline{X}}$ of a size 149x7.

The objective is to reveal the "hidden" (but simulated and, thus, known) structure P^* of the dataset $\overline{\overline{X}}$ by identifying groups of data with similar functional behaviors, representative of different operational conditions of the turbine. Without loss of generality, it is assumed that the operational conditions of the *NPP* turbine are M=7: 1) three classes of normal condition (*NC1*, *NC2*, *NC3*), 2) three classes of abnormal condition (*AC1*, *AC2*, *AC3*), and 3) one class of outliers (i.e., unknown behaviours). The dataset $\overline{\overline{X}}$ is pictorially shown in Figure 5: data with similar characteristics, e.g., vibration signals, and environmental and operational conditions which can influence the turbine behavior, e.g., vacuum and temperature signals, have been grouped together and will be treated within the same base clustering.

As shown in Figure 5, H=3 sets of features $\overline{\overline{X}}_j$ have been simulated and considered: the set of features 1, 2, and 3 (j=1), that of features 4 and 5 (j=2), and that of features 6 and 7 (j=3). This is found by a filter approach for which the optimal subsets of features are selected on the basis of statistical properties.

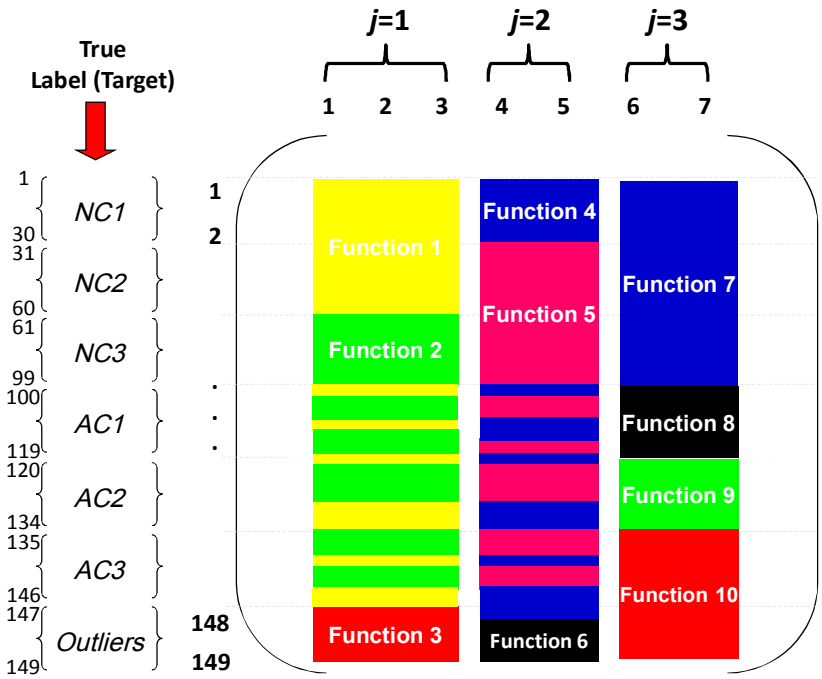

Figure 5: The seven operational conditions of the artificial case study.

The values of the features for different classes of data have been created by randomly sampling their realization from different multivariate normal and log-normal distribution functions (1 to 10 in Figure 5), whose combination characterizes the class.

Figure 6 shows the sampled data of the three sets of features: it is worth noticing that clustering each *j*-th set of features independently may reveal only some groups of the "hidden" *NPP* turbine operational conditions indicated in Figure 5, whereas only a final consensus clustering would enlighten all the M=7 clusters. In particular:

1. Figure 6 (Left) shows the dataset of the j=1 set of features: clusters can be seen for *NC1* and *NC2* in squares, *NC3* in diamonds, and there are also three outliers (147-149). Base clustering of this set of features cannot reveal any abnormal operational condition.

2. Figure 6 (Middle) shows the dataset of the j=2 set of features: clusters can be seen for *NC1* in squares, *NC2* and *NC3* in diamonds, and there is also one outlier (149). Again, base clustering of this set of features cannot reveal any abnormal operational condition.

3. Figure 6 (Right) shows the dataset of the j=3 set of features: clusters can be seen for all normal operational conditions in squares, and abnormal operational conditions *AC1* in diamonds, *AC2* in circles, and *AC3* in triangles. Base clustering of this set of features cannot reveal any outlier.

The objective is to aggregate these base clusterings into a final consensus clustering P^*, capable of identifying the "true" grouping of the shut-down transients of the *NPP* turbine.

To mine the clusters shown in Figure 6, the *j*-th base clustering outcomes are obtained by the unsupervised Fuzzy *C*-Means (*FCM*) algorithm (Baraldi et al. 2013c).

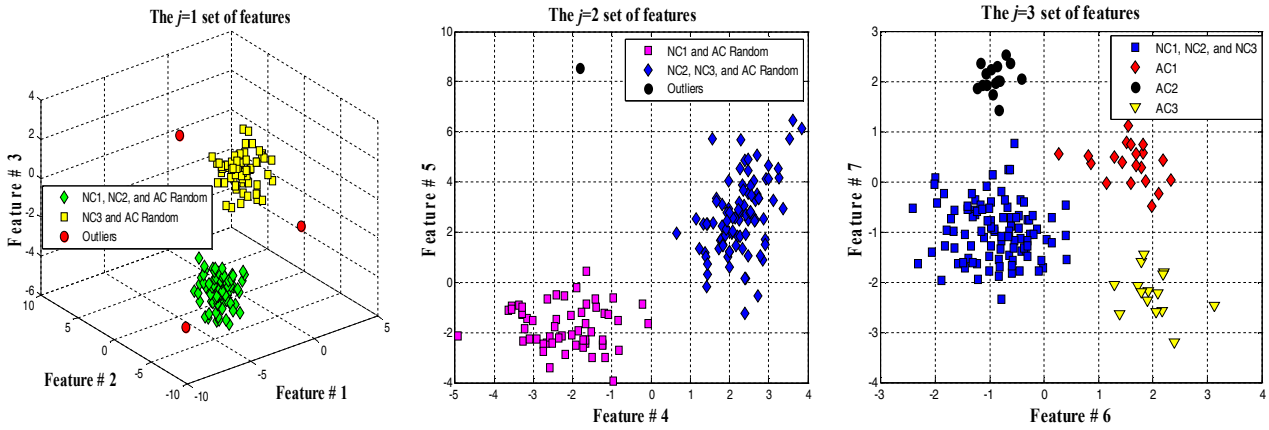

Figure 6: The artificial datasets of the three sets of features.

For identifying the correct number of clusters C_{opt}^{j} for each base clustering, single clustering validity index (e.g., Silhouette, Davies-Bouldin (*DB*), etc.) or a combination of different validity indices can be used (Onanena, Oukhellou, come, Jemei, Candusso, Hissel, & Aknin, 2013). In this work, Davies-Bouldin (*DB*) validity criterion has been considered for mining the clusters of the base clusterings (Davies & Bouldin, 1979) (whereas, the Silhouette validity index is used for identifying the optimum number of clusters in the final consensus clustering). The Davies-Bouldin (*DB*) criterion is based on the ratio of within-cluster and between-cluster distances: the optimal clustering, which gives optimal separation and compactness of the obtained clusters, has the smallest *DB* index value (Davies & Bouldin, 1979; Legány, Juhász, & Babos, 2006; Onanena et al. 2013).

Figure 7 shows the *DB* values for different numbers of clusters in the range of [2,10], for each *j*-th set of features: the star indicates the optimum number of clusters C_{opt}^{j}. For validation of the *DB* validity criterion to decide C_{opt}^{j}, we use the information on the "simulated" classes to which the data belong, to calculate the misclassification rate (Table 3) (it is worth noticing that in real industrial applications the real class is unknown).

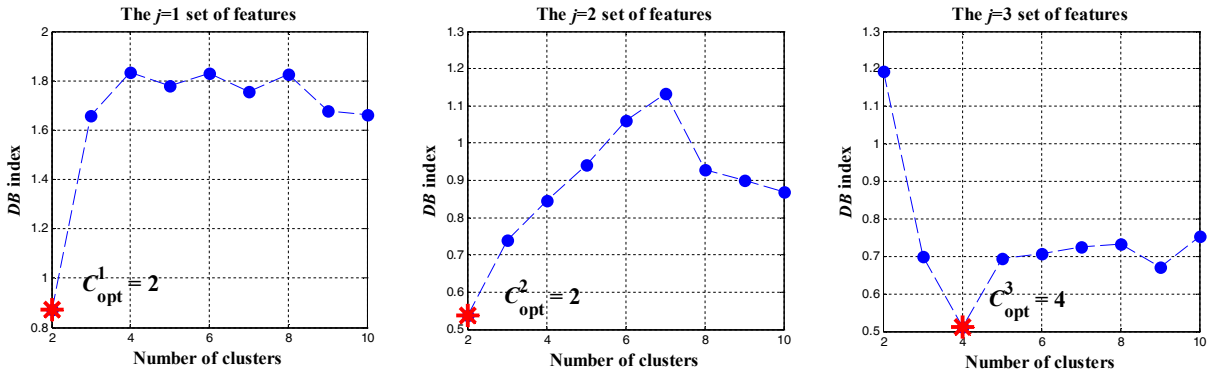

Figure 7: *DB* values vs. cluster numbers for the three sets of features.

Table 3: Optimum numbers of clusters and misclassification rates of clustering for the three sets of features.

Set of features	C_{opt}^{j}	*Misclassification rate*
j=1	*2*	*8.1%*
j=2	*2*	*5.3%*
j=3	*4*	*6.1%*

The obtained base clustering labels for each set of features have been, then, stored in a matrix $\overline{\overline{Y}}$ of size 149x3. The application of the clustering ensemble approach aims at finding the final consensus clustering of the data. In Section 4.1 and Section 4.2 the *CSPA-METIS* approach and the proposed approach are applied, respectively.

4.1. Application of *CSPA-METIS* approach

The application of the *CSPA-METIS* approach is here described according to the steps illustrated in Section 2: the overall adjacency matrix $\overline{\overline{A}}$ and the overall similarity matrix $\overline{\overline{S}}$ have been computed (Steps 1 and 2), respectively. A graph is obtained from $\overline{\overline{S}}$ and *METIS* is used to produce a final consensus clustering (Strehl & Ghosh, 2002; Topchy et al. 2004).

To this aim, the number of clusters *M*=7 in the final consensus clustering is assumed to be known "a priori". Figure 8 shows the obtained results of the aggregation *P** (left) compared to the true clustering (right).

The Figure shows the *N*=149 data (middle) in chronological order from top to bottom, with the associated true clustering labels located on the right coordinate, i.e., *NC1*, *NC2*, *NC3*, *AC1*, *AC2*, *AC3*, and *Outliers* with different color shades for their transients allocations. A fully symmetric plot would mean 100% of correct label assignment, whereas the blurrier the plot, the larger the misclassification rate. The application of *CSPA-METIS* leads us to distinguish mainly three clusters, i.e., *NC1*, *NC2* and *NC3*, whereas the remaining data have not been correctly clustered. Comparing

the obtained clustering results with the true "simulated" clustering, one can calculate the misclassification rate to be equal to 41.6% (62 out of 149 data incorrectly classified), which is not a satisfactory result.

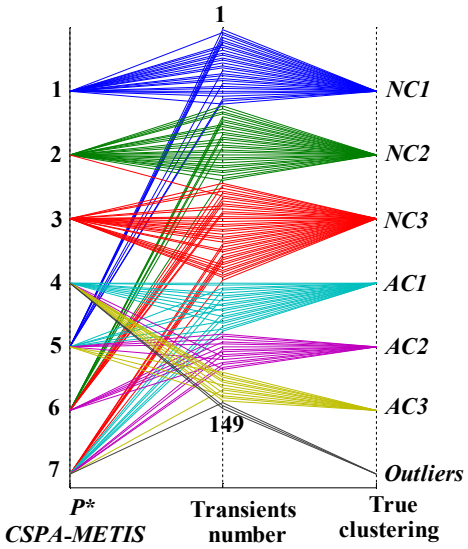

Figure 8: The obtained final consensus clustering by *CSPA-METIS* for *M*=7 vs. the true clustering.

One might be wondering whether the result would change if a different validity index would be used at this stage of the approach. For completeness, we use the Silhouette for selecting the number of clusters from the interval [2,16], where the lower bound (2) is the minimum number of base clusters (see Table 3), whereas the upper bound (16) is the number of the largest combination of the three base clusters (i.e., 2x2x4). The optimum number of clusters C^* in the final consensus clustering is found for the value at which the Silhouette measure is maximized, i.e., $C^*=3$ (star in Figure 9) (for which the obtained clusters are well separated and compacted). Despite that, again the clusters are not representative of the true "simulated" clustering, i.e., $M=7$.

The obtained results of the aggregation P^*, compared with the true clustering are shown in Figure 10 (left and right, respectively). Comparing the obtained clustering results with the true "simulated" clustering, one can calculate the misclassification rate to be equal to 36.9% (55 out of 149 data incorrectly classified).

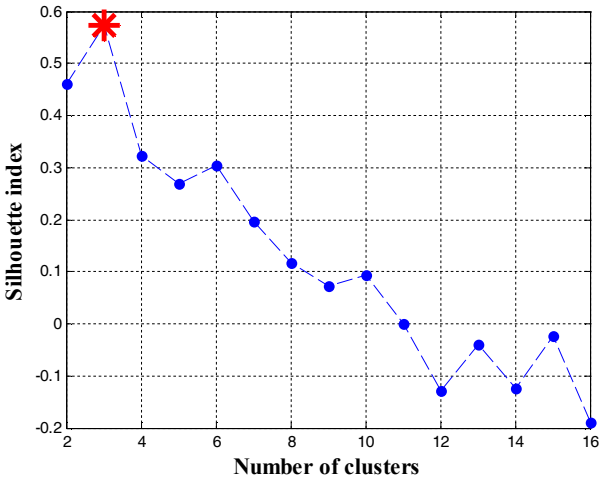

Figure 9: Silhouette values vs. cluster numbers.

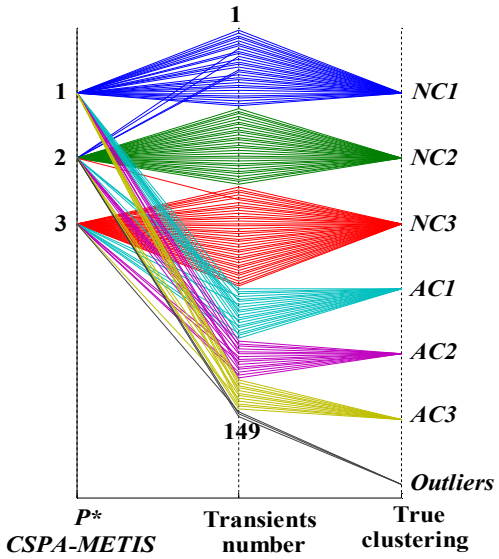

Figure 10: The obtained final consensus clustering by *CSPA-METIS* for *C = 3 vs. the true clustering.**

In the following Section, the application of the developed approach is shown to improve the final consensus clustering.

4.2. Application of the proposed ensemble clustering approach

The application of the proposed ensemble clustering is here described according to the steps presented in Section 3: the method entails a similar procedure of *CSPA-METIS* for calculating $\overline{\overline{S}}$ and a procedure to identify the final consensus clustering *P**.

Given the similarity matrix $\overline{\overline{S}}$, we calculate $\overline{\overline{L}}_{rs}$ and its eigenvectors $\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_{C_{candidate}}, \ldots, \overline{u}_{149}$, and the corresponding eigenvalue $\lambda_1, \lambda_2, \ldots, \lambda_{C_{candidate}}, \ldots, \lambda_{149}$. The obtained eigenvectors are stored in the matrix $\overline{\overline{U}}$ with size 149x149 (see also Appendix A.1). The number *M* of clusters in the final consensus

clustering is selected according to the values of the Silhouette index for different numbers of clusters $C_{candidate}$ that span the interval [2,16], where the lower bound (2) is the minimum number of base clusters (see Table 3), whereas the upper bound (16) is the number of the largest combination of the three base clusters (i.e., 2x2x4): the optimum number of clusters C^* in the final consensus clustering is the value at which the Silhouette is maximized, i.e., C^* =6 (star in Figure 11).

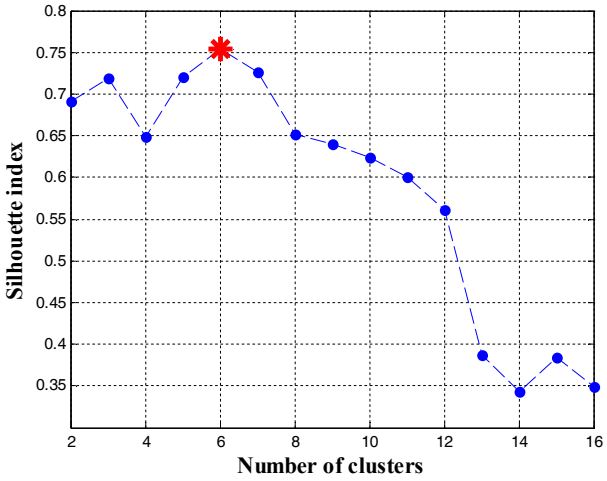

Figure 11: Silhouette values vs. cluster numbers.

The results of the application of the proposed method to the artificial case study are represented in Figure 12. Comparing the obtained clustering results (left) with the true "simulated" clustering (right), one can recognize that the misclassification rate has been reduced to 4.03% (6 out of 149 data incorrectly classified).

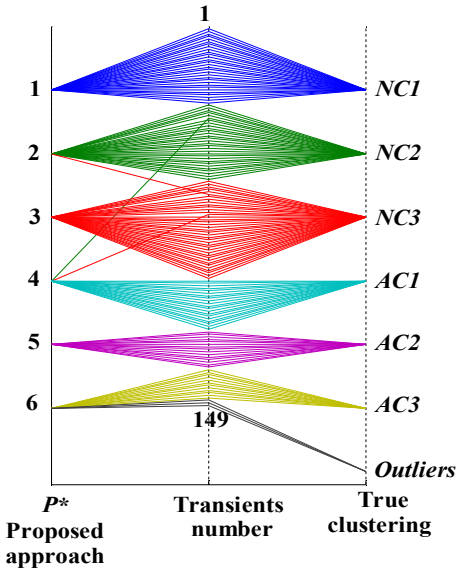

Figure 12: The obtained final consensus clustering by the proposed approach vs. the true clustering.

It is worth noticing that only six out of seven operational conditions have been recognized (C^* =6, while M=7). The outliers (three transients – class 7) have not been grouped together: this depends on the capability of the base clustering algorithm in recognizing the outliers (Topchy et al. 2004; Topchy et al. 2005; Serir et al. 2012).

For example, the optimum number of clusters for the j=1 set of features is $C_{opt}^{1}=2$ (see Figure 7), whereas it should be equal to 3 (see Figure 5). This sensitivity to the quality of the data at hand calls for an investigation on the robustness of the proposed method to different dataset characteristics, as it will be discussed in the following Section.

5. Robustness of the ensemble clustering approach to clustering overlapping

To verify the robustness of the proposed approach, a controlled sensitivity test has been designed. By robustness, here we intend the property of the approach to provide final consensus clustering with low misclassification rate even in case of a large overlap or separation of the real clusters.

With this aim, the clusters of Figure 6 have been modified by changing the parameters of the multivariate distributions from which the data are sampled, as follows:

1. **Case I (Large separation)**: in this case, the clusters of the j-th set of features, j=1,…,3 are designed to be well separated and compacted.
2. **Case II**: this is typically the case of Section 4. In this case, the clusters of the j-th set of features, j=1,…,3, are slightly overlapped compared to Case I.
3. **Case III (Large overlap)**: in this case, the obtained clusters from the j-th set of features, j=1,…,3, are overlapped and less compact.

Figure 13 shows the three cases for the three sets of features. As long as we are moving from Case I to Case III, the clusters identified start overlapping and become less compact.

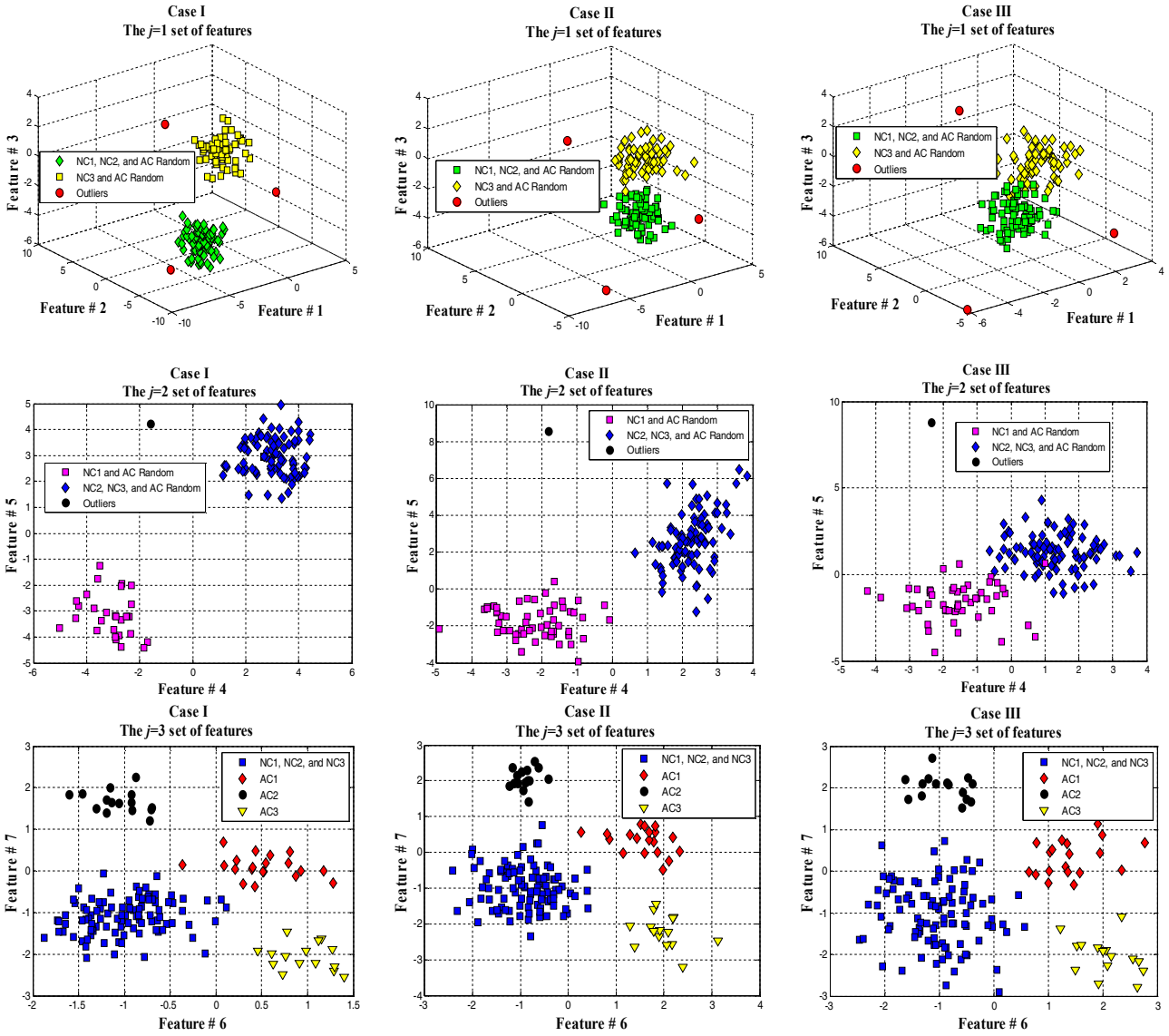

Figure 13: The three controlled cases for the three sets of features.

Figure 14 shows the results of the application of the proposed method to Cases I, II and III. The maximum Silhouette values (star in Figure 14 (left)) of the three cases indicate that the optimum number of clusters C^* in the final consensus clustering is still equal to 6.

The corresponding final consensus clustering (Figure 14 (right)) is compared with the one obtained by *CSPA-METIS* for the predetermined value M=7 (Figure 14 (middle)). It is interesting to notice that the clusters of the final consensus clustering obtained by the proposed approach are well representative of the true clusters, contrarily to the final consensus clustering obtained by *CSPA-METIS*.

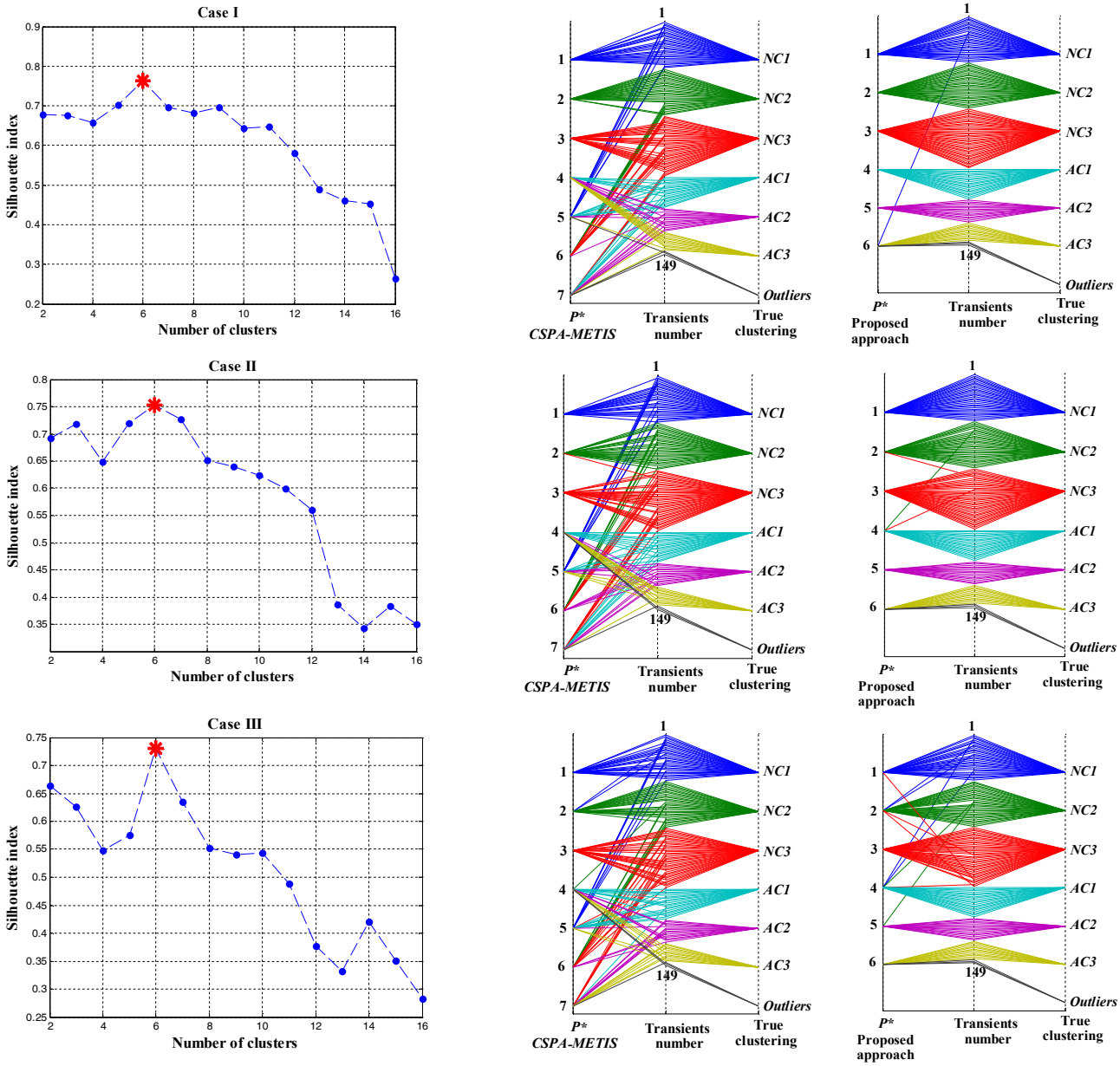

Figure 14: Silhouette values (left) and the final consensus clustering obtained for the three artificial cases by the proposed approach (right) and *CSPA-METIS* (middle).

The performances of the two approaches can be more precisely compared by calculating the misclassification rates in the three test cases by using the information on the real classes to which the data belong. The misclassification rates for the three cases using the two approaches are reported in Table 4.

Table 4: The misclassification rates of the proposed and *CSPA-METIS* approaches for the three test cases.

	The Proposed approach	**The *CSPA-METIS* approach**
Case 1	2.7%	36.9%
Case 2	4.0%	40.3%
Case 3	8.7%	43.6%

Furthermore, as the clusters of the sets of features are overlapped and spread (Case III), the performance of the proposed approach decreases compared to Case I, as expected. In conclusion, we can state that the proposed approach is superior to *CSPA-METIS*, for this particular dataset.

6. The real case study

The proposed approach has been applied to a real industrial case concerning N=149 real shut-down multidimensional transients of a *NPP* turbine. The generic i-th transient is a multidimensional transient in a Z=70 dimensional signal space with a time horizon of N_p=4500 time steps (2.5 hours).

The objective is to partition the N=149 multidimensional transients into M ("a priori" unknown) dissimilar groups, such that transients belonging to the same group are more similar than those belonging to the other groups. Engineering and experts judgment suggest a set of H=2 base clusterings:

1. Clustering of data representative of the turbine condition (j=1): seven signals of the turbine shaft vibrations have been considered (taken from sensors located at different stages of the turbine, whose detailed characteristics cannot be provided, due to confidentiality reasons), since vibration data contains signatures which, if properly interpreted, can reveal the operational condition of the turbine (Betta, Liguori, Paolillo, & Pietrosanto, 2002; Baraldi et al. 2013a). The similarity between the transients is measured by computing the pointwise difference between all seven vibration signals values. Then, a Spectral Clustering technique, embedding the unsupervised Fuzzy C-Means (*FCM*) algorithm, is applied to the obtained similarity matrix. Five different groups of transients $C_{opt}^{1}=5$ representing different operational conditions have been identified thanks to the Eigengap heuristic theory (see Appendix A.1 – Step 3).
2. Clustering of data representative of the environmental and operational conditions that can influence the turbine behavior (j=2): the values of turbine shaft speed, vacuum and structural

temperature signals have been considered (Baraldi et al. 2013b) (taken from different locations of the turbine, whose details cannot be disseminated, due to confidentiality reasons). The optimum numbers of clusters is found to be $C_{opt}^{2}=6$.

The base clusterings results have been aggregated in a matrix $\overline{\overline{Y}}$ with a size of 149x2 and the proposed approach has been applied following the steps illustrated in Section 3. The optimum number of clusters C^* in the final consensus clustering is selected according to the Silhouette values for different numbers of clusters $C_{candidate}$ that span in the interval [5,30], where the lower bound (5) is the minimum between C_{opt}^{1} and C_{opt}^{2}, and the upper bound (30) is the number of the largest combination of the two base clusters (i.e., 5x6).

It is important to point out that neither a too large nor a too small number of clusters can be considered as a valuable result from the practical point of view of linking turbine conditions with environmental and operational conditions: a large number of clusters makes the explanation of the turbine conditions too vague, whereas a small number is at risk of poor specification of the obtained clusters. In this analysis, the optimum number of clusters C^* in the final consensus clustering is found to be $C^*=14$, at which the Silhouette measure is maximized (star in Figure 15): this is a good compromise between small and large numbers of clusters. Figure 15 shows, indeed, that the Silhouette values for small and large numbers of C^* are much worse than for $C^*=14$, due to the dissimilarity of the data (inappropriately) assigned to the same clusters.

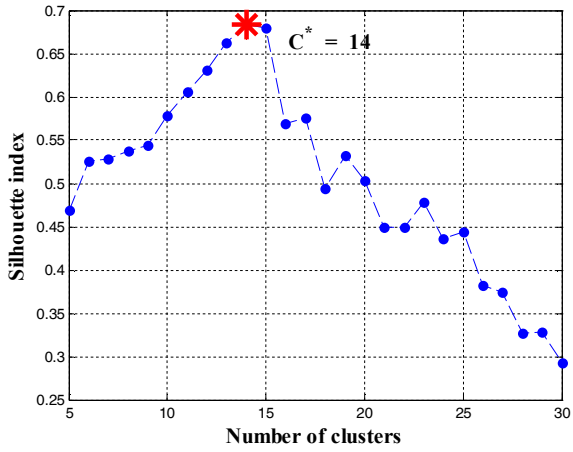

Figure 15: Silhouette values vs. cluster numbers.

Results of the application to the real case study are shown in Figure 16, where the N=149 transients are plotted in chronological order on the horizontal axis along with the j=1 base clustering results (the vertical axis) and the j=2 base clustering results represented by six different markers (square, diamond, star, triangle, circle, and dot).

Looking to the *j*=1 base clustering results, one can clearly identify four blocks of different labels (C_1^1, C_2^1, C_3^1 and C_4^1). Since the transients are numbered in increasing order with respect to their "calendar" occurrence, it has been possible to infer from the experts that the functional behavior of the turbine is different in the four clusters because of major maintenance interventions that have been undertaken at the specific calendar times and have resulted in radical changes of the turbine behaviour.

Among these main blocks, 17 transients (1, 3, 20, 24, 25, 26, 28, 31, 51, 70, 114, 115, 130, 136, 145, 146, and 147) are classified as outliers, since they are not clustered together with the previous 4 groups and, thus, could be representative of different faulty conditions in the turbine (C_5^1) (Baraldi et al. 2013a).

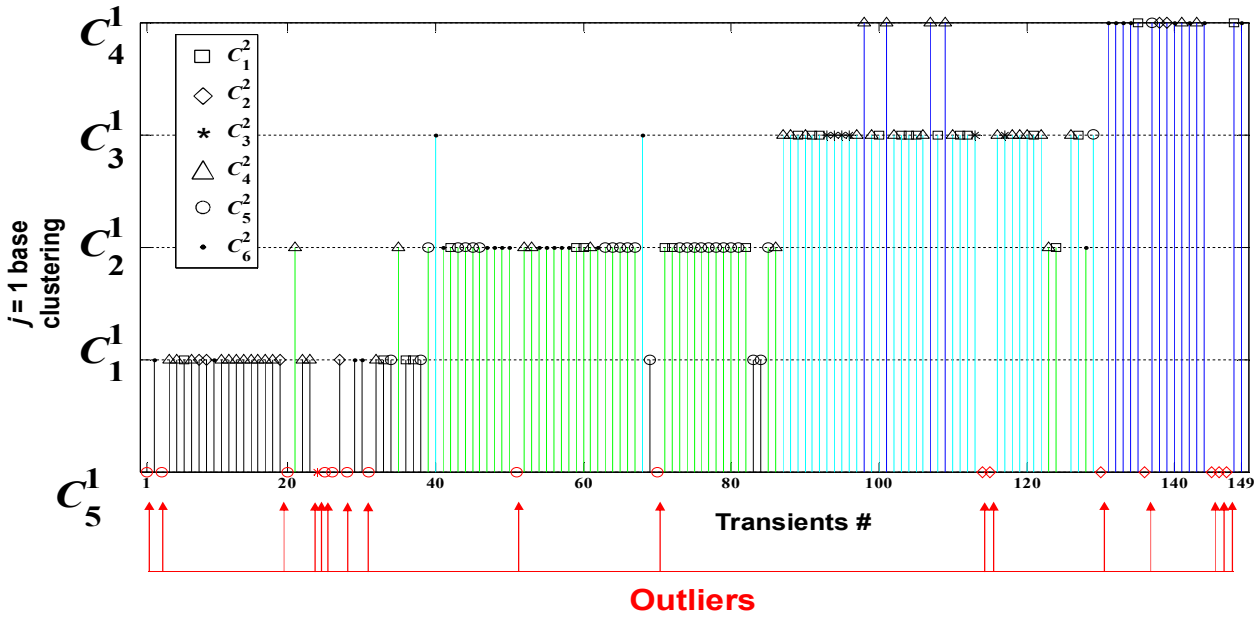

Figure 16: The 149 transients in chronological order along with the *j*=1 and *j*=2 base clustering results.

For the ease of clarity, we only consider vibration signal 1 as an example of vibration signal evolution of the *j*=1 base clustering results for the 5 clusters $C_1^1, C_2^1, C_3^1, C_4^1$ and C_5^1 and the corresponding turbine speed values (Figure 17).

One can recognize that, on one side, the functional behaviors of transients belonging to clusters 1 to 4 (C_1^1, C_2^1, C_3^1 and C_4^1) are similar, with some peculiarities that lead to their splitting into 4 clusters rather than being clustered together, whereas the transients of cluster 5 (C_5^1) greatly differ from the others (outliers) (Baraldi et al. 2013a).

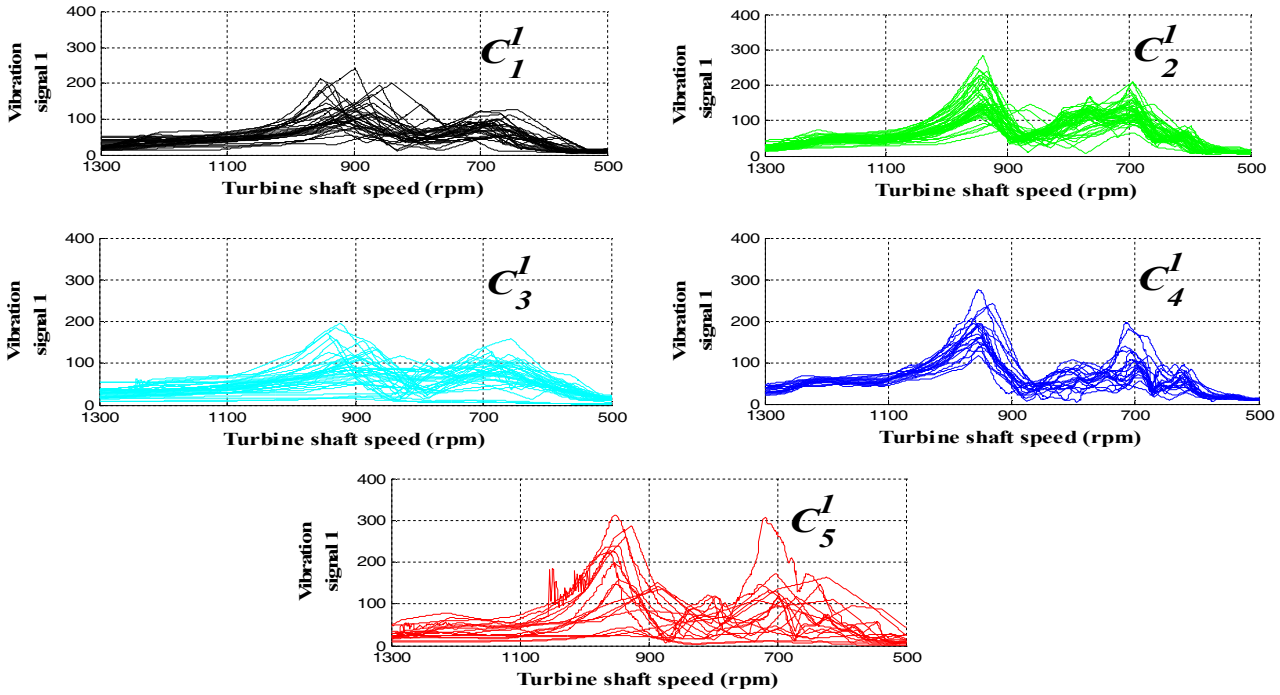

Figure 17: The evolution of vibration signal 1 of the 5 obtained clusters of the *j*=1 base clustering and the corresponding turbine speed values.

It is worth mentioning that the consensus clustering *P** can provide us with more insights than the *j*=1 base clustering. In fact, *j*=2 base clustering helps explaining the characteristics of C_1^1, C_2^1, C_3^1 and C_4^1 (of Figure 17) on the basis of the environmental and operational conditions.

In fact, looking at the environmental and operational conditions obtained by the *j*=2 base clustering in Figure 16, one can recognize that transients of each cluster obtained by the *j*=1 base clustering are influenced by different environmental and operational conditions that are obtained by the *j*=2 base clustering.

For example, Figure 18 shows pictorially that the transients belonging to C_2^1 of the *j*=1 base clustering have been splitted into four different final consensus clusters $(P_3^*, P_7^*, P_{10}^*$, and $P_{13}^*)$, each one due to a different environmental and operational conditions $(C_5^2, C_1^2, C_6^2$, and $C_4^2)$ as recognized by the *j*=2 base clustering (circle, square, dot and triangle markers, respectively in Figure 16).

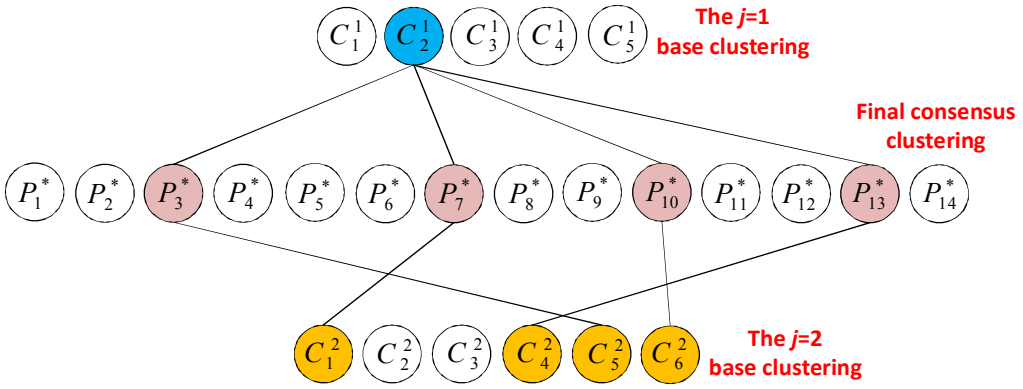

Figure 18: Characteristics of cluster 2 of the *j*=1 base clustering in the final consensus clustering on the basis of four environmental and operational conditions of the *j*=2 base clustering.

Figure 19 (top) shows the evolution of vibration signal 1 and the corresponding turbine speed for the transients belonging to C_2^1 of the *j*=1 base clustering splitted into four clusters (P_3^*, P_7^*, P_{10}^*, and P_{13}^*) obtained in the final consensus clustering (Figure 19 (bottom)): the transients indeed have similar functional behaviors as obtained by the *j*=1 base clustering, but they are further divided since they are influenced by different environmental and operational conditions obtained by the *j*=2 base clustering.

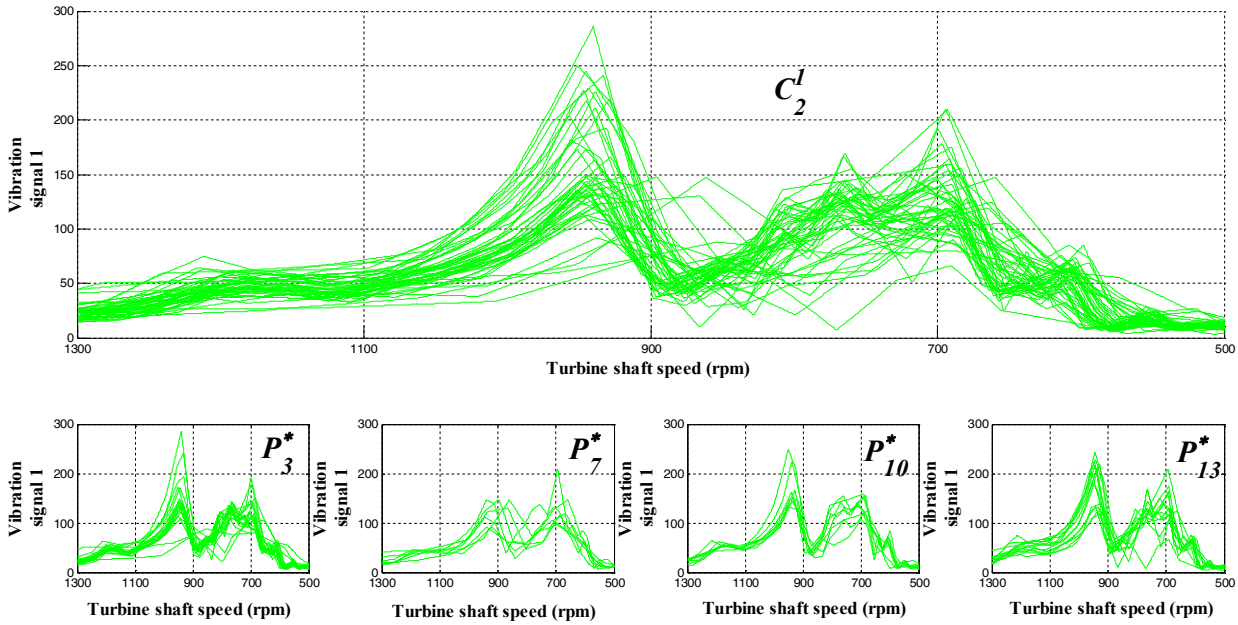

Figure 19: The evolution of vibration signal 1 of cluster 2 obtained by the *j*=1 base clustering with respect to the 4 clusters obtained in the final consensus clustering.

As last remark, it is worth mentioning that two clusters (P_2^* and P_{12}^*) of the final consensus clustering aggregate most of the outliers which belong to C_5^1 of the *j*=1 base clustering (all these

transients are explained by the environmental and operational conditions C_2^2 and C_5^2 of the *j*=2 base clustering).

This lead us to distinguish, in the set of outlier transients with peculiar behavior of the turbine, two representative faulty conditions at two different environmental and operational conditions (P_2^* and P_{12}^*).

Figure 20 shows the evolution of vibration signal 1 and the corresponding turbine speed for the transients of the two final consensus clusters (P_2^* and P_{12}^*): despite that these transients are sufficiently similar in functional behaviour to belong to C_5^1 of the *j*=1 base clustering, their grouping into only two consensus clusters is driven (and can be explained) by the two different environmental and operational conditions (C_2^2 and C_5^2) obtained by the *j*=2 base clustering.

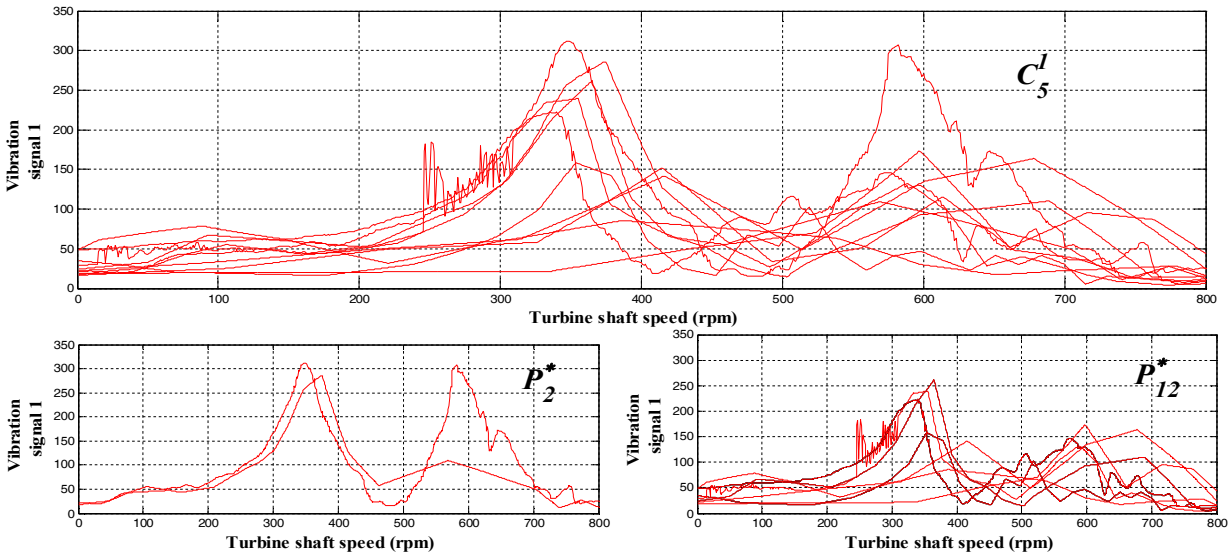

Figure 20: The evolution of vibration signal 1 of the transients aggregated in the two clusters (P_2^* and P_{12}^*) of the final consensus clustering.

The ability of the proposed approach to distinguish the different operational conditions of the turbine and recognize different faulty conditions of the turbine is an indication of the good performance of the proposed approach.

7. Conclusions

In this work, an approach to build a consensus clustering of individual base clusterings is proposed, based on Spectral Clustering and Silhouette validity index. First, the base clustering results are

summarized in a co-association matrix by pairwise similarity computation. Then, a Spectral Clustering technique, embedding the unsupervised *K*-Means algorithm, is applied to the matrix of similarity values so that the clusters are formed by the most similar data. The optimum number of clusters is selected among several candidates based on the morphology of the obtained clusters, measured by the Silhouette validity index that gives reason of the similarity of data belonging to the same cluster and the dissimilarity with those in the other clusters.

The proposed approach has been successfully applied to an artificial case study "properly" designed to reproduce the signal trend behavior of a Nuclear Power Plant (*NPP*) turbine during shut-down transients. The results obtained have been shown satisfactory by comparison to those obtained by the *CSPA-METIS* approach of literature. Further, three controlled datasets containing *M* sparse or overlapping clusters have been analyzed to verify the robustness with respect to clustering overlapping.

Finally, the proposed approach has been applied to a real industrial case concerning the multidimensional signals of 149 shut-down transients of a *NPP* turbine. Different base clusterings representative of different groupings of the shut-down transients of the turbine have been obtained by using multiple, different sources of data (features), such as vibration, turbine shaft speed, temperature, and vacuum signals. The approach has led to distinguishing 14 different operational conditions of the turbine, representative of different behaviors under different environmental and operational conditions. Two peculiar behaviors of the turbine have been identified, representative of two faulty conditions at two different environmental and operational conditions.

Acknowledgements

This research has been carried out under the project "Processing condition monitoring data for diagnosis and prognosis of components in a fleet of electricity production plants" in collaboration between Électricité de France (EDF) and Politecnico di Milano. The participation of Sameer Al-Dahidi and Piero Baraldi has been possible within the European Union Project INNovation through Human Factors in risk analysis and management (INNHF, www.innhf.eu) funded by the 7th framework program FP7-PEOPLE-2011-Initial Training Network: Marie-Curie Action.

The authors would like to thank all the reviewers for their valuable comments to improve the quality of this paper.

Nomenclature

NPP	Nuclear Power Plant	$\overline{\overline{I}}$	Eigenvectors of $\overline{\overline{L}}_{rs}$
CSPA	Cluster-based Similarity Partitioning Algorithm	F	Number of features (columns) of $\overline{\overline{X}}$
	Original space dataset matrix		
METIS	Serial Graph Partitioning and Fill-reducing Matrix Ordering Algorithm	Z	Number of signals of each *i*-th transient
NC	Normal operational conditions	$\overline{\overline{Y}}_j$	*j*-th base clustering result, $j=1,\ldots,H$
AC	Abnormal operational conditions	C^*	Optimum number of clusters in the final consensus clustering
SOM	Self-Organizing Maps	$P^*_{C_{candidate}}$	Final consensus clustering with $C_{candidate}$ clusters, $C_{candidate} \in [C_{min}, C_{max}]$
FCM	Fuzzy *C*-Means	$P^*_{C^*}$	Final consensus clustering at the optimum number of clusters, C^*
HMMs	Hidden Markov Models	M	True number of clusters in the final consensus clustering
*P**	Final consensus clustering	*DB*	Davies-Bouldin criteria
$\overline{\overline{X}}$	Original space dataset matrix	$\overline{\overline{X}}_j$	*j*-th set of features of the original dataset, $j=1,\ldots,H$
$\overline{\overline{Y}}$	Labels aggregation matrix (base clustering results)	$SV_{C_{candidate}}$	Silhouette validity value at $C_{candidate}$, $C_{candidate} \in [C_{min}, C_{max}]$
H	Number of base clusterings	a^i	Average distance of the *i*-th datum from the other data belonging to the same cluster
j	Index of base clustering	b^i	Minimum average distance of the *i*-th datum from the data belonging to a different cluster
N	Number of data (rows) of $\overline{\overline{X}}$	S_m	Mean Silhouette value for the *m*-th cluster
i	Index of a datum (transient) belonging to $\overline{\overline{X}}$	S_{ij}	Pairwise similarity value between the *i*-th and *j*-th data
C^j_{opt}	Optimum number of clusters of the *j*-th set of features	C_m	*m*-th cluster in the final consensus clustering
$\overline{\overline{A}}^j$	Adjacency binary similarity matrix of the *j*-th base clustering, $j=1,\ldots,H$	n_m	Total number of data in the *m*-th cluster in the final consensus clustering
μ	Pairwise binary similarity value	S^i	Silhouette value of the *i*-th datum
$\overline{\overline{S}}$	Co-association matrix	$\overline{\overline{D}}$	*i*-th entry of the diagonal matrix $\overline{\overline{D}}$
$C_{\min}$	Minimum number of clusters in the final consensus clustering *P**	$\overline{\overline{L}}_{rs}$	Diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_N$
$C_{\max}$	Maximum number of clusters in the final consensus clustering *P**	$\overline{u}_{C_{candidate}}$	Normalized Laplacian Matrix
$C_{candidate}$	Possible number of clusters in the final consensus clustering *P**, $C_{candidate} \in [C_{min}, C_{max}]$	λ	The $C_{candidate}$-th eigenvector of $\overline{\overline{L}}_{rs}$
		$\overline{\overline{U}}$	Eigenvalue of $\overline{\overline{L}}_{rs}$

References

(Ahuja et al., 2012) Ahuja, S., & Dhanya, C. T. (2012). Regionalization of Rainfall Using RCDA Cluster Ensemble Algorithm in India. *Journal of Software Engineering and Applications*, vol. 5 (8), pp. 568-573. doi: 10.4236/jsea.2012.58065

(Al-Dahidi et al., 2014) Al-Dahidi, S., Baraldi, P., Di Maio, F., & Zio, E. (2014). A novel fault detection system taking into account uncertainties in the reconstructed signals. *Annals of Nuclear Energy*, vol. 73, pp. 131–144. doi:10.1016/j.anucene.2014.06.036

(Al-Dahidi, 2014) Al-Dahidi, S. (2014). The Use of Self Organizing Maps for Diagnosing Faults in Motor Bearings. *Safety and Reliability: Methodology and Applications- Proceedings of the European Safety and Reliability Conference, ESREL 2014 (*895-902), September 14-18, Wroclaw, Poland.

(Ayad et al., 2010) Ayad, H. G., & Kamel, M. S. (2010). On voting-based consensus of cluster ensembles. *Pattern Recognition*, vol. 43(5), pp. 1943-1953.

(Baraldi et al., 2012) Baraldi, P., Di Maio, F., & Zio, E. (2012). Unsupervised clustering for fault diagnosis. *Proceedings of Prognostics and System Health Management Conference (PHM-2012 IEEE Conference) (1-9)*, May 23-25, Beijing, China.

(Baraldi et al., 2013a) Baraldi, P., Di Maio, F., Rigamonti, M., Zio, E., & Seraoui, R. (2013a). Unsupervised clustering of vibration signals for identifying anomalous conditions in a nuclear turbine. *Special Issue RACR2013, on the Journal of Intelligent and Fuzzy Systems (JIFS)*. doi: 10.3233/IFS-141459

(Baraldi et al., 2013b) Baraldi, P., Di Maio, F., Rigamonti, M., Zio, E., & Seraoui, R. (2013b). Clustering for unsupervised fault diagnosis in nuclear turbine shut-down transients. *Mechanical Systems and Signal Processing*, Available online 16 January 2015. doi: 10.1016/j.ymssp.2014.12.018

(Baraldi et al., 2013c) Baraldi, P., Di Maio, F., & Zio, E. (2013c). Unsupervised Clustering for Fault Diagnosis in Nuclear Power Plant Components. *International Journal of Computational Intelligence Systems*, vol. 6 (4), pp. 764-777.

(Barnard et al., 1994) Barnard, S. T., & Simon, H. D. (1994). Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. *Concurrency: Practice and Experience*, vol. 6(2), pp. 101-117.

(Baruah et al., 2005) Baruah, P., & Chinnam, R. B. (2005). HMMs for diagnostics and prognostics in machining processes. International Journal of Production Research, vol. 43(6), pp. 1275-1293.

(Betta et al., 2002) Betta, G., Liguori, C., Paolillo, A., Pietrosanto, A. (2002). A DSP-based FFT-Analyzer for the fault diagnosis of rotating machine based on vibration analysis. *IEEE Transactions on instrumentation and measurements*, vol. 51(6), pp. 1316-1322.

(Bezdek, 1981) Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, New York.

(Bhavaraju et al., 2010) Bhavaraju K. M., Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2010). A Comparative Study on Bearings Faults Classification by Artificial Neural Networks and Self-Organizing Maps using Wavelets. *International Journal of Engineering Science and Technology,* vol. 2(5), pp. 1001-1008.

(Bocaniala et al., 2004) Bocaniala, C.D., Sa Da Costa, J., & Palade, V. (2004). A novel fuzzy classification solution for fault diagnosis. *Journal of Intelligent and Fuzzy Systems*, vol. 15 (3-4), pp. 195-205.

(Bolotin et al., 1998) Bolotin, V.V., & Shipkov, A.A. (1998). A model of the environmentally affected growth of fatigue cracks. *Journal of Applied Mathematics and Mechanics*, vol. 62(2), pp. 289-296. doi:10.1016/S0021-8928(98)00037-9

(Bui et al., 1993) Bui, T., & Jones, C. (1993). A Heuristic for Reducing Fill-In in Sparse Matrix Factorization. *In 6th SIAM Conference Parallel Processing for Scientific Computing (*445–452), March 22-24, Norfolk, Virginia, USA.

(Chakaravathy et al., 1996) Chakaravathy, S. V., & Ghosh, J. (1996). Scale based clustering using a radial basis function network. *IEEE Transactions on Neural Networks*, vol. 2(5), pp. 1250–61. doi: 10.1109/72.536318

(Chaovalit et al., 2005) Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. *In System Sciences*, *2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on (112c-112c). IEEE, January 3-6, Big Island, Hawaii. doi: 10.1109/HICSS.2005.445*

(Charrad et al., 2010) Charrad, M., Lechevallier, Y., Ahmed, M. B., Saporta, G. (2010). On the Number of Clusters in Block Clustering Algorithms. *In 23rd International FLAIRS Conference (392-397), May 19-21, Florida, USA*.

(Chatterjee et al., 2013) Chatterjee, S., & Mukhopadhyay, A. (2013). Clustering Ensemble: A Multiobjective Genetic Algorithm based Approach. *Procedia Technology*, vol. 10, pp. 443-449. doi:10.1016/j.protcy.2013.12.381

(Chen, 2007) Chen, K. (2007). *Trends in neural computation*. Springer.

(Datta et al., 2007) Datta, A., Mavroidis, C., & Hosek, M. (2007). A Role of Unsupervised Clustering for Intelligent Fault Diagnosis. *In ASME 2007 International Mechanical Engineering Congress and Exposition*. , vol. 9: Mechanical Systems and Control, pp. 687-695. doi:10.1115/IMECE2007-43492.

(Davies et al., 1979) Davies, D.L., & Bouldin, D.W. (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. *PAMI*-1(2), pp. 224-227. doi: 10.1109/TPAMI.1979.4766909

(Di Maio et al., 2012) Di Maio, F., Hu, J., Tse, P., Pecht, M., Tsui, K., & Zio, E. (2012). Ensemble-approaches for clustering health status of oil sand pumps. *Expert Systems with Applications,* vol. 39(5), pp. 4847-4859.

(Di Maio et al., 2014) Di Maio, F., Nicola, G., Zio, E., & Yu, Y. (2014). Ensemble-based sensitivity analysis of a Best Estimate Thermal Hydraulics model: Application to a Passive Containment Cooling System of an AP1000 Nuclear Power Plant. *Annals of Nuclear Energy*, vol. 73, November 2014, pp. 200-210. doi:10.1016/j.anucene.2014.06.043

(Dimitriadou et al., 2001) Dimitriadou, E., Weingessel, A., & Homik, K. (2001). Voting-merging: an ensemble method for clustering. *In Proc. 2001 International Conference Artificial Neural Networks (ICANN'01) (*217-224), August 21–25, Vienna, Austria. doi : 10.1007/3-540-44668-0_31

(Dudoit et al., 2003) Dudoit, S., & Fridlyand, J. (2003). Bagging to improve the accuracy of a clustering procedure. *Bioinformatics*, vol. 19(9), pp. 1090-1099.

(Fern et al., 2008) Fern, X. Z., & Lin, W. (2008). Cluster ensemble selection. *Statistical Analysis and Data Mining*, vol. 1(3), pp. 128-141.

(Figueiredo et al., 2002) Figueiredo, M. A., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24(3), pp. 381-396. doi: 10.1109/34.990138

(Fred et al., 2005) Fred, A. L., & Jain, A. K. (2005). Combining multiple clusterings using evidence accumulation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27(6), pp. 835-850.

(Ghaemi et al., 2009) Ghaemi, R., Sulaiman, M. N., Ibrahim, H., & Mustapha, N. (2009). A survey: clustering ensembles techniques. *World Academy of Science, Engineering and Technology*, vol. 50, pp. 636-645.

(Ghaemi et al., 2011) Ghaemi, R., bin Sulaiman, N., Ibrahim, H., & Mustapha, N. (2011). A review: accuracy optimization in clustering ensembles using genetic algorithms. *Artificial Intelligence Review*, vol. 35(4), pp. 287-318.

(Gonçalves et al., 2011) Gonçalves, L. F., Bosa, J. L., Balen, T. R., Lubaszewski, M. S., Schneider, E. L., & Henriques, R. V. (2011). Fault detection, diagnosis and prediction in electrical valves using self-organizing maps. *Journal of Electronic Testing,* vol. 27(4), pp. 551-564.

(Greene et al., 2007) Greene, D., & Cunningham, P. (2007). Constraint selection by committee: An ensemble approach to identifying informative constraints for semi-supervised clustering. *In Machine Learning: ECML 2007*, pp. 140-151. Springer Berlin Heidelberg.

(Hartigan, 1975) Hartigan, J. (1975). CLUSTERING ALGORITHMS. New York, Wiley.

(Iqbal et al., 2012) Iqbal, A. M., Moh'd, A., & Khan, Z. (2012). Semi-supervised clustering ensemble by voting. *In: Proceeding of the International Conference on Information and Communication Systems (ICICS 2009) (1–5), December 8-10, Macau, China.*

(Jardine et al., 2006) Jardine, A.K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, vol. 20(7), pp. 1483–1510. doi:10.1016/j.ymssp.2005.09.012

(Johnson, 1967) Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, vol. 32(3), pp. 241-254.

(Karypis et al., 1995) Karypis, G., & Kumar, V. (1995). METIS - *Unstructured Graph Partitioning and Sparse Matrix Ordering System*, Version 2.0 (Technical report).

(Karypis et al., 1997) Karypis, G., Aggarwal, R., & Kumar, V., Shekhar, S. (1997). Multilevel Hypergraph Partitioning: Applications in VLSI Design, *In Proc. ACM/IEEE Design Automation Conference*, pages 526-529.

(Karypis et al., 1998) Karypis, G., & Kumar, V. (1998). A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal of Scientific Computing*, vol. 20(1), pp. 359-392.

(Legány et al., 2006) Legány, C., Juhász, S., & Babos, A. (2006). Cluster validity measurement techniques. *Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases* (388-393), December 16-18, Tenerife, Canary Islands, Spain

(Leguizamón et al., 1996) Leguizamón, S., Pelgrum, H., & Azzali, S. (1996). Unsupervised Fuzzy C-means classification for the determination of dynamically homogeneous areas. *Revista SELPER*, vol. 12(12), pp. 20-24.

(Li et al., 2011) Li, Y. S., & Chen, K. C. (2011). Graph partition and identification of cluster number in data analysis. *In Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC '11) (5)*, vol. 62(5), February 21-23, Seoul, korea. doi=10.1145/1968613.1968688.

(Lin et al., 2013) Lin, Y., Chen, M., & Zhou, D. (2013). Online probabilistic operational safety assessment of multi-mode engineering systems using Bayesian methods. *Reliability Engineering & System Safety*, vol. 119, pp. 150-157. doi:10.1016/j.ress.2013.05.018

(Mohar, 1997) Mohar, B. (1997). Some Applications of Laplace Eigenvalues of Graphs. *Graph Symmetry: Algebraic Methods and Applications*, vol. 497, pp. 225-275. doi: 10.1007/978-94-015-8937-6_6

(Muller et al., 2008) Muller, A., Suhner, M. C., & Iung, B. (2008). Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *Reliability Engineering & System Safety*, vol. 93(2), pp. 234-253. doi:10.1016/j.ress.2006.12.004

(Ng et al., 2001) Ng, A.Y., Jordan, M.I., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems (NIPS)*, vol. 14, pp. 849-856.

(Onanena et al., 2013) Onanena, R., Oukhellou, L., come, E., Jemei, S., Candusso, D., Hissel, D., & Aknin, P. (2013). Fuel Cell Health Monitoring Using Self Organizing Maps. *Chemical Engineering Transactions*, vol. 33, pp. 1021-1026. doi: 10.3303/CET1333171

(Rousseeuw, 1987) Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65. doi:10.1016/0377-0427(87)90125-7

(Serir et al., 2012) Serir, L., Ramasso, E., & Zerhouni, N. (2012). Evidential evolving Gustafson–Kessel algorithm for online data streams partitioning using belief function theory. *International journal of approximate reasoning*, vol. 53(5), pp. 747-768. doi:10.1016/j.ijar.2012.01.009

(Serir et al., 2013) Serir, L., Ramasso, E., Nectoux, P., & Zerhouni, N. (2013). E2GKpro: An evidential evolving multi-modeling approach for system behavior prediction with applications. *Mechanical Systems and Signal Processing*, vol. 37(1), pp. 213-228. doi:10.1016/j.ymssp.2012.06.023

(Siegel et al., 2011) Siegel, D., & Lee, J. (2011). An Auto-Associative Residual Processing and K-means Clustering Approach for Anemometer Health Assessment. *International Journal of Prognostics and Health Management*, vol. 2(2) 014, pp. 1-12. ISSN 2153-2648

(Strehl et al., 200) Strehl, A., & Ghosh, J. (2002). Cluster ensembles-a knowledge reuse framework for combining partitions. *The Journal of Machine Learning Research*, vol. 3, pp. 583-617. doi: 10.1162/153244303321897735

(Salvador, 2002) Salvador, A. (2002). Faults diagnosis in industrial processes with a hybrid diagnostic system. *In MICAI 2002: Advances in Artificial Intelligence*, vol. 2313, pp. 536-545. Springer Berlin Heidelberg. doi: 10.1007/3-540-46016-0_56

(Su et al., 2001) Su, M. C., & Chou, C. H. (2001). A modified version of the K-means algorithm with a distance based on cluster symmetry. *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23(6), pp. 674-680.

(Topchy et al., 2004) Topchy, A., Jain, P., & Punch, W. (2004). A Mixture Model for Clustering Ensembles. *Proceedings of the 2004 SIAM International Conference on Data Mining (*379-390), April 22-24, Florida. doi: 10.1137/1.9781611972740.35

(Topchy et al., 2005) Topchy, A., Jain, A. K., & Punch, W. (2005). Clustering ensembles: Models of consensus and weak partitions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27(12), pp. 1866-1881.

(Van Wijk et al., 1999) Van Wijk, J., & Van Selow, E. (1999). Cluster and calendar based visualization of time series data. *Proceedings of IEEE Symposium on Information Visualization* (4-9), October 24-29, San Francisco, CA. doi: 10.1109/INFVIS.1999.801851

(Vega-Pons et al., 2011) Vega-Pons, S., & Ruiz-Shulcloper, J. (2011). A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 25(03), pp. 337-372.

(Vlachos et al., 2003) Vlachos, M., Lin, J., Eamonn K., & Dimitrios G. (2003). A wavelet-based anytime algorithm for k-means clustering of time series. *In Proc. Workshop on Clustering High Dimensionality Data and Its Applications (23-30)*, San Francisco, CA.

(Von Luxburg, 2007) Von Luxburg, U. (2007). A Tutorial on Spectral Clustering. *Statistics and Computing*, vol. 17(4), pp. 395-416.

(Wang et al., 2008) Wang, T., Yu, J., Siegel, D., & Lee, J. (2008). A similarity based prognostics approach for remaining useful life estimation of engineered systems. *In Prognostics and Health Management, 2008. PHM 2008. International Conference on*, (1-6). IEEE., October 6-9, Denver, CO. doi: 10.1109/PHM.2008.4711421

(Wang, 2010) Wang, T. (2010). Trajectory Similarity Based Prediction for Remaining Useful Life Estimation. Doctoral dissertation. University of Cincinnati, U.S. http://gradworks.umi.com/3432353.pdf

(Wu et al., 2011) Wu, F., & Lee, J. (2011). Information Reconstruction Method for Improved Clustering and Diagnosis of Generic Gearbox Signals. *International Journal of the Prognostics and Health Management Society*, vol. 2(1) 004, 9 pages. ISSN 2153-2648

(Xiufeng et al., 2010) Xiufeng, G., & Changzheng, X. (2010). K-means Multiple Clustering Research Based on Pseudo Parallel Genetic Algorithm. *In Information Technology and Applications (IFITA)*, 2010 International Forum on (1, pp. 30-33). IEEE, July 16-18, Kunming. doi : 10.1109/IFITA.2010.186

(Zhou et al., 2004) Zhou, S., Zhang, J., & Wang, S. (2004). Fault diagnosis in industrial processes using principal component analysis and hidden Markov model. *In American Control Conference, 2004. Proceedings of the 2004*, vol. 6, pp. 5680-5685. IEEE.

(Zhao et al., 2007) Zhao, Z., & Liu, H. (2007). Spectral feature selection for supervised and unsupervised learning. *Proceedings of the 24th international conference on Machine learning* (1151-1157), June 20-24, Corvalis, Oregon.

Appendices

Appendix A.1 Unsupervised Spectral clustering

Spectral Clustering technique uses the spectrum (eigenvalues) of the similarity matrix of the data to perform dimensionality reduction before clustering in fewer dimensions (Baraldi et al. 2012; Baraldi et al. 2013c). In this work, the similarity matrix $\overline{\overline{S}}$ of size *N*x*N* is computed by Cluster-based Similarity Partition Algorithm (*CSPA*). The Spectral Clustering technique entails four steps (Baraldi et al. 2013a):

Step 1: *Normalized Laplacian Matrix*. Starting from the similarity matrix $\overline{\overline{S}}$, the degree matrix $\overline{\overline{D}}$ is calculated, whose entries $d_1, d_2, \ldots, d_N$ are:

$$d_i = \sum_{j=1}^{N} S_{ij}, i = 1,2,...,N \quad \text{(A1)}$$

Based on $\overline{\overline{D}}$, the normalized Laplacian matrix $\overline{\overline{L}}_{rs}$, is calculated:

$$\overline{\overline{L}}_{rs} = \overline{\overline{D}}^{-1}\overline{\overline{L}} = \overline{\overline{I}} - \overline{\overline{D}}^{-1}\overline{\overline{S}} \quad \text{(A2)}$$

where $\overline{\overline{L}} = \overline{\overline{D}} - \overline{\overline{S}}$ and $\overline{\overline{I}}$ is the identity matrix of size [*N*, *N*].

Step 2: *Eigenvalues and eigenvectors of* $\overline{\overline{L}}_{rs}$. Given $\overline{\overline{L}}_{rs}$, compute the eigenvectors $\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_N$. The first *C* eigenvalues are such that they are very small whereas λ_{C+1} is relatively large (Ng, Jordan, & Weiss, 2001; Von Luxburg, 2007; Zhao & Liu, 2007).

Step 3: *Number of clusters*. The number of clusters is set equal to *C*, according to the Eigengap heuristic theory (Mohar, 1997).

Step 4: *Feature extraction*. The relevant information on the structure of the matrix $\overline{\overline{S}}$ is obtained by considering the eigenvectors $\bar{u}_1,\bar{u}_2,\ldots,\bar{u}_N$ associated to the C smallest eigenvalues of its laplacian matrix $\overline{\overline{L}}_{rs}$. The square matrix $\overline{\overline{S}}$ is transformed into a matrix $\overline{\overline{U}}$ of size [N, C], in which the C columns of $\overline{\overline{U}}$ are the eigenvectors (Von Luxburg, 2007).

Appendix A.2 Silhouette validity index

To evaluate the optimal number of clusters C^* among several clusters candidates, Silhouette validity index has been adopted. The silhouette value for the i-th datum, i=1,…,N, is a measure of how similar/dissimilar that datum is to others in its own cluster and to the other clusters, respectively. The silhouette value for the i-th datum S^i is defined as (Rousseeuw, 1987):

$$S^i = \left(b_i - a_i\right) / \max\left(a_i, b_i\right) \tag{A3}$$

where a_i is the average distance from the i-th datum to the others in the same cluster, and b_i is the minimum average distance from the i-th datum to the others in a different cluster, minimized over clusters.

The mean of the silhouette values for the m-th cluster C_m is called the cluster mean silhouette and is denoted as S_m (Eq. (A4)):

$$S_m = \frac{1}{n_m} \sum_{i \in C_m} S^i \tag{A4}$$

where n_m is total number of data in the m-th cluster. Finally, the global silhouette index $SV_{C_{candidate}}$ is the mean of the mean silhouettes (Eq. (A5)) through all the clusters.

$$SV_{C_{candidate}} = \frac{1}{C_{candidate}} \sum_{m=1}^{C_{candidate}} S_m \tag{A5}$$

The Silhouette value ranges from -1 to +1. A high Silhouette value SV_{C^*} indicates that the C^* clusters of the final consensus clustering are well separated and compacted.