HAL
open science

Clustering for unsupervised fault diagnosis in nuclear turbine shut-down transients

Piero Baraldi, Francesco Di Maio, Marco Rigamonti, Enrico Zio, Redouane Seraoui

To cite this version:

Piero Baraldi, Francesco Di Maio, Marco Rigamonti, Enrico Zio, Redouane Seraoui. Clustering for unsupervised fault diagnosis in nuclear turbine shut-down transients. Mechanical Systems and Signal Processing, 2015, 58-59, pp.160-178. 10.1016/j.ymssp.2014.12.018 . hal-01265643

HAL Id: hal-01265643

https://hal.science/hal-01265643v1

Submitted on 1 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CLUSTERING FOR UNSUPERVISED FAULT DIAGNOSIS IN NUCLEAR TURBINE SHUT-DOWN TRANSIENTS

Piero Baraldi[1], Francesco Di Maio[1], Marco Rigamonti[1], Enrico Zio[1,2], Redouane Seraoui[3]

[1]*Energy Department, Politecnico di Milano*
Via Ponzio 34/3, 20133 Milano, Italy
piero.baraldi@polimi.it *francesco.dimaio@polimi.it* *enrico.zio@polimi.it*

[2]*Chair on Systems Science and the Energetic challenge*
Ecole Centrale Paris and Supelec
enrico.zio@ecp.fr enrico.zio@supelec.fr

[3]*EDF-R&D\STEP*
Simulation et Traitement de l' information pour l' exploitation des systèmes de production,
6 quai Watier, Chatou Cedex, France

ABSTRACT

Empirical methods for fault diagnosis usually entail a process of supervised training based on a set of examples of signal evolutions "labeled" with the corresponding, known classes of fault. However, in practice, the signals collected during plant operation may be, very often, "unlabeled", i.e., the information on the corresponding type of occurred fault is not available. To cope with this practical situation, in this paper we develop a methodology for the identification of transient signals showing similar characteristics, under the conjecture that operational/faulty transient conditions of the same type lead to similar behavior in the measured signals evolution. The methodology is founded on a feature extraction procedure, which feeds a spectral clustering technique, embedding the unsupervised Fuzzy C-Means (FCM) algorithm, that evaluates the functional similarity among the different operational/faulty transients. A procedure for validating the plausibility of the obtained clusters is also propounded based on physical considerations. The methodology is applied to a real industrial case, on the basis of 148 shut-down transients of a Nuclear Power Plant (NPP) steam turbine.

Key Words:

Fault Diagnosis, Steam Turbine, Nuclear Power Plant, Unsupervised Clustering, Fuzzy-based Slope Analysis, Fuzzy Similarity, Spectral Clustering, Fuzzy C-Means.

1. Introduction

Methodological approaches have been proposed for fault diagnosis in components of Nuclear Power Plants (NPPs) [Cheon et al., 1993; Kim et al., 1996; Reifman, 1997; Zio et al., 2006; Di Maio et al., 2011]. These are attractive alternatives for tackling the complexity of the fault diagnosis task when the phenomena that determine the equipment behavior show highly non-linear interrelationships between the causes and the signal evolutions. However, application of these approaches is limited in practice because of lack of examples to be used for the classification models (classifiers) training, i.e., the parameters need to be tuned through an iterative process based on a set of examples constituted by signals "labeled" with the corresponding class of fault under which conditions they have been measured [Reifman, 1997; Zio et al., 2007]. These "labeled" examples can be difficult to be collected even if a model is available and commonly used for Probabilistic Safety Assessment (PSA), because i) some failures can be "a priori" unknown and thus cannot be simulated [Di Maio et al., 2014] or ii) data collected during plant faulty conditions do not have any fault class information (or it is missing), i.e., plant operation is "unlabeled".

The objective of this work, that evolved from the particular industrial need of Electricitè de France (EDF), is to develop a methodology for the identification of "unlabeled" transients originated by operational/faulty transient conditions of the same class. The work is based on the conjecture that these different classes of conditions lead to similar behaviors of the measured signals. The outcomes of the developed method can thus be used as "labels" for a supervised classifier whose knowledge is enriched by the "labels" provided on previously unseen failures (and thus impossible to be modeled).

In general terms, the methods to do this can be divided into two categories: model-based and pattern recognition techniques [Venkatasubramian et al., 2003, Sheng et al., 2004]. Model-based methods need expert specific domain knowledge for building the classification model; in practice, they may be difficult to develop and have difficulties in providing a timely recognition of operational anomalies [Tian et al., 2007]. On the contrary, pattern recognition methods seem to be able to satisfy a number of practical requirements, such as short calculation time and high classification accuracy [Zio, 2007].

For this reason, in this work we focus on the development of a new pattern recognition method for distinguishing different signal evolution behaviors due to different operational/faulty transients, which are "a priori" unknown. Therefore, the problem is one of unsupervised classification/clustering, in which the vectors of measured signal values need to be partitioned into a number of homogeneous clusters so that those vectors belonging to the same cluster are very similar to each other and dissimilar to those of the other clusters. The final target is to be able to label these

families (clusters) of different behavior and to use them as the reference trajectories for a supervised classifier, thus reducing the time for fault detection and diagnosis and facilitating the decision on how to intervene to avoid or mitigate the consequences of the fault [Zio et al., 2010c].

The methodology is based on the typical tasks of i) collecting data, ii) extracting and selecting the characteristic features from these and iii) classifying the data [Di Maio et al., 2011].

Several techniques of feature extraction can be used, such as statistical indicators [Anker et al., 2003], Haar wavelet transforms [Subrumani et al., 2006], pointwise difference [Baraldi et al., 2013a], correlation [Guyon et al., 2003; Baraldi et al., 2011], Principal Component Analysis (PCA) [Baraldi et al., 2010], but they show problems when dealing with misaligned signals [Secchi et al., 2008], i.e., amplified and/or delayed transients data. This work is concerned with the clustering of misaligned transient data.

In this work, we originally propose a Fuzzy-based slope analysis for feature extraction, as an extension of the initial proposal in [Baraldi et al., 2013b]. The rationale behind this technique is that functional behaviors of transients can be approximated by fuzzy sets (terms) such as high/low negative slope, high/low positive slope. Fuzzy logic is exploited to handle the imprecise linguistic concepts, such as “low” and “high”.

The methodology is applied to 70 signals from 148 shut-down transients of a real nuclear steam turbine, by measuring their fuzzy similarity [Zio et al., 2010a] with respect to the features extracted by the Fuzzy-based slope analysis. A similarity graph [von Luxburg, 2007] is built, in which each vertex represents a transient and the weight associated to the edge connecting two vertices is the value of (fuzzy) similarity between the two corresponding transients. A spectral clustering technique, embedding the Fuzzy C-Means (FCM) algorithm, is finally applied in order to find an optimal partition of the graph [von Luxburg, 2007; Baraldi et al., 2013a].

In order to discuss the obtained results and their plausibility based on physical considerations, we look at the representation of the identified clusters in feature spaces completely different from those which have driven the clustering: the rationale is that if in alternative feature spaces the transients tend to form clusters similar to those identified, this would be a confirmation of the correctness of the identified clusters.

The rest of the paper is organized as follows: Section 2 states the problem; in Section 3, the proposed methodology is described in details; in Section 4, the methodology is applied to the case study of real shut-down transients; in Section 5, the results are discussed and validated; finally, in Section 6, some conclusions and remarks are drawn.

2. Problem statement

Let us assume that the values of Z signals at different times have been measured during N plant transients originated by C different classes of behavior, e.g., different operational conditions, different faults. In practice, the generic i-th transient can be seen as a transient into the Z-dimensional signal space and represented by the matrix of values $\overline{\overline{X}}^{i}$ whose component x_{lk}^{i} represents the value of signal k taken at time t_l, $k = 1,...,Z$ and $l = 0,...,T_i - 1$, where T_i is the number of available measurements for the i-th transient.

Without loss of generality, we assume that:

- measurements are taken from an arbitrary time t_0;
- measurements are taken at fixed time steps, Δt (i.e., corresponding the smallest sampling rate of the measured signals); thus, $t_l = l \cdot \Delta t$;
- all transients have the same lenght $(T-1)\cdot \Delta t$, i.e., $T_i = T$, for any i=1,…,N.

These restrictions (i.e., different sampling rates and different transients length) can be easily overcome by properly applying time warping techniques [Park et al., 2000] to relax the assumptions done in this work.

The objective of the present work is to partition the N transients $\overline{\overline{X}}^{i}$ into an unknown number of clusters, C, each one containing transients of similar behavior, that should mean that they have been originated by the same physical causes (operational conditions or faults).

3. Methodology

The methodology here proposed for clustering operational transients is based on spectral clustering [Strang et al., 1996]. The main characteristic of spectral clustering is that it allows partitioning objects (in our case, vectors of measured signals) into clusters by using a measure of similarity between them. A similarity graph $G = (V, E)$ is introduced, in which each vertex v_i in the graph represents an object and a weight is associated to each edge p_{ij} connecting vertices i and j, to measure the similarity between objects i and j [von Luxburg, 2007]. Clustering aims at finding a partition of the graph such that the edges between elements belonging to different groups of the partition have small weights (which means that objects in different clusters are dissimilar from each other) and the edges connecting elements within the same group have large weights (which means that objects within the same cluster are similar to each other) [Alpert et al., 1999].

Section 3.1 illustrates the feature extraction technique proposed to measure the similarity among transients; Section 3.2 illustrates the details of the similarity measure evaluation, whereas Section 3.3 focuses on the spectral clustering.

3.1 Fuzzy-based slope analysis

The definition of the similarity measure between two transients should consider the functional characteristics of the signal transients, e.g., form, slope, curvature [Joentgen et al., 1999] rather than being based on the magnitude of the signal values, which may be dissimilar due to the presence of outliers, intensity of the faults, plant operational regimes, measurement noise [Angstenberger, 2001].

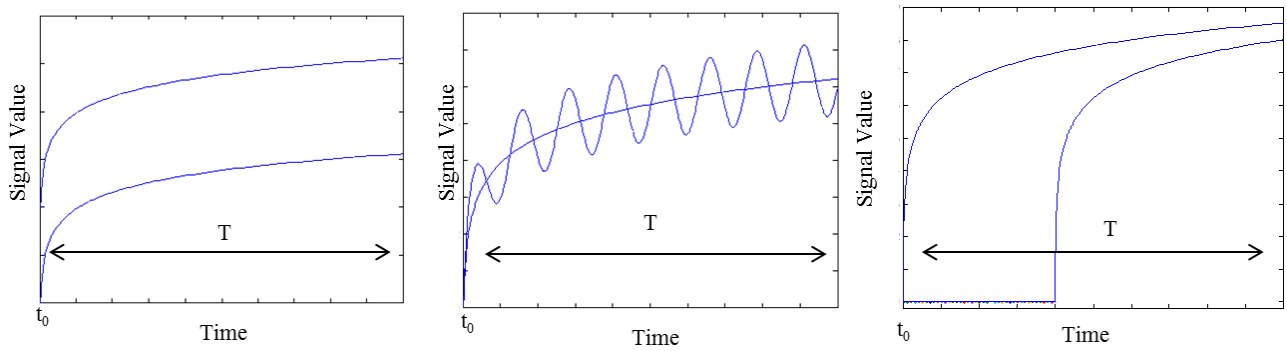

Figure 1. (Left) two transients with the same functional characteristics and different operation regimes; (Centre) two transients with different functional characteristics and same operation regimes; (Right) two misaligned transients with the same operation regimes.

For ease of clarity, Figure 1 (left) shows two normal transients of the same length T at different operational regimes (i.e., characterized by the same functional behaviour and different magnitudes), Figure 1 (right) shows two misaligned transients of the same length T at the same operational regimes (i.e., characterized by the same functional behaviour and magnitudes), whereas Figure 1 (centre) shows a normal transients and a possible associated failed transients (i.e., characterized by different functional behavior) of the same length T and all starting at an arbitrary time t_0. In particular, in this work we deal with misaligned transients, among which we aim at identifying different functional behaviours (independently from the time) by extracting some meaningful feature from the raw signal values.

For extracting functional characteristics from the raw signals, we consider the Fuzzy-based slope analysis feature extraction technique. Specifically developed in this work, this feature extraction technique is based on a fuzzy pre-processing of the data to handle misaligned transients for which several techniques fail to properly extract meaningful information from raw signals [Secchi et al., 2008]. This technique consists in five steps, which are applied to the collected values of each k-th signal, k=1,...,Z, without resorting to any type of further pre-processing:

1. **Signal slope computation:** divide the *i*-th transient of the *k*-th signal in *R* intervals of length $L = \frac{T\Delta t}{R}$ (Figure 2). For each *r*-th interval, the slope ${}^{r}\Delta_{k}^{i}$ of the signal is computed using the least squares regression method [Frank, 1987] .

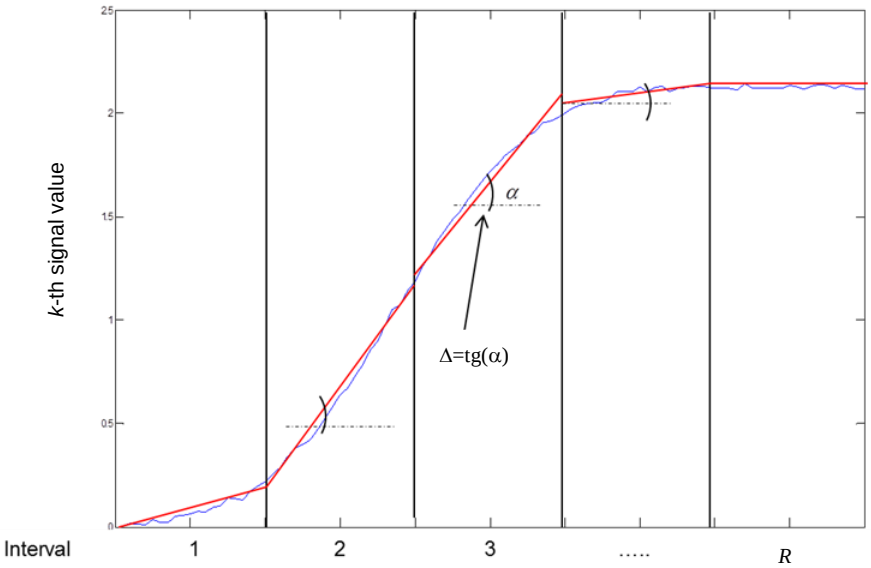

Figure 2: Computation of the slope for each *r-th* interval

2. **Signal slope distribution:** for each *k*-th signal, collect the *RxN* slope values of all the *N* transients and build the signal slope empirical distributions for the positive and negative values.

3. **Percentiles computation:** calculate the 1st and the 50th percentile of the negative slopes distribution (N_1 and N_{50}), and the 50th and the 99th percentile of the positive slopes distribution (P_{50} and P_{99}), as shown in Figures 3 and 4, respectively. These percentile values are considered representative of high and low negative slopes (i.e., decreasing transient), and low and high positive slopes (i.e., increasing transient), respectively.

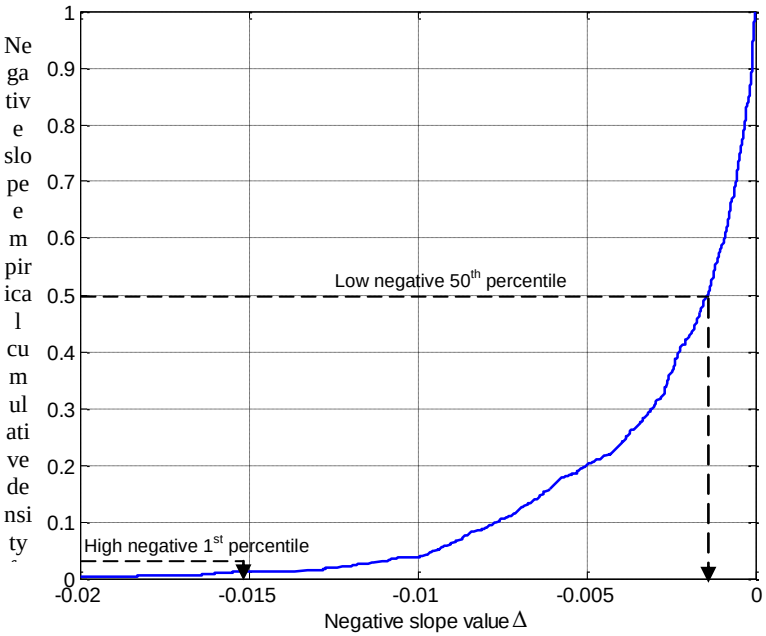

Figure 3: Empirical cumulative distribution of the negative signal slope values

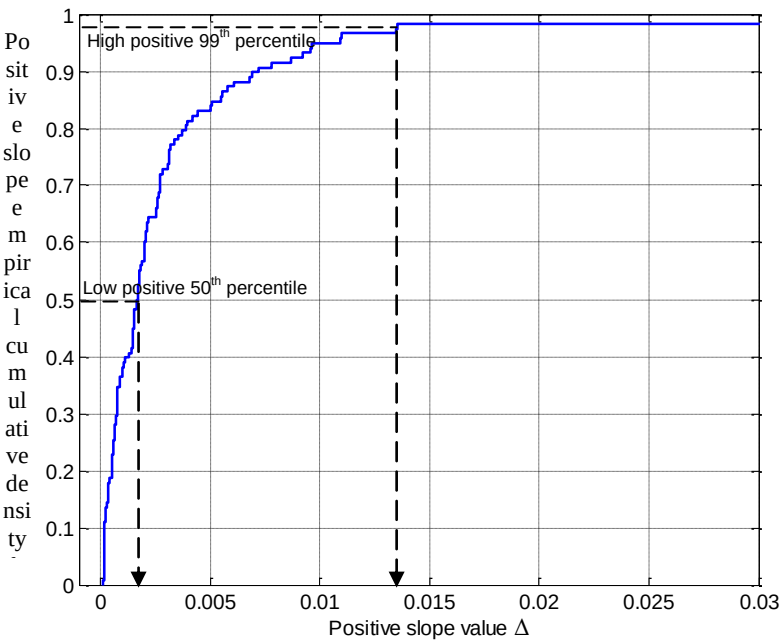

Figure 4: Empirical cumulative distribution of the positive signal slope values

4. **Fuzzy sets and membership functions creation**: consider the signal slope, Δ, as a linguistic variable that can be approximated within a fuzzy framework defined by the fuzzy sets (terms) high negative, low negative, low positive and high positive.

   In this case (shown in Figure 5), the membership functions of the *k*-th signal ${}_s\mu_k$, *s*=1,2,3,4,5, which define the fuzzy sets, are asymmetric and unevenly spaced triangular

functions centered on the percentile values (N_1, N_{50}, P_{50} and P_{99}) computed at step 3 and on the zero value (V_0):

- $$ {}_1\mu_k(\Delta)_{High_Negative} = \begin{cases} 1 & \Delta < N_1 \\ \dfrac{\Delta - N_{50}}{N_1 - N_{50}} & N_1 < \Delta < N_{50} \end{cases} \quad (1) $$

- $$ {}_2\mu_k(\Delta)_{Low_Negative} = \begin{cases} \dfrac{N_1 - \Delta}{N_1 - N_{50}} & N_1 < \Delta < N_{50} \\ \dfrac{\Delta - V_0}{N_{50} - V_0} & N_{50} < \Delta < V_0 \end{cases} \quad (2) $$

- $$ {}_3\mu_k(\Delta)_{Zero} = \begin{cases} \dfrac{N_{50} - \Delta}{N_{50} - V_0} & N_{50} < \Delta < V_0 \\ \dfrac{\Delta - P_{50}}{V_0 - P_{50}} & V_0 < \Delta < P_{50} \end{cases} \quad (3) $$

- $$ {}_4\mu_k(\Delta)_{Low_Positive} = \begin{cases} \dfrac{V_0 - \Delta}{V_0 - P_{50}} & V_0 < \Delta < P_{50} \\ \dfrac{\Delta - P_{99}}{P_{50} - P_{99}} & P_{50} < \Delta < P_{99} \end{cases} \quad (4) $$

- $$ {}_5\mu_k(\Delta)_{High_Positive} = \begin{cases} \dfrac{P_{50} - \Delta}{P_{99} - P_{50}} & P_{50} < \Delta < P_{99} \\ 1 & \Delta > P_{99} \end{cases} \quad (5) $$

We choose triangular functions because they consist of two linear segments joined at the peak, which makes graphical representation and operations very simple, can be constructed easily on the basis of little information and the sum of membership values of the fuzzy sets is 1 for any value [Bojadziev et al., 1995]. We resort to asymmetric, unevenly spaced triangular fuzzy sets because the use of symmetric, evenly spaced triangular fuzzy sets may give inappropriate results when the values are not scattered all over the universe, leading to empty fuzzy partitions [Baldwin et al., 2003].

Once the membership functions have been defined, each computed signal slope can be characterized by its degrees of membership to the fuzzy sets which represent the level to which the slope verifies the extent properties of the set, as shown in Figure 5.

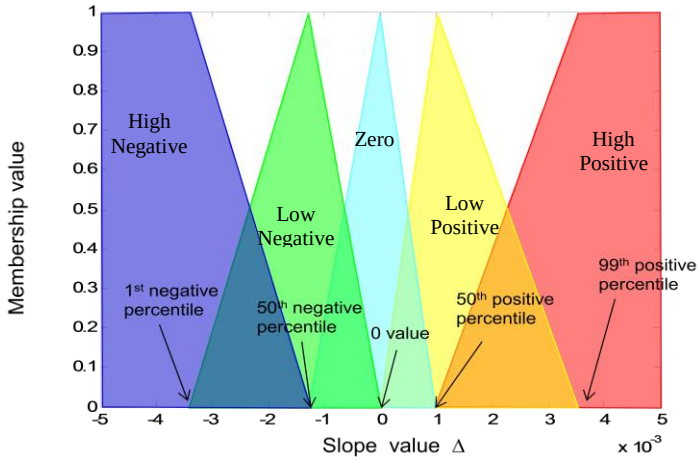

Figure 5: Membership functions of the linguistic variable slope

5. **Feature extraction**: for each *i*-th transient, compute the mean membership ${}_s\overline{\mu}_k^i$ of the *k*-th signal to each of the *S*=5 fuzzy sets:

$$ {}_s\overline{\mu}_k^i = \frac{1}{R}\sum_{r=1}^{R} {}_s^r\mu_k^i \qquad \begin{array}{l} s=1,..,5 \\ i=1,..,N \\ k=1,..,Z \\ r=1,..,R \end{array} \qquad (6) $$

Since five different membership features are obtained for each signal, the matrix $\overline{\overline{X}}^i$ of size [*T*-1, *Z*] describing a transient is transformed into a vector $\overline{Y}^i$ of size 5*Z* which constitutes the new representation of transient *i*, as shown in Figure 6.

It is worth pointing out that this representation i) associates to each transient a time-independent indicator (i.e., the mean membership averaged on time) of the signal slope and ii) normalizes all the *k* signals of each *i*-th transient with respect to the identified fuzzy sets, allowing for a fair comparison of transients *i* and *j* with respect to the mean slope (i.e., the mean membership to the fuzzy set) of all the *k* signals, as we shall see in the following.

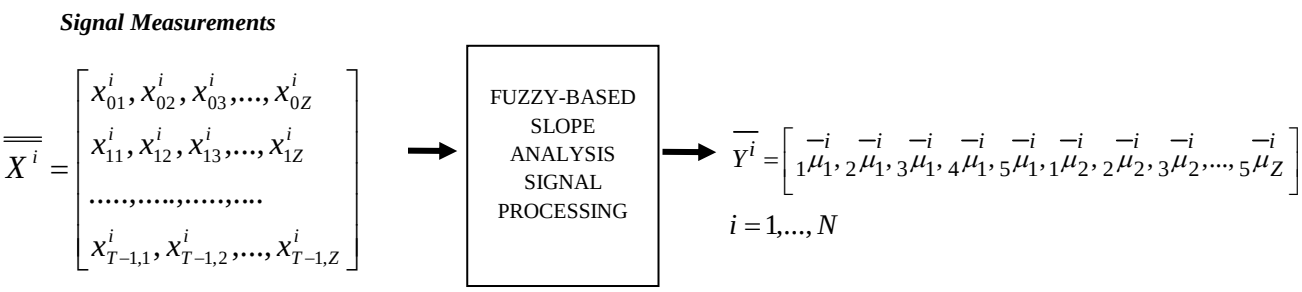

Figure 6: Sketch of the feature extraction technique based on the fuzzy-based slope analysis

3.2 Similarity-based pattern matching

After the data pre-processing, the similarity between transient i and transient j can be computed by considering the corresponding vectors of extracted features $\overline{Y^i}$ and $\overline{Y^j}$. To this aim, a fuzzy similarity measure is considered to determine the degree of closeness of the two transients with reference to the pointwise difference between the $5Z$ mean membership ${}_s\overline{\mu}_k^i$ computed in step 5 of Section 3.1. [Zio et al., 2010a]. Without loss of generality, let us define the pointwise difference δ_{ij} between the transients $\overline{Y^i}$ and $\overline{Y^j}$:

$$\delta_{ij} = \sqrt{\sum_{k=1}^{5Z}\left(y_k^i - y_k^j\right)^2} \qquad (7)$$

The pointwise difference of the two transients is done for each k-th normalized signal and then aggregated with respect to all the Z signals and evaluated with reference to an "approximately zero" fuzzy set (FS) specified by a function which maps δ_{ij} into a value μ_{ij} of membership to the condition of "approximately zero": values of μ_{ij} close to 0 indicate that the signal evolutions in the two transients i and j are very different, whereas values close to 1 indicate high similarity [Zio et al., 2010b].

In this work, the following bell-shaped function is used:

$$\mu_{ij} = e^{-\frac{\delta_{ij}^2}{\sigma^2}} \qquad (8)$$

The arbitrary parameter σ can be set by the analyst to shape the desired interpretation of similarity into the fuzzy set: the larger the value of σ, the narrower the fuzzy set and the stronger the definition of similarity [Zio et al., 2010a].

3.3 Spectral clustering

The computation of the fuzzy similarity between all possible pairs of transients originates the similarity matrix $\overline{\overline{W}}$ of size [N, N], whose generic element μ_{ij} represents the fuzzy similarity between transients i and j. The diagonal components μ_{ij} are set to 1 and the matrix is symmetric $\left(\mu_{ij} = \mu_{ji}\right)$. From the matrix $\overline{\overline{W}}$ a similarity graph $G = (V, E)$ is constructed, where each vertex v_i represents the i-th transient and the weight associated to the edge p_{ij} connecting the two vertices i

and j is the similarity value μ_{ij} [von Luxburg, 2007]. The original problem of identifying groups of similar transients can be reformulated in that of finding a partition of the similarity graph such that the edges connecting elements of different groups have small weights and the edges connecting elements within a group have large weights [Alpert et al., 1999]. The spectral clustering algorithm is based on the Fuzzy C-Means algorithm and the steps are described in details in Appendix A.

4. Application to NPP turbine shut-down transients

In this Section, we apply the proposed fuzzy-based slope analysis to a real industrial case regarding N=148 shut-down transients of a NPP turbine. Let us consider the measured values of $Z=70$ signals taken at $T=4500$ time steps during the N shut-down transients. In practice, the generic i-th transient can be seen as a multidimensional transient into the $Z=70$ dimensional signal space and represented by the matrix of values $\overline{\overline{X}}^{i}$ whose component x_{lk}^{i} represents the value of signal k taken at time t_l, $k=1,...,70$ and $l=0,...,4499$. The objective of the work is to partition the *148* multidimensional transients $\overline{\overline{X}}^{i}$ into an a priori unknown number of clusters, C, each one containing transients characterized by similar functional behavior under the conjecture that plant operational states, working conditions, faults and malfunctioning of the same type lead to similar behavior in the measured signals evolution.

4.1 Data pre-treatment

Each transient consists in 4500 measurements of 70 signals: such a large amount of data in high dimensionality can be troublesome to handle for identifying similarities between transients [Baraldi et al., 2011]. Thus, a correlation matrix of size [70, 70] (Figure 7) has been fed to the spectral clustering algorithm described in Appendix A: we take (high) correlation as good indicator of relevant information for the clustering of the transients. The rationale behind this choice is that in different operational (or failure) conditions the highly correlated signals should behave in the same manner being affected by the same physical causes. This is supposed to facilitate the recognition of possible anomalies in the system behavior that could be shadowed by the erratic evolution of uncorrelated signals.

This procedure has allowed identifying six groups of signals. Each signal is characterized by an high degree of correlation with the signals of the same group and a low degree of correlation with the signals of other groups. In the following analyses, we refer to the largest group of signals that is composed by 27 signals.

Notice the possible limitation of the approach of selecting a subgroup of highly correlated signals and considering only them for the identification of clusters of transients: since we are using only a part of the available information, we may not be able to identify operational conditions, failures or anomalies which affect signals other than those considered. Thus, for a complete analysis, the same procedure of unsupervised clustering of the transients described in Section 4.3 should be sequentially applied to other subgroups of the 70 signals.

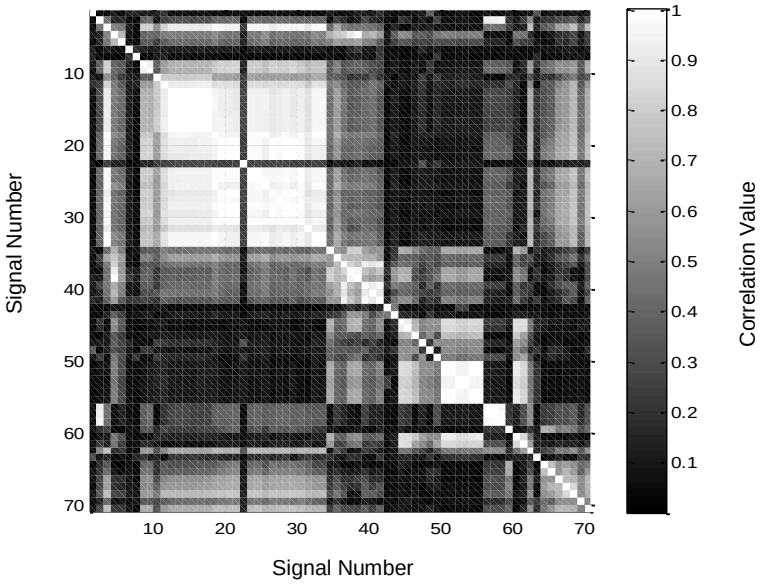

Figure 7: Correlation map between the 70 signals

4.2 Application of the clustering based on the Fuzzy slope analysis

The fuzzy-based slope analysis technique is applied to the transients $\overline{\overline{X}}^{i}$ [4500, 27]. Using Eq.(6), we obtain $\overline{Y}^{i}$ [135] for each i-th transient, i=1,2,...,148. Then, applying eq.(7) and eq.(8), we construct the similarity matrix $\overline{\overline{W}}$ of size [148,148] represented in Figure 8: the larger the similarity μ_{ij}, the brighter the shade of the ij-th element of the matrix. The similarity values (eq.(8)) have been obtained with an optimized value of the bell-shaped function parameter σ=0.55.

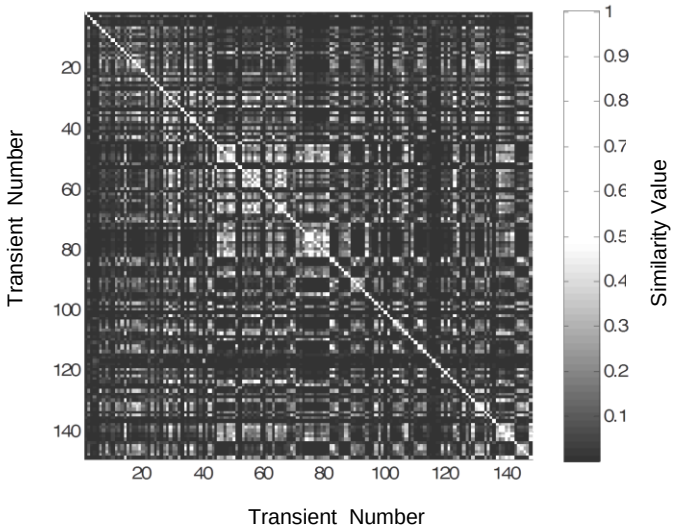

Figure 8: The similarity matrix $\overline{\overline{W}}$: the larger the similarity, the brighter the shade of the *ij*-th element

Figure 9 shows the 148 eigenvalues obtained by applying the proposed spectral analysis method to matrix $\overline{\overline{W}}$, as described in Appendix A: according to the eigengap theory (von Luxburg, 2007) (i.e., the first four eigenvalues are closer to zero and the gap between the fourth and the fifth is the largest), the number of clusters *C* is set equal to 4.

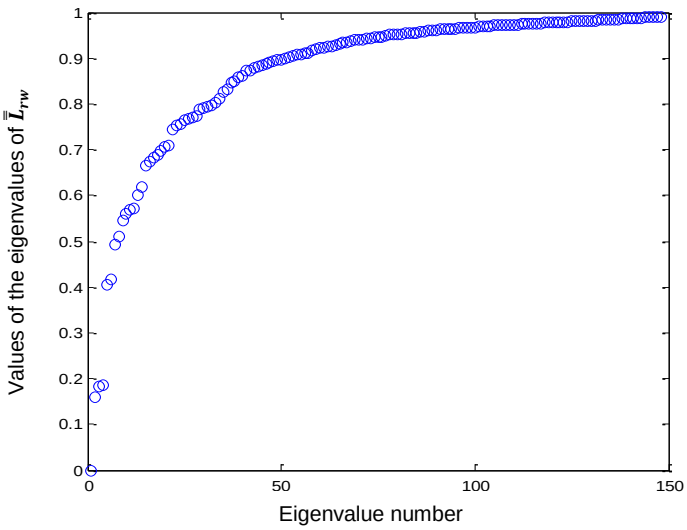

Figure 9: The 148 eigenvalues of L_{rw}

The relevant information on the structure of the matrix $\overline{\overline{W}}$ is obtained by considering the eigenvectors $\overline{u}_1, \overline{u}_2, \ldots, \overline{u}_C$ associated to the *C* smallest eigenvalues of its laplacian matrix $\overline{\overline{L}}_{rw}$ (eq.

(2a) in Appendix A). The square matrix $\overline{\overline{W}}$ is transformed into a reduced matrix $\overline{\overline{U}}$ of size [N, C], in which the C columns of $\overline{\overline{U}}$ are the eigenvectors $\bar{u}_1, \bar{u}_2, ..., \bar{u}_C$ (Step 2 in Appendix A).

The problem of clustering the 148 transients $\overline{\overline{X}}^i$ is finally reduced to the problem of finding four clusters among the 148 4-dimensional vectors where, for each i-th transient, $\bar{u}^i = \{u_1^i, u_2^i, ..., u_C^i\}$ constitutes a reduced representation of $\overline{\overline{X}}^i$. It is important to point out that the first eigenvector is constant and consequently does not carry any information about the clustering of the data.

The FCM partitioning of the obtained vectors $\bar{u}^i$ provides the memberships μ_{ic} of the i-th transient, i=1,2,3,...,148, to the c-th cluster, c=1,2,3,4. The transient belonging to a cluster with the largest value of membership is called prototypical transient, and its functional behavior can be taken as most characteristic of the cluster.

4.3 Analysis and interpretation of the results

The matrix of values μ_{ic} containing the degree of membership of each transient, i=1,...,148 to each cluster c=1,...,4 has been analyzed for distinguishing two categories of transients:

- Operational transients: transients assigned to one cluster with large confidence (max (μ_{ic}) $\geq$ 0.7).
- Malfunctions: transients not assigned to any cluster with enough confidence (max (μ_{ic}) < 0.7).

We firstly consider the representation of the identified clusters of transients in the feature space of the eigenvectors u_2, u_3 and u_4 identified in Section 4.2, which are at the basis of the clustering of the transients obtained (remember that u_1 is constant and, thus, is useless to the clustering task). Figure 10 shows i) a clear separation between the transients assigned to clusters 2 and 3 and the remaining transients, and ii) transients of clusters 1 and 4 seem to be slightly separated, although there is a zone of overlapping between them. This overlapping zone is due to the fact that the two involved clusters do not greatly differ in the space of the first 3 eigenvectors of L_{rw}: the transition from one cluster to the other is smooth and, thus, the assignment to any of the cluster is not robust (i.e., transients in the overlapping zone are expected to be characterized by a border-line functional behaviour, due to the sharing of some functional characteristics of both the identified clusters (as it will be shown in Section 4.3.1)).

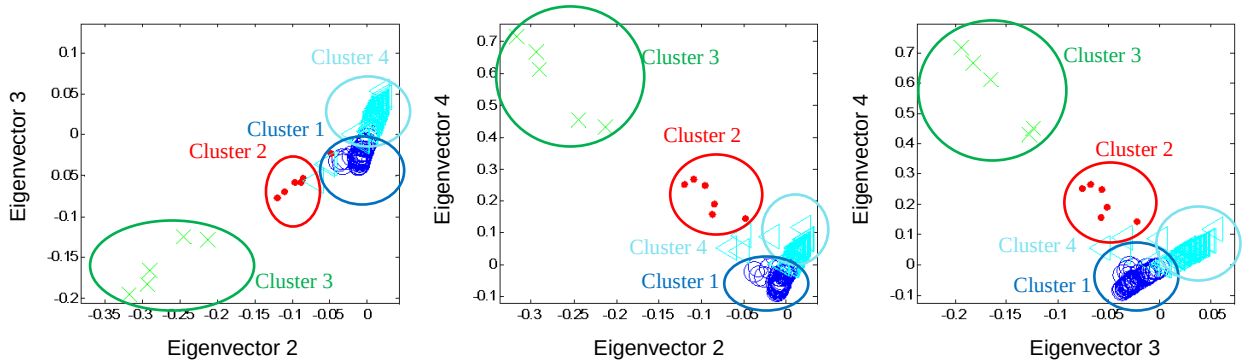

Figure 10: Representation of the clusters in the space of the first 3 eigenvectors of L_{rw}

In what follows, we critically discuss the obtained results and their plausibility based on physical considerations, and engineering and expert knowledge.

Figure 11 shows the evolution of signal 3 for the transients assigned to one of the four identified clusters with large confidence. These transients are confirmed by EDF experts to be due to shut-down transients occurred in nominal operational conditions: differences may be due to different working conditions (e.g., loads, tests,…). It is worth noticing that all the transients belonging to the same cluster can be considered with similar shape but lagged in time. Thus, we show the capability of the method in dealing with misaligned transients that have the same functional characteristics as in Figure 1 (right).

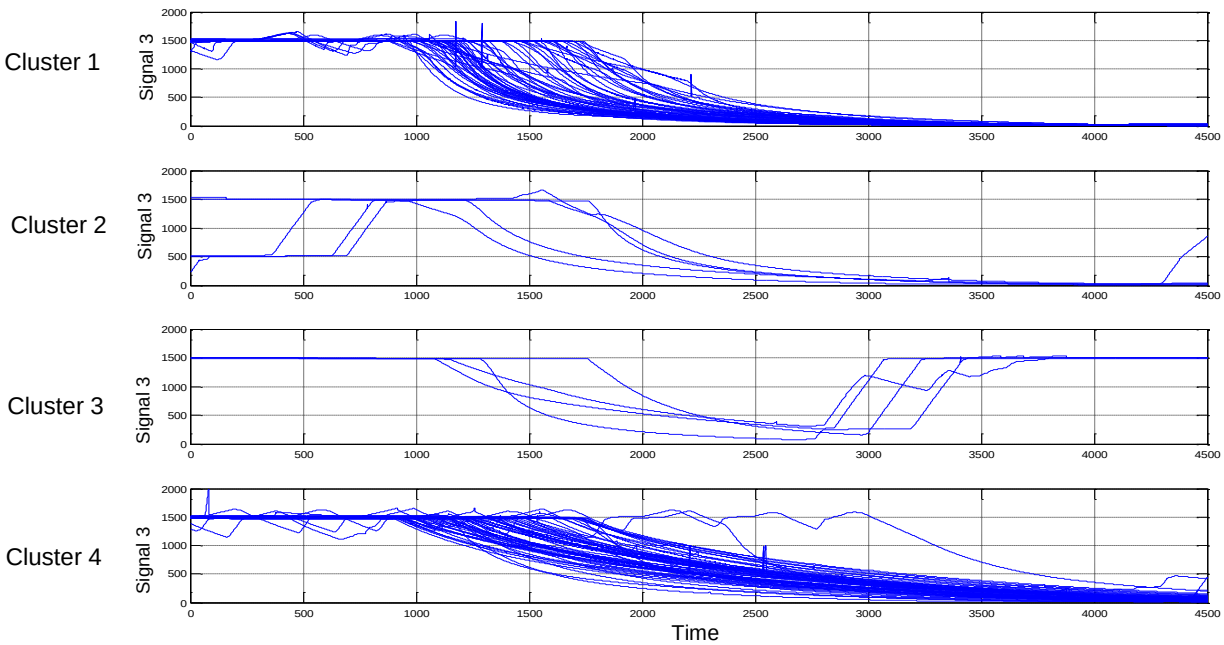

Figure 11: Evolutions of the signal 3 in the transients assigned to the four clusters

Clusters 1 and 4 mainly differ in the rate of decrease of the signals: those of transients belonging to cluster 1 are characterized by a sharper, less smooth decrease than those of cluster 4. These transients are representative of tests of successful emergency shut-down of the turbine and planned shut-down transients for preventive maintenance, respectively [Leyzerovich, 1997]. Furthermore, it is possible to note that, even if the distinction between the more characteristic transients belonging to cluster 1 and cluster 4 is clearly marked, there are several transients at the border between the two clusters for which the signal behavior seems to be very similar. On the contrary, clusters 2 and 3 show peculiar functional behaviors which allow distinguishing these transients from the others. Those belonging to cluster 2 are characterized by a low mean value of the signal due to the majority of the operational time spent at low regime (i.e., long period of time characterized by small signal values and short period of time characterized by large signal values): these are representative of a test where a planned and slow shut-down transient follows a cold start-up transient, which is characterized by an initial plateau necessary to warm up the turbine structural material in order to avoid destructive thermal stresses [Brink et al., 1994]. Conversely, those belonging to cluster 3 are characterized by a large mean value of the signal due to the majority of the operational time spent at high regime (i.e., long period of time characterized by large signal values and short period of time characterized by small signal values: these are representative of a test where an emergency shut-down is followed by an hot start-up, which is faster than the cold start-up because there are less problem connected with thermal stresses induced in the structural materials of the turbine [Brink et al., 1994]. For ease of comprehension, the membership values of the prototypical trajectories of each cluster are reported in Table I, whereas the membership values of the operational transients assigned with the lowest degree of membership are reported in Table II.

TABLE I: degrees of membership of the prototypical trajectories of each cluster

Transient number	**Degree of membership to cluster 1(μ_{i1})**	**Degree of membership to cluster 2(μ_{i2})**	**Degree of membership to cluster 3(μ_{i3})**	**Degree of membership to cluster 4(μ_{i4})**	**Type of transients**
147	**0,9993**	0,0001	0	0,0006	Prototype 1
5	0,0078	**0, 9758**	0,0024	0,0140	Prototype 2
44	0,0009	0,0020	**0,9961**	0,0010	Prototype 3
141	0,0006	0,0002	0	**0,9992**	Prototype 4

TABLE II: degrees of membership of the operational transients assigned with the lowest degree of membership to each cluster

Transient number	Degree of membership to cluster 1(μ_{i1})	Degree of membership to cluster 2(μ_{i2})	Degree of membership to cluster 3(μ_{i3})	Degree of membership to cluster 4(μ_{i4})	Assigned Cluster
27	**0,7015**	0,0238	0,0029	0, 2718	1
116	0,0417	**0,8595**	0,0331	0,0657	2
33	0,0507	0,1806	**0,7045**	0,0642	3
56	0,2356	0,0255	0,0029	**0,7360**	4

4.3.1 Analysis of the malfunctions

Table III reports the transients that have not been assigned to any cluster with enough confidence, i.e., their degrees of membership to all four clusters are lower than 0.7.

TABLE III: Not assigned transients (In bold, the largest degree of membership).

Transient number	Degree of membership to cluster 1(μ_{i1})	Degree of membership to cluster 2(μ_{i2})	Degree of membership to cluster 3(μ_{i3})	Degree of membership to cluster 4(μ_{i4})	Type of transients
16	**0,674**	0,025	0,003	0,298	Failure
21	**0,564**	0,028	0,003	0,404	Failure
100	**0,286**	0,251	0,191	0,272	Failure
106	**0,680**	0,025	0,003	0,292	Border-line
35	0,099	**0,634**	0,015	0,253	Border-line
60	0,066	0,262	**0,586**	0,086	Border-line
3	0,339	0,245	0,016	**0,400**	Failure
4	0,202	0,314	0,014	**0,469**	Border-line
7	0,314	0,029	0,003	**0,653**	Border-line
39	0,298	0,028	0,003	**0,671**	Border-line
43	0,457	0,030	0,003	**0,510**	Border-line
54	0,407	0,030	0,003	**0,560**	Border-line
96	0,450	0,035	0,004	**0,511**	Border-line
115	0,462	0,036	0,004	**0,499**	Failure
128	0,356	0,151	0,011	**0,482**	Failure

A degree of membership $\mu_{ic} < 0.7$ can be due to:

- Border-line effects: transients belonging to more than one cluster, because being at the border of two clusters they share some characteristics with both clusters (may be due to border-line operational conditions).

- Failures: transients different from those belonging to the four identified clusters, i.e., they do not share any characteristics with any transients and may be due to failures occurred during the transients.

In order to identify whether these transients are “border-line” or “failures” avoiding inconsistency phenomena and giving robust results, we have developed an Auto-Associative Kernel Regression model (AAKR) [Baraldi et al., 2011] that is trained on a set of examples formed by the patterns of the “operational” transient (Figure 12): this method has been chosen because, even though computationally intensive, it has been specifically requested by EDF due to its prompt availability, “out-of-the-shelf”, and qualified (within EDF environment) characteristics.

When the “malfunctions” transients are fed to the developed AAKR model, two cases may arise: the signal reconstructions (output of the AAKR) are similar to the signal behaviour of the “malfunction” transient (input to the AAKR) or they are different. In the former case, we can conclude that the “malfunction” transient is similar to those used for the model training, i.e., the transients of the four clusters, and, thus, the transient is at the border of two clusters. Contrarily, in case of remarkable difference between the reconstructed and “malfunction” transients, we can conclude that the transient is atypical with respect to the training transients and, thus, it is a “failure” characterized by anomalous signal behaviors. According to this analysis, 6 out of the 15 transients are “failures”, whereas the remaining 9 transients are “border-line” operational transients (see Table III, last column).

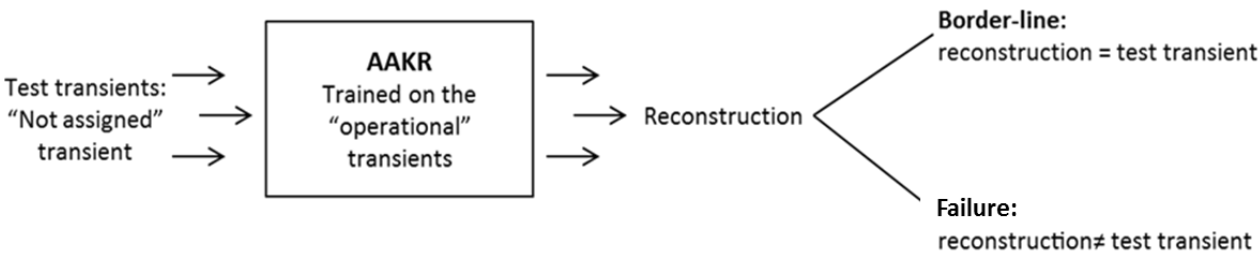

Figure 12: Sketch of the fault detection method based on the Auto-Associative Kernel Regression (AAKR) model

To provide a visual and heuristic interpretation of the results given by the application of the AAKR we have also performed a multivariate boxplot analysis of the values of the eigenvectors characterizing the “not assigned” transients. This generally confirms the results provided by the AAKR. For clarity of representation, Figure 13 shows only the monovariate boxplot of the values of the eigenvector 3 for transients associated to clusters 1 and 4.

In general terms, it is possible to notice that: i) operational transients belonging to cluster 4 are characterized by larger values of the Eigenvector 3 with respect to cluster 1; ii) border-line transients associated to cluster 1 (circles in Figure 13) and border-line transients associated to cluster 4 (triangles in Figure 13) lie within (or very close to) the whiskers of the corresponding operational cluster (1 and 4, respectively), highlighting the typical sharing of the physical characteristic of border-line transients; iii) failure transients (represented by the crosses and the stars, depending on the cluster associated with the largest degree of membership) fall out of the whiskers (i.e., are outliers) of the distribution of the most similar cluster, highlighting a different functional behaviour from the operational transients.

Despite this, it has to be pointed out that, in some cases, the boxplot analysis for the outlier identification is not as straightforward as the AAKR when the number of eigenvectors to be treated increases, because monovariate boxplots of different eigenvectors might provide conflicting results, such as for one of the border-line transients belonging to cluster 4 (triangle of Figure 13 which lies outside the whiskers of the boxplot of cluster 4 and within the whiskers of the boxplot of cluster 1): with respect to a monovariate boxplot analysis, this should be classified as failure (in conflict with the results provided by the AAKR and reported in Table III, last column); however, by a non-trivial multivariate analysis of the boxplots of the other eigenvectors, it is possible to correctly identify the border-line characteristic of the considered transient.

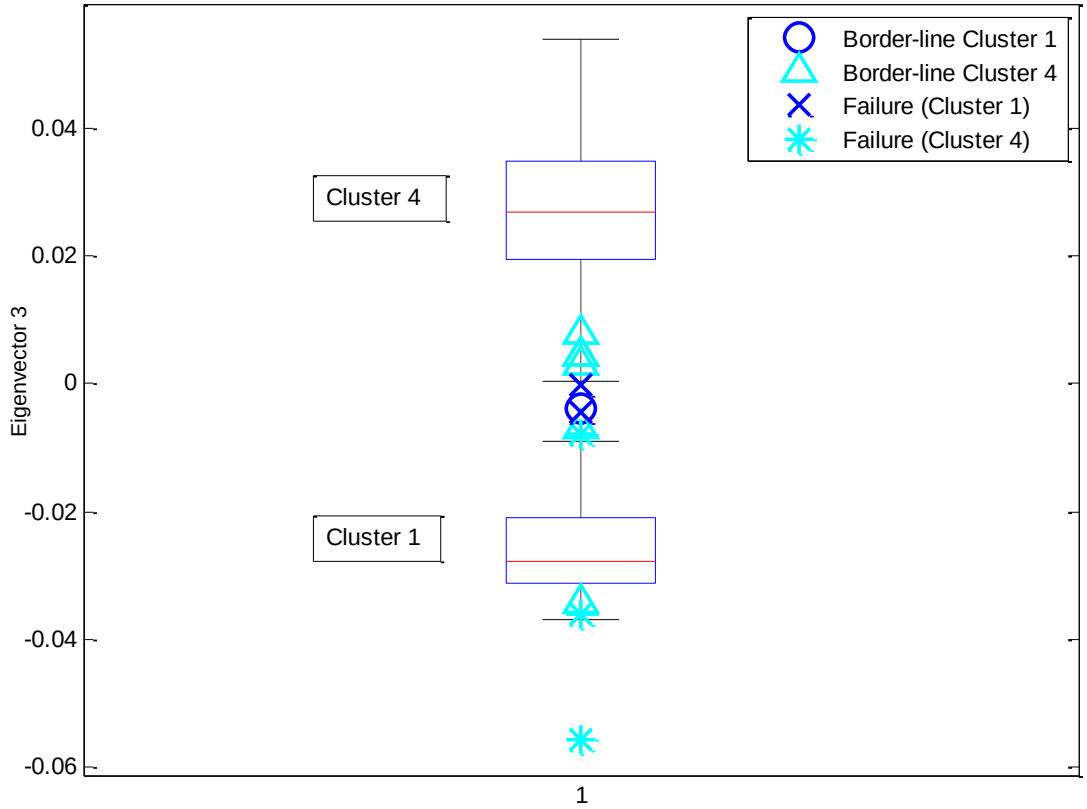

Figure 13: Boxplots of the values of Eigenvector 3 with respect to the transients associated to cluster 1 and cluster 4

As an example of "border-line" transients, let us consider transient 106: according to Table III, the two clusters to which this transient belong with the largest degrees of membership are clusters 1 and 4. Figure 14 shows the evolutions of signals 3, 24 and 32 for this transient (dashed line) and compare them with the evolutions of the signals in the prototypical transients of clusters 1 and 4 (continuous and dotted lines, respectively). Notice that the signal evolutions of the "border-line" transients are in the middle of the signal evolution of the prototypical transients and, therefore, the rate of the decreasing part of the signal cannot be univocally defined as "sharp" or "smooth", as it is for clusters 1 and 4, respectively.

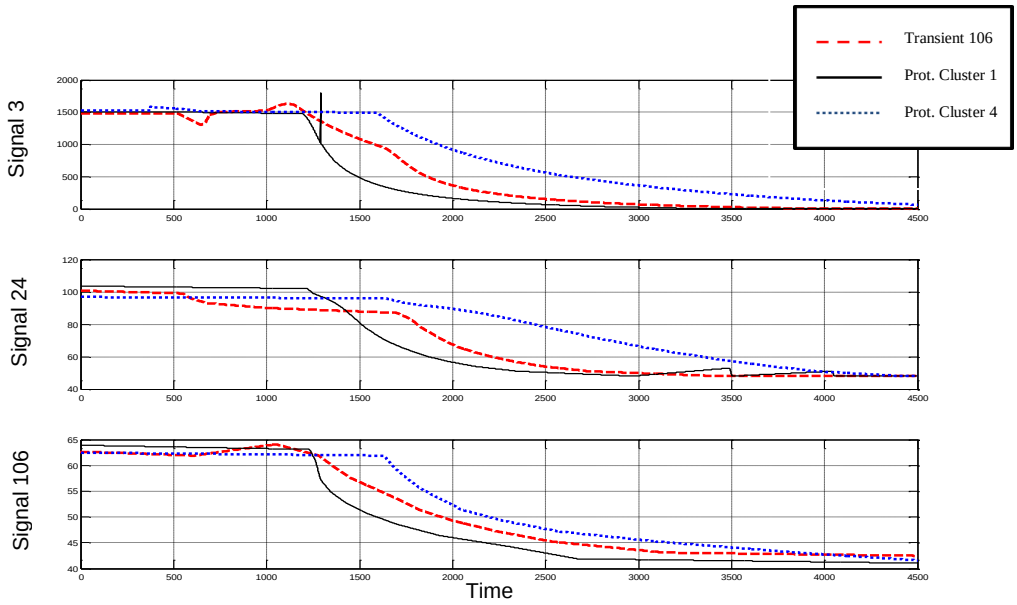

Figure 14: Evolution of signals 3, 24 and 32 in transient 106, and in the prototypical transients of clusters 1 and 4

On the other hand, Figure 15 shows the evolutions of signals 11, 13 and 17 in the "failure" transient 100: notice that the difference between the functional behavior of the signals in this transient and in the prototypical transient of the most similar cluster is remarkable. This anomalous behavior has been confirmed by EDF experts to be due to failures occurred during the transient (whose characteristics cannot be discussed, due to confidentiality reasons).

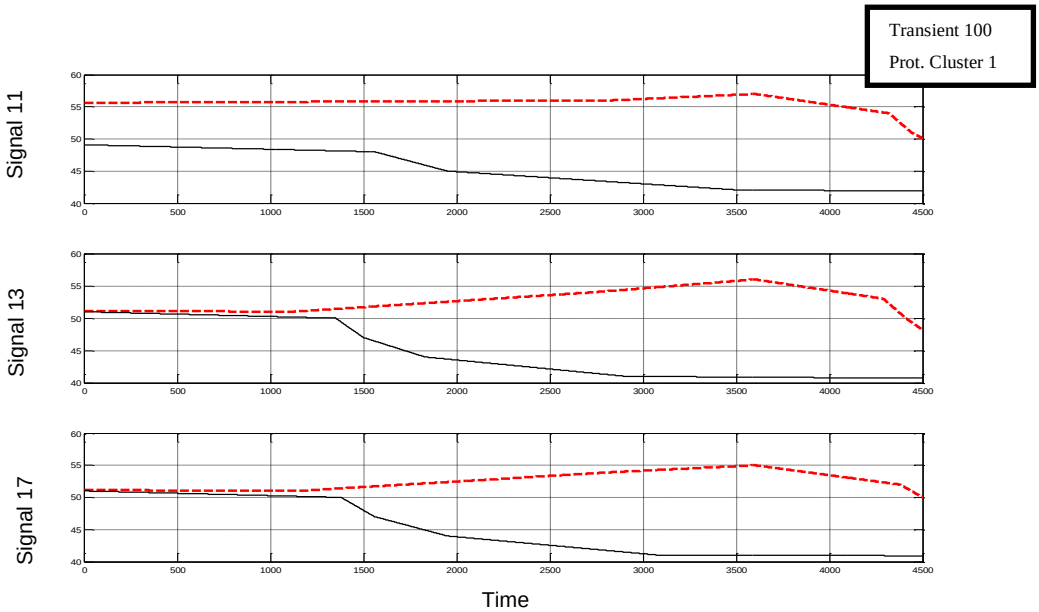

Figure 15: Evolutions of signals 11, 13 and 17 in transient 100 and in the prototypical transient of cluster 1

5. Validation of the results

We consider the representation of the identified clusters of transients (Figure 10) in feature spaces completely different from those which have driven (supervised) the clustering: the rationale behind this is the fact that if in alternative feature spaces the transients tend to form clusters similar to those identified by the applied methodology, then the correctness of the identified clusters would be further confirmed. In particular, we consider two different feature spaces, one formed by statistical indicators of the signal evolutions [Anker eta al., 2003] and the other by the Haar wavelet transforms [Subrumani et al., 2006].

In the feature space of the statistical indicators, each original transient $\overline{\overline{X}}^i$ [4500, 27] is transformed into $\overline{Y}^i$ composed by 81 features. In practice, the 4500 measurements of the *k*-th signal in transient *i* are substituted by their mean m_k^i, standard deviation σ_k^i and skewness s_k^i values.

Similarly, in the feature space of the wavelet transforms, each original transient $\overline{\overline{X}}^i$ [4500, 27] is transformed into a vector, $\overline{Y}^i$, of 81 features. In this case, the features (three for each signal) which catch the evolution of a signal in a transient are: the mean value of the signal (w_1), the maximum wavelet coefficient over all the scales (w_2), the minimum wavelet coefficient over all the scales (w_3). The application of Haar wavelet transforms requires setting the window size T_s equal to a wavelet dyadic decomposition value (i.e., powers of 2): in this case, T_s is set to 4096. In order to limit at minimum any distortion that may be introduced by the compression from 4500 to 4096 measurements, we *i*) divide the first 4444 measurements into 404 intervals of length equal to 11, *ii*) compress the signals by removing the 11th measurement of every interval, *iii*) add the last 56 measurements of the original transients (steady-state condition), *iv*) obtain transients with the desired length of 4096 measurements.

Both feature spaces are formed by 81 features and it is difficult to provide representations of the patterns $\overline{Y}^i$ representing the transients $\overline{\overline{X}}^i$ [4500, 27]. Therefore, we have proceeded at identifying the most meaningful features for the representation of the transients by resorting to a fuzzy method for sensitivity analysis [Zio et al., 2003]: an index of the importance of each one of the 81 inputs with respect to its clustering capability is found, measuring the contribution of the input in separating the transients $\overline{\overline{X}}^i$ into *C* clusters.

Figure 16 shows the projection of the transients in the space of the four identified most important features of the statistical indicators. The different symbols represent the clusters to which the transients have been associated with largest degree of membership in Section 4, thus considering the raw signal values. Notice that clusters 1 and 4 (empty circles and triangles,

respectively) can be clearly distinguished by considering these four statistical indicators. In particular, although boundaries of clusters 2 (dots) and 3 (crosses) cannot be easily identified, because in this feature space these two clusters are shadowed by clusters 1 and 4, the capability in identifying clusters 1 and 4 in this feature space is due to the fact that these two clusters differ for the rate of decrease of the considered signals, which is well captured by features such as the mean and the standard deviation of the signals. With respect to the other two clusters, one can observe that the transients of cluster 2 do not tend to be located in the same zone, whereas transients of cluster 3 are characterized by the lowest values of the standard deviation features.

On the other side, considering the Haar wavelet transform features, the transients of cluster 2 and cluster 3 are well separated from the others (as shown in Figure 17), whereas the separation between transients of clusters 1 and 4 do not appear.

In conclusion, the two obtained projections of the transients in other feature spaces (Figures 16-17) lead us to assert that also in these alternative feature spaces the considered transients tend to form clusters similar to those previously identified in Section 4 by our methodology, and this can be considered as a confirmation of the correctness of the identified clusters.

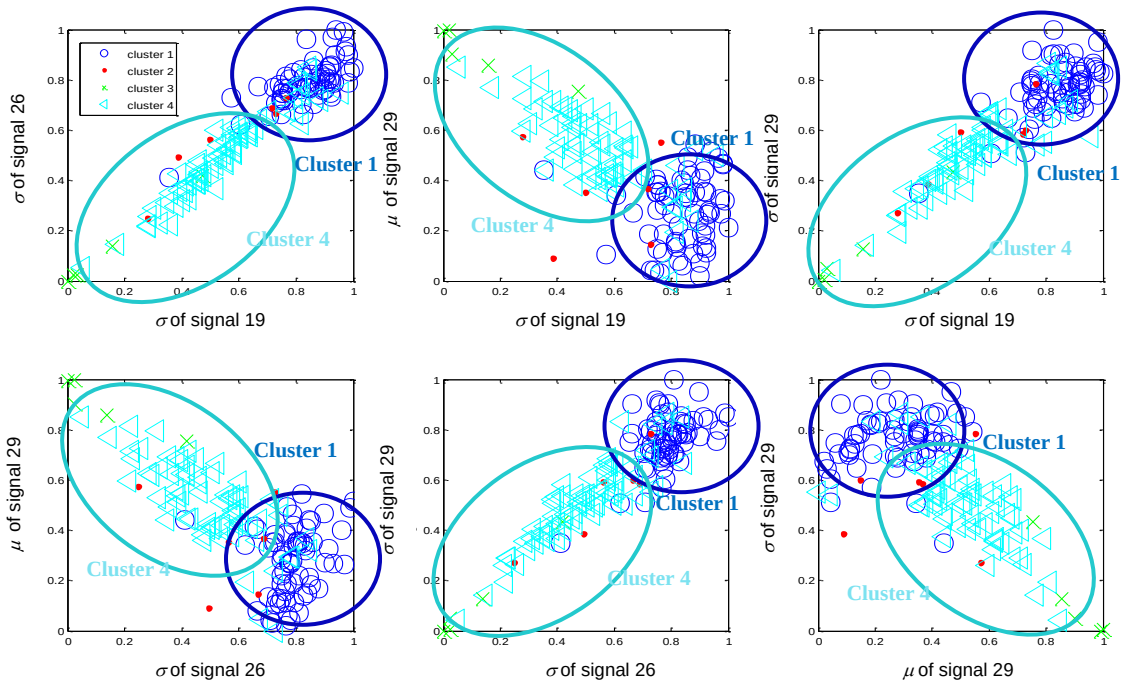

Figure 16: Projection of the transients in the space of the 4 most relevant statistical indicators features

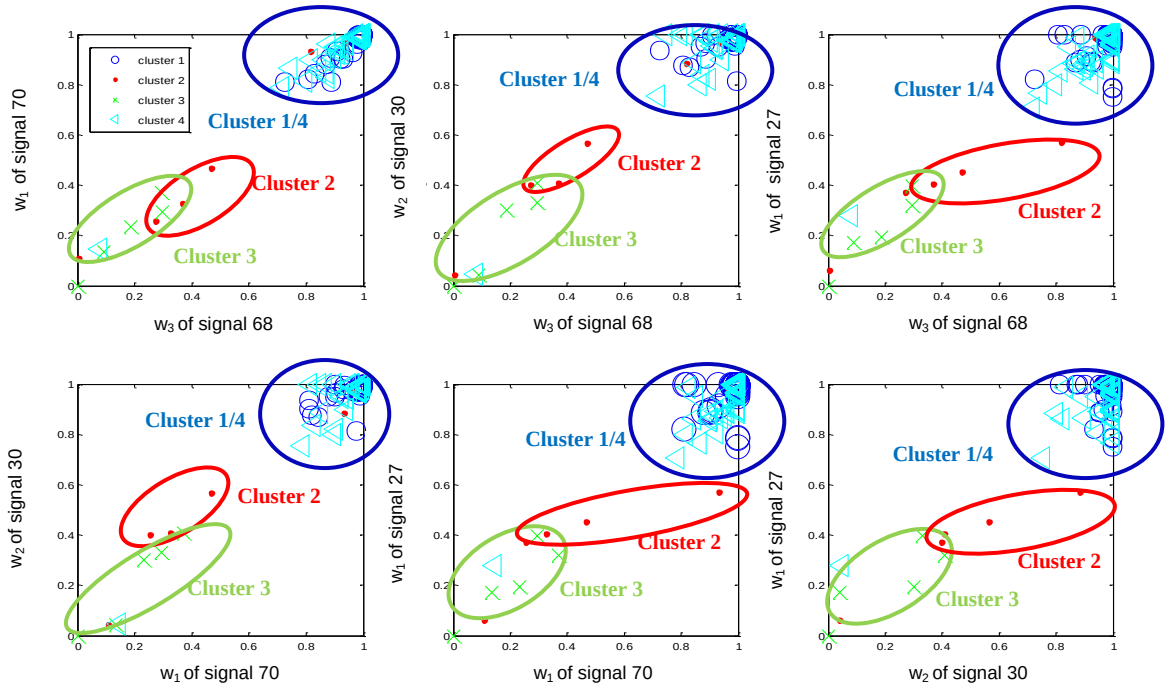

Figure 17: Projection of the transients in the space of the 4 most relevant wavelet features

6. Conclusions

In this work, we have presented a methodology for unsupervised classification of transients with similar behavior, for distinguishing among different operational/faulty conditions. We have embraced a feature extraction technique capable of dealing with misaligned and amplified data.

The clustering of the transients is then obtained by computing the fuzzy similarity among the extracted features of the transients. A similarity graph is built, in which each vertex represents a transient and the weight associated to the edge connecting two vertices is the value of the fuzzy similarity between the two corresponding transients. An FCM-based spectral analysis algorithm is applied in order to find an optimal partition of the graph that leads to the identification of different groups of transients characterized by strong similarity between them.

An application of the proposed methodology to a real industrial case study has been shown with reference to 148 shut-down transients of a NPP turbine. Four groups of different transients have been identified, which differ for the rate of decrease of signals correlated to the turbine speed. In addition, two other groups have been extracted and analyzed by EDF experts: one is characterized by transients with anomalous behaviors for some signals due to peculiar working conditions, whereas the other is due to failures occurred during the transients.

For validation of the obtained clusters, the results have also been analyzed in spaces described by features different from those which have driven the clustering: therein, the transients tend to form clusters indeed similar to those identified by the applied methodology.

References

[Alpert et al., 1999] C. Alpert, A. Kahng, S. Yao, "Spectral partitioning: the more eigenvectors, the better", Discrete Applied Math, Volume 90, pages 3-26, 1999.

[Angstenberger, 2001] L. Angstenberger, "Dynamic fuzzy pattern recognition, international series in intelligent technologies", vol. 17. Dordrecht: Kluwer Academic Publishers; 2001.

[Anker et al., 2003] R. Anker, I. Chernyshev, P. Egger, F. Mehran, J.A. Ritter, "Measuring decent work with statistical indicators", International Labour Review, Vol. 142 , No. 2, 2003.

[Baldwin et al., 2003] J. F. Baldwin, S. B. Karale, "Asymmetric Triangular Fuzzy Sets for Classification Models", Lecture Notes in Computer Science, 2003, Volume 2773, p.364-370, 2003.

[Baraldi et al., 2010] P. Baraldi, A. Cammi, F. Mangili, E. Zio, "Local Fusion of an Ensemble of Models for the Reconstruction of Faulty Signals", IEEE Transactions on nuclear science, Vol. 57, No. 2, Milan, Italy, April 2010.

[Baraldi et al., 2011] P. Baraldi, F. Di Maio, L. Pappaglione, E. Zio, R. Seraoui, "Condition Monitoring of Electrical Power Plant Components During Operational Transients", Proceedings of the Institution of Mechanical Engineers, Part O, Journal of Risk and Reliability, 226(6) 568–583, 2012.

[Baraldi et al., 2013a] P. Baraldi, F. Di Maio, E. Zio, "Unsupervised Clustering for Fault Diagnosis in Nuclear Power Plant Components and Systems", in print, International Journal of Computational Intelligence Systems, 2013.

[Baraldi et al., 2013b] P. Baraldi, F. Di Maio, M. Rigamonti, E. Zio, R. Seraoui, " Transients Analysis of a Nuclear Power Plant Component for Fault Diagnosis", accepted for publication on the proceedings of Prognostics and System Health Management Conference (PHM-2013), Milan, Italy, 8-11 September 2013.

[Bojadziev et al., 1995] G. Bojadziev, M. Bojadziev, "Fuzzy sets, fuzzy logic, applications", Advances in Fuzzy System - Applications and Theory Vol. 5, World Scientific Publishing, Singapore, p.35-41, 1995.

[Brink et al., 1994] J. J. Brink, H. Bohl, "Steam turbine – Generator set start-up and operation", Engineering Conference, p. 295-310, 1994.

[Cheon et al., 1993] Cheon, S. W., Chang, S. H., Chung, H. Y., Bien, Z. N., "Application of neural networks to multiple alarm processing and diagnosis in nuclear power plants", IEEE Transactions on Nuclear Science 40 (1) , pp. 11-20, 1993.

[Di Maio et al., 2011] F. Di Maio, E. Zio, M. Pecht, P. Tse, K. Tsui, "Ensemble of Unsupervised Fuzzy C-Means classifiers for clustering health status of oil sand pumps", Expert Systems with Applications 39, p.4847–4859, 2011.

[Di Maio et al., 2013] Di Maio, F., Baraldi, P., Zio, E., Seraoui, R. "Fault detection in nuclear power plants components by a combination of statistical methods", IEEE Transactions on Reliability, 62 (4), art. no. 6632931, pp. 833-845, 2013.

[Di Maio et al., 2014] F. Di Maio, S. Baronchelli, E. Zio, "Hierarchical Differential Evolution for Minimal Cut Sets Identification: Application to Nuclear Safety Systems", available on-line, European Journal of Operational Research

[Frank, 1987] I. E. Frank, "Intermediate least squares regression method", Chemometrics and Intelligent Laboratory Systems, 1: 233–242, 1987.

[Guyon et al., 2003] I. Guyon, A. Elisseeff, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3, 1157-1182, 2003.

[Joentgen et al., 1999] A. Joentgen, L. Mikenina, R. Weber, H.J. Zimmermann, " Dynamic Fuzzy Data Analysis based on Similarity between Functions", Fuzzy Sets and Systems, 105 (1), 1999, p. 81-90, 1999.

[Kim et al., 1996] Kim, K., Bartlett, E.B., "Nuclear power plant fault diagnosis using neural networks with error estimation by series association", IEEE Transactions on Nuclear Science 43 (4 PART 2) , pp. 2373-2388, 1996.

[Leyzerovich, 1997] Leyzerovich, A., "Large power steam turbines", p.855-998, PennWell Books, Oklahoma, 1997.

[Park et al., 2000] Park, S., Chu, W. W., Yoon, J., Hsu, C., "Efficient searches for similar subsequences of different lengths in sequence databases", Proceedings - International Conference on Data Engineering, pp. 23-32, 2000.

[Reifman, 1997] Reifman, J., "Survey of artificial intelligence methods for detection and identification of component faults in nuclear power plants", Nucl. Technol. 119, 76–97, 1997.

[Secchi et al., 2008] L.M. Sangalli, P. Secchi, S. Vantini, V. Vitelli, "K-mean alignment for curve clustering", Communications in Applied and Industrial Mathematics, ISSN 2038-0909, 1, 1, 204-224, Milano, Italy, 2008.

[Sheng et al., 2004] Sheng, Y., Rovnyak, S.M, Decision Tree-Based Methodology for High Impedance Fault Detection, IEEE Transactions on Power Delivery, Vol. 19, No. 2, 2004.

[Strang et al., 1996] G. Strang, T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge, 1996.

[Subramani et al., 2006] P. Subramani, R. Sahu, S. Verma," Feature selection using Haar wavelet power spectrum", ABV-Indian Institute of Information Technology and Management, Gwalior, India, 2006.

[Tian et al., 2007] Tian, J., Gao, M., Li, K., Zhou, H., 2007, Fault Detection of Oil Pump Based on Classify Support Vector Machine, *2007 IEEE International Conference on Control and Automation WeD5-4*, Guangzhou, China, May 30–June 1, 2007.

[Venkatasubramanian et al., 2003] Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-based Methods, Computers & Chemical Engineering, Vol. 27, pp. 293-311, 2003.

[von Luxburg, 2007] U. von Luxburg, "A Tutorial on Spectral Clustering", Statistics and Computing, Volume 17(4), pages395-416, 2007.

[Zio et al., 2003] E. Zio, P. Baraldi, "Sensitivity analysis and fuzzy modelling for passive systems reliability assessment", Annals of Nuclear Energy, Vol. 31, Pages 277–301, 2004.

[Zio et al., 2006] Zio, E., Gola, G., Neuro-fuzzy Pattern Classification for Fault Diagnosis in Nuclear Components, Annals of Nuclear Energy, Vol. 33, Issue 5, pp. 415-426, 2006.

[Zio, 2007] Zio, E., Soft Computing Methods Applied to Condition Monitoring and Fault Diagnosis for Maintenance, Reliability: Theory and Applications, Vol. 3, 2007.

[Zio et al., 2010a] E. Zio, F. Di Maio, "A Data-Driven Fuzzy Approach for Predicting the Remaining Useful Life in Dynamic Failure Scenarios of a Nuclear System", Reliability Engineering and System Safety, RESS, Volume 95(1), Pages 49-57, 2010.

[Zio et al., 2010b] E. Zio, F. Di Maio, "A Fuzzy Similarity-Based Method for Failure Detection and Recovery Time Estimation", International Journal of Performability Engineering, Vol. 6, No. 5, September 2010.

[Zio et al., 2010c] E. Zio, F. Di Maio, M. Stasi, "A Data-driven Approach for Predicting Failure Scenarios in Nuclear Systems", Annals of Nuclear Energy, 37, 482–491, 2010.

Appendix A: Unsupervised spectral Clustering

- ***Step 1: normalized Graph Laplacian Matrix***

  Compute:

  - the degree matrix $\overline{\overline{D}}$ which is a diagonal matrix with diagonal entries d_1, d_2,…, d_N defined by

$$d_i = \sum_{j=1}^{N} \mu_{ij} \ , \qquad i = 1,2,...,N \tag{1a}$$

  - the normalized graph Laplacian matrix:

$$\overline{\overline{L}}_{rw} = \overline{\overline{D}}^{-1}\overline{\overline{L}} = \overline{\overline{I}} - \overline{\overline{D}}^{-1}\overline{\overline{W}} \tag{2a}$$

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

- ***Step 2: eigenvalues and eigenvectors of L_{rw}***

The information on the structure of a graph can be obtained from its spectrum [Zhao et al., 2007]. Given $\overline{\overline{L}}_{rw}$, compute the first C eigenvalues $\lambda_1, \lambda_2, ..., \lambda_C$ and corresponding eigenvectors $\overline{u}_1, \overline{u}_2, ..., \overline{u}_c$. The first C eigenvalues are such that they are very small whereas λ_{C+1} is relatively large [von Luxburg, 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 $\overline{\overline{W}}$ is obtained by considering the eigenvectors $\overline{u}_1, \overline{u}_2, ..., \overline{u}_c$ associated to the C smallest eigenvalues of its laplacian matrix $\overline{\overline{L}}_{rw}$. The square matrix $\overline{\overline{W}}$ is transformed into a reduced matrix $\overline{\overline{U}}$ of size [N, C], in which the C columns of $\overline{\overline{U}}$ are the eigenvectors $\overline{u}_1, \overline{u}_2, ..., \overline{u}_c$. Thus, the i-th transient similarity with other transients is captured in the C-dimensional vector $\overline{u}_i$ corresponding to the i-th row of the matrix $\overline{\overline{U}}$. It has been shown that this change of representation enhances the cluster properties in the data, so that clusters can be more easily identified [von Luxburg, 2007].

- ***Step 5: Unsupervised clustering***

In this work, we resort to the Fuzzy C-Means (FCM) algorithm to partition the data into C clusters [Bezdek, 1981; Leguizamon et al., 1996; Alata et al., 2008]. FCM originates from hard C-Means clustering: the difference is that it allows elements (transients, in our case) to belong to two or more clusters [Klir et al., 1995]. For each i-th element, the algorithms provides its membership m_{ic} to all clusters, $c = 1, 2, ..., C$. If needed, crisp assignment can be obtained, e.g., by considering the cluster to whom the element belongs with the largest membership value. A prototypical transient can be identified for each cluster by considering the transient with the largest membership value to the cluster.