Online coherency identification and stability condition for large interconnected power systems using an unsupervised data mining technique

Emilio Barocio[1] ✉, Petr Korba[2], Walter Sattinger[3], Felix Rafael Segundo Sevilla[2]

[1]*Graduated Program in Electrical Engineering, Universidad de Guadalajara, Guadalajara, Mexico*
[2]*Institute of Energy Systems and Fluid Engineering, Zurich University of Applied Sciences, Winterthur, Switzerland*
[3]*Swissgrid, Aarau, Switzerland*
✉ *E-mail: emilio.barocio@cucei.udg.mx*

Abstract: Identification of coherent generators and the determination of the stability system condition in large interconnected power system is one of the key steps to carry out different control system strategies to avoid a partial or complete blackout of a power system. However, the oscillatory trends, the larger amount data available and the non-linear dynamic behaviour of the frequency measurements often mislead the appropriate knowledge of the actual coherent groups, making wide-area coherency monitoring a challenging task. This paper presents a novel online unsupervised data mining technique to identify coherent groups, to detect the power system disturbance event and determine status stability condition of the system. The innovative part of the proposed approach resides on combining traditional plain algorithms such as singular value decomposition (SVD) and *K*-means for clustering together with new concept based on clustering slopes. The proposed combination provides an added value to other applications relying on similar algorithms available in the literature. To validate the effectiveness of the proposed method, two case studies are presented, where data is extracted from the large and comprehensive initial dynamic model of ENTSO-E and the results compared to other alternative methods available in the literature.

1 Introduction

Energy demand is constantly growing as a result of a more digital and technological society. Consequently, in order to satisfy the growing demand, the need for interconnecting existing power grids into a larger and more complex network is necessary. For this reason, challenges caused by exceeding power flows at the extra high-voltage network are a primary objective to be studied. Interconnections of electrical systems not only allow energy trading but also provide flexibility when balancing generation and demand. Nevertheless, interconnections introduce operational complexities for utilities and make the system more prone to oscillate when malfunctions or negative events occur such as three-phase short-circuits on transmission lines, generator trips or variations on loads just to mention a few. Inter-area oscillations, which are known to be in the range of 0.1–1 Hz [1], appear as soon as a group of generators located far away from each other (distant geographical locations), oscillate against other groups. These collections of synchronous machines swinging in the same direction are also known as coherent groups. Identification of groups of coherent generators is an important task for dynamic system studies and is neither a simple nor obvious assignment [2].

Possessing such valuable information can prevent to limit the spread of cascading events by partitioning the system into multiple independent islands or circuits [2, 3]. Another possibility to profit from having the coherency groups is to identify the most critical nodes to design damping controllers as well as identification of the optimal location of phasor measurement units (PMUs) [2, 3]. Similarly, there are two possible paths for determination of the coherent groups, methods based on models and methods based on measurements [3, 4]. Since one of the milestones in this work is to provide online tools, special attention is given to methods based on measurements.

In addition to the problem of identifying coherent groups, the actual size of the power system by itself introduces an extra degree of complexity, which means that the groups might change according to disturbance location, loading conditions and the

structure of the network [5, 6]. Large power grids generate significant amounts of information, which has to be stored in local servers requiring significant amounts of central processing unit (CPU) storage. This problem is known as big data and although it is not a new concept but rather a challenge that has always existed in different domains, it has drawn a lot of attention in recent years within the power system community. The algorithms presented in this brief introduce potential solutions to big data problems in power systems such as the proposed unsupervised data mining technique. Existing works in the literature offer offline algorithms for identification of coherency groups [7–9]. Also available are the online approaches and the first main difference among online and offline applications is related to the nature of the input signals used. Most of these methods require generator rotor angles and rotor speeds, which in some cases are estimated directly from synchrophasor measurements as in [10], and in other cases, the authors assume to have direct access to these variables [11, 12]. Alternatively, to machine parameters, bus voltages angles, and frequency of the system [3, 13], which are signals directly measured from the PMUs, are also used to determine these notorious coherency groups. Online algorithms also contrast among them in terms of the complexity in the mathematical formulation to provide assurance in the results.

Among the most successful multivariate methods for coherency identification are those that can be applied directly to time-series, features or objects such as the principal component analysis (PCA) [8, 14], which rely on orthogonal transformations in order to find linear correlations between variables of the dataset. Comparatively, singular value decomposition (SVD) yields a matrix decomposition, which allows extracting the most representative data information using the singular value information [15]. Similarly, independent component analysis (ICA) [5, 13] is a computational method for separating a multivariate signal into additive subcomponents. The success of ICA is dependent on conditions such as non-Gaussian signals and datasets statistically independent of each other. Those methods (PCA, SVD, and ICA) use spatial signatures that are contained in their modal
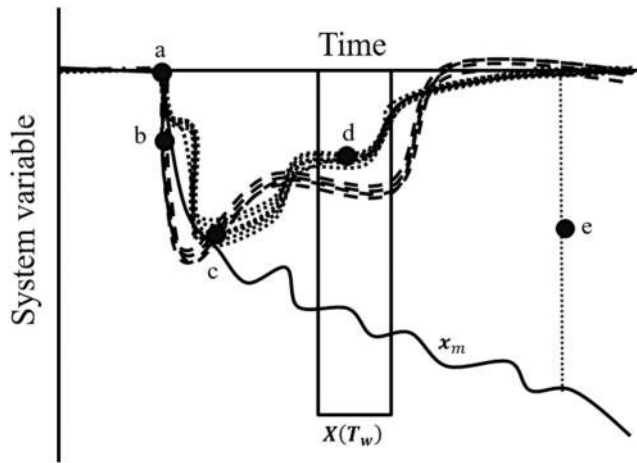
Fig. 1 *Illustration of the unstable generator's frequency response*

decomposition vectors for coherency identification; which usually are the three most representative vectors. However, the detection of the onset of transient dynamic behaviour and the power system stability status is not directly measured. Recently, kernel PCA (KPCA), which is a robust variant of PCA, and clustering analysis based on affinity propagation (AP) was proposed for coherency detection in power systems with extensive penetration of renewable generation source [16]. KPCA + AP find correlations among multiples indexes using an AP-based clustering algorithm and Prony to identify the coherent groups and also low-frequency modes, which require a large period of time up to 4 s to identify low-frequency modes of 0.2 Hz. More advance methods based on concepts such as Koopman operator have been used for coherency identification and determination of stability margins of the system [17, 18]. The application of procedures such as Koopman mode (KM) decomposition and dynamic mode decomposition (DMD) combine the ability of mode identification with multivariate statistical techniques to extract modal parameter information, determine dominant coherency process and provide a stability condition of the system. Despite the benefits of KM and DMD, these sophisticated methods are just applied during the oscillatory process and require a large period of time to be successfully applied.

Another widespread family of approaches is those based on Euclidean distances, such as K-means [12]; hierarchical clusterings such as the agglomerative nesting [11], density models such as density-based spatial clustering of applications with noise and hybrid algorithms such as CLIQUE that combines density and grid-based space data [7]. Pattern recognition formulations have also been successfully applied in [2, 10, 19], however, these sophisticated methods present visualisation problems as the number of coherent areas of collected signals increase. Parallel to the already stated methods, those based on Fourier analysis such as [3, 20], which are popular for providing modal information such as frequency and damping of oscillatory modes, can also be found in the literature. However, it has been proved that Fourier-based methods could be inadequate for non-stationary signals and require an appropriate proportion between the sampling frequency and size of the window to provide an accurate estimation. Despite the effectiveness to find coherent groups from formulations displayed before, complex methods derived from graph theory [21, 22], neural networks [6] and formulations based on different optimisation strategies in [23, 24] can also be found but on a minor scale.

The aim of what to do with the knowledge of the coherence groups is the ultimate argument to select the algorithm to be applied and consequently the required nature of the input signals. To this end, some of the potential directions of how to use such information can be developing security indexes for stability assessment [3], implementation of clustering directly to existing wide-area monitoring systems (WAMSs) [10, 25], determine if an islanding event has occurred [14, 26], identify disturbance events

[7] or as in few cases, improve existing algorithms or present an alternative approach for determination of coherence groups [11].

In addition to the challenges described before, the length of the time window plays an important role to carry out effective online onset transient and coherency detection as well as to determine the stability condition of the system. Some existing methods process instantaneous values measured directly from PMUs or from the synthetic data to define coherency dynamic procedures [2, 14, 19, 21]. These algorithms are fastest to provide information for control actions but may be affected by missing data, outliers and incorrect measurements. Other algorithms present time delays caused by the size of the time window (1–10 s) selected to classify [6, 27, 28] and extract the modal information [3, 8, 13, 16–18, 20].

Based on the above review, in this work, we present an alternative to attempt to cope with challenging problems regarding online islanding and coherency identification as well as the stability condition of the system. The proposed online data mining methodology brings several contributions and novelties as listed below:

- A security index based on clustering that can be used for online islanding detection. The proposed metric uses as input frequency measurements from PMUs, which are used to detect: (a) when a transient problem occurs and (b) identify the triggering of unstable events.
- The novel security indicator exploits the concept of acceleration and deceleration of the synchronous machines to estimate online the coherent groups and classify the islanding.
- Online visual dynamic information to awareness the power system stability status and evolving oscillatory coherency phenomenon mechanism.

The paper is organised as follows: Section 2 introduces the proposed tool for online application. Section 3 synthesises and breaks down the proposed approach and includes a detailed description. Section 4 introduces the system under investigation, presents the results of two different study cases and discuss the main results obtained. Section 5 describes the future work and conclusions of this work. Finally, an appendix to compute a threshold introduced in Section 2 is presented.

2 Proposed unsupervised online clustering data mining algorithm

2.1 Problem definition of clustering dynamic trajectory motion in power systems

The concept of clustering consists of dividing a set of objects into smaller groups that have similar aspects or attributes. In a formal definition, cluster analysis is about finding groups in a set of objects or features distributed on Euclidean space $U = \{u_1 \ \dots \ u_r\} \in \mathbb{R}^{m \times r}$, which is defined by m objects and r data vectors. Let us assume that Ω is the set of all possible clusters C of the data set U, where $C = \{c_1 \ \dots \ c_k\}$, such that $\cup_{j=1}^{K} c_j = U$ and $c_j \cap c_i^c = \varnothing, \forall i, j = 1, 2, \dots, K, \quad \forall i \neq j$ and considering that $K \in \{1, 2, \dots, \text{card}(C)\}$. Consequently, every object $u_{ij} \in U$ is contained in exactly one and only one set c_j. The number of clusters K may be fixed in advance or defined by an iterative search. To carry out this important task, successful algorithms have been developed to cope with this challenging problem in diverse disciplines [29].

In the context of this work, the so-called objects can be associated with dynamic trajectories related to generators frequency responses, which can be monitored by PMUs or simulated using state-of-the art commercial stability calculation software tools. Fig. 1 displays a conceptual description of the collected system dynamic data.

To introduce the online clustering problem, Fig. 1 is used to illustrate the complex behaviour of the system following a severe disturbance and the different stages to extract valuable information. Point (a) indicates the initial time of a fault, point (b) determines the turbine-generator inertial response, point (c) depicts the governor control response, point (d) defines the typical reaction
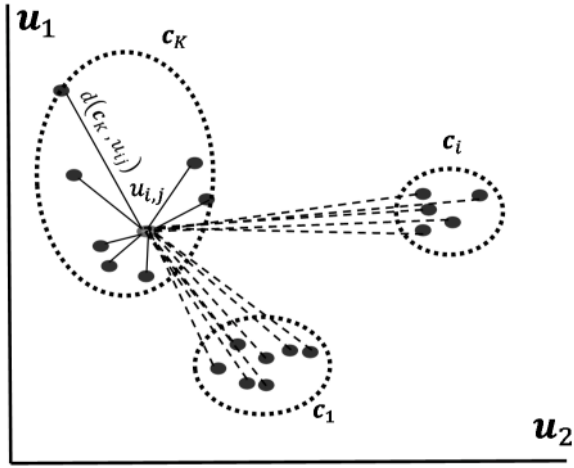
Fig. 2 *Example: 2D clustering data process*

time of the automatic generation control (AGC) response and point (e) suggests the appropriate instant for application of metrics to indicate unstable conditions. These dynamic responses have non-linear characteristics associated with each dynamic trajectory motion, the time scale of control actions and the increasing amount of data collected by the PMUs.

In this study, short time windows (0.25 s) are used to propose a global index that helps to understand the mechanism of the dynamic coherency process as well to detect the onset of the transient. To this end, let us assume a matrix that describes the motion trajectory, which corresponds to the objects collected in discrete time and is represented by $X(T_\mathrm{w})$. Following the nomenclature in [17], assume that $x_j(t_1)$, $j = 1, \ldots, m$ denotes an element of observation, at the jth measurement point, and t_i, $i = 1, 2 \ldots l, \ldots, w$ is the time at which the observations are made

$$X(T_\mathrm{w}) = \begin{bmatrix} x_1(t_1)\ldots & x_1(t_\mathrm{w}) \\ x_2(t_1)\ldots & x_2(t_\mathrm{w}) \\ \vdots & \vdots \\ x_j(t_1)\ldots & x_j(t_\mathrm{w}) \\ \vdots & \vdots \\ x_m(t_1)\ldots & x_m(t_\mathrm{w}) \end{bmatrix} \in \mathbb{R}^{m \times w} \qquad (1)$$

When $m \gg w$ implies that an extra-large amount of data has to be processed simultaneously; requesting demand for high-computational resources and complexity data analysis in the online tools. To tackle this problem, the proposed approach involves the three major steps:

i. Feature extraction
ii. Clustering features
iii. Stability status system

The proposed algorithm provides information on each step: detection of the transient period, cluster visualisation, and intuitive indexes to assess the system status and the dynamic coherency phenomenon mechanism.

2.2 Feature extraction by dimensional reduction data technique

A dimensional reduction technique is first used to extract the relevant information contained in (1) and provide Euclidian space outputs that facilitate the clustering process. In this regard, SVD is used to obtain a low-dimensional spatial decomposition of a high-dimensional transient process during a limited time window T_w. Mathematically, this is equivalent to finding the best low-rank approximation (in L2 norm) of the data. Let r be the true rank of the data matrix $X(T_\mathrm{w})$ using the SVD as given by

$$X(T_\mathrm{w}) = U\Sigma V^\mathrm{T} = [U][\Sigma_r \quad 0] \begin{bmatrix} V_r^\mathrm{T} \\ V_s^\mathrm{T} \end{bmatrix}, r \leq m \gg w \qquad (2)$$

where Σ_r is a diagonal matrix that contains the singular values denoted as

$$\Sigma_r = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}, \ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \qquad (3)$$

With σ_j ranked in descending order of energy. It can be noted that the columns of $U \in \mathbb{R}^{m \times m}$ are the eigenmodes (pseudo mode shapes) that holds information on the modal spatial and the matrix product of $\Sigma_r V_r^\mathrm{T} \in \mathbb{R}^{r \times w}$ are related to the temporal structure. Based on these facts, (2) can be rewritten in a more useful form

$$\bar{X}(T_\mathrm{w}) \cong u_1 a_1(t) + \cdots + u_j a_j(t) + \cdots + u_r a_r(t) \qquad (4)$$

where $a_j(t) = \sigma_j \sum_{l=1}^{w} V_{r(j,l)}^\mathrm{T}$ are the temporal amplitudes and $u_j = [u_{1j} \quad \ldots \quad u_{mj}]^\mathrm{T}$ are the spatial modes. Considering that vectors u_i in matrix U are mutually orthogonal and each eigenvector u_j is weighted by each mode energy extracted; based on this, the information contained in (1) may be represented by the reduced model given in (4) with the first two or three principal components (r), providing the outputs in a Euclidian space (2D or 3D) for the clustering process.

2.3 Clustering feature extraction data

According to (4), the most dominant eigenmodes are selected to extract the main features of data contained in (1). Usually, two or three eigenmodes are enough to characterise the whole data given in (1). Considering the properties announced in [17] and without loss of generality, let us assume as outputs that $U = \{u_1 \ u_2\}^\mathrm{T}$ be the data set of objects and assume that $C = \{c_1 \ c_2, \ldots, c_K\}^\mathrm{T}$ be its clustering into K clusters, and the distance between each centroid and each data point u_{ij} be defined as $d(c_K, u_{ij})$ as is described in a 2D plot shown in Fig. 2.

There exist a wide range of techniques available to determine the number of clusters existing within a range of data [29]. In this work, the K-means clustering algorithm has been selected for its computational efficiency and effectiveness to process Euclidean distance among spherical cluster data. The algorithm inputs are data set features $u_i \in U$ defined for each data point u_{ij} and the number of clusters K.

To initialize the algorithms, first, the initial estimates for the K centroids are carried out. These can be randomly generated or selected a priori from the data set. Then, the algorithm iterates between two processes described below:

i. *Data assignment*: Each centroid defines one of the clusters. Here, each data point u_{ij} is assigned to its nearest centroid c_K, based on the following equation:

$$\underset{c_K \in C}{\arg \min} \ d(c_K, u_{ij}) \qquad (5)$$

where $d(c_K, u_{ij})$ is the standard (L2) Euclidean distance, which can also be used for different types of metrics [29]. Let c_K be the set of data point u_{ij} assignments for each Kth cluster centroid.

ii. *Centroid update*: During the second step, the centroids are computed again by taking the mean of all data points assigned to that Kth centroid's cluster as follows:

$$c_K = \frac{1}{p_K} \sum_{u_{ij} \in c_K} u_{ij} \qquad (6)$$

where p_K denotes the number of objects in the Kth cluster. The algorithm iterates between the data assignment and the centroid update until a stopping criterion is met.

The algorithm described above finds the clusters and data set labels for a particular pre-chosen K. To determine the optimal number of clusters that best fit the dataset U, the unsupervised Silhouette index (SI) was selected based on its ability to examine compactness and separation of the clusters [30]. Following the illustrative example in Fig. 2, the SI assigns a quality measure to each data point at u_{ij} in c_K known as the Silhouette width. The Silhouette width is a confidence indicator in the membership of the u_{ij} data in cluster c_k and is defined as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}} \tag{7}$$

where

$$a(i) = \underset{d \in c_K}{\text{avg}}\ \delta(d_i, d\ ),\ d_i \in c_K \tag{8}$$

$$b(i) = \underset{C \neq c_K}{\min} \left\{ \underset{d \in c_K}{\text{avg}}\ \delta(d_i, d\ ) \right\},\ d_i \in c_K, \tag{9}$$

and where $a(i)$ denotes the average distance between the u_{ij} and all data features in the same cluster c_K (continuous lines in Fig. 2) and $b(i)$ is the minimum average distance between the u_{ij} and all data features in the closest another cluster (dashed lines in Fig. 2). From expressions (7), (8) and (9) it may be inferred that variable $s(i)$ will take values in the range of $-1 \leq s(i) \leq 1$. The Silhouette coefficient is expected to be positive $a(i) < b(i)$ with values of $a(i)$ as close to zero as possible. The coefficient assumes its maximum value of 1 when $a(i) = 0$.

Thus, for a given cluster c_k, it is possible to calculate a cluster Silhouette $s(i)$, which characterises the compactness and separation of such a cluster. Moreover, for any partition, a global SI $G_{s(i)}$ can be used as an effective validity index for a partition K

$$G_{s(T_w)} = \frac{1}{K} \sum_{i=1}^{K} s(i) \tag{10}$$

where $s(i)$ is the Silhouette coefficient. In this case, the maximum SI value is taken as the optimal partition, indicating the best clustering.

To synthesise the clustering procedure, K-means operates a number of times for a different number of clusters. Then its respective curve of validation index $G_{s(i)}$ is outlined and automatic search of the significant 'knee' within the diagram is performed. The number of clusters at which the 'knee' is observed indicates the optimum clustering for the selected data set. The index is also entitled to obtain its corresponding labels, which allow to evaluate the quality of the compactness and separation of each cluster as well as provide visual information.

2.4 Determination of stability status using a global stability index (GSI)

After the number of clusters has been determined for the time window T_w, the signals are grouped at $c_K(x_j(T_w))$ by its corresponding labels provided by the clustering process described in Section 3.3. Next, the number of clusters and its respective trajectories at T_w are grouped as illustrated in Fig. 3a.

Now, to simplify the analysis, the average value trajectories corresponding to $c_K(x_j(T_w))$ are calculated as follows:

$$\tilde{c}_k(T_w) = \frac{1}{l} \sum_{j=1}^{l} c_K(x_j(T_w)), \quad j = 1, 2, \ldots, l \tag{11}$$
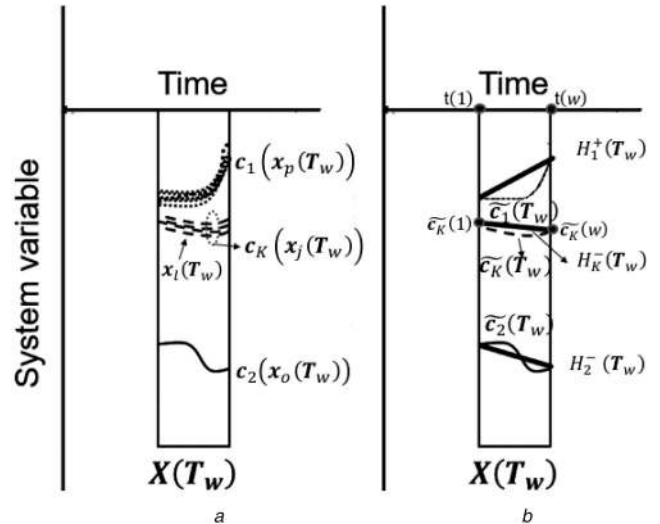

Fig. 3 *Time window analysis*
(a) Clusters $c_K(x_j(T_w))$ and their labelled dynamic trajectories, *(b)* Average clustering $\tilde{c}_k(T_w)$ and its, respectively, slopes $H_i^{\pm}(T_w)$

From (11), it can be noticed that values $\tilde{c}_k(T_w)$ may represent a descending, ascending or constant dynamic trajectory cluster as depicted in Fig. 3b. Based on these trajectory variations, a measure to quantify the changes that experiment the average clusters is then proposed

The suggested measure depends on the average values defined in (11) and is based on the computation of their slopes as follows:

$$H_i^{\pm}(T_w) = \frac{(\tilde{c}_k(w) - \tilde{c}_k(1))}{(t(w) - t(1))} \tag{12}$$

where $\tilde{c}_k(w)$ denotes the sample at the end of T_w, while $\tilde{c}_k(1)$ indicates the first sample of T_w, similarly $t(w)$ and $t(1)$ express the last and first samples in T_w with respect to the vector of time. From the analysis of Fig. 3b and expression (12), the following remarks can be obtained:

The values of the average slopes can be positive $(H_i^+(T_w))$ or negative $(H_i^-(T_w))$. These binary variations represent dynamic trajectories ascending or descending, respectively. At the same time, they represent a physical increment or decrement of the generator frequency.
If $H_i^{\pm}(T_w) = H_j^{\pm}(T_w)$, $j, i = 1, \ldots, K$, $i \neq j$, two slopes are in parallel indicating that $\tilde{c}_i(T_w)$ and $\tilde{c}_j(T_w)$ are evolving in the same direction.
If $H_i^{\pm}(T_w) \neq H_j^{\pm}(T_w)$, $j, i = 1, \ldots, K$, $i \neq j$, two slopes are different from each other and are possible to have the following combinations:

(a) $H_i^-(T_w) \neq H_j^-(T_w)$ or $H_i^+(T_w) \neq H_j^+(T_w)$, imply that $\tilde{c}_i(T_w)$ and $\tilde{c}_j(T_w)$ might cross each other as the oscillation is evolving.

(b) $H_i^-(T_w) \neq H_j^+(T_w)$ or $H_i^+(T_w) \neq H_j^-(T_w)$, produces an eventual dynamic separation between $\tilde{c}_i(T_w)$ and $\tilde{c}_j(T_w)$.

(c) $H_i^-(T_w) \ll H_j^+(T_w)$ or $H_i^-(T_w) \gg H_j^-(T_w)$, indicates a large separation between $\tilde{c}_i(T_w)$ and $\tilde{c}_j(T_w)$, suggesting an eventual unstable operating condition.

The information provided by $H_i^{\pm}(T_w)$ is important and represent a key factor to understand how the oscillatory process evolves during the coherency process. From the aforementioned concepts of ascending and descending 'slopes', two novel indexes can be introduced as the following:

$$Iac(T_w) = \frac{1}{q} \sum_{i=1}^{q} H_i^+(T_w) \tag{13}$$

$$Idc(\boldsymbol{T}_w) = \frac{1}{p}\sum_{i=1}^{p} H_i^-(\boldsymbol{T}_w) \qquad (14)$$

Equations (13) and (14) represent the average ascending and descending slopes of each average cluster, indicating the time global changes that experiment the average clusters during the oscillatory process. Based on these equations a GSI is now proposed

$$GSI(\boldsymbol{T}_w) = \sqrt{Iac(\boldsymbol{T}_w)^2 + Idc(\boldsymbol{T}_w)^2} \qquad (15)$$

Note that (15) is the Euclidean distant metric between the ascending and descending average slopes defined in (13) and (14), respectively. The GSI index is used to track the stability status of the system as follows:

$$\Gamma(\boldsymbol{T}_w)\begin{cases} 1.0 & \text{if} \quad GSI(\boldsymbol{T}_w) \geq \gamma \\ 0.0 & \text{otherwise} \end{cases} \qquad (16)$$

A decision threshold higher than $GSI(\boldsymbol{T}_w) \geq \gamma$ indicates 'the onset of transient behaviour'. More generally, if the warning flag $\Gamma(\boldsymbol{T}_w)$ holds for several consecutive analysed windows, a sudden change in the system behaviour is detected and the control system initiates its protocols to re-establish the state operation condition. In real-time situations, this algorithm is expected to capture emerging transient processes or locally-occurring events associated with major system changes and reduce false alarms. It is worth mentioning that the most appropriate value for the threshold γ is not random but systematically selected. In this work, a statistical process is proposed and presented in the Appendix.

3 General description of proposed unsupervised online clustering data mining algorithm

Next, a general overview of the proposed approach is summarised in four stages where a detailed description of each step is also included.

Stage 1: data processing

- An appropriate time window for sampling the generator frequency is selected and the collected data is organised in a matrix structure $\boldsymbol{X}(\boldsymbol{T}_w)$

Stage 2: dimensionality reduction data

- The outputs $\boldsymbol{U}$ from the low-dimensional map are acquired, which are extracted from the data of the time window $\boldsymbol{X}(\boldsymbol{T}_w)$ using SVD.

Stage 3: clustering data

- The optimal number of clusters hidden in the Euclidean outputs $\boldsymbol{U}$ are determined using K-means and the proposed clustering validity index.

Stage 4: Stability condition

- The decision threshold γ is calibrated following the proposed statistical procedure (see the Appendix).
- The average clusters are calculated to create the ascending and descending slopes.
- The *GSI* (15) is computed to indicate the stability status of the system.

This procedure allows evaluating the performance of the proposed algorithm. The general diagram depicted in Fig. 4 provides a comprehensive graphical description of the information related to the different stages. This procedure was implemented in Matlab version *R*2015. Although non-linear islanding event detection algorithms such as in [14], offer important benefits such as adaptive thresholds for event detection and noise resilience, these algorithms require a correct selection and precise parametric tuning of kernel functions to handle with several non-linearities. In addition, Liu *et al*. [14] concentrated on detecting the separation and condition of the system without providing information about coherency. The main results of this work are presented in the following section.

4 Results and analysis

4.1 Coherency problem on the ENTSO-E system

Fig. 5 depicts in green the countries that constitute the ENTSO-E region of continental Europe. The dynamic model used in this work is a comprehensive representation of the area and mimics the general dynamic behaviour of this region.

Permanent measurements through the actual WAMS confirm the presence of two main oscillatory modes in this model [1, 31]. These modes have been denominated as the East–West and the North–South modes with frequencies (presently) of ∼0.15 and 0.25 Hz, respectively. The interconnection of Turkey (TR) in December 2010 [32], incremented the size and complexity of the original system and as a consequence, the frequency of the East–West mode was directly affected, reducing its original value of ∼0.2 Hz [31].

4.2 Description of presented study cases

To validate the methodology presented in Section 2, we use the initial dynamic model of continental Europe described before. For sake of simplicity, in the following subsections, the proposed methodology is evaluated only for the selected areas; namely Spain (ES), Italy (IT), Switzerland (CH), Germany (DE), and Turkey (TR). The selection of these areas was based on practical experience and the knowledge of the modal analysis behaviour of the interconnected power system. All simulations have been performed using the professional software DIgSILENT PowerFactory 2018 SP1 with a fixed sampling time of 100 ms [33, 34].

Although a large number of case studies were conducted, only two illustrative examples are presented. These representative simulation cases exhibit the most significant properties of the proposed approach and therefore are reported in this study. To test the proposed online data mining algorithm described in Section 2. A sampling frequency of 100 Hz recommended in the IEEE Standard for Synchrophasor Measurements for Power Systems
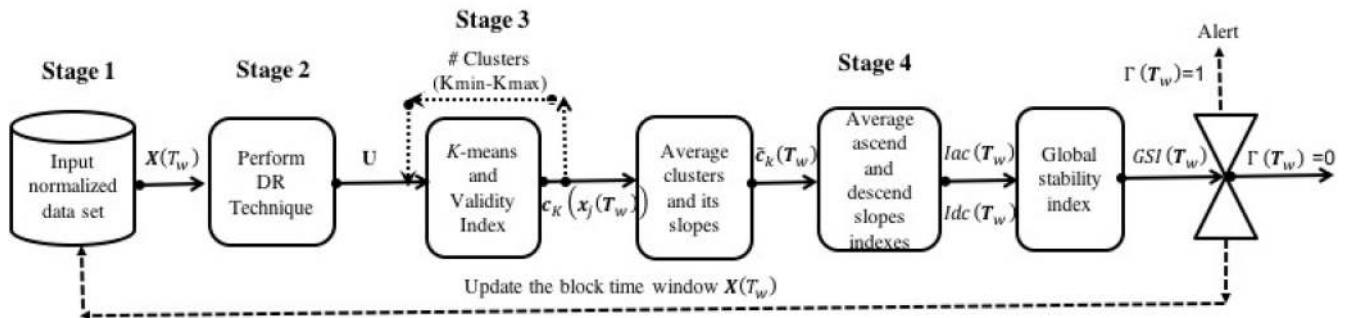

Fig. 4 *Flow chart for the proposed online data mining coherency and status stability system identification approach*

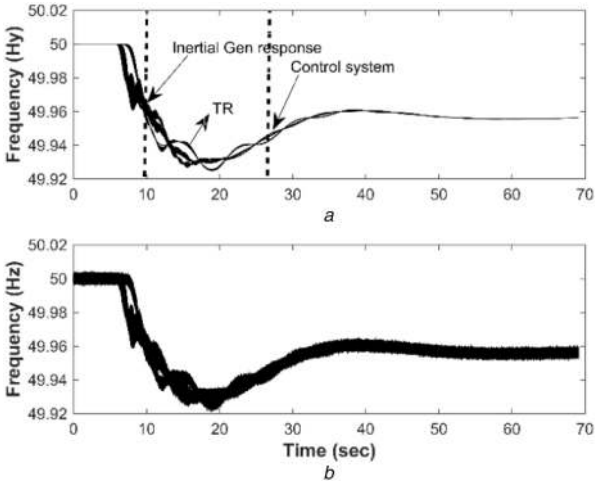**Fig. 5** *ENTSO-E regions of continental Europe highlighted in green*

Fig. 6 *Frequency response to the loss of a large synchronous generator of 1.4 GW in France*

Fig. 7 *Case 1 stable: online indexes*
(a) Ascent and descent of the slopes, *(b)* GSI, *(c)* Number of clusters determined by $G_{s(T_w)}$, *(d)* Alarm index to indicate the transient period

C37.118.1–2011 is used [35]. A uniform stochastic variation on the frequency response of around 0.005 Hz was included to simulate small balancing response of the system generation. To effectively detect the onset of a transient period and monitor the coherency electromechanical oscillatory behaviour of the system, a short time window was selected. Taking into account the IEEE standard 1547-2003, the time response of islanding detection should be <3 s, [36] and the IEC 61727 standard for photovoltaic systems [37] consider a time response <2 s; in this brief, a short time window of $X(T_w) = 0.25$ s was selected to give enough time to discriminate false alarms and to avoid false triggering events. A threshold of $\gamma = 25 \times 10^{-3}$ was selected following the statistical procedure described in the Appendix. Finally, a searching clustering procedure between the minimum and the maximum number of clusters (3–10) was selected in order to track online coherency processes. Next, the two selected study cases are introduced and discussed.

4.3 Study case one: stable response

In this study case, the system is exposed to a severe fault in France. A simulation of 70 s is performed, where at 6 s, a large synchronous generator of 1.4 GW is suddenly disconnected from the three phases (generator trip). The result of this event is depicted in Fig. 6, where the frequency traces of the 653 synchronous machines from the selected countries are shown. The number of measurements per country is as follows: 70 in ES, 144 in IT, 20 in CH, 292 in DE and 127 in TR, conforming a matrix data of $X \in \mathbb{R}^{653x7000}$. From the simulation results depicted in Fig. 6a it can be noticed the interaction of the different groups (oscillations), particularly how these traces vary as the simulation time reach the end.

Following this event, all frequency responses converge at the same point after 70 s. Visual inspection of the traces in Fig. 6a suggests the presence of electromechanical oscillations; TR
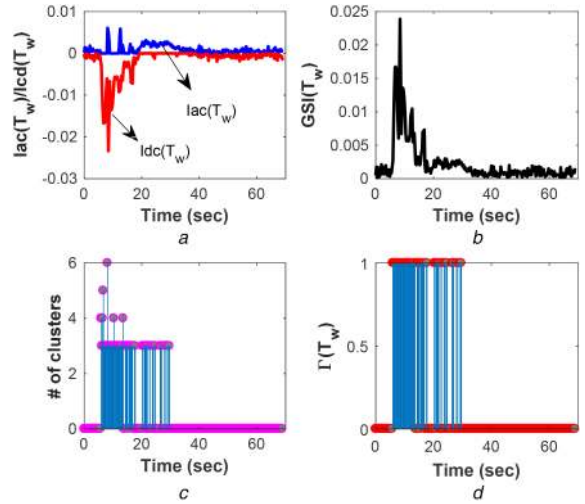
generators and the rest of the European countries under investigation (ES, IT, CH, and DE) present a well-defined cluster motion trajectory in the opposite direction, particularly during the time window between 10 and 30 s. Nevertheless, by pure visual inspection, it cannot be concluded how many groups exist outside the aforementioned window of time. In addition, if the traces are corrupted with noise as it could be expected on a real PMU measurement (Fig. 6b), it is not even possible to know the interaction among these groups.

Fig. 7 presents the result of the online indexes proposed in Section 2.4. A block time window of $X(T_w) \in \mathbb{R}^{653 \times 25}$ is used to process the information online. The computational processing time for the proposed method is calculated following reference [38]. In this case, it has been assumed several time delays related to the collected block time window (T_{Bl}), the latency on the communication system (T_{com}) and the computational time spent by the proposed methodology (T_{cal}). The total computational processing time is calculated as $T = T_{Bl} + T_{com} + T_{cal}$. Based on this expression, the proposed method requires around 0.95 s ($T = 0.25 + 0.1 + 0.60$) of computational time to perform the evaluation. With this time, the proposed method can calculate an effective detection of the transient period and monitor the coherency electromechanical oscillatory behaviour of the system in comparison with [14, 16]. At the same time, the algorithm meets the technical requirements of the IEEE standards 1547-2003 and IEC 61727, that consider a time response for islanding detection <2 and 3 s, respectively [36, 37].

From multiple studies, the dimensional reduction algorithm captures 99% variance with three singular values, reducing the model to a matrix $\bar{X} \in \mathbb{R}^{653 \times 3}$. Fig. 7a displays the average ascending $Iac(T_w)$ and descending $Iad(T_w)$ dynamic slopes; both indexes identify the time periods when the frequency of the generators starts to speed up and slow down, respectively. The decelerating period is identified to be in the time period of 6–20 s, after that the control starts to influence the system response and the acceleration period begins. Similar dynamic behaviours are identified using (15) and visualised in Fig. 7b. Note that after oscillations on the system disappear, the distance between the indexes (13) and (14) is close to zero, indicating that the system is converging to an equilibrium point. Two additional indexes are displayed on Figs. 7c and d, respectively, which describe the number of clusters and the alarms raised to detect the transient period.

Fig. 8 shows the online slope clustering results; three well-defined coherent groups were identified. Two oscillatory electromechanical processes are clearly defined; generators from TR swinging against the rest of EU and ES machines including some generators in DE swinging against CH, IT and the rest of DE.
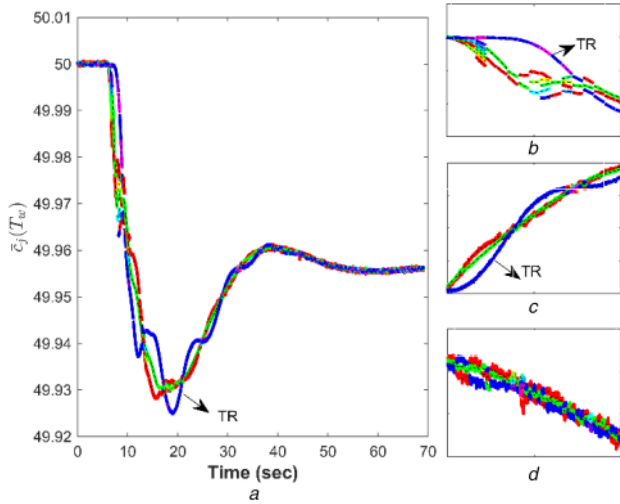
Fig. 8 *Case 1 stable*
(a) Online slope clustering, *(b)* Zoom 6–10 s of slope clustering, *(c)* Zoom 25–35 s of slope clustering, *(d)* Zoom 40–50 s of slope clustering
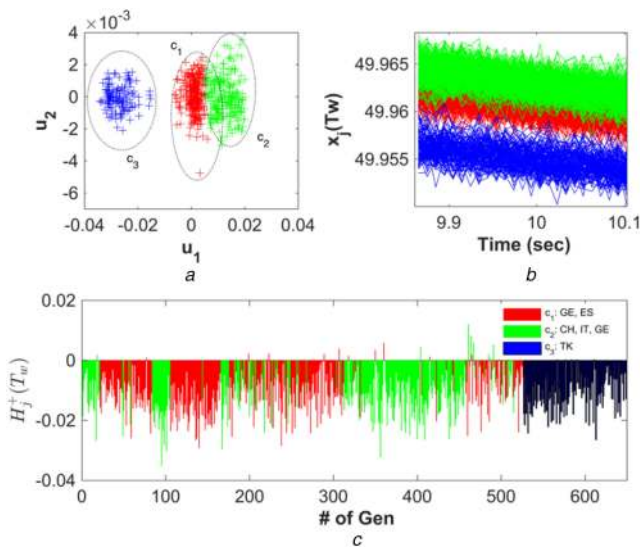

Fig. 9 *Case 1 stable: clustering at turbine-generator inertia response*
(a) Clustering visualisation 2D, *(b)* Labelled dynamic trajectories, *(c)* Average slope generator

The zoom of the online clustering slope results is displayed in Figs. 8b–d, respectively. Fig. 8b presents the inertial frequency response period, where several groups (4 or 6) can be distinguished. After the controls respond to the system fault, three groups are conformed and the results are depicted in Fig. 8c. Finally, Fig. 8d displays the response of the system after it has converged to a new equilibrium point and the system has reached steady state.

To gain more insight into the coherency analysis, special attention is placed to the different phases related to the control actions that impact the generator frequency response as introduced in Section 1; the inertial response (6–19 s), the effects of the controls associated with the governor and the response associated with the AGC (20–55 s). Fig. 9 presents the results applied to the time window related to the turbine-generator inertia response, which is depicted as a dotted line in Fig. 6a. During this phase, the turbine-generator resists against having a frequency drop, resulting in a complex dynamic behaviour caused by different inertial responses of each generator. As a consequence, it is difficult to identify the number of clusters. Fig. 9a exhibits three well-defined coherent groups namely c_1, c_2 and c_3. Furthermore, Fig. 9b presents their corresponding labelled trajectories, which confirm the existence of different dynamic behaviours associate with each cluster. To assess even more closely the contribution of the slopes
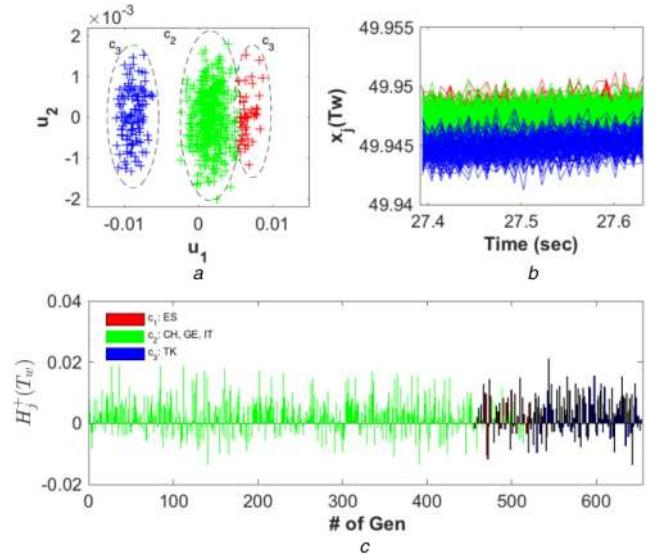

Fig. 10 *Case 1 stable, clustering at control system response*
(a) Clustering visualisation 2D, *(b)* Labelled dynamic trajectories, *(c)* Average slope generator

associated with each generator, an individual slope index is defined as follows:

$$\bar{H}_j^{\pm}(T_w) = \frac{(x_j(w, i) - x_j(1, i))}{(t(w) - t(1))}, \quad i = 1, 2, ..., K \tag{17}$$

where $x_j(w, i)$ is the w sample of the jth trajectory that belongs to the ith cluster. Fig. 9c displays the 'generator contribution'; in this regard, all generators frequency trajectories are decelerating conforming groups in the function of each slope and its respective label. Note that all generators in TR conform group c_3, generators in IT, CH, and DE conform group c_2, and generators in DE exhibit high participation in group c_1.

The second case of interest is related to the control action of the system. The governor of each unit participating in primary control starts to actuate automatically to restore frequency when it deviates at least >10 mHz. The 5% governor droop calls for full output within 3 Hz (5% of 50 Hz) [39]. This mode of regulation is known as speed-droop regulation and is activated to increase turbine output power and regulate the generator frequency.

Similarly, as in the inertia response, in Figs. 10a and b three coherent groups and their associate dynamic trajectories can be observed. The uniform coherent behaviour is the result of the control influence over the system. Additionally, from Fig. 10c the following can be concluded: cluster c_2 is the largest group, which is mostly integrated by generators in CH, DE, and IT, cluster c_1 is formed entirely by ES and, as in the previous case, c_3 is constituted only by generators in TR. Fig. 11 shows the geographical information of these clusters.

To complement this analysis, the average trajectory clusters and the average slopes defined in (12) were computed for both previous cases and the results are presented in Fig. 12. The average trajectory clusters displayed in Figs. 12a and b illustrate the effect of deceleration and acceleration of the frequency responses. In addition, the magnitudes of slope values $H_i^{\pm}(T_w)$ are displayed in Figs. 12c and d. From the subplots, it can be noticed that the oscillatory process is decreasing, resulting in a new equilibrium point. These results confirm previous dynamic studies of the ENTSO-E system [33, 34], where low-frequency modes were identified.

4.4 Study case two: unstable response

Although the actual system is highly robust and stable, the objective of this subsection is to test the algorithms developed in Section 3 under extreme conditions. For this reason, a few model simplifications were performed (mainly with respect to automatic
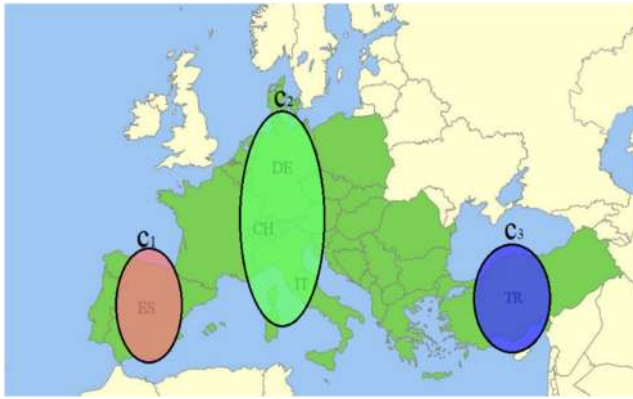
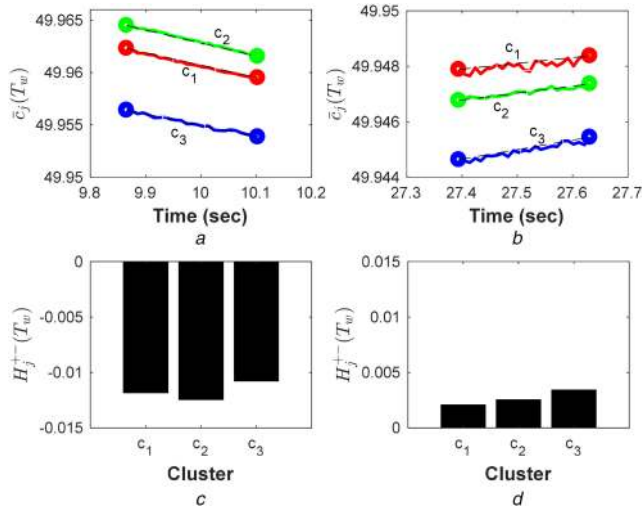**Fig. 11** *Geographical location of clusters at control system response*

Fig. 12 *Case 1 stable*
(a), (c) Average cluster and average slope at the inertial response, *(b), (d)* Average cluster and slope at the control action system

Fig. 13 *Frequency response fault as the CH-IT_1 line trip*

Fig. 14 *Case 2 unstable: online indexes*
(a) Ascent and descent of the slopes, *(b)* GSI, *(c)* Number of clusters determined by $G_{s(T_w)}$, *(d)* Alarm index to indicate the transient period

voltage regulator (AVR) settings) to illustrate a fictitious operating condition of the system between Switzerland (CH) and Italy (IT). Based on this, one of the CH-IT interconnections was selected to simulate a three-phase line trip, which subsequently is going to be referred to as the CH-IT_1 line. Fig. 13 depicts 3 s of simulation time for the selected areas. The first 5 s have been omitted and the evolution and effects of the fault are displayed. Similarly, as in the previous case, a line trip on CH-IT_1 is applied at 6 s and is assumed that the line is not back in service again. As is observed, the system becomes immediately unstable after a few milliseconds. The traces show that some machines in IT and CH are the most affected elements after this severe event.

Following the validation of the proposed algorithms, two different simulation stages have been selected to be evaluated under unstable conditions and are highlighted in Fig. 13 as (a) fault initiation and (b) fault evolution.

Fig. 14 depicts the online indexes proposed in Section 2.4. Fig. 14*a* displays the average ascent $Iac(T_w)$ and descent $Iad(T_w)$ dynamic slopes; from the results, it is possible to observe that $Iad(T_w)$ has identified correctly the unstable dynamic behaviour at 6 s. This result is confirmed by the GSI, which can be observed in Fig. 14*b*. Two additional indexes are displayed in Figs. 14*c* and *d*, respectively. The two indexes describe the number of clusters and the alarm raised to detect the transient period.

Similarly, Fig. 15 presents the results for the two simulation stages described above and depicted in Fig. 14. At the fault initiation stage; Fig. 15*a* displays the number of clusters identified (three clusters) and Fig. 15*b* shows the large separation of cluster c_3 (IT(310)) from the rest of the system. Furthermore, in the analysis of the evolution of the fault; three well-defined groups are shown in Fig. 15*c*, two of them c_3 (IT (310)) and c_1 (CH (3)) are the most
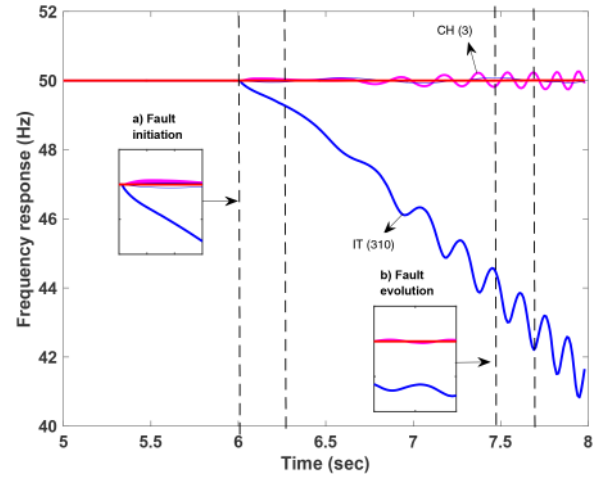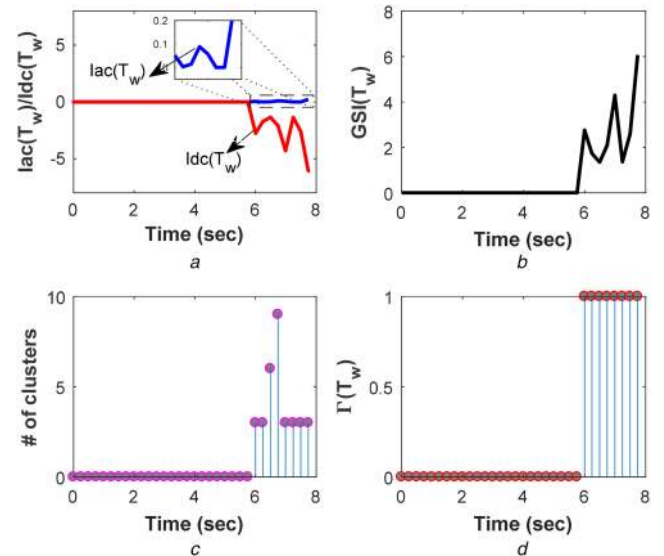
notorious groups and it was expected c_3 (IT (310)) remains separated from the rest of the system.

To complement the previous observations, the average trajectory cluster and the average slope cluster $H_i^+(T_w)$ were computed for both selected periods. The results are depicted in Figs. 16*c* and *d*, respectively. From Figs. 16*a* and *c*, it can be noticed that that clusters c_1 (CH) and c_3(IT(310)) represent the most pronounced decrements on frequencies responses, in addition, generators from cluster c_2 (DE, IT, ES, and TR), increment their frequency responses attempting to compensate the unbalanced condition.

The graphical location of these clusters is depicted in Fig. 17. Consequently, during the fault evolution, the controllers try to return the system to balance; Figs. 16*b* and *d* show two parallel oscillatory processes. The first shows how cluster c_1 (CH(3)) is increasing its speed up with respect to cluster c_2 (IT, DE, ES, TR), and the second one confirms the separation of c_3 (IT(310)) indicating the unstable condition of the system.

4.5 Comparisons with other multivariate clustering methods

In order to demonstrate the effectiveness of the proposed approach (SVD + Slope) a performance comparison with other multivariate clustering methods, such as PCA [8], KM [18] and DMD [17] is presented. As described in [17], the spatial signatures of the proposed approach, PCA, KM, and DMD, contained in the modal vectors u_j, $\tilde{\phi}_j$, $\Re\{v_j\}$ and $\Re\{\phi_j\}$, provide valuable information
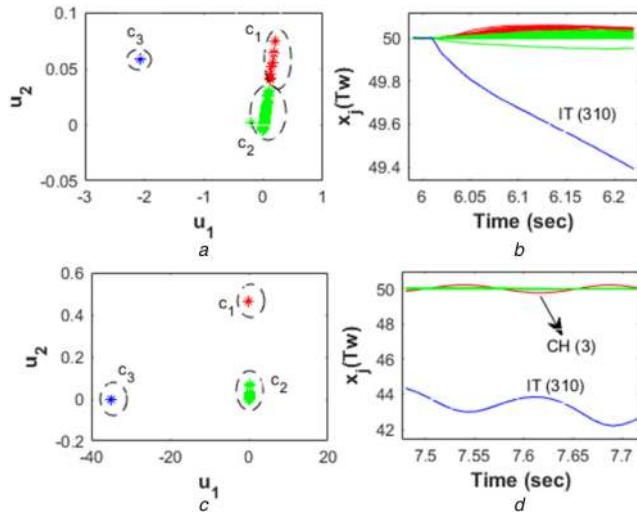
Fig. 15 *Case 2 unstable*
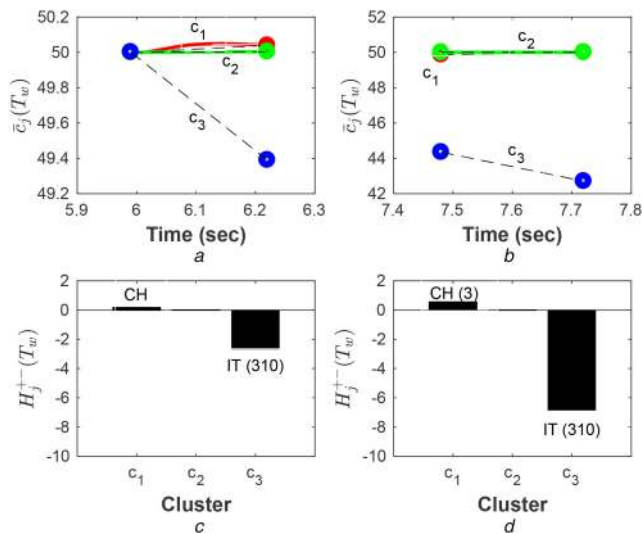*(a), (b)* Clustering visualisation at fault system, *(c), (d)* Clustering visualisation at fault evolution

Fig. 16 *Case 2 unstable*
(a), *(c)* Average slope cluster at fault initiation, *(b)*, *(d)* Average slope cluster fault evolution
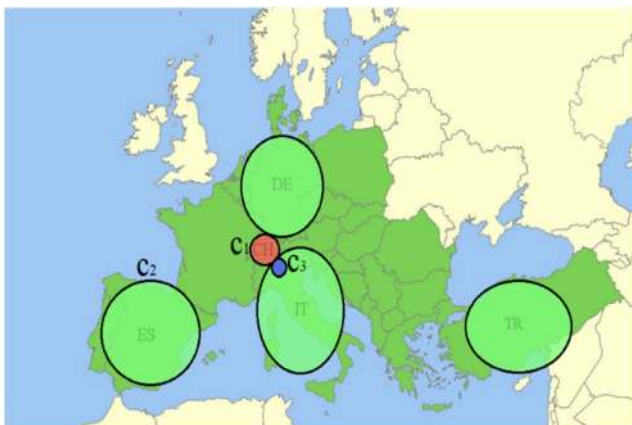

Fig. 17 *Geographical location of clusters at fault*

Table 1 Coherent groups and their CPU time comparison

Method	Proposed	PCA	KM	DMD
#cluster	c_1, c_2, c_3	c_1, c_2, c_3	c_1, c_2, c_3	c_1, c_2, c_3
CPU time	$\simeq 0.15$ s	$\simeq 0.06$ s	$\simeq 5.0$ s	$\simeq 1.20$ s

The CPU time used by the K-means algorithm was not included in this table.

about the coherency groups involved in the oscillatory dynamic process. Since the coherency identification process using the aforementioned multivariate methods is not straightforward, K-means algorithm is uniformly used to process the spatial signature information obtained from the proposed and alternative multivariate methods to identify the optimal number of groups that participate in the coherency process.

Table 1 shows the final number of coherent groups identified by each algorithm and the CPU time required. The algorithms were applied to the simulation presented in Section 4.3, which corresponds to the stable case. A window of 10 s length (6–16 s) was used to analyse the development of the coherency groups during the different control phases of the frequency response. The length of the window (10 s) was selected to allow all algorithms to identify the low-frequency modes that characterise the oscillatory process on the ENTSO-E system. It is worth mentioning that the same comparison is omitted for the unstable case because most of the alternative approaches require a long time window (post-fault) in order to capture the dynamic behaviour of the low-frequency oscillations, which is not possible to have for the unstable case (see Fig. 13).

It can be noticed that all multivariate methods identified three coherent groups, where all generators in TR constitute group c_3, generators in IT, CH, and DE conform group c_2, and generators in ES conform group c_1. The results of Table 1 are depicted in Figs. 18 in the form of 2D clustering visualisation plots. Figs. 18a and b present the score plots of the proposed approach and PCA, respectively. Figs. 18c and d show the score plots of the alternative methods KM and DMD, respectively, all methods for the two low-frequency modes ($\simeq 0.10$ and $\simeq 0.22$ Hz) presented in the ENTSO-E system [31, 34].

As shown in Table 1, PCA presents the shorter CPU time in comparison with the rest of the algorithms. However, the other multivariate approaches (proposed, KM, and DMD) present a better separation among the groups as it can be seen in Figs. 18a, c, and d, respectively. The benefit of the separation is reflected during the application of K-means, which results in more accurate identification of the groups involved during the coherency process. Moreover, these methods provide additional information such as the slope of the signal (proposed), and frequency, damping, energy mode and participation factors (KM and DMD), which cannot be obtained with PCA.

The main drawback with KM and DMD algorithms is the length of the window required (larger) to capture slow electromechanical modes (in the range of 0.1 Hz). On the other hand, SVD based and PCA algorithms can process information faster, using shorter windows of time due to their ability to reduce high-dimensional datasets into smaller dimensions while retaining the most significant information. For these reasons, SVD and PCA have the potential to be applied in combination with other methods such as in the proposed approach. The result is to provide new ways to develop systematic tools to identify coherency groups and global stability indexes. By combining algorithms in an innovative way as demonstrated in this brief, it is possible to create more sophisticated and robust PCA variants such as the one proposed in the literature for only coherency identification [14, 16].

5 Conclusion

This study presents a novel technique for coherency identification and stability condition for large interconnected power systems based on analysing the generator frequency responses from multiple generators simultaneously. The proposed online data mining methodology makes several contributions and its novelty resides mainly on three different factors: (i) a security index based on clustering that can be used for online islanding detection, (ii) the novel security indicator exploits the concept of acceleration and deceleration of the synchronous machines to estimate online the coherent groups and classify the islanding and (iii) an online visual dynamic information to determine the power system stability status and evolving oscillatory coherency phenomenon mechanism.

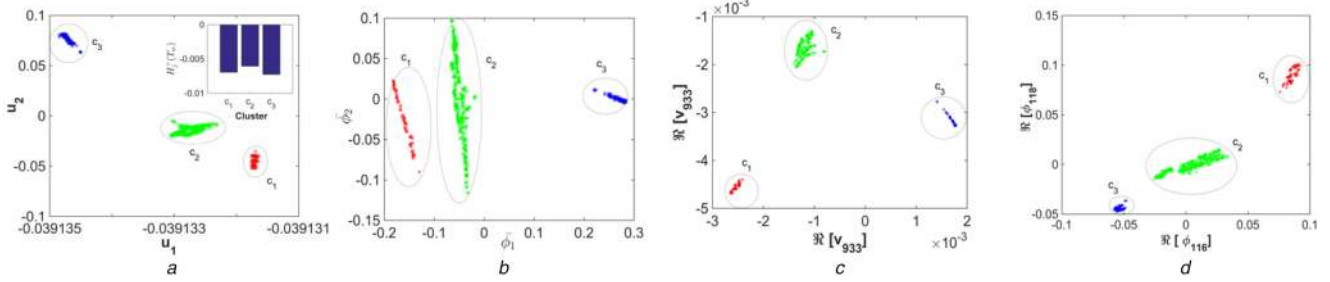The effectiveness of the proposed method using an unsupervised data mining technique was verified through

IET Gener. Transm. Distrib., 2019, Vol. 13 Iss. 15, pp. 3323-3333

3331

Fig. 18 *Comparison of clustering among methods for case 1; 2D clustering visualisation:18*
(a) Proposed, *(b)* PCA, *(c)* KM, *(d)* DMD

DIgSILENT PowerFactory simulations using the complex initial dynamic model of ENTSO-E. Two case studies were selected to highlight the contributions of the proposed method with respect to the existing state-of-the-art methods; the proposed approach detects onset of the transient events, determines the stability status of the system under investigation, provides coherency dynamic information and analytical information about how the electromechanical oscillations are evolving; enabling the development of a systematic tool for an adaptive scheme of wide area anti-islanding protection.

Future work includes to extend the proposed algorithm to improve its robustness in the presence of high noise using more sophisticated and robust dimensional reduction techniques and cascading failures; to develop an adaptive threshold γ. In addition, non-linear dimensionality reduction techniques could also be incorporated to improve the separation information and facilitate the clustering process.

6 Acknowledgments

7 References

[1] Grebe, E., Kabouris, J., Lopez Barba, S., *et al.*: 'Low frequency oscillations in the interconnected system of continental Europe'. IEEE PES General Meeting, 2010, Calgary, AB, Canada, July 2009, pp. 1–7

[2] Khalil, A.M., Iravani, R.: 'A dynamic coherency identification method based on frequency deviation signals', *IEEE Trans. Power Syst.*, 2016, **31**, (3), pp. 1779–1787

[3] Ayon, J.J., Barocio, E., Cabrera, I.R., *et al.*: 'Identification of coherent areas using a power spectral density algorithm', *Electr. Eng.*, 2017, **100**, (2), pp. 1–11

[4] Mei, K., Rovnyak, S.M., Ong, C.M.: 'Clustering-based dynamic event location using wide-area phasor measurements', *IEEE Trans. Power Syst.*, 2008, **23**, (2), pp. 673–679

[5] Giannuzzi, G., Pisani, C., Sattinger, W.: 'Generator coherency analysis in ENTSO-E continental system: current status and ongoing developments in Italian and Swiss case', *IFAC-PapersOnLine*, 2016, **49**, (27), pp. 400–406

[6] Isazadeh, G., Khodabakhshian, A., Gholipour, E.: 'A new intelligent wide area controlled islanding detection method in interconnected power systems', *Int. Trans. Electr. Energy Syst.*, 2017, **27**, (7), pp. 1–11

[7] Dahal, O.P., Member, S., Brahma, S.M., *et al.*: 'Comprehensive clustering of disturbance events recorded by phasor measurement units', *IEEE Trans. Power Deliv.*, 2014, **29**, (3), pp. 1390–1397

[8] Anaparthi, K.K., Chaudhuri, B., Thornhill, N., *et al.*: 'Coherency identification in power systems through principal component analysis', *IEEE Trans. Power Syst.*, 2005, **352**, (3), pp. 30–44

[9] Miah, A.M.: 'Simple dynamic equivalent for fast online transient stability assessment', *IEE Proc., Gener. Transm. Distrib.*, 1998, **145**, p. 49

[10] Aghamohammadi, M.R., Tabandeh, S.M.: 'A new approach for online coherency identification in power systems based on correlation characteristics of generators rotor oscillations', *Int. J. Electr. Power Energy Syst.*, 2016, **83**, pp. 470–484

[11] Tang, K., Venayagamoorthy, G.K.: 'Online coherency analysis of synchronous generators in a power system'. IEEE PES Innovative Smart Grid Technologies Conf. (ISGT 2014), Washington, DC, USA, 2014, pp. 1–5

[12] Thakallapelli, A., Hossain, S.J., Kamalasadan, S.: 'Coherency and online signal selection based wide area control of wind integrated power grid', *IEEE Trans. Ind. Appl.*, 2018, **9994**, (c), pp. 1–11

[13] Ariff, M.A.M., Pal, B.: 'Coherency identification in interconnected power system-an independent component analysis approach', *IEEE Trans. Power Syst.*, 2012, **28**, (2), p. 1

[14] Liu, X., Kennedy, J.M., Laverty, D.M., *et al.*: 'Wide-area phase-angle measurements for islanding detection-an adaptive nonlinear approach', *IEEE Trans. Power Deliv.*, 2016, **31**, (4), pp. 1901–1911

[15] Zhu, Q., Chen, J., Duan, X., *et al.*: 'A method for coherency identification based on singular value decomposition'. 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 2016, pp. 1–5

[16] Lin, Z., Wen, F., Member, S., *et al.*: 'WAMS-based coherency detection for situational awareness in power systems with renewables', *IEEE Trans. Power Syst.*, 2018, **33**, (5), pp. 5410–5426

[17] Barocio, E., Pal, B.C., Thornhill, N.F., *et al.*: 'A dynamic mode decomposition framework for global power system oscillation analysis', *IEEE Trans. Power Syst.*, 2015, **30**, (6), pp. 2902–2912

[18] Susuki, Y., Mezic, I.: 'Nonlinear Koopman modes and coherency identification of coupled swing dynamics', *IEEE Trans. Power Syst.*, 2011, **26**, (4), pp. 1894–1904

[19] Juarez, C., Member, S., Messina, A.R., *et al.*: 'Characterization of multimachine system behavior using a hierarchical trajectory cluster analysis', *IEEE Trans. Power Syst.*, 2011, **26**, (3), pp. 972–981

[20] Jonsson, M., Begovic, M., Daalder, J.: 'A new method suitable for real-time generator coherency determination', *IEEE Trans. Power Syst.*, 2004, **19**, (3), pp. 1473–1482

[21] Gomez, O., Rios, M.A.: 'Real time identification of coherent groups for controlled islanding based on graph theory', *IET Gener. Transm. Distrib.*, 2015, **9**, (8), pp. 748–758

[22] Lin, Z., Wen, F., Ding, Y., *et al.*: 'Data-driven coherency identification for generators based on spectral clustering', *IEEE Trans. Ind. Inf.*, 2018, **14**, (3), pp. 1275–1285

[23] Agrawal, R., Thukaram, D.: 'Support vector clustering-based direct coherency identification of generators in a multi-machine power system', *IET Gener. Transm. Distrib.*, 2013, **7**, (12), pp. 1357–1366

[24] Wei, J., Kundur, D., Butler-Purry, K.L.: 'A novel bio-inspired technique for rapid real-time generator coherency identification', *IEEE Trans. Smart Grid*, 2015, **6**, (1), pp. 178–188

[25] Lin, Z., Wen, F., Ding, Y., *et al.*: 'Wide-area coherency identification of generators in interconnected power systems with renewables', *IET Gener. Transm. Distrib.*, 2017, **11**, (18), pp. 4444–4455

[26] Mahat, P., Chen, Z., Bak-Jensen, B.: 'Review on islanding operation of distribution system with distributed generation'. 2011 IEEE Power and Energy Society General Meeting IEEE, Detroit, MI, USA, 2011, pp. 1–8

[27] Kamwa, I., Samantaray, S.R., Joós, G.: 'Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records', *IEEE Trans. Power Syst.*, 2009, **24**, (1), pp. 258–270

[28] Alam, M.R., Muttaqi, K.M., Bouzerdoum, A.: 'An approach for assessing the effectiveness of multiple-feature-based SVM method for islanding detection of distributed generation', *IEEE Trans. Ind. Appl.*, 2014, **50**, (4), pp. 2844–2852

[29] Aggarwal, C.C., Reddy, C.K.: 'DATA CLUSTERING algorithms and applications edited by', 2014

[30] Rousseeuw, P.J.: 'Silhouettes: a graphical aid to the interpretation and validation of cluster analysis', *J. Comput. Appl. Math.*, 1987, **20**, pp. 53–65

[31] ENTSO-E: 'Analysis of Continental Europe Inter-Area Oscillations on 1st December 2016', 2017

[32] Group T.E.-E.P.: 'Report on Blackout in Turkey on 31st March 2015', 2015

[33] Semerow, A., Sebastian, H., Luther, M., *et al.*: 'Dynamic study model for the interconnected power system of continental Europe in different simulation tools', *IEEE Eindhoven PowerTech*, 2015, **977**, pp. 1–6

[34] Segundo Sevilla, F.R., Korba, P., Uhlen, K., *et al.*: 'Evaluation of the ENTSO-e initial dynamic model of continental Europe subject to parameter variations'. 2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conf. (ISGT 2017), Arlington, VA, USA, 2017, pp. 1–2

[35] IEEE: 'Standard for synchrophasor measurements for power systems', Revision of IEEE Std. C37.118–2005, 2011, pp. 1–61

[36] IEEE: 'Standard for interconnecting distributed resources with electric power systems', June, Revision IEEE 1547–2003, 2003, pp. 1–28

[37] IEC: 'Standard photovoltaic (PV) systems – characteristics of the utility interface' 2009, IEC Std. 61727

[38] Ge, H., Guo, Q., Sun, H., *et al.*: 'Multivariate statistical analysis-based power-grid-partitioning method', *IET Gener. Transm. Distrib.*, 2015, **16**, pp. 1023–1031

[39] ENTSO-E: 'Frequency measurement requirements and usage', January, Final Version 7, 2018, pp. 1–11

3332

IET Gener. Transm. Distrib., 2019, Vol. 13 Iss. 15, pp. 3323-3333

8 Appendix

In order to calibrate the threshold γ of the proposed method and detect a transient event, an off-line statistical procedure is proposed as follows:

1. Add white Gaussian noise to each vector signal $(x_j(t),\ j = 1, 2\ldots, m)$ using the Matlab function 'awgn'.
2. Generate multiple data sets by adding independent noise to model (1).
3. Compute GSI using (15).
4. Construct a Q–Q plot with maximum values of the GSI obtained from the data sets generated at step 2.
5. Estimate the histogram with maximum values of the GSI obtained from the data sets generated at step 2.
6. Determine the value of the threshold γ using the statistical information provided in steps 4 and 5.

For clarity, in this study, 1000 runs for $T_w = 10$ s and a sampling frequency of 100 Hz were performed. In all cases, Gaussian noise of 60 dB was added to each set to simulate ambient operation conditions.

Fig. 19a shows the GSI values for the 1000 data sets simulated with independent noise. Fig. 18b displays the Q–Q plot that allows evaluating if the maximum values of the GSI follow a linear distribution and Fig. 19c displays the computed histogram. From Figs. 19a and b, the linear and Gaussian distribution of the GSI values can be observed. Based on these facts, a value of
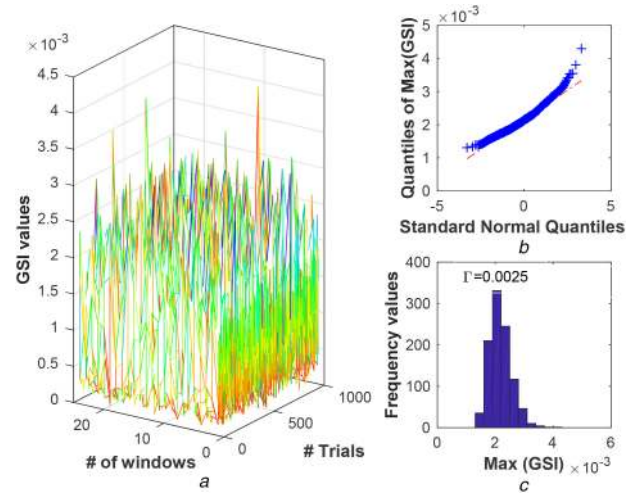

Fig. 19 *Calibration of the threshold γ*
(a) 1000 simulations to compute GSI under noisy signals, *(b)* Q–Q-plot of the maximum values of GSI trial, *(c)* Histogram of maximum values of each trial GSI

$\gamma = 25 \times 10^{-3}$ is selected from the information provided in Fig. 16c. This value guarantees that 90% of the GSI data obtained from the ambient operation simulations follow the linear and Gaussian properties and remain below the selected threshold. Based on this statistical analysis, it is expected to reduce the number of false alarms during the online tracking process.