IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 October 2023, accepted 24 October 2023, date of publication 26 October 2023, date of current version 3 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327916

RESEARCH ARTICLE

A Robust Spectral Clustering Method Based on PMU Measurements for Coherent Areas Identification

ANNALISA LICCARDO [1], (Member, IEEE), DAVIDE LAURIA [2], FRANCESCO BONAVOLONTÃ [1], GIORGIO MARIA GIANNUZZI [3], COSIMO PISANI [3], AND SALVATORE TESSITORE [3]

[1]Department of Electrical Engineering and Information Technology (DIETI), University of Naples Federico II, 80125 Naples, Italy
[2]Department of Industrial Engineering (DII), University of Naples Federico II, 80125 Naples, Italy
[3]Terna Rete Italia, 00156 Rome, Italy

Corresponding author: Annalisa Liccardo (annalisa.liccardo@unina.it)

This work was supported by the Research Agreement " Wide Area Monitoring System (WAMS) Digital Twin for Comparison and Optimization of Algorithms for Parameter Estimation of Interarea Oscillations" between the Department of Electrical Engineering and Information Technology and Department of Industrial Engineering and the National Inter-University Consortium for Energy and Electrical Systems (EnSiEL).

ABSTRACT The paper deals with the separation of power system into coherent areas; this is a relevant issue for managing the network in both normal operating conditions and during anomalous events. In particular, the attention is focused on partitioning the power system in such a way as to group together frequency signals, measured by means of phasor measurement units (PMU), exhibiting similar oscillatory behavior after the occurrence of a fault or disturbance. Unfortunately, the increasingly massive presence of renewable energy sources is undermining the clustering methods defined so far, requiring new solutions to the problem. To overcome the considered drawbacks, the authors propose hereinafter to (*i*) improve the grouping capabilities of an iterative spectral clustering method thanks to the definition of new parameters for similarity estimation (Modified Bray Curtis index) and cluster thresholding (weighted Fiedler value) as well as (*ii*) enhance its robustness with respect to both measurement noise and uncertainty affecting the PMUs by means of a deep test procedure. To this aim, particular attention is paid in the design and assessment stage to the definition of both filtering algorithm and measurement parameters (e.g., the length of the analysis window). Once defined these parameters, the method is capable of 100% correctly separating transmission network sections oscillating with similar trends in a number of tests conducted on simulated and actual signals, so highlighting the promising performance of the method highlighting its reliability and efficacy in different test conditions.

INDEX TERMS Frequency oscillations, interarea oscillations, PMU measurements, power transmission network, spectral clustering.

I. INTRODUCTION

The division of a power system into coherent electrical areas is a fundamental issue, that allows to effectively manage the network in normal operation as well as during anomalous events.

During normal operating conditions, the network partitioning, for example, allows model reduction to simplify its state analysis; in addition, the inertia of a given area can be estimated. Given the increasing integration of renewable energy sources, in fact, it becomes imperative to dynamically estimate the inertia of each individual area within the power system [1], [2], in order to predict its capability to dampen frequency fluctuations and take possible measures to increase inertia [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Padmanabh Thakur.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

As regards the management of network events, this approach enables the implementation of advanced strategies such as islanding, which involves isolating specific sections of the power network to maintain stability and minimize the impact of disturbances. By dividing the system into distinct coherent areas, operators can strategically control and optimize power flow to ensure reliable and secure operation.

The authors' research is particularly focused on the use of partitioning in coherent areas with the aim of damping interarea oscillatory modes [4], [5], [6], which are low-frequency oscillations that can propagate across multiple regions of the power system. By identifying and analyzing the coherent groups within each area, control actions can be coordinated to suppress these oscillations and maintain system stability.

A coherent group can be identified as a set of generators that exhibit similar oscillatory behavior after the occurrence of a disturbance is usually indicated as coherent group [7]. Generators within a coherent area display similar amplitude and phase characteristics when participating in a common set of inter-area modes ranging from 0.1 to 1 Hz. This inherent coherence within the group enables efficient coherency detection, allowing operators to identify and analyze the behavior of these interconnected generators as a collective unit.

In the presence of intraarea oscillations (involving frequency components lying within 1-3 Hz), identifying coherent areas turns out to be more challenging. The reason must be found in the different behavior that generators exhibit with respect to interarea modes (that are highly correlated) and intraarea oscillations (that can vary significantly). This variability in intraarea oscillations poses a significant hurdle in accurately determining coherency and necessitates advanced algorithms and techniques for precise detection and analysis [8].

As a further complication, dynamic variations in system coherency are introduced by the integration of wind and solar energy sources. Their intermittent nature and inherent fluctuations in generation levels can affect the coherence of the interconnected generators. As a result, coherency detection algorithms and methodologies need to adapt to these dynamic changes in order to maintain an accurate understanding of the system's behavior and ensure reliable operation.

As the integration of renewable energy sources continues to grow, it becomes increasingly important to develop coherency detection techniques that account for the unique characteristics and challenges posed by renewable intermittency. Future research efforts should focus on refining existing methods and developing novel approaches that consider the dynamic nature of renewable generation and ensure robust coherency detection in modern power systems.

In the proposed approach, the paradigm of spectral clustering is followed, thus providing a complete change of viewpoint in terms of analysis domain. In particular, the proposed algorithm primarily relies on spectral analysis of the Laplacian of the graph generated from the similarity matrix [9]. A similar solution to the problem of coherent area identification has been presented in [10], where the spectral clustering approach is applied on the equivalent graph of transmission network to single out the optimal location for placing a PMU; the main drawback is the need of knowledge of the network model. On the contrary, the proposed method only exploit the frequency measurements of PMU already deployed on the network thus making it independent from the network model and/or topology.

Differently from approaches based on spectral clustering available in literature, the coefficient of the similarity matrix are evaluated according to Bray Curtis index, in order to obtain indices within 0 and 1, expressing the similarity between measured data. This coefficient are evaluated by directly processing the frequency measures obtained by PMUs. By leveraging the Laplacian matrix, which characterizes the graph, the spectral decomposition is performed, enabling the identification of spectral mappings and subsequent partitioning of the data and PMUs into coherent clusters exhibiting homogeneous oscillatory behavior.

Through this approach, the algorithm effectively captures the underlying patterns and dynamics present in the PMU data. By analyzing the spectral properties of the Laplacian matrix, it becomes possible to identify coherent groups or clusters that exhibit similar oscillation characteristics. This partitioning process enables a more accurate and detailed understanding of the system's behavior, facilitating efficient analysis and control strategies.

Particular attention has been paid on making the proposed method robust with respect to both measurement noise and uncertainties affecting the PMU, that usually make the literature solutions unreliable. To this aim, the data coming from the PMU are first bandpass filtered by means of a filter whose frequency range is suitably tailored with the band of interest. Moreover, a proper study is conducted to identify the best length of the observation interval capable to provide the most reliable results in the final coherent area separation. The paper is organized as follows: solutions already available in the literature are first presented and discussed in Section II, while the theoretical background as well as the operating steps of the proposed method are described in Section III; the results obtained by the application of the method on numerical, simulated networks and actual data are presented and discussed in Section IV, V and VI. Conclusions are finally drawn in Section VII.

II. RELATED WORKS

Methods for identifying coherent areas are usually based on the measurement data provided by PMUs [11], [12]. With their high data delivery rate, they provide a comprehensive view of system dynamics and allows for a detailed visualization of how coherency evolves over time [13]. However, there are significant challenges associated with leveraging PMU data to support corrective control strategies at the area level.

One critical issue revolves around the accuracy of estimation. PMUs need to deliver precise measurements while effectively dealing with distortion levels and rapid frequency variations [14]. Additionally, during transient phenomena characterized by a continuous spectrum, PMUs often assume that the fundamental component can be approximated as a narrowband component. This assumption may lead to partial or unsatisfactory information derived from PMU measurements [15].

Coherency detection techniques in the context of power systems have traditionally relied on multivariate data analysis methods such as Principal Component Analysis (PCA) [16] and Independent Component Analysis (ICA) [17]. However, these techniques have predominantly been tested and applied in conventional power systems, where the presence or predominance of renewable energy sources is not accounted for.

To address the challenges posed by renewable integration, alternative methods have emerged that utilize advanced signal processing techniques. As an example, the Hilbert-Huang Transform [18] and Wavelet Transform (WT) [19] have been employed in coherency detection approaches. Some measurement methods leverage similarity features combined with clustering techniques, such as Discrete Fourier Transform (DFT) with subtractive clustering [20]. While these approaches may consider the presence of renewable sources in some cases, they often overlook the intermittent nature and dynamic characteristics of renewable generation.

Several available methods, including the Radial Basis Neural Network [21] and Dynamic Coherency Detection (DCD) [22], rely on threshold-based approaches, which can be challenging to set appropriately in the presence of variable renewable generation. To optimize the performance of such threshold-dependent algorithms, it is essential to configure the algorithm to achieve the widest possible threshold range, minimizing the potential for algorithm confusion.

More recently, other approaches for coherent areas clustering have been presented in the literature. As an example, in [23] the optimal partition for islanding is defined, giving a specific score to each partition. The Laplacian is evaluated on the weights of each line and cluster extension; therefore, the method does not exploit actual measurement data and requires that the distribution line model has to be a-priori known. Clustering and controlled islanding are carried out in [24] and [25] by operating on the generators rotor angle, a signal characterized by high signal-to-noise ratio. Unfortunately, all the provided results refer to only simulation study (as also in [26]) and no example of methods application to measurements data coming from actual devices and network is given. Rotor angles and speed deviations are the inputs of the method presented in [27]. Two limitations are related to (*i*) the need of knowing the network model (the system is partitioned into two areas whose equivalent power and inertia is estimated) and (*ii*) the exploited thresholding approach that should fail in the presence of noisy input signals.

III. THE PROPOSED METHOD

A. THE SPECTRAL CLUSTERING APPROACH

Spectral clustering (SC) is graph-based method, that transforms a clustering problem into a graph segmentation problem, without any assumption on the number or the form of the data clusters. The graph is defined by its M vertices (or nodes) v_i and by the edges linking the nodes; the edges are characterized by their weights, i.e. a value within 0 and 1, indicating the strength of the link between the vertices [28].

The SC goal is separating the graph in two sub-graphs in order to group nodes that are similar to each other. At this aim, the edges value are set according to a similarity matrix $\mathbf{W}$, whose entries $\omega_{ij} = \omega_{ij}$ quantify the similarity degree between each pair of nodes v_i and v_j.

In the proposed method, the vertices are the sets of frequency measures $x^{(k)}$, with $k = 1, \ldots, N$, provided by the M PMUs of the network. The method aims at grouping together in the same sub-graph frequency measures that evolve over time with similar behavior.

Several approaches exist to construct the similarity matrix. Since the authors aim at singling out areas with the same frequency fluctuations, the Modified Bray Curtis (MBC) index is used [29]. The Bray Curtis index is defined as [30]:

$$BC_{ij} = \frac{\sum_{k=1}^{N} \left| x_i^{(k)} - x_j^{(k)} \right|}{\sum_{k=1}^{N} \left| x_i^{(k)} + x_j^{(k)} \right|} \tag{1}$$

The more similar the signal pattern over time, the lower the BC index. In order to obtain a similarity parameter between 0 and 1, MBC, defined in (2) is evaluated as the BC complement of 1. If, however, the signals are very different from each other, the Bray Curtis index becomes greater than 1 and its 1-complement has no meaning. Therefore, Bray Curtis values greater than 1 are discarded (forcing the corresponding MBC value to 0). The MBC values are exploited as entries of the similarity matrix W.

$$MBC_{ij} = \begin{cases} 1 - BC_{ij} & if \quad BC_{ij} \leq 1 \\ 0 & if \quad BC_{ij} > 1 \end{cases} \tag{2}$$

With this index, it is also possible to separate areas that contribute to the same oscillation, but with different phases; as an example, this typically occurs when groups of generators swing against each other, exhibiting opposite frequency variations. In this case, in fact, the BC assumes very high value (since the denominator tends to zero) and, consequently, the MBC goes to zero. If a similarity index based on correlation coefficient were used (as in [31], [32]), the absolute value would have to be made, because according to the spectral clustering theory, the edges value has to be non negative; as a consequence, those generators would be considered part of a single cluster.

From the similarity matrix, the degree matrix $\mathbf{D}$ is evaluated. It is a diagonal matrix, whose entries on the main diagonal are the sum of the edge weights related to that node:

$$d_i = \sum_{j=1}^{M} \omega_{ij} \tag{3}$$

In graph theory, a fundamental role is played by the Laplacian. In this paper, the non-normalized Laplacian defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \tag{4}$$

has been taken into account. The spectrum of the Laplacian, i.e. the eigenvalues λ_i $(i = 1, .., M)$ and eigenvectors Ψ_i obtained from the eigendecomposition of the Laplacian, exhibits important properties [33]. In particular, if the first eigenvalue λ_1 is zero, the graph is partitionable. More specifically, the number of eigenvalues equal to zero (or approximately zero) defines how many clusters ideally make up the graph. The number of clusters, therefore, can be identified by locating the so-called spectral gap, i.e. the jump between the near zero eigenvalues and those with value significantly different from zero. The second eigenvalue, λ_2, which is referred to as Fiedler value or algebraic value in the literature, is strongly related to the graph connectivity and, also, it informs about the intensity of the connections between the nodes of the graph. Low values of λ_2 show that the graph can be partitioned since it involves nodes with low similarity [28].

The second eigenvector of the Laplacian Ψ_2 (called Fiedler vector), on the other hand, can be used to perform partitioning. This, in fact, consists of negative values (which are poorly connected to the first node) and non-negative values, which are more connected to the first node, thus providing a straightforward criterion for cluster separation.

B. CLUSTER PARTITIONING

The proposed method examines the Fiedler value for deciding whether a cluster is partitionable or not; in particular, if this value is sufficiently low, the cluster is partitionable. To determine whether λ_2 is sufficiently low or not, a comparison must be made with a predefined threshold λ_2^{thr}. As in all criteria using a threshold, the selection of the right threshold is an essential aspect. In the case of Fiedler value, this aspect is even more crucial because λ_2 also takes into account the strength of the connections in the group and can be sensitive to the number of nodes (i.e. PMUs). For this purpose, an additional criterion was introduced to increase the reliability of the clustering. In particular, the cut value is defined as follows

$$Cut(A, B) = \sum_{i \in A, j \in B} \omega_{ij} \tag{5}$$

where A and B are the two groups in which the entire cluster has been separated according to the associated Fiedler vector [36]. The cut value expresses the total weight of the edges connecting the nodes belonging to the sub-parts A and B and, as such, it defines the quality of the partition. However, according to (5), it also depends on the number of nodes that make up the two sub-groups. Then a further parameter taken into account is the weighted Fiedler value, $w\lambda_2$, evaluated as the ratio between the Fiedler and the cut value:

$$w\lambda_2 = \frac{\lambda_2}{Cut(A, B)} \tag{6}$$

The algorithm, hence, will perform the partition if both the two following conditions are met:

$$\lambda_2 \leq \lambda_2^{thr} \quad and \quad w\lambda_2 \leq w\lambda_2^{thr} \tag{7}$$

C. DEVELOPED ALGORITHM

Spectral clustering is a bi-partitioning method. This means that if there are more than two clusters, an iterative procedure is required. A flow chart of the realized algorithm is shown in Fig. 1. The algorithm works on a queue of clusters to be analyzed, which is updated during execution. Initially, the queue contains only the initial group of PMUs, called *cluster*, which is removed from the queue and analyzed. All the signals involved in the initial cluster are subjected to a pre-processing phase, described in [5], in which the following operations are performed: 1) identification of any missing data and their filling using a linear interpolation method; 2) removal of the offset, to obtain zero mean oscillations; 3) detrend to remove slow trends that lead to a bias in the offset of the signals.

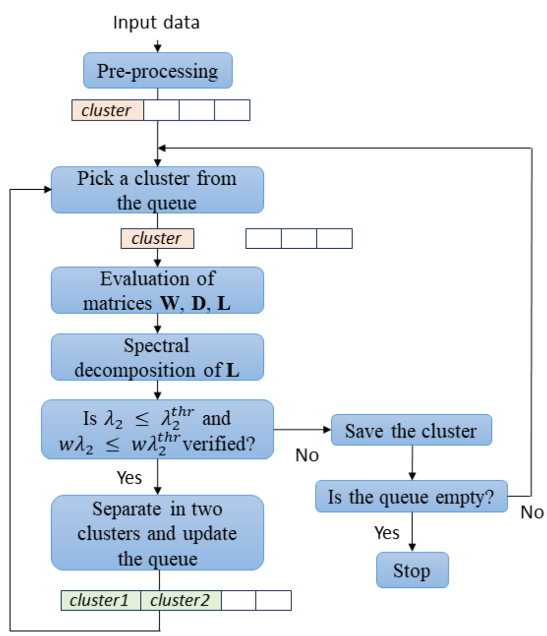

FIGURE 1. Flow chart of the developed algorithm.

Successively, the similarity matrix, the degree matrix and the Laplacian are evaluated. Then the eigenvalue decomposition of the Laplacian is performed. If both the Fiedler value and the weighted Fiedler value are sufficiently low, i.e. below the predetermined thresholds, this means that the group is separable into two clusters. These will be placed in the queue under the name *cluster1* and *cluster2*. Conversely, if λ_2 and $w\lambda_2$, are above their thresholds, the analyzed group is not separable. If the queue of clusters to be analyzed is empty, the algorithm stops. Otherwise, a new cluster is taken from the queue and analyzed for further separation. For example, if *cluster* has been separated into *cluster1* and *cluster2*, the algorithm is repeated by taking *cluster1* from the queue. If, this time, the Fiedler parameters are above the threshold, *cluster1* cannot be separated and it is recorded as a single group. In the queue, however, there is still *cluster2* to be examined. If *cluster2* is separable, two new clusters are created, named *cluster21* and *cluster22*, which, in turn, are placed in the queue to check whether they are further separable or not.

The algorithm is applied until the queue is empty, i.e., all the frequency measures have been suitably separated into coherent clusters that cannot be further partitioned.

It is worth noting that the proposed method can be applied on every transmission networks, whatever the specific nature of the power sources, either traditional of renewable. Differently from the other solutions already available in the literature, the method, in fact, is capable of carrying out the real-time clustering by only leveraging on frequency measurements of the PMU without the need of assumptions about the power flows and load characteristic.

IV. METHOD ASSESSMENT WITH SYNTHESIZED SIGNALS

In order to preliminary assess the proposed method and to select the proper threshold values, several tests have been carried out on synthesized data, representing the frequency measures provided by 12 PMUs. The tests have been designed considering the characteristics of the real measurement data in different operating conditions in terms of noise and oscillations parameters [14], [15], [16]. The sampling rate was set equal to 10 S/s, in agreement with the typical data rate adopted by PMUs of the European network.

Simulations have been carried out in Mathworks MATLAB® environment by considering the PMU as belonging to three different groups (coherent areas) each of which characterized by a reference oscillation frequency. For each PMU of the same group, so, a sinusoidal oscillation of given frequency and amplitude was synthesized. Frequency values close to those characterizing the modes exhibited by the European transmission network were chosen [17], [18]. More specifically, for the PMUs referred to as 1 to 5 in the first group, the reference frequency is set equal to 0.11 Hz; for the second group, consisting of PMUs referred to as 6 to 8, the frequency is 0.21 Hz; and PMUs referred to as 9 to 12, constituting the third group, the reference frequency is 0.33 Hz. To make the measurement sets more compliant with actual measures, the oscillation frequency values of the PMU belonging to the same group were slightly varied. In particular, the frequency of each PMU within a specific group was drawn from a random variable characterized by a Gaussian probability density function (pdf), characterized by mean value equal to the reference frequency of the group and standard deviation equal to 1% of the mean.

$$f_{PMU_i} = \mathcal{N}\left(f_j, 0.01 \cdot f_j\right) \quad i = 1, .., 12 \quad and \quad j = 1, .., 3 \tag{8}$$

A variability in values was introduced also for the amplitude. The amplitude reference value was set equal to 0.001 Hz, while the instantaneous amplitude value was randomly selected from a Gaussian pdf centered on the reference value and having standard deviation equal to its 2%.

$$A_{PMU_i} = \mathcal{N}\left(1mHz, 0.01 \cdot 1mHz\right) \quad i = 1, .., 12 \tag{9}$$

In order to emulate the typical evolution versus time of the actual data received from PMUs in the transmission network, all measured set have been affected by colored noise, that has been obtained by performing the Wiener process [19], i.e. the integral of a zero-averaged additive white Gaussian noise (AWGN), whose standard deviation value was specifically tailored according to the desired Signal-to-Noise Ratio (SNR). It has to be highlighted that, for each PMU, the seed of the pseudo-random sequence for extracting values from the pdfs is set according to the system time; this ensures that uncorrelated noisy signals are generated.

The tests have been carried out with different duration T_w of the time window, equal respectively to 60 s, 120 s, 180 s, 240 s, and 300 s. As regards the SNR, values equal to -10 dB, 0 dB, 10 dB, and 20 dB have been considered.

Fig. 2 shows, as an example, the set of data synthesized setting a 120 s time window and an SNR equal to 0 dB. The clusters detected by the algorithm are shown in Figure 3. The initial cluster is first separated in *cluster2*, including the three PMUs oscillating at 0.21 Hz and *cluster1*, including the remaining PMUs. Successively, *cluster1* is further separated in *cluster11* and *cluster12*, containing, respectively, the 4 PMUs at 0.33 Hz and the 5 PMUs at 0.11 Hz.

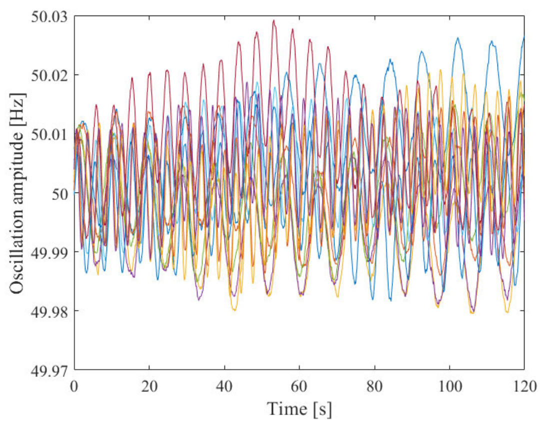

FIGURE 2. Synthesized signal, with Tw=120 s and SNR=0 dB.

The values of λ_2 experienced in the various tested conditions for all the examined clusters have been plotted in

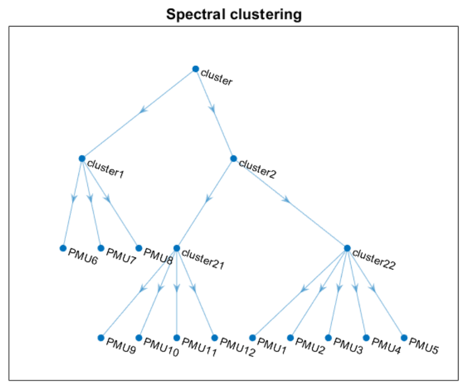

FIGURE 3. **Clusters detected by the developed algorithm.**

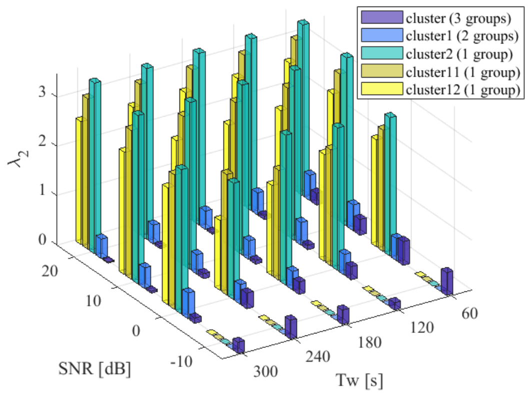

FIGURE 4. **Fiedler values observed in different test conditions.**

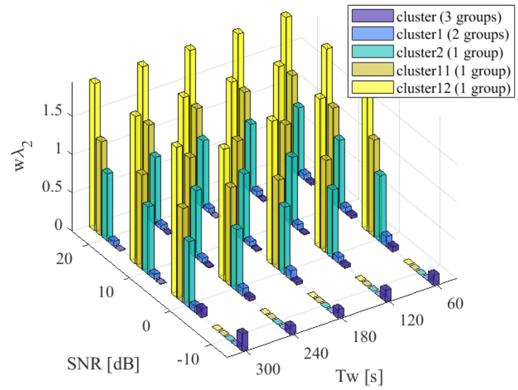

FIGURE 5. **Weighted Fiedler values observed in different test conditions.**

Fig. 4. It is worth noting that Fiedler values associated with clusters that have to be separated because they contain two or more different groups of PMUs (*cluster* and *cluster1*), have a value between 0 and 0.5; the Fiedler values associated with clusters that have not to be separated exhibit significantly higher amplitude, all exceeding 1.2. If the noise is such that the SNR does not subceed 0 dB, the larger the time window, the lower the λ_2 of the clusters to be separated; high T_w values, thus, make the detection of dissimilar clusters more straightforward. As far as the effect of noise is concerned, this generally increases the Fiedler values, tending to mask the similarity between the signals. When the SNR equal to -10 dB, only the Fiedler value of the initial cluster was reported, since the separation fails, grouping the signals incorrectly.

It can be noticed that the values of λ_2 for the initial cluster including all the PMUs are all close to each other for all the test conditions. There is no way, by observing the λ_2 values, to know in advance that the separation will fail and the resulting groups will be wrong.

The same conclusions can be drawn by observing the weighted Fiedler values, shown in Fig. 5.

The graph shows that the use of this parameter adds reliability to the algorithm, because the distance between the values of λ_2 of the clusters to be separated and those of the clusters not to be separated is even higher. But even in this case, if the SNR is -10 dB, the $w\lambda_2$ values of clusters and cluster 1 continue to be very low even though the separation fails.

In order to make the algorithm correctly separate clusters even in the presence of higher noise level, a filtering process is required; this implies an inevitable increase in processing time and the use of longer time windows, to remove the transient effect due to the numerical filter.

A numerical bandpass filter, suitably tuned on the frequency band of interest, was therefore implemented, whose order is 320 samples, with lower and upper cut-off frequencies equal to 0.05 Hz and 0.6 Hz respectively, suitable tailored for the considered frequency band of interest. The synthesized data was first processed with this filter and then sent as input to the implemented clustering algorithm. The number of samples must be at least 960, as 320 samples must be removed at the begin and the end of the data frame. Thus, at a sampling rate of 10 Hz, the window duration must be at least 96 s, so the condition Tw=60 s was no longer considered.

The results for both λ_2 and $w\lambda_2$ are shown in Fig. 6 and Fig. 7 respectively. Thanks to the filter, the algorithm provides reliable results even if the SNR reaches -10 dB. Even for less severe noise conditions, however, the filtering process helps to better discriminate separable clusters from those that are not.

This test phase also made it possible to determine the optimal parameters of the algorithm. In particular, the optimal duration of the time window is 180 s; this duration combines the performance of the algorithm with the need to reduce the algorithm's response time. Moreover, as highlighted in Fig. 6 and Fig. 7, the values $\lambda_2 = 0.8$ and $w\lambda_2 = 0.1$ can be chosen as optimal thresholds.

V. TESTS WITH MODIFIED KUNDUR MODEL

In order to assess the method also with typical signal observable on transmission network, tests have been carried out on a modified version of the Kundur model. In particular, authors have realized a four-area eight-machine model,

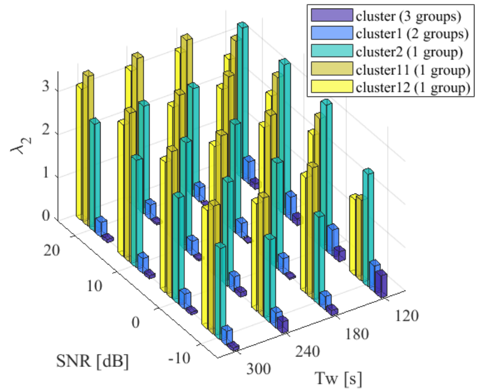

FIGURE 6. Fiedler values with the filtering process.

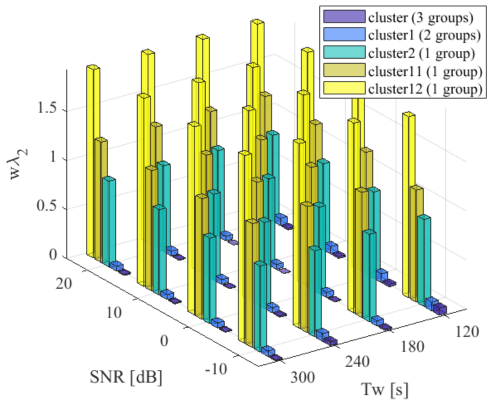

FIGURE 7. Weighted Fiedler values with the filtering process.

by connecting two traditional Kundur networks through of a transmission line, whose length has been set equal to 100 km. The loads in the Kundur network are modeled as constant power loads, which, as noted in [34], are the most severe. However, the model chosen for the load does not affect the performance of the clustering method which is PMU-based and it is influenced only by measurement data. The model of the test network is shown in Fig. 8.

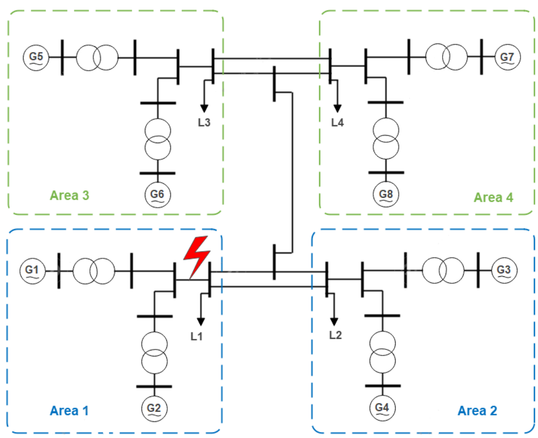

FIGURE 8. Modified Kundur model for the clustering assessment.

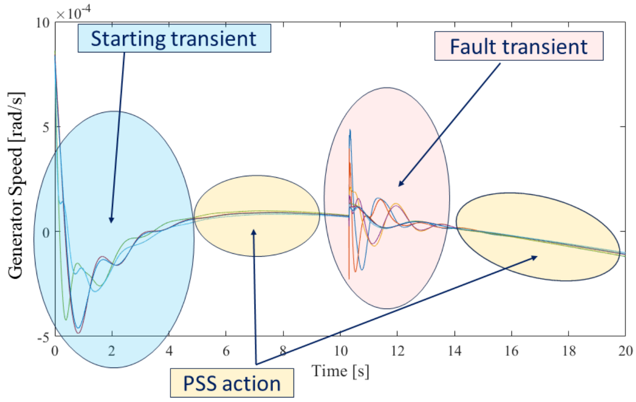

FIGURE 9. Time evolution of the generators speed of the simulated model.

The speeds of the 8 generators are used as signals to be processed for testing the clustering method. The model is run with a simulation step of 50 μs and, to reduce the simulation time, a window of 20 s was considered. The Kundur network is very useful for the assessment of measurement methods on transmission networks, as it represents a realistic model that takes all network components into account [35]. The disadvantage is that the oscillatory signals consist of intra-area oscillations (frequency between 1 and 3 Hz); moreover, due to the presence of Power System Stabilizers (PSSs) these oscillations are quickly damped. Therefore, the proposed method has to be assessed with shorter time windows, whose duration is about few seconds. However, the need of proportionally scaling the duration of the time window and the filter bandwidth does not affect the generality of the proposed method and the observed results.

In the model, a three-phase ground fault in Area 1, lasting 0.1 s, was simulated in the Area 1 at time t=10.3 s, in order to observe the transient oscillatory modes on the network following an abrupt change in load conditions. The observed generator speeds are shown in Fig. 9. As shown in the plot, several zones can be distinguished, relating to the different phases of the simulated network. At the beginning of the simulation, the initial transient can be observed, during which the network is brought to an equilibrium condition. The action of the PSS at each generator dampens the speed variations, bringing the machines to rotate at the synchronous speed.

The clustering algorithm, performed on this portion of the signal, correctly returns the clusters shown in Fig. 10, corresponding to the signals plotted in Fig. 11. The first cluster includes the speeds of the generators in Area 1 and Area 3 (G1, G2, G5, G6); the second cluster includes generators of Area 2 and Area 4 (G3, G4, G7, G8). Note that, given the symmetry of the simulated system, the speeds of Area 1 are perfectly coinciding with the speeds of Area 3, as the speeds of Area 2 exhibit the same time evolution of the speeds of Area 4. In fact, at this stage, the pair consisting

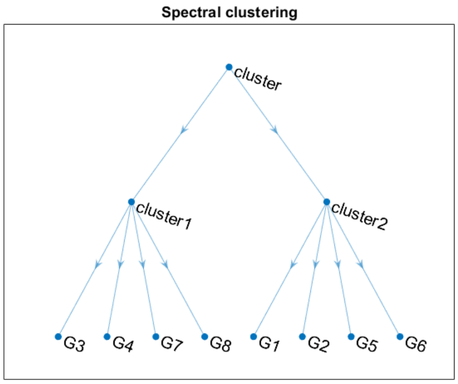

FIGURE 10. Clusters detected during the starting transient.

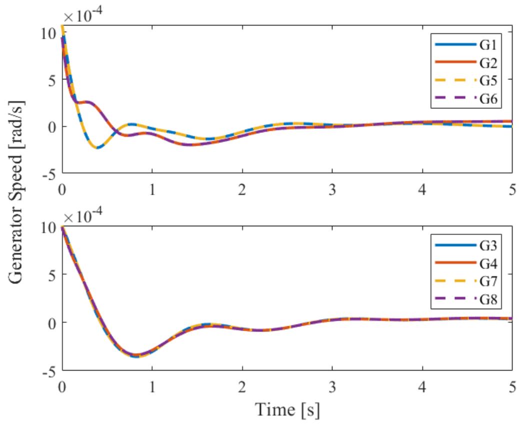

FIGURE 11. Speed signals of the clusters recognized during starting transient.

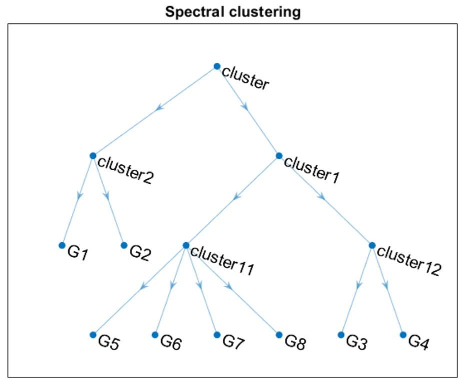

FIGURE 12. Clusters detected during the fault transient.

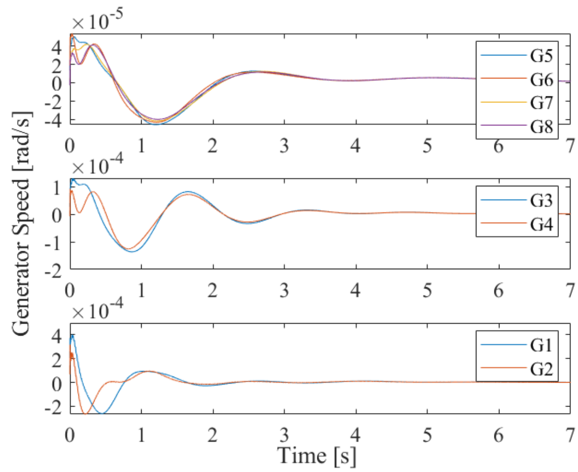

FIGURE 13. Speed signals of the clusters recognized during fault transient.

of Area 1 and Area 2 and the pair consisting of Area 3 and Area 4 are two decoupled systems. Therefore, Area 1 and Area 3 have the same behavior, as do Area 2 and Area 4, as if the connection bus at the center of the network did not exist.

The scenario changes during the fault. In this phase, in fact, the network must find a new synchronous condition following the abrupt unbalance between generated power and power absorbed by the loads; therefore, a transient is triggered where, this time, all the generators in the network are involved. Since there is no longer a symmetry condition, due to the fault location in Area 1, all areas oscillate differently from each other. Again, the proposed algorithm behaves in the expected manner, as it returns the clusters shown in Fig. 12 and Fig. 13. The output of the algorithm confirms that, during the fault transient, Area 1 and Area 2 oscillate against each other; Area 3 and Area 4, which are less affected by the fault, constitute a unique area whose velocities oscillate coherently.

Finally, the algorithm was also tested in the steady state zones, in which the generators are all synchronous; here too the method exhibited excellent reliability, recognizing that the 8 generators belong to a single cluster. It is worth noting that Kundur model has been exploited as test in other recent proposals, as an example [22]; the proposed method, however, outperforms those solutions thanks to the higher robustness brought by the adopted choice of the performance index and the optimal selection of the algorithm configuration parameters, thus improving the results of the traditional spectral clustering.

VI. ASSESSMENT WITH DATA FROM REAL PMUS

Finally, further tests were performed on real data acquired from PMUs of the European monitoring network, provided by Terna. Specifically, the time evolution of the measured frequency provided by 18 PMUs are given as input to the proposed clustering method. For security issues, neither the dates on which the measurements were taken, nor the exact geographical location of the PMUs can be given; the PMUs are referred to as with the generic name $PMU_i, i = 1, \ldots, 18$.

Fig. 14 shows the data frame, consisting of a 180 s time window containing the frequency measurements of all the PMUs. It can easily be seen that due to the noise, it is difficult to perceive the oscillations that characterize the frequency measures.

After the pre-processing phase described in III-C, the dataset is processed by the algorithm. This identifies

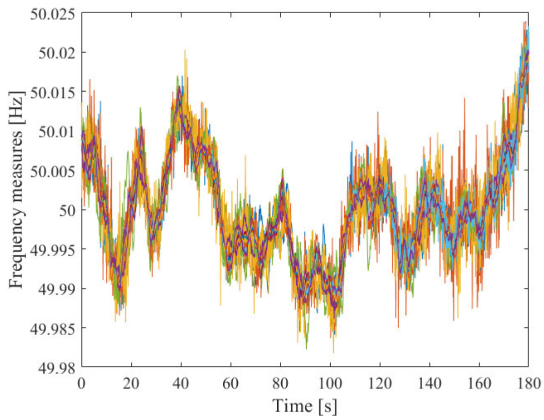

FIGURE 14. Frequency measures acquired by real PMUs for the method assessment.

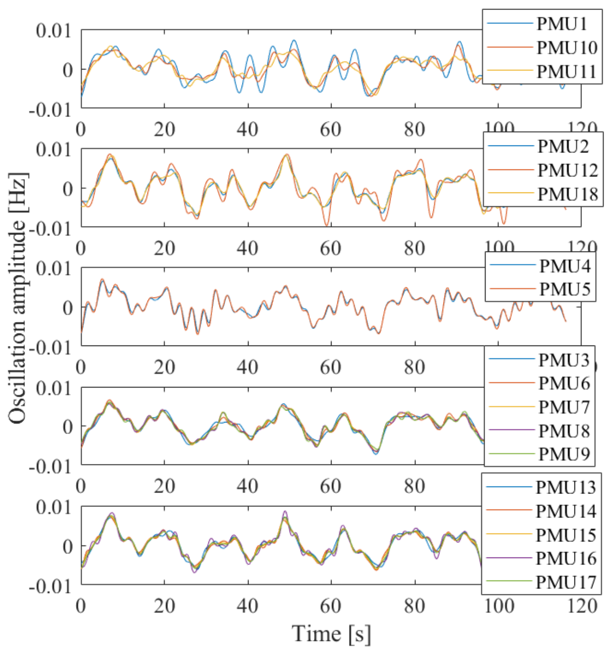

FIGURE 15. Frequency oscillations of the clusters detected by the algorithm.

5 clusters, i.e. 5 coherent areas. The signals of the obtained clusters are shown in Fig. 15, from which the strong correlation between the signals belonging to the same cluster is clearly visible.

From the graphs, it can be seen that the PMUs observe different oscillatory modes, characterized by different frequency, amplitude and phase. Some PMUs even observe the composition of two oscillatory modes of different frequencies. This characteristic generally makes it more difficult to recognize coherent areas. The proposed method is able to effectively identify clusters of PMUs, recognizing similar oscillatory modes, and to separate the PMUs to highlight coherent areas oscillating together.

As a proof of the correctness of the observed output, the geographical area of the PMUs whose measurements were

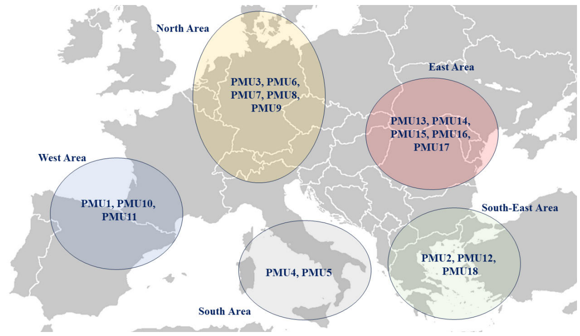

FIGURE 16. Geographic locations of the clustered PMUs.

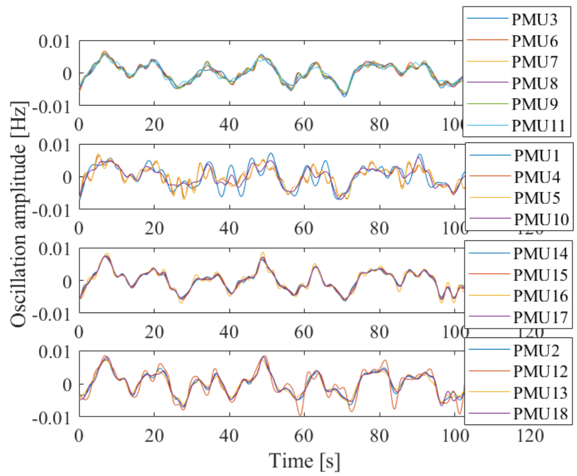

FIGURE 17. Frequency oscillations of the clusters detected by the k-means algorithm.

taken into account has been highlighted in Fig. 16. Due to data security issues, the exact location of the PMUs cannot be shown; indeed, the clusters identified by the proposed method is able to group PMUs that are located in areas that are geographically close to each other. As it can be expected, PMUs located at neighboring states in the network observe similar oscillatory phenomena.

Many tests were repeated, at different signal amplitudes and changing the set of PMUs processed. In all cases, the method proved to effectively group coherent areas. To better highlight the performance of the proposed method, the same dataset was processed using the k-means algorithm (an approach similar to that presented in [37]), which is widely adopted for data clustering [38], [39], [40]. This approach assumes that the number of clusters is given as input and, therefore, it is already known [41]. For this reason, an iterative bi-partitioning algorithm, similar to the one implemented in III-C was developed. The signals of the identified clusters are shown in Fig. 17 The main difference can be observed in cluster2, where both PMUs from South Area and West Area were included. The k-means approach does not recognize the differences between the oscillations that have similar amplitudes and trends, but different frequencies. The same problem is observed for PMU13, which was included in the South-East Area group, although it is characterized by an additional oscillatory mode with a different frequency than is not present in the signals provided by the other PMUs of the cluster.

VII. CONCLUSION

The paper dealt with the problem of suitably grouping the frequency signals measured by PMUs involved in the stability assessment of transmission networks according to their coherent operation areas. To this aim, an innovative method based on the iterated application of spectral clustering and optimized cluster partitioning has been defined and implemented; in particular, the method exploits a modified Bray Curtis index to define the entries of the similarity matrix. Starting from that matrix, after some calculations described above, the eigenvalues and eigenvectors of the non-normalized Laplacian allow to evaluate two parameters, referred to as the Fiedler and weighted Fiedler value. Comparing those values with two thresholds suitably defined, it is possible to decide if the considered cluster of frequency measures should be partitioned or not. By iterating this process on all the clusters that are successively generated, it is possible to separate all the frequency measures belonging to coherent areas, i.e., characterized by correlated evolution versus time.

Method performance has been assessed in a number of tests conducted on both simulated and actual measurement data. In particular, numerical simulations conducted on different conditions of SNR and observation interval have allowed to tune the values of the Fiedler and weighted Fiedler thresholds (equally respectively to 0.8 and 0.1) as well as defining and assessing both a proper filtering stage (consisting of a numerical band-pass filter suitably tuned with the frequency interval of interest) and observation interval (greater than 96s at a sample rate of 10Hz) in order to improve the robustness of the method. Tests conducted on a modified Kundur model have highlighted the capability of the proposed method of correctly grouping the measured data in all the operating condition of the network (starting transient, PSS action and fault transient). In all test conditions, the method has been capable of correctly separate coherent areas with a success probability equal to 100%. Finally, the method has been applied on actual frequency measurements, acquired by 18 PMUs spread throughout the Europe. The obtained results are of remarkable interest, since the coherent area separation perfectly matches the geographical distribution of the PMUs, as it can be expected. Moreover, the proposed method has clearly outperformed a typical solution presented in the literature, being capable of recognizing and separating oscillations characterized by similar amplitude but different frequencies. Ongoing activities are mainly focused on the exploitation of the proposed metohd for (*i*) the estimation of the area inertia and (*ii*) the early identification of possible critical operating conditions of the network.

REFERENCES

[1] G. R. Moraes, V. Ilea, A. Berizzi, C. Pisani, G. Giannuzzi, and R. Zaottini, "A perturbation-based methodology to estimate the equivalent inertia of an area monitored by PMUs," *Energies*, vol. 14, no. 24, p. 8477, Dec. 2021.

[2] F. Allella, E. Chiodo, G. M. Giannuzzi, D. Lauria, and F. Mottola, "On-line estimation assessment of power systems inertia with high penetration of renewable generation," *IEEE Access*, vol. 8, pp. 62689–62697, 2020.

[3] E. M. Carlini, F. Del Pizzo, G. M. Giannuzzi, D. Lauria, F. Mottola, and C. Pisani, "Online analysis and prediction of the inertia in power systems with renewable power generation based on a minimum variance harmonic finite impulse response filter," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107042.

[4] F. Bonavolonta, L. P. D. Noia, A. Liccardo, S. Tessitore, and D. Lauria, "A PSO-MMA method for the parameters estimation of interarea oscillations in electrical grids," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 11, pp. 8853–8865, Nov. 2020.

[5] A. Liccardo, S. Tessitore, F. Bonavolonta, S. Cristiano, L. P. D. Noia, G. M. Giannuzzi, and C. Pisani, "Detection and analysis of inter-area oscillations through a dynamic-order DMD approach," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–14, 2022.

[6] G. Giannuzzi, D. Lauria, C. Pisani, and D. Villacci, "Real-time tracking of electromechanical oscillations in ENTSO-e continental European synchronous area," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 1147–1158, Jan. 2015.

[7] I. Kamwa, A. K. Pradhan, G. Joos, and S. R. Samantaray, "Fuzzy partitioning of a real power system for dynamic vulnerability assessment," *IEEE Trans. Power Syst.*, vol. 24, no. 3, pp. 1356–1365, Aug. 2009.

[8] L. Ding, Z. Ma, P. Wall, and V. Terzija, "Graph spectra based controlled islanding for low inertia power systems," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 302–309, Feb. 2017.

[9] C. W. Wu, "Algebraic connectivity of directed graphs," *Linear Multilinear Algebra*, vol. 53, no. 3, pp. 203–223, Jun. 2005.

[10] M. A. Rios and O. Gómez, "Identification of coherent groups and PMU placement for inter-area monitoring based on graph theory," in *Proc. IEEE PES Conf. Innov. Smart Grid Technol. Latin Amer. (ISGT LA)*, Medellin, Colombia, Oct. 2011, pp. 1–7.

[11] E. Barocio, P. Korba, W. Sattinger, and F. R. S. Sevilla, "Online coherency identification and stability condition for large interconnected power systems using an unsupervised data mining technique," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 15, pp. 3323–3333, Aug. 2019.

[12] G. Giannuzzi, C. Pisani, and W. Sattinger, "Generator coherency analysis in ENTSO-E continental system: Current status and ongoing developments in Italian and Swiss case," *IFAC-PapersOnLine*, vol. 49, no. 27, pp. 400–406, 2016.

[13] P. M. Ashton, G. A. Taylor, A. M. Carter, M. E. Bradley, and W. Hung, "Application of phasor measurement units to estimate power system inertial frequency response," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2013, pp. 1–5.

[14] D. Li, N. Dong, Y. Yao, B. Xu, and D. W. Gao, "Area inertia estimation of power system containing wind power considering dispersion of frequency response based on measured area frequency," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 22, pp. 4640–4651, Nov. 2022.

[15] G. Frigo, A. Derviskadic, Y. Zuo, and M. Paolone, "PMU-based ROCOF measurements: Uncertainty limits and metrological significance in power system applications," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 3810–3822, Oct. 2019.

[16] K. K. Anaparthi, B. Chaudhuri, N. F. Thornhill, and B. C. Pal, "Coherency identification in power systems through principal component analysis," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1658–1660, Aug. 2005.

[17] M. A. M. Ariff and B. C. Pal, "Coherency identification in interconnected power system—An independent component analysis approach," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1747–1755, May 2013.

[18] N. Senroy, "Generator coherency using the Hilbert–Huang transform," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1701–1708, Nov. 2008.

[19] S. Avdakovic, E. Becirovic, A. Nuhanovic, and M. Kusljugic, "Generator coherency using the wavelet phase difference approach," *IEEE Trans. Power Syst.*, vol. 29, no. 1, pp. 271–278, Jan. 2014.

[20] M. H. Rezaeian, S. Esmaeili, and R. Fadaeinedjad, "Generator coherency and network partitioning for dynamic equivalencing using subtractive clustering algorithm," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3085–3095, Dec. 2018.

[21] J. Raitoharju, S. Kiranyaz, and M. Gabbouj, "Training radial basis function neural networks for classification via class-specific clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2458–2471, Dec. 2016.

[22] L. Lugnani, M. R. A. Paternina, D. Dotta, J. H. Chow, and Y. Liu, "Power system coherency detection from wide-area measurements by typicality-based data analysis," *IEEE Trans. Power Syst.*, vol. 37, no. 1, pp. 388–401, Jan. 2022.

[23] R. Rocchetta, "Enhancing the resilience of critical infrastructures: Statistical analysis of power grid spectral clustering and post-contingency vulnerability metrics," *Renew. Sustain. Energy Rev.*, vol. 159, May 2022, Art. no. 112185.

[24] S. F. Mahdavizadeh, M. R. Aghamohammadi, and S. Ranjbar, "Frequency stability-based controlled islanding scheme based on clustering algorithm and electrical distance using real-time dynamic criteria from WAMS data," *Sustain. Energy, Grids Netw.*, vol. 30, Jun. 2022, Art. no. 100560.

[25] A. A. Badr, A. Safari, and S. N. Ravadanegh, "Segmentation of interconnected power systems considering microgrids and the uncertainty of renewable energy sources," *IET Gener., Transmiss. Distrib.*, vol. 17, no. 17, pp. 3814–3827, Sep. 2023.

[26] W. J. Farmer and A. J. Rix, "Evaluating power system network inertia using spectral clustering to define local area stability," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107404.

[27] S. Ranjbar, "Adaptive criteria of estimating power system separation times based on inter-area signal," *IET Gener., Transmiss. Distrib.*, vol. 17, no. 3, pp. 573–588, Feb. 2023.

[28] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[29] D. Lauria and C. Pisani, "Real time generator coherency evaluation via Hilbert transform and signals morphological similarity," in *Proc. Int. Symp. Power Electron., Electr. Drives, Autom. Motion*, Jun. 2014, pp. 78–83.

[30] M. G. Michie, "Use of the bray-curtis similarity measure in cluster analysis of foraminiferal data," *J. Int. Assoc. Math. Geol.*, vol. 14, no. 6, pp. 661–667, Dec. 1982.

[31] M. B. Blaschko and C. H. Lampert, "Correlational spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[32] K. Fujiwara, M. Kano, and S. Hasebe, "Correlation-based spectral clustering for flexible process monitoring," *J. Process Control*, vol. 21, no. 10, pp. 1438–1448, Dec. 2011.

[33] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 14, 2001.

[34] I. D. Pasiopoulou, E. O. Kontis, T. A. Papadopoulos, and G. K. Papagiannis, "Effect of load modeling on power system stability studies," *Electr. Power Syst. Res.*, vol. 207, Jun. 2022, Art. no. 107846.

[35] P. S. Kundur and O. P. Malik, *Power System Stability and Control*. McGraw-Hill Education, 2022.

[36] M. Yuan and Q. Zhu, "Spectral clustering algorithm based on fast search of natural neighbors," *IEEE Access*, vol. 8, pp. 67277–67288, 2020.

[37] F. Raak, Y. Susuki, and T. Hikihara, "Data-driven partitioning of power networks via Koopman mode analysis," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2799–2808, Jul. 2016.

[38] J. Shan, Y. Zhang, Q. Zhao, and J. Lin, "A K-means clustering and triangulation-based scheme for accurate detection of multiple adjacent through-the-wall human targets," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[39] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[40] A. Aligholian, A. Shahsavari, E. M. Stewart, E. Cortez, and H. Mohsenian-Rad, "Unsupervised event detection, clustering, and use case exposition in micro-PMU measurements," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3624–3636, Jul. 2021.

[41] K. P. Sinaga, I. Hussain, and M.-S. Yang, "Entropy K-means clustering with feature reduction under unknown number of clusters," *IEEE Access*, vol. 9, pp. 67736–67751, 2021.

ANNALISA LICCARDO (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Naples Federico II, Naples, Italy, in 2003 and 2006, respectively. Since 2016, she has been an Associate Professor in measurement with the Department of Electrical Engineering and Information Technology, University of Naples Federico II. She has founded the spin off ARCADIA, for the realization of AR environments for remote control of measurement instruments. Her main current research interests include advanced measurements for monitoring and protection of electrical power systems, the IoT sensors for electrical measurements, distributed measurement systems, and AR-based remote laboratory.

DAVIDE LAURIA received the M.Sc. degree (Hons.) in electrotechnical engineering and the M.Sc. degree (Hons.) in mathematics from the University of Naples, Naples, Italy, in 1987 and 1995, respectively. He is currently a Full Professor in electrical power systems with the Department of Industrial Engineering, University of Naples Federico II, Naples. He is the coauthor of more than 200 scientific articles. He is also the coauthor of the chapter Probabilistic Transient Stability Assessment and On-Line Bayes Estimation of the book *Innovations in Power Systems Reliability* (Springer, 2011). His main research interests include power system analysis, stability, control, electric transportation systems, electrical system reliability, and application of power electronics to electrical power systems.

FRANCESCO BONAVOLONTÃ received the master's and Ph.D. degrees in electrical engineering from the University of Naples Federico II, Naples, Italy, in 2011 and 2015, respectively. He is currently a Research Fellow with the Department of Electrical and Information Technologies, University of Naples Federico II. He has founded the spin off ARCADIA, for the realization of AR environments for remote control of measurement instruments. His research interests include the area of instrumentation and measurement and can be divided into three main areas: remote control of measurement instruments, measurement methods based on compressive sampling, distributed measurement systems for monitoring and protecting electrical networks, and development of innovative measurement sensors based on artificial intelligence algorithms. He is a member of the Technical Commitee TC-37-Measurements and Networking of IEEE.

GIORGIO MARIA GIANNUZZI received the master's degree in electrical engineering from the University of Rome, Rome, Italy, in 1996. Until December 2000, he was with ABB, Zürich, Switzerland, where he was in charge of network studies, protection, and control applications. Since 2001, he has been with Terna, Rome, as an Expert in defense plans/systems, dynamic studies, protection, telecontrol, and substation automation. From 2004 to 2011, he coordinated the study, design, and activation of wide area defense system (including interruptible customers system) and wide area monitoring systems. Until 2009, he was a member of the UCTE Expert Group on Power System Stability. In 2010, he joined ENTSO-E, Bruxells, Belgium, System Protection and Dynamics Group, where he has been the Convenor, since 2014, coordinating the European evaluation over dispersed generation impact on system security and load shedding guidelines. He is currently responsible of the Engineering Department, National Dispatching Centre.

COSIMO PISANI received the Ph.D. degree in electrical engineering from the University of Naples Federico II, Naples, Italy, in 2014. During the Ph.D. degree in collaboration with Terna, Rome, Italy, he investigated some dynamic stability issues of large interconnected power system, such as the European one. He is currently the Head of Stability and Network Logics with the Dispatching and Switching Department of Terna SpA. He is also the Leader of Wide Area Monitoring System (WAMS) Task Force within ENTSO-E System Protection and Dynamic and Italian representative of CIGRE Study Committee C4 System Technical Performance. He is the author or coauthor of more than 80 scientific articles in IEEE/CIGRE community. His research interests include applications of dynamic stability of power systems, wide area monitoring and protection systems, high-voltage direct current systems, and power system restoration.

SALVATORE TESSITORE received the M.Sc. degree in electrical engineering and the Ph.D. degree from the University of Naples Federico II, Naples, Italy, in 2019 and 2022, respectively. His Ph.D. thesis titled "Detection and Measurement of Inter-Area Oscillations for Power System Stability." From 2019 to 2022, he attended the Ph.D. School of Information Technology and Electrical Engineering, currently carrying out his research activity in collaboration with Terna Rete Italia, Rome, Italy. His research interests include measurements and signal processing for electrical transmission networks.

Open Access funding provided by 'Università degli Studi di Napoli "Federico II"' within the CRUI CARE Agreement