POLITECNICO DI TORINO
Repository ISTITUZIONALE

Model-based Identification of Alternative Bidding Zone Configurations from Clustering Algorithms Applied on Locational Marginal Prices

(Article begins on next page)

10 November 2024

Model-based identification of alternative Bidding Zone Configurations from Clustering Algorithms applied on Locational Marginal Prices

Pietro Colella, Andrea Mazza, Ettore Bompard,
Gianfranco Chicco, Angela Russo
ENSIEL - Politecnico di Torino
Dipartimento Energia "Galileo Ferraris"
Torino, Italy
{pietro.colella,andrea.mazza,ettore.bompard,gianfranco.chicco,
angela.russo}@polito.it

Enrico Maria Carlini, Mauro Caprabianca,
Federico Quaglia, Luca Luzi, Giuseppina Nuzzo
Terna Rete Italia Spa
Dispacciamento e Conduzione
Roma, Italy
federico.quaglia@terna.it

Abstract—**This paper deals with the application of clustering methods to assist the bidding zone review processes in Italy, considering the Locational Marginal Prices (*LMP*s) as the relevant features. A novel approach based on the definition of the input data for clustering, depending on a number of scenarios defined by the Transmission System Operator, is exploited. The problem under analysis requires additional procedures to solve the challenging issue of incorporating node connection constraints in the clustering algorithm. A dedicated procedure, based on the definition of specific functions, is then applied to develop customised versions of k-means and hierarchical clustering. The customised procedures implemented can identify both wide clusters and outliers, whose location depends on the assessed scenarios.**

Keywords—*Clustering, bidding zones, locational marginal prices, scenarios.*

I. INTRODUCTION

A simplified representation of the transmission grid is applied in the framework of the European electricity markets, which are based on a zonal approach. In particular, electrical nodes are aggregated into Bidding Zones (BZs): energy trades are freely to occur inside each BZ, while cross-zonal exchanges are limited to a maximum amount (set by the cross-border capacities made available to the market). Currently, the large majority of the European BZs correspond to a single Member State: only Italy and Nordic countries (Sweden, Norway and Denmark) are split into several internal BZs.

The BZ configuration has a relevant impact on the efficiency of the electricity markets and on power system security (and/or the amount of remedial actions to be applied by Transmission System Operators - TSOs) [1]. For this reason, the European Guideline on Congestion Management and Capacity Allocation [2] introduced a monitoring and a review process and, in the framework of the recent Clean Energy Package, the Regulation (EU) 2019/943 [3] started a European BZ review (expected to be completed during 2021).

In the framework of the review processes, identifying alternative BZ configurations to be compared with the current one is a crucial step. For this scope, several Stakeholders expressed their interest in the application of model-based algorithms, which could support TSOs providing quantitative analysis on the expected scenarios. In Italy, the National Regulatory Authority asked to the TSO to work on this field in order to develop a proper methodology to be adopted in the future when the internal Italian BZs will be reviewed again (a BZ review has been

already completed in 2018) [4-6]. Fig. 1 shows the present structure of the BZs in Italy, as well as some interconnections with neighbouring countries.
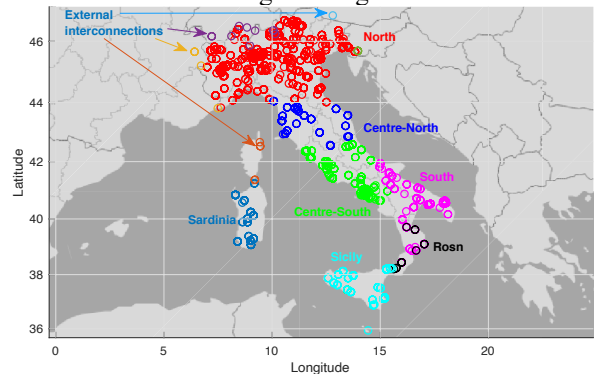

Fig. 1. Present bidding zones in Italy and external interconnections.

In previous studies (recalled in Section III), clustering algorithms have been used to identify the possible BZs on the basis of the Locational Marginal Prices (*LMP*s) calculated from an optimal power flow applied either on a limited set of very realistic scenarios or on a large set (8760) of simplified scenarios (where load and RES nodal distribution has been determined based on typical profiles). In both cases, a grid topology where all elements are available has been considered. In these studies, the input data were taken from the *LMP*s determined at the network nodes for the all the simulated cases/hours of the year. These data were then reduced by removing all the hours in which all the *LMP*s were equal. In fact, the presence of equal *LMP*s would have no contribution to the components of the distance between the *LMP*s in all pairs of nodes at the corresponding hour. However, this approach has a conceptual drawback: if the number of congested hours during the year is very low, the partitioning of the BZs would be based only on what happens on a few hours, thus making the general validity of the result questionable. It is then crucial to rethink the way to form the features used in the clustering procedures, in such a way to make them representative and avoid massive filtering of the input data.

To bridge these research gaps, this paper introduces a new approach where a probability of occurrence is assigned to each scenario on which *LMP*s are computed. For each of these scenarios, variations are elaborated to consider that one (or more) grid elements could be out of service for planned outages (or long-lasting fault). This

approach requires setting up the calculation of the *LMP*s for the grid status-based scenarios in an appropriate way [7] and customising the clustering algorithms to make them able to deal with different scenarios. The importance of the scenarios is established by using suitably defined weighting factors (reflecting the occurrence probability).

II. GENERATION OF THE SCENARIOS

In order to derive reliable *LMP*s that could represent a large range of system working conditions, a relevant set of scenarios in terms load and RES generation have been identified. In particular, 2018 actual scenarios experienced by the Italian power system have been clustered using an unsupervised machine learning approach (k-means algorithm), using zonal load, solar infeed, wind infeed, hydro infeed and import from other countries as relevant variables. In order to trade off accuracy and computation time, 20 clusters have been identified as suitable compromise for this study and the centroids of these clusters have been selected as base scenarios for the *LMP*s computation. Keeping the correspondent real-time snapshots of the power system, a realistic nodal distribution of load and distributed generation representing a wide range of system operating conditions has been obtained. Then, for each scenario 5 variations have been considered: a full network availability scenario and 4 planned maintenance cases identified for testing purposes by the TSO on the basis of detailed knowledge on the network operation. For each scenario, an Optimal Power Flow that explicitly incorporates the *N*-1 security criteria has been run to compute realistic LMPs for a significantly wide range of operating conditions (the details are indicated in [7]). Furthermore, the TSO assigned a weight to each scenario, to represent its relative importance (e.g., probability of occurrence estimated according to historical data). The scenario-based definition of the features to be used as inputs for the clustering procedure in different with respect to the *LMP* time series generally used, and is a specific contribution provided in this paper.

III. APPLICATION OF CLUSTERING METHODS TO THE FORMATION OF THE BIDDING ZONES

A. Background on clustering methods for BZ formation

An overview of the clustering methods is presented in [8], where it is indicated that the most used unsupervised approaches are the methods k-means and hierarchical clustering (HC), both based on distances, with penalty factors introduced to represent the distances between non-connected pairs of nodes.

Previous results in the application of the clustering methods to form the BZs (reached in the activity carried out by the authors) led to the following outcomes:

- The application of the basic versions of the clustering algorithms found for example in Matlab® are insufficient to form BZs where all nodes are connected, because the basic algorithms embed no topological check [9]. If a dedicated procedure to identify the connected zones is executed at the end of the basic clustering algorithms [10], the numbers of BZs generally increases.

- The implementation of customised versions of the clustering algorithms that contain an internal check of the interconnections during the execution of the clustering procedure has improved the results. In this case, high penalty factors have been associated to the pairs of non-interconnected nodes, in such a way to discourage merging the clusters when there is no interconnection among their nodes. In this way, for the clustering methods that require the number of clusters among the input data, the final number of clusters generally remains the same as the initial one.

- The solutions obtained highly depend on the clustering method adopted. Even some variants of the same method can lead to very different results. This has been shown in [9] for four variants of the customised agglomerative HC algorithm, based on Euclidean distance, in which the only difference has been the linkage criterion adopted. Taking two groups of nodes, the linkage criterion determines how to calculate the distance between the two groups [11]. Using classical linkage criteria such as single, average, and Ward [12], the results obtained with the same number of BZs were significantly different. The single linkage criterion created a big cluster and had the trend of isolating some outliers. The Ward linkage criterion again led to the creation of highly populated clusters and a few smaller clusters. The average linkage criterion created more uniform partitions. These results are in line with other applications of the linkage criteria [13].

B. Challenges for applying the clustering methods

The traditional clustering methods are not suitable to obtain the partitions of the BZs based on the *LMP*s, because they do not consider the connections among the nodes. The extension of the clustering algorithms to incorporate the node connection is not straightforward. The introduction of topological information concerning the network structures in the clustering algorithms fits with the distance-based calculation framework by introducing appropriate penalty factors applied to the pairs of non-connected nodes.

Another solution to improve clustering accuracy by adding knowledge on the problem domain has been suggested in [14], by considering *constrained clustering algorithms* based on graphs, in particular applied to the k-means algorithm [15]. Two types of constraints are introduced, namely, (i) *must-link* constraints (e.g., pairs of nodes that must belong to the same cluster based on previously known information), ad (ii) *cannot-link* constraints (e.g., pairs of nodes that cannot belong to the same cluster based on previously known information). In the modified version of k-means developed in [14], called *COP-kmeans*, each component (e.g., node in the case of our paper) is assigned to the closest cluster that does not violate the constraints, and the absence of available clusters results in clustering failure. On the one hand, the introduction of the constraints referring to the domain knowledge changes the basic rationale of the clustering algorithms, with consequences on the computation times and convergence characteristics that depend on the effect of the constraint. On the other hand, the introduction of the constraints clustering can make clustering

performance less sensitive to the algorithm chosen, especially when the constraints (e.g., topological) drive the solution in specific directions.

The application of must-link and cannot-link concepts has been used in some problems referring to electrical networks such as the formation of intentional islands [16], in which the relations between nodes belonging or not to an island is known a priori. However, these concepts cannot be directly incorporated in the algorithms for BZ clustering, because the previous information only concerns the network topology, while there is no pre-established decision referring to the location on pairs of nodes within the same cluster or in different clusters. In Chapter 7 of a recent book [17], the discussion that mentions the existence of constrained clustering algorithms indicates the possibility of accommodating contiguity constraints to ensure that the clusters form connected sub-graphs, adding "while we did not encounter such problems in practice". The problem under consideration in this paper is indeed requesting the presence of connected sub-graphs in each cluster, thus indicating how the studies in progress address more general challenging aspects in the clustering domain.

C. General aspects and notation

The input data available refer to the network structure, the selected scenarios and the corresponding *LMP*s. The general notation is indicated below.

a) Variables:
  D Number of initial *LMP* data for each node
  F Number of input features for clustering
  K Number of centroids
  N Number of nodes
  S Number of scenarios

b) Vectors and matrices:
- Node adjacency matrix $\mathbf{A} \in \mathbb{N}^{N,N}$
- Centroid matrix $\mathbf{C} \in \mathbb{N}^{K,F}$
- Initial data matrix $\mathbf{D} \in \mathbb{R}^{N,D}$
- Cluster adjacency matrix $\mathbf{M} \in \mathbb{N}^{K,K}$
- Input data matrix for clustering $\mathbf{X} \in \mathbb{R}^{N,F}$
- Cluster location vector $\mathbf{v} \in \mathbb{R}^{N,1}$
- Vector containing the weighting factors for the scenarios $\mathbf{w} \in \mathbb{R}^{S,1}$

D. Input data pre-processing

The adjacency matrix $\mathbf{A}$ contains the connection among all pairs of nodes, coded in binary form (0 = not connected; 1 = connected). The nodes are defined in such a way to represent all the possible interconnections of the busbars inside the power substation (i.e., multiple nodes are used to represent separate operation of the busbars in different scenarios). However, the matrix $\mathbf{A}$ does not change in all the calculations carried out for clustering. Possible situations in which one of the nodes (e.g., a busbar) is not connected to the rest of the system are handled by assigning to this node the same *LMP* of the node corresponding to the other busbar in the substation.

The presence of multiple scenarios is addressed by first forming a set of features that depends on the *LMP*s indicated by the TSO for each scenario. In this way, the data input vector is formed as a data matrix $\mathbf{D}$ in which the *LMPs* corresponding to the various scenarios are added to the columns, resulting in D columns. Then, each column

of the matrix $\mathbf{D}$ belonging to the same scenario is multiplied by the corresponding weighting factor included in the vector $\mathbf{w}$. Finally, the columns in which all *LMP* values are equal are eliminated, as they would not contribute to the calculation of the distances between nodes in the clustering procedure. In this way, the input data used for clustering is $\mathbf{X}$, with the number of columns equal to the number of features $F \leq D$.

E. Specific functions for customising the clustering algorithms

A number of functions have been set up to incorporate domain knowledge into the clustering algorithms. These functions perform the following calculations:
- F1 (*Remove empty clusters*): in some cases (e.g., k-means) in which the number of clusters is set up as one of the inputs, it may happen that the final number of clusters is lower than the desired one. This function takes the cluster location vector $\mathbf{v}$ as input, and provides as the output the new version of the vector $\mathbf{v}$ in which the empty clusters have been removed and the clusters have been renumbered from 1 to the maximum number of resulting clusters.
- F2 (*Form connected clusters*): checks the within-cluster connection between the nodes belonging to the same cluster. The input data are the cluster vector $\mathbf{v}$ and the adjacency matrix $\mathbf{A}$. For each cluster, if all the nodes in the cluster are connected nothing changes; otherwise, the initial cluster is split into connected clusters, with an increase in the total number of clusters. The output is the modified cluster location vector $\mathbf{v}$, together with the new number of clusters.
- F3 (*Cluster connections*): identifies the branches that connect the clusters. The input data are the cluster vector $\mathbf{v}$ and the adjacency matrix $\mathbf{A}$. The outputs are the list of branches that connect the various clusters, the numbers of the connected clusters, and the cluster adjacency matrix $\mathbf{M}$, whose entries are equal to unity if the clusters are connected, and zero elsewhere.
- F4 (*Cluster grouping*): reduces the number of clusters. The inputs are the cluster location vector $\mathbf{v}$, the adjacency matrix $\mathbf{A}$, the input data matrix $\mathbf{X}$, the centroid matrix $\mathbf{C}$, and the desired final number of clusters K. The clusters are grouped by successively merging the pair of clusters with the lower distance between the centroids. Each time the new centroids are recalculated starting from the entries of the input data matrix $\mathbf{X}$ that belong to the new cluster formed. The output is the updated cluster location vector $\mathbf{v}$.

F. Customisation of the hierarchical clustering algorithm

A given number of clusters is requested as input data. At the beginning, all nodes are assigned to a separate cluster. The classical HC is based on three basic steps, namely: (i) the calculation of the distances between the clusters (typically the Euclidean distance), (ii) the definition of the linkage criterion (with a number of variants indicated as single, complete, centroid, average, and Ward), and (iii) the bottom-up pairwise grouping of the clusters. These steps are executed until the requested number of clusters is reached.

Two HC variants have been implemented:

a) *Penalty-based HC*: the classical structure of the HC is maintained, in which the first step of calculation of the distances is modified by adding to the classical distances a high penalty factor to all combinations of nodes that are not directly connected. The consequence of this addition is mainly reflected on the application of the linkage criteria. In particular, the linkage criteria are affected by the presence of very high terms that bring high values to the entries. The single linkage criterion is the one less affected, because it considers only the nearest distances among the components of pairs of clusters.

b) *Constrained HC*: in this case, the specific knowledge on the domain is included by allowing only operations between connected areas. The centroid linkage is used as the relevant linkage criterion. The iterative process is set up as follows, starting from the initial data and from $K = N$ centroids:

> ➤ identify the connections between clusters (function F3), calculate the distances between centroids, use the cluster adjacency matrix to extract only the pairs of clusters connected, and determine the minimum distance between pairs of clusters;
> ➤ check and remove the empty clusters (function F1);
> ➤ recalculate the centroids;
> ➤ apply the stop criterion: at each iteration the number of clusters is reduced by one; the process stops when the number of clusters reaches the requested one, otherwise the next iteration is executed.

G. Customisation of the k-means algorithm

Two different types of customised versions of the k-means algorithm have been implemented. In both cases, the k-means++ algorithm[1] [18] is used to choose the initial centroids, to avoid total random initialisation.

The two versions are:

a) *Constrained k-means*: an iterative process is set up as follows, starting from the initial data and centroids:

> ➤ for each node, calculate the distances of the features from the centroids, and determine the location of the centroids for which there is the minimum distance (as in the classical k-means);
> ➤ check and remove the empty clusters (function F1);
> ➤ check the within-cluster connection and form connected clusters (function F2);
> ➤ identify the connections between clusters (function F3);
> ➤ reduce the number of clusters (function F4);
> ➤ recalculate the centroids;
> ➤ apply the stop criterion: stabilisation of the vector **v**, or maximum number of iterations; if the stop criterion is not satisfied, the next iteration is run.

b) *Post-processing of the classical k-means results*: after the execution of k-means, the functions F1, F2, F3 and F4 are applied once, then the final centroids are recalculated.

---

[1] The centroid of the first cluster is randomly chosen from the data. The other centroids are randomly chosen from the other data by using probability weights proportional to the distances from the closest cluster centroid.

IV. CASE STUDY APPLICATIONS AND RESULTS

The application of the customised and constrained clustering procedures with scenario-based *LMP* data to the Italian network is shown in this section. The network nodes considered include the continental part of Italy and the Sicily island, while the nodes at the external interconnections and inside the Sardinia island are not included (in this study, the Sardinia island is considered as a BZ by definition since it is also a different synchronous area). The network model has 918 nodes and 1317 branches. Five scenarios are considered, in which the weights depend on the historical occurrence of the planned outages (and of similar ones) [7]:

Scenario 1 (TSO weight 5%): planned outage of a 380 kV line on the Adriatic path. This outage limits in a significant way the South-North transmission capacity.

Scenario 2 (TSO weight 40%): planned outage of a 380 kV line in the Southern part of Italy. This outage limits the transmission capacity from the South to the Central part of Italy.

Scenario 3 (TSO weight 5%): planned outage of a 380 kV line in the Central part of Italy. This outage does not typically impact in a significant way the transmission capacity between the existing BZs.

Scenario 4 (TSO weight 20%): planned outage of a 380 kV line in the North-Western part of Italy. This outage does not typically impact in a significant way transmission capacity between existing BZs, but during this unavailability intra-zonal congestions could occur.

Scenario 5 (TSO weight 30%): all grid elements are considered fully available. The weight of this case is the difference between 100% and the sum of the weights of the other scenarios.

The initial set of data includes $D = 100$ initial data for each node, taken from the $S = 5$ scenarios. After the elimination of the data having equal *LMP* values for all the nodes, the remaining input data contain $F = 65$ features, with a total of 8003 different *LMP*s. Fig. 2 shows the *LMP*s corresponding to these features. High *LMP*s appear in a few nodes, depending on the grid status-scenarios defined.
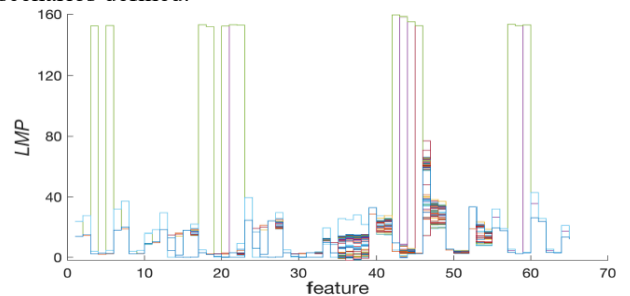

Fig. 2. *LMP*s corresponding to the features used for clustering.

C. Results from the hierarchical clustering variants

C.1. Penalty-based HC

This variant of the HC has been run by using the *average linkage* criterion and setting up a penalty factor equal to 10^{20} (i.e., five degrees of magnitude higher than the sum of the absolute differences between the *LMP*s). A warning has to be noticed in the use of the average linkage criterion. The average is calculated by considering all the distances between pairs of nodes in the group. Thereby,

this average depends on the number of nodes not connected among them, as each of these pairs of nodes provides a contribution to the distance affected by the penalty. The final result then reflects the number of pairs of non-connected nodes in the groups. Nonetheless, the results obtained have a relatively uniform distribution of the final clusters (Fig. 3). The corresponding cluster centroids are shown in Fig. 4, from which the averaging effect is clearly seen from the reduction of the maximum *LMP*s with respect to the initial values shown in Fig. 2.
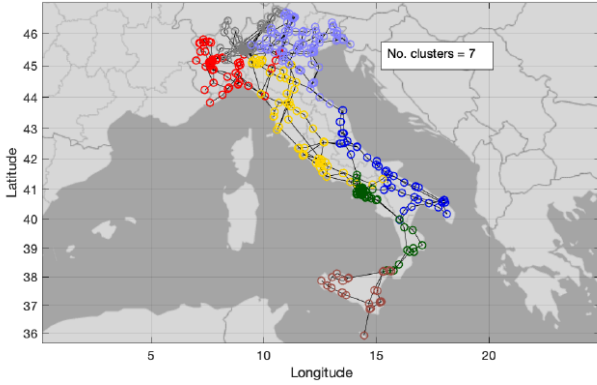

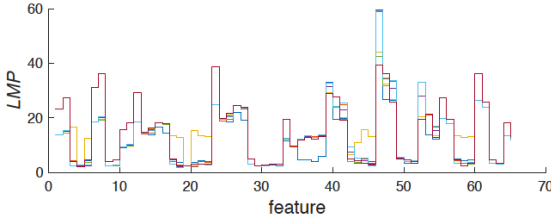Fig. 3. Results of the Penalty-based *average linkage* HC (7 clusters).

Fig. 4. Centroids of the 7 clusters obtained from the Penalty-based HC with *average linkage* criterion.

This warning does not apply to the *single linkage* criterion, as this criterion considers the minimum distance among the components of different groups (i.e., the nearest neighbours). However, in this case, at the beginning of the grouping the clusters are formed by merging the most similar *LMP* patterns; then, the further clusters are obtained by merging together the most similar *LMP* patterns in two different groups, regardless of the other *LMP* patterns belonging to these clusters. At the end, the clustering procedure stops and isolates the *LMP* patterns that resemble less to any of the *LMP* patterns belonging to the existing clusters. Hence, the final clusters are typically composed of a large group and a few outliers (Fig. 5). The corresponding cluster centroids are shown in Fig. 6 and indicate the prevailing role of the outliers to determine the clustering results.
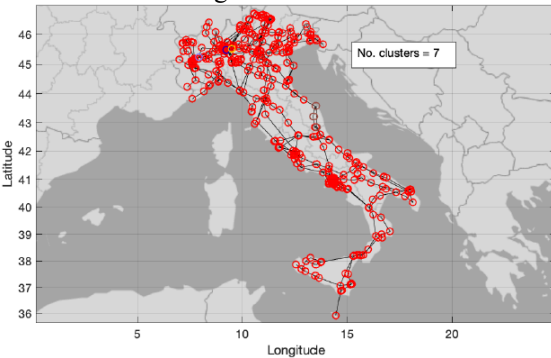

Fig. 5. Results of the Penalty-based *single linkage* HC (7 clusters).

C.2. Constrained HC

The Constrained HC adopts a centroid-based mechanism for the calculation of the distances, which differs from the classical centroid linkage because only the connected clusters are considered. When the numbers of imposed clusters reduces, conceptually the distances among the closest centroids should increase. From the zoom reported in Fig. 7 from 2 to 100 clusters, an effect of applying the Constrained HC algorithm is the non-monotonic variation of the *minimum* distance between the cluster centroids (the one relevant for merging the clusters; the mean and maximum distances are reported as well for the sake of completeness). In fact, even pairs of clusters with close distance based on the features, but not connected at a given point during the execution, cannot be merged in the same cluster. However, further grouping could make these clusters connected through another group, thus making the connection available for the progress of the algorithm.
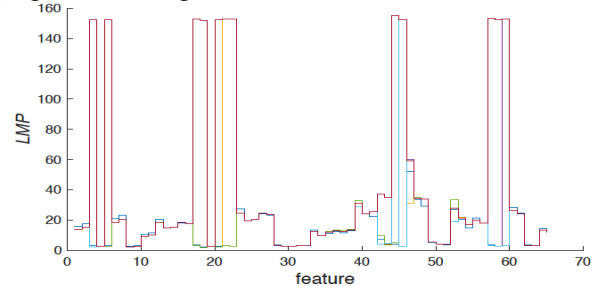

Fig. 6. Centroids of the 7 clusters obtained from the Penalty-based HC with *single linkage* criterion.
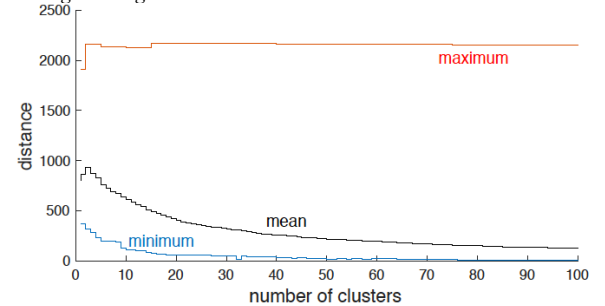

Fig. 7. Minimum, mean and maximum distance between the cluster centroids for various numbers of clusters in the Constrained HC method (zoom from 2 to 100 clusters).

The execution of the Constrained HC on the dataset used shows the ability of the clustering algorithm to identify a number of outliers (Fig. 8). This result and the related centroids (Fig. 9) are similar (but not equal) to outcomes of the Penalty-based HC with single linkage.
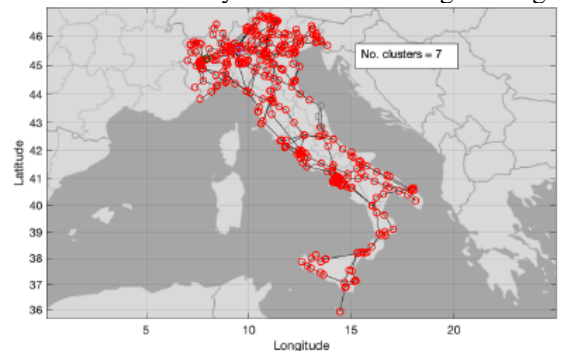

Fig. 8. Results of the Constrained HC (7 clusters).

C. Results from the k-means variants

The k-means variants adopt the same set of functions introduced for the HC variants. In this way, the typical properties of the k-means clustering (i.e., the ability of forming relatively uniform groups) are not yet found in general. The solution is mainly driven by the functions applied to introduce the constraints. The example shown in Fig. 11 for the Constrained k-means with 7 clusters indicates a solution with similar attitude to isolate the outliers to the ones found for the Constrained HC and the Penalty-based with single linkage criterion. Also the nature of the centroids (not shown here) is similar.
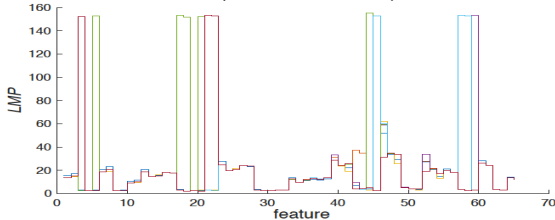

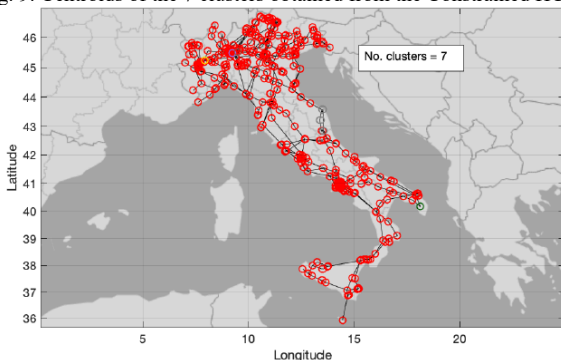Fig. 9. Centroids of the 7 clusters obtained from the Constrained HC.

Fig. 10. Results of the Constrained k-means (7 clusters).

V. Conclusions

This paper has extended the application of clustering algorithms for the purpose of grouping together the network nodes with similar evolution in time of the *LMP*s, considering a dataset based on weighted scenarios that represent different operating conditions of the Italian power system. The scenarios have been selected in order to well represent different load and distributed generation conditions experienced by the power system, and five network topological conditions are considered.

The nature of the problem was not directly tractable with conventional clustering algorithms, due to the needs of considering the physical links among the network nodes and of guaranteeing the connectivity of the nodes in the resulting clusters. For this purpose, customised versions of the k-means and hierarchical clustering algorithms have been developed and applied.

From the results obtained, it is possible to formulate some general and specific considerations. As general considerations, the results from the clustering algorithms are always data-driven, and there is no single clustering algorithm that provides the final solution. The outcomes from different clustering algorithms have to be interpreted and discussed. The specific results found in this paper indicate that the choice of the scenarios can drive the clustering solutions towards specific outcomes, suggesting that a sufficiently large set of cases have to be considered when applying this approach. In particular, for the clustering techniques more aimed to identifying the outliers, the individual scenarios in which there are more

discontinuities in the *LMPs* leave a footprint that has an impact on the formation of the zones. However, since these outliers are typically composed of a few nodes, these groups of nodes are too small to be considered as separate BZs. More likely, these nodes could be considered as separators among different BZs.

The clustering algorithms used in this paper depend on the final number of clusters provided as one of the inputs. Further stop criteria could be considering for clustering, for example based on a distance criterion, concluding the procedure when the clusters become too far apart. However, the tuning of the limit distance depends on the data and it is not simple to be addressed. Moreover, in the constrained versions of the procedures there could be a non-monotonic evolution of the distances used to form the clusters. Further criteria and analyses with more scenarios will have to be adopted to establish the final structure of the BZs. The results referring to these aspects will be presented in future contributions.

References

[1] A. Pototschnig, "The Importance of a Sound Bidding-Zone Review for the Efficient Functioning of the Internal Electricity Market", *Florence School of Regulation - Energy*, Issue 2020/22, May 2020.

[2] European Commission, Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management, web: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R1222&from=EN.

[3] Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the Internal Market for Electricity.

[4] ARERA, Decision 386/2018/R/EEL "*Disposizioni in merito alla suddivisione della rete rilevante in zone, in esito al processo di revisione svolto ai sensi del regolamento UE 2015/1222 (CACM)*" (in Italian), web: https://www.arera.it/allegati/docs/18/386-18.pdf.

[5] ARERA, Decision 103/2019/R/EEL "*Ulteriori disposizioni in merito alla suddivisione della rete rilevante in zone, in esito al processo di revisione svolto ai sensi del Regolamento (UE) 2015/1222 (CACM)*" (in Italian), web: https://www.arera.it/allegati/docs/19/103-19.pdf.

[6] ARERA, Decision 496/2017/R/EEL "*Disposizioni in merito alla revisione della suddivisione della rete rilevante in zone*" (in Italian), web: https://www.arera.it/allegati/docs/17/496-17.pdf.

[7] C. Bovo, V. Ilea, M. Caprabianca, F. Quaglia, L. Luzi, G. Nuzzo, "Optimal computation of Network indicators for Electricity Market Bidding Zones configuration", *55th International Universities Power Engineering Conference (UPEC 2020)*, Torino, Italy, 1-4 September 2020, under review.

[8] G. Chicco, P. Colella, A. Griffone, A. Russo, Y. Zhang, E.M. Carlini, M. Caprabianca, F. Quaglia, L. Luzi, and G. Nuzzo, "Overview of the clustering algorithms for the formation of the bidding zones", Proc. *54th UPEC,* Bucharest, Romania, 3-6 September 2019.

[9] L. Michi, F. Quaglia, E. Bompard, A. Griffone, C. Bovo, E.M. Carlini, L. Luzi, G. Chicco, A. Mazza, V. Ilea, M. Caprabianca, G. Nuzzo, P. Colella, and A. Russo, "Optimal Bidding Zone Configuration: Investigation on Model-based Algorithms and their Application to the Italian Power System", *Proc. AEIT Int. Annual Conference*, Florence, Italy, 18-20 Sept. 2019.

[10] A. Griffone, A. Mazza, and G. Chicco, "Applications of clustering techniques to the definition of the bidding zones", Proc. *54th UPEC*, Bucharest, Romania, 3-6 September 2019.

[11] M.R. Anderberg, *Cluster analysis for applications*, New York: Academic Press, 1973.

[12] J.H. Ward, "Hierarchical Grouping to Optimize an Objective Function", *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[13] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping", *Energy*, vol. 42, no. 1, pp. 68–80, June 2012.

[14] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge", *Proc. 18th Int. Conference on Machine Learning (ICML '01)*, pp. 577–584, 2001.

[15] J.B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *Proc. 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, vol. 1, pp. 281-297, 1967.

[16] A. Esmaeilian and H. Kezunovic, "Prevention of power grid blackouts using intentional islanding scheme", *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 622–629, 2017.

[17] F. Fouss, M. Saerens, and M. Shimbo, *Algorithms and Models for Network Data and Link Analysis*, Cambridge University Press, Cambridge, UK, 2016.

[18] D. Arthur and S. Vassilvitskii, "K-means++: The Advantages of Careful Seeding", SODA '07: *Proc. Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.