(Article begins on next page)

# Overview of the Clustering Algorithms for the Formation of the Bidding Zones

Gianfranco Chicco, Pietro Colella, Andrea Griffone, Angela Russo, Yang Zhang
Politecnico di Torino, Dipartimento Energia "Galileo Ferraris"
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
{gianfranco.chicco,pietro.colella,angela.russo,yang.zhang}@polito.it

Enrico Maria Carlini, Mauro Caprabianca, Federico Quaglia, Luca Luzi, Giuseppina Nuzzo
Terna Rete Italia SpA
Viale Egidio Galbani, 70 – 00156 Rome, Italy
federico.quaglia@terna.it

***Abstract*—This paper presents an overview of the concepts, features and methods used in model-based approaches for the determination of the bidding zones in multi-regional interconnected networks. The main aspects are discussed on the basis of a set of selected articles taken from the scientific literature. The solution schemes are mainly based on clustering algorithms. The main conclusions are that no prevailing solution emerges, and that further insights are needed, also with the incorporation of specific knowledge taken from the nature of the problem in the solution methods and indicators.**

***Index Terms*–bidding zones, clustering, locational marginal process, power transfer distribution factors, zonal prices.**

## I. Introduction

The current European electricity market design is based on a zonal approach, where cross-zonal trades and exchanges are limited according to available transfer capacities calculated by the Transmission System Operators (TSOs), while internal trades inside bidding zones are considered as unrestricted. A properly designed bidding zone configuration is hence a key enabler factor for achieving system security and market efficiency.

For this reason, the Commission Regulation (EU) 1222/2015 of 24th July 2015, also known as Guideline on Capacity Allocation and Congestion Management (CACM) [1], established a European framework for assessing the efficiency of the existing Bidding Zone Configuration and, when relevant, for reviewing this configuration.

When a Bidding Zone Review process is launched, involved TSOs are requested to deliver a proposal on the methodology to be applied in the assessment as well as a set of alternative configurations to be considered and compared to the current one.

For the latter task, two main methodologies (or a combination of them) have been identified up to now [2]:

- *Expert-based approach*: TSO's expert can define alternative configurations according to their knowledge as well as according to actual data and future expectations.
- *Model-based approach*: dedicated algorithms are developed in order to identify alternative configurations.

The first approach allows TSOs to deliver acceptable results in a timely manner and to include all their expert knowledge in an easy way. Relevant data and results included in the ENTSO-E triennial Bidding Zones technical report [3] and in the Ten Year Network Development Plan [4] (or in similar national reports) are considered in this method.

The second approach allows to identify (in a highly automated way) the most suitable configuration according to a predefined objective function or measure of quality. Less degrees of freedom are left to the expert, improving external acceptability (even if transparency for "non-technical" Stakeholders could be endangered, since some algorithms could be perceived as "black boxes").

Several model-based methods have been proposed in the literature: the goal of this paper is then to explore and discuss proposed solutions, trying to identify a subset of most promising solution to be considered for the Italian Power System case (according to the relevant ARERA guideline [5]).

The discussion on the data and solution methods is based on 31 articles selected from the scientific literature.

The next sections of this paper are organized as follows. Section II refer to the input data in terms of the features and time intervals used for the analysis. Section III recalls the metrics used in the studies. Section IV discusses the solution methods, with specific focus on the clustering algorithms. The last section contains the concluding remarks.

## II. Input Data

### A. Main input features

The main features used in the reviewed literature are the Locational Marginal Prices (*LMP*s) and the Power Transfer Distribution Factors (*PTDF*s) [6][7] of relevant CBCO[1]s (Fig. 1). The details are indicated in Table I. Other features used in a limited number of contributions include:

- Available Transmission Capacity (*ATC*)
- Principal Component Analysis (PCA), to perform data compression on the initial data (e.g., *LMP*s)
- Sensitivity of the nodal power injections (*SNPI*)
- Electrical parameters (EL_PAR)
- Transmission Congestion Distribution Factors (*TCDF*)
- Nodal price differences (*NPD*s), or nodal marginal costs
- Power injections responsible of the loop flows (*PILF*)

[1] A CBCO represents a combination of a Critical Branch (the element on which the flow is being monitored) and of a Critical Outage (the simulated contingency after which the flow on the CB has to be verified).

Electrical parameters, sensitivities, *ATC* and power injections are technical inputs used to construct solution schemes related to the network structure. *NPD*s are used in [11] together with network information. The *TCDF*s are defined by using the results of the full AC power flow instead of the DC power flow considered to assess the *PTDF*s. The *PTDF*s are generally limited to the most congested CBCOs, and in [21] are weighted by using congestion rate factors. Weighting factors are also used in [14] on the basis of the importance of the nodes (depending on load and generation) to obtain more similar sizes of the zones. A composition of electrical and geographical distances is used in [25] and [26].

In the *LMP* context, the PCA has been applied by rescaling the data to get the average hourly zonal prices equal to unity, and further transforming the data to operate with the correlation matrix [24]. In this way, the interdependencies among the data are considered [23]. However, with the PCA transformations the physical meaning of the entries is lost.

TABLE I. MAIN INPUT FEATURES

| *Reference* / *Input feature* | *LMP* | *PTDF* | PCA | *SNPI* | ATC | EL_PAR | *TCDF* | *NPD* | *PILF* |
|---|---|---|---|---|---|---|---|---|---|
| [8] | √ | | | | | | | | |
| [9] | √ | | | | | | | | |
| [10] | √ | | | | | | | | |
| [11] | | | | | | | | √ | |
| [12] | | | | | | √ | | | |
| [13] | √ | | | | | | | | |
| [14] | √ | | | | | | | | |
| [15] | √ | | | | | | | | |
| [16] | √ | | | | | | | | |
| [17] | √ | | | | | | | | |
| [18] | √ | | | | | | | | |
| [19] | √ | | | | | | | | |
| [20] | | √ | | | | | | | |
| [21] | | √ | | | | | | | |
| [22] | | | | | | | | | √ |
| [23] | √ | | √ | | | | | | |
| [24] | | | √ | | | | | | |
| [25] | | | | | | √ | | | |
| [26] | √ | | | | | | | | |
| [27] | | | | | | √ | | | |
| [28] | | | | | √ | | | | |
| [29] | | √ | | | | | | | |
| [30] | | √ | | | | | | | |
| [31] | | √ | | | | | | | |
| [32] | √ | | | | | | | | |
| [33] | | | | √ | | | | | |
| [34] | | | | √ | | | | | |
| [35] | | | | √ | | | | | |
| [36] | √ | | | | | | | | |
| [37] | | | | | | | √ | | |
| [38] | √ | | | | | | | | |

*B. Time intervals*

The determination of the bidding zones is based on the analysis of the power system operation during a given time horizon, carried out with a certain time step. In the literature references reviewed, the longest time horizon considered to represent the features is one year for nine articles, while other two articles conduct the analysis over one month. Furthermore, fifteen articles carry out *one-shot* analysis on a specific system solution. The typical time step (resolution) used for the analysis is one hour (15 minutes in one reference). Moreover, nine articles adopt a scenario approach.

With respect to the time intervals, *one-shot* calculations seem to be very limited, especially to study power systems with high penetration of renewable energy sources (RES), and in case the bidding zone review has a large geographical scope. When scenarios are considered, load profiles are typically used, but also information on RES infeeds has to be incorporated in an accurate way [11].

For large systems, the computational burden of carrying out a long-term analysis in which the details of the power system operation are considered may become very high. In practical applications, case-by-case identification of the best compromise between time horizon/resolution and computation time is needed.

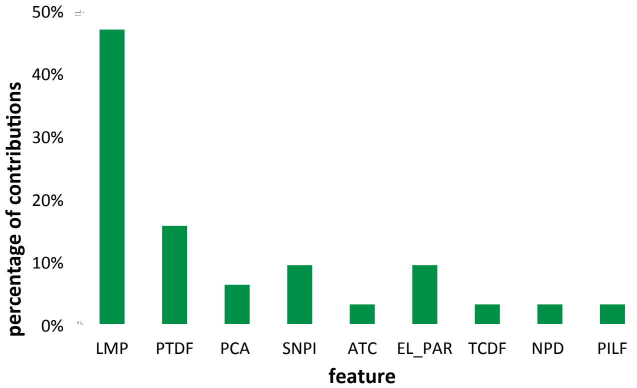


Fig. 1. Partitioning of the input features in the contributions.

## III. METRICS

The formulation of the solution methods generally requires the adoption of a metric or distance to compare the entries with each other. Fig. 2 shows the partitioning of the choices made in the reviewed literature. Table II indicates more detailed information. The Euclidean distance is the most used in the solution algorithms (even though it is not largely prevailing as in other clustering applications). Other metrics have been tested, such as boolean-based, electrical distance, fuzzy membership, log-likelihood, Manhattan distance, spectral distance, and the Ward criterion. The variety of metrics adopted indicates the plurality of attempts to create dedicated frameworks of analysis.

Specific metrics and indicators are used also for validation of the results. In this case, general-purpose metrics used for clustering validity assessment are typically adopted. It is of particular interest to introduce metrics that take into account the nature of the problem, i.e., technical and economic aspects of the bidding zone partitioning, such as the generation costs, redispatch costs, potential of market power, and violation of network security considered in [9].

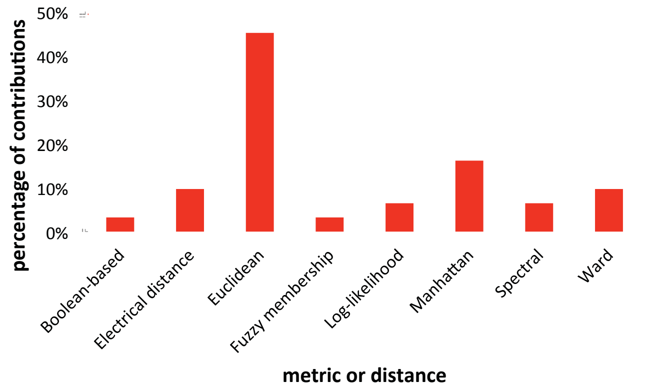


Fig. 2. Partitioning of the different distances/metrics in the contributions.

TABLE II. METRICS USED

| *Reference* / *Metric* | Boolean-based | Electrical distance | Euclidean | Fuzzy membership | Log-likelihood | Manhattan | Spectral distance | Ward's criterion |
|---|---|---|---|---|---|---|---|---|
| [8] | | | | | | √ | | |
| [9] | | | | | | √ | | |
| [10] | | | | | | | | √ |
| [11] | | | | | | | √ | |
| [12] | | √ | | | | | | |
| [13] | | | √ | | | | | |
| [14] | | | √ | | | | | |
| [15] | | | √ | | √ | | | |
| [16] | | | √ | | √ | | | |
| [17] | | | √ | | | | | |
| [18] | | | √ | | | | | |
| [19] | | | √ | | | | | |
| [20] | | √ | | | | | | |
| [21] | | | √ | | | | | |
| [22] | | | √ | | | | | |
| [23] | | | √ | | | | | |
| [24] | | | √ | | | | | |
| [25] | | √ | | | | | | |
| [26] | | | √ | | | | | |
| [27] | | | | | | | √ | |
| [28] | | | √ | | | | | |
| [29] | | | | | | | | √ |
| [30] | √ | | | | | | | |
| [31] | | | | | | √ | | |
| [32] | | | | | | | | √ |
| [33] | | | | | | √ | | |
| [34] | | | | | | √ | | |
| [35] | | | | √ | | | | |
| [36] | | | √ | | | | | |
| [37] | | | √ | | | | | |
| [38] | | | √ | | | | | |

For the clustering methods, regardless of the feature used as input, the clustering validity assessment has to be carried out starting by using the same features for all the methods to be compared. For example, the analysis on the same system can be conducted with *LMP*s or *PTDF*s as features, obtaining in both cases the vectors containing the allocation of each node to a cluster. For the comparisons, the clustering validity indices have to be computed in both cases by using the same feature (either the *LMP* or the *PTDF*) to calculate the relevant distances.

## IV. SOLUTION METHODS

The most used solution methods are unsupervised clustering approaches (Fig. 3), especially with the classical methods k-means and hierarchical clustering (HC). Other solution schemes have also been proposed (Table III), such as the BubbleClust algorithm, Consensus clustering, Fuzzy c-means (FCM), Genetic algorithm, Geographical clustering, Hierarchical clustering, k-medoids, Price differential clustering (PDC), Spectral clustering, Tabu search embedded algorithm (TSEA), a Two-step algorithm, ATC-based changing partitions, as well as a Network partition scheme (NPS) not based on clustering.

The main aspects concerning the methods used refer to:

- *Network topology*: the nodes belonging to the same price zone must be electrically connected. However, in general the original clustering methods do not incorporate topology-based constraints. The physical connection of the zones is guaranteed in the geographical clustering [18] and in the NPS [20]. Information on the topology (i.e., incidence and adjacency matrices) can be exploited in the spectral clustering. In HC it is possible to analyse the final dendrogram that indicates the sequential aggregation of the entries and the corresponding value of the variable used in the linkage criterion. The introduction of network topology constraints requires either a post-processing of the clustering results, or the modification of the standard clustering algorithms. The latter is easy to be implemented in HC to condition the merging of a pair of clusters [14]. In [25] the medoids are CBCOs determined in such a way that their terminals do not belong to the same zone.
- *Number of clusters*: in most methods it has to be defined *a priori*. In NPS it is defined in an initial pre-optimisation phase. In PDC and TSEA it depends on a user-defined input threshold on the difference between average *LMP*s. In the geographical clustering the number of clusters depends on the CBCOs with higher probability of congestion.
- *Initialisation of the clusters or zones*: this is a crucial point that may heavily impact on the effectiveness of the solution process. Random initialisations are rather ineffective. In HC all the nodes are the initial clusters. With k-means some improvements in the centroid definition are possible by maximising the geographic distance [31].

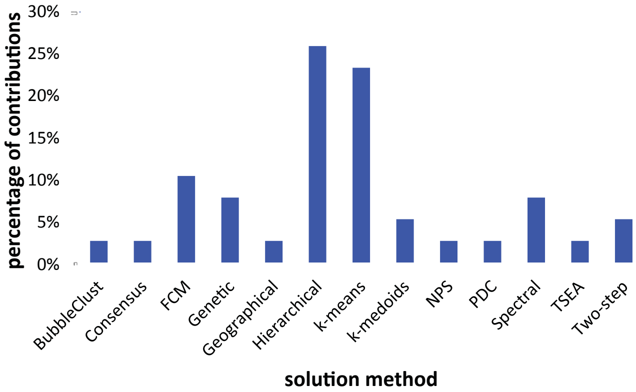

Fig. 3. Partitioning of the solution methods in the reviewed contributions.

TABLE III. SOLUTION METHODS

| *Reference* / *Method* | BubbleClust | Consensus | FCM | Genetic | Geographical | Hierarchical | k-means | k-medoids | NPS | PDC | Spectral | TSEA | Two-step | ATC-based |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [8] | | | | √ | | | | | | | | | | |
| [9] | | | | √ | | | | | | | | | | |
| [10] | | | | | | √ | | | | | | | | |
| [11] | | | | | | | | | | | √ | | | |
| [12] | | | | √ | | | √ | | | | | | | |
| [13] | | | | | | √ | | | | | | | | |
| [14] | | | | | | √ | | | | | | | | |
| [15] | | | | | | | √ | | | | | | √ | |
| [16][2] | | | | | | | √ | | | | | | √ | |
| [17] | | | √ | | | | | | | | | | | |
| [18] | | | √ | | √ | | | | | √ | | | | |
| [19] | | | √ | | | | | | | | | | | |
| [20] | | | | | | | | | √ | | | | | |
| [21] | √ | | | | | | | | | | | | | |
| [22] | | | | | | √ | | | | | | | | |
| [23] | | | | | | | √ | | | | | | | |
| [24] | | | | | | | √ | | | | | | | |
| [25] | | | | | | | | √ | | | | | | |
| [26] | | | | | | √ | √ | √ | | | | | | |
| [27] | | | | | | | | | | | √ | | | |
| [28] | | | | | | | √ | | | | √ | | | |
| [29] | | | | | | √ | | | | | | | | |
| [30] | | | | | | √ | | | | | | | | |
| [31] | | | | | | | √ | | | | | | | |
| [32] | | √ | | | | √ | | | | | | | | |
| [33] | | | | | | √ | | | | | | | | |
| [34] | | | | | | √ | | | | | | | | |
| [35] | | | √ | | | | | | | | | | | |
| [36] | | | | | | | √ | | | | | | | |
| [37] | | | | | | | | | | | | | | √ |
| [38] | | | | | | | | | | | | √ | | |

[2] The objective is the characterisation of the zonal prices to support transmission system planning and generation investments, without determining the partitioning into network zones.

Possible issues in the results include:

- The creation of small clusters, not acceptable as bidding zones because of high market power. This may happen in PDC and TSEA for small values of the threshold [18], and in general for all methods if there is no explicit check during the solution process.
- The possible instability of the borders of the bidding areas during time. This may be reduced with multi-scenario analyses, with the application of a consensus algorithm [32], or with the use of sensitivities of the nodal power injections with respect to the power flows in the congested CBCOs [33].

The methods indicated below have been able to show solutions to the formation of bidding zones. However, all methods have their drawbacks. The most common one is that the physical connection of the nodes belonging to the same cluster has to be verified a posteriori (with the exception of handling the network topology as indicated before). Some other drawbacks are summarised as follows:

- BubbleClust: approximations of the load and generation levels appear in the construction of the generation shift key matrix used to pass from zonal injections to power flows, and the same matrix is undefined for self-sufficient zones.
- Consensus: the computational burden increases with respect to a single clustering method, and the creation of the bidding zones needs a careful comparison among the clusters formed by the clustering method considered.
- FCM: dependence of the result on the choice of the initial centroids and on the parameter (fuzzification level).
- Genetic algorithm: need for suitable initialisation (initial clusters with connected nodes), difficult implementation of the crossover operator to avoid losing the node connection, while an efficient implementation of the mutation operator allows swapping the cluster only for nodes located at the border of two clusters. The use of a penalty factor to make it difficult to merge non-connected nodes during the solution process slows down the process and cannot guarantee the final connection of the nodes in all zones.
- Geographical: only the case with cross-border congestion has been studied [18]. High dispersion of the *LMP*s in the same cluster may be found, bringing ineffective economic signals.
- Hierarchical: computation times higher than k-means and k-medoids, possible scalability limits due to the use of the pair-to-pair comparison [26], and congested lines not always located at the borders of the zones [39].
- k-means: dependence of the result on the choice of the initial centroids, difficult implementation of the node connection requirement.
- k-medoids: dependence of the result on the choice of the initial medoids, difficult implementation of the node connection requirement.
- NPS: the nodes without generators are not considered, so the method is not suitable to be used for a market with generation offers and demand bids, and the zone definition does not take into account the evolutions in time.

- PDC: in the application to time series of data, an efficient determination of the threshold is needed.
- Spectral: challenging choice of the threshold used to form the graph, and need for the application of a clustering algorithm (e.g., k-means) applied to the most significant eigenvectors of the incidence matrix.
- TSEA: sensitivity of the results to the threshold, and low stability of the zones in time [38].
- Two-step: the formation of too many subclusters in the first step may slow down the second step [15][16].

## V. Concluding Remarks

This paper has presented an overview of the solution methods considered in the literature for the partitioning of the interconnected power system into distinct bidding zones.

From the analysis of the literature, no dominant solution algorithm emerges. The solution techniques mainly include clustering algorithms adapted to the specific problem, and other customised algorithms that exploit the knowledge of the network topology, constraints and operating conditions. In general, the network-based information is particularly important, so as to make it impractical to use any clustering algorithm in its classical version. In fact:

1. The classical clustering algorithms do not include mechanisms to verify the physical connection among the network nodes, with the exception of the PDC, in which the node connection is explicitly checked before merging the nodes that satisfy the price difference criterion. If a classical clustering algorithm is used, the physical connection of the nodes can be verified a posteriori, with the drawback that the number of clusters may change with respect to the predefined number (if any). If the computational code of the clustering algorithm is available, the node connection check can be included in a customised version of the solver [40]. In this case, the incorporation of this check may be easier for agglomerative clustering algorithms (such as HC) than for centroid-based algorithms such as k-means. Moreover, the use of penalty factors for pairs of nodes physically not connected, aimed at avoiding the merging of these nodes during the clustering process, does not appear generally satisfactory to guarantee the creation of separate zones, making it necessary to carry out a final check.
2. There is no clear-cut criterion for the definition of the input value for the number of clusters (when it is directly requested) or for the threshold to be used when the number of clusters is not included among the input data. The definition of the (optimal) number of clusters is a known issue for the clustering applications, and there is no unique solution. For a large number of nodes, the available techniques (e.g., depending on the parametric analysis of specific indicators when the number of cluster changes) could lead to numbers of clusters too high with respect to the number of zones reasonably acceptable for the system. Regulatory authorities and TSOs could provide indicative ranges of the number of zones. In this way, it would be possible to drive the solution methods to the definition of an acceptable number of clusters, bearing in mind that the number of clusters is not necessarily equal to the number of bidding zones (that can be confirmed after a post-processing check, to take into account specific situations such as refinements due to the merging of excessively small clusters into larger zones, or to reshaping the zones to take into account the transnational borders).
3. Some algorithms require the initialisation of the entries in the clusters or of the centroids. In these cases, a random initialisation of the entries is inappropriate, as the nodes included in the same cluster could be located anywhere, without respecting the connection of the zones. Better solutions could maintain the zone connection in the initial conditions. This aspect needs to be further investigated.

Further studies are needed to provide insights in different directions, namely:

- The check of the convenience to adopt LMPs, PTDFs of the most congested CBCOs, or other choices, as the input data for the solution procedures.
- The use of information on the network structure and operation coming from extended modelling of the power system, taking into account the approximations in the power system models that may be useful without limiting the effectiveness of the results [41][42], and the issues depending on network topology and in particular to topology changes over time.
- The extension of the solution methods applied to other algorithms available in the literature, provided that these algorithms are able to easily incorporate the characteristics of the problem under analysis, mainly to keep the connection of the nodes belonging to the same zone, as well as the possibility of applying effective initialisations. Further research is needed to formulate and implement methods that embed both the computational effectiveness of viable clustering algorithms and the knowledge of the power system domain to define appropriate features and constraints.
- The identification of the most suitable method also depends on desired properties of the final bidding zone configuration (e.g., the consideration of the physical borders between countries, the criterion to define robust price zones, and the incorporation of system-based security constraints).
- The effectiveness of the methods depends on the data structure, and on the definition of suitable indicators that represent also problem-related aspects. The definition and testing of dedicated metrics and indicators that incorporate the nature of the problem in a more specific way (e.g., the minimum size of the zones, also depending on the consideration of market power aspects [40]).
- The assessment of the possible benefits of alternative metrics (e.g., the Minkowski distances as a generalisation of the Manhattan and Euclidean distances, or the Mahalanobis distance used in an extended framework that takes into account also the correlations between random time series of the relevant variables, or others [43]).

## REFERENCES


[1] European Commission, Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management, Official Journal of the European Union, 25 July 2015, L 24-72.

[2] ENTSO-E (European Network of Transmission System Operators for Electricity), First edition of the bidding zone review - Final report, March 2018, online: https://docstore.entsoe.eu/Documents/News/bz-review/2018-03_First_Edition_of_the_Bidding_Zone_Review.pdf, accessed 14.05.2019.

[3] ENTSO-E, Bidding Zone Configuration Technical Report 2018, online: https://docstore.entsoe.eu/Documents/Events/2018/BZ_report/20181015_BZ_TR_FINAL.pdf, accessed 14.05.2019.

[4] ENTSO-E, Ten-Year Network Development Plan (TYNDP) 2018 – Network Development Plan to 2025, 2030 and 2040, online: http://tyndp.entsoe.eu/maps-data/, accessed 14.05.2019.

[5] Italian Authority for Electricity, Gas and the Hydro System, Disposizioni in merito alla revisione della suddivisione della rete rilevante in zone, Decision 496/2017/R/EEL (in Italian), 28 June 2017, online: www.arera.it/allegati/docs/17/496-17.pdf, accessed 14.05.2019.

[6] A.J. Wood and B.F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. New York: Wiley-Interscience, 1984.

[7] J-B. Bart and M. Andreewsky, Network Modelling for Congestion Management: Zonal Representation Versus Nodal Representation, *Proc. 15th PSCC*, Liège, Belgium, 2005.

[8] C. Breuer, N. Seeger, and A. Moser, Determination of alternative bidding areas based on a full nodal pricing approach, *Proc. IEEE PES General Meeting*, Vancouver, BC, 2013.

[9] C. Breuer and A. Moser, Optimized bidding area delimitations and their impact on electricity markets and congestion management, *Proc. 11th Int. Conf. European Energy Market (EEM14)*, Krakow, Poland, 2014.

[10] B. Burstedde, From nodal to zonal pricing: A bottom-up approach to the second-best, *2012 9th International Conference on the European Energy Market*, Florence, Italy, 2012.

[11] K.-K. Cao, J. Metzdorf, and S. Birbalta, Incorporating Power Transmission Bottlenecks into Aggregated Energy System Models, *Sustainability*, vol. 10, no. 6, 2018.

[12] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, S. Blumsack, and M. Patel, Multi-Attribute Partitioning of Power Networks Based on Electrical Distance, *IEEE Trans. Power Systems*, vol. 28, no. 4, pp. 4979-4987, 2013.

[13] T. Felling and C. Weber, Identifying price zones using nodal prices and supply & demand weighted nodes, *Proc. IEEE ENERGYCON*, Leuven, Belgium, 2016.

[14] T. Felling and C. Weber, Consistent and robust delimitation of price zones under uncertainty with an application to Central Western Europe, *Energy Economics*, vol. 75, pp. 583-601, 2018.

[15] J. Ferreira, S. Ramos, Z. Vale, and J. Soares, Zonal prices analysis supported by a data mining based methodology, *Proc. IEEE PES General Meeting*, Providence, RI, 25-29 July 2010.

[16] J. Ferreira, S. Ramos, Z. Vale, and J. Soares, A Data-Mining-Based Methodology for Transmission Expansion Planning, *IEEE Intelligent Systems*, vol. 26, no. 2, 2011.

[17] Y.-Y. Hong and C.-Y. Hsiao, Locational marginal price forecasting in deregulated electricity markets using artificial intelligence, *IEE Proc – Gener., Transm. and Distrib.*, vol. 149, no. 5, pp. 621-626, 2002.

[18] M. Imran, and J. W. Bialek, Effectiveness of zonal congestion management in the European electricity market, *Proc. IEEE 2nd Int. Power and Energy Conference*, Johor Bahru, Malaysia, 2008, pp. 7-12.

[19] S.-H. Jang, J.-H. Kim, S.-H. Lee, and J.-H. Park, Zone Clustering LMP with Location information using an Improved Fuzzy C-Mean, *Proc. Int. Conf. on Intelligent Systems Applications to Power Systems*, Niigata, Japan, 5-8 November 2007.

[20] C.Q. Kang, Q.X. Chen, W.M. Lin, Y.R. Hong, Q. Xia, Z.X. Chen, Y. Wu, and J.B. Xin, Zonal marginal pricing approach based on sequential network partition and congestion contribution identification, *Int. Journal of Electr. Power & Energy Syst.*, vol. 51, pp. 321-328, 2013.

[21] M. Kłos, K. Wawrzyniak, M. Jakubek, and G. Oryńczak, The scheme of a novel methodology for zonal division based on power transfer distribution factors, *Proc. 40th Annual Conference of the IEEE Industrial Electronics Society*, Dallas, TX, 2014, pp. 3598-3604.

[22] M. Klos, K. Wawrzyniak, and M. Jakubek, Decomposition of power flow used for optimizing zonal configurations of energy market, *2015 12th Int. Conf. European Energy Market*, Lisbon, Portugal, 2015.

[23] D. Kiran, A.R. Abhyankar, and B.K. Panigrahi, Zonal price based clustering of bidding zones, *IEEE 6th Int. Conf. on Power Systems (ICPS)*, New Delhi, India, 4-6 March 2016.

[24] M. Koivisto, P. Heine, I. Mellin, and M. Lehtonen, Clustering of Connection Points and Load Modeling in Distribution Systems, *IEEE Trans. on Power Systems*, vol. 28, no. 2, pp. 1255-1265, 2013.

[25] S. Lumbreras, A. Ramos, L. Olmos, F. Echavarren, F. Banez-Chicharro, M. Rivier, P. Panciatici, J. Maeght, and C. Pache, Network partition based on critical branches for large-scale transmission expansion planning, *Proc. IEEE PowerTech*, Eindhoven, The Netherlands, 29 June-2 July 2015.

[26] N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, Redispatch index for assessing bidding zone delineation, *IET Generation, Transmission & Distribution*, vol. 11, no. 17, pp. 4248-4255, 2017.

[27] R. J. Sánchez-García *et al.*, Hierarchical Spectral Clustering of Power Grids, *IEEE Trans. Power Syst.* vol. 29, no. 5, pp. 2229-2237, 2014.

[28] E. Shayesteh, B.F. Hobbs, L. Söder and M. Amelin, ATC-Based System Reduction for Planning Power Systems With Correlated Wind and Loads, *IEEE Trans. Power Syst.* vol. 30, no. 1, pp. 429-438, 2015.

[29] K. Van den Bergh, C. Wijsen, E. Delarue, and W. D'haeseleer, The impact of bidding zone configurations on electricity market outcomes, *Proc. IEEE ENERGYCON*, Leuven, Belgium, 4-8 April 2016.

[30] D.V. Volodin and T.A. Vaskovskaya, Clustering approach for determination of congestion zones on nodal electricity markets in long term periods, *Proc. IEEE PowerTech*, Eindhoven, The Netherlands, 29 June-2 July 2015.

[31] F. Wang and K.W. Hedman, Reserve zone determination based on statistical clustering methods, *Proc. NAPS*, Champaign, IL, 2012.

[32] K. Wawrzyniak, G. Oryńczak, M. Kłos, A. Goska, and M. Jakubek, Division of the energy market into zones in variable weather conditions using Locational Marginal Prices, *Proc. 39th IECON*, Vienna, Austria, 2013, pp. 2027-2032.

[33] H. Yang, A new clustering method for partitioning price zone in power market environment, *Periodica Polytechnica Ser. El. Eng.*, vol. 48, no. 3-4, pp. 183-195, 2004.

[34] H. Yang, R. Zhou, and J. Liu, A RBFN hierarchical clustering based network partitioning method for zonal pricing, *Proc. 2nd International Conf. on Electrical and Electronics Engineering*, Mexico City, Mexico, 9 September 2005, pp. 282–285.

[35] H. Yang and R. Zhou, Monte Carlo Simulation Based Price Zone Partitioning Considering Market Uncertainty, *Proc. PMAPS*, Stockholm, Sweden, 11-15 June 2006.

[36] N. Yao, J. Wu, K. Liu, and J. Cai, Dynamic locational marginal prices based zonal division in large-scale regional electricity markets, *Proc. 12th WCICA*, Guilin, China, 12-15 June 2016, pp. 2443-2448.

[37] Y.T. Yoon, J.R. Arce, K.K. Collison, and M.D. Ilic, *Implementation of Cluster-based Congestion Management Systems*, Energy Laboratory Publication # MIT EL 00-001 WP, Massachusetts Institute of Technology, 2000.

[38] R. Zhang, D. Wang, and W.Y. Yun, Power-Grid-Partitioning Model and its Tabu-Search-Embedded Algorithm for Zonal Pricing, *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 15927-15932, 2008.

[39] M. Jakubek, K. Wawrzyniak, M. Kłos, and M. Blachnik, Are locational marginal prices a good heuristic to divide energy market into bidding zones?, *Proc. 12th Int. Conf. European Energy Market*, Lisbon, Portugal, 2015.

[40] A. Griffone, A. Mazza and G. Chicco, Applications of Clustering Techniques to the Definition of the Bidding Zones, submitted to *UPEC 2019*, Bucharest, Romania, 3-6 September 2019.

[41] A. Mazza *et al.*, Creation of a computational framework for the European transmission grid with Power-to-Gas, submitted to *UPEC 2019*, Bucharest, Romania, 3-6 September 2019.

[42] C. Bovo *et al*., Review of the Mathematic Models to Calculate the Network Indicators to Define the Bidding Zones, submitted to *UPEC 2019*, Bucharest, Romania, 3-6 September 2019.

[43] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O.P. Patel, A. Tiwari, M. Joo Er, W. Ding, C.-T. Lin, A review of clustering techniques and developments, *Neurocomputing*, vol. 267, pp. 664-681, 2017.