Optimal Bidding Zone Configuration: Investigation on Model-based Algorithms and their Application to the Italian Power System

Optimal Bidding Zone Configuration: Investigation on Model-based Algorithms and their Application to the Italian Power System

Luigi Michi
Strategy, Development and System Operation
Terna
Rome, Italy
luigi.michi@terna.it

Federico Quaglia
Dispatching and Operation
Terna Rete Italia
Rome, Italy
federico.quaglia@terna.it

Ettore Bompard
Politecnico di Torino
Dipartimento Energia
Turin, Italy
ettore.bompard@polito.it

Andrea Griffone
Politecnico di Torino
Dipartimento Energia
Turin, Italy
andrea.griffone@polito.it

Cristian Bovo
Politecnico di Milano
Dipartimento di Energia
Milan, Italy
cristian.bovo@polimi.it

Enrico Maria Carlini
Dispatching and Operation
Terna Rete Italia
Rome, Italy
enricomaria.carlini@terna.it

Luca Luzi
Dispatching and Operation
Terna Rete Italia
Rome, Italy
luca.luzi@terna.it

Gianfranco Chicco
Politecnico di Torino
Dipartimento Energia
Turin, Italy
gianfranco.chicco@polito.it

Andrea Mazza
Politecnico di Torino
Dipartimento Energia
Turin, Italy
andrea.mazza@polito.it

Valentin Ilea
Politecnico di Milano
Dipartimento di Energia
Milan, Italy
cristian.bovo@polimi.it

Mauro Caprabianca
Dispatching and Operation
Terna Rete Italia
Rome, Italy
mauro.caprabianca@terna.it

Giuseppina Nuzzo
Dispatching and Operation
Terna Rete Italia
Rome, Italy
giuseppina.nuzzo@terna.it

Pietro Colella
Politecnico di Torino
Dipartimento Energia
Turin, Italy
pietro.colella@polito.it

Angela Russo
Politecnico di Torino
Dipartimento Energia
Turin, Italy
angela.russo@polito.it

Abstract — **This paper focuses on model-based approaches that could be adopted for identifying alternative configurations to be considered in a bidding zone review process. Considering the complexity of this task, automated procedures can significantly help transmission system operators, allowing them to assess a large amount of possible system conditions and future scenarios. These methodologies are based on a 2-step approach: in the first step relevant nodal quantities are computed; then, in the second step, nodes are aggregated into zones using proper clustering algorithms. This paper starts with a critical review of the existing proposals, highlighting advantages and disadvantages of each of them, focusing on their practical implementation on a wide-area power system and in the context of the current electricity market framework.**

Some promising options are then identified for the Italian Power System case. In particular, an improved security constrained optimal power flow algorithm for computing Locational Marginal Prices (LMPs) and for identifying relevant critical branches (to be considered in the Power Transfer Distribution Factors computation) has been developed. Then, a selected set of clustering algorithms has been implemented and tested to check their effectiveness in forming LMP-based bidding zones.

Keywords — bidding zones, model-based, locational marginal process, power transfer distribution factors.

I. Introduction

The European Internal Energy Market model for electricity is based on a zonal approach: the European Power System (or part of it) is represented as composed of several interconnected zones ("bidding zones") in the context of the energy market clearing algorithms. Trades between different bidding zones are constrained according to cross-border capacities defined by relevant Transmission System Operators (TSOs), whereas trades are freely to occur inside each bidding zone (in contrast to"nodal" models, where all network constraints are represented in a detailed way).

In this framework, a high level of overall market efficiency can be achieved only if the implemented bidding zone configuration is able to correctly reflect most relevant congestion patterns and the effect of the trades on them, lowering the impact of the unconstrained internal trades assumption and achieving an optimal trade-off between energy market efficiency and redispatching costs. A well-designed structure is crucial in order to provide correct locational price signals, driving short-term dispatching decisions and (assuming other factors being equal) long-term planning investments. Significant price deviations could arise among different bidding zones.

The task of designing the optimal configuration is getting more and more complex in the context of the ongoing energy transition. Huge penetration of distributed generation and demand response resources, wind and solar variable infeed, decommissioning of big thermal base-load power plants, increased demand sensitivity to weather conditions (e.g., temperature), increased frequency of occurrence of simultaneous scarcity situations in two or more neighboring market zones because of unavailability of several transmission lines or power plants or due to extreme weather events observed and expected as consequence of the ongoing climate change, are making power flow patterns more and more variable (season by season, day by day, hour by hour). Hence, critical network sections are changing their locations according to weather conditions and commodity prices, requiring a bidding zone configuration able to cope with a large set of possible system conditions.

All of these elements are well reflected in the Commission Regulation (EU) 1222/2015 of 24th July 2015, also known as Guideline on Capacity Allocation and Congestion Management (CACM) [1], where a European framework for monitoring and/or reviewing existing Bidding Zone Configuration is defined. CACM requires TSOs to identify the configuration that provides the best trade-off between three main criteria: network security, market efficiency and long-term stability and robustness.

The first CACM compliant Italian Bidding Zone Review has been completed in 2018: small changes[1] to the Bidding Zone configuration have been implemented starting from the 1st of January 2019 [2], while most relevant modifications[2] will be applied in 2021 [3]. In this study Terna, the Italian TSO, identified and tested alternative Bidding Zone configurations according to an expert-based approach: using historical data and relevant information available, several options have been identified according to expert opinions.

This paper focuses on model-based approaches that could complement expert-based assessments in the task of identifying alternative configurations (as requested also in [4]): due to the complexity of this activity, automated procedures can significantly help the TSOs, allowing them to assess a large amount of possible system conditions and future scenarios. These methodologies are generally based on a 2-step approach. In the first step, relevant nodal indicators are computed. Then, in the second step, nodes are aggregated into zones by using proper clustering algorithms.

The next part of this paper starts with a critical review of the nodal indicators (Section 2) and clustering algorithms (Section 3), focusing on their practical implementation on a large-scale power system and in the context of the current electricity market framework. In Section 4, some promising options are identified and applied to an Italian Power System study case; the relevant results are presented and critically discussed. The last section contains the conclusions.

II. NODAL INDICATORS

A. Nodal indicators computation review

Due to its paramount role in the electricity market operation, optimal bidding zone configuration is a crucial element in electricity market design [5][6][7]. Hence advanced approaches could be introduced in bidding zone review processes, using computational models that accurately represent (i) the transmission network with relevant security limits, and (ii) the macroeconomic environment and the market rules. These imply the use of advanced market models, i.e., Security Constrained Optimal Power Flow (SCOPF) models.

A recent literature review [8] shows that the most used indicators adopted to define the bidding zone configuration are the Locational Marginal Prices (LMPs) and the Power Transfer Distribution Factors (PTDFs). The LMPs are obtained directly from the results of the SCOPF model, while the PTDFs represent first order sensitivities of the power transits with respect to nodal power injections, sensitivities calculated in the solution of the SCOPF.

Therefore, the quality of the indicators strongly depends on the structure of the SCOPF. Reference [8] shows that the transmission network is generally modeled using a DC PF representation, thus losing any information regarding potential voltage problems. An additional critical aspect is the modelling of the relevant, i.e., real power flow related, network security criterion: in general, only the N security criterion is considered in the literature [8]. This limitation is sometimes (only partially) overcome by indirectly considering the N-1 security criterion in the same formal structure of the N security criterion, i.e. by reducing the current limits of each branch. According to [8], an explicit representation of N-1 security criterion has not yet been used. Moreover, also the corrective remedial actions and their impact on the network indicators have not been considered: they can only be included if the N-1 security criterion is explicitly adopted.

Due to the impact in terms of computational effort linked to a full AC model, in this paper an improved DC SCOPF for computing LMPs is presented: the related AC extension is currently under development.

B. A DC SCOPF model with explicit N-1 security

Following, a DC SCOPF model with N-1 corrective security criteria is proposed. Here, the generation and demand real power profiles are represented through market bids laid down by rules of the Italian Day-Ahead Market [9] where few step-wise price-quantity bids for each production or consumption unit are adopted. Thus, for a given time interval, the goal of the SCOPF is to maximize the social welfare:

$$\max \sum_{\substack{k \in \boldsymbol{LOD} \\ m \in \boldsymbol{BIDd}}} C_{D,k}^{m} \cdot \Omega_{D,k}^{m} - \sum_{\substack{k \in \boldsymbol{GEN} \\ m \in \boldsymbol{BIDg}}} C_{G,k}^{m} \cdot \Omega_{G,k}^{m} \tag{1}$$

where $\boldsymbol{LOD}$ and $\boldsymbol{GEN}$ are the sets of demands and generators in the grid, respectively; $\boldsymbol{BIDd}$ and $\boldsymbol{BIDg}$ are the sets of demand and generator bids, respectively; $C_{D,k}^{m}$ and $C_{G,k}^{m}$ are the submitted prices in the demand and generator bids, respectively; $\Omega_{D,k}^{m}$ and $\Omega_{G,k}^{m}$ are the accepted quantities of the demand and generator bids, respectively – positive variables upper bounded by the value of the submitted bids.

The generators production and demand consumption is:

$$P_{G,k} = \sum_{m \in \boldsymbol{BIDg}} \Omega_{G,k}^{m}, \forall k \in \boldsymbol{GEN} \tag{2,a}$$

$$P_{D,k} = \sum_{m \in \boldsymbol{BIDd}} \Omega_{D,k}^{m}, \forall k \in \boldsymbol{LOD} \tag{2,b}$$

The convexity of the bids [9] combined with (1) will give a successive acceptance of the offered quantities according to the competitiveness of the bid prices.

The objective function (1) is subject to transmission constraints. The N security, normal operating, conditions are represented by the DC PF equations:

$$\boldsymbol{P} = \boldsymbol{B}_{\boldsymbol{bus}}^{\boldsymbol{N}} \cdot \boldsymbol{\theta}^{\boldsymbol{N}} + \boldsymbol{B}_{\boldsymbol{PST}}^{\boldsymbol{N}} \cdot \boldsymbol{\varphi}_{\boldsymbol{PST}}^{\boldsymbol{N}} \tag{3}$$

where $\boldsymbol{P}$ is the vector of nodal power injections; $\boldsymbol{B}_{\boldsymbol{bus}}^{\boldsymbol{N}}$ is the bus admittance matrix of the DC PF method; $\boldsymbol{B}_{\boldsymbol{PST}}^{\boldsymbol{N}}$ is the bus to phase-shifting transformers (PST) admittance matrix; $\boldsymbol{\theta}^{\boldsymbol{N}}$ is the vector of unknowns of the nodal voltage phasors and $\boldsymbol{\varphi}_{\boldsymbol{PST}}^{\boldsymbol{N}}$ is the vector of unknowns of the PSTs phases.

The power flows through the network's branches $\boldsymbol{F}^{\boldsymbol{N}}$ are:

$$\boldsymbol{F}^{\boldsymbol{N}} = \boldsymbol{B}_{\boldsymbol{br}}^{\boldsymbol{N}} \cdot \boldsymbol{\theta}^{\boldsymbol{N}} + \boldsymbol{B}_{\boldsymbol{brPST}}^{\boldsymbol{N}} \cdot \boldsymbol{\varphi}_{\boldsymbol{PST}}^{\boldsymbol{N}} \tag{4}$$

where $\boldsymbol{B}_{\boldsymbol{br}}^{\boldsymbol{N}}$ and $\boldsymbol{B}_{\boldsymbol{brPST}}^{\boldsymbol{N}}$ are the branch-to-nodes and branch-to-PSTs admittance matrices, respectively.

The maximum branch power flow constraints are:

[1] 3 out of the 4 previously existing national virtual Bidding Zones have been merged to the adjacent geographical Bidding Zone.

[2] Umbria region will be moved from Central North to Central South BZ, a new "Calabria" Bidding Zone will be created and the last national virtual Bidding Zone will be eliminated ("Rossano").

$$-\boldsymbol{F}^{N}_{max} \leq \boldsymbol{F}^{N} \leq \boldsymbol{F}^{N}_{max} \quad (5)$$

where $\boldsymbol{F}^{N}_{max}$ is the vector of maximum power flow allowed through the branches of the network.

The generators and PST capability constraints are:

$$\boldsymbol{P}^{min}_{g} \leq \boldsymbol{P}_{g} \leq \boldsymbol{P}^{max}_{g} \quad (6,a)$$

$$\boldsymbol{\varphi}^{min}_{PST} \leq \boldsymbol{\varphi}^{N}_{PST} \leq \boldsymbol{\varphi}^{max}_{PST} \quad (6,b)$$

where $\boldsymbol{P}^{min}_{g}$, $\boldsymbol{P}^{max}_{g}$ are the vectors of minimum, maximum power outputs of generators, respectively; $\boldsymbol{\varphi}^{min}_{PST}$, $\boldsymbol{\varphi}^{max}_{PST}$ are the vectors of minimum, maximum PST phases, respectively.

The application of corrective actions following the outage of a branch involves two stages. First, the outage occurres but not enough time passes to implement the corrective actions: the network finds itself in conditions analogous to N-1 preventive security, since only the outage impact on the network operation is present; the N-1 preventive security constraints need to be formulated but they can be relaxed (reflecting the temporary admissible loading capacity of the elements). In the second stage, corrective actions are applied to solve the short-term violations to guarantee long-term security (e,g. longer than 20 min).

The effect of the outage of branch b on the power flow in branch a is assessed using the Linear Outage Distribution Factors, i.e., DF^{b}_{a} [10]. Thus, the equivalent N-1 preventive security conditions are expressed as:

$$-\boldsymbol{k}^{prev}_{a,b} \cdot \boldsymbol{F}^{N}_{max,a} \leq \boldsymbol{F}^{N}_{a} + \boldsymbol{DF}^{b}_{a} \cdot \boldsymbol{F}^{N}_{b} \leq \boldsymbol{k}^{prev}_{a,b} \cdot \boldsymbol{F}^{N}_{max,a} \quad (7)$$

Eq. (7) is written for all (a, b) combinations that result critical for the grid. In (7), $\boldsymbol{F}^{N}_{a}$ and $\boldsymbol{F}^{N}_{b}$ are vectors of $\boldsymbol{F}^{N}$ variables defined according to the (a, b) combinations; $\boldsymbol{F}^{N}_{max,a}$ is the vector of maximum power flow allowed through the a branches, while $\boldsymbol{k}^{prev}_{a,b}$ is a coefficients vector defined by the user and allowing the scaling of $\boldsymbol{F}^{N}_{max,a}$.

The impact of the corrective actions on the power flow of branch a at the outage of branch b, ΔF^{b}_{a}, is assessed using linear sensitivities calculated on the DC PF model. Thus:

$$\Delta F^{b}_{a} = \boldsymbol{DF}^{b}_{g} \cdot \boldsymbol{\Delta P}^{b}_{g} + \boldsymbol{DF}^{b}_{PST} \cdot \boldsymbol{\Delta\varphi}^{b}_{PST} \; \forall(a, b) \quad (8)$$

where $\boldsymbol{\Delta P}^{b}_{g}$ is the vector of corrective adjustments in the output of the generating units following the outage of branch b; $\boldsymbol{\Delta\varphi}^{b}_{PST}$ is the vector of corrective adjustments in the PST phases following the outage of branch b, while $\boldsymbol{DF}^{b}_{g}$ and $\boldsymbol{DF}^{b}_{PST}$ are matrices of first order sensitivities calculated using the DC PF model of the network in N-1 conditions.

Thus, the N-1 corrective security constraints considering all the (a, b) combinations that are critical for the grid are:

$$-\boldsymbol{k}^{cor}_{a,b} \cdot \boldsymbol{F}^{N}_{max,a} \leq \boldsymbol{F}^{N}_{a} + \boldsymbol{DF}^{b}_{a} \cdot \boldsymbol{F}^{N}_{b} + \boldsymbol{\Delta F}^{b}_{a} \quad (9,a)$$

$$\boldsymbol{F}^{N}_{a} + \boldsymbol{DF}^{b}_{a} \cdot \boldsymbol{F}^{N}_{b} + \boldsymbol{\Delta F}^{b}_{a} \leq \boldsymbol{k}^{cor}_{a,b} \cdot \boldsymbol{F}^{N}_{max,a} \quad (9,b)$$

where $\boldsymbol{\Delta F}^{b}_{a}$ is the vector of ΔF^{b}_{a} for all (a, b) combinations, $\boldsymbol{k}^{cor}_{a,b}$ is a vector of coefficients defined by the user allowing the proper scaling of $\boldsymbol{F}^{N}_{max,a}$. Constraints (9) consist of the superposition of the two stages, since the corrective actions are applied only in the second stage after the outage; mathematically, this superposition is possible due to the use of the DC PF model, which is a linear model.

It is also necessary to constrain the corrective actions to the capability of the devices that apply them:

$$\boldsymbol{\varphi}^{min}_{PST} \leq \boldsymbol{\varphi}^{N}_{PST} + \boldsymbol{\Delta\varphi}^{b}_{PST} \leq \boldsymbol{\varphi}^{max}_{PST} \quad \forall(a, b) \quad (10,a)$$

$$\boldsymbol{P}^{min}_{g} \leq \boldsymbol{P}_{g} + \boldsymbol{\Delta P}^{b}_{g} \leq \boldsymbol{P}^{max}_{g} \quad \forall(a, b) \quad (10,b)$$

Finally, the ramp constraints on the corrective actions of the generators are introduced:

$$\boldsymbol{\Delta P}^{\downarrow}_{g} \leq \boldsymbol{\Delta P}^{b}_{g} \leq \boldsymbol{\Delta P}^{\uparrow}_{g} \;\; \forall(a, b) \quad (11)$$

where $\boldsymbol{\Delta P}^{\downarrow}_{g}$, $\boldsymbol{\Delta P}^{\uparrow}_{g}$ are the vectors representing the ramp up and ramp down limits of the generators, respectively.

The set of (a, b) combinations considered by the model is defined following the iterative procedure:

- *Step 0*: the $\boldsymbol{DF}^{b}_{a}$ matrix is calculated for all (a, b) combinations; the set of (a, b) combinations used in the DC SCOPF problem is considered initially null;
- *Step 1*: the DC SCOPF solution is found;
- *Step 2*: a (a, b) combination is included in the set if conditions (12) are both satisfied:

$$DF^{b}_{a} \cdot F^{N}_{b} \geq e_1 \cdot F^{N}_{max,a} \quad (12,a)$$

$$F^{N}_{a} + DF^{b}_{a} \cdot F^{N}_{b} \geq e_2 \cdot F^{N}_{max,a} \quad (12,b)$$

  where e_1 and e_2 are user-defined parameters. In simple terms, a pair (a, b) is considered in the DC SCOPF problem if the outage of b has significant impact on a, (12,a) and if it brings the power flow in a in a critical region, (12,b).
- *Step 3*: if the set of (a, b) combinations is updated then the process resumes from *Step 1*, otherwise it stops.

Finally, the LMPs are given by the Lagrangian multipliers associated to the DC PF equations in N security conditions (3): implicitly they include the effects of the N-1 security constraints since the involved variables, i.e., $\boldsymbol{P}_{g}$ and $\boldsymbol{\varphi}^{N}_{PST}$, are present also in N-1 security constraints, e.g. (10).

III. Clustering Algorithms

A. Role of clustering algorithms

Clustering methods are widely used for grouping a number of entities on the basis of data that represent their characteristics or behaviour [11]. In short, the grouping process applied to time series of data used in the electricity sector [12] contains the following phases:

- *data gathering and processing*: measurement or calculation of the data for the M entities under study, and treatment of bad data or missing data;
- *pre-clustering*: data pre-screening, selection of H representative features, and formation of the MxH input data matrix;
- *clustering*: selection of the clustering algorithm(s) to be used, formation of the clusters by using the clustering algorithm(s), formation of the cluster centroids (if needed), and computation of the clustering validity indicators (if a comparison among the results from different algorithms or variants is needed);
- *post-clustering*: formation of the final groups taking into account possible external constraints or links among the entries belonging to the resulting clusters.

Each clustering method has some pros that make it suitable to identify specific regularities in the data structures. However, there is no one-size-fits-all clustering approach. The results are always data-driven. Thereby, the role of the expert of the domain is always essential, to avoid troubles or misleading solutions.

In the cases analysed in this paper, the entities under study are the *M* nodes of the electrical transmission system, and the LMPs resulting at each node for a given number of hourly time steps are the relevant entries. The dataset used comes from a real system and does not contain any missing nor bad data.

B. Pre-clustering phase

In the specific case of LMPs, useful considerations come from a pre-screening of the data available, carried out before running any clustering procedure, with the aim of endeavouring possible simplifications and avoiding useless and time consuming executions.

In this pre-screening, one of the relevant aspects is the number of time steps at which the LMPs differ in different nodes. The standard deviation at each time step is calculated, then all time steps with standard deviation null (or lower than a predefined threshold) are removed from the dataset, as they do not provide useful information for highlighting the differences among the nodes. In this way, the LMPs in the *H* remaining time steps can be used directly as clustering features. The number (and percentage) of the remaining time steps is recorded to keep track of the portion of data eliminated, as the same number of resulting bidding zones could be more significant when it comes from a larger percentage of time steps analysed. An alternative is to explore the possibility to define the features by resorting to data compression techniques; the analysis of this possibility is outside the scope of this paper. The input data matrix is then formed by the MxH entries.

The *H* time steps are then subject to a further check, to identify possible regularities in the spatial distribution of the LMPs. Let us consider an example taken from real data, in which in 22 hours over 24 the LMPs are always equal, and only in the other two hours there is a variability of the LMPs across the nodes, but with only *two* different variants for the LMP distribution. In this case, it is possible to get a solution from clustering only when the number of bidding zones does not exceed 2. This is a key limiting factor for using the clustering algorithm that need a predefined number of clusters. In the case indicated, if the predefined number of clusters is set to a value higher than 2, there will be no convergence of the clustering algorithm. In other words, a pre-screening is needed to ensure that there is a sufficient variety of LMPs in the data set, to avoid running a clustering algorithm that would never converge.

C. Clustering and post-clustering phases

The clustering algorithms have been used for the definition of the bidding zones in various works. An overall view is presented in [13], while specific aspects of the implementation of various clustering algorithms are discussed in [14]. In particular, the clustering algorithms have been generally executed by using as features the LMPs or the PTDFs of the most critical elements. In this paper, the focus is set on the use of LMPs, and on the testing of a set of specific clustering algorithms that have provided good performance in the study carried out on a reduced model of the European transmission network [14], namely:

- *k-means* (KM) clustering [15], in the version that adopts the squared Euclidean distance metric and mitigates the issue of random initialisation by using the k-means++ algorithm [16] for cluster center initialization; and,
- *hierarchical* clustering [11] with different linkage criteria, that is, different ways to calculate the distance between clusters to decide how to merge pairs of clusters on the basis of the minimum distance; in particular, the cases tested include the Ward's minimum variance linkage criterion (HW) [17], the single linkage or nearest neighbour criterion (HS), and the average linkage criterion (HA).

For both algorithms, the results of the basic versions implemented in the Matlab® suite are considered. However, these versions form the zones only on the basis of the numerical values of the LMPs, and do not consider the connections among the nodes. As such, it is very likely to obtain non-connected zones. This issue may be tackled in different ways. One possibility is to operate a post-processing of the clustering results in which, starting from the predefined number of clusters, the non-connected zones are splitted, leading to the increase in the number of final zones. Another way is to modify the code of the clustering algorithm, by incorporating the node connection check inside the algorithm, in which a distance matrix is used to represent how the pairs of nodes are connected, and a penalty factor is applied to the entries of a distance matrix when the nodes are not connected. This makes less likely to merge non-connected nodes in the clustering procedure. On these bases, two further versions of the clustering algorithms that use the distance matrix with penalty factors have been implemented, denoted as:

- customised k-means clustering (KM_C); and,
- customised hierarchical clustering (HW_C).

IV. Case Study

A. Real system used in the case study

The model of the electrical transmission system used in this paper is taken from the Italian High Voltage network, composed of $M = 1535$ nodes and $L = 3483$ lines. The nodes are relevant for the analysis with the LMPs, while the lines are used to verify the physical connection among the nodes belonging to the same zone. The number of nodes is higher than the number of physical locations (substations), as the possible splitting of the busbars to allow multi-node operation inside the same substation is taken into account.

Fig. 1 shows the current Italian bidding zone configuration. There are 7 main zones (North, Centre-North, Centre-South, South, Rosn, Sicily and Sardinia) and five other sets of nodes located at the cross-border connections with Austria, Corsica, France (mainland), Slovenia, and Switzerland). This configuration (defined with an expert-based approach and selected after a detailed comparison with other expert-based alternatives [3]) is taken as a reference for the study carried out in this paper.

B. Input data

The data used for the tests contain the hourly LMPs for four days (scenarios) that represent different system conditions, namely:

- D1, a scenario with very low demand (January 1st, 2019);
- D2, a scenario with high demand (January 15th, 2019);
- D3, a scenario with several planned outages of network branches (March 7th, 2019); and,
- D4, a scenario with high production from photovoltaic systems (March 29th, 2019).

From data pre-screening based on the calculation of the number of hourly time steps with standard deviation across the network nodes higher than $\varepsilon_\sigma = 10^{-5}$ is 2 for D1, 15 for D2, 10 for D3, and 24 for D4, for a total of $H = 51$ time steps (53% of the available time steps) originated from merging the four days. The clustering analysis is then carried out by considering these 51 time steps. The input data matrix that contains the LMP values has dimensions 1535x51. The adjacency matrix that contains the information on the pairs of connected nodes is also provided.

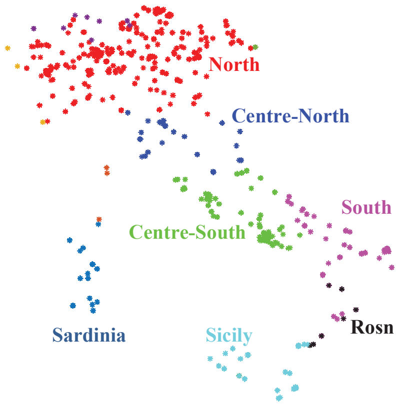

Fig. 1 Present bidding zones in Italy.

C. Clustering results and discussion

All the clustering algorithms tested require to predefine the number of clusters. The results reported in this section are obtained by setting the number of clusters to the number of the main zones in the present situation, $K = 7$.

The results shown in Fig. 2 contain the partitioning obtained by running the basic clustering algorithms without checking the connections of the final zones (i.e., with $K = 7$ and possible non-connected zones), and after the application of zone splitting on the basis of the previous results, to obtain a number $K' \geq K$ of connected zones.

In particular, the final number of zones K' becomes 10 for KM, and 9 for HW, HA and HS. However, the solutions found are rather different. KM and HW tend to create a partition in which the North and Centre-North zones are merged, while HA and HS tend to create a large group and to single out some zones that contain a few nodes. Sicily and Sardinia are in general kept as separate zones.

Fig. 3 reports the partitioning obtained by using the customised versions of the clustering algorithms, in which the post-processing check on the connection of the zones has been activated to verify the effectiveness of using the internal connection check based on penalty factors for the non-connected nodes. The results show that the KM_C version (with 11 zones) provides results consistent with KM (10 final zones). For the hierarchical clustering algorithms, HW_C and tends to create a bigger number of zones (20 zones, with respect to the 7 initial zones), thus failing in the appropriate application of the penalty factor, while HA_C and HS_C maintain the predefined number of 7 final zones, but with different partitionings, with HS_C confirming the attitude of the single linkage criterion to create a large group and to single out some zones that contain a few nodes.

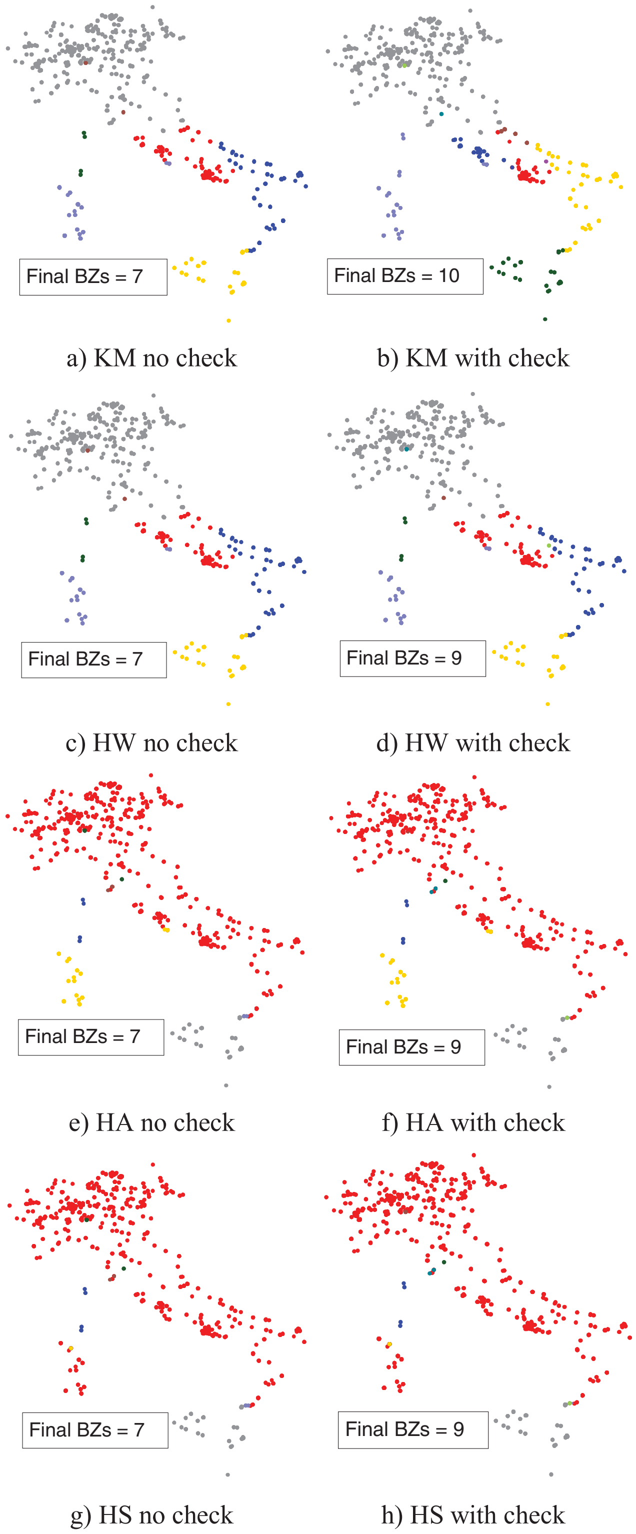

Fig. 2. Partitioning by using the basic versions of the clustering algorithms without and with the check on the bidding zone connection.

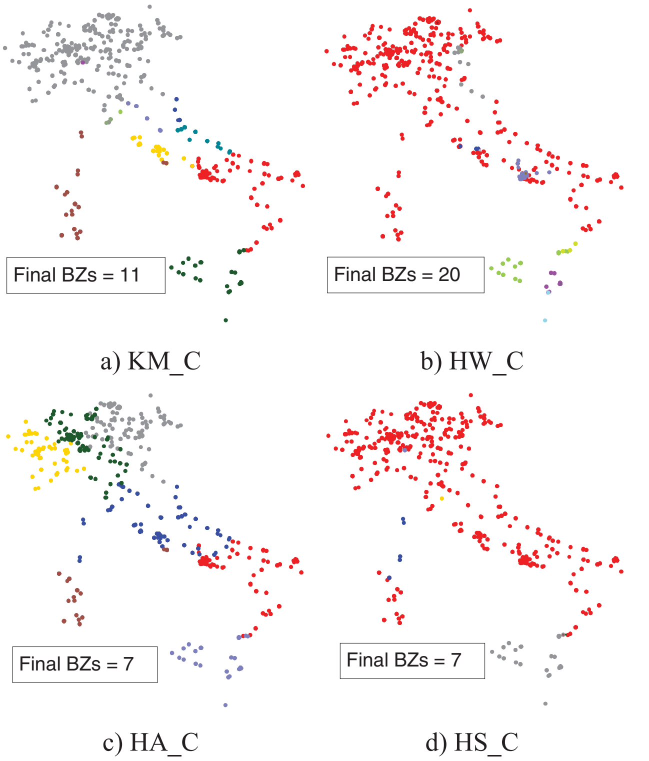

Fig. 3. Partitioning of the bidding zones by using the customised versions of the clustering algorithms with final check of the zone connection.

V. Conclusions

Today, a large set of automated methods (clustering algorithms) is available for defining alternative bidding zone configurations starting from nodal information (LMPs or PTDFs). In this paper, the results of selected standard clustering approaches and of some customized versions have been presented for a limited set of Italian Power System scenarios. In general, results obtained from the KM (k-means clustering with and without topological check) and from HW (hierarchical clustering with and without topological check) clustering algorithms are substantially aligned with the current bidding zone configuration: only North and Central North bidding zones tend to be merged, having considered only a limited set of input scenario in this test (without relevant congestions in this area). Instead HA (average linkage criterion) and HS (single linkage or nearest neighbour criterion) tend to produce very different aggregation that seem not compatible with Terna's expectations (in light of the input LMP data). The introduction of distance matrix penalty factors resulted in very different aggregation for all the clustering algorithms except for the KM (which resulted to be the most robust approach considering the test cases under assessment). In all cases, some very small zones are found from the clustering algorithms. The information on these small zones, if confirmed using a larger set of scenarios, may be useful for strategic reasoning, even though these clustered zones may be too small to be proposed as bidding zones. Further insights on these aspects will be provided in the next (future) work to extent this paper. The selection of a proper algorithm is then confirmed to be a crucial data-driven exercise: different algorithms (or even the same algorithm in its different versions) could lead to very different clusters/bidding zones. The role of the power system expert in the interpretation of the results is therefore essential.

Further work is in progress for the refinement of the clustering algorithms used in this paper and for the application to the Italian system of the clustering algorithms by using the PTDFs as features. A larger set of scenarios will be also introduced in the assessment in order to take into account all the potential system working conditions and produce stable and reliable results for an improved comparison with the existing bidding zone configuration.

References

[1] European Commission, Commission Regulation (EU) 2015/1222 of 24 July 2015, web: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R1222&from=EN (accessed 12 August 2019).

[2] 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 (accessed 12 August 2019).

[3] 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 (accessed 12 August 2019).

[4] 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 (accessed 12 August 2019).

[5] L.H. Lam, V. Ilea, C. Bovo, "Impact of the price coupling of regions project on the day-ahead electricity market in Italy", in *Proc. 2017 IEEE Manchester PowerTech*, Powertech 2017, pp. 1-6.

[6] L. H. Lam, V. Ilea, and C. Bovo, "European day-ahead electricity market coupling: Discussion, modeling, and case study," *Electric Power Systems Research*, vol. 155, no.Supplement C, pp.80-92, 2018.

[7] L. H. Lam, V. Ilea, and C. Bovo, "Integrated European intra-day electricity market: Rules, modeling and analysis", *Applied Energy*, vol. 238, 15 March 2019, pp. 258-273.

[8] C. Bovo, V. Ilea, E. M. Carlini, M. Caprabianca, F. Quaglia, L. Luizi, G. Nuzzi, "Review of the Mathematic Models to Calculate Network Indicators for Bidding Zones Definition", *54th International Universities Power Engineering Conference (UPEC 2019),* Bucharest, Romania, 3-6 September 2019.

[9] GME, "*Spot Electricity Market (MPE)*", web: https://www.mercatoelettrico.org/En/Mercati/MercatoElettrico/MPE.aspx (accessed 12 August 2019).

[10] M.K. Enns, J.J. Quada, B. Sackett, "Fast linear contingency analysis", IEEE Transactions on Power Apparatus and Systems, vol. 101, No. 4, 1982.

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

[12] 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.

[13] 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", *54th International Universities Power Engineering Conference (UPEC 2019),* Bucharest, Romania, 3-6 September 2019.

[14] A. Griffone, A. Mazza, and G. Chicco, "Applications of clustering techniques to the definition of the bidding zones", *54th International Universities Power Engineering Conference (UPEC 2019),* Bucharest, Romania, 3-6 September 2019.

[15] J.T. Tou and R.C. Gonzalez, *Pattern recognition principles*, London, UK: Addison-Wesley, 1974.

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

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