

Advanced Clustering of Flow-Based Domains for Adequacy Study Purposes

Bashir Bakhshideh Zad, Behzad Vatandoust, Jean-François Toubeau, Zacharie De Grève, François Vallée
Power Systems and Markets Research Group
University of Mons, Belgium

Abstract— The cross-border capacities are prominent input data in the adequacy assessment of an interconnected electric power system. In order to incorporate the interconnection capacities through the Flow-Based (FB) domains in the adequacy assessments of the Central Western Europe (CWE) electricity system, the main challenge lies in the dependence of FB domains on factors that are unknown over the long-term horizon of the adequacy study. To tackle this challenge, the CWE Transmission System Operators (TSOs) employ a two-step methodology, which consists in first clustering the FB domains, then correlating the obtained FB clusters with relevant factors. This paper focuses on the first step of the methodology and studies advanced clustering techniques to improve the partition of FB domains. To this end, different dedicated distance measures, clustering approaches and cluster validation metrics are proposed. Our clustering studies conducted on the historical FB domains demonstrate that the fuzzy clustering technique in combination with the proposed dissimilarity distance measure lead to the cluster results, which are considerably improved with respect to the ones obtained by the application of k -medoids clustering method and Hausdorff distance, traditionally used by the TSOs.

Index Terms— adequacy assessment, flow-based market coupling, clustering techniques, interconnection capacities.

I. INTRODUCTION

Adequacy study evaluates the ability of an electric power system to meet the load demand over the studied horizon. Traditionally, the adequacy assessment has been carried out according to a deterministic formulation based on the amounts of peak load and available generation considering conservative contingency scenarios. In order to efficiently cope with the increasing uncertainty in the adequacy study (e.g., arisen from the renewable-based generations [1], [2]) while addressing the cost-effectiveness aspect (avoiding the reliance on conservative scenarios), the adequacy assessment methodologies are changing from deterministic towards probabilistic-based approaches [3]. In the latter case, Monte Carlo simulations are usually carried out to capture the uncertain nature of load and generation as well as the unplanned outages. The probabilistic (known also as risk-based) adequacy assessment determines through Monte Carlo simulations and economic dispatch, the probability, duration, and amount of energy shortfall in the studied scenarios [3].

Besides the available domestic generation and load demand as well as the unplanned outages, it exists another important factor in the adequacy study of an interconnected electric power system, namely the cross-border exchange capacities. The latter defines the amount of achievable import or export via the interconnections. Generally, the electricity systems (markets) are interconnected to benefit from the available resources in other zones and countries to eventually improve the social welfare of the involved countries. Currently, there are two approaches for incorporating the interconnection capacities in the electricity market in Europe, introduced as follows. The Net Transfer Capacity (NTC) approach that assumes a commercial capacity between two market zones, and the Flow-Based (FB) approach, which more accurately considers the physical grid constraints. The FB method aims to consider the interdependencies between the flows crossing different borders by relying on a joint methodology, shared by all the involved countries. Reference [4] describes concepts and definitions of the Flow-Based Market Coupling (FBMC). The NTC is the traditional capacity allocation approach still being used on specific borders in Europe, while the Central Western Europe (CWE) region, which is the focus of the current paper, has moved towards the FBMC since 2015. The FBMC is the target model to be applied to other regions of Europe [5].

In order to incorporate interconnection capacities through the FB domains in the adequacy assessments, the main challenge consists in finding the FB domains that efficiently represent the network constraints and exogenous conditions over the long-term horizon of the adequacy study (typically between one to several years ahead). In practice, the FB domains depend on factors such as network operating points and exogenous conditions (e.g., meteorological conditions that affect the load demand and renewable generation) [6], which cannot be known precisely over such a long-time horizon. To address this challenge, the French TSO (RTE), Belgian TSO (Elia) and European TSOs (ENTSO-E) employ a two-step methodology, which consists in first clustering the FB domains, then correlating the obtained FB clusters with the relevant important factors [5], [7]-[9]. More precisely, the objective of the clustering task is to group the FB domains into a reduced number of clusters, and to select the representative object (prototype) of each cluster. The goal of the correlation phase is to determine a link between the clustered FB data and the

This work is supported via the energy transition fund project 'ADABEL' organized by the FPS economy, S.M.E.s, Self-employed and Energy, Belgium.

external factors to assign the proper cluster representative to each sampled scenario of the risk-based adequacy study.

In the current paper, we conduct a gap analysis of the current clustering methodology employed by the TSOs, and we propose suggestions and solutions to improve its performance. This will lead to a better calibration of one of the most important adequacy assessment inputs, i.e., the interconnection capacities. Indeed, relying on wrong FB domains leads to consideration of unrealistic import and export capacities in the adequacy assessment that will certainly mislead the results.

This paper discusses the possible improvements in the clustering phase of the methodology (employed by the TSOs) by proposing tailored distance measures for comparison of FB domains, studying different clustering techniques, investigating impact of changing the number of clusters, and validating the clustered results by meaningful cluster validation metrics.

The remainder of this paper is structured as follows. In section II, the principles and motivation of the flow-based market coupling are discussed. The current methodology used by the TSOs for the adequacy assessment incorporating the FB domains is presented in section III. Then, the proposed improvements regarding the choice of proper dissimilarity measures and efficient clustering techniques are explained in sections IV and V, respectively. Afterwards, section VI presents the proposed cluster validation techniques, and section VII discusses the studied clustering cases and the obtained results. Finally, the last section is dedicated to the conclusions.

II. FLOW-BASED MARKET COUPLING (FBMC)

The FBMC aims at modelling the physically feasible power exchanges for the trade considering the interdependencies of flows within zones (countries). In practice, there is a fundamental difference between commercial energy trades and physical flows as in an electric power network, the power flows through the existing paths according to the Kirchhoff's laws. Consequently, the exchange capacity between two market zones cannot be fully allocated to the commercial trade between them, as some of the capacity will be used by flows resulting from the trade of other market zones [4]. In FBMC, physical possible flows for the trade are determined according to the FB domains, which are represented by a set of linear constraints as:

$$\mathbf{A} \times \mathbf{x} \leq \mathbf{b} \quad (1)$$

where \mathbf{A} is a matrix containing the Power Transfer Distribution Factors (PTDFs), \mathbf{x} is the net position (= export - import) of each zone or country, and \mathbf{b} is vector of the Remaining Available (capacity) Margin (RAM) of grid elements. The PTDF coefficients indicate the incremental physical flows induced on transmission lines as a result of a power exchange between two zones. Each row of the system of linear constraints (1) corresponds to one selected grid element. The FB domain, at a given hour corresponds to the intersection of all half-spaces created by the system of linear constraints (1), which eventually creates a N -dimensional polytope, where N is the number of countries involved in the FBMC. Currently, five zones (countries) in the CWE region participate in the FBMC, i.e., Germany, France, Belgium, Austria, and the Netherlands. Figure 1 illustrates a typical FB domain projected on various 2-dimensional (2D) planes. It shows the interconnection capacities between the (two) selected countries while

neglecting the coordinates relating to 3 other zones. It should be noted that in practice, a feasible exchange is determined in accordance with coordinates of all the (5) dimensions, i.e., an exchange to be feasible must be placed inside the FB polytope.

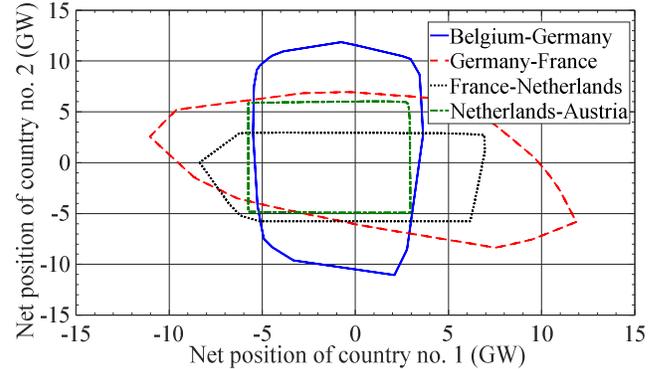


Figure 1. Projection of a typical FB domain on various 2D planes. The first (second) country appearing in the figure legend corresponds to the horizontal (vertical) axis.

III. ADEQUACY STUDY INCORPORATING FB DOMAINS

FB domains modelled through a set of linear constraints are integrated into the adequacy study by adding (1) in the optimization problem of the adequacy assessment.

A. What is the Challenge?

Integrating FB domains in the adequacy assessment with time horizons from one to several years poses several challenges. The main difficulty consists in finding, for each time step (generated scenario), the FB domain that correctly represents its network constraints and exogenous conditions. Indeed, the FB domain depends on factors such as network operating points and climatic conditions, which are unknown over such a long horizon.

B. Current Methodology Used by the TSOs

To address the aforementioned challenge, the French TSO (RTE), Belgian TSO (Elia) and European TSOs (ENTSO-E) rely on a strategy based on a two-step clustering-correlation procedure of the FB domains [5], [7]-[9], summarized below.

FB domains are firstly clustered into k groups according to their geometrical resemblance. A partitional clustering algorithm (i.e., k -medoids, see below) is employed to that end, in combination with a dissimilarity measure (or distance), which compares the geometrical shapes of the FB domains. More precisely, the distance between two arbitrary FB domains A and B is computed using the coordinates of the polytope vertices (shown in Figure 1 in 2D plane). To this end, the Euclidean distances between each vertex of A and the corresponding closest vertex of B are summed to constitute the total distance between A and B. The above procedure is applied to calculate all distances between every possible pair of FB domains. In the end, a square matrix is constructed that includes all distances (dissimilarities) of the FB domains in the studied dataset. The k -medoids clustering algorithm is then applied [5], [8] using the distance exposed above. It consists in a partitional clustering algorithm, which structures the input space by assigning each object to the cluster with the closest medoid.

The aim of the correlation phase is to identify a link between the partitioned FB domains and the external factors. Indeed,

adequacy analyses rely on Monte Carlo sampling of such factors. Considering a FB domain that is in line with the sampled scenarios is thus crucial in such a context. The shape of a FB domain is affected by several factors of different importance levels [6]. The objective is to carry out the correlation study with the most important factors affecting the FB domains. The final selected factors are correlated with the cluster memberships obtained in the previous phase, to obtain the probability of occurrence of each FB cluster for each factor combination (e.g., high, medium, and low levels). Once a scenario is generated within the (risk-based) adequacy study process, according to its corresponding factor combination, medoid of the cluster with the highest probability is employed, and the linear constraints encoded by that medoid are integrated into the economic dispatch problem.

C. Proposed Improvements

This paper focuses on the clustering phase of the above methodology. It aims to improve the quality of clustered FB data by defining new dissimilarity measures, examining various clustering methods, studying impact of changing the number of clusters, and defining efficient cluster validation techniques.

IV. PROPOSED DISTANCE MEASURES FOR DISSIMILARITY EVALUATION OF FB DOMAINS

The first direction of investigation on possible improvements to the methodology presented in the previous section is dedicated to the choice of the dissimilarity metric, or distance measure, which is needed to evaluate the degree of dissimilarity between two FB domains. The Hausdorff distance (HD), traditionally employed in computer vision (see e.g., [10]) is firstly proposed. Slight variations of the classical HD, which rely e.g., on a squared Euclidean norm, are then discussed. Another original distance measure based on the volume calculation of the compared FB polytopes is finally presented.

A. Hausdorff Distance (HD)

The Hausdorff distance is usually employed in computer vision (e.g., for shape matching and object recognition) in order to identify the physical shapes of objects [10]. Let A and B be two sets of points, with $A = \{a_1, a_2, \dots, a_n\}$, and $B = \{b_1, b_2, \dots, b_m\}$, which can represent (vertices of) two polytopes. The Hausdorff distance from A to B is defined as follows.

$$HD(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (2)$$

with the $\|\cdot\|$ referring to the Euclidean norm operator. The above expression is equivalent to find for every point of A, its closest point from B. The maximum of the above minimum distances gives the Hausdorff distance.

B. Modified Hausdorff Distance (MHD)

The modified Hausdorff distance is the metric employed by the TSOs in the framework of the methodology presented in section III.B. In contrast to the original HD, instead of finding the maximum of all the minimum distances, the sum of all the minimum distances is calculated as the MHD.

C. Modified Hausdorff Distance based on the Squared Euclidean Norm (MHD_SQE)

The third distance measure is similar to the MHD but uses the squared Euclidean norm $\|\cdot\|^2$ instead. Consequently, it

intensifies the large discrepancies between polytope shapes, which can bring an opportunity for a better clustering performance.

D. Volume-Based (VB) Metric

FB domains are polytopes with N dimensions, centered around the origin. They occupy a volume that can be calculated considering the convex hull of their vertices [11]. A new dissimilarity measure based on the volume calculation of compared polytopes is proposed in this paper according to

$$VB = 1 - \frac{\text{volume}(FB_1 \cap FB_2)}{\text{volume}(FB_1 \cup FB_2)} \quad (3)$$

The VB metric can take values between 0 and 1. A distance of 0 corresponds to two identical FB domains, whereas higher values demonstrate an increasing dissimilarity.

E. Combination of Hausdorff Distance and Volume-Based (HD2V) Metric

To provide a better distinction when comparing FB domains, a new dissimilarity measure is defined by combining the MHD_SQE and the VB distances as:

$$HD2V = \frac{\text{MHD_SQE}}{1 - VB} \quad (4)$$

The numerator of (4) is the modified Hausdorff distance with the squared Euclidean norm. Its denominator represents the similarity of two FB domains according to their volume calculation. In practice, this new index intensifies the distances between FB domains that are not similar while for rather similar objects, its impact is less. Indeed, the denominator of (4) can vary between 0 and 1, and for its small values, indicating little similarities, it nonlinearly increases the index value with respect to the MHD_SQE distance.

V. STUDIED CLUSTERING TECHNIQUES

Regarding the choice of clustering techniques, firstly, we consider three different k -medoids algorithms. In particular, the performance of the classical Partitioning Around Medoids (PAM) method is investigated alongside two more recent algorithms offering better exploration capabilities according to the data science literature. In addition, fuzzy and bottom-up clustering approaches are studied here to further evaluate the possible improvements in this direction.

A. k -medoids Algorithms

Similarly to k -means, k -medoids belongs to the family of partitional clustering algorithms, which structure the input space by assigning each data object to the cluster with the closest representative object (or prototype), according to a given distance measure. With k -medoids, the cluster representative objects are their respective medoids, i.e., the data objects that minimize the sum of distances with all the objects of the considered cluster, contrarily to k -means where the cluster prototypes are computed by taking the average of the cluster objects. Cluster representative objects are therefore existing physical objects with k -medoids, whereas prototypes with k -means are artificial objects, which may show undesirable non-physical properties.

The methodology developed by the TSOs relies on the k -medoids clustering. The clustering process is achieved by minimizing a non-convex loss function and may therefore be

trapped in local minima (depending on the algorithm initialization). The time complexity of the classical k -medoids technique poses furthermore problems with large datasets. More recent versions of the k -medoids algorithm, able to more effectively deal with these issues, have been proposed in the literature. The performance of the traditional k -medoids method is examined here alongside two of these recent algorithms to evaluate the possible improvement in this direction.

1) *Partitioning Around Medoid (PAM)*: PAM is the traditional method of the k -medoids family, developed in 1990 [12]. It consists of two phases named build and swap. The build phase chooses k times the point which leads to the smallest sum of object distances. The swap phase considers all possible changes to the set of selected medoids to improve the initialization step of the build phase. It includes replacing (swapping) medoid with non-medoid objects.

2) *A simple and fast algorithm for k -medoid (Fastkmed)*: Expressing that the PAM method faces difficulties when dealing with large datasets due to its time complexity, the Fastkmed algorithm is introduced in [13] as a simple and fast alternative to the PAM. This algorithm consists of three steps. In the first step, using its proposed formulation, the initial medoids are selected, which are the k most middle objects in the dataset. Then, the objects are assigned to their nearest initial medoids. In step 2, the initial medoids are updated by finding the objects, which minimize the total distance to other objects in each cluster. In step 3, objects are assigned to new selected medoids of step 2. If an improvement is found, we return to step 2, and repeat the same procedure as long as a reduction in the cluster distances is achieved.

3) *Ranked k -medoid (Rankkmed)*: Stating that the Fastkmed is sensitive to the initialization step, and that it can get trapped in local optimum points, [14] proposes a ranked k -medoids algorithm. Practically, a new function is introduced that ranks objects according to their similarities. Also, a hostility measure is used to evaluate how dissimilar are objects in the selected group. The Rankkmed algorithm in the initialization step, randomly selects k medoids from the dataset. Then, using the ranked objects, it chooses the group of the most similar objects to each selected medoids. Also, the hostility measure defines the most dissimilar object of the group that should be updated for constituting the new medoids.

B. Fuzzy Clustering (FANNY)

In fuzzy clustering, each observation is spread out over the various clusters. In other words, each object does not explicitly belong to one specific cluster, but to all clusters according to membership values expressing different levels of belonging. In the literature, fuzzy clustering is also called as soft clustering, in contrast to hard clustering (which links each object to one specific cluster). In the current paper, one particular implementation of fuzzy clustering, i.e., FANNY algorithm [12] is studied. It aims to minimize an objective function that includes the memberships of objects i and j to each cluster as well as the distance between i and j (d_{ij}) as follows [12].

$$\text{Min} \sum_{v=1}^k \frac{\sum_{i=1}^n \sum_{j=1}^n u_{iv}^r u_{jv}^r d_{ij}}{2 \sum_{j=1}^n u_{jv}^r} \quad (5)$$

where n , k and r denote the number of observations in the dataset, number of clusters, and the membership exponent, respectively. Also, u_{iv} (u_{jv}) is the membership of observation i (j) to cluster v . The memberships are nonnegative, and for each observation, they sum to 1. The membership values are the decision variables of above optimization problem. Once the problem is solved, the membership results can define the explicit assignment (hard clustering) assuming that each object belongs to the cluster with the greatest membership value.

C. Bottom-up clustering

The bottom-up clustering algorithm, or agglomerative hierarchical clustering, constructs a hierarchy of similar objects with the number of clusters ranging from one to the number of observations in the dataset. In this approach, at first, each observation is a small cluster by itself. Then, clusters are merged successively until only one large cluster remains, which contains all the objects. In this paper, Agglomerative Nesting (AGNES) algorithm presented in [12] is used for clustering process. The complete linkage method is applied, which means the largest dissimilarity between two objects placed in two different clusters is selected when merging the clusters.

VI. CLUSTER VALIDATION TECHNIQUES

In the literature, different cluster validation techniques are found that can be grouped into the following categories: internal validation, external validation, and visual assessment. Internal validation relies on the computation of indices, which quantify the quality of a partition without any information on the ideal solution. Reference [15] presents a comprehensive review of different internal validation indices. External validation indices aim at comparing the clustering results with some reference or external correct data, when it is available. In the cluster visualization, human intuition via visual assessment is utilized to judge the quality of partitioned data.

In this paper, we rely on internal validation measures to evaluate performance of the clustering results due to the following reasons. Firstly, the external validation technique is not possible since there is no available reference or previous correct clustering results for the sake of comparison. Visual assessment is neither an effective method in our case. This would require the generation and comparison of numerous 2D plots, representing the 2D projections of the N -dimensional FB polytopes (5-dimensional in our case) for each combination of clustering algorithm/distance measure/number of clusters, which will not be convenient for the analysis and interpretation.

A. Silhouette Width

The Silhouette index [12] is based on measures of the cluster compactness and separation. The compactness of object i ($\alpha(i)$) is computed as the average distance between object i and all the other objects in the same cluster. The separation of object i ($\beta(i)$) is evaluated based on the average distance between object i and all the objects placed in the nearest cluster (called neighbor cluster). The Silhouette width of object i ($s(i)$) is obtained by:

$$s(i) = \frac{\beta(i) - \alpha(i)}{\max\{\alpha(i), \beta(i)\}} \quad (6)$$

The Silhouette width lies in $[-1, 1]$, and should be maximized. The values smaller than zero express that the object

i is wrongly assigned to the selected cluster, and that it should have been placed in its neighbor cluster. In this work, the average of the Silhouette widths of all the objects is employed as a global index assessing the quality of the clustering solution.

B. Average Compactness to Average Separation Ratio

Separation of object i is measured in the Silhouette index with respect to its closest neighbor cluster only (i.e., Silhouette encodes the worst separation). We propose here to have an additional insight on the results by computing the ratio between the average intra-cluster distances (compactness) and the average inter-cluster distances (separation). The average distance between clusters I and J is equal to the average of the distances between every object of cluster I and all objects of cluster J . The final average inter-cluster distance is the average of all above distances between all pairs of I and J ($I \neq J$). Lower intra-cluster distance and larger separation are preferred; thus, smaller values of this index show an improved cluster quality.

VII. CASE STUDIES

In this section, performance of the proposed dissimilarity measures and investigated clustering algorithms is tested. The clustering analysis is carried out on 2760 hourly FB domains corresponding to the first 115 days of 2020, which relate to 5 CWE countries. The source of the FB data is the joint allocation office [16], where the FB data can be obtained with their hourly PTFD coefficients and RAM values. Given that the vertices of the polytopes representing the FB domains are needed for computing the distance measures presented earlier, and that data in [16] is provided under the form of linear equations, a vertex enumeration operation [17] is carried out to obtain these vertices. The quality of the clustering results is evaluated through the average Silhouette width (of all the objects) as well as the proposed index based on the average compactness to average separation ratio. Two cases are investigated as follows.

A. Investigation on the Proposed Dissimilarity Measures and Studied Clustering Techniques

In the first case, the objective is to evaluate the performance of the proposed dissimilarity measures and studied clustering algorithms for a predefined number of clusters. Table I presents the cluster results evaluated by the average Silhouette index when the 2760 FB domains are grouped into 5 clusters ($k=5$). Table II gives the results assessed by the index based on the average compactness to average separation ratio. The best result of each category is shown hereafter in bold for a better distinction.

TABLE I. CLUSTERING RESULTS EVALUATED BY THE AVERAGE SILHOUETTE INDEX ($k=5$)

| | <i>HD</i> | <i>MHD</i> | <i>MHD_SQE</i> | <i>VB</i> | <i>HD2V</i> |
|-----------|-------------|-------------|----------------|-------------|-------------|
| PAM | 0.03 | - 0.05 | 0.04 | 0.07 | 0.14 |
| Fastkmed | 0.046 | 0.006 | 0.143 | 0.05 | 0.1 |
| Rankkmed | 0.055 | 0.08 | 0.152 | 0.073 | 0.19 |
| Fuzzy | 0.06 | 0.16 | 0.26 | 0.09 | 0.36 |
| Bottom-up | 0.04 | 0.06 | 0.09 | -0.08 | 0.1 |

Considering the results given in Table I, it is seen that the fuzzy clustering leads to the highest average Silhouette values using all the studied dissimilarity measures. This is explained by the fact that the fuzzy method tries to minimize the sum of distances between all the objects, while in the k -medoids approaches, the objective is to minimize the object distances with respect to their medoids. The former formulation can offer a better cluster compactness (i.e., smaller cluster diameter).

Furthermore, in Table I, it can also be observed that the dissimilarity measure based on the ratio of Hausdorff distance to volume (HD2V) has a superior performance (when used in the clustering process) with respect to others, using all the studied clustering methods. This is achieved thanks to the improved distinction possibility of this metric. For the same reason, the MHD_SQE measure being based on the squared Euclidean distance leads to better results compared to the MHD (i.e., the one being used by the TSOs). In addition, it is seen that the VB measure (equation (3)) putting all dissimilarities between 0 and 1 cannot provide high-quality cluster results. Similarly, being based on the maximum of minimum distances, the HD metric does not provide a good distinction between the objects; consequently, it is not able to result in proper cluster results. In Table I, it is also noticed that using the recent versions of the k -medoids family, higher Silhouette values are obtained compared to the ones of the PAM method.

TABLE II. CLUSTERING RESULTS EVALUATED BY THE AVERAGE COMPACTNESS TO AVERAGE SEPARATION RATIO ($k=5$)

| | <i>HD</i> | <i>MHD</i> | <i>MHD_SQE</i> | <i>VB</i> | <i>HD2V</i> |
|-----------|--------------|-------------|----------------|--------------|--------------|
| PAM | 0.421 | 0.38 | 0.284 | 0.424 | 0.195 |
| Fastkmed | 0.455 | 0.375 | 0.351 | 0.445 | 0.353 |
| Rankkmed | 0.448 | 0.37 | 0.339 | 0.42 | 0.327 |
| Fuzzy | 0.399 | 0.28 | 0.212 | 0.415 | 0.163 |
| Bottom-up | 0.404 | 0.33 | 0.277 | 0.459 | 0.222 |

The results presented in Table II (evaluating cluster quality by the index based on average compactness to average separation ratio) are in line with the ones of Table I, and similar discussions can be applied to Table II as well regarding superior performances of the HD2V and the fuzzy clustering technique. It should be noted that smaller values in Table II show the improved cluster results (unlike the results of Table I evaluated by the Silhouette index).

B. Investigation on the Impact of Number of Clusters

The main goal of the second studied case is to investigate impact of changing the number of clusters (k) on the cluster validation indices. In this regard, the clustering analysis is performed using the introduced clustering algorithms when the number of clusters (k) is increased to 10, 15 and 20. Due to the space limit, the investigation of this part is only carried out on two dissimilarity measures, namely the MHD (used by the TSOs) and the HD2V proposed in this work. Table III presents the clustering results using the MHD measure, and Table IV gives the results obtained by the HD2V metric. In Table III and Table IV, Ind. 1 relates to the average Silhouette, and Ind. 2 represents the index based on the average compactness to average separation ratio.

TABLE III. CLUSTERING RESULTS WITH DIFFERENT NUMBERS OF CLUSTER USING THE MHD

| | $k=10$ | | $k=15$ | | $k=20$ | |
|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| | <i>Ind. 1</i> | <i>Ind. 2</i> | <i>Ind. 1</i> | <i>Ind. 2</i> | <i>Ind. 1</i> | <i>Ind. 2</i> |
| PAM | -0.05 | 0.324 | -0.09 | 0.315 | -0.1 | 0.3 |
| Fastkmed | 0.05 | 0.365 | 0.02 | 0.348 | 0.1 | 0.33 |
| Rankkmed | 0.06 | 0.35 | 0.05 | 0.3 | 0.1 | 0.29 |
| Fuzzy | 0.13 | 0.258 | 0.11 | 0.249 | 0.11 | 0.242 |
| Bottom-up | 0.01 | 0.311 | -0.01 | 0.295 | -0.01 | 0.28 |

As it can be seen in Table III and Table IV, when increasing the number of clusters (k), the cluster quality according to Ind. 1 is worsening (in most of the cases) while Ind. 2 shows the improvement of results. This is due to the fact that the Silhouette index measures the cluster separation based on the distance with respect to the nearest cluster while the other studied index (Ind. 2), evaluates the average cluster separations. Increasing the number of clusters generally improves the clusters compactness (homogeneity) but it can worsen the cluster separation with respect to its nearest cluster. The final choice of the proper number of clusters depends on the studied validation index, and the main purpose of the clustering analysis. In our application, the main objective of the clustering task is to gather the similar FB domains in one group such that the clustered FB domains (placed in that group) would lead to (relatively) similar adequacy results. In this regard, the cohesion (homogeneity) of the clustered data is the most important factor. The obtained results in this section show that increasing the number of clusters can improve the cluster cohesion (compactness).

TABLE IV. CLUSTERING RESULTS WITH DIFFERENT NUMBERS OF CLUSTER USING THE HD2V

| | $k=10$ | | $k=15$ | | $k=20$ | |
|-----------|---------------|---------------|---------------|---------------|---------------|---------------|
| | <i>Ind. 1</i> | <i>Ind. 2</i> | <i>Ind. 1</i> | <i>Ind. 2</i> | <i>Ind. 1</i> | <i>Ind. 2</i> |
| PAM | -0.01 | 0.175 | -0.03 | 0.154 | -0.06 | 0.145 |
| Fastkmed | 0.08 | 0.288 | 0.06 | 0.275 | 0.09 | 0.27 |
| Rankkmed | 0.16 | 0.246 | 0.1 | 0.184 | 0.08 | 0.178 |
| Fuzzy | 0.25 | 0.136 | 0.24 | 0.123 | 0.23 | 0.114 |
| Bottom-up | 0.05 | 0.191 | 0.01 | 0.173 | -0.02 | 0.15 |

In addition, comparing the indices reported in Table III with their counterparts given in Table IV, one can observe that using HD2V measure (Table IV) proposed in this paper, the quality of clustered data has been improved. Furthermore, in both Table III and Table IV, fuzzy clustering technique leads to the best performance among the studied clustering methods. These outcomes are in line with the results of the previous section.

Finally, taking into account the conducted analyses of this paper, it is concluded that the fuzzy clustering technique in combination with the dissimilarity metric based on the ratio of Hausdorff distance to volume provide the best results. This proposed solution considerably improves the clusters quality compared to application of the k -medoids method and the modified Hausdorff distance, traditionally used by the TSOs.

VIII. CONCLUSIONS

In order to incorporate the interconnection capacities through the FB domains in the adequacy assessment, the CWE TSOs rely on a methodology based on the clustering-correlation analysis of the FB domains. This paper proposes different distance measures, and studies various clustering techniques aiming at improving performance of the clustering phase. The obtained results confirm that the proposed metric based on the ratio of the Hausdorff distance to the volume can improve the quality of the clustered data. Furthermore, it is demonstrated that the fuzzy clustering method can lead to better results compared to the ones obtained by the medoids-based approaches. The combination of the fuzzy clustering and this proposed dissimilarity measure improves the average Silhouette value from 0.08 (obtained according to the methodology employed by the TSOs) to 0.36 (for $k=5$). The investigation also shows that by increasing the number of clusters, the average compactness to separation of the clusters is improved while the cluster separation with respect to its nearest neighbor can worsen. Finally, it is shown that the medoids-based approaches are sensitive to the initialization step, and their performance depends on the exploration capability of the selected algorithm. Although this paper focuses on advanced clustering of FB domains in context of an adequacy assessment, the highlighted suggestions and improvements of this work can be extended to other studies incorporating the FBMC.

REFERENCES

- [1] J-F. Toubeau, B. Bakhshideh Zad, M. Hupez, Z. De Grève, F. Vallée, "Deep reinforcement learning-based voltage control to deal with model uncertainties in distribution networks," *Energies*, vol. 13, no. 15, 2020.
- [2] J-F. Toubeau, J. Bottieau, F. Vallée, Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, 2018.
- [3] Identification of appropriate generation and system adequacy standards for the internal electricity market, European commission, 2016.
- [4] K. Van den Bergh, J. Boury, E. Delarue, "The flow-based market coupling in central western Europe: concepts and definitions," *The Electricity Journal*, vol. 29, no. 1, pp. 24-29, 2016.
- [5] Adequacy and flexibility study for Belgium 2020-2030, Elia, 2019.
- [6] Modelling of flow-based domains in Antares for adequacy studies, RTE, 2017.
- [7] Mid-term adequacy forecast, ENTSOE, 2019.
- [8] Typical flow-based days selection, RTE, 2017.
- [9] Bilan prévisionnel de l'équilibre offre-demande d'électricité en France, RTE, 2019.
- [10] D. P. Huttenlocher, G. A. Klanderman and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, 1993.
- [11] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469-483, 1996.
- [12] L. Kaufman, P. J. Rousseeuw, "Finding groups in data: an introduction to cluster analysis, vol. 39, Wiley Online Library, 1990.
- [13] HS. Park, C. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Systems with Applications*, vol. 36, pp. 3336-3341, 2009.
- [14] S. M. Razavi Zadegan, M. Mirzaei, F. Sadoughi, "Ranked k-medoids: a fast and accurate rank-based partitioning algorithm for clustering large datasets," *Knowledge-Based Systems*, vol. 39, pp. 133-143, 2013.
- [15] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243-256, 2013.
- [16] Joint allocation office, <https://www.jao.eu/>
- [17] D. Avis, "A revised implementation of the reverse search vertex enumeration algorithm, in Polytopes - Combinatorics and Computation," pp. 177-198, 2000.