Defining the Optimal Number of Demand Response Programs and Tariffs Using Clustering Methods

Cátia Silva, Pedro Faria, Zita Vale
GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development,
PP - Polytechnic of Porto
Porto, Portugal
cvcds@isep.ipp.pt; pnf@isep.ipp.pt; zav@isep.ipp.pt

***Abstract*— Nowadays, the data can be considered an asset when properly managed. An entity with the right tool to analyse the amount of data existent and withdraw crucial information will have the power to obliterate the competition. In the Energy sector, with Smart Grid introduction, small resources have more influence in the market through Demand Response and bidirectional communication. However, none of the actual business models is prepared to deal with the uncertainty related to these resources. The authors, in order to find a solution for this complex problem, proposed a methodology which the goal is to minimize operation costs and give fair compensation for resources who participate in the management of local markets. With this fair payment, it is expected continuous participation. Through clustering methods, remuneration groups are created. In the present paper, a study about the optimal number of clusters is performed. The information gives the Aggregator control in results of the following phases, understanding the impact in the remuneration of the resources.**

***Keywords*— Clustering; Aggregation; Consumers; Remuneration; Energy Market**

NOMENCLATURE

c	Number of consumers (c = 1,2, …, C)
C	Maximum number of consumers c
$C_{DG}(p)$	Distributed generation unit p cost
$C_{IDR}(c)$	Incentive based Demand Response cost for consumer c
C_{NSP}	Non-supplied power cost
$C_{Supa}(sa)$	Regular sa supplier cost
$C_{Supr}(sr)$	Additional sr supplier cost
p	Number of generation units (p = 1,2, …, P)
P	Maximum number of producers p
$P^{add}_{Suplier}(sa)$	Scheduled power for additional *sa* supplier
$P^{add\ Max}_{Suplier}(sa)$	Maximum power from an additional supplier
$P^{addTotal}_{Suplier}(sa)$	Maximum allowed total power from all the additional suppliers
$P_{DG}(p)$	Scheduled power for Distributed Generation unit *p*
$P_{DR}(c)$	Scheduled power reduction Demand Response program for consumer *c*
$P^{Initial}_{Load}(c)$	Initial consumption of the consumers
P^{Max}_{DG}	Maximum power schedule in a Distributed Generation resource
P_{NSP}	Non-supplied power
$P^{reg\ Max}_{Suplier}(sr)$	Maximum power from a regular supplier
$P^{reg\ Total}_{Suplier}(sr)$	Maximum allowed total power from all the regular suppliers
$P^{reg}_{Suplier}(sr)$	Scheduled power for a regular *sr* supplier
$P^{TotalMax}_{DG}$	Maximum allowed total power from all the Distributed Generation units
S	Maximum number of suppliers s
sa	Number of additional suppliers (sa = 1,2, …, Sa)
sr	Number of regular suppliers (sr = 1,2, …, Sr)

The present work was done and funded in the scope of the following projects: COLORS Project, CEECIND/02887/2017, and UID/EEA/00760/2019 funded by FEDER Funds through COMPETE program and by National Funds through FCT. Cátia Silva is supported by Fundação para a Ciência e a Tecnologia (FCT), grant SFRH/BD/144200/2019.

I. INTRODUCTION

With the evolution of technologies and communication means, the amount of data collected daily and from several sources becomes massive and hard to manage. In this way, it is crucial to find adequate tools for analysing such data and gather useful information – used as knowledge to different areas like business, science and engineering, medicine, among others [1].

Regarding Energy Sector, the introduction of Smart Grids definition origin a revolution in the management of the market, underlining the possibility of Demand Response (DR) and the increasing penetration of Distributed Generation (DG) resources, namely Renewable based. These concepts introduce several advantages and enable small resources in the market transactions increasing their interaction and knowledge, sometimes in real-time, thanks to advances in communication, namely bi-directional communication. Anyway, there is a necessity to assign an entity to manage these small resources and the Virtual Power Players (VPP) definition is presented, dealing with the local communities' market. Deal with these resources becomes a priority but nowadays the Business Models applied don't consider these concepts. Although, in several countries' efforts are being made, none of them as able to fully implement and be successful. The amount of data to manage associated with the uncertainty levels becomes a complex challenge to undertake and be a focus in several studies. Also, the compensation for participation in DR programs is a significant matter. Finding strategies to maintain the continuous contribution on the market transactions from small consumers part, reducing the uncertainty, the remuneration must be engaging, and the business models are not yet prepared to deal with this transformation

In the present paper, a methodology has been proposed to find an adequate Business Model to successful include small resources in the energy market, is the prosecution of previous works [2]–[4]. Counting 5 different phases – Profiling, Scheduling, Aggregation, Remuneration and Classification, the methodology is able to manage small consumers with DR programs, small DG units and also the combination of the two – prosumers. In fact, one important feature in this proposal is the flexibility to be used not only for planning (days, weeks, months) but also in an operation (several minutes) approach easing the Aggregator tasks. The main idea is to optimally schedule all the local community resources and assign to each one a remuneration group according to the amount of participation in the market transaction – for example, in the consumers' case, the amount of reduction from DR programs. With this approach, both parties have advantages: VPP minimize its operation costs with an optimal schedule and the resources are remunerate fairly when participating in the management of the local market. The focus of the present paper is the number of groups to be formed concerning the database and the situation. Analysing the optimal number of groups, performing several methods with this goal and for three different scenarios, studying the influence of the weekdays.

The present paper is divided into five sections. Section I introduces the theme and the principal focus. Section II details the proposed methodology. Section III presents the case study and the different scenarios to be studied. In Section IV the analysis from the results are presented. Lastly, Section V shows the principal conclusions.

II. Proposed Methodology

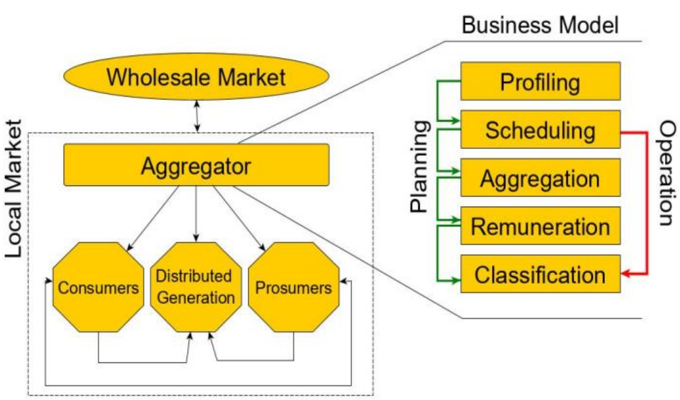


Fig. 1. Proposed Methodology

As can be seen in Fig. 1, the authors added a feature to the proposed methodology regarding previous works, giving enough flexibility to perform different time horizons: a couple of minutes (operation – red path) to weeks or even months (planning – green path), depending on the information used as input and previously analysed in the Profiling phase. In a planning approach, having more time, the first phase is done in a more detailed way, studying the resources and delineate their profile. For operation, only the needed inputs for Scheduling are considered. The flexibility feature from the proposed methodology not only includes diverse time horizons but also different sizes and types of resources: the VPP can aggregate small consumers that participate in DR programs, DG units and prosumers. Aggregators with only one type of resource will also be able to use the methodology.

Studying further each phase, Profiling considers all the information regarding the resources, e.g. the maximum capacity of the DG units, the reduction capacity of the consumers belonging to DR programs, as well as the consumption tariffs associated with each resource. The selected information is used as input for optimal Scheduling. The VPP goal optimizes the management of the local community and minimizes operation costs. From the point of view of the resources, when participating in the market transaction, be compensated with fair remuneration according to the actual amount of reduction. It is important to highlight that through the present paper the Scheduling phase may be referred Optimization phase. The objective function is given by Equation (1).

$$MinOF = \sum_{p=1}^{P} P_{DG(p)} C_{DG(p)} + \sum_{sr=1}^{Sr} P^{reg}_{Supplier(sr)} C_{Supplier(sr)} + \sum_{sa=1}^{Sa} P^{add}_{Supplier(sa)} C_{Supplier(sa)} + \sum_{c=1}^{C} P_{DR(c)} C_{DR(c)} + P_{NSP} C_{NSP} \qquad (1)$$

Due to the uncertainty associated with the small resources, the VPP must also consider suppliers to suppress the demand when the production from DG units is not enough to suppress the demand and maintain the balance between consumption and generation. Two types of suppliers were considered: regular and additional. Let p, sr, sa and c be the number of DG units, regular suppliers, additional suppliers and consumers, respectively. Equation (2) presents the balance constraint considered, where the sum of the possible reduction of each consumer through DR programs (PDR) to its initial load (PInitialLoad) should equal the sum of all DG units (PDG) and suppliers, both additional (PaddSupplier) and regular (PregSupplier).

$$\sum_{c=1}^{C} [P^{Initial}_{Load(c)} - P_{DR(c)}] = \sum_{p=1}^{P} P_{DG(p)} + \sum_{sr=1}^{Sr} P^{reg}_{Supplier(sr)} + \sum_{sa=1}^{Sa} P^{add}_{Supplier(sa)} + P_{NSP} \qquad (2)$$

Also, in Equation (2), PNSP is a variable that refers to the value of Non-Supplied Power (NSP) and where the goal for an optimum network flow it to keep this value as close to null as possible, proving that the grid is being well managed and none of the consumption was not supplied.

Regarding boundaries for the small consumers belonging to DR programs, it was considered Equation (3), where the maximum value of reduction for each consumer restrict the amount of reduction that can be considered.

$$P_{DR(c)} \le P_{DR(c)}^{Max}, \ \ \forall c \in \{1, \dots, C\} \tag{3}$$

DG units are limited by Equations (4)-(6). Considering Equation (4) and (5), the upper and lower bounds are applied to all the units. Also, it is considered a maximum contribution from these technologies, giving the Aggregation some control with Equation (6).

$$P_{DG(p)} \le P_{DG(p)}^{Max}, \ \ \forall p \in \{1, \dots, P\} \tag{4}$$

$$P_{DG(p)} \ge P_{DG(p)}^{Min}, \ \ \forall p \in \{1, \dots, P\} \tag{5}$$

$$\sum_{p=1}^{P} P_{DG(p)} \le P_{DG}^{TotalMax} \tag{6}$$

The technical limits for both suppliers are shown in Equations (7)-(10). Similar to DG units, the suppliers have an upper bound and are also limited regarding the total amount that each type of external suppliers can have available to this network: regular (Equation 7 and Equation 8) and additional (Equation 9 and Equation 10).

$$P_{Supplier(sr)}^{reg} \le P_{Supplier(sr)}^{regMAX}, \forall \in \{1, \dots, Sr\} \tag{7}$$

$$\sum_{sr=1}^{Sr} P_{Supplier(sr)}^{reg} \le P_{Supplier(sr)}^{regTOTAL} \tag{8}$$

$$P_{Supplier(sa)}^{add} \le P_{Supplier(sa)}^{addMAX}, \forall \in \{1, \dots, Sa\} \tag{9}$$

$$\sum_{sa=1}^{Sa} P_{Supplier(sa)}^{add} \le P_{Supplier(sa)}^{addTOTAL} \tag{10}$$

Concluding the Optimization, the output is collected and both Aggregation and Remuneration phase is introduced. The goal is to assign a remuneration group (and tariff) to each resource according to the participation in the management of the local market, making compensation fair. The resources with higher participation belong to the same group and, consequently, will be remunerated at the same tariff being this approach used as motivation the continuous participation. For performing the Aggregation phase, clustering methods where used, namely k-means. This is a partitioning method and it's defined by the search of a centroid value that represents each group. The centroid is found when the distance between this element and the remaining is minimal. Several functions can be applied, for example, the Euclidean distance. One of the inputs of this method is the number of k cluster to be formed. This parameter will affect the output and a range of k must be studied. In fact, the k value is important in the proposed methodology because will influence the final remuneration of the resources. In this way, the present paper aims to study and compare the optimal number of clusters (kopt) in a database and the ideal for VPP considering several scenarios. The database is organized in matrix form, where the columns correspond to periods and the rows to the result from the Optimization phase from each resource to be studied. Several methods can be used to find the optimal number of k clusters for a giving database [5]. In the present paper, the authors analysed two types of methods: direct and statistical testing. The first consists in to optimize a criterion, for example, the within-cluster sum of squares (wss) for Elbow Method and for Silhouette Method, the average silhouette, two of the most known methods.

To find kopt with Elbow Method, a range of k must be chosen, and a clustering method must be performed with a step of 1. For each k the total wss is calculated and a curve is drawn according to the k range: the location of the "elbow" marks the kopt. In other words, the k where the cost drastically reduces and then hits the plateau is considered the optimal [6]. It is important to mention that Elbow Method can be ambiguous considering only the plot curve, depending on the perspective of the analyst. In order to avoid this problem, the authors choose to compute the elbow strength for each k of the selected range. Depending on how strong the elbow signal is, the kopt is found maximizing the relative strength instead of the visualization of an "elbow". The feature is considered an advance when compared with previous works by the authors [7].

An alternative, in direct methods, is the Silhouette Method, proposed by Kaufman and Rousseeuw in 1990 [8]. The criterion for this method, average silhouette, determine the quality of the cluster – the k with higher average silhouette width is considered the optimal cluster in the selected range. The curve of all average silhouette can also be drawn, and the maximum value is kopt. Both direct methods presented can be applied to any clustering method.

Regarding Statistical Testing Method, the Gap statistic was selected and has been published by R. Tibshirani, G. Walther, and T. Hastie in 2001 [9]. This method performs a comparison between the total intra-cluster variation for a range of k with their expected values under null reference distribution of the data (no apparent grouping). The kopt is found when maximizing the gap statistic, meaning that the selected clustering structure is the farthest from the random uniform distribution of points. So, the k with the higher value of gap statistic is considered the optimal.

Returning to the proposed methodology, Classification phase is presented when the operation approach was considered by the authors. The results from planning approach are used to gather a training database in order to produce several rules. The main goal is to assign a remuneration group to each resource, considering only their performance in the Scheduling phase.

III. Case Study

In this section, the case study and the three scenarios are detailed. A 30 kV Real Portuguese distribution network fed by a substation of (60/30kV) with a maximum capacity of 90 MVA was chosen to prove the viability of the proposed methodology. With a total of 548 DG units, this network contains seven different types of technologies: small hydro, waste-to-energy, wind, photovoltaic, biomass, fuel cell and co-generation as showed in Table. 1. For each type, the tariff applied, and the number of units is presented.

Table. 1. Distributed Generation units' characterization

Type	Tariff [m.u./kWh]	# Units
Biomass	0.1206	25
Co-generation	0.0975	16
Fuel Cell	0.0945	13
Photovoltaic	0.2889	208
Small Hydro	0.0961	25
Waste-to-energy	0.0900	7
Wind	0.0988	254
	Total	548

Regarding consumers, it was considered that all of 20.310 participate in DR programs and, with this, are able to reduce some amount of consumption. Considering five different types of consumers, the principal characteristics are presented in Table. 2. The tariffs applied are divided according to a schedule, and there are three types of schedules used by Portuguese energy traders. With this, single is when the tariff is the same all day; double when the day is divided into off-valley periods and valley periods; triple when there are peak, off-peak and valley periods. Tariffs are different depending on the day of the week and season.

Table. 2. Consumers' characterization

Type of Consumer	Type of Schedule	Period	Tariff [m.u./kW]	# Units
Domestic	Single ≤ 2.3 kVA	-	0.1426	10,168
Small Commerce	Single ≥ 2.3 kVA	-	0.1652	9828
Medium Commerce	Double	Off-valley hours	0.1948	82
		Valley hours	0.1016	
Large Commerce	Triple	Peak hours	0.2253	85
		Off-valley	0.1765	
Industrial		Valley hours	0.1016	147
			Total	20,310

To a successful application of the proposed methodology, the aggregator is responsible for providing all the input characteristics for Scheduling phase. This is the goal of the Profiling phase, where all information is analysed with adequate data mining tools. The aggregation phase is the focus study of the present paper and the selected clustering method is k-means. The goal is, given a database with the results from consumers reduction in a chosen week, perform a comparison of three methods to determine the optimal number of clusters – kopt. In fact, three scenarios were also created to understand the influence of the weekdays in the study. So, Scenario 1 considers results for working days; Scenario 2 considers results for weekend and finally, Scenario 3 which consider results for the whole week. The week chosen for this study was divided into periods of 15 minutes. The size of the database of each scenario is presented in Table. 3.

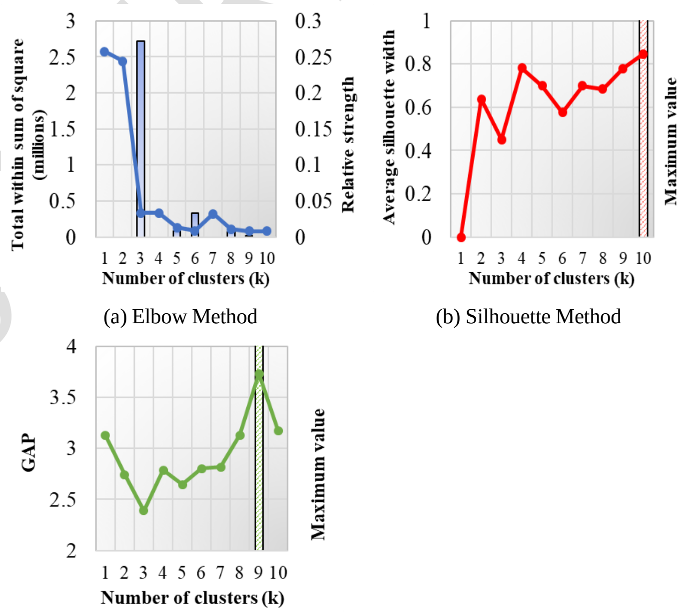
Table. 3. Scenarios Definition

Scenario	Definition	Weekdays	Database Size
1	Working Days	Monday to Friday	480
2	Weekend	Saturday to Sunday	192
3	Whole Week	Monday to Sunday	672

IV. Results

In the present section, the authors analyse the results for three different methods used to determine the optimal number of clusters. It is crucial to highlight the importance of the number of clusters in the proposed methodology. Once selected, the results of the following phases will be influenced: the remuneration tariffs and the final remuneration depend on the Aggregator choosing the most profitable solution for both point of views. In fact, giving a range of k clusters to be studied and finding the optimal will ease the Aggregator task. In this way, three scenarios were created and analysed for k=1 to k=10, being Scenario 1, Scenario 2 and Scenario 3 results presented by Fig. 2, Fig. 3 and Fig. 4, respectively.

In the first Scenario, for Elbow Method, k = 3 was considered the kopt. The total wss for the selected k was 336,968.18 and the relative strength was 0.27. Regarding Silhouette Method, the highest average silhouette width was achieved in the last k studied, kopt = 10, with 0.85. For the Gap Statistic Method, the k with the largest gap was found in k = 9, where the resulting value was 3.73.



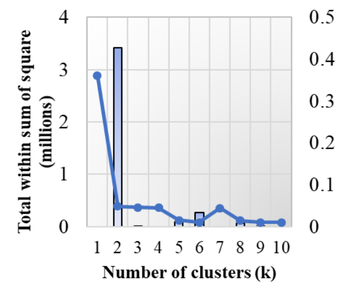
(a) Elbow Method

(b) Silhouette Method

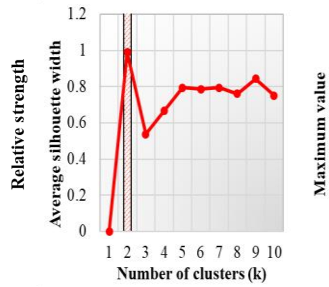
(c) Gap Statistic Method

Fig. 2. Methods comparison for Scenario 1: Week

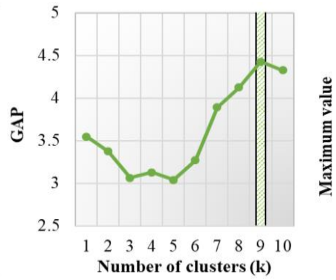
In Scenario 2, the Elbow Method considered k = 2 as the optimal cluster. The relative strength of this k was 0.43 and the total wss 365,156.6. The higher value of average, in the Silhouette Method, was also in k = 2 (0.99). For the third method studied, the largest gap was found again in k=9, this time with 4.43.



(a) Elbow Method



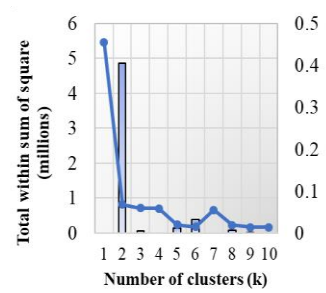
(b) Silhouette Method



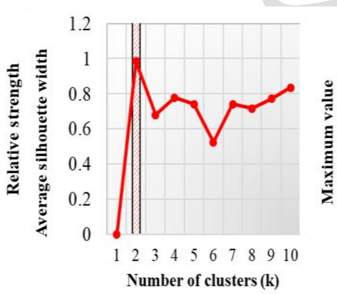
(c) Gap Statistic Method

Fig. 3. Methods comparison for Scenario 2: Weekend

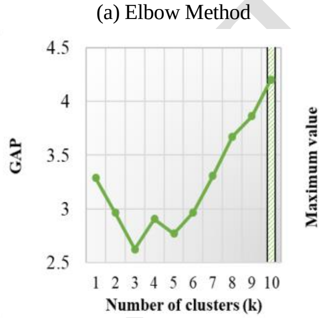
For the last scenario, the relative strength of the kopt = 2 was lower than the previous scenario being 0.41 although the total wss was superior to the double (819,598.40). Maintaining the kopt = 2, the result for Silhouette Method was once more 0.99. The maximum value for gap statistic was achieved, this time, in k = 10, with a value of 4.20.



(a) Elbow Method



(b) Silhouette Method



(c) Gap Statistic Method

Fig. 4. Methods comparison for Scenario 3: Whole Week

Direct methods had similar results in two of the scenarios but, those results were not approved by the Statistical testing method selected, where the outcomes were between 9 and 10 number k of clusters. Although the range was reduced, it was not possible to achieve a solution and it can be concluded that weekdays had influence. In this way, the final remuneration was compared for all kopt found by the selected methods, in order to understand which results in a higher benefit to VPP point of view without forgetting the fair remuneration for the associated resources in the local community. Table. 4 presents the results from the selected kopt for each scenario. Highlighting the fact that remuneration tariff for each group is the maximum tariff found in the mention.

Table. 4. Final Remuneration

Method	Scenario 1	Scenario 2	Scenario 3
Elbow	892,842.8	321,402.1	1,249,564.4
Silhouette	704,702.3	321,402.1	1,249,564.4
Gap Statistic	722,055.8	302.110.0	1,008,649.8

The values of remuneration are also influenced by the number of days in each scenario, being unable to compare between scenarios.

For Scenario 1, Elbow method opt for kopt = 3 resulting in 892,842.8 m.u. in total remuneration. Gap statistic resulting remuneration was 722,055.8 m.u. for a kopt of 9. The lowest value was the Silhouette Method with 704,702.3 m.u. where the kopt was 10. The results for Scenario 2 were 321,402.1 m.u. for Elbow and Silhouette method (both with kopt = 2) and 302.110.0 m.u. for Gap Statistic method, obtaining the lowest remuneration value. The last scenario had the highest values of remuneration (also the scenario where more days were considered). Again, the elbow and silhouette method had the same final remuneration with 1,249,564.4 m.u. and Gap Statistic the lower value with 1,008,649.8 m.u. In this way, it is possible to affirm, the method with higher kopt number had the lowest value of total of remuneration. Also, to prove the fair compensation, the values of an individual remuneration are presented (Table. 5), and the comparison proves that the method proposed had higher remuneration values, being an incentive to continuous participation.

Table. 5. Individual Remuneration

Scenario 1	Scenario 2	Scenario 3
651,633.7	284,870.0	936,503.7

V. Conclusions

In the present paper, a methodology to manage a local market with small resources was proposed. The goal is to ease the Aggregator task when handling the uncertainty created by these resources. The method proposed has five main phases and can be applied in both planning and operation approaches. The focus of the present paper was the Aggregation phase where remuneration groups are created. The goal for this phase is to created groups according to the actual participation of the resources in the management of the local market. Clustering methods were used, namely k-means. This method needs as input the number of clusters to be formed. Since it's vital information, influencing the following phases, the authors consider that comparing different method to find the optimal number of clusters for a given database was an important study in their investigation.

About the results, the authors deliberate that analysis between the optimal and the ideal number of clusters is important for the Aggregator. In fact, the optimal can give more benefits to VPP but may not consider the fair remuneration for the small resources, one of the main goals of the proposed methodology. Also, the results from the selected methods were not conclusive since the kopt was very different not reaching consensus. As future work, the authors intend to investigate further the cluster formation, cluster validation or even another clustering method, for the Aggregation Phase, crucial in the proposed method.

References

[1] S. Zhang, C. Zhang, and Q. Yang, “Data preparation for data mining,” *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 375–381, 2003.

[2] C. Silva, P. Faria, and Z. Vale, “Discussing Different Clustering Methods for the Aggregation of Demand Response and Distributed Generation,” in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2018.

[3] C. Silva, P. Faria, and Z. Vale, “Assessment of Distributed Generation Units Remuneration Using Different Clustering Methods for Aggregation,” *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*, Aalborg, Denmark, 2018.

[4] C. Silva, P. Faria, and Z. Vale, “Clustering Support for an Aggregator in a Smart Grid Context,” in *18th International Conference on Hybrid Intelligent Systems (HIS)*, 2018.

[5] B. Purnima and K. Arvind, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.

[6] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 336, no. 1.

[7] C. Silva, P. Faria, and Z. Vale, “Multi-Period Observation Clustering for Tariff Definition on a Weekly Basis Remuneration of Demand Response,” *Energies*, vol. 12, no. 7, p. 1248, 2019.

[8] L. Kaufman and P. J. Rousseeuw, *Finding groups in data : an introduction to cluster analysis*. Wiley-Interscience, 1990.

[9] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic.” 2001.