

# Clustering Support for an Aggregator in a Smart Grid Context

Cátia Silva, Pedro Faria and Zita Vale

GECAD – Research Group on Intelligent Engineering and Computing  
for Advanced Innovation and Development,  
Rua Dr. António Bernardino de Almeida,  
4200-072 Porto, Portugal

cvcds@isep.ipp.pt;pnf@isep.ipp.pt;zav@isep.ipp.pt

## Abstract.

The future of the industry foresees the automation and allocation of more intelligence to processes. A revolution in relation to the present. With this, new challenges and consequently more complexity is added to the management of the sectors. In the electric sector is introduced the theme of the Smart grids and so all the concepts aggregated with it. The possibility of the existence of demand response programs and the expansion of the distributed generation units for small players are key concepts and with enormous influence in the management of the markets belonging to this sector. Thus, a method is proposed that would help manage these resources through their aggregation, opening a new port for business models based on this idea. The benefit will be to take advantage of a more effective and efficient way the energy potential present in each group that is formed. Thus, in this paper will be explored the potential of clustering methods for the aggregation of resources.

**Keywords:** Clustering; Smart Grid; Aggregation; Business Model

## 1 Introduction

One of main drivers of the revolution that the world industry is going through is environmental concern. With the increase in the campaign to reduce the emissions of harmful gases, the solution will have to go through resources and processes that are more environmentally friendly, [1]. With regard to the electric sector, the expansion and promotion of the idea of endogenous renewable energies makes the concept of generation distributed one of the variables that will certainly have to be included in the new market models, [2]. In addition, the concept of Smart grids opens the possibility for consumers to have more access to information about their energy. It can be pointed out that they have access to the real-time price or even the possibility of selling to the energy market produced through the micro-production that they possess.

However, all these new hypotheses are variables of complication for the business models in the electricity markets, [3]. From now on, demand response programs will have to be considered and, given that the wholesale market is defined by competition between several companies, as efficiently as the solution found that includes these variables, the better will be the results. However, the uncertainty that these variables introduce, for example the microgeneration, may hinder this task, [4].

Regarding the infrastructure of the existing distribution network, it is not yet prepared for the future that lies ahead despite efforts. It will still be necessary to go a long way to allow conventional networks to be able to feed their loads in a controlled and smarter way and, in addition, to enable the addition of the elements distributed, [5].

However, there is a prediction of multiple advantages in the promotion and implementation of these concepts. One can speak for example in reducing the losses in the transport of energy in the distribution network. With the possibility of the energy being consumed immediately by consumers closest to the production sites, the distance that it travels will be lower, thus reducing losses. In this way there is an enormous advantage in the promotion of distributed generation and an incentive for producers and consumers to increase their participation in demand response programs.

Regarding consumers, this new chance of realizing in real time the price of energy in the market could change their behavior. Through signs and changes, both reductions and increases, the end-user can review their consumption. Thus, virtual power players will need to understand the changes in the load diagram and work to achieve the goal of reducing the load peak. Efficient management will be crucial to achieving all the goals, [6].

In this way the methodology proposed in [7], used in this paper, allows the virtual power player to aggregate small resources, including consumers participating in demand response programs.

Different works, like the authors of this paper, try to fill this gap in the actual business models, including in their studies the concepts introduced by Smart Grids, like in [8] - [11] and many more. For example in [8] was developed a solution through a hierarchical multi-agent system with 3 levels to control renewable energy resources and demand response units, including advanced inverters for PVs. One of those layers makes the schedules for flexible energy sources, based on its optimization criteria, either maximizing the profit or minimizing energy differences between local production and consumption. The results from scheduling depends on the system rules and boundary conditions, which include technical, economic, market, and regulatory provisions.

The method proposed in this paper will be presented in more detail in the following section and it has been designed in such a way as to allocate the fairer remuneration for each group of aggregate resources, adding a different perspective to the actual business models. Different numbers of groups and aggregation methods are tested in this paper through a tool that was designed to understand what will be most beneficial in prior decision making of the optimum number that enables the minimization of operating costs by the aggregator. In this way, the results from optimization phase are used as input for this tool. This paper will present the tool and results from one chosen k cluster. The main objective is to show the viability of it and how can be useful for the aggregator.

## 2 Approach

Through the methodology proposed in [7] it will be possible to understand the role of the aggregator in the network infrastructure and, thus, a way of dealing with the challenges that the new paradigm is imposing. Fig. 1 presents each of the phases that make up the methodology and its logical sequence of events.

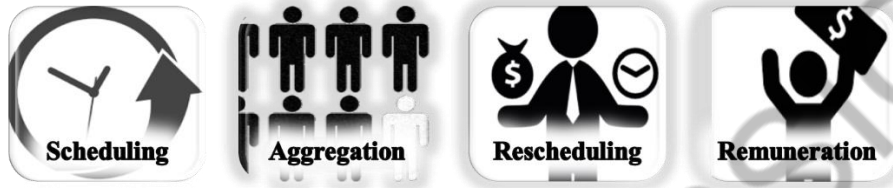


Fig. 1. Proposed Methodology

At an early stage, an optimal scheduling of all resources associated with the aggregator will be carried out, namely, distributed generation units (DG), consumers who belong to the demand response and supplier programs and suppliers. The last one may be used if DG units do not meet the needs of the consumer. Profiles used as input data for this optimization include, for example, the elasticity of the demand price, the levels of comfort sought by each of them, the possibilities of direct control of load and the existence of production of either heat or electricity, ... Price and operation constraints were considered in this optimization, also operational constraints imposed by virtual power player to achieve its goals. The MATLAB software potentials were used through the toolbox, TOMLAB, for the optimization.

The second phase represents the aggregation of resources. Through a method of clustering are defined groups considering the results obtained in the first phase – schedule power for each resource. By grouping these small resources, VPP will have a considerable amount of energy for the negotiations in the market and allows the expansion of their participation in DR programs. It will be an innovative thing since so far, the participation of such resources would only be made in an indirect way. In [12] classification methods were studied using this methodology. In this paper and to test the potential of the clustering methods, four were compared for several  $k$  – number of clusters. The methods and the case study will be presented in the following sections more explicitly. This part of the study was carried out using software R and Shiny package.

The last phase, the remuneration stage, is used for continued collaboration of all resources with the aggregator in the network operation, it is used as compensation. Here, taking into account the tariffs that were designed specifically for each group, through the maximum of each one, all resource will be paid at the end of the schedule.

### 3 Case Study

The case of study presented to test the feasibility of the proposed method is composed of a real distribution network composed of about 548 distributed producers and 20310 end-user consumers. To manage the consumers from demand response programs, two main programs were considered. The first one is called Real Time Pricing (RTP), based on price, where consumers change their load by responding to changes in the price of electricity in real time. The second one is based on incentives, in which consumers are paid at a fixed price per kW of reduced load. Incentive based programs (Reduce, Cut) and RTP were applied in this study to distinct types of consumers.

Regarding the distributed generation, this case was studied Wind, Biomass, Small hydro, co-generation, Photovoltaic, Fuel cell and Waste-to-energy. Table 1 presents the detailed information of these units, showing the unit number by type, the unit operating price in m.u./kWh and the total available capacity.

**Table 1.** Distributed Generation Characterization

Designation	N° of units	Capacity (kWh)	Price (m.u./kWh)
Wind	254	5 866.09	0.071
Co-generation	16	6 910.10	0.00106
Waste-to-energy	7	53.10	0.056
Photovoltaic	208	7 061.28	0.150
Biomass	25	2 826.58	0.086
Fuel cell	13	2 457.60	0.098
Small hydro	25	214.05	0.042
Total DG	548	25 388.79 kWh	

Regarding the Table 2. shows the characterization for the types of consumers that participate in demand response programs: Domestic (DM), Small Commerce (SM), Medium Commerce (MC), Large Commerce (LC) and industrial (ID). It also shows the different possibilities of participation.

**Table 2.** Demand Response Consumers Characterization

Designation	Reduce	Cut	RTP	Initial price (m.u./kWh)
Domestic (DM)	•			0.12 (0.20)
Small commerce (SM)	•			0.18 (0.16)
Medium commerce (MC)		•		0.2 (0.20)
Large commerce (LC)		•		0.19 (0.20)
Industrial (ID)			•	0.15 (0.53)
Total N° of DR	19 996	167	147	20 310
Total Capacity (kWh)	8 676	1 106	11 571	21 354.36

The main goal is to aggregate the resources considering the results of the optimization performed in the first phase. In this paper, the authors choose to study the behavior of the tool considering only consumers. So, for testing, in this paper will only be exposed part of the aggregated resources (DG and DR units) and will not be considered the ones that got a null result in the optimization phase. Consequently, in the chosen scenario, only two types of consumers who belong to programs of DR were tested. The incentive-based program were studied, resulting in consumers belonging only to Small Commerce and Medium Commerce.

## 4 Demonstration

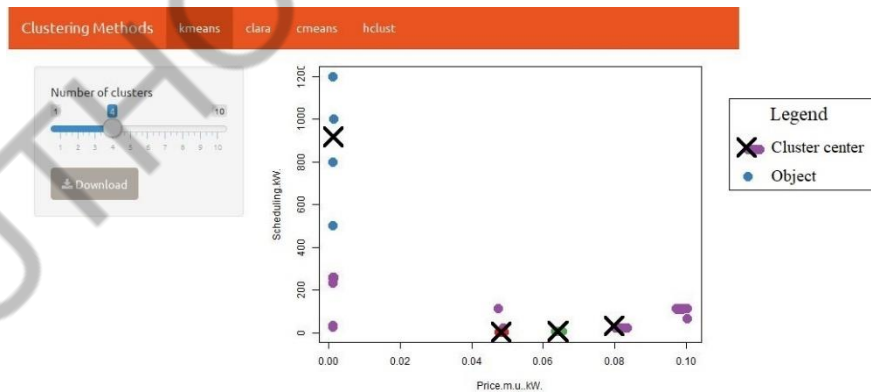
Through Fig. 2 will be able to perceive the tool interface built for comparing clustering methods.

This interactive tool has been developed by giving use to the software R through the Shiny package. The main intention of this tool will be to facilitate and assist in the task of the Analyst in this type of business model. Thus, a clearer vision and analysis will be possible for the decision in the selection of the method to be used as the problem it faces.

One of the advantages goes through help in the decision of the optimal number of clusters, since it is usually a task of the analyst, through the visualization of the database graphically. As the number of K cluster changes, it also varies the group to which each point belongs and the center of the cluster. There is the possibility of variation between 1 and 10 clusters.

Thus, four clustering methods were selected to be used and compared in this tool: two methods of partitional clustering – k-means and CLARA, one method of fuzzy clustering – c-means and one method of hierarchical clustering.

After finding the appropriated k cluster, this tool also the possibility to download the results to an external Excel file to be used later.



**Fig. 2.** Clustering Methods: Tool to compare different clustering methods

Firstly, K-means is presented as it is one of the most common unsupervised machine learning methods. The variation of this method used in this tool is the one presented by Hartigan-Wong and its used by default in R software, [13]. The algorithm is defined as the total variation within a cluster is then taken as the sum of the squares of Euclidean distance between a point and the center of the cluster, and then assigns the point to the nearest cluster. In this paper this method was performed only one time to obtain results. Each cluster is represented by a new center, the centroid, that is represented by a cross in Fig.2, which corresponds to the average of the points assigned to the  $k$  cluster in question. This method, however, presents a problem that can be considered serious for some cases since it cannot handle noise and outliers.

CLARA (Clustering Large Applications) is an extension of Partitioning Around Medoids (PAM), for a larger database. The algorithm is defined by the iterative search of objects that can represent a cluster, medoids, also represented by a cross. The exchange between a medoid and a non-medoid is validated only when there is an improvement in the criterion of the objective function - the minimization of the sum of the dissimilarities of all objects relative to the nearest medoid, [14]. As already mentioned, the advantage of this method is that it can deal with larger sets than PAM. However, the efficient performance of CLARA depends upon the size of the data base and a biased sample data may lead to poor results, [15].

Regarding C-means, this method is considered as a soft clustering that, unlike realizing whether or not an object belongs to a cluster, assigns a likeness of this point to the assigned cluster, [16]. To perform this clustering method in R software was chosen the algorithm *cmeans*. The parameter  $m$  is the degree of fuzzification and only support real values greater than 1. The bigger it is, the more fuzzy the membership values of the clustered data points are. In this paper,  $m = 20$ . Similar to k-means, this type of fuzzy clustering method can also represent the centers of each cluster, again presented with a cross

Finally, hierarchical clustering (hclust), as the name implies, structures the database in a hierarchical way, according to a proximity matrix. This clustering method can be divided into hierarchical agglomeration algorithms and divisive clustering. In R, an agglomeration method was used and was chosen *complete* linkage for this paper. In this way, each object is assigned to cluster and then, iteratively, join the two most similar clusters, continuing until there is just a single cluster with all the objects. At each iteration, distances between clusters are recomputed by the Lance-Williams dissimilarity update formula. It will not be possible to decide a priori a number of the clusters. So, the analyst according to the division that finds most interesting for the situation, can split the resulting dendrogram, [17].

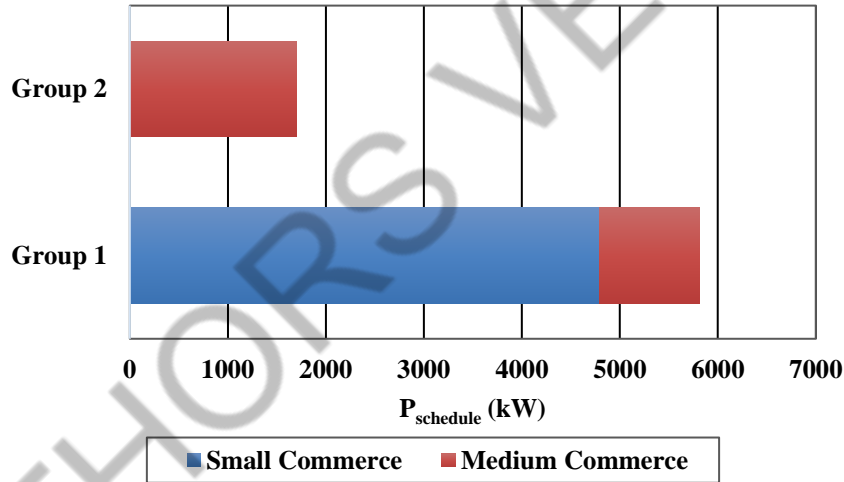
As previously said, the usefulness and feasibility of this tool was tested in the second phase of the methodology where it is proposed to compare several clustering methods for the aggregation of the various resources associated with the aggregator, specifically in this paper, small and medium commerce type of consumers. Initially, a database with the results obtained in MATLAB of the scheduling proposed in phase one, and with the tool, all the resources were grouped according to the number  $k$  of clusters.

## 5 Results

In this paper and through the following figures, the authors choose to expose the results of the aggregation for only two types of consumers associated with the programs of DR: Small Commerce and Medium commerce. These consumers represent the highest percentage in the data base so, a total of 9910 consumers were analyzed. The figures represent the total power according to each type of consumer for each of the groups formed.

Using the results obtained in the clustering tool, the four methods group these types of consumer according to the results obtained in the first phase of the proposed methodology. Thus, the inputs for this tool were the scheduled power of each of the resources. In this paper, the authors choose to only study one of the  $k$  clusters for the different methods chosen, to compare them in the same base. Since there are two types of consumers to be studied, the results for  $k = 2$  clusters are presented through this section to understand how these methods behave.

Fig. 3 refers to the K-means method. This method included all small commerce consumers in Group 1. Regarding the consumers of medium commerce, only one minority was included in the group 1. The remaining composes the group 2.



**Fig. 3.** K-means: results for  $k=2$

Fig. 4 shows the results for the C-means method. Group 1 is composed, in majority, by small commerce elements, presenting only a minimum percentage, 4 elements, of medium commerce. The group 2 is constituted in its entirety by elements of medium commerce.

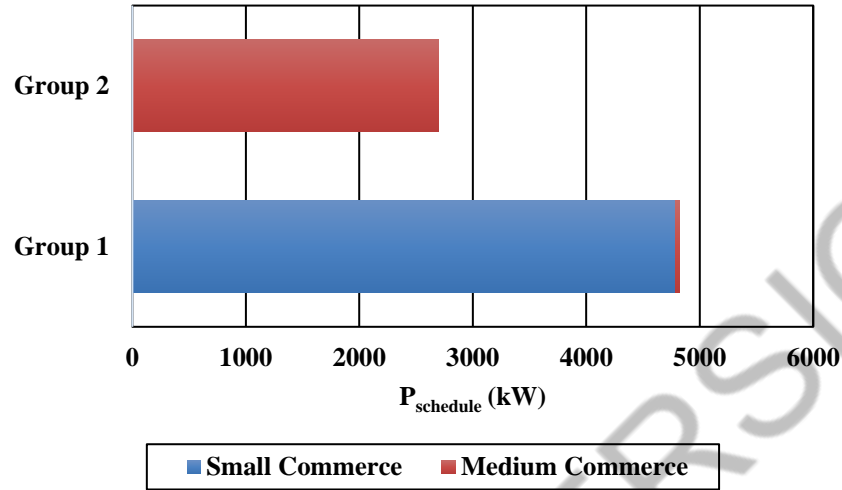


Fig. 4. C-means: results for k=2

The results obtained through the hierarchical method are shown in Fig. 5. As it is possible to repair, group 1 is a major part of the small consumers. Group 2 is only made up of 3 elements that add up to each other about 343.94 kW. In this paper, since the purpose was to compare k=2 clusters for all the methods, the result dendrogram is not analyzed.

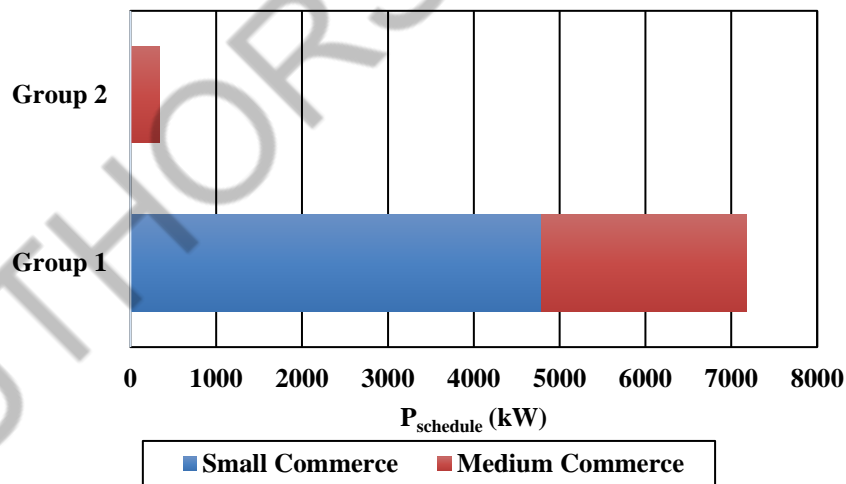


Fig. 5. Hclust: results for k=2

About the CLARA method, Fig. 6 shows its results to K = 2. Group 1 is again constituted mostly by Small Commerce. The medium commerce represents only

0.006% of group 1. In relation to the other group, the remaining 78 elements of medium commerce are what constitute it.

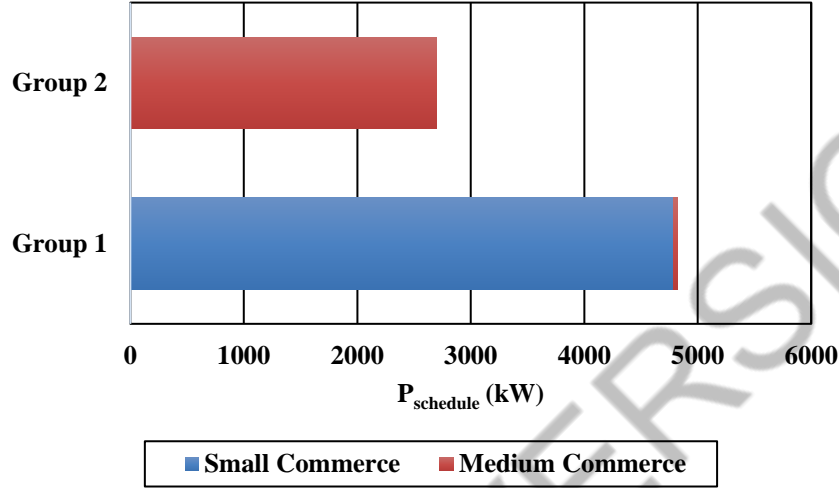


Fig. 6. CLARA: results for k=2

Now, comparing and analyzing the four methods presented, we can conclude that CLARA and c-means almost divided the data base by type of consumer, attributing group 1 to small commerce and group 2 to medium commerce. Regarding k-means and hclust, both assign a lot of members of medium commerce to group 1. One way to provide better results on hclust may be changing the distance metric and the linkage criteria. Regarding k-means, the initial seeds have a strong impact in the final result, in this way, changing and add some randomness to start may result in different outcomes, [15].

## 6 Conclusions

A radical change approaches all industry. Intelligent and automatic processes are required for the future. The structuring of the electricity sector is now changing to achieve a full introduction of the smart grid context. The market is not an exception, and players of the market will have to create new business models that include all the variables. Thus, the suggestion of aggregation of small resources and the idea of using their potential is suggested as a viable solution in a methodology proposed. The goal for the tool presented is to help decision making made by the virtual power player, as aggregator. By aggregating similar resources, the resulting groups will perform a fairer remuneration. The tool presented in this paper can simulates several situations, e.g., the number of clusters in which the database can be divided and compare different methods in order to decide the optimal solution.

## 7 Acknowledgments

The present work was done and funded in the scope of the following projects: CONTEST Project (P2020 - 23575), and UID/EEA/00760/2013 funded by FEDER Funds through COMPETE program and by National Funds through FCT.

## 8 References

- [1] D. Alahakoon and X. Yu, "Smart Electricity Meter Data Intelligence for Future Energy Systems: A Survey," *IEEE Trans. Ind. Informatics*, vol. 12, no. 1, pp. 425–436, 2016.
- [2] P. Faria, Z. Vale, and J. Baptista, "Demand Response Programs Design and Use Considering Intensive Penetration of Distributed Generation," *Energies*, vol. 8, no. 6, pp. 6230–6246, Jun. 2015.
- [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. Clastres, "Smart grids: Another step towards competition, energy security and climate change objectives," *Energy Policy*, vol. 39, no. 9, pp. 5399–5408, Sep. 2011.
- [5] R. Deng, Z. Yang, M.-Y. Chow, and J. Chen, "A Survey on Demand Response in Smart Grids: Mathematical Models and Approaches," *IEEE Trans. Ind. Informatics*, vol. 11, no. 3, pp. 1–1, 2015.
- [6] S. Lin, F. Li, E. Tian, Y. Fu, and D. Li, "Clustering Load Profiles for Demand Response Applications," *IEEE Trans. Smart Grid*, vol. 3053, no. c, pp. 1–9, 2017.
- [7] P. Faria, J. Spínola, and Z. Vale, "Aggregation and Remuneration of Electricity Consumers and Producers for the Definition of Demand-Response Programs," *IEEE Trans. Ind. Informatics*, vol. 12, no. 3, pp. 952–961, 2016.
- [8] J. Zupancic *et al.*, "Market-based business model for flexible energy aggregators in distribution networks," *Int. Conf. Eur. Energy Mark. EEM*, 2017.
- [9] R. Rezaian and W. Prügler, "Business models for the integration of electric vehicles into the Austrian energy system," *9th Int. Conf. Eur. Energy Mark. EEM 12*, pp. 1–8, 2012.
- [10] I. Ilieva and J. Rajasekharan, "Energy storage as a trigger for business model innovation in the energy sector," *2018 IEEE Int. Energy Conf. ENERGYCON 2018*, pp. 1–6, 2018.
- [11] Q. Zhang, J. Zhang, X. Han, X. Jin, H. Fu, and T. Zang, "Business model status of distributed power supply and its concerns in future development," in *2017 IEEE Conference on Energy Internet and Energy System Integration, EI2 2017 - Proceedings*, 2018, vol. 2018–Janua, pp. 1–5.
- [12] C. Silva, P. Faria, and Z. Vale, "Classification Approaches to Foster the Use of Distributed Generation with Improved Remuneration," 2013.
- [13] L. Morissette and S. Chartier, "The k-means clustering technique: General considerations and implementation in Mathematica," *Tutor. Quant. Methods Psychol.*, vol. 9, no. 1, pp. 15–24, 2013.
- [14] A. Kassambara, *Practical Guide to Cluster Analysis in R*, 1st ed. .
- [15] S. J. Saket and S. Pandya, "An Overview of Partitioning Algorithms in Clustering Techniques," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 5, no. 6, 2016.
- [16] D. Jyoti Bora and A. Kumar Gupta, "A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm," *Int. J. Comput. Trends Technol.*, vol. 10, no. 2, 2014.
- [17] P. Wirasanti, E. Ortjohann, A. Schmelter, and D. Morton, "Clustering power systems strategy the future of distributed generation," *Int. Symp. Power Electron. Power Electron. Electr. Drives, Autom. Motion*, pp. 679–683, 2012.