

## Article

# Local Renewable Energy Communities: Classification and Sizing

Bruno Canizes <sup>1,\*</sup> , João Costa <sup>2</sup>, Diego Bairrão <sup>2</sup> and Zita Vale <sup>1</sup> 

<sup>1</sup> GECAD Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, LASI—Intelligent Systems Associate Laboratory, Polytechnic of Porto, R. Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal

<sup>2</sup> School of Engineering, Polytechnic of Porto, R. Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal

\* Correspondence: bmc@isep.ipp.pt

**Abstract:** The transition from the current energy architecture to a new model is evident and inevitable. The coming future promises innovative and increasingly rigorous projects and challenges for everyone involved in this value chain. Technological developments have allowed the emergence of new concepts, such as renewable energy communities, decentralized renewable energy production, and even energy storage. These factors have incited consumers to play a more active role in the electricity sector and contribute considerably to the achievement of environmental objectives. With the introduction of renewable energy communities, the need to develop new management and optimization tools, mainly in generation and load management, arises. Thus, this paper proposes a platform capable of clustering consumers and prosumers according to their energy and geographical characteristics to create renewable energy communities. Thus, this paper proposes a platform capable of clustering consumers and prosumers according to their energy and geographical characteristics to create renewable energy communities. Moreover, through this platform, the identification (homogeneous energy communities, mixed energy communities, and self-sufficient energy communities) and the size of each community are also obtained. Three algorithms are considered to achieve this purpose: K-means, density-based spatial clustering of applications with noise, and linkage algorithms (single-link, complete-link, average-link, and Wards' method). With this work, it is possible to verify each algorithm's behavior and effectiveness in clustering the players into communities. A total of 233 members from 9 cities in the northern region of Portugal (Porto District) were considered to demonstrate the application of the proposed platform. The results demonstrate that the linkage algorithms presented the best classification performance, achieving 0.631 by complete-link in the Silhouette score, 2124.174 by Ward's method in the Calinski-Harabasz index, and 0.329 by single-link on the Davies-Bouldin index. Additionally, the developed platform demonstrated adequacy, versatility, and robustness concerning the classification and sizing of renewable energy communities.

**Keywords:** classification and sizing; clustering algorithms; clustering evaluation metrics; decentralized renewable energy generation; renewable energy communities



**Citation:** Canizes, B.; Costa, J.; Bairrão, D.; Vale, Z. Local Renewable Energy Communities: Classification and Sizing. *Energies* **2023**, *16*, 2389. <https://doi.org/10.3390/en16052389>

Academic Editor: Marco Merlo

Received: 29 December 2022

Revised: 25 February 2023

Accepted: 27 February 2023

Published: 2 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Renewable and distributed energy sources, such as wind and solar sources, dispersed along the distribution grid are gaining great importance and contribute to policy and environmental objectives. However, the variability and intermittent nature of these energy sources pose new challenges to power grid management and planning. Currently, the global energy demand is increasing, and the transition to electric vehicles is growing every year. Consequently, energy consumption in the power grid is expected to increase considerably, mostly powered by non-renewable energy sources. Thus, there is an inevitable need to minimize the subsequent carbon footprint by implementing large-scale renewable energy generation and energy storage systems [1–3].

The desire to increase the use of renewable sources and achieve carbon neutrality resulted in the development of renewable energy communities (RECs). RECs, in recent years, have been the subject of much interest both at the industrial and research levels since they are the fundamental pillars for building a smart energy system [4,5]. The expectations are high since it could bring solutions to problems we are currently facing, such as high energy demand and load management [6–10]. Moreover, some cooperative models between distributed RECs have been investigated in several applications, such as:

- Optimizing mechanisms against power loss by replacing central power management units with smart devices [11];
- Load management by exchanging and sharing locally generated energy [12–16].

Members of an REC may include public or private companies, domestic consumers, businesses, municipalities, and industries, among others. It should be managed by its participants and be located near renewable energy projects [17,18]. The possible integration of several types of consumers in RECs allows participants to be included in the energy sector chain, becoming, in this way, more active in what contributes to electricity generation [19]. Moreover, with RECs, a reduction in the power flowing in the transmission and distribution lines can be achieved, leading to a reduction in congestion and power loss [20]. RECs will allow greater flexibility in using locally generated energy, making it possible to control energy sharing and load balancing. As a result, the net energy resulting from these communities is nothing more than the difference between the energy generated and the energy consumed during a certain period [21,22].

The problems related to the identification and sizing of RECs should consider not only the spatial distances of the members in the electricity grid but also the net energy of each one over a time period [9]. An REC can be divided into three types according to their net energies:

- Homogeneous energy communities (HEC);
- Mixed energy communities (MEC);
- Self-sufficient energy communities (SEC).

The net energy value is the difference between energy generation and consumption in a considered time period. If the difference is negative, it means that net energy is negative, i.e., consumption is greater than the generation. On the other hand, if the difference is positive, the net energy is positive, meaning that consumption is lower than the generation.

An HEC is defined as a set of members whose net energies are exclusively positive or negative over a time period. The RECs with a set of members whose net energies are mixed with positive and negative are known as an MEC. In this REC, the members can share their locally generated energy with other adjacent members, avoiding resorting to the power grid and ensuring better resilience and reliability of the power supply. Finally, the SEC corresponds to a set of members with total positive net energy (the individual net energy of each member is irrelevant).

In this context, developing tools that can analyze and group these vast amounts of data is essential. Clustering is one of the techniques capable of analyzing and grouping huge amounts of data into small groups based on their characteristics. This technique is defined through the similarity function, which proceeds to aggregate data with similarities and divides them into groups or clusters. Clustering can be performed in several ways as many categories of clustering algorithms exist. However, it is important to note that each category has its strengths and weaknesses, so specific algorithms may be better suited to an input dataset than others.

### 1.1. Related Work

To the best of the authors' knowledge, there is a lack of work concerning the identification and clustering of RECs in the specialized literature. Reference [23] proposes an optimization model that simulates an electric energy market involving prosumers and electric vehicles. Furthermore, it presents an energy community with several categories of prosumers: household, commercial, and industrial. These prosumers were equipped

with photovoltaic panels and battery storage systems. Faia et al. [24] propose an LEM model that considers an energy community with a large concentration of electric vehicles and who were capable of facilitating prosumer-to-vehicle (P2V) transactions. Every energy community participant can sell and buy electricity from the community's retailer as well as other members. A mixed-integer linear programming (MILP) formulation is used to model the problem, and it is solved using a decentralized and iterative procedure. A local market mechanism is proposed in [25] in which end-users (consumers, small producers, and prosumers) exchange energy amongst themselves. However, due to the possible low market liquidity, the method implies that end-users satisfy their energy needs through bilateral contracts with an aggregator/retailer with wholesale market access. Reference [26] introduces a methodology based on demand response (DR) participation in a citizen energy community environment, employing unsupervised learning methods, including convolutional neural networks and K-means. Using end-user flexibility, it can assess future grid events and balance consumption and generation by analyzing the grid's future occurrences. Invitations to the DR event were extended to end-users based on their position and evaluated by three metrics. These metrics were energy flexibility, participation rate, and the end-users past flexibility. Hong et al. [27], propose an approach for REC identification and clustering by considering a power grid where the generation and energy sharing are predominate. However, there were limitations regarding the range of clustering algorithms studied and the lack of analysis and comparison between them to determine which ones are more effective in clustering. A summary of related work in the literature is presented in Table 1.

**Table 1.** Summary of related work.

Ref.	Topic	Technique/Method	Data Sources	Main Findings	Limitations
[23]	Optimization model for an electric energy market with prosumers and electric vehicles (EV).	Mixed-Integer Linear Programming	Hypothetical energy community with 15 prosumers and 20 EVs.	Energy costs reduction in energy communities that include prosumers and EVs; increase the efficiency of the energy flow.	Based on several assumptions, such as the availability of real-time energy data, which may not always be practical or feasible; assumes that all prosumers and EVs will participate in the local electricity market, which may not be the case in real-world scenarios; small-scale energy community and may not be generalizable to larger communities or different geographical regions.
[24]	A decentralized iterative approach for implementing local electricity markets (LEM) to manage the charging and discharging of electric vehicles.	Mixed-Integer Linear Programming	A generated dataset of EV driving patterns based on real-world data from a Portuguese mobility survey; realistic-world electricity market prices data.	Manage the EVs charging and discharging in a LEM, considering factors such as EV driving patterns, battery degradation, and market price fluctuations; economic efficiency of the LEM by enabling price signals to incentivize EV owners to charge or discharge; improve the environmental performance of the LEM by reducing the need for peak power generation and increasing the use of renewable energy sources.	The study focuses on a small-scale LEM with a limited number of EVs; the authors made several assumptions to simplify the modeling of the system; unclear how the proposed approach would perform in practice and what challenges would arise.

Table 1. Cont.

Ref.	Topic	Technique/Method	Data Sources	Main Findings	Limitations
[25]	Bidding strategies in local electricity markets and the potential effects of cascading wholesale market integration on those markets.	Bi-level optimization using computational intelligence through ant colony optimization, a variant of differential evolution called HyDE-DE, vortex search algorithm, and an estimation distribution algorithm called CUMDANCauchy++.	Hypothetical power profiles of residential houses and PV; realistic wholesale market (WSM) prices based on MIBEL market.	Showed that LEMs can reduce user costs and increase the profits of small producers in realistic case studies; take advantage of energy traded at the local level; benefits for both users and the aggregator.	Assumed an aggregator fixed tariff calculated using the forecast of the WSM prices; LEM transactions are done within a voltage limit and that the aggregator tariff offered to the consumers considers grid fees somehow.
[26]	Development of evaluation metrics for identifying the most suitable energy community end-users to participate in demand response (DR) programs	Clustering analysis by K-means and hierarchical clustering.	Realistic data collected from an energy community.	Clustering analysis can be an effective way to group energy community end-users based on their energy consumption patterns; the authors found that the K-means clustering algorithm is more effective than hierarchical clustering in grouping energy community end-users for demand response programs.	Limited number of performance metrics; the findings of the research are based on a specific energy community.
[27]	Energy communities for microgrids on the power grid.	A combination of clustering algorithms K-means and DBSCAN.	Real world datasets: National Imagery and Mapping Agency [28]; Richardson et al. [29] in East Midlands, UK; Barker et al. [30] in Massachusetts, USA.	Group energy consumers and producers based on their energy usage patterns and optimize the microgrid community based on energy flow and cost; enable the collaboration between energy consumers and producers to balance energy production and consumption.	The method relies on accurate data on energy consumption and production, which may not always be available.

### 1.2. Research Gap

To our best knowledge, since no studies reflect and evaluate the performance of different algorithms in REC clustering, it is of the utmost significance to determine which algorithms produce the best outcomes in this procedure. In this way, the following question arises: What is the best approach for clustering and classifying RECs to make them energy-efficient and balanced?

This research work will consider three clustering algorithms, with the main objective of grouping the different members of a database into several RECs, according to the characteristics and similarities of each member:

- K-means—an iterative unsupervised learning algorithm that attempts to partition a dataset into  $K$  distinct and non-overlapping subgroups or clusters, where each point only belongs to one group. It is characterized by having an ambitious approach to discovering clustering that minimizes the sum of squared error, converging to a local solution rather than a globally optimal solution [31];

- Density-based spatial clustering of applications with noise (DBSCAN)—a classic algorithm and one of the most important spatial density-based clustering algorithms. It can be applied to large datasets with outliers and simultaneously discover clusters with varied shapes with acceptable efficiency, even in the presence of noise. Furthermore, DBSCAN can find clusters that considers the characteristics of the data, and it is not necessary to define the number of clusters in advance. Hence, it allows the formation of groups with arbitrary formats [32];
- Linkage algorithms—an approach that uses the agglomerative hierarchical clustering method, which considers that the fusion of clusters is based on the distance between them [33]. There are several types of linkage algorithms, and these are divided into two major groups, the algorithms based on graphical methods and the algorithms based on geometric methods. Within the algorithms based on graphical methods, the most important and most widely used are single-link, average-link, and complete-link [33]. On the other hand, the most commonly used algorithm based on geometric methods is Ward’s method [34]. All of them are based on the similarity and Euclidean distance between the various data points, but the main difference is how this same distance is calculated.

These algorithms incorporate several clustering strategies, allowing a critical and comparative analysis of each algorithm’s performance to determine the most suitable method for clustering datasets of this sort.

It is worth noting that the clustering techniques are useful because they provide a way to identify structures in the data that might not be immediately apparent through traditional data analysis techniques (e.g., statistical analysis and regression analysis). Moreover, in the case of geographic location, net energy, and member identification, the clustering techniques would help to identify clusters of data points that share similar attributes or characteristics, such as similar geographic locations or similar net energy levels. This information can then be used to make more informed decisions about where to invest resources or what types of members to target for outreach. Additionally, clustering techniques can be useful for identifying anomalies or outliers in the data, which can be important for understanding the data and making predictions. For example, members in a cluster with significantly different net energy levels compared to the other members in the cluster might indicate an issue that needs to be investigated.

### 1.3. Contributions

The proposed platform can contribute to the minimization of power losses and congestion, mitigate power failures, and maximize local energy generation and sharing.

Considering the existing research gaps in the previous works, this paper presents the following contributions:

- Development of a clustering and classification model of RECs;
- Identification of an REC taking into consideration its energy characteristics;
- Understanding to what extent the correct formation of an REC contributes to a more stable and less congested power grid;
- Analyzing and comparing the efficiency of several clustering algorithms.

Furthermore, several performance metrics will also be addressed to enable a better evaluation of the clustering quality.

### 1.4. Paper Organization

After this introduction, Section 2 covers the proposed methodology and the details of its operation. A case study has been conducted and described in Section 3 to verify the performance of the proposed methodology. In Section 4, the results and their discussion are presented. Finally, Section 5 presents the most pertinent findings.

## 2. Proposed Methodology

This section presents a detailed description of the proposed methodology used in this research work. First, Section 2.1 discusses the methodology used to identify and classify different types of RECs. Next, Section 2.2 provides information about the developed model, which can group several members into RECs according to their net energy and geographical characteristics. Finally, Section 2.3 presents the three evaluation metrics used to assess the performance of each clustering model.

### 2.1. Identification and Classification of Renewable Energy Communities

We define  $N$  as the number of members existing in a community,  $E(t)$  as the net energy per unit of time  $t$ , and  $D$  as the spatial distance between two members in an electrical network. The net energy value is the difference between energy generation and consumption. It can be observed that if  $E(t) < 0$ , then the member needs to request external energy to sustain itself energetically; if  $E(t) > 0$ , then the member holds an excessive amount of energy that can be shared. Based on these variables, a distinction will be made between the different types of energy communities.

This research work also includes the sizing of renewable energy communities, i.e., determining the appropriate scale of the community's renewable energy infrastructure to meet its energy needs by trying to approximate the net energy closer to zero. This involves considering factors such as the community's current energy demand, local generation, and the availability of renewable energy resources (e.g., solar). In other words, the sizing aims to find the ideal number of members that form an REC and the right balance between energy generation and consumption to ensure that the community is able to generate enough renewable energy to meet its needs while avoiding overproduction or energy waste.

#### 2.1.1. Homogeneous Energy Community (HEC)

An HEC classification is relatively easy when compared to the others (namely, MEC and SEC) since it can be defined by a set of members whose net energies are only positive ( $E(t) > 0$ ) or only negative ( $E(t) < 0$ ).

For the identification of the different HECs in the electrical network, two variables are necessary: the net energy ( $E(t)$ ) and the geographical distances between members ( $D$ ). Therefore, the problem of identifying the energy community can be considered a clustering problem based on geographical distances between members, where the number of groups is given by  $K$ , and the net energy of the HEC can be aggregated. However, other constraints will have to be considered in the implementation of this concept since the  $E(t)$  of each HEC is limited, i.e., the power supply depends on the capacity of the installed generation units. In these cases, the number of RECs is unknown, and the following additional constraints (1) should be applied: if  $\text{HEC} \sum E(t)$  is positive, it may not exceed a positive upper limit; if  $\text{HEC} \sum E(t)$  is negative, it may not be less than a lower limit.

$$\sum_{i=1}^N E(t) = E_1(t) + E_2(t) + E_3(t) + (\dots) + E(t)_N \quad (1)$$

where:

- $E(t)$ — is the difference between the value of generated and consumed energy;
- $N$ —is the number of members that form an REC;
- $i$ —is the member number.

#### 2.1.2. Mixed Energy Community (MEC)

As previously verified, some members have positive  $E(t)$ , i.e., excess energy that can be shared or stored. In contrast, others have negative  $E(t)$ , i.e., they need energy from the grid. In this way, these members can connect and form an MEC. In an MEC, the members can share their locally generated energy surplus. In this power-sharing process, not only members with negative  $E(t)$  are favored as they can acquire cheaper energy, but

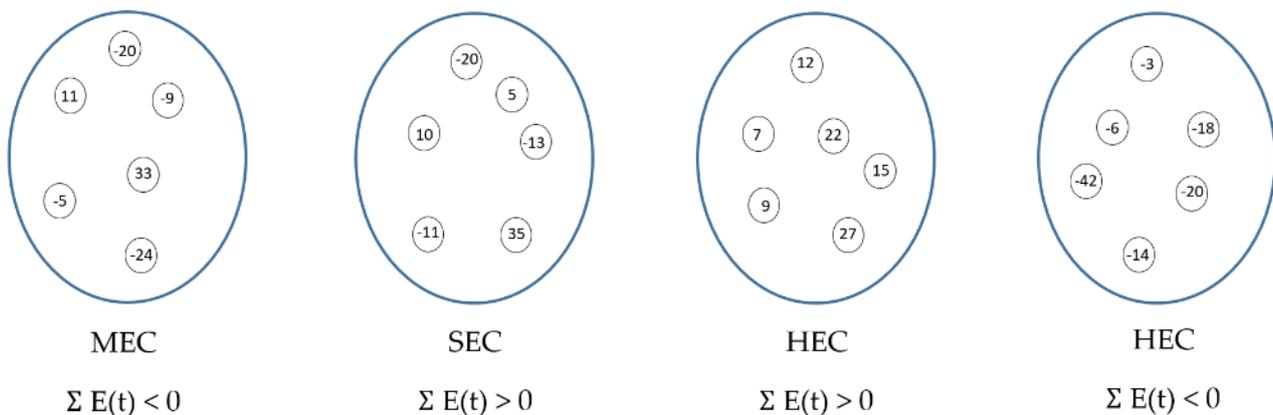
also members with positive  $E(t)$  are favored once they can increase the profitability of their generation units by selling the surplus.

The ideal MEC can be defined as a set of members geographically close to each other, with a balance between energy generation and consumption. Therefore, all locally generated energy is consumed in the vicinity, avoiding energy waste or injection into the grid. For instance, a member with positive  $E(t)$  and another with negative  $E(t)$  can mutually balance their energy needs using locally produced energy.

### 2.1.3. Self-Sufficient Energy Community (SEC)

The SEC can be classified as a special case within the MEC group since they are also composed of positive and negative  $E(t)$  members. However, there is an important difference: the members who constitute an SEC can fully balance the energy demand with locally generated energy through their generation units, mainly used for self-consumption. In this case, it can be considered that the  $E(t) > 0$ , which makes the study of these communities quite interesting since they do not depend, in part, on the electricity grid, thus creating several advantages for those who belong to them (e.g., they are not affected in case of contingencies of the main grid).

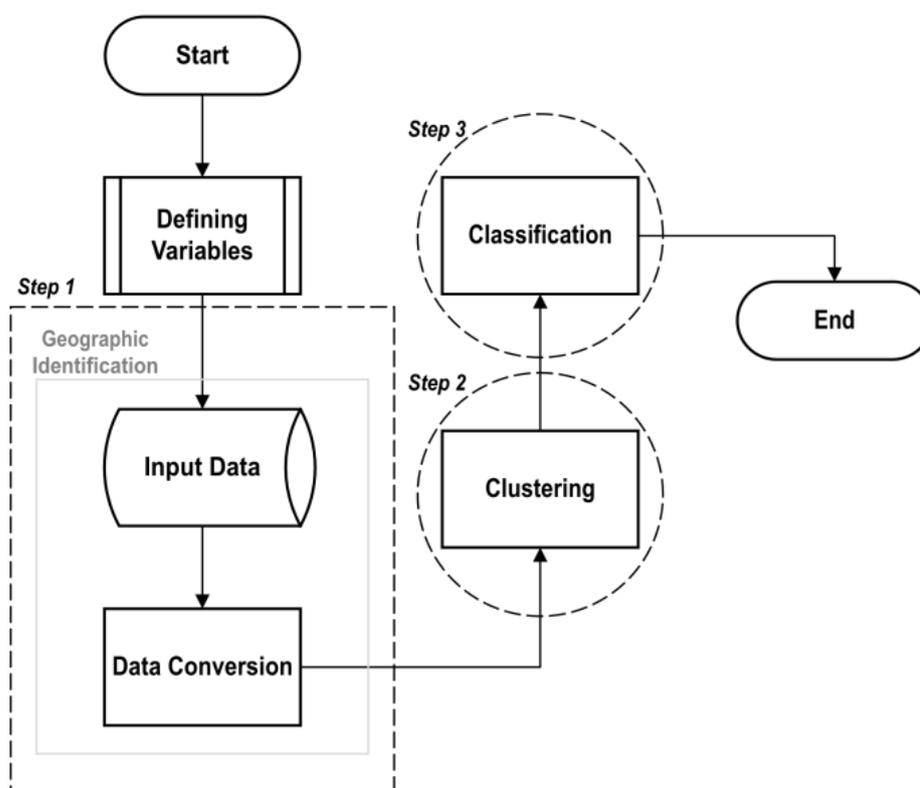
To consolidate the concepts related to identifying the types of RECs discussed above, Figure 1 shows how this classification is done, taking into account the net energy of each REC. As can be seen, an HEC is only formed by members with  $E(t) > 0$  or  $E(t) < 0$ . On the other hand, an MEC is constituted by members whose net energies can be negative or positive, while SECs require  $\sum E(t) > 0$ .



**Figure 1.** Comparison of the net energy of the different types of RECs.

## 2.2. Clustering and Classification Model

The development of three processes characterizes the proposed model, as can be seen in the flowchart in Figure 2. Initially (Step 1), the identification of the geographic coordinates of each member is done, enabling their graphical georeferencing. After that (Step 2), each algorithm (K-means, DBSCAN, and linkage algorithms) can run, and group (cluster) all the considered members. Finally (Step 3), each cluster (REC) is classified at the end of the process.



**Figure 2.** Diagram of the proposed platform—clustering and classification model.

With the need to geographically represent each member of the community, it was indispensable to collect each member's latitude and longitude values and store them in a database. In addition, it was also necessary to collect the energy generation and consumption values to calculate each member's net energy. The database stores the following data:

- Identification of each member;
- Annual net energy of each member (MWh/year);
- Latitude location of each member;
- Longitude location of each member.

The proposed model imports the data previously stored in the database. Initially, the latitude and longitude of each member are converted from the WGS84 (WGS 84 (World Geodetic System 1984) is a 3D geodetic datum that provides a consistent coordinate system for the entire earth and is the reference coordinate system used by the Global Positioning System (GPS) [35].) reference system to the UTM (UTM (Universal Transverse Mercator) is a type of map projection that is used to project the surface of the earth onto a two-dimensional plane. It divides the earth into 60 zones, each 6 degrees of longitude wide. Furthermore, it provides a convenient method for specifying positions on the earth's surface using a rectangular coordinate system [36].) system. This conversion allows each member to be represented in a Cartesian referential. After this process, the data are properly processed, allowing their importation to each considered algorithm.

As seen in Figure 3, the model starts by normalizing the values relative to each community's net energy (correlated energy) for the correct classification. After that, the classification of each REC starts. Thus, if all members belong to an REC with only positive net energy, it is classified as an HEC positive. On the other hand, if all members belong to an REC with only negative net energy, it is classified as an HEC negative. Moreover, an REC is classified as an MEC if its members have negative and positive net energies.

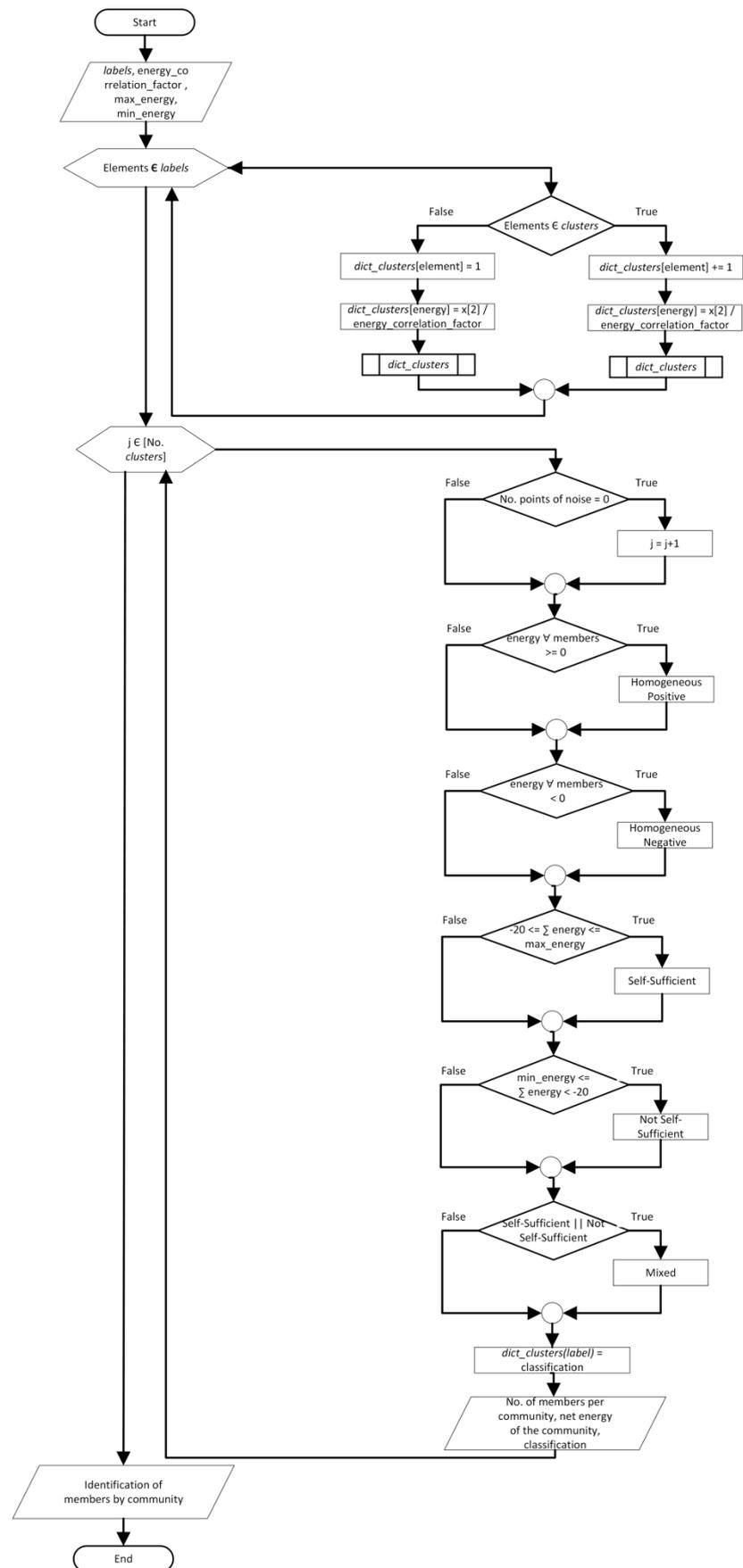


Figure 3. Flowchart of the classification model.

According to the sum of its members' net energies, an REC classification can be one of two types: self-sufficient and non-self-sufficient. For an REC to be considered self-sufficient, it must have a value of  $E(t) \geq 0$ . However, to speed up the classification process, the authors included two data intervals for the net energies summation. In this way, if the sum is within the range of  $[-20, \text{max\_energy}]$  it is classified as self-sufficient. Conversely, if the summation value falls between the range of  $[\text{min\_energy}, -20]$ , it is classified as not self-sufficient.

### 2.3. Metrics for Clustering Assessment

It is essential to thoroughly analyze the effectiveness of different clustering techniques. Therefore, we typically turn to cluster performance evaluation metrics to acquire more accurate findings that are easy to compare. This is because numerous situations require more than a manual and qualitative evaluation. To determine which algorithms produced the best clustering results, all of the algorithms that were being examined were subjected to the following three evaluation metrics:

- Silhouette score index: Used to measure how far apart clusters are from each other. It shows how close each point in a cluster is to each point in the clusters next to it and gives back the average silhouette coefficient used on all the samples. The silhouette coefficient is determined by taking the average distance between each sample's intra-cluster and its nearest cluster. The range of this coefficient is from  $-1$  to  $1$ . The more space between clusters, the higher the coefficients are (the closer they are to  $+1$ ). If the value is  $0$ , the sample is on or very close to the line that divides two nearby clusters. If the value is negative, the samples might have been put in the wrong cluster [37].
- Calinski-Harabasz index: The variance ratio theory is the basis for the Calinski-Harabasz index. Within-cluster diffusion and between-cluster dispersion are the factors used to determine this ratio. Clustering is more accurate for a higher index [38].
- Davies-Bouldin index: This index is defined as the average similarity measure of each cluster with its most similar cluster. The similarity is defined as the ratio of within-cluster and between-cluster distances. As a result, clusters that are farther apart and less dispersed will score higher. Conversely, the lowest possible score is  $0$ , and unlike most performance metrics, a lower value corresponds to a greater clustering performance [39].

### 2.4. Clustering Algorithms—Objectives and Constraints

K-means is a well-known clustering algorithm that aims to partition a set of data points into  $K$  clusters, where each data point belongs to the cluster with the closest mean.

The objective for K-means is to minimize the sum of squared distances between each data point (member) and the mean of its assigned cluster (REC). To achieve this objective, the K-means algorithm works by iteratively updating the cluster and reassigning the members to the closest REC. The algorithm repeats these steps until the cluster assignments no longer change, or a maximum number of iterations is reached. The constraints for the problem are: (i) the number of  $K$  clusters (REC) must be specified; (ii) it is assumed that the members are continuous variables; and (iii) it is assumed that the variance of each REC is spherical, meaning that each dimension of the feature space has an equal variance.

The math of K-means involves the calculation of distances and centroids.

The distance between two points (members)  $i$  and  $j$  is calculated using the Euclidean distance (the Euclidean distance is a measure of the distance between two points in Euclidean space. It is the straight-line distance between two points in a 2- or higher-dimensional space.) (2):

$$D_{i,j} = \sqrt{\sum [(i_p - j_p)^2]} \quad (2)$$

where  $D_{i,j}$  is the Euclidean distance between points  $i$  and  $j$ ,  $i_p$  and  $j_p$  are the  $p$ th dimensions of points  $i$  and  $j$ , respectively.

The centroid of a cluster is the mean of all the data points in the cluster (3):

$$C = \frac{1}{A} \cdot \sum B \quad (3)$$

where:  $A$  is the number of data points in the cluster,  $B$  is the set of data points in the cluster.

The objective function of K-means is to minimize the sum of the squared distances between each data point and its assigned centroid (4):

$$OF = \min \sum D_{i,C(i)}^2 \quad (4)$$

where  $i$  is a data point,  $C_i$  is the centroid of the cluster to which  $i$  belongs, and  $D_{i,C(i)}$  is the distance between  $i$  and its centroid.

DBSCAN groups members close to each other in a dense region. The objective of this algorithm in our work is to create an REC with closely packed members. For this, some constraints must be satisfied: (i) a minimum number of members to form a dense region (in our study, the minimum number of members to create a cluster (REC) is two) and (ii) a metric distance to determine the proximity of members to each other (it is considered 2 km of distance—see Section 4.1).

The DBSCAN involves the calculation of distances and densities between data points (members) and clusters (RECs).

A point  $i$  is considered a core point if it has at least  $minP$  other points within a radius  $\delta$  around it (5):

$$X_i = \{j \mid D_{i,j} \leq \delta\} \text{ if } |X_i| \geq minP ; i \rightarrow \text{core point} \quad (5)$$

where  $X_i$  is the set of points within  $\delta$  distance from point  $i$ ,  $minP$  is the minimum number of points required to create a cluster, and  $\delta$  is the radius within which  $minP$  points are counted.

A point  $j$  is a border point if it is not a core point but is within  $\delta$  distance of a core point (6):

$$\text{if } |X_i| \leq minP ; j \in \varphi \quad (6)$$

where  $\varphi$  is the set of all points in the dataset.

A point  $m$  is a noise point if it is neither a core point nor a border point (7):

$$\text{if } m \notin \varphi \quad (7)$$

Finally, two points  $i$  and  $j$  are density-connected if there exists a core point  $n$  such that both  $i$  and  $j$  are reachable from  $n$ . Additionally, a point  $i$  is reachable from a core point  $n$  if there exists a sequence of core points  $\{i_1, i_2, i_3, \dots, i_z\}$  such that  $i_1 = n$ ,  $i_z = i$  and  $i_{p+1}$  is directly density-reachable from  $i_p$ .

The linkage algorithms are a class of hierarchical clustering algorithms. The objective is aimed at the minimization of the distance between members within an REC while maximizing the distance between the RECs. This objective is subject to the following constraints: (i) determining which members can be grouped together in an REC. In single linkage, for instance, only the closest members are grouped together, while in complete linkage, all members within a certain distance threshold are grouped together; (ii) ensuring that the hierarchy of RECs is consistent with the metric distance used. In other words, for instance, if members 1 and 2 are closer to each other than 1 and 3, then the algorithm should not group 1 with 3 before grouping it with 2.

Each linkage algorithm differs in how they calculate the distance between clusters (RECs).

Single linkage is the simplest and fastest of the linkage algorithms. It calculates the distance between two clusters as the minimum distance between any two points in the two clusters (8):

$$D_{I,J} = \min D_{i,j}; \forall i \in I, j \in J \quad (8)$$

where  $I$  and  $J$  are two clusters, and  $i$  and  $j$  are individual points within the clusters.

Complete linkage calculates the distance between two clusters as the maximum distance between any two points in the two clusters (9):

$$D_{I,J} = \max D_{i,j}; \forall i \in I, j \in J \quad (9)$$

Average linkage calculates the distance between two clusters as the average distance between all pairs of points in the two clusters (10):

$$D_{I,J} = \frac{\sum D_{i,j}}{a \cdot b}; \forall i \in I, j \in J \quad (10)$$

where  $a$  and  $b$  are the number of points in clusters  $I$  and  $J$ , respectively.

Ward's linkage minimizes the increase in the total within-cluster sum of squares when two clusters are merged. It calculates the distance between two clusters as (11):

$$D_{I,J} = \frac{\|C_I - C_J\|^2}{a + b} \quad (11)$$

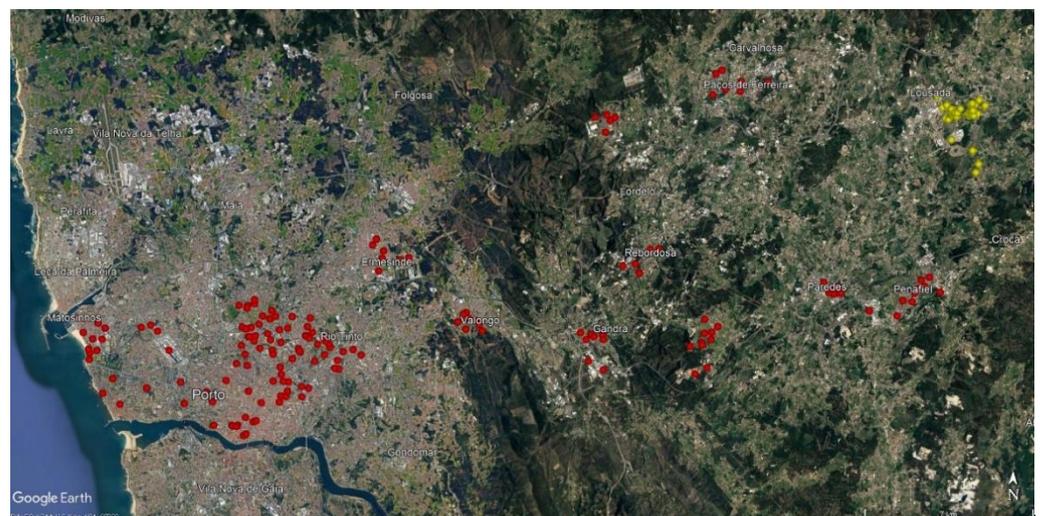
where  $C_I$  and  $C_J$  are the centroids of clusters  $I$  and  $J$ , respectively, and  $\|C_I - C_J\|$  denotes the Euclidean distance between the centroids.

### 3. Case Study

The subject addressed in this research work is relatively unknown and innovative. Consequently, it was only possible to obtain a small quantity of real data (from an existent renewable energy community in the north of Portugal), and most of the data of generation and consumption created considers several installations typologies.

#### 3.1. Data Analysis and Characterization

The necessary data for 170 members in the Porto district in Portugal were created to show how the proposed model is applied. Each member's electricity generation and consumption values and their geographical positions (latitude and longitude) were determined. To complement the study, the authors also used real data (from a renewable energy community project located in the north of Portugal) in conjunction with the generated data. These data were obtained through the community members' invoices. Figure 4 demonstrates the geographic location of each member considered in this case study. The red dots denote the considered/created members, and the yellow dots represent the real members.



**Figure 4.** Geographic location of the members under study (Google Earth Pro) (red dots are the Created Members-, yellow dots are the Real Members).

The idea to include created data (from fictitious members) in conjunction with real data (from real members) is to have the opportunity to consider in the proposed platform more cities in the same region (north of Portugal), many more members, make the problem much more complex in computing terms, obtain more reliable conclusions, and stress the proposed platform.

It is important to note that the latitude and longitude of each member is transformed from the WGS84 reference system to the UTM system, with a reference of 0 degrees for both latitude and longitude. Furthermore, the correlated energy is generated using a factor of 1000. This is critical for the built model to avoid issues during execution when dealing with energy quantities in MWh and kilometers for latitude and longitude.

The members were distributed by the zones of greater population/industrial density within the district of Porto. As shown in Figure 5, around 43% of the members are located in Porto city, with the remaining 57% distributed in neighboring regions, namely in Valongo, Lousada, Ermesinde, Paredes, Paços de Ferreira, Penafiel, Rebordosa, and Gandra.

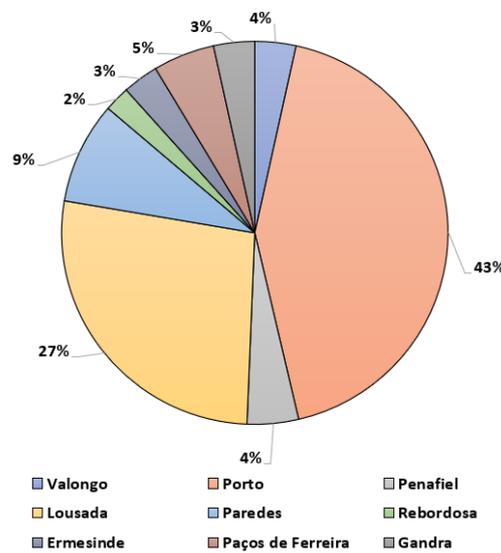


Figure 5. Geographical distribution of members by region.

The created energy consumption and generation vary between [0–400] MWh/year, referring to several types of installations, from private consumers to industries. Figure 6 shows the energy intervals distribution of members for annual generation and demand.

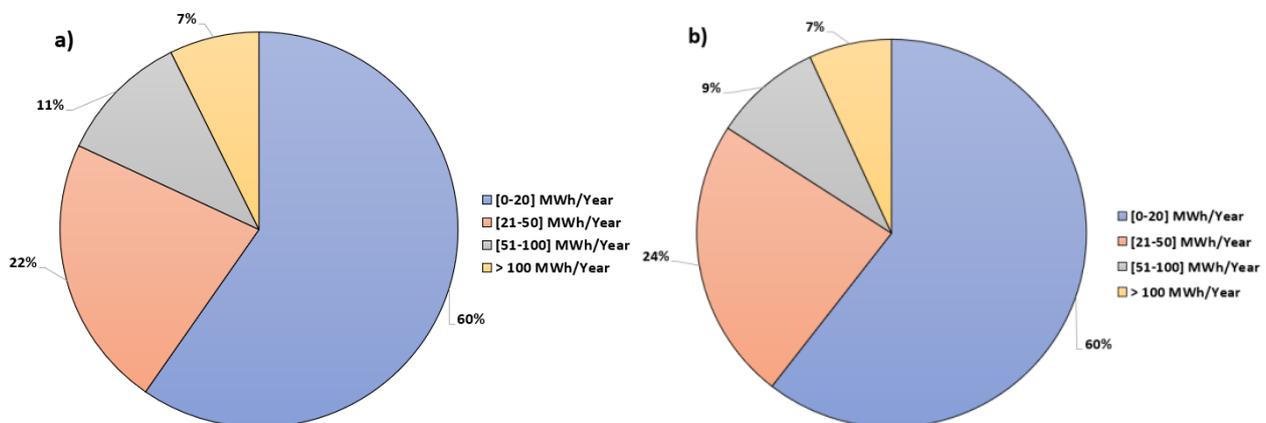


Figure 6. Members' annual generation and demand percentage—(a) annual generation (b) members' annual consumption values.

### 3.2. Renewable Energy Community—North of Portugal

The real REC is characterized as a mixed community, with about 14 electro-producing centers (photovoltaics) and about 63 consumer members. This set of members considers several types of facilities, from residential to commercial and industrial. For those, real data was used, namely the demand and local generation in MWh/year and their exact locations (latitude and longitude). Unfortunately, and for confidentiality reasons, it is not possible to identify the members, the entity responsible for the REC, and their exact location.

The photovoltaic (PV) installations considered the estimated electricity consumption profile relative to all REC members, ensuring that most of the energy generated is consumed within it (aiming to self-consumption and energy sharing). It was also necessary to calculate the energy generated annually and, consequently, estimate the average number of hours of solar radiation per day in the considered region, allowing the authors to calculate the average energy generated annually (513 MWh/year). Figure 7 shows the percentage of energy generated by each producer in the REC in the north of Portugal. The members' energy bills were used to calculate the REC average annual consumption, representing a total of 402 MWh/year (Figure 8).

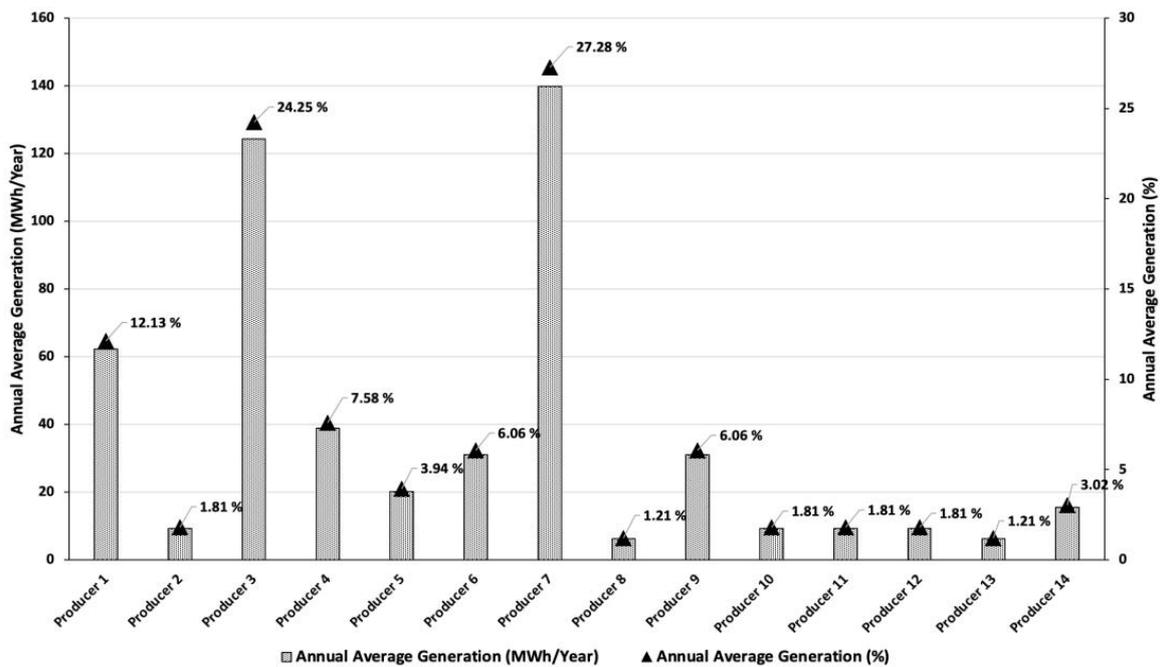


Figure 7. Estimated generation chart of electro-producer centers in the north of Portugal REC.

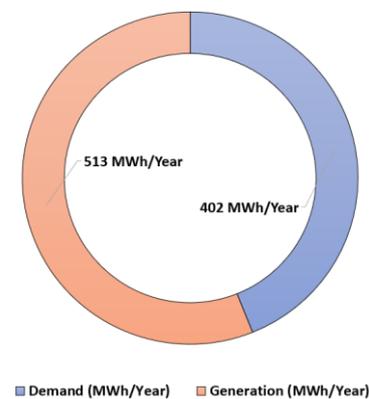
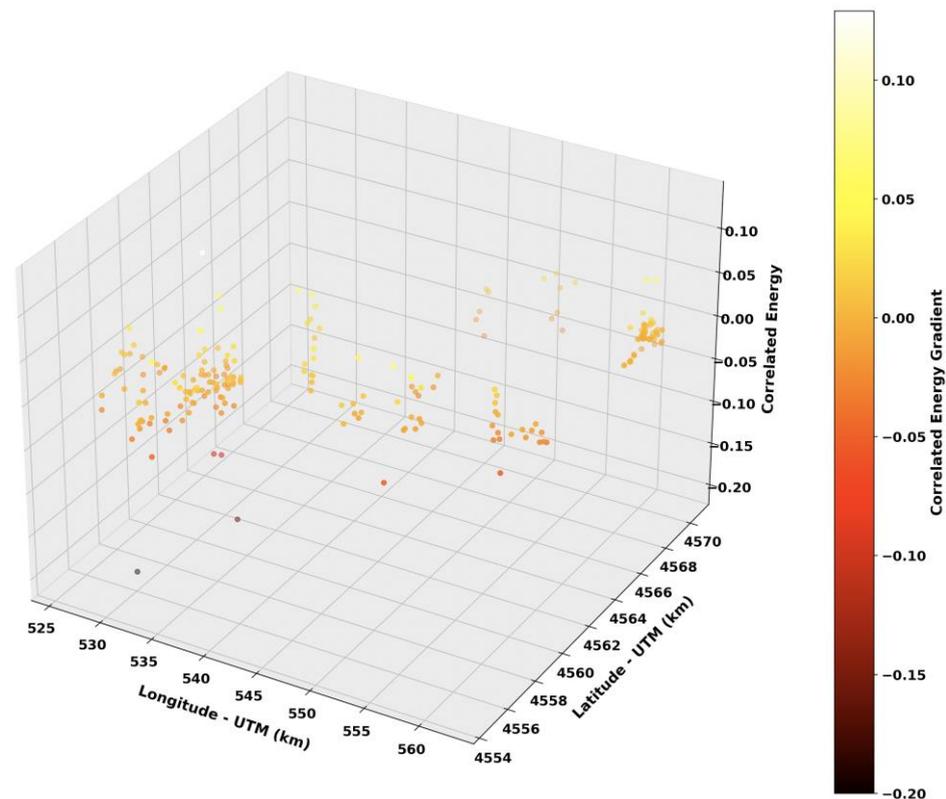


Figure 8. REC's estimated consumption and generation chart.

#### 4. Results and Discussion

The proposed methodology has been applied to the case study presented in Section 3 to show its applicability. To have an equal comparison base between all algorithms, it was necessary to predefine some variables before running the algorithms, namely the maximum distance between members (DBSCAN algorithm) and the predefined number of clusters for the K-means and linkage algorithms.

Figure 9 shows how the different members are distributed at the spatial and energetic levels. All members are represented in a Cartesian referential of 3 dimensions, where the x-axis represents the longitude, the y-axis the latitude, and the z-axis the correlated net energy. It can also be seen that overall, no cluster members stand out for the net energy value since the color gradient is uniformly distributed.



**Figure 9.** Net energy and geographic location comparison between members.

##### 4.1. DBSCAN Results

The study was developed considering the fact that all generation units are connected to a low-voltage distribution network. In this way, and according to Portuguese law, the geographical distance between the power plants and the consumer cannot be greater than 2 km in a REC [40]. Furthermore, only two members are required for an REC to be formed. Thus, the input parameters for the DBSCAN algorithm are:

- maximum distance between points = 2 (km);
- minimum number of points = 2.

Figures 10 and 11 show that the DBSCAN algorithm has grouped the members into ten RECs of which four RECs were classified as non-self-sufficient, and six RECs were classified as self-sufficient. In addition, it was also verified that seven of those RECs were classified as mixed, two as positive homogeneous, and one as negative homogeneous. As a result, the RECs' net energy values range between  $-497$  MWh/year and  $249$  MWh/year.

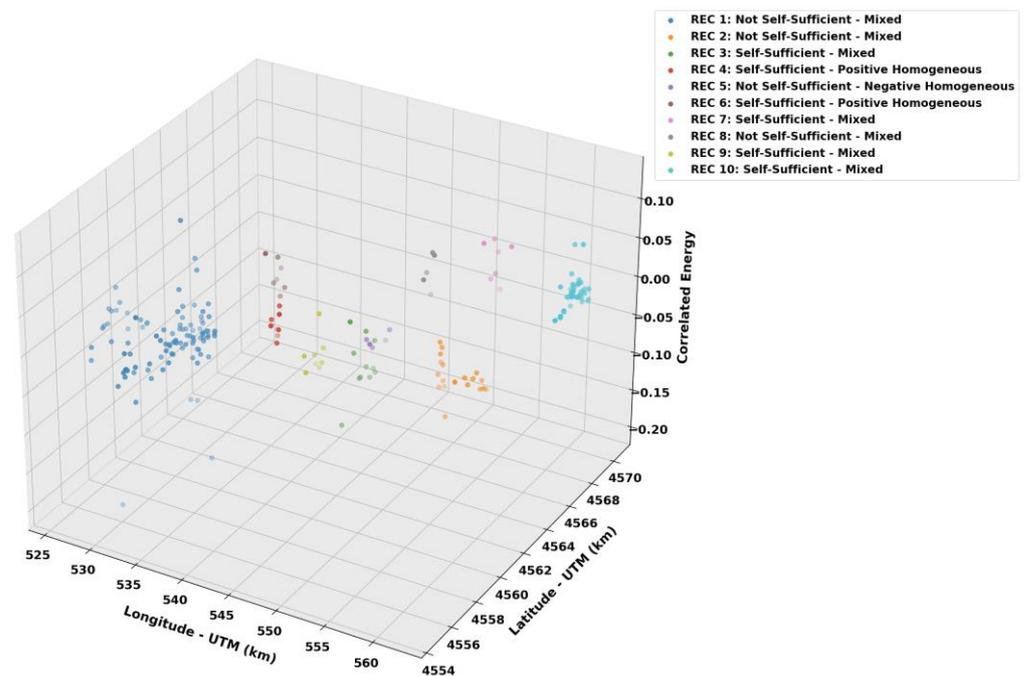


Figure 10. Members distributed by REC—DBSCAN.

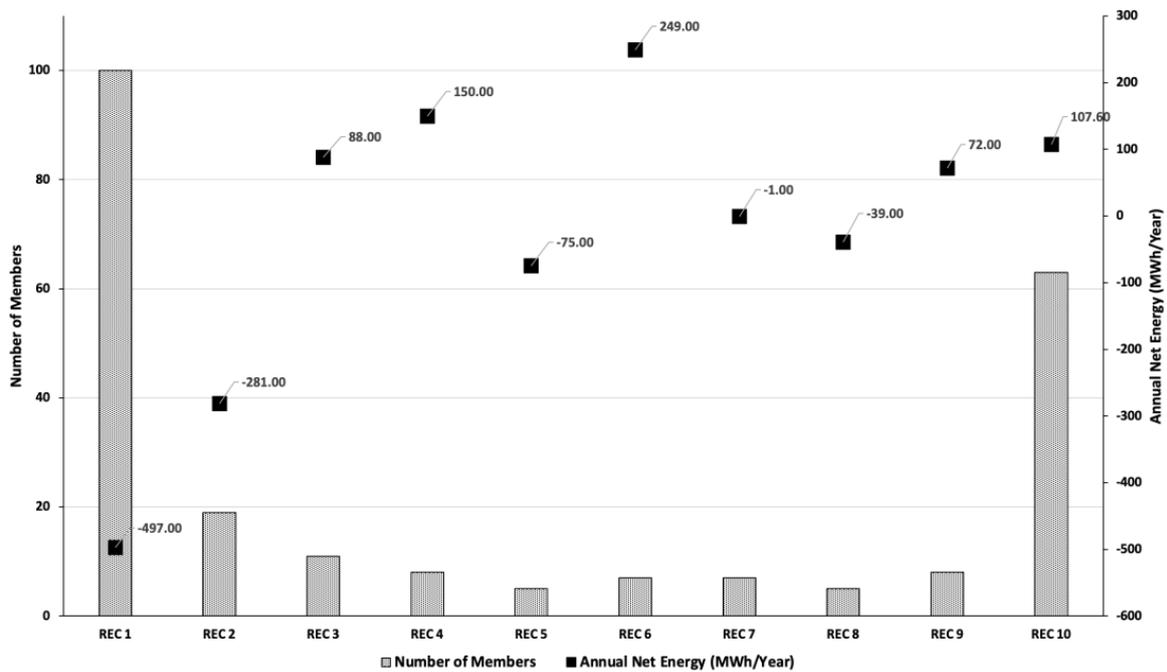


Figure 11. Number of members and annual net energy by the RECs—DBSCAN.

For a proper and efficient methods comparison, the number of clusters to be used in the other methods will be equal to the number of clusters given by DBSCAN, i.e., ten clusters (communities).

#### 4.2. K-Means Results

The K-means algorithm, beyond the data relating to members, requires the number of desired clusters.

The results obtained by the K-means algorithm were quite interesting. It was possible to see a more balanced distribution of members among the RECs at a quantitative level

compared to the DBSCAN algorithm (Figures 12 and 13). The K-means algorithm uses the partitional clustering technique, i.e., it is a non-deterministic algorithm. Thus, it generates different results in each run. As a result, in some runs, it was possible to verify that the K-means could not identify the real REC. This example is shown in Figures 14 and 15.

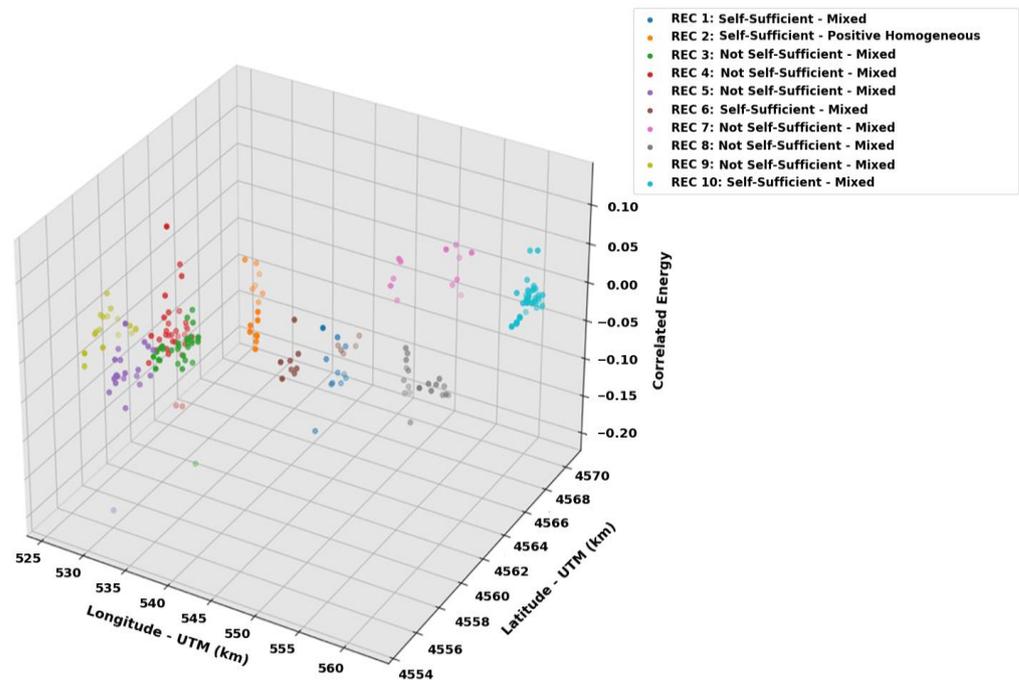


Figure 12. Members distributed by REC (with the real REC identified)—K-means.

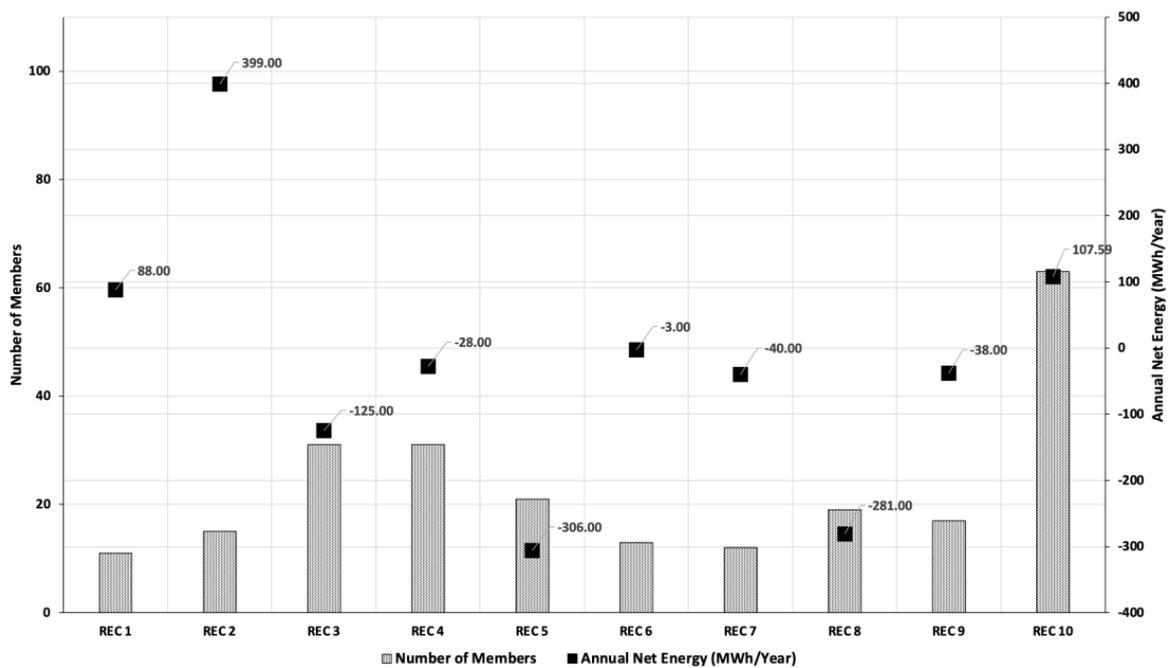


Figure 13. Number of members and annual net energy by REC (with the real REC identified)—K-means.

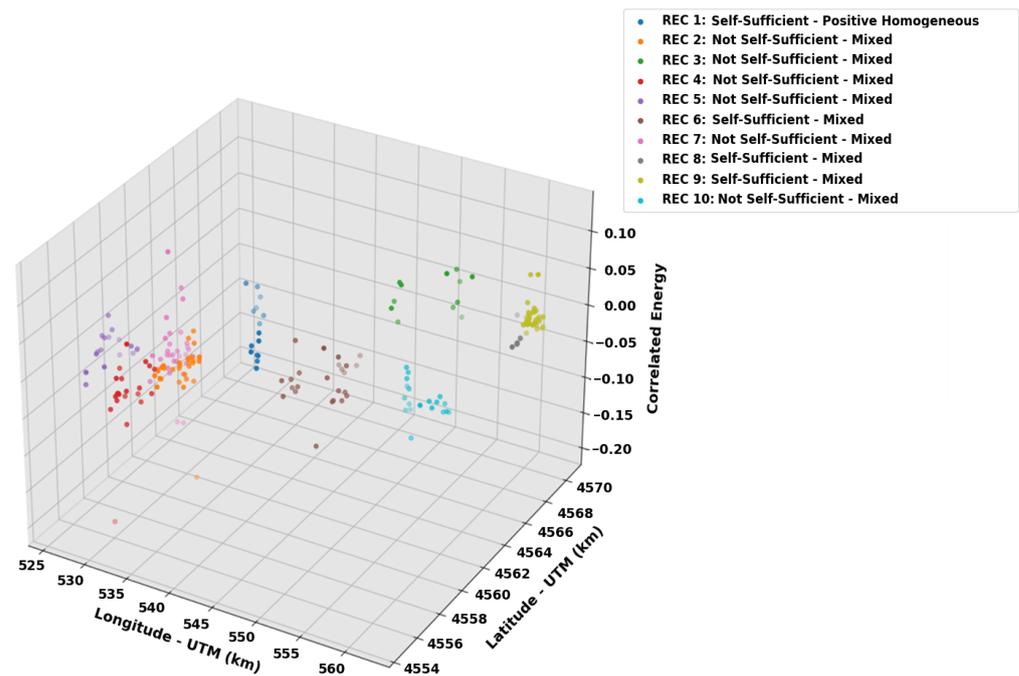


Figure 14. Members distributed by REC (real REC not identified)—K-means.

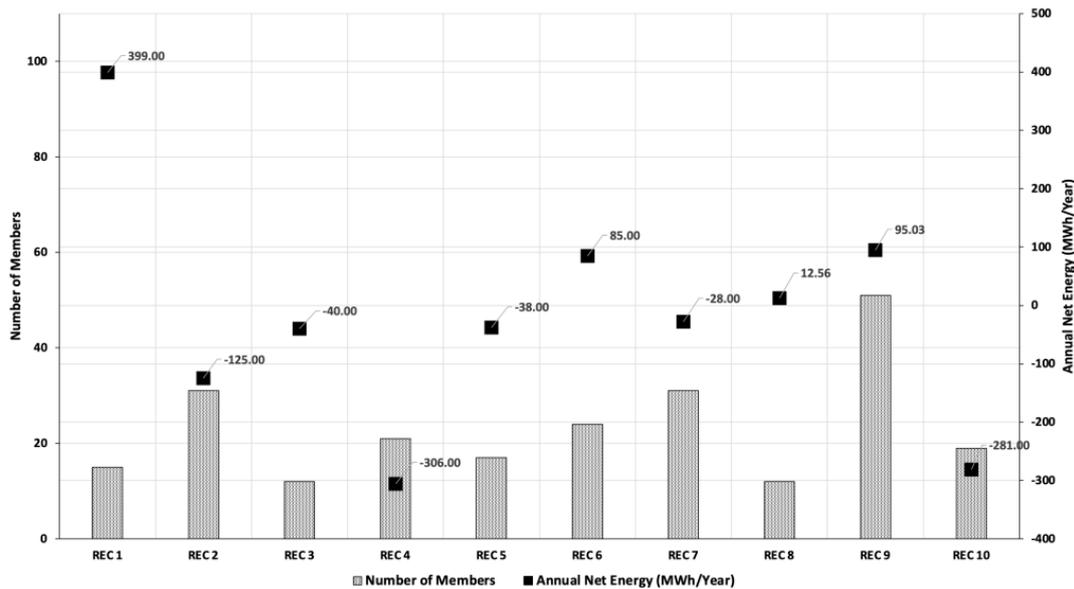


Figure 15. Number of members and annual net energy by REC (real REC not identified)—K-means.

Most of the results obtained from the K-means algorithm (when the real REC was identified) were very similar to those obtained by the Ward’s method (Section 4.3.1).

### 4.3. Linkage Algorithms

Similar to the K-means algorithm, the linkage algorithms only need the number of clusters to perform the data clustering.

#### 4.3.1. Ward’s Method Results

Ward’s method presents three self-sufficient RECs and seven non-self-sufficient RECs, of which one was classified as positively homogeneous and the others as mixed. As can be seen in Figure 16, this method presented more similar communities compared to the other

methods, in terms of the number of members and clusters/community size (no formation of large clusters).

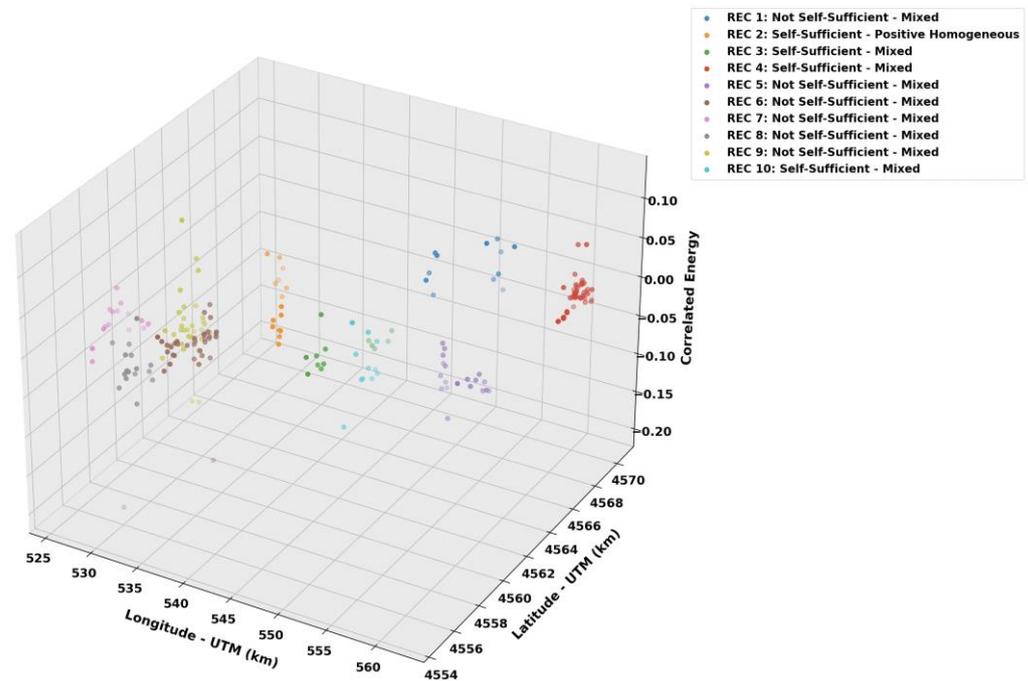


Figure 16. Members distributed by REC—Ward’s method.

However, at the energy level, Ward’s method presented the largest number of non-self-sufficient communities, ranging from  $-347$  MWh/year to  $399$  MWh/year (Figure 17).

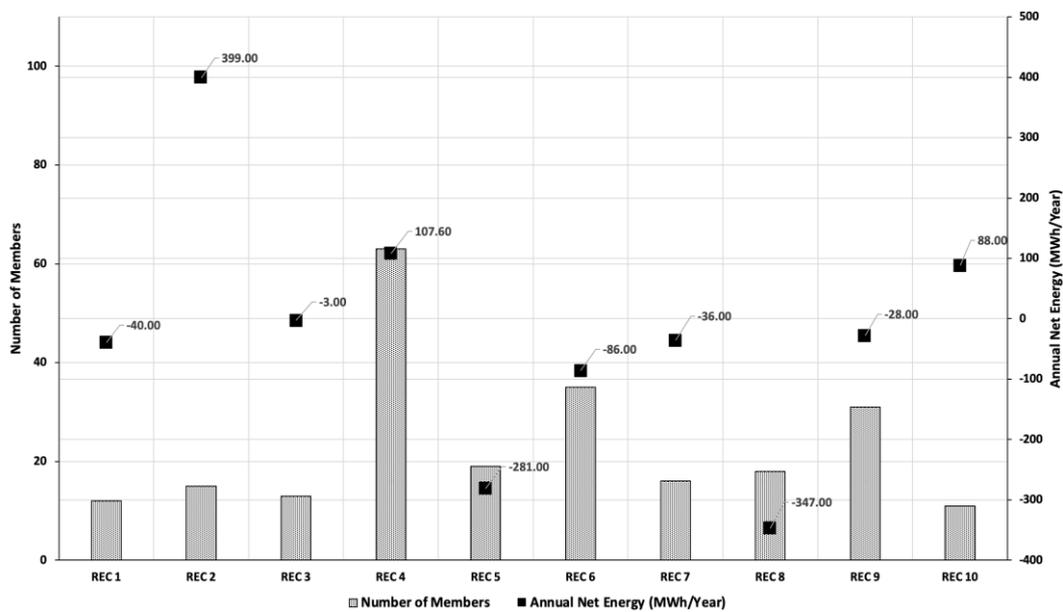


Figure 17. Number of members and annual net energy by REC—Ward’s method.

#### 4.3.2. Complete-Link Results

The complete-link method created five self-sufficient and five non-self-sufficient RECs. Two were classified as positive homogeneous, and the others as mixed (Figure 18). Analyzing the RECs created, this method presents the smallest variation in terms of net energy, varying between  $-281$  MWh/year (REC 4) and  $249$  MWh/year (REC 5), as can be seen in Figure 19.

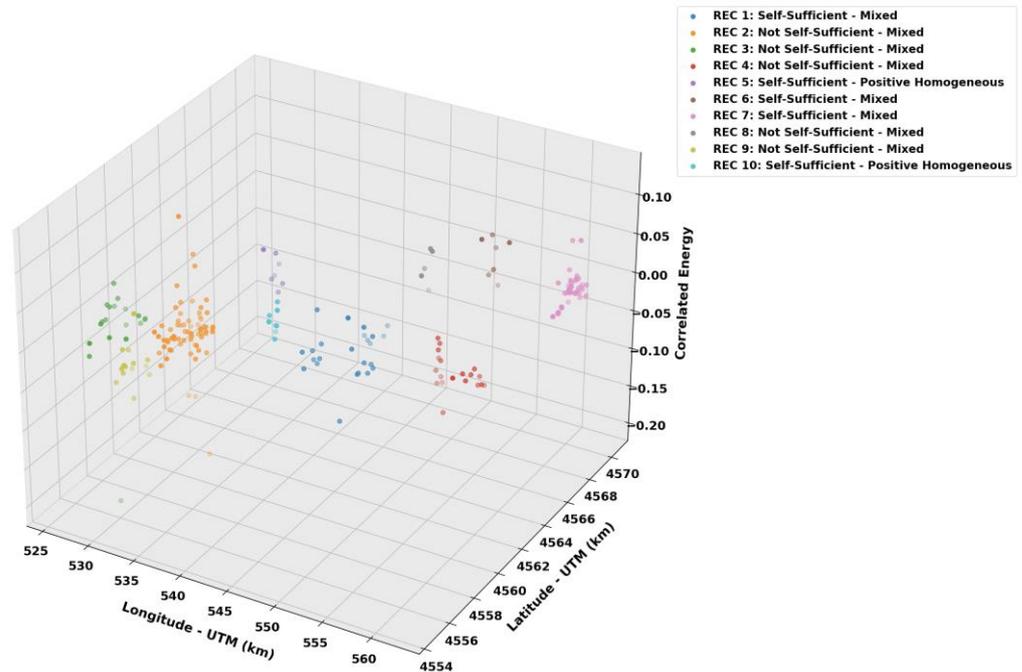


Figure 18. Members distributed by REC—complete-link.

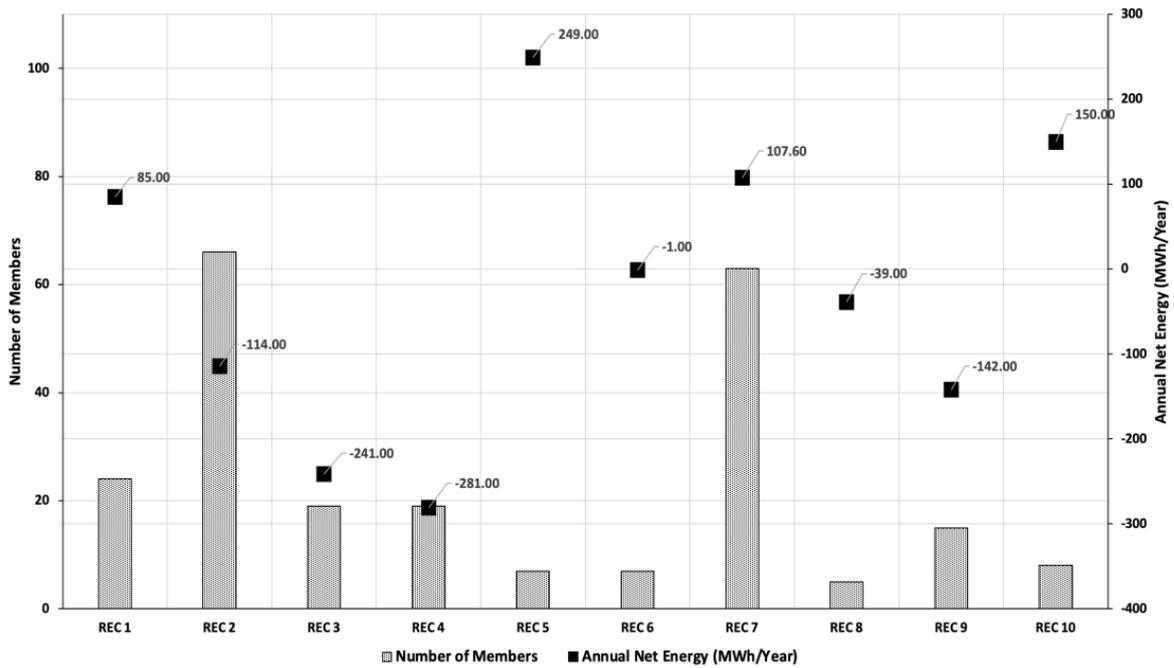


Figure 19. Number of members and annual net energy by REC—complete-link.

### 4.3.3. Average-Link Results

The average-link method presents six self-sufficient and four non-self-sufficient RECs (Figure 20). Two are positive homogeneous, and eight are mixed. This method showed more self-sufficient communities when compared to the complete-link method. Furthermore, this method presents more RECs with the net energy closest to zero, as shown in Figure 21.

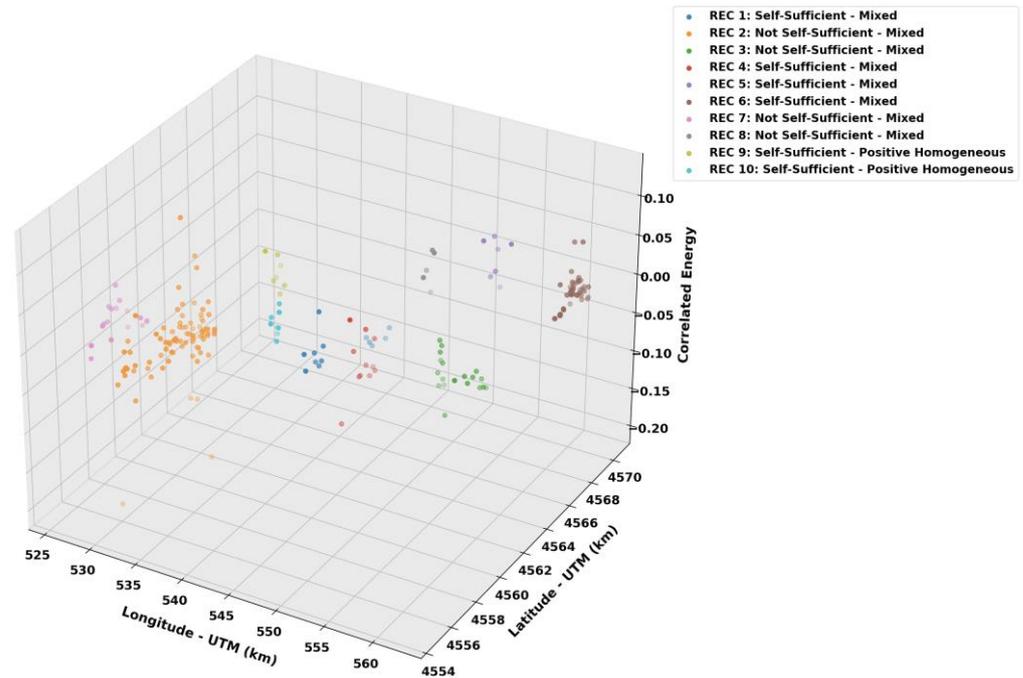


Figure 20. Members distributed by REC—average-link.

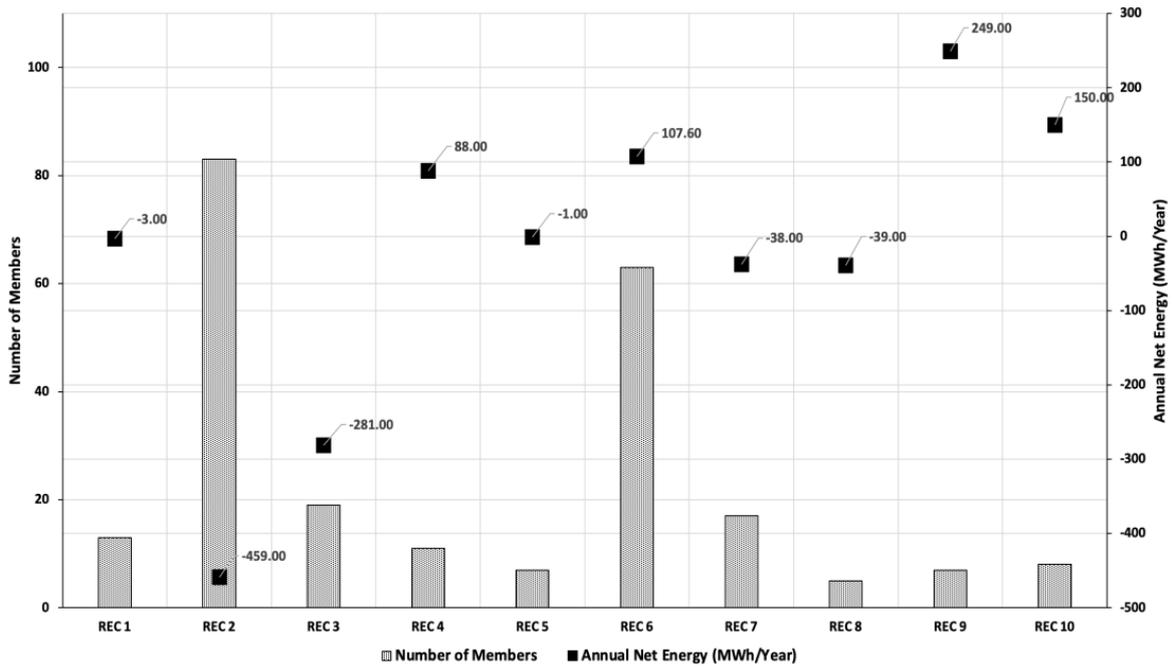


Figure 21. Number of members and annual net energy by REC—average-link.

#### 4.3.4. Single-Link Results

Figures 22 and 23 depict the single-link method results. This method presents the same number of self-sufficient and non-self-sufficient communities as the average-link method. However, with a small difference in classification, as it created two positive homogeneous RECs, one negative homogeneous, and seven mixed. This method’s results present some similarities with the DBSCAN algorithm. Both found more homogeneous RECs and created the RECs with the most members and the lowest net energy.

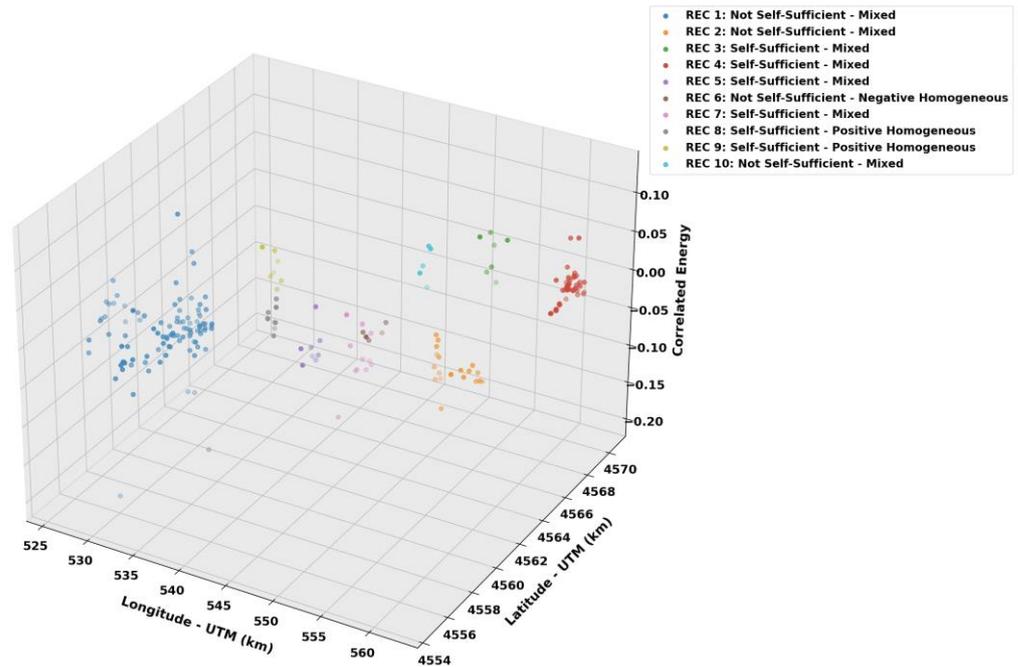


Figure 22. Members distributed by REC—single-link.

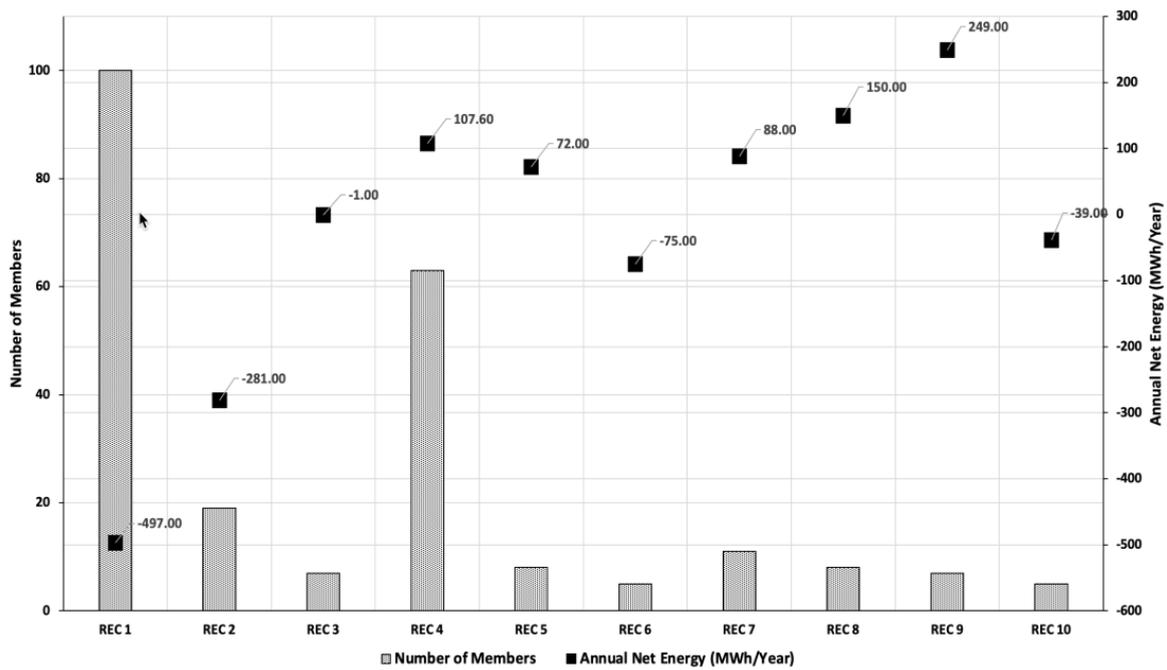


Figure 23. Number of members and annual net energy by REC—single-link.

#### 4.4. Clustering Evaluation Metrics

There were substantial differences in performances for the algorithms under study, as shown by the clustering evaluation metrics presented in Section 2.3. Table 2 shows the classifications obtained by the clustering techniques under consideration. The Silhouette, Calinski-Harabasz, and Davies-Bouldin indexes are commonly used to evaluate the quality of clustering results.

**Table 2.** Performance evaluation of clustering algorithms.

	Silhouette Score Index	Calinski-Harabasz Index	Davies-Bouldin Index
K-means	0.568	1951.153	0.578
DBSCAN	0.653	897.255	0.329
Ward's method	0.619	2124.174	0.590
Complete-link	0.631	1980.555	0.451
Average-link	0.671	1760.073	0.429
Single-link	0.653	897.255	0.329

Analyzing the Silhouette score index, it is possible to see that the average-link method is the one that achieves the best classification—the higher value (0.671), indicating that this method chose the members that were more well-matched to its own cluster. Moreover, the single-link and DBSCAN methods follow the average-link very closely, presenting a score of 0.653.

Regarding Calinski-Harabasz index, Ward's method obtained the best classification with 2124.174. This high ratio means that the clusters are well-separated (the between-cluster variance is large, and the within-cluster variance is small), presenting the most suitable clustering solution.

Concerning the Davies-Bouldin index, the single-link and DBSCAN methods are the ones that show the lower values (0.329), i.e., the lower average similarity score between each cluster and its most similar cluster. This means that the single-link and DBSCAN methods present the best results with the clusters well separated compared to the other methods.

Additionally, it is possible to see in Table 2 that the K-means method was the one with the worst overall performance, presenting the classification of 0.568, 1951.153, and 0.578 for the Silhouette score index, Calinski-Harabasz index, and Davies-Bouldin index, respectively.

It is worth noting that the Silhouette score index measures how well each data point fits into its assigned cluster, the Calinski-Harabasz index measures the separation between the different clusters, and the Davies-Bouldin index measures the distance between the clusters. However, the results of using these indexes to evaluate the quality of clustering results will depend on the specific parameters used in the clustering algorithm, such as the number of clusters, the metric distance used, and the initialization method. Therefore, it is important to carefully select these parameters to obtain meaningful and reliable results. For instance, increasing the number of clusters can improve the Silhouette score, as each cluster will have a smaller number of data points, leading to a better fit. However, increasing the number of clusters can also decrease the Calinski-Harabasz and Davies-Bouldin indexes, as the clusters may become less separated or more overlapped. Similarly, changing the metric distance used can affect the clustering results and, therefore, the evaluation indexes. For example, Euclidean distance can be appropriate for clustering continuous variables, while Hamming distance (in clustering algorithms, the Hamming distance is used as a measure of dissimilarity between two data points that are represented as binary vectors. The Hamming distance is simply the number of positions in which the two binary vectors differ.) can be appropriate for clustering categorical variables.

In general, these metrics can be used to compare different clustering algorithms or to tune the parameters of a single algorithm. A good clustering algorithm should produce high Silhouette scores, high Calinski-Harabasz indexes, and low Davies-Bouldin indexes, indicating that the clusters are well-separated and distinct.

#### 4.5. Classification and Sizing of Renewable Energy Communities—Adequacy, Versatility, and Robustness

By accurately classifying and sizing renewable energy communities, the adequacy of the renewable energy system can be improved, which means that it can be designed to meet the energy demands of the community without over or under-sizing the system. This can result in a more reliable and cost-effective renewable energy system. Moreover, classifying and sizing RECs can also improve the versatility of the renewable energy system. By understanding the energy demands and characteristics of the community, it

can be designed to incorporate multiple renewable energy sources and energy storage technologies. This can increase the flexibility and adaptability of the renewable energy system, making it more versatile to changes in energy demand and supply.

On the other hand, the robustness of the renewable energy system can be improved. Thus, the system can be designed to be more resilient to environmental changes and fluctuations in energy supply by accurately predicting the energy demand and supply. This can improve its reliability and resilience, making it more robust and capable of withstanding unforeseen events.

## 5. Conclusions

Due to the cooperation between network members and energy communities, the study of renewable energy communities with dispersed energy resources would improve the energy management of power grids. Several clustering and identification strategies for renewable energy communities, including homogenous, mixed, and self-sufficient energy communities, were proposed in this research work. We also tested the methods' effectiveness and efficiency using real and created datasets.

DBSCAN and single-link proved to be the algorithms capable of detecting the greatest number of energy communities with identical characteristics. Consequently, these algorithms could provide significant support in network energy management. They could predict, for instance, the optimal location for installing new generation unit resources with energy supply capacity in order to balance homogeneous energy communities with high energy consumption, thereby preventing excessive energy requests to the grid. The DBSCAN algorithm, on the other hand, was shown to be ineffective when clustering large high-density clusters since it only considers the distance between each cluster member and does not consider the distance between the most distant members of the cluster. Due to this, not all renewable energy communities generated by this algorithm satisfied the maximum distance requirement between generation units and customer installations.

Regarding the real CER, there was no problem for most of the algorithms under study in identifying it, except the K-means algorithm (as it generates a different solution on each run). Therefore, this REC is classified as mixed self-sufficient, confirming the pre-analysis for this REC.

Comparing the results obtained from the clustering evaluation metrics and the results obtained graphically, we can conclude that the results are identical, i.e., the linkage algorithms obtained the best results. In addition, they established the largest number of self-sufficient communities with net energies near zero, and the largest number of communities with homogenous energy. Regarding classification, this strategy proved to be effective, appropriately categorizing all formed renewable energy communities. Moreover, the linkage algorithms also stood out positively by presenting the best evaluation metrics classifications, i.e., 0.631 (complete-link) in the Silhouette score, 2124.174 (Ward's method) in the Calinski-Harabasz index, and 0.329 (single-link) on the Davies-Bouldin index.

The main limitations of the proposed model are: (a) it does not include any stochastic model for generation and demand forecast to improve the efficiency of each algorithm search; (b) it does not include a previous identification of the possibility of each community member supplying energy flexibility to improve the cluster's creation; and (c) since data gathering from all members is necessary for energy community classification and size, privacy issues could arise.

The findings indicate that identifying, classifying, and clustering renewable energy communities can be a valuable tool for distribution network planning, operation, and management.

**Author Contributions:** Conceptualization, B.C. and J.C.; methodology, B.C. and J.C.; software, B.C. and J.C.; validation, B.C., D.B. and Z.V.; formal analysis, B.C. and D.B.; investigation, B.C. and J.C.; resources, Z.V.; data curation, B.C. and D.B.; writing—original draft preparation, J.C.; writing—review and editing, B.C. and D.B.; supervision, B.C. and Z.V.; project administration, Z.V.; funding acquisition, Z.V. and B.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research has received funding from NGS Innovation Pact—New Generation Storage (C644936001-00000045), by the NGS consortium, co-financed by NextGeneration EU, through the Incentive System Agendas for Business Innovation, within the scope of the Recovery and Resilience Plan (PRR).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors acknowledge the work facilities and equipment provided by GECAD research center (UIDB/00760/2020, FEDER Operational Programme for Competitiveness and Internationalization (COMPETE 2020)) to the project team.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Canizes, B.; Soares, J.; Vale, Z.; Corchado, J. Optimal Distribution Grid Operation Using DLMP-Based Pricing for Electric Vehicle Charging Infrastructure in a Smart City. *Energies* **2019**, *12*, 686. [CrossRef]
2. Li, R.; Wu, Q.; Oren, S.S. Distribution Locational Marginal Pricing for Optimal Electric Vehicle Charging Management. *IEEE Trans. Power Syst.* **2014**, *29*, 203–211. [CrossRef]
3. Liu, Z.; Wu, Q.; Oren, S.S.; Huang, S.; Li, R.; Cheng, L. Distribution Locational Marginal Pricing for Optimal Electric Vehicle Charging Through Chance Constrained Mixed-Integer Programming. *IEEE Trans. Smart Grid* **2018**, *9*, 644–654. [CrossRef]
4. Yiasoumas, G.; Psara, K.; Georghiou, G.E. A Review of Energy Communities: Definitions, Technologies, Data Management. In Proceedings of the 2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), Thessaloniki, Greece, 17–19 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6.
5. de São José, D.; Faria, P.; Vale, Z. Smart Energy Community: A Systematic Review with Metanalysis. *Energy Strategy Rev.* **2021**, *36*, 100678. [CrossRef]
6. Ma, O.; Alkadi, N.; Cappers, P.; Denholm, P.; Dudley, J.; Goli, S.; Hummon, M.; Kiliccote, S.; MacDonald, J.; Matson, N.; et al. Demand Response for Ancillary Services. *IEEE Trans. Smart Grid* **2013**, *4*, 1988–1995. [CrossRef]
7. Sardi, J.; Mithulananthan, N. Community Energy Storage, a Critical Element in Smart Grid: A Review of Technology, Prospect, Challenges and Opportunity. In Proceedings of the 2014 4th International Conference on Engineering Technology and Technopreneurship (ICE2T), Kuala Lumpur, Malaysia, 26–28 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 125–130.
8. Tulabing, R.; Yin, R.; DeForest, N.; Li, Y.; Wang, K.; Yong, T.; Stadler, M. Modeling Study on Flexible Load's Demand Response Potentials for Providing Ancillary Services at the Substation Level. *Electr. Power Syst. Res.* **2016**, *140*, 240–252. [CrossRef]
9. Kennedy, J.; Ciufu, P.; Agalgaonkar, A. Intelligent Load Management in Microgrids. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–8.
10. Canizes, B.; Silveira, V.; Vale, Z. Demand Response and Dispatchable Generation as Ancillary Services to Support the Low Voltage Distribution Network Operation. *Energy Rep.* **2022**, *8*, 7–15. [CrossRef]
11. Rahbari-Asr, N.; Ojha, U.; Zhang, Z.; Chow, M.-Y. Incremental Welfare Consensus Algorithm for Cooperative Distributed Generation/Demand Response in Smart Grid. *IEEE Trans. Smart Grid* **2014**, *5*, 2836–2845. [CrossRef]
12. Sui, B.; Zhu, Y.; Gong, J.; Liu, B. Optimal Grid-Connected Storage Planning in Distribution Network Integrated with Renewable Distributed Generations. In Proceedings of the 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2), Changsha, China, 8–10 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 61–66.
13. Liu, D.; Shang, C.; Cheng, H. A Two-Stage Robust Optimization for Coordinated Planning of Generation and Energy Storage Systems. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–5.
14. Hong, Y.; Vaidya, J.; Lu, H.; Karras, P.; Goel, S. Collaborative Search Log Sanitization: Toward Differential Privacy and Boosted Utility. *IEEE Trans. Dependable Secur. Comput.* **2015**, *12*, 504–518. [CrossRef]
15. Saad, W.; Han, Z.; Poor, H.; Basar, T. Game-Theoretic Methods for the Smart Grid: An Overview of Microgrid Systems, Demand-Side Management, and Smart Grid Communications. *IEEE Signal Process. Mag.* **2012**, *29*, 86–105. [CrossRef]
16. Canizes, B.; Soares, J.; Lezama, F.; Silva, C.; Vale, Z.; Corchado, J.M. Optimal Expansion Planning Considering Storage Investment and Seasonal Effect of Demand and Renewable Generation. *Renew. Energy* **2019**, *138*, 937–954. [CrossRef]
17. Presidency of the Council of Ministers. *DECREE LAW No. 162/2019*; Presidency of the Council of Ministers: Lisbon, Portugal, 2019; Volume 162/2019, pp. 45–62.
18. Sale, H.; Morch, A.; Buonanno, A.; Caliano, M.; Di Somma, M.; Papadimitriou, C. Development of Energy Communities in Europe. In Proceedings of the 2022 18th International Conference on the European Energy Market (EEM), Ljubljana, Slovenia, 13–15 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
19. Faia, R.; Pinto, T.; Lezama, F.; Vale, Z.; Corchado, J.M.; González-Briones, A. Prosumers Flexibility as Support for Ancillary Services in Low Voltage Level. *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.* **2022**, *11*, 65–80. [CrossRef]
20. Union, E. Directive 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the Promotion of the Use of Energy from Renewable Sources. *Off. J. Eur. Union* **2018**. Document 32018L2001. Available online: <http://data.europa.eu/eli/dir/2018/2001/oj> (accessed on 1 February 2023).

21. Carlisle, N.; Van Geet, O.; Pless, S. *Definition of a “Zero Net Energy” Community*; National Renewable Energy Lab.: Golden, CO, USA, 2009.
22. Gjorgievski, V.Z.; Velkovski, B.; Cundeva, S. Quantification of the Shared Energy in Energy Communities. In Proceedings of the 2022 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Novi Sad, Serbia, 10–12 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
23. Faia, R.; Soares, J.; Vale, Z.; Corchado, J.M. An Optimization Model for Energy Community Costs Minimization Considering a Local Electricity Market between Prosumers and Electric Vehicles. *Electronics* **2021**, *10*, 129. [[CrossRef](#)]
24. Faia, R.; Soares, J.; Fotouhi Ghazvini, M.A.; Franco, J.F.; Vale, Z. Local Electricity Markets for Electric Vehicles: An Application Study Using a Decentralized Iterative Approach. *Front. Energy Res.* **2021**, *9*, 705066. [[CrossRef](#)]
25. Lezama, F.; Soares, J.; Faia, R.; Vale, Z.; Kilkki, O.; Repo, S.; Segerstam, J. Bidding in Local Electricity Markets with Cascading Wholesale Market Integration. *Int. J. Electr. Power Energy Syst.* **2021**, *131*, 107045. [[CrossRef](#)]
26. Barreto, R.; Gonçalves, C.; Gomes, L.; Faria, P.; Vale, Z. Evaluation Metrics to Assess the Most Suitable Energy Community End-Users to Participate in Demand Response. *Energies* **2022**, *15*, 2380. [[CrossRef](#)]
27. Hong, Y.; Goel, S.; Lu, H.; Wang, S. Discovering Energy Communities for Microgrids on the Power Grid. In Proceedings of the 2017 IEEE International Conference on Smart Grid Communications (SmartGridComm), Singapore, 23–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 64–70.
28. NGA National Imagery and Mapping Agency. Available online: [https://www.nga.mil/defining-moments/National\\_Imagery\\_and\\_Mapping\\_Agency.html](https://www.nga.mil/defining-moments/National_Imagery_and_Mapping_Agency.html) (accessed on 1 February 2023).
29. Richardson, I.; Thomson, M.; Infield, D.; Clifford, C. Domestic Electricity Use: A High-Resolution Energy Demand Model. *Energy Build.* **2010**, *42*, 1878–1887. [[CrossRef](#)]
30. Barker, S.; Mishra, A.; Irwin, D.; Cecchet, E.; Shenoy, P.; Albrecht, J. Smart\*: An Open Data Set and Tools for Enabling Research in Sustainable Homes. In *Proceedings of the Workshop on Data Mining Applications in Sustainability*; Association for Computing Machinery: New York, NY, USA, 2012.
31. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* **1979**, *28*, 100. [[CrossRef](#)]
32. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; AAAI Press: Palo Alto, CA, USA, 1996; pp. 226–231.
33. Rani, Y.; Rohil, H. A Study of Hierarchical Clustering Algorithm. *Int. J. Inf. Comput. Technol.* **2013**, *3*, 1225–1232.
34. Murtagh, F.; Contreras, P. Algorithms for Hierarchical Clustering: An Overview. *WIREs Data Min. Knowl. Discov.* **2012**, *2*, 86–97. [[CrossRef](#)]
35. *National Geospatial-Intelligence Agency (NGA) Standardization Document—Its Definition and Relationships with Local Geodetic Systems*; National Geospatial-Intelligence Agency: Boulder, CO, USA, 2014.
36. PROJ. PROJ Coordinate Transformation Software Library—Open Source Geospatial Foundation. Available online: <https://proj.org/operations/projections/utm.html> (accessed on 1 February 2023).
37. Belyadi, H.; Haghghat, A. Unsupervised Machine Learning: Clustering Algorithms. In *Machine Learning Guide for Oil and Gas Using Python*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 125–168.
38. Karna, A.; Gibert, K. Automatic Identification of the Number of Clusters in Hierarchical Clustering. *Neural Comput. Appl.* **2022**, *34*, 119–134. [[CrossRef](#)]
39. Xiao, J.; Lu, J.; Li, X. Davies Bouldin Index Based Hierarchical Initialization K-Means. *Intell. Data Anal.* **2017**, *21*, 1327–1338. [[CrossRef](#)]
40. Presidency of the Council of Ministers. *DECREE LAW No. 15/2022*; Presidency of the Council of Ministers: Lisbon, Portugal, 2022; pp. 3–185.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.