

Article

Contributions to Power Grid System Analysis Based on Clustering Techniques

Gheorghe Grigoraș^{1,*}, Maria Simona Raboaca^{2,3,4} , Catalin Dumitrescu^{5,*} , Daniela Lucia Manea⁶,
Traian Candin Mihaltan⁷, Violeta-Carolina Niculescu^{2,*}  and Bogdan Constantin Neagu^{1,*} 

- ¹ Department of Power Engineering, “Gheorghe Asachi” Technical University of Iasi, 700050 Iasi, Romania
² National Research and Development Institute for Cryogenic and Isotopic Technologies—ICSI Rm. Vâlcea, Uzinei Street, No. 4, P.O. Box 7 Râureni, 240050 Ramnicu Valcea, Romania
³ Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava, 720229 Suceava, Romania
⁴ Doctoral School Polytechnic, University of Bucharest, 060042 Bucharest, Romania
⁵ Department Telematics and Electronics for Transports, University “Politehnica” of Bucharest, 060042 Bucharest, Romania
⁶ Faculty of Civil Engineering, Technical University of Cluj-Napoca, Constantin Daicoviciu Street, No. 15, 400020 Cluj-Napoca, Romania
⁷ Faculty of Building Services Engineering, Technical University of Cluj—Napoca, Bd. 21 Decembrie 1989, No. 128-130, 400604 Cluj-Napoca, Romania
* Correspondence: ggrigor@ee.tuiasi.ro (G.G.); catalin.dumitrescu@upb.ro (C.D.); violeta.niculescu@icsi.ro (V.-C.N.); bogdan.neagu@tuiasi.ro (B.C.N.)

Abstract: The topic addressed in this article is part of the current concerns of modernizing power systems by promoting and implementing the concept of smart grid(s). The concepts of smart metering, a smart home, and an electric car are developing simultaneously with the idea of a smart city by developing high-performance electrical equipment and systems, telecommunications technologies, and computing and infrastructure based on artificial intelligence algorithms. The article presents contributions regarding the modeling of consumer classification and load profiling in electrical power networks and the efficiency of clustering techniques in their profiling as well as the simulation of the load of medium-voltage/low-voltage network distribution transformers to electricity meters.

Keywords: smart grid; clustering techniques; pattern clustering; power distribution planning; regression algorithms



Citation: Grigoraș, G.; Raboaca, M.S.; Dumitrescu, C.; Manea, D.L.; Mihaltan, T.C.; Niculescu, V.-C.; Neagu, B.C. Contributions to Power Grid System Analysis Based on Clustering Techniques. *Sensors* **2023**, *23*, 1895. <https://doi.org/10.3390/s23041895>

Academic Editors: Ching-Hung Lee and Lian-Wang Lee

Received: 17 December 2022

Revised: 4 February 2023

Accepted: 4 February 2023

Published: 8 February 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

A few of the changes and problems the world’s population is currently facing are related to the climate, electricity, food, water, transport, utilities, health, education, administration, and industry. Cities use 75% of the energy produced and are answerable for 80% of all dioxide emissions, although they only cover 2% of the planet’s surface. Future cities will need to adapt in order to counteract the effects of factors such as environmental change, population expansion, and social mobility, together with migration, human conflicts and unfairness, economic globalization, technological advancements, food, water, and energy vulnerability, geostrategic shifts, etc. Future cities must manage infrastructure and resources more intelligently to meet people’s needs both now and, in the future, as the world becomes increasingly “urban” [1,2].

Mobility is at the core of modern society when the aforementioned factors are considered. Over the next 20 years, there will be a lot of changes in this area as global auto markets and the transportation industry are reshaped by electrification, shared mobility, vehicle networking, and autonomous vehicles. This transformation is supported by technological advancement and other crucial variables, such as legislative directives pointing these two sectors in the direction of low-carbon solutions and increased fuel efficiency [3–6].

In the medium and long term, automakers and major EV operators are ensuring that the decarbonization goals are treated to a higher standard. Around the world, there are currently more than 7 million electric vehicles (EVs) in use, and other aspects of road transportation, such as the freight industry, are being electrified.

The electric vehicle is not a recent invention; it first emerged alongside internal combustion engine vehicles. Electric cars outperformed all other car types between 1890 and 1900. Because they were less noisy and polluted than cars with internal combustion engines, they significantly increased in popularity at the beginning of the 20th century. However, the short battery life of electric automobiles meant that their owners could travel great distances.

In recent years, technological advancements and concerns related to climate change have progressively sparked the resurgence of electric automobiles. The transformation from internal combustion to electric engines, which is now occurring in the automotive area, represents the most significant change. To adjust to changing market conditions, automakers have undertaken massive financial investments and unforeseen alliances [7–9].

EVs have been more and more popular in recent years due to their capacity to provide a variety of benefits, including [9]:

- Energy efficiency: Electric vehicles use less energy than vehicles powered by traditional internal combustion engines (ICEs).
- Electric mobility improves energy security because the road transportation industry is so reliant on petroleum-based fuels. Additionally, electricity may be created using a range of materials and fuels and is frequently produced locally.
- Air pollution: Since electric vehicles produce no emissions, they are an excellent solution to the issue of air pollution, particularly in densely populated regions and those nearby, where many citizens can be exposed to dangerous toxins from transit vehicles.
- Greenhouse gas (GHG) emissions: Combined with a progressive increase in the production of low-carbon energy, increasing electric mobility can result in considerable reductions in GHG emissions from the transportation infrastructure when compared to other traditional vehicles. Additionally, electric vehicles can behave according to the integration of renewable energy, which is often unpredictable for generating electricity, and offer flexible services for power systems.
- Noise reduction: Electric vehicles, especially those with two or three wheels, are quieter than ICE vehicles.

Industrial advancement: Given the relevance of energy storage for the switch to “clean” electricity, electric vehicles also have the capability of storing generated energy. In essence, battery technology, one of the fundamental factors in industrial competitiveness, is a possible facilitator for the significant reduction in cost in the electric car industry.

In many cities across the world today, one can find personal automobiles, common transport, car sharing, taxis, municipal parking lots, two- and three-wheeled machines (mainly electric scooters), as well as an expanding number of commercial and freight vehicle sectors. The proposed study topic fits within the background and is related to actual policies in the decarbonization of transportation areas, minimization of urban pollution, and the integration of EVs into electrical systems [10–12].

This study’s objectives were to analyze the impact and potential effects of integrating many EVs into the power network, as the efficiency could be influenced in both positive and negative ways by the need to charge all these EVs, as well as suggest several strategies and measures to integrate them. The article discusses the effects of integrating electric vehicles into the electric distribution networks and analyses those effects from a technical standpoint in terms of how they affect urban electric distribution networks and how that affects changes in electricity usage.

One challenge is connected to energy management, the impact of which may be reduced if the charging of EVs is achieved outside the peak period, which is likely to occur when big fleets of electric vehicles are incorporated into power networks. Electric vehicles (EVs) can be viewed from the perspective of electrical networks as either basic tasks

(with continuous consumption) or using the Vehicle-to-Grid (V2G) idea, where storage devices plan recharging intervals or they can inject grid energy using their energy storage devices [9,10]. In the case of a significant uptake of EVs, it will be important to coordinate their functioning both as a potential source of revenue and as a new aggregator that affects the electrical network by managing demand.

If EVs are gathered or aggregated, they can contribute to the balancing process by lowering grid usage or injecting electricity into the grid as needed by helping to coordinate the load.

The article is organized in the following way: a presentation of the clustering methods used in this study is first given; electrical load profiling, load-type profiles correlated with low voltage (LV) consumers, and a distribution network loading simulation are discussed along with the results obtained, and then the conclusions are presented.

In recent years, technological developments in the energy domain, including the introduction of smart meters and the transition to the concept of “Smart Grids”, have provided transmission and distribution operators with opportunities to forecast the load required by the system, modeling consumers to take into account their behavior, for the prevention of unplanned outages, the optimal load planning of generating units, etc. [13]. In these cases, operators must manage a lot of data and perform complex analyses in order to make the best decisions regarding the optimal planning and operation of electricity networks. For the efficient management of large databases, there are two aspects that need to be considered: data extraction (data analysis to obtain specific knowledge, patterns, or models) and database management (data storage, processing, and querying). Both related concepts are crucial in the energy networks’ decision-making process [14]. The choice of technology and the organization of installation, processing, and maintenance operations are the first steps in the adoption of smart metering systems. These technologies mark a significant development in the interaction between users and network operators. If consumer-mounted smart meters were totally integrated into a modern metering infrastructure and data were adapted, then distribution operators (ODs) could have access to full monitoring, which would make it easier to estimate the state of distribution networks [15–17].

The transition to the “Smart Grids” concept can lead to the implementation of smart monitoring and remote communications equipment necessary for optimal power systems operation and planning that will lead to maximizing economic benefits and minimizing the environmental impact. The phrase “Smart Grids” represents a hyperbole that involves the management of the EEA without the intervention of the human factor. The key components of this idea include two-way communication with consumers and all other market participants, as well as digital control of the energy transmission and distribution network. This smart infrastructure will enable various energy services, markets, integrated distributed energy sources, and control systems. The world economy in the future will be supported by the smart grid. It suggests that, in many respects, electricity generation, transmission and distribution companies, regulators, and institutions, indeed, all levels of government, face a real challenge in terms of the energy sector which is the driving force behind the world economy [18–22].

Starting from the aspects highlighted above, the efficient solution to many problems related to the management of the electric power system goes through the elaboration of solutions based on one form or another of artificial intelligence. Artificial Intelligence (AI) techniques aim to create intelligent computing systems, systems based on the characteristics of human intelligence: reasoning, ability to learn, solve and communicate, systems for problems for which there is no classical computational algorithm. Over time, with the development of AI techniques, hybrid algorithms have been developed and perfected, such as fuzzy logic-controlled neural networks, fuzzy genetic algorithms or expert systems, and artificial neural networks generated by genetic algorithms or neuro systems, having already proven their effectiveness [23–25].

Because an AI system has more complex tasks to solve, the knowledge that needs to be represented in it increases (facts, rules and heuristics of the field, general concepts, and

theories). In general, a system may work well, in line with the goal set by the knowledge provided, but any move outside its competence causes its performance to decline rapidly. This phenomenon is also called the fragility of knowledge [26].

Several machine learning techniques are presented in the studied literature, including supervised and unsupervised approaches that have been utilized for energy consumption level predictions. Among the unsupervised learning techniques, clustering is considered one of the most frequently applied techniques in data mining and machine learning. Clustering involves partitioning objects with similar patterns under observation into different groups. A vast number of works on clustering electricity usage patterns have been presented by researchers. The article [27] generated a typical load profile from data measured with automatic meter reading systems, then performed cluster analysis using three clustering algorithms, specifically, the hierarchical, k-means, and fuzzy c-means algorithms. In [28], classified daily load curves in industrial parks, which can be regarded as micro-grids from the energy network perspective using SOM, then exploited k-means to obtain a number of clusters, and [29] demonstrated the possibility of applying disaggregation techniques on smart meter data via fuzzy c-means clustering. Similar work is in [30], where they utilized the k-means algorithm to group residential houses with similar hourly electricity use profiles, and in [31] where they proposed a method to characterize medium voltage electricity consumers by using several clustering algorithms. In order to choose the best one among the typical load profiles, they measured the performance of the clustering algorithms in terms of eight clustering validity indices. To deal with the scalability and computational complexity of the power consumption profiling process, the authors of [32] proposed a multi-layered clustering method for power consumption profiling. First, they acquired local power consumption profiles using k-means, considering clusters with a low number of patterns as abnormal power consumption behavior. In the second stage, a global power consumption profile was derived from the local ones. Furthermore, Refs. [33,34] applied an improved k-means algorithm with particle swarm optimization (PSO) to open residential buildings datasets to divide their electricity consumption in an entire region into different levels. The authors of [35] developed a methodology in which one-dimensional time series smart meter data were reshaped to two-dimensional arrays called load profile images. After performing image processing techniques on those images, they derived the class load image profiles via clustering algorithms. In addition, [36] partitioned customers into electricity user groups based on similar electricity usage behavior with the SOM, k-means, and hierarchical clustering algorithms. Similar to group electricity consumption profiles, the authors of [37] investigated a shape-based clustering method.

The discussed literature reveals several limitations of the employed techniques from various perspectives of energy consumption prediction. The literature lacks focus on capturing the recognizable patterns in building smart sensing data, which has a limited number of features. These features can be represented in low-dimensional feature space and may affect the overall performance of data analytic tasks. Many of the existing techniques enquire about the number of clusters to differentiate among distinct categories of data. In addition, the presentation of energy consumption for data analysts and common individuals is a common problem that has not been tackled effectively in the existing literature.

Therefore, based on the above-mentioned problems in household energy predictions, this paper presents a new framework with the following main contributions:

- There is not a lot of energy consumption data obtained from smart sensors for residential buildings, and the presence of missing data is a difficult problem in statistical analysis. Less than 1% of missing data can be considered a common problem, and between 1 and 5% can be considered a solvable problem. If the percentage is greater than 5%, difficulties may arise in solving the respective problems. Thus, if the values are between 5 and 15%, the problem requires the application of sophisticated solution methods, and if the value exceeds 15%, the problem may have difficulties in interpretation. Finding recognizable patterns in such data is very difficult, which affects the performance of electricity consumption analysis. To solve this problem, five

hierarchical clustering methods were used in the clustering process: average distance (Average Method), the center of gravity (Centroid Method), minimum distance (Single Linkage Method), maximum distance (Complete Linkage Method), and Ward, which were the basis for obtaining the load-type profiles presented in the article. Based on the clustering methods used, we propose a method based on load-type profiles which is more robust, reduces calculation errors for off-peak time/peak time, requires less computation volume, and converts small model representations from data with reduced dimensions in high-level representations, comparable to using large databases.

- Clustering algorithms require an input parameter to divide data into multiple clusters. This article, by using the method of simultaneous layers in the hierarchical clustering process, average distance (Average Method), the center of gravity (Centroid Method), minimum distance (Single Linkage Method), maximum distance (Complete Linkage Method), and Ward, obtained an adaptive grouping to organize large data.
- After dividing the data into several clusters, the regression model based on first- and second-degree polynomials for every consumption class was applied, performing a predictive statistical analysis on the data to determine which buildings had a high, medium, and low level of energy consumption.

2. Materials and Methods

2.1. Clustering Techniques

The use of clustering algorithms allows for the spatial distribution of characteristic vectors to be used as a basis for grouping input data. Each element connected with a set of data will be characterized by a vector whose components are represented by the representative qualities or attributes of the vector in order to examine the similarity or differences between them and to categorize them. The determination of the characteristic/attribute number and their definition requires a deeper analysis of the database designed on the available information and considering the expertise of specialists [38,39]. The vectors connected to the input data are grouped throughout the clustering phase based on the estimated distance between each of them. Depending on the analyzed topic, the clustering process will result in one or more clusters (groups, patterns, models, or classes), which describe the spatial position of the qualities taken into account for the process' elements. Within each cluster, the elements are closer to a common center when compared to other centers belonging to other groups. This aspect is exemplified in Figure 1.

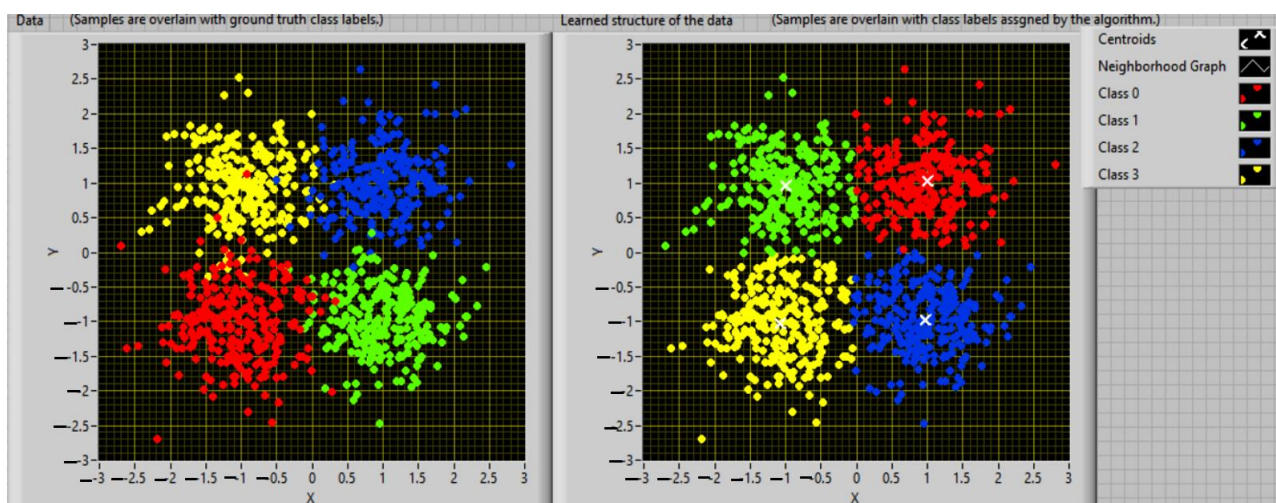


Figure 1. Example of clustering processing.

In Figure 1, the elements, represented by vectors with two characteristics (x_i, y_i) , $i = 1, \dots, N$ (where N is the maximum number of elements of the database subjected to the clustering procedure) were assembled using a similarity principle defined by the distance

calculated between vectors. In this mode, two or more components can be associated with the same pattern if the distance between them is smaller in relation to the distance from the elements of another cluster. Finally, each cluster will be characterized by a representative element, determined by mediating the characteristics of the elements that make up the cluster [40].

2.2. Stages of the Clustering Procedure

Clustering processes can be applied in various domains in order to group unlabeled components. These domains already involve various assumptions, terms, or techniques, related to clustering procedure phases, as a function of the addressed problems.

The steps that must be covered are described below and represented in Figure 2 [40,41].

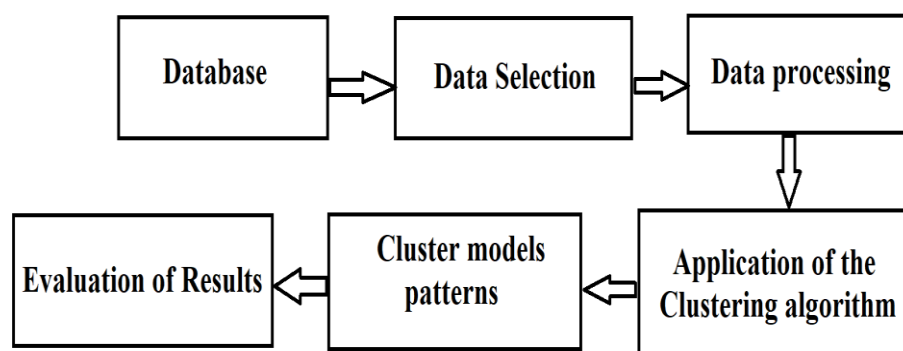


Figure 2. Clustering process phases.

Step 1. The components subjected to the clustering procedure must be established. In this phase, it is necessary to consider the option from the database that is best suited to the aim of the problem. The type and size of the attributes available for the clustering procedure can be chosen.

Step 2. The attributes/characteristics of the components subjected to the process must be extracted. The identification of the most useful and important characteristics must be achieved. During this process, one or more component transformations can be accomplished to obtain new dominant attributes.

Step 3. A similarity measure must be defined. Usually, the similarity can be determined by measuring the distance connecting adjacent items. Once a vector has been attributed to each component, this length may indicate how similar two elements are. The literature has defined several different distance measurement techniques, with Euclidean distance being the most used.

Step 4. This phase represents the actual clustering procedure. It can be achieved in various modes depending upon the techniques chosen by the decisional factor. All clustering techniques should conduct several clusters for any input data set. If no clusters resulted from the process, other techniques can be applied in order to obtain the desired results. The obtained results can be “clear”, meaning that the separation of the components is achieved in well-defined clusters, or they can be “fuzzy”, meaning that each component has a degree of dependence on each cluster.

Step 5. The results must be extracted. For this, an accurate interpretation of the results is necessary so that their rendering can be achieved in a simple way that is easy to interpret by the decisional factor. In this scenario, either from the point of view of automatic analysis (where a computational system can perform further data processing effectively) or from the perspective of a human, simplicity is required (the representation used for results is easier to understand by decision makers). The extraction of the results from the clustering method is a brief illustration of each cluster using representative components.

Step 6. Evaluation of results. An assessment of the clustering procedure is taken into account when analyzing the validity of the outcomes (represented by clusters), and this evaluation often employs an optimization criterion. This impartial analysis examines

whether the outcomes are accurate. If a cluster does not occur unintentionally or for other reasons, it is validated.

2.3. Clustering Methods

Next, certain terms will be defined and the notions that will be used will be briefly described in the next section.

The element x^j , $j = 1, \dots, N$ (where N is the total number of elements), is a unique object used in the clustering process. It is usually represented by an n -dimensional vector $x^j = [x_1^j \ x_2^j \ \dots \ x_n^j]$.

Scalar components x_i^j , $i = 1, \dots, n$, are called characteristics/attributes of the vector x^j and are established by the decision maker.

The distance between vectors is a metric in the space of the attributes x_i^j , $i = 1, \dots, n$, corresponding to input vectors x^j , $j = 1, \dots, N$, and used to determine the similarity between elements. The most used is the Euclidean distance.

For a database X , consisting of the vectors x^j , $j = 1, \dots, N$, with n characteristics, $x^j = [x_1^j \ x_2^j \ \dots \ x_n^j]$, different distances between vectors can be defined. Thus, if two vectors x^r and x^s are taken into account, the distance can be calculated with the relation:

$$d(x^r, x^s) = \sqrt{(x^r - x^s)(x^r - x^s)^t} \quad (1)$$

There are several ways to classify clustering methods in the literature. The most used classifications are given in [40,42]: hierarchical methods and the K-means method.

Hierarchical clustering methods can be subdivided, according to their meaning, into methods of agglomeration and division. In the case of agglomeration methods, for example, we start from the k clusters, each containing a single element x^j , $j = 1, \dots, N$, and by successive mergers, form a single cluster, containing all N elements. In the case of division methods, the direction of deployment is inverse, i.e., starting from a single cluster containing all the x^j , $j = 1, \dots, N$ elements, we reach k clusters, each containing a single element x^j . Agglomeration techniques are usually used more frequently. As shown above, in the hierarchical spatial grouping, an agglomeration process goes through a series of mergers/couplings of groups/classes, $P_n, P_{n-1}, \dots, P_1$. The first, P_n , consists of n "groups" with a single element/object, and the last P_1 includes a single group having all n elements/objects. At each stage, the method couples two close groups (at the first level, of course, this means the coupling of two elements/objects that are close to each other (in distance), since at the initial stage, each group has an element) [40,43].

The clustering process can be illustrated as a two-dimensional diagram named a dendrogram. These methods are suitable for small tables, having a few hundred rows. The desired number of clusters can be chosen after the proper shaft is designed by imposing a threshold [30,40].

The difference among the agglomeration techniques is given by the method for defining the distance between the clusters.

Considering two clusters, C_r and C_s , containing n_r and n_s elements, the average distance $d(C_r, C_s)$ is calculated on an Euclidean distance base:

$$d(x_r^f, x_s^h) = \sqrt{\sum_{k=1}^N (x_{rk}^f - x_{sk}^h)^2} \quad (2)$$

can be expressed through the relation:

$$d(C_r, C_s) = \frac{1}{n_r n_s} \sum_{f=1}^{n_r} \sum_{h=1}^{n_s} d(x_r^f, x_s^h) \quad (3)$$

Starting from the above-mentioned, the most used hierarchical clustering methods are briefly presented below.

The following figure shows the result obtained by applying the hierarchical clustering algorithm (Figure 3).

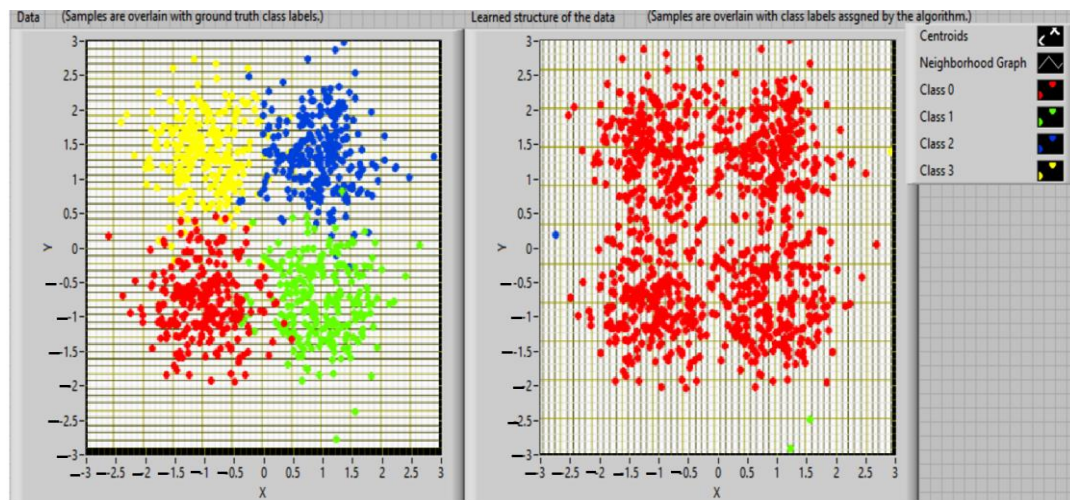


Figure 3. The representation of hierarchical clustering algorithms.

a. Minimum distance method (minimum method)

This method is the simplest, being based on the minimum distance, also known as the method to the closest neighbor. In such cases, the distance between the clusters represents the distance between the closest items:

$$d(C_r, C_s) = \min\{d(f, h)\} \quad (4)$$

where the f component is attributed to the C_r cluster and the element h to cluster C_s . In this case, the distance is calculated between each possible pair of elements (f, h) . The minimum value is the distance between the cluster's C_r and C_s . In other words, the distance between two clusters is given by the shortest link value. At every stage of the clustering process, the C_r and C_s clusters, for which $d(C_r, C_s)$ is minimal, will be coupled. A graphical interpretation of the minimum distance between the clusters is shown in Figure 4.

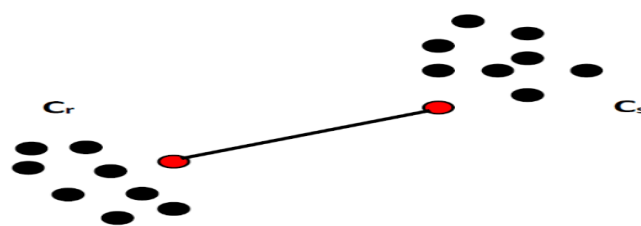


Figure 4. A representation of the minimum distance between two clusters.

b. Maximum distance method (maximum method)

This method, also known as the farthest neighbor method, is based on distance maximum, being the opposite of the minimum method. In this technique, the distance $d(C_r, C_s)$ is calculated using the equation:

$$d(C_r, C_s) = \max\{d(f, h)\} \quad (5)$$

where the f element is attributed to the C_r cluster and the h element to the cluster C_s . In such a case, the distance between two clusters is given by the longest link value. At each phase of the spatial hierarchical grouping, the clusters C_r and C_s , for which $d(C_r, C_s)$ is maximum, will be coupled. Figure 5 presents a graphical interpretation of the distance between the clusters.

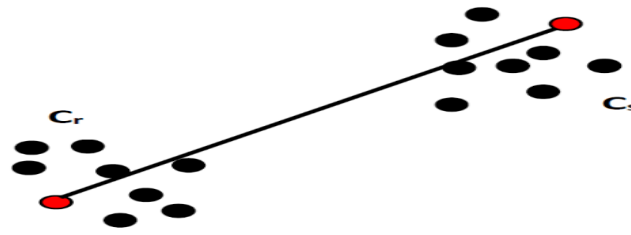


Figure 5. A representation of the maximum distance between two clusters.

c. The average distance method

In this method, the distance between two clusters is defined as the average of the distances between all element pairs, where each pair contains one element from each cluster. The average distance $d(C_r, C_s)$ is calculated with the relation:

$$d(C_r, C_s) = \frac{T_{r,s}}{(n_r \times n_s)} \quad (6)$$

where $T_{r,s}$ is the sum of all possible distances between the elements of C_r cluster and the elements of C_s cluster and n_r and n_s represent the number of elements in the cluster C_r and C_s , respectively.

At each stage of the clustering process, the C_r and C_s clusters for which the distance $d(C_r, C_s)$ is minimum, are coupled. Figure 6 illustrates how the average distance is defined.

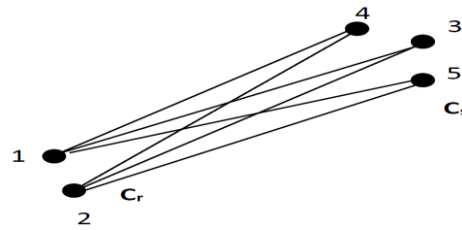


Figure 6. Representation of the average distance between two clusters.

d. Center of weighted method

In this technique, the distance between two clusters is described as the square Euclidean distance between their centers of weight.

$$d(C_r, C_s) = \|\bar{x}_r - \bar{x}_s\|^2 \quad (7)$$

where $\bar{x}_r$ and $\bar{x}_s$ are the mean vectors for the C_r and C_s clusters.

This method is much more robust, deviating from the average more than other methods of hierarchical clustering, but in other situations may not give as good results as the Ward method or the average distance method.

e. Ward method

This method seeks to form $P_n, P_{n-1}, \dots, P_1$ partitions in a way that minimizes the information loss associated with each cluster and measures them in an easy-to-use interpreted form. At each step of the analysis, two clusters are combined, the fusion of which leads to results that minimize the increase in “lost information”. The lost information is defined by Ward in the conditions of the criterion of the square sum of the error.

$$d(C_r, C_s) = \frac{\|\bar{x}_r - \bar{x}_s\|^2}{\frac{1}{n_r} + \frac{1}{n_s}} \quad (8)$$

where $\bar{x}_r$ and $\bar{x}_s$ are the mean vectors for the C_r and C_s clusters and n_r and n_s represent the number of elements from the C_r and C_s clusters, respectively.

f. K-means method

This method involves a simple and easy mechanism to classify the input data set into several K clusters (K fixed a priori). The basic plan implies defining K centers of *weight*, one by one for each group. These centers of *weight* must be rationally fixed because different locations lead to different results. The best choice is to fix them, if possible, as far apart as possible from each other. The next step is to take each element of the input data set and link it with the closest center of *weight*. The first grouping stage ends when there are no more ungrouped items. At this point, it is mandatory to recalculate new K centers of the groups arising from the previous phase. The process continues until the positions of the new centers no longer change significantly.

The objective of the method is to minimize an objective function (quadratic error function) given by the expression:

$$J = \sum_{k=1}^K \sum_{l=1}^{n_k} \|x_l^k - c_k\|^2 \quad (9)$$

where $\|x_l^k - c_k\|^2$ is the distance measured between point x_l^k , $l = 1, \dots, n_k$, where n_k is the total number of components of the k cluster, and the center of the group c_k , $k = 1, \dots, K$.

2.4. Validation of Results

The evaluation of the results obtained from the clustering process is the main concern of cluster validation. At this stage, the density, size and shape, separation of clusters, and robustness of classification were examined. The literature mentions the following tests to validate the clustering process [39,40,44,45]:

- External tests—data not included in the basic ranking are compared with the categorization results of the input data.
- Internal tests—only input data are utilized to evaluate the classification's quality; each cluster's separate validation is carried out using this test.
- Relative tests—this approach takes into account several classifications of the database, the results being analyzed using the same clustering method, but with various input data.

Internal cluster validation tests are more common and effective in real-world settings. Testing based on the creation of a global silhouette index of clusters is one of them and is also one of the most popular. This test determines the average shape width for each cluster, the median shape width of each element, and the average shape width of the entire collection of input data. With this method, each cluster might have a “shape” that is based on comparing its separation and density. The clustering procedure is then validated using the shape's average width, and the ideal cluster number will also be set using the same information.

$$GSI = \frac{1}{N_k} \sum_{k=1}^K F_k \quad (10)$$

F_k , the local silhouette coefficient, is calculated using the relation:

$$F_k = \frac{1}{r_k} \sum_{l=1}^{r_k} f_l \quad (11)$$

f_l , the silhouette width coefficient for element l , is calculated by:

$$f_l = \frac{b_l - a_l}{\max\{b_l, a_l\}} \quad (12)$$

where a_l is the average distance from element l and the elements from cluster k and b_l is the minimum median distance from component l and the components in the closest cluster k .

In Equation (12), if the element l is unique inside a cluster, then $f_l = 0$. The literature proposes the following explanation of the GSI coefficient [38,40]:

- 0.71 to 1.00 (a strong structure was highlighted);
- 0.51 to 0.70 (a reasonable structure was obtained);
- 0.26 to 0.50 (the structure is weak and might be artificial);
- <0.25 (no substantial structure was noticed).

Verifying the clustering process quality is one of the main steps in analyzing the inherent database characteristics. Its purpose is to evaluate the results of the process clustering and select the schema that best fits the elements in the database.

2.5. Electric Load Profiling

The loads from the nodes within the electrical networks range in consumption time and place. Consequently, distribution operators (ODs) need details regarding the load of fed consumers so that they will be able to optimally plan and operate the network and ensure proper power supply and operation modes, load management, and proper billing [40,46–48]. The load demanded by consumers depends on various parameters such as:

- Consumer type: consumption type, with/without electric heating, or size of the building;
- Time factor: time of day, weekday, and month;
- Climatic factors: humidity, temperature, cloudiness, wind speed, etc.;
- Other electrical charges related to the analyzed load.

For a certain consumer, his behavior is determined by a load profile correlated with the electricity consumption for each interval. The accessibility of this data is dependent on the type of consumer. In general, small consumers (such as residential ones) have an uncertain behavior, because the implementation of smart metering on a large scale would lead to large investments whose recovery time from their energy consumption would be too long. For these consumers, there is only the consumption of electricity at certain periods of time each year. For large consumers (such as industrial consumers), the installation of smart meters is facilitated by advantages related to billing (done every month) and high electricity consumption (justifying the investment by the fact that the detailed recording of consumption allows for the application of differentiated tariffs varying with consumption period).

In the traditional strategy for a distribution system plan, load profiles are employed to evaluate the maximum necessary load, in correlation with the simultaneity coefficient of the consumers coupled to the network node. Despite the fact that this strategy is appropriate, some major disadvantages arise:

- There are inherent inaccuracies, due to the simultaneity coefficients, which must be highlighted;
- The energy consumption and losses calculation does not have an increased precision;
- The voltage within the network nodes from various hours is not known;
- The load profiles of nodes having arbitrary variations in power requirement cannot be accurately modeled or evaluated.

Utilizing modern techniques of load analysis, load forecasts and the calculation of power and energy losses (for any period) can be achieved. The use of load profiles has some advantages [40,49]:

- It is not necessary to estimate or calculate concurrency coefficients, as load profiles already include the information;
- The energy consumption and power/energy losses calculation can be correctly achieved at any point within the network;
- The main voltage and charging are known for any period;
- The optimal position of the transformer plot can be determined, both for the peak load period, as well as for other times of the day;
- The effect of overloading or increasing the load is modeled more accurately than in the traditional method.

Medium voltage (MV) and low voltage (LV) networks, mainly urban ones, have many nodes even if only the nodes where electric distribution substations (EDSs) are located are taken into account. Thus, monitoring the system consumption for every node can become overwhelming, sometimes virtually impossible. This is practically overcome if, in the organization studies corresponding to the networks, the load-type profiles are associated with the node groups.

2.6. Load-Type Profiles Correlated with Nodes in Electrical Distribution Networks

This paragraph presents an approach based on hierarchical clustering techniques (presented in Section 2) for calculating the load-type profiles (LTP) correlated with nodes' high voltage (HV) and medium voltage (MV) electrical distribution networks. Through knowledge of the load profile of the nodes, the OD can clarify the procedure for estimating the requirement in a certain sector.

In this respect, it is mandatory to understand the daily loading profiles. The load diagram of the nodes must be reconstructed by applying the standard load profile and daily demand. This standard profile is characterized by 24 or 48 load values.

The type and season influence the shape of load profiles. Because many profile responsibilities are correlated with various nodes of the network and may complicate the problem, they must be grouped into clusters, taking into consideration some similarities among them. For every cluster, the typical load profile can be established.

In this sense, all the measurements performed must be processed, by arrangement and normalization, using a convenient normalization factor (average power, peak power, or more frequently, the energy consumption of the studied period):

$$p_i^h = \frac{P_i^h}{\sum_{h=1}^T P_i^h}, i = 1, \dots, N \quad (13)$$

where p_i^h is the normal value of the power in i node at h hour, P_i^h is the real value of the power in i node at h hour, and $\sum_{h=1}^T P_i^h$ is the total energy consumption in the interval T (24 h).

It is worth mentioning that after the clustering procedure is applied, clusters can be acquired as coherent and representative so that the diagrams within the same cluster are similar [40,49,50]. In the end, every cluster will be related to a typical task profile, estimated using the graphs median.

$$m_{C_k}^h = \frac{\sum_{i=1}^{N_{C_k}} p_i^h}{N_{C_k}}; h = 1, \dots, 24; k = 1, \dots, N_K \quad (14)$$

where N_K is the cluster numbers derived from the node's classification in concordance with the absorbed load (active power) and N_{C_k} is the node number from each cluster C_k , $k = 1, \dots, N_K$.

2.7. Load-Type Profiles Correlated to Low-Voltage Consumers

In recent years, distribution operators have increasingly used smart metering systems (Smart Metering System) to monitor the electricity consumption of consumers. The development of these systems begins with technology selection and planning for installation, operation, and maintenance. In general, the implementation of residential or non-residential consumer categories is quite discrete in the case of many countries in the European Union. Currently, there are two alternative solutions to solve the problem of metering electricity consumption for consumers.

The first solution envisages the installation of smart meters for all consumers, being expensive and uneconomical, but the most accurate. The second solution envisages the continued use of traditional meters and the attachment of load-type profiles to the monthly energy consumption, which is then distributed over days and hours.

For large consumers, in the category of industrial ones, the presence of smart meters is necessary for several reasons: invoicing is done every month, the consumption of electricity is high (which implies rapid amortization of the investment), and it provides a detailed recording of consumption necessary for the process of invoicing.

This section presents an algorithm for determining the type of load profiles associated with LV consumers according to the category of energy consumption in which they fall. Consumption categories are identified from historical information and can be updated following changes in consumer behavior.

The algorithm phases are [40]:

Phase 1. In this phase of load analysis and database formation, a representative specimen must be identified from the crowd of consumers who have installed smart meters and the sampling step for the purchase of load schedules must be described. Consequently, a database with the registered load program and the consumer category is created.

Phase 2. Technical problems in the pre-processing of the data may affect the quality of the database in real cases of monitoring consumer charges, requiring many meters spread over a large geographic region over an extended period. The most relevant and frequent problems are communication issues, interruptions, meter failure, and, occasionally, the irregular behavior of individual consumers. These issues will influence the records of the database, appearing as null values or exceeding a particular threshold set by the connection notice. Such records must be identified and working techniques must be applied to obtain the missing data, resulting in the substitution of missing or equal to zero data with some estimated values so the database can be made ready to obtain clusters.

Phase 3. The database with records of load schedules must be divided into clusters of consumption, defined by the consumer's type: residential, commercial, or industrial.

Phase 4. This phase represents the clustering procedure. A hierarchization of clusters is achieved, taking into consideration the daily consumers' energy consumption within every consumption macro-category. In this respect, the K-media clustering technique is applied. For every cluster, the representative load profile is determined by applying the average of the load graphs' hourly values.

Phase 5. Task-type profiles must be assigned. For every customer's class, a typical task profile is assigned as a function of the activity macro-category to which the consumption belongs.

The suggested algorithm was evaluated using a database with 296 load diagrams. After the macro-categories were divided, 147 consumers were distributed in the residential consumers class, 97 in the commercial consumers class, and 52 in the industrial consumers class.

The utilization of the k-means clustering technique within every activity macro-category resulted in five clusters within the residential consumers class and three clusters within the commercial and industrial consumers ones.

Using these standard profiles, consumers can be better delineated in connection with load modification than the standard profiles correlated to the complete activity macro-category. This detail can be emphasized if a correlation is achieved between the load-type profiles of every activity macro-category and the load-type profiles linked to these macro-categories (Figure 7).

The change in consumers' energy consumption determines important problems in the planning of activities associated with the technological procedure, in terms of adopting the optimal power supply and operation solutions. Solving these issues can be efficiently achieved by utilizing consumption profiles correlated to energy carriers.

This section proposes a perspective based on clustering methods for the determination of load-type profiles for electric vehicle charging networks. The standard profile's forms, which take into account both the type of electrical equipment or installations supplied and the time of year the survey is conducted, represent the specifics of how consumers use electricity.

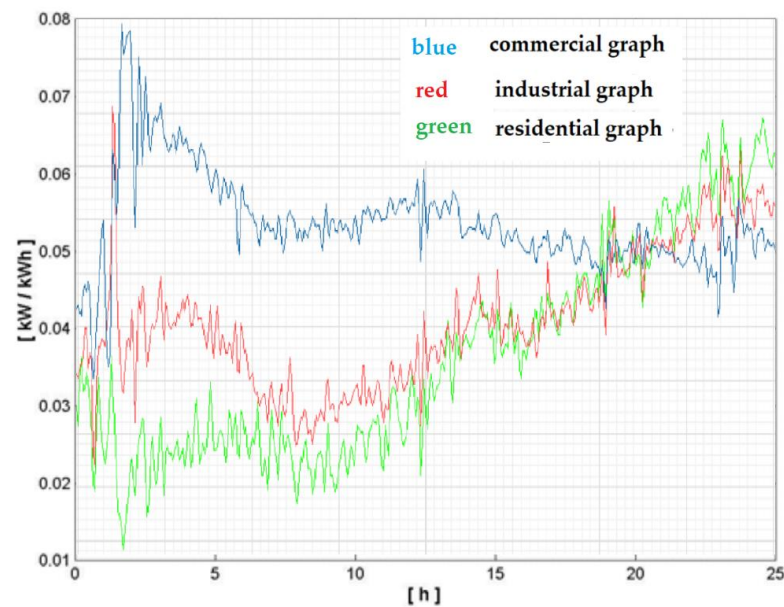


Figure 7. Load-type profiles of activity macro-categories.

The methodology proposed for the load profiling process for electrical vehicles as (industrial) consumers is presented in Figure 8 [40].

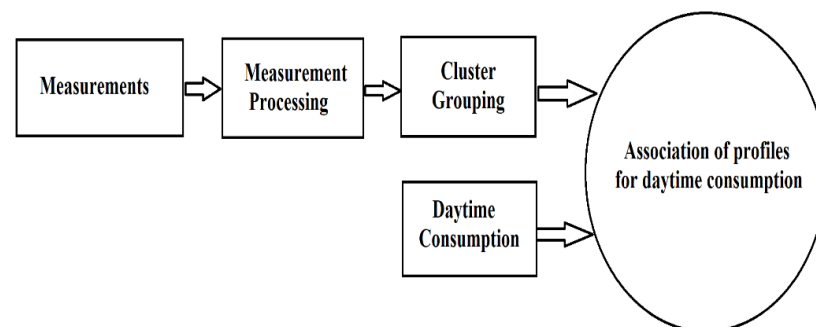


Figure 8. Scheme for the determination of load-type profiles in electric vehicles considered as industrial consumers.

The importance of every phase from Figure 7 is the same as for all profiling procedures that are operated with clustering methods from the previous paragraphs.

2.8. Distribution Network Loading Simulation

In the distribution network, the basic component represents the loading simulation of MV/LV transformers. Because tens of thousands of transformers are positioned in the distribution matrix, their hourly load is hard to determine, due to the many distribution networks, including current and voltage sensors mounted in devices, transformers, and MV connections unequipped with recording meters with transmission capacity, remote, real-time recording, and load level. Consequently, it is hard to identify those transformers operating at overload or to estimate the loads of MV connections intended for transfer between distributors without a simulation.

The most effective route to evaluate a load of transformers, without performing real measurements, is represented by the utilization of simulation programs, with some of the following elements being taken into consideration [40]:

1. The number of consumers coupled to every transformer;
2. The consumer type;
3. The annual energy consumption for every consumer;

4. Task type diagrams correlated with every consumer class;
5. Software able to calculate the load of transformers.

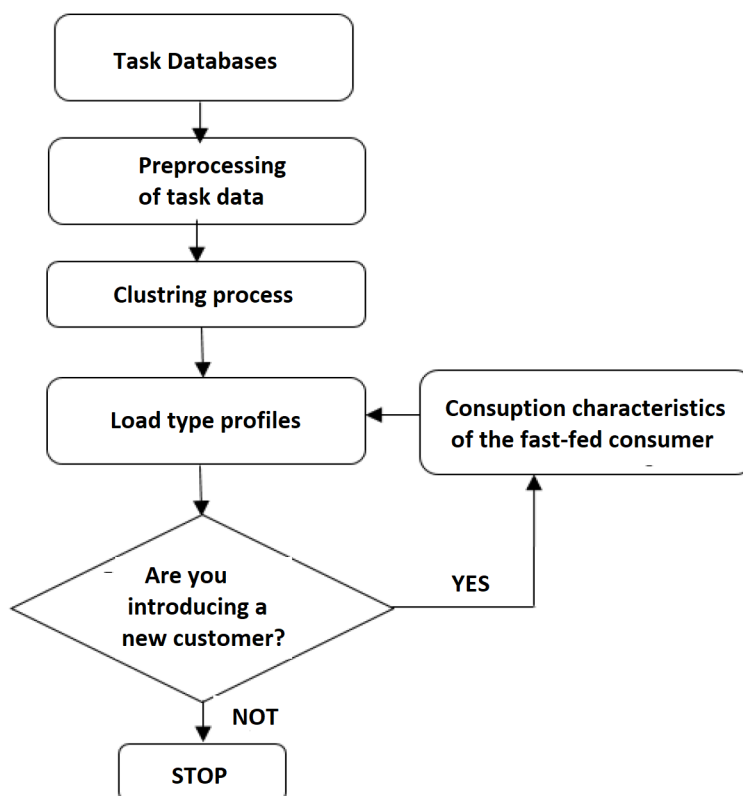
It must be emphasized that, when a load of transformers is estimated, the maximum and hourly active powers (within low voltage/medium voltage side) are computed on peak days and varying typical intervals (winter, summer, average working days of the week, etc.).

2.9. MV/LV Distribution Transformers Load Simulation by Clustering Procedure

The database structure that is necessary to simulate the load of the distribution transformers in the MV/LV substations includes:

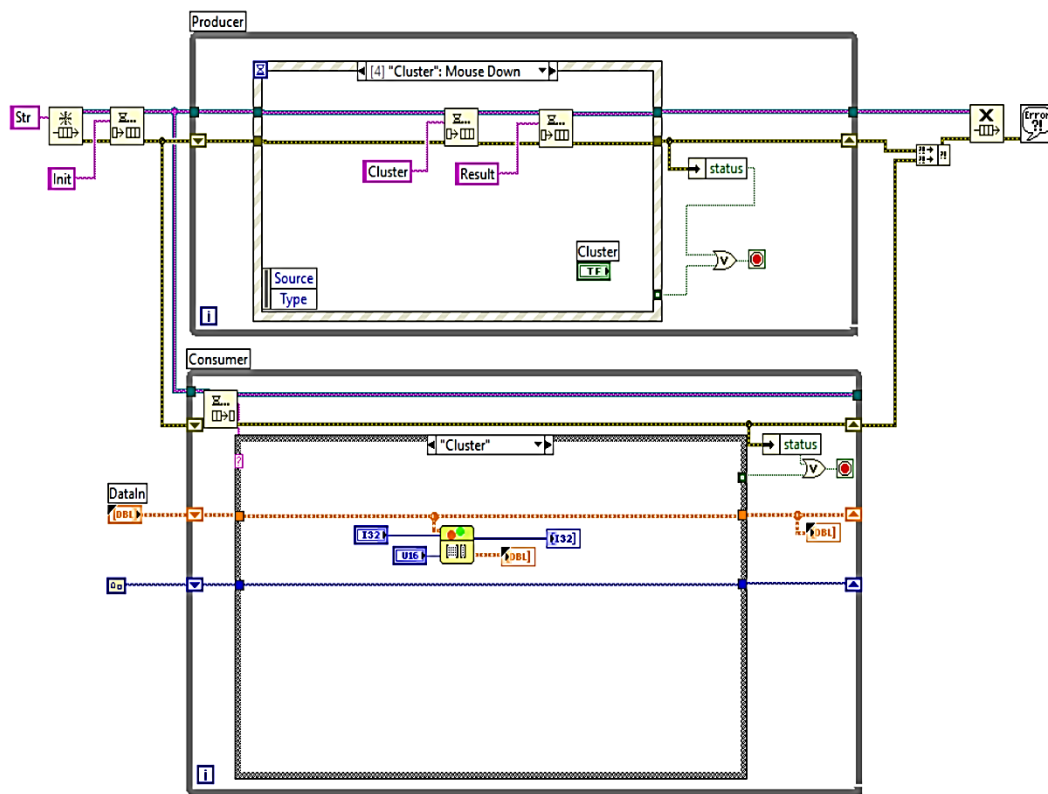
- The LV database having a “consumer link-substation (CS)” (number and class of consumers from every transformer in CS);
- Basic profiles containing task-type profiles for all consumers classes;
- A consumption database containing data on the annual energy consumption and consumer class.

In order to establish the typical load profiles correlated with consumers in LV networks, a database is required to include as many registered load diagrams as possible, representing all consumption categories. The decisional procedure of associating a typical task profile with a certain consumer constitutes a complicated problem. Consequently, a load profiling algorithm using clustering methods has been suggested for residential and non-residential consumers. Figure 9 shows the implementation diagram for determining load-type profiles for residential and non-residential LV consumers [40] and the software pseudocode. In the following figures, the software implementation for the proposed algorithm is represented. The software development was carried out in the LabVIEW programming language developed by National Instruments.

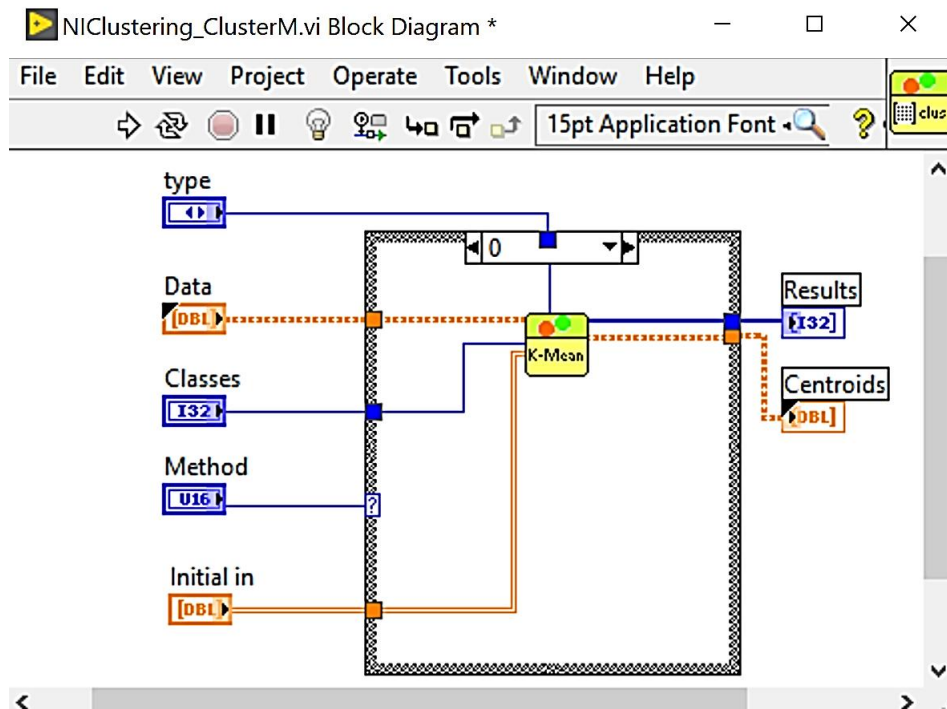


(a)

Figure 9. Cont.



(b)



(c)

Figure 9. Cont.

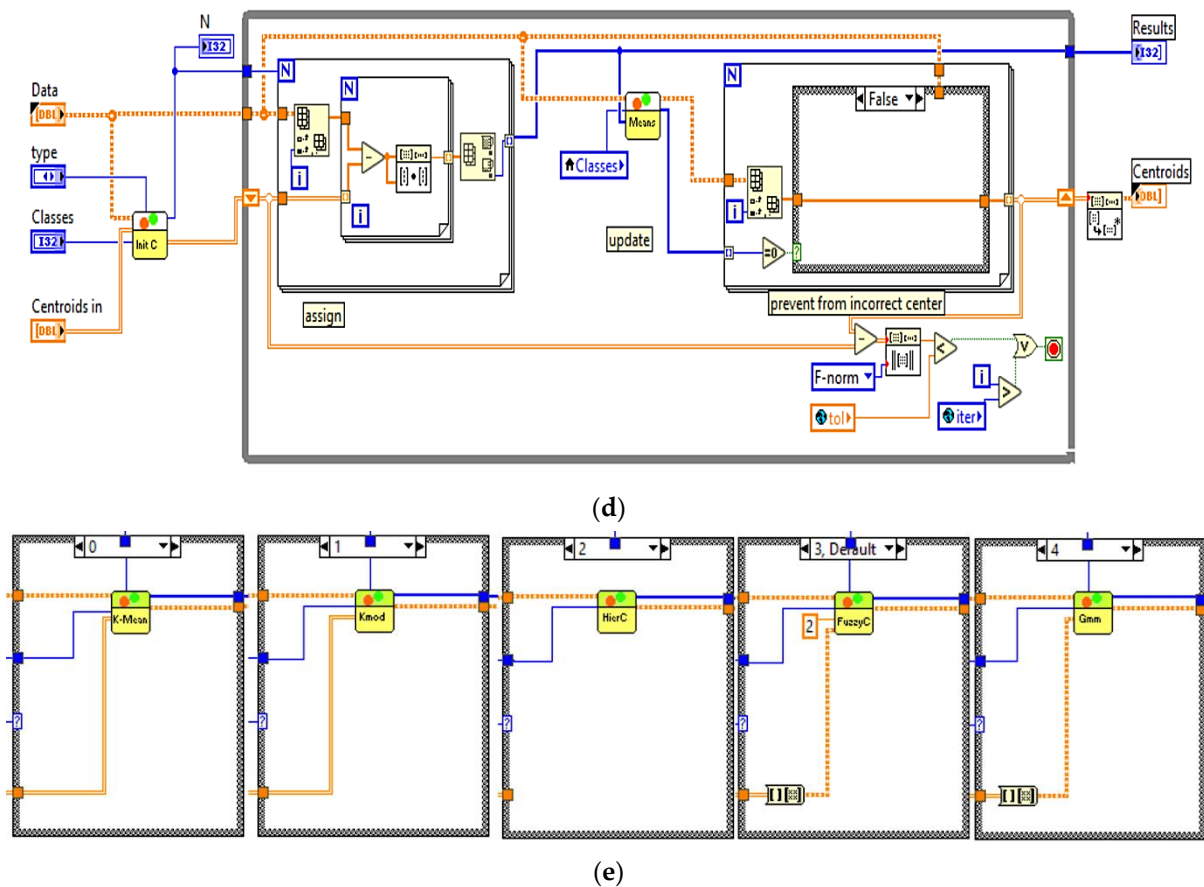


Figure 9. Implementation diagram for determining the load-type profiles for residential and non-residential LV (a) consumers and the software pseudocode: (b) is master code, (c) implementation code for k-means, (d) implementation code for hierarchical cluster and (e) represent software modules for all clustering methods.

3. Results

An important issue in the optimal operation and planning of electric distribution networks by electric companies is the estimation of the maximum load of consumers. This problem occurs especially with home consumers. The decision to determine the optimal number of consumption categories within the same class of consumers and the estimation of the maximum load is a complex problem. Therefore, an improved regression method with clustering techniques is presented. Within the proposed method, to define the consumption categories of consumers, a classification can be made considering the monthly energy consumption and the maximum load (data obtained from the recorded load graphs). After identifying the consumption category, the maximum load of each consumer is estimated using a regression model corresponding to each consumption class. The major advantage of the proposed method is the exploration of the data using clustering techniques in order to obtain models/patterns/categories of consumption.

The contributions to modeling loads in electrical networks, the effectiveness of clustering techniques in their profiling, and the load simulation for MV/LV distribution transformers using clustering procedures are highlighted in the study.

3.1. Estimating the Maximum Load of LV Users (Consumers) Using Clustering Techniques

The estimation of the maximum load of users is a key parameter in the effective operation and planning of electricity distribution networks by energy providers. This problem occurs especially in household consumers, but also in the integration of electric vehicle charging. From analyzing the built databases and using data exploration techniques, load

patterns/models can be identified that can be extrapolated to other potential consumers such as electric vehicle charging networks.

The decision to estimate the appropriate number of consumption categories within the same consumer class and the maximum burden is a complex issue. Thus, a customized regression method using clustering techniques is presented.

The monthly energy consumption, as well as the maximum load, can be classified using the suggested approach to determine the consumption groups of users (data obtained from the recorded load graphs). After determining the consumption category, a regression pattern matching every consumption class is used to estimate the maximum load of each consumer. The main benefit of the suggested approach is the data exploration utilizing clustering techniques to produce models, trends, and consumption categories that would aid in consumer modeling.

Its stages are presented below [40]:

Stage 1. Database. For consumers in the selected pilot area where smart meters have been installed, the load graphs for the analyzed time period are recorded. For each consumer, the variables that characterize the consumption category (maximum load, monthly and annual energy) are extracted.

Stage 2. Data pre-processing. All records from the database are analyzed, and those that contain missing values, equal to zero or atypical, are subjected to the process of treatment with missing data techniques. After all processing, the data are used to classify consumption categories.

Stage 3. Data exploration process. The first step in this stage is the use of clustering techniques in order to obtain consumption categories according to the maximum load and monthly energy consumption. The K-media technique is utilized for the clustering process. Then, a regression model is built for each consumption category to estimate the maximum load related to consumers.

Step 4. Estimation of peak load. The regression models obtained in Stage 3 are used to estimate the maximum load absorbed by the monitored consumers by means of classical meters.

The structure of the database includes information on the type of consumers (residential, industrial, commercial, and public) and their daily/monthly/annual consumption respective to the maximum load absorbed. This information can be obtained using smart meters. The information acquired is represented by the load curves that describe the consumer's behavior during the day. The processing of the load curves allows us to determine the data on energy consumption respective to the maximum load absorbed by each consumer. For the method implementation, the database was split in two, a working base and a testing base. The application of the working base will lead to mathematical regression models able to conduct the estimation of the maximum load for each type of consumer within the same category of energy consumption. Energy consumption categories are then acquired using the K-average clustering procedure for each consumer type (residential, industrial, commercial, and public). The outcomes of the clustering technique are confirmed using a group quality assessment, based on the shape coefficient. The method is then tested to estimate the maximum load absorbed by consumers belonging to the testing base.

The method was applied using an initial database containing records of load schedules for 1160 household consumers located in a countryside region over a one-month period. For each consumer, the characteristic variables related to the peak load ($P_{\max}$) and monthly energy consumption (W_{luna}) were extracted.

In Stage 2, the analysis of load graphs and characteristic variables led to the elimination of 15 consumers due to zero energy consumption. Thus, in the next phase, the database consisted only of the associated information for 1145 consumers. This base was split in two, a working and a testing base. The size of the working and testing bases varies in the literature and primarily depends on the total amount of records in the original database. The split between the two bases ranges from 90/10 to 60/40. In this instance, the working

database contained 814 consumers (representing 66 percent of the database), leaving the testing database with the remaining customers (331 consumers, 33 percent of the database).

In Stage 3, the data exploration process was initialized through the K-media clustering method to obtain the consumption categories in which the consumers in the work base will be integrated.

The first phase of the K-average method consisted of the determination of the maximum cluster number (consumption categories) with the relation:

$$C_{k\max} = \sqrt{N} = \sqrt{841} = 29 \quad (15)$$

Then, for each $K = 2, \dots, 29$, the K-averages technique was applied. For the clustering processes initiated for the K value of the number of clusters, the quality of the grouping process was estimated by calculating the global silhouette index (GSI). The values of the shape index obtained for each $K = 2, \dots, 29$ are represented in Figure 10. This grouping is the best result of the clustering technique since the GSI has the greatest value for $K = 5$. The results for the global silhouette index ($GSI = 0.775$) show that the clustering technique was quite effective. The graphical representation of the clusters obtained in the case $K_{optim} = 5$ is illustrated in Figure 11.

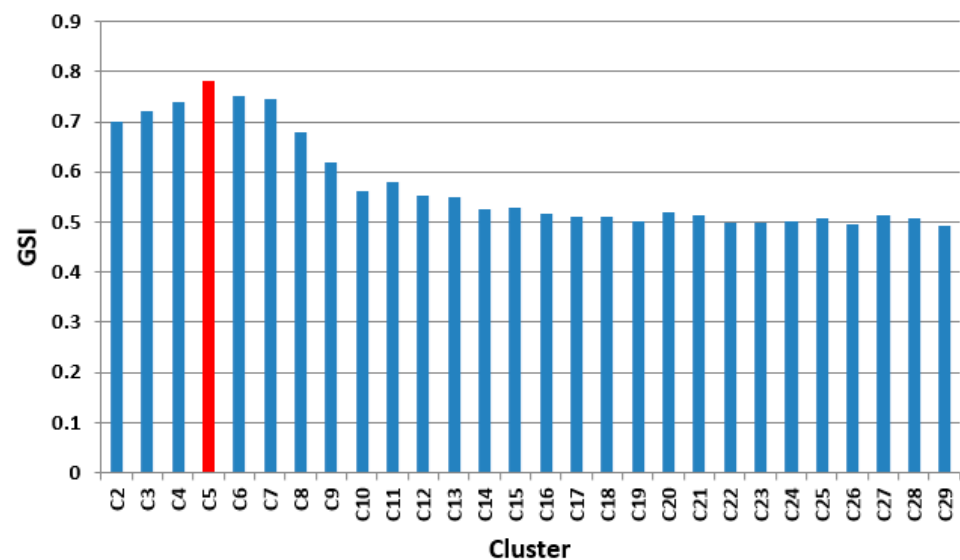


Figure 10. GSI values for cluster ranges between C2 and C29.

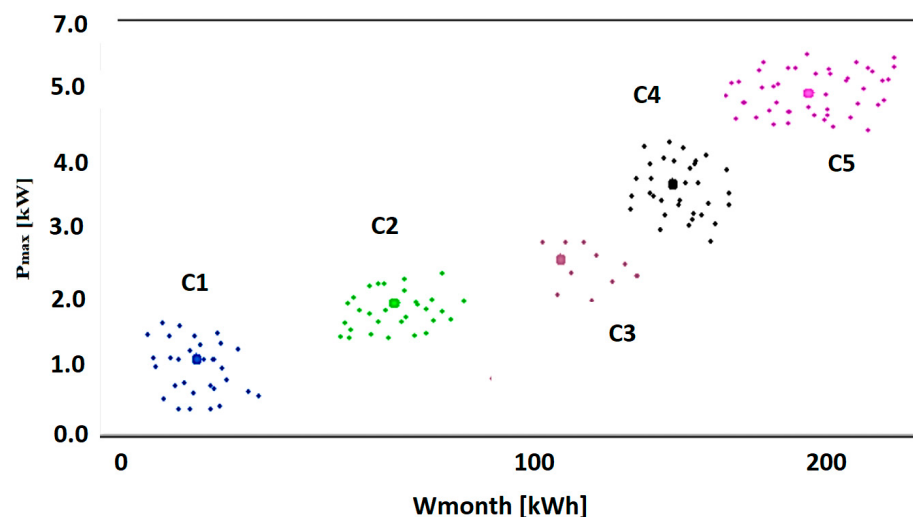


Figure 11. Graphical representation of clusters for the case $K_{optim} = 5$.

The characteristic variables represented by the mean values (m) and the dispersion (σ) corresponding to the maximum load ($P_{\max}$) and the monthly energy (W_{month}) for each cluster (consumption category) are summarized in Table 1.

Table 1. Statistical variables of the characteristic variables associated with the consumption categories.

Consumption Class	Number of Consumers	$P_{\max}$ (kW)		W_{month} (kWh)	
		m	σ	m	σ
C1	31	1.51	0.32	192.20	13.66
C2	104	1.06	0.13	115.62	11.83
C3	387	0.43	0.09	30.70	12.93
C4	279	0.87	0.17	88.64	11.83
C5	12	4.78	0.92	252.96	19.70

The analysis of data from each consumption category indicated a correlation between the maximum load and the monthly energy consumption that can be mathematically modeled by applying regression models. In Figures 12–14, regression models are represented based on first- and second-degree polynomials for every consumption class.

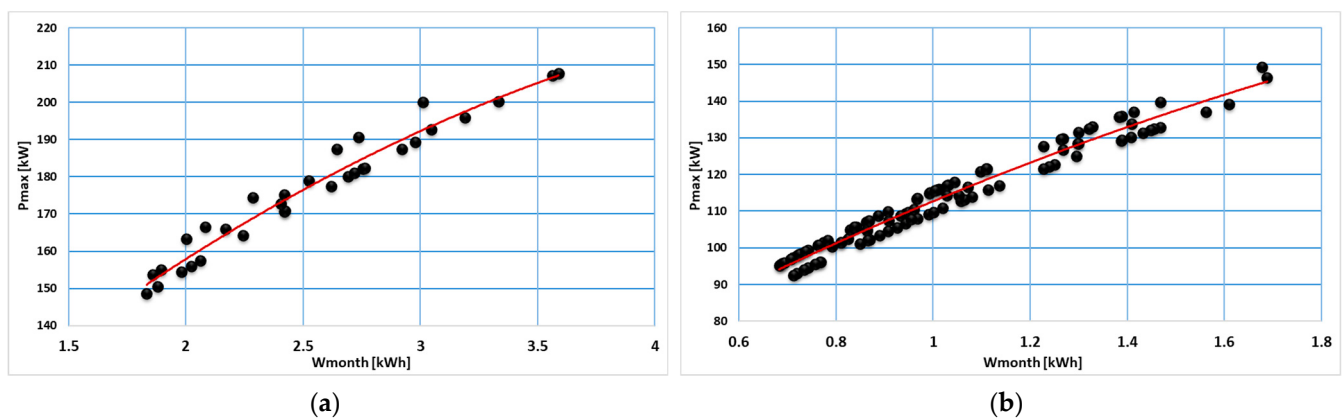


Figure 12. Regression models correlated to consumption class C1 (a) and C2 (b).

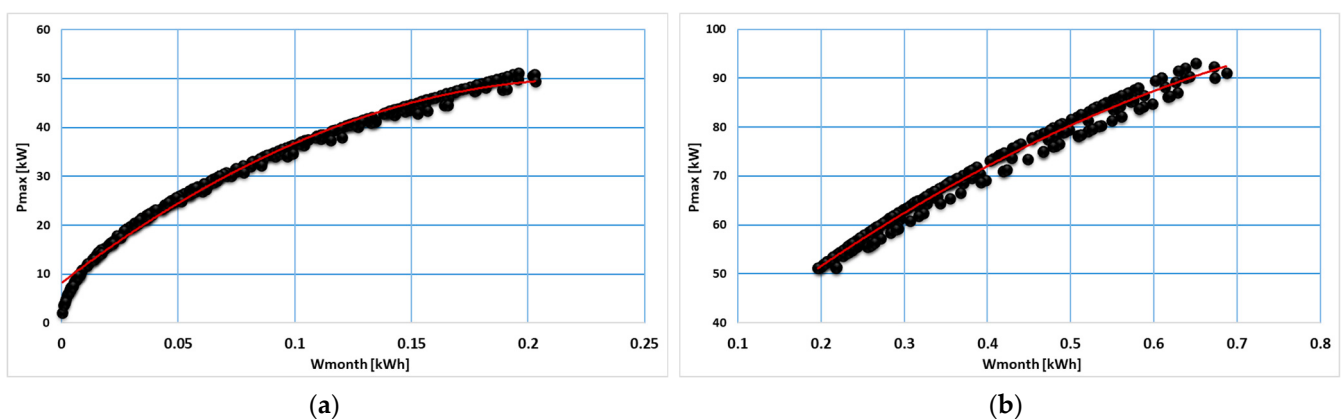


Figure 13. Regression models associated with consumption class C3 (a) and C4 (b).

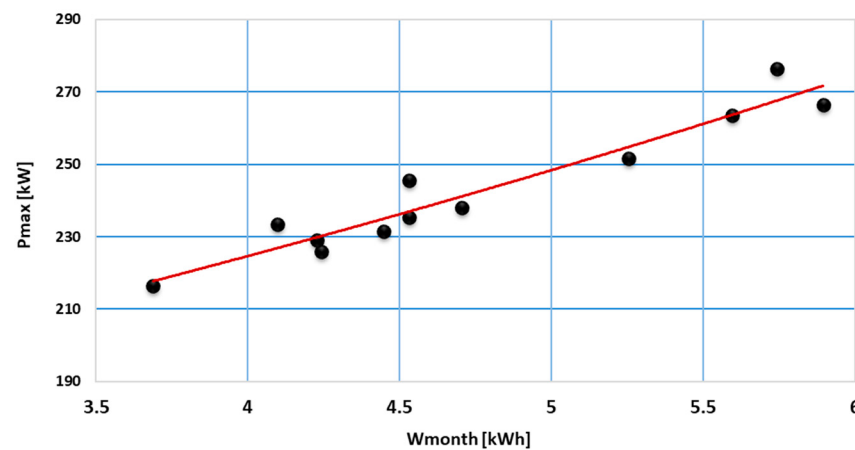


Figure 14. Regression models associated with consumption class C5.

Regression models based on the degree II polynomial led to a better approximation than the degree I polynomial. Thus, regression models based on the degree II polynomial were adopted to determine the $P_{\max}$ (W_{month}) dependence; coefficients of the regression models determined for each consumption class are summarized in Table 2.

Table 2. Regression model coefficients associated with consumption categories.

Consumption Class	$a (\times 10^{-5})$	$b (\times 10^{-3})$	c
C1	13	22	2.6
C2	6.2	3.1	0.7
C3	7.3	0.023	0.00008
C4	8.2	0.45	0.027
C5	15	105	17

In Stage 4, for all consumers in the test base, the maximum load is estimated by applying the regression model of the consumption class associated with every consumer, depending on the monthly energy consumed.

Figure 15 presents the real and forecasted values associated with the maximum consumer tasks within the test base, assembled according to the consumption class.

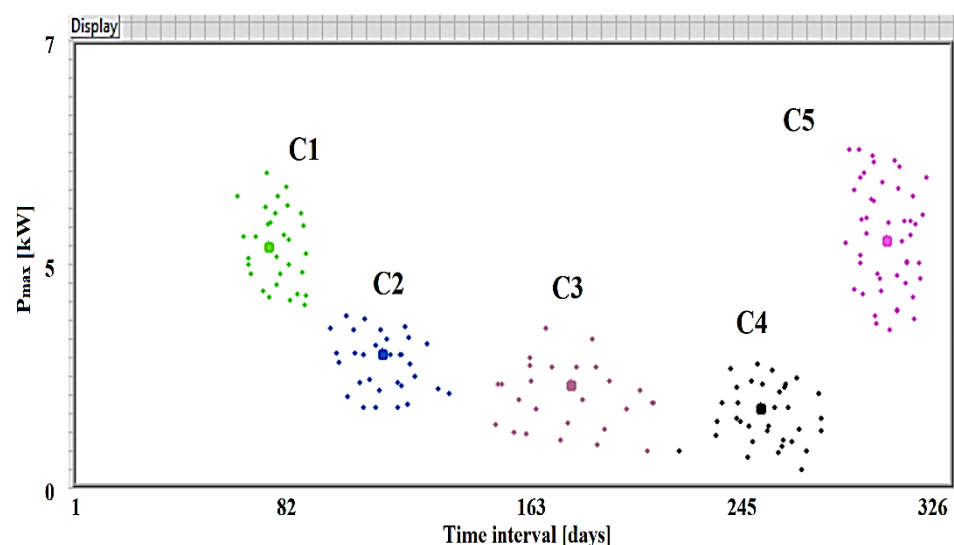


Figure 15. Real and estimated maximum load values of the clustering testing base (C1–C5 represent the clusters used for the analysis).

Of the 331 testing base consumers, 206 (approximately 62%) forecast errors were $\leq 3\%$, 77 (approximately 23%) errors were between 3 and 7%, and 48 (approximately 15%) errors were between 7 and 10% (Figure 16); thus, the average estimation error in the testing base was 4.01%. This value is reasonable in the context that most consumers do not have permanent monitoring.

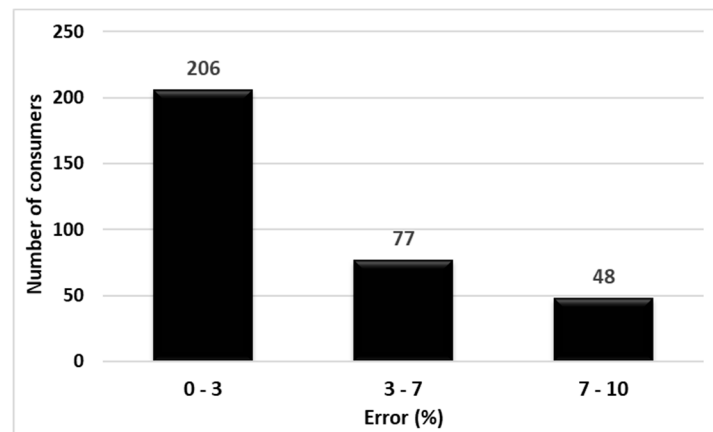


Figure 16. Classification of maximum load forecast errors for testing base consumers.

3.2. Simulation of a Load of MV/LV Distribution Transformers by Clustering Procedure Application

The future power supply networks of electric vehicles will be powered by medium and low-voltage networks.

The following algorithm is suggested for simulating the load of the distribution transformers from the MV/LV transformation stations [40].

Phase 1. Database: representative consumers with smart meters will be sorted out from the database. Typical load diagrams will be recorded for every consumer, then, the main characteristic variables will be detached, i.e., minimum ($P_{\min}$) and maximum active power ($P_{\max}$), daily energy (W_z), and consumption class.

Phase 2. Pre-processing load diagrams: all records involving missing data or values will be excluded or subjected to processing. After being pre-processed and reduced, the results will be applied to obtain the classification into consumption categories (clusters) using the clustering procedure.

Phase 3. The division into consumption macro-categories: the database with the records of load schedules will be split into clusters described by the consumer's type: residential, commercial, and industrial.

Phase 4. Clustering procedure: a clustering technique will be applied for load-type profile determination to determine the optimum results. In the end, a typical load profile for each consumption class will be obtained by applying the average of the hourly values for the load diagrams.

Phase 5. Determining the load profiles: a typical load profile will be attributed to every consumer's class, depending on their consumption class.

Phase 6. Estimating the load of the MV/LV transformer: a simulation protocol will then be proposed based on the Equation [40]:

$$P^h = \sum_{k=1}^{C_k} n_k W_{med\ k} p_k^h + \sqrt{\sum_{k=1}^{C_k} n_k (W_{med\ k} \sigma_k^h)^2}, h = 1, \dots, 24 [\text{kW}] \quad (16)$$

where:

P^h is the MV/LV transformer load from the transformer station at h , (kW);

n_k is the consumer number, where k is the consumption class;

$W_{med\ k}$ is the average energy consumption, where k is the consumption class, (kWh);

p_k^h is the hourly coefficient of transformation for energy consumed by the consumers, (kW/kWh);

σ_k is the standard deviation of the power distribution necessary to the cluster consumption (kW/kWh);

C_k is the cluster number (consumption class) correlated to the consumer's feed.

The weight center technique was applied to evaluate the load-type profiles through a database containing 180 load curves registered by smart meters from residential consumers in a distribution system from the LV pilot.

To evaluate the representative load profiles, the gravity center was applied, taking into consideration the scattering of the LV pilot located in a region from Romania [40]. Every load graph is set by 48 hourly values correlated with consumer behavior over one day. Missing load curves or abnormal values of zero throughout the day must be eliminated from the procedure, with only 144 consumers remaining eligible. The clustering procedure resulted in five consumption classes (clusters) (Table 3).

Table 3. Consumption class characteristics.

Consumption Class	Number of Consumers	$P_{\max}$ (kW)		$P_{\min}$ (kW)		W (kWh)	
		m	σ	m	σ	m	σ
C1	15	0.21	0.02	0.06	0.01	3.41	0.82
C2	5	0.51	0.05	0.02	0.02	4.35	1.20
C3	22	0.46	0.12	0.03	0.01	2.98	0.54
C4	30	0.04	0.06	0.02	0.03	0.25	0.43
C5	72	0.17	0.06	0.03	0.02	1.95	0.41

It can be observed that the most representative class of consumption is C5 (50% of the total consumer number), the least representative being C2 (only 3.5%). The load-type profiles correlated to every cluster (consumption class) are depicted in Figures 17–19 and the consumer distribution is presented in Figure 20.

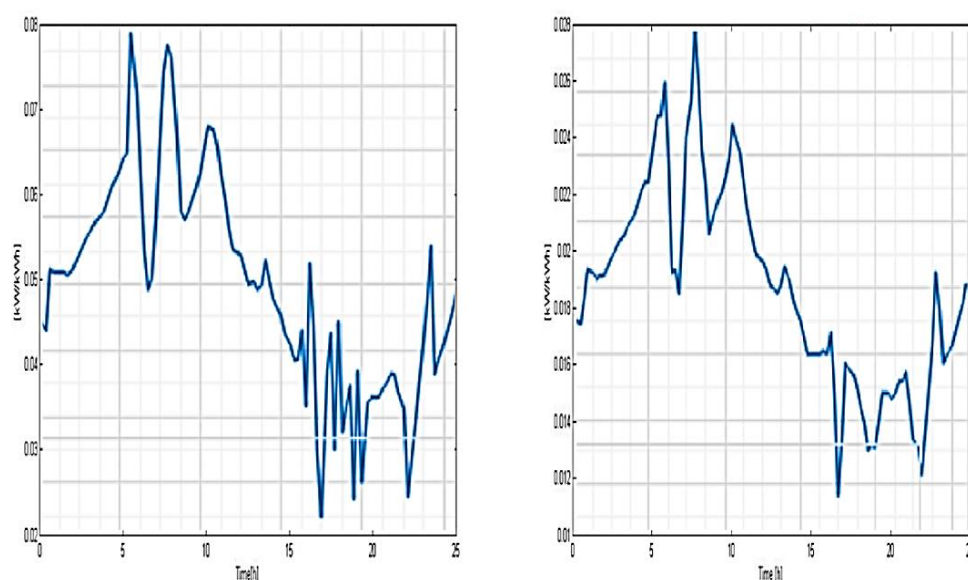


Figure 17. Load-type profile related to C1 and C2 consumption classes.

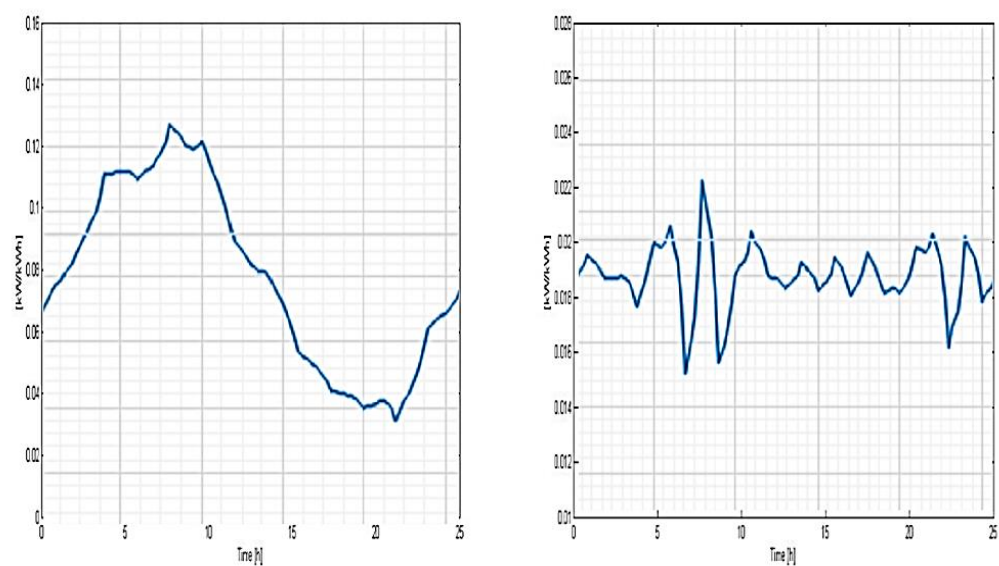


Figure 18. Load-type profile related to C3 and C4 consumption classes.

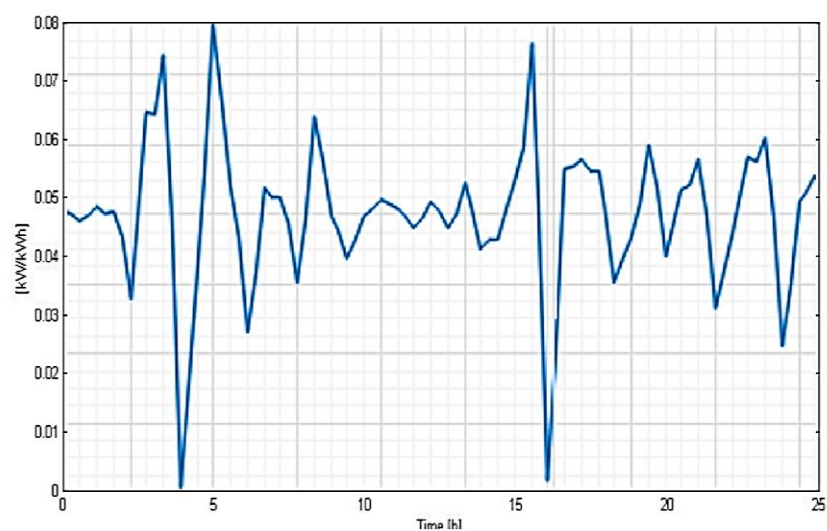


Figure 19. Load-type profile related to C5 consumption class.

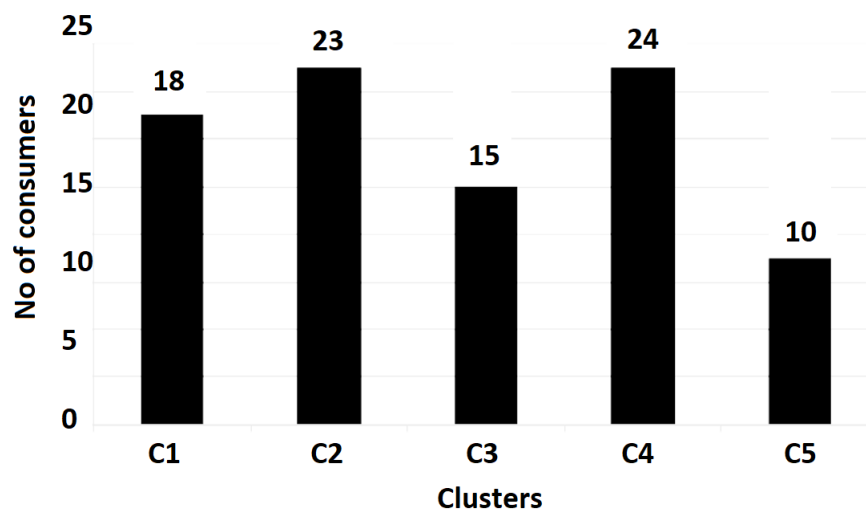


Figure 20. Consumer associations resulting from the testing base.

4. Discussion

Numerical simulations related to the methodologies, algorithms, and calculation programs developed in this paper have shown that the intelligent distribution of consumers in Smart Grid distribution systems can help smooth the charging curve that can lead to lower electricity prices and facilitate the integration of renewable energy sources, resulting in a much safer and more economical operation of Smart Grid networks. The authors in [37,38] perform an analysis of data extraction techniques from the perspectives of different technical approaches to achieve consumer profiles using direct clustering, indirect clustering, clustering evaluation criteria, and customer segmentation. The article [39] presents an approach to the consumer profile from the perspective of time series, and in [38], the issue discussed is approached with the Bayes model and k-means clustering. As can be seen from the literature, clustering algorithms are frequently used in the energy field for profiling consumers. The method proposed in this article combines grouping algorithms by clustering techniques and evaluation criteria for the clustering results using the regression algorithm with second-order polynomials (logistic regression). By separating consumer behaviors, the relationship between them can be simplified. The use of hierarchical clusters (hierarchical classification) can significantly reduce the influence of external factors (e.g., region, weather, time, day, and social activities) on classifier performance. The results of the case study showed that the model proposed in this paper achieves a better classification of electricity consumption. In addition, the technique presented in this article contributes to an overall improvement in the profiling of consumers, as the proposed method achieves a better classification using fewer training samples. The performance of the results presented suggests that the proposed data-based model can be used as an effective tool in real-time. The idea of load demand variability is key information for the load monitoring control unit, thus the proposed task prediction models will help energy management. Based on a more accurately forecasted load demand, different optimization techniques for demand response applications can be developed. In addition, the classification of the model proposed in this paper depends on the completeness and reliability of the data. By combining these two algorithms in the next stage, we plan to develop a supervised machine learning algorithm that will automatically determine the profile of consumers based on historical data and data acquired in real-time (data mining).

5. Conclusions

Following the proposed study that formed the basis of this article, some conclusions can be made:

- Urban areas have significant issues in several areas, including the economy, water supply, energy, buildings transit, environmental protection, and basic services as a result of the phenomenon known as “global urbanization”.
- Municipalities are encouraged to employ smart ideas and try various smart infrastructure approaches in order to address these problems, thereby becoming the future smart cities or “Smart Cities”.
- Urban transportation issues are a key component of the Smart City idea, and the approximately 7.2 million electric passenger and freight vehicles demonstrate that electrification of the transportation sector is the undeniable future of mobility.
- The restrains in regulations on the use of conventional fossil fuels in Europe and China caused the automotive industry to quickly realign to multiple EV and BEV models, with more than 442 new products being available at this moment, leading to an 87% drop in Lithium-Ion battery prices, per kWh, between 2010 and 2019. These factors all have contributed to the rapid growth of the number of EVs. When considering the above-mentioned factors, electric vehicles have emerged as one of the primary solutions for decarbonizing the transportation industry and using renewable energy sources to generate electricity. Their impact on electrical networks, however, cannot be disregarded. The quest for low-emission mobility around the globe is expected to drive a major increase in the electrification of road transportation in the next decades.

The volume of the world's electricity demand may shift as a result of the rise in electric vehicles, posing serious problems for the infrastructure supporting electricity production, transmission, and distribution.

- The integration of too many EVs will significantly impact the electric power systems; however, by coordinating EV charging, flexibility services in the electric power network can be achieved, and the required investments in infrastructure can be kept to a minimum level.

Future study directions suggested to continue the research outlined in this paper include:

- Participation of EVs or charging stations equipped with converters that use power electronics in reactive power regulation services for EVs.
- Offering support services by coordinating EV charging via LV power grids.
- Variations in the voltage level caused by EV fleets since it is equal to the electrical charge throughout the steady state operation or photovoltaic renewable energy sources while supplying energy to the grid.
- Analysis of hybrid solutions utilizing battery energy storage systems for the necessary integration of ultra-fast charging stations with capacities of up to 350 kW in metropolitan electricity networks.

Author Contributions: Conceptualization, G.G., M.S.R., C.D. and V.-C.N.; Data curation, G.G., C.D., D.L.M. and B.C.N.; Formal analysis, G.G., C.D., D.L.M., T.C.M., V.-C.N. and B.C.N.; Funding acquisition, M.S.R.; In-vestigation, G.G., C.D., D.L.M., T.C.M. and V.-C.N.; Methodology, G.G., M.S.R., C.D., T.C.M. and B.C.N.; Project administration, G.G., M.S.R. and B.C.N.; Resources, M.S.R.; Software, M.S.R., T.C.M. and B.C.N.; Supervision, G.G. and B.C.N.; Validation, G.G., M.S.R., T.C.M. and B.C.N.; Visualization, G.G., C.D., D.L.M., V.-C.N. and B.C.N.; Writing—original draft, M.S.R., C.D. and V.-C.N.; Writing—review and editing, G.G., M.S.R., D.L.M., V.-C.N. and B.C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by “Gheorghe Asachi” Technical University of Iasi, Romania. This paper was partially supported by UEFISCDI Romania and MCI through BEIA project AutoDecS, AISTOR, Hydro3D, PREVENTION, DAFCC, EREMI, FinSESCO, CREATE and by Europe research and innovation program under grant agreement No. 101084323 (BLOW).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work is supported by Ministry of Research, Innovation, Digitization from Romania by the National Plan of R&D, Project PN 1911, Subprogram 1.1, Contract No. 19PFE/30.12.2021.

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

References

1. Institute of Communication & Computer Systems of the National Technical University of Athens ICCS-NTUA. *Study on Cost Benefit Analysis of in EU Member States Smart Metering Systems*; European Commision, Final Report; Institute of Communication & Computer Systems of the National Technical University of Athens ICCS-NTUA: Athina, Greece, 2015.
2. IEA Secretariat Energy Efficiency Working Party, Proposal for an International Energy Association Initiative to Promote Energy-Efficient Distribution Transformers. Available online: <http://www.copperinfo.com/energy/transformers.proposal.html> (accessed on 12 January 2023).
3. McDonald, J. Adaptive Intelligent Power Systems: Active Distribution Networks. *Energy Policy* **2008**, *36*, 4346–4351. [CrossRef]
4. ANRE. *Rapoarte de Evaluare a Potențialului de Creștere a Eficienței Energetice a Rețelelor de Energie Electrică și Gaze Naturale, în ceea ce Privește Transportul, Distribuția, Gestiunea Sarcinii și Interoperabilitatea, Precum și Racordarea Capacităților de Producere, Inclusiv a Microgeneratoarelor*; Evaluation Report of the Romanian Energy Agency: Romania, 2014.
5. Iordache, M.; Conecini, I. *Electric Power Quality (in Romanian)*; Ed. Tehnica: Bucharest, Romania, 1997.
6. ANRE. Standardul de Performanță Pentru Serviciul de Distribuție a Energiei Electrice. Available online: <https://www.anre.ro/ro/energieelectrica/legislatie/norme-tehnice/standarde-de-performanta1387201290> (accessed on 14 January 2023).
7. Charalampopoulos, C.; Psomopoulos, C.; Ioannidis, G.C. *Implementing the E.U. Ecodesign Directive on Distribution Transformers*; IET MedPower Conference: Athens, Greece, 2014.

8. International Energy Agency—Global EV Outlook 2020, June 2020. Available online: <https://www.iea.org/reports/global-ev-outlook-2020> (accessed on 13 January 2023).
9. Lopes, J.A.P.; Soares, F.J.; Almeida, P.M.R. Integration of Electric Vehicles in the Electric Power System. *Proc. IEEE* **2011**, *99*, 168–183.
10. Degefa, M.Z.; Millar, R.J.; Koivisto, M.; Humayun, M.; Lehtonen, M. Load Flow Analysis Framework for Active Distribution Networks Based on Smart Meter Reading System. *Engineering* **2013**, *5*, 1–8.
11. Loughlin, F.M.; Duffy, A.; Conlon, M. *Analyzing Domestic Electricity Smart Metering Data Using Self Organizing Maps*; CIRED Workshop: Lisbon, Portugal, 29–30 May 2012.
12. Leite, D.R.V.; Lamin, H.; de Albuquerque, J.M.C.; Camargo, I.M.T. Regulatory Impact Analysis of Smart Meters Implementation in Brazil. In Proceedings of the Innovative Smart Grid Technologies (ISGT), Washington, DC, USA, 13–14 February 2012.
13. Richardson, D.B. Electric Vehicles and the Electric Grid: A Review of Modeling Approaches, Impacts and Renewable Energy Integration. In *Renewable and Sustainable Energy Reviews*; Elsevier: Amsterdam, The Netherlands, 2012; Volume 19, pp. 247–254.
14. Li, W.; Li, Y.; Deng, H.; Bao, L. Planning of Electric Public Transport System under Battery Swap Mode. *Sustainability* **2018**, *10*, 2528. [\[CrossRef\]](#)
15. Radu, A.T.; Eremia, M.; Toma, L. Optimal Electrical Vehicle Charging Strategy for Operating Conditions Improvement in Distribution Electrical Grid. In Proceedings of the International Conference on Condition Monitoring, Diagnosis and Maintenance—CMDM 2017, (4th Edition), Bucharest, Romania, 21 September 2017.
16. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data Clustering: A Review. Available online: www.cermics.enpc.fr/~jkeriven/vision/articles (accessed on 13 January 2023).
17. Mutanen, A.; Ruska, M.; Repo, S.; Järventausta, P. Customer Classification and Load Profiling Method for Distribution Systems. *IEEE Trans. Power Deliv.* **2011**, *26*, 1755–1763.
18. Guo, Z.; Wang, Z.J.; Kashani, A. Home Appliance Load Modeling from Aggregated Smart Meter Data. *IEEE Trans. Power Syst.* **2014**, *in press*.
19. Piacentini, R. Modernizing Power Grids with Distributed Intelligence and Smart Grid-Ready Instrumentation. In Proceedings of the Innovative Smart Grid Technologies (ISGT), Washington, DC, USA, 13–14 February 2012.
20. Turker, H.; Radu, A.; Bacha, S.; Frey, D.; Richer, J.; Lebrusq, P. Optimal charge control of electric vehicles in parking stations for cost minimization in V2G concept. In Proceedings of the 2014 International Conference on Renewable Energy Research and Application (ICRERA), Milwaukee, WI, USA, 17 March 2014; pp. 945–951.
21. Eremia, M.; Shahidehpour, M. (Eds.) *Electrical Power System Dynamics. Modeling, Stability, and Control*; John Wiley&Sons: Hoboken, NJ, USA, 2013.
22. Fidalgo, J.N.; Matosb, M.A.; Ribeiro, L. A new clustering algorithm for load profiling based on billing data. *Electr. Power Syst. Res.* **2012**, *82*, 27–33.
23. Hussein, T. Mouftah, Melike Erol-Kantarci, Sameh Sorour, *Connected and Autonomous Vehicles in Smart Cities*; CRC Press: Boca Raton, FL, USA, 2020; Volume 516, p. 242. ISBN 9780367350345.
24. Rigas, E.S.; Ramchurn, S.D.; Bassiliades, N. Managing Electric Vehicles in the Smart Grid Using Artificial Intelligence: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2015**, 16–25. [\[CrossRef\]](#)
25. Marijana, Z.; Mitrović, S.; Has, A. Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *Int. J. Inf. Manag.* **2021**, *58*, 102074. [\[CrossRef\]](#)
26. Howell, S.; Rezgui, Y.; Hippolyte, J.L.; Jayan, B.; Li, H. Towards the next generation of smart grids: Semantic and holonic multiagent management of distributed energy resources. *Elsevier J. Renew. Sustain. Energy Rev.* **2017**, *77*, 193–214. [\[CrossRef\]](#)
27. Kim, Y.-I.; Ko, J.-M.; Choi, S.-H. Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence Applications in Smart Grid (CIASG), Paris, France, 11–15 April 2011; pp. 1–6.
28. Hernandez, L.; Baladron, C.; Aguiar, J.; Carro, B.; Sánchez-Esguevillas, A. Classification and clustering of electricity demand patterns in industrial parks. *Energies* **2012**, *5*, 5215–5228. [\[CrossRef\]](#)
29. Ford, V.; Siraj, A. Clustering of smart meter data for disaggregation. In Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, 3–5 December 2013; pp. 507–510.
30. Rhodes, J.D.; Cole, W.J.; Upshaw, C.R.; Edgar, T.F.; Webber, M.E. Clustering analysis of residential electricity demand profiles. *Appl. Energ.* **2014**, *135*, 461–471. [\[CrossRef\]](#)
31. Ramos, S.; Duarte, J.M.; Duarte, F.J.; Vale, Z. A data-mining-based methodology to support MV electricity customers’ characterization. *Energy Build.* **2015**, *91*, 16–25. [\[CrossRef\]](#)
32. Al-Jarrah, O.Y.; Al-Hammadi, Y.; Yoo, P.D.; Muhaidat, S. Multi-layered clustering for power consumption profiling in smart grids. *IEEE Access* **2017**, *5*, 18459–18468.
33. Cai, H.; Shen, S.; Lin, Q.; Li, X.; Xiao, H. Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management. *IEEE Access* **2019**, *7*, 30386–30397. [\[CrossRef\]](#)
34. Nordahl, C.; Boeva, V.; Grahn, H.; Netz, M.P. Profiling of Household Residents’ Electricity Consumption Behavior Using Clustering Analysis. In Proceedings of the International Conference on Computational Science, Faro, Portugal, 12–14 June 2019; pp. 779–786.

35. Park, K.-J.; Son, S.-Y. A Novel Load Image Profile-Based Electricity Load Clustering Methodology. *IEEE Access* **2019**, *7*, 59048–59058.
36. Rasanen, T.; Voukantsis, D.; Niska, H.; Karatzas, K.; Kolehmainen, M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Appl. Energ.* **2010**, *87*, 3538–3545.
37. Wen, L.; Zhou, K.; Yang, S. A shape-based clustering method for pattern recognition of residential electricity consumption. *J. Clean. Prod.* **2019**, *212*, 475–488.
38. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
39. Yatskiv, I.; Gusarova, L. The Methods of Cluster Analysis Results Validation. In Proceedings of the International Conference RelStat'04, Latvia, Riga, 13–15 October 2005; pp. 75–80.
40. Grigoraș, G. Contributions on the Use of Artificial Intelligence in Supervision and Control of the Power Systems. Habilitation Thesis, “Gheorghe Asachi” Technical University of Iasi, Iasi, Romania, 2019. Available online: https://doctorat.tuiasi.ro/doc/Teze_abilitare/IEEIA/Grigoras/Teza_abilitare_Gheorghe_Grigoras.pdf (accessed on 13 January 2023).
41. Gerbec, D.; Gasperic, S.; Simon, I.; Gubina, F. Hierarchic Clustering Methods for Consumers Load Profile Determination. In Proceedings of the 2nd Balkan Power Conference, Belgrad, Yugoslavia; 2002; pp. 9–15.
42. Gordon, A.D. *Classification*, 2nd ed.; Chapman & Hall: New York, NY, USA, 1999.
43. Clustering: An introduction. Available online: www.elet.polimi.it/upload/matteucc/Clustering/tutorial_html (accessed on 13 January 2023).
44. Seppälä, A. Load Research and Load Estimation in Electricity Distribution. Available online: <https://cris.vtt.fi/en/publications/load-research-and-load-estimation-in-electricity-distribution-dis> (accessed on 13 January 2023).
45. Carter-Brown, C.G. Load profile modeling for integrated energy planning. In Proceedings of the Domestic Use of Electrical Energy Conference, Rome, Italy, 1999; pp. 13–18.
46. Nazarko, J.; Zalewski, W. The Fuzzy Regression Approach to Peak Load Estimation in Power Distribution Systems. *IEEE Trans. Power Syst.* **1999**, *14*, 809–814.
47. Wang, Y.; Chen, Q.; Kang, C.; Zhang, M.; Wang, K.; Zhao, Y. Load Profiling and Its Application to Demand Response: A Review. *Tsinghua Sci. Technol.* **2015**, *20*, 117–129.
48. Leong, H.; Jung, J.; Kang, B. Development of Operational Strategies of Energy Storage System Using Classification of Customer Load Profiles under Time-of-Use Tariffs in South Korea. *Energ. J.* **2020**, *13*, 1723. [[CrossRef](#)]
49. Issi, F.; Kaplan, O. The Determination of Load Profiles and Power Consumptions of Home Appliances. *Energ. J.* **2018**, *11*, 607. [[CrossRef](#)]
50. Wen, Z.; Li, P.; Lang, X.; Shen, X.; Cao, M.; Hua, X. Hierarchical classification method of electricity consumption industries through TNPE and Bayes, Measurement and Control. *J. Sage Pub.* **2021**, *54*, 346–359. [[CrossRef](#)]

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.