International Research Journal of Advanced Engineering and Science
ISSN (Online): 2455-9024

Clustering Process Based on Daily Load Profiles With 20 kV Voltage Level for Electric Distribution Networks in Mahajanga City, Madagascar

Rakotomalala Lovasoa Feno Fanantenana[1], Andriniriniaimalaza Philibert[1], Totozafiny Theodore[1], Randrianirina Jean Marc Fabien Sitraka[2], Rakotoarimanana Liva Graffin[2], Zely Randriamanantany[2]
[1]ISSTM and FSTE department, Mahajanga University, Mahajanga, Madagascar
[2]Physics department/ IME, Antananarivo University, LTTC Laboratory, Antananarivo, Madagascar
[1]ralovas@gmail.com, theodore.totozafiny@gmail.com, philibert.andriniriniaimalaza@gmail.com
randrianirinajeanfabien@gmail.com, graffinliva@gmail.com, zelyran@yahoo.fr

***Abstract*— *This study focuses on the classification of daily electrical load profiles in Mahajanga, Madagascar, where the 20 kV distribution network supplies five interconnected zones (ZI 1 to ZI 5). Real consumption data, measured on 24 hours on a weekday in March 2025, reveal clear variability in demand, especially during evening peak hours.To analyze this behavior, two hierarchical clustering methods Centroid, Average Linkage and k-means Cluster were applied to group similar profiles. The clustering process resulted in six distinct consumption groups for both methods, with slight differences in group membership and structure. Strong correlations were identified between ZI 4 and ZI 5 (CC = 0.9735), as well as between ZI 2 and ZI 4, indicating similar load behaviors. The highest average load was observed in ZI 2 (3823.37 kW), while the lowest was recorded in ZI 1 (1118.02 kW), based on LSMeans estimates. A consistent result by K-means Clustering forecasts a Typical Load Profiles (TLP) by adapting the load demand off all ZI by one single significant curve including the variation of daily load consumption on 24 hours. These results demonstrate the effectiveness of clustering in detecting consumption patterns and optimizing grid management. The approach provides a practical tool for demand forecasting and energy planning in urban tropical regions with limited monitoring infrastructure.*

***Keywords*— *Load demand, Clustering, Electric Grid Optimization, Typical load Profiles, Energy Forecasting.*

I. INTRODUCTION

Energy demand in urban tropical regions continues to grow due to demographic expansion and climate-related factors. Mahajanga, a coastal city in western Madagascar, faces strong variability in daily electricity consumption, with temperatures often exceeding 30°C. The electricity distribution network, managed by JIRAMA, operates at 20 kV and supplies five interconnected zones with distinct usage characteristics. These variations create challenges in predicting load demand and optimizing energy distribution. This study aims to analyze daily load profiles recorded in Mahajanga and to group them into consistent patterns using clustering techniques. The purpose of this approach is to improve the understanding of consumption behavior and to support more efficient planning of energy resources. The classification of demand curves helps the Distribution Network Operator (DNO) to anticipate peak loads, reduce energy losses, and guide infrastructure investment [1].

To reach this objective, the research uses real load data collected over 15-minute intervals on a weekday in March 2025. Two hierarchical clustering methods Centroid, Average Linkage and K-means clustering are applied to group the load profiles based on their similarity [2][3]. This data-driven approach captures patterns in electricity usage and identifies zones with correlated demand behavior. For example, the analysis reveals that ZI 2 and ZI 4 consistently experience high consumption during peak evening hours, likely due to population density and commercial activity. The quality of clustering is evaluated through statistical indicators such as the Cophenetic Correlation Coefficient (CC) and LSMeans, which provide insights into the internal structure and accuracy of the grouping [4]. A high CC value confirms that the clusters reflect real similarities in load behavior.

This research demonstrates how machine learning techniques like clustering can enhance energy management in developing regions. It also shows how data classification supports decision-making in distribution networks without relying on advanced metering infrastructure. The findings contribute to future strategies for smart planning, energy loss reduction, and demand forecasting in tropical electricity systems.

II. METHOD DESCRIPTIONS

The first work in this research is to collect the daily load profile data for this year 2025. Once the data is acquired, the manner of grouping the load profile was accompanied by using two pieces of clustering method. After that, clustering analysis are carried out to classify the data load profiles with a comparison of two (02) hierarchical methods. Finally, the results of the clustering are discussed for an eventual perspective.

I.1 Load profiles data collection

In this paper, the collection of the data and the simulation is provides by the company JIRAMA which product and distribute Electricity network in Madagascar. We obtained a daily load profiles, sample data for a weekday on march 2025 on Electric Network Distribution of JIRAMA Mahajanga.

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

Mahajanga town is localized in the west coast of Madagascar, characterized by a tropical climate which is considered the warmest region of the island with a temperature for about 27°C to 35°C., the geographical coordinates are 15°43S and 46°19E. The data load profiles describe the consumption of five (05) interconnected zone (ZI) in the city feed by 20 kV voltage level. Figure.1 shows the variation of load demand in depend of the hour of the day and every fifteen minutes' intervals for the peak hour demand. This peak hour, the load demand is higher, is from 07:00 pm to 09:00 pm and it's depend on the area. The ZI in very popular area are recorded a considerable fluctuation of the load demand such as ZI 2 and ZI 4, a slight variation is observed for the others ZI. The highest value of consumption is almost 5600 kW during the peak period for ZI 2, it's lower for ZI 1,3, 4 and 5.

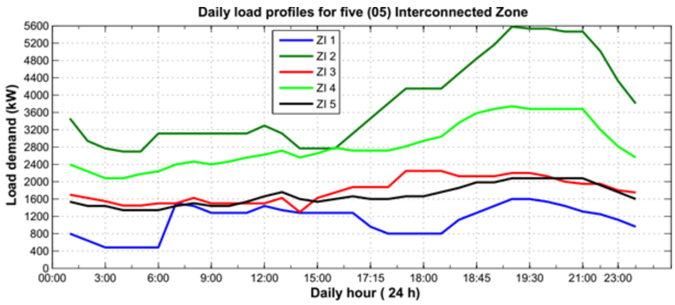

Figure 1. Daily load demand for the five (05) ZI in Mahajanga city.

An investigation of that load profiles seems indispensable to enhance the production and the networks distribution of electricity. Clustering is a part of data analysis to distinguish the difference of the attribute for its classification and improving the grid of electric distribution in our case of study.

I.2 Clustering approach

Clustering is an approach used in unsupervised machine learning models. The term cluster study refers to a group of statistical techniques that are specifically designed to identify classifications in complicated data sets. The purpose of cluster analysis is to group objects into clusters so that objects within one cluster share more in common with one another than they do with the objects of other clusters. Consequently, the analysis arranges objects into similar groups based on multivariate observations [4][5].

The objective of Clustering data is to capture the structure in a heterogeneous group of data. Several studies have explored the application of clustering methods with a machine learning approach on Energy Load Profiles (ELPs) based on electricity distribution network [6].

In our case, we use two (02) methods of clustering in hierarchical to classify that load profile shown in fig.1. Centroid and Average linkage, the most used in several paper are our choice to compare the dendrogram of the ZI load demand [4][5].

- *Hierarchical Clustering:* the hierarchical clustering method is based on a tree structure known as the dendrogram. This can be done in a top down approach known as the divisive method which starts at a single cluster and performs binary splits until all clusters only have one member [1].

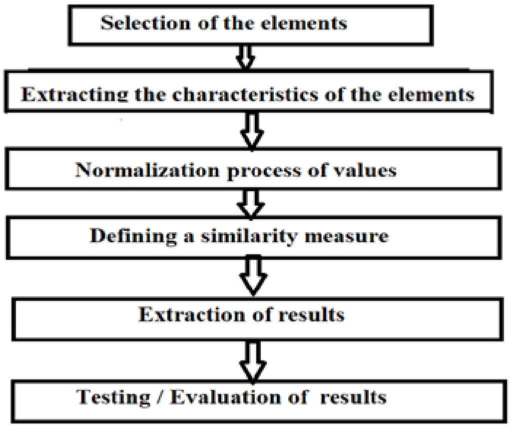

Figure 2. The stages of clustering process [1]

- The average linkage method is seen as being one of the most robust methods and is the average distance between all pairs of data points where one comes from each group. This is in contrast to simpler methods known as single linkage (nearest neighbor) and complete linkage (furthest neighbor) where distance is simply calculated by the nearest or furthest points in each group.
- The centroid method, the distance between two clusters is defined as the squared Euclidean distance between their means. The centroid method is more robust to outliers than most other hierarchical methods but in other respects may not perform as well as Ward's method or average linkage [4].

Because the classification of the load profiles through the visual comparison of the graphics is subjective and impractical, the cluster analysis method was applied to solve this problem [1]. Difference between the two methods (average linkage and centroid) arise to the different criteria in defining the distance between clusters. The results of clustering process by centroid and average are presented as dendrograms in Figure 3.

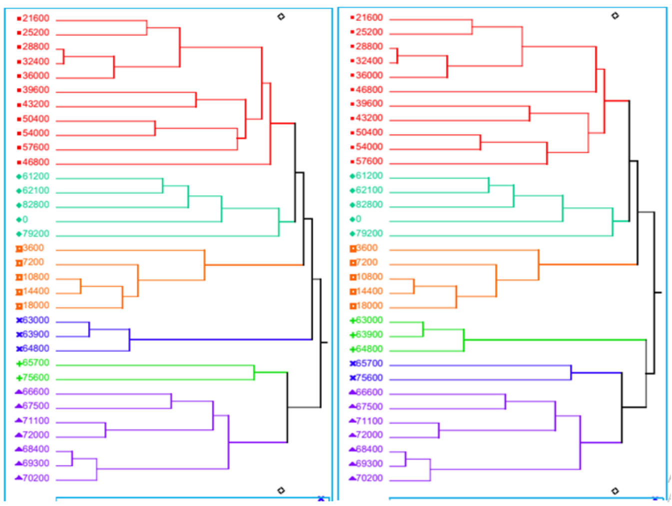

Figure 3. The dendrograms of Centroid (left) and Average linkage (right) methods.

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

The electric network distribution of the case of study contain five (05) interconnected area or ZI, the dendrograms shows six (06) groups of cluster for Centroid and Average linkage methods. As can be seen from the figures below, the slight difference between both dendrograms refers to the number of member of each group by his power value in kW and the order of the values, the number seen in the figure. Each group is distinguished by its marks and its colors. The distance between clusters are not shown in the figure but in the table below (table.1)

Following the dendrogram, it may wish to confirm that the height of the link reflects the actual distance between the two clusters that include those two things. The height represents the linkage of objects in the cluster tree. This linkage should correlate strongly with the distances between objects in the distance vector. For Average linkage method the Euclidian distance is calculated by [1]:

$$d(C_r, C_s) = \frac{\sum_{\substack{X_r \in C_r \\ X_s \in C_s}} d(X_r, X_r)}{|N_r|.|N_s|}$$

For Centroid method: $d(C_r, C_s) = \left\|\overline{X}_r - \bar{X}_s\right\|^2$
where $\overline{X}_r$and $\overline{X}_s$ are the average vectors of the cluster Cr et Cs[1].

Table 1 shows the distance and the number of clusters associated with each (leader, joiner) pair, defined by the load demand (kW), for both the Centroid and Average linkage methods. The values observed are almost identical for the first six cluster numbers. The difference appears starting from the seventh distance (26), where the value increases for the Average linkage method, except for cluster numbers 16 and 19, where both methods yield equal values. Cluster number 1 has the highest distance value and will be considered as an outlier. In terms of (leader/joiner) pairs, the values alternate and are almost aligned for the two hierarchical methods. This dissimilarity reflects the different ways in which the two methods classify the data.

TABLE 1. Clustering history

Cluster Centroid				Cluster Average linkage			
Number of Clusters	Distance	Leader	Joiner	Number of Clusters	Distance	Leader	Joiner
32	0.1213218896	28800	32400	32	0.1213218896	28800	32400
31	0.1290607924	68400	69300	31	0.1290607924	68400	69300
30	0.1819828344	10800	14400	30	0.1819828344	10800	14400
29	0.2426437792	63000	63900	29	0.2426437792	63000	63900
28	0.3232666903	68400	70200	28	0.3296445435	68400	70200
27	0.4039209278	71100	72000	27	0.4039209278	71100	72000
26	0.4676290463	28800	36000	26	0.4715471081	28800	36000
25	0.498054551	10800	18000	25	0.5062981076	10800	18000
24	0.5017914676	63000	64800	24	0.5162496275	63000	64800
23	0.5228713131	7200	10800	23	0.5524283824	21600	25200
22	0.5524283824	21600	25200	22	0.5621296428	50400	54000
21	0.5621296428	50400	54000	21	0.5729072695	61200	62100
20	0.5729072695	61200	62100	20	0.5779603014	7200	10800
19	0.5854334193	66600	67500	19	0.5854334193	66600	67500
18	0.5942760453	21600	28800	18	0.6649783646	61200	82800
17	0.6001171056	61200	82800	17	0.6931856435	21600	28800
16	0.6979881858	39600	43200	16	0.6979881858	39600	43200
15	0.7558793044	3600	7200	15	0.8173724669	3600	7200
14	0.8078931808	66600	71100	14	0.8816596076	50400	57600
13	0.8271444451	61200	0	13	0.8827018141	66600	71100
12	0.8335247984	66600	68400	12	0.9049325688	61200	0
11	0.8356591589	50400	57600	11	0.9647425736	65700	75600
10	0.9220455632	39600	50400	10	0.9732269137	66600	68400
9	0.9647425736	65700	75600	9	1.0861921837	39600	50400
8	1.0084729574	21600	39600	8	1.1128560017	21600	46800
7	1.1047219825	21600	46800	7	1.2868705513	21600	39600
6	1.2856725941	61200	79200	6	1.3719645667	61200	79200
5	1.4392483461	65700	66600	5	1.6170289256	65700	66600
4	1.6057424064	21600	61200	4	1.8963769715	21600	61200
3	2.3823652836	21600	3600	3	2.6325305524	21600	3600
2	2.726958382	21600	63000	2	2.901042876	63000	65700
1	3.8013325668	21600	65700	1	4.0643763956	21600	63000

- *K-mean clustering*: This method uses a simple and straightforward mechanism to classify the input dataset into *K* clusters (*K* being an integer fixed in advance). The goal is to group load profile data in order to determine the number of clusters and identify a central point (centroid) for each cluster. Once the centroid of each cluster is determined, each data point is assigned to the nearest centroid. The centroids are then recalculated, and the process is repeated

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

iteratively until the centroids reach a stable position. The grouping is performed by minimizing the sum of the squared distances between the data points and their corresponding cluster centroids [7][8].

$$J = \min\left(\sum_{k=1}^{K}\sum_{l=1}^{n_k}\left\|X_l^{\,k} - c_k\right\|^2\right)$$

Where $\|X_l^k - c_k\|^2$ represents the distance between the vector X_l^k and the centroid of cluster c_k , k=1,2,3…,K=1,2,3…,n_k represents the number of the vectors from the cluster k.

In this work, K-means clustering is applied to identify the Typical Load Profile (TLP) from all the ZI load demand data. This TLP-based approach offers an alternative to the smart metering-based method. Based on the TLP, the Distribution Network Operator (DNO) can extract the main indicators of the load curves to determine the nodes in the electrical distribution network using Smart Meter information. By knowing the load profiles of the nodes, DNOs can simplify demand forecasting within the supply area, thereby improving the efficiency of their marketing strategies [1][8].

- *Cophenetic correlation coefficient (CC):* If the clustering is valid, the linkage of objects in the cluster tree should strongly correlate with the distances between objects in the distance vector[9]. The more accurate the clustering solution reflects the data, the closer the cophenetic correlation coefficient (CC) value is to 1. The CC is calculated using the following formula:

$$CC = \frac{\sum(d(X_r,X_s)-\overline{d(X_r-X_s)})(d(T_r,T_s)-\overline{t(T_r,T_s)})}{\sqrt{\sum(d(X_r,X_s)-\overline{d(X_r-X_s)})^2(\sum(d(T_r,T_s)-\overline{t(T_r,T_s)})^2)}}$$

$\overline{d(X_r-X_s)}, \overline{t(T_r,T_s)}$ represents the average distances.

$t(T_r, T_s)$: the dendrogram distance between the models [T_r] and [T_s], r≠s from the dendrogram [T] [1][3].

III. Results And Discussions

II.1- Validation of Clustering results

a- Correlation matrix

Cluster validation is primarily concerned with assessing the outcomes of the clustering procedure. Internal cluster validation can be performed using the Correlation Coefficient (CC) or the Silhouette Coefficient (SC). The average cluster width for each cluster, the median width for each element, and the average width for the full dataset are all measured by the SC test. The values of both CC and SC range from -1 to 1. In fact, some literature provides the following interpretations for CC and SC [10] [11]:

- 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.

The first validation test verifies the correlation between the ZI load profiles. Figure 4 shows a correlation matrix of the ZIs, highlighting a high correlation between ZI4 and ZI 5 with a CC of 0.9735. These two interconnected areas have similar load demand values according to the clustering process. A strong correlation is also observed between ZI 4 and ZI 2, with a CC of 0.9323. For ZI 5 and ZI 2, a slight decrease is noted with a CC of 0.9286, but this still indicates a strong relationship in load demand values from the clustering analysis. Thus, a strong correlation among ZI 4, ZI 5, and ZI 2 is evident. These three interconnected areas could be merged into one or, at most, two interconnected areas. However, this concept is subjective, and it is advisable to consider other types of analysis for confirmation. Moreover, the load demanded by consumers depends on various factors such as time (weekday, weekend, holiday, events), climatic conditions (humidity, temperature, wind speed), and consumer type.

Correlations

Variable	ZI 1 (kW)	ZI 2 (kW)	ZI 3 (kW)	ZI 4 (kW)	ZI 5 (kW)
ZI 1 (kW)	1.0000	0.4505	0.2128	0.6109	0.6102
ZI 2 (kW)	0.4505	1.0000	0.8120	0.9323	0.9286
ZI 3 (kW)	0.2128	0.8120	1.0000	0.8016	0.7616
ZI 4 (kW)	0.6109	0.9323	0.8016	1.0000	0.9735
ZI 5 (kW)	0.6102	0.9286	0.7616	0.9735	1.0000

Figure 4. Correlation Matrix of the ZI

b- Intercept matrix

In the context of clustering, especially when applying discriminant analysis or MANOVA (Multivariate Analysis of Variance) post-clustering, the overall E & H matrices concept is used. The E and H matrices are components of a decomposition that evaluates how well groups (or clusters) differ from each other in terms of multivariate means. These matrices include the Hypothesis Matrix (H), the Error Matrix (E), and the Intercept Matrix (I) [12][13].

In this case study, the Intercept Matrix is an important index supporting the Correlation Matrix. The Intercept Matrix represents the interactions between the ZIs. Figure 4 shows these interaction or influence terms between zones. Compared to the Correlation Matrix, the Intercept Matrix displays larger component values with higher cross-dependencies. All the ZIs are related, as shown in Figure 5 by the framed values with similar colors. However, only the highest values are considered to confirm the strength of the correlations.

Apart from the strong correlation between ZI 2, ZI 4, and ZI 5, the Intercept Matrix also shows a link between ZI 1 and ZI 2. Additionally, the Intercept Matrix increases the correlation between ZI 3 and ZI 2, while decreasing the correlation between ZI 4 and ZI 5. Therefore, according to this matrix, all ZIs are highly correlated, which is less precise compared to the Correlation Matrix. The accuracy of the Intercept Matrix is less significant for result validation.

Intercept	ZI 1 (kW)	ZI 2 (kW)	ZI 3 (kW)	ZI 4 (kW)	ZI 5 (kW)
ZI 1 (kW)	4.18926e10	1.43263e11	6.75148e10	1.06223e11	6.2972e+10
ZI 2 (kW)	1.43263e11	4.89927e11	2.30885e11	3.63258e11	2.1535e+11
ZI 3 (kW)	6.75148e10	2.30885e11	1.08808e11	1.7119e+11	1.01487e11
ZI 4 (kW)	1.06223e11	3.63258e11	1.7119e+11	2.69339e11	1.59672e11
ZI 5 (kW)	6.2972e+10	2.1535e+11	1.01487e11	1.59672e11	9.46581e10

Figure 5. Intercept matrix of the ZI

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

c- Scatterplot matrix

To illustrate the potential of the proposed approach using the Correlation Matrix and Intercept Matrix, another clustering technique is presented in Figure 6. A scatterplot matrix of the data is shown to visualize the distribution of each cluster for the load demand groups in the ZIs. Based on the obtained clusters, the structure of the scatterplot determines the partition quality of each ZI. In this study, this is confirmed through the correlation coefficient matrix [12] [14].

The scatterplot matrix of the five interconnected areas shows that each point within a cluster roughly follows linear regression curves. The distances between clusters are validated, with each ZI divided into six groups marked by different symbols and colors. The scatterplot also indicates the number of load demand points in each group, ranging from a minimum of two points to a maximum of eight to ten points, depending on the ZIs. Each ZI is defined by its load demand value intervals (in kW) along the two axes of the scatterplot matrix.

Furthermore, Figure 6 highlights the strong correlation between ZI 4 and ZI 5, indicated by a red parallel line in the scatterplot. This correlation confirms the quality of clustering in these interconnected areas, where clusters show similar alignment and robustness, and the distances between clusters are small and consistent.

On the other hand, the clustering distribution for ZI 1 and ZI 3 shows significant distances between clusters, with each cluster containing some outlier points. Consequently, these two areas have a lower correlation with the others.

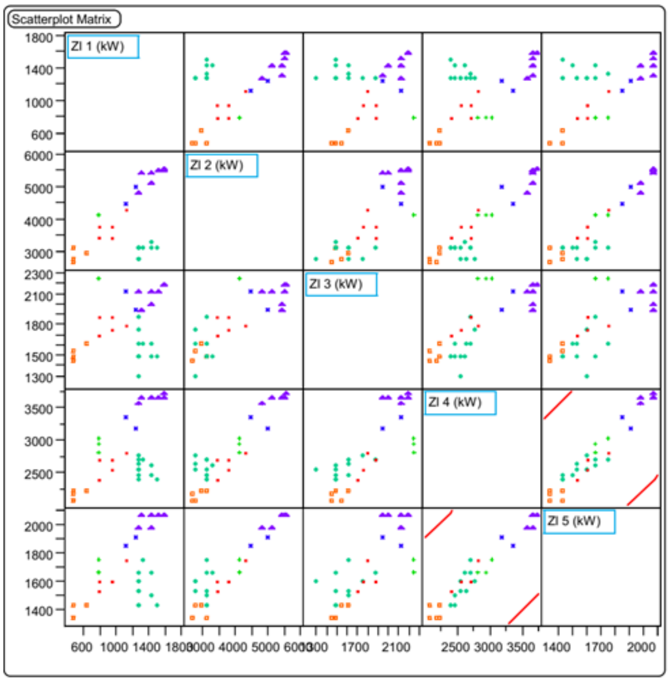

Figure 6. scatterplot matrix with the five (05) interconnected zone (ZI)

d- Outliers distance

In terms of load profiles in electric distribution networks, an outlier is detected as an irregular consumption pattern. Outliers impact statistical results and may lead to ambiguous conclusions. Outliers are defined as values that differ greatly from the typical dataset [1].

First, outliers must be detected and then properly treated. Two methods can be used for outlier detection: data mining techniques such as clustering, and steady-state estimation [1][14].

Using the hierarchical clustering process, Figure 7 shows the outlier distances (Mahalanobis distance) for the six clusters. The row number corresponds to the cluster number (see Table 1), and the distances represent values to be analyzed. The distances indicate that outliers are not present in all clusters, as the values remain below 1.0 for each cluster, with point distributions being nearly uniform.

An outlier is detected when some points fall outside this distance range and exhibit significantly higher distances compared to others. Figure 7 confirms this principle of outlier detection.

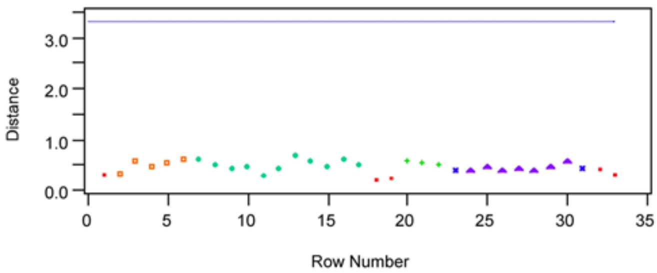

Figure 7. Outlier distance analysis.

II-2. The load profiling process

This section presents a clustering approach aimed at determining the Typical Load Profiles (TLP) of nodes in the electric distribution network operating at a 20 kV voltage level. As previously outlined in the clustering process section, the TLP is a key outcome of the clustering process. It consolidates the load curves of the five interconnected zones (ZI) into a single representative profile based on their similarities [8].

Before obtaining the TLP, the Least Squares Means (LSMeans) are extracted from the clustering process, as shown in Figure 8. LSMeans are adjusted means derived from linear models such as regression or ANOVA. They represent the average expected response for each factor level, taking into account the influence of all other model variables. The overall LSMeans are the estimated group means averaged across the levels of other factors, based on predicted values.

In our case study, the LSMeans for ZI 1 reflects the average predicted energy consumption for that zone, adjusted for time and interaction effects. The overall LSMeans represent the adjusted average across all ZIs. Figure 8 displays the overall LSMeans values for the five ZIs: the lowest value is for ZI 1 at 1118.0188 kW, while the highest is for ZI 2 at 3823.36879 kW.

The resulting prediction curve resembles an assembly of four linear regression segments. By analogy, these overall LSMeans values are essential for comparing the estimated values of the Typical Load Profiles (TLP).

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

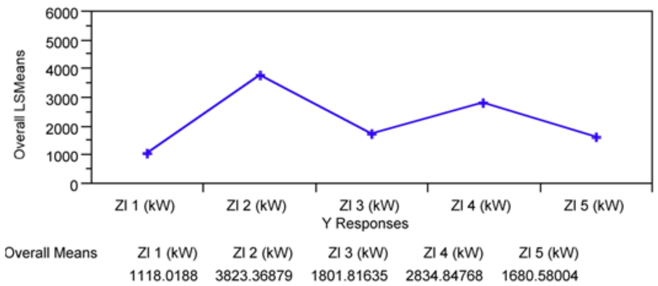

Figure 8. Overall LSMeans for the ZI

Figure 1 illustrates the load consumption patterns on a typical summer weekday, showing the level of activity driving the demand. Interconnected areas with a high number of residential neighborhoods, such as ZI 2 with seventeen (17) popular neighborhoods, and ZI 4 with twenty-five (25) neighborhoods spread over a large area, exhibit significantly higher load demand compared to the others.

Figure 9 presents the Typical Load Profiles (TLP) for the five (05) ZIs. The TLP reveals consistent fluctuation in load demand over a 24-hour period:

- From 00:00 to 03:00, the load demand decreases steadily, dropping from around 2300 kW to a minimum of about 800 kW;
- From 03:00 to 12:00, the demand rises again, reaching up to 2350 kW;
- Between 12:00 and 13:00, a slight decrease is observed, falling below 2100 kW;
- From 13:00 to 19:00, the demand intensifies, peaking at approximately 3400 kW;
- Between 19:00 and 00:00, the load decreases again, continuing until it reaches its lowest point around 03:00.

Overall, the daily load variation can be divided into three main intervals:

- The largest variation, approximately 1600 kW, occurs between 19:00 and 03:00;
- The second highest variation, around 1300 kW, takes place between 13:00 and 19:00;
- The third variation, about 1200 kW, is observed between 03:00 and 13:00.

This TLP is generated by merging the five individual load profiles from the Mahajanga electric distribution network using a clustering-based forecasting method. The variation patterns and load ranges derived from the TLP can help estimate energy loss coefficients. The load factor reflects how load demand fluctuates at the feeder level, while the loss factor allows estimation of average energy loss in the distribution network without requiring full load flow analysis.

Understanding consumer load behavior plays a key role in managing and controlling electricity demand. Distribution Network Operators (DNOs) can use this insight to forecast and prepare for load variations on specific days.

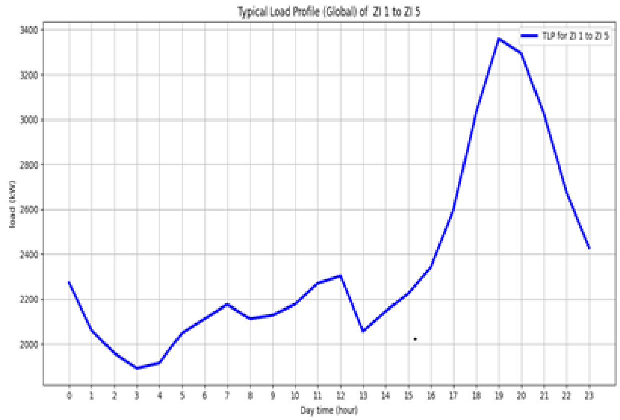

Figure 8. Typical Load Profile (Global) of ZI 1to ZI 5

IV. CONCLUSIONS

This study investigated the classification of daily load profiles in Mahajanga's medium-voltage electricity distribution network using Clustering techniques. The use of Centroid, Average Linkage and k-Means methods allowed for the identification of distinct consumption behaviors among five interconnected zones. The analysis confirmed strong correlations between certain areas, particularly ZI 4 and ZI 5, which share similar demand patterns during peak hours. By applying statistical validation tools such as the Cophenetic Correlation Coefficient and LSMeans, the clustering process demonstrated a reliable internal structure. These results highlight the value of data-driven approaches in managing electric distribution systems in tropical urban contexts such as the proposal TLP (Typical Load Profiles) which combines all the Interconnected Zone (ZI). This TLP provides the Distribution Network Operator (DNO) with essential information for improving demand forecasting, minimizing energy losses, and optimizing resource allocation. Moreover, the approach offers a practical alternative in environments where smart metering systems are limited or unavailable. Furthermore, clustering techniques represent a robust and scalable solution for understanding load behavior and supporting efficient grid planning. Future research may combine these methods with real-time monitoring, meteorological data, or socio-economic indicators to further enhance forecasting accuracy and operational reliability.

REFERENCES

[1] YI-Tung. C, An introduction to approaches and modern applications with ensemble learning, NOVA science publishers New York USA, pp. 1-109, 2020, ISBN: 978-1-53618-680-2 .

[2] Gheorghe.G, Gheorghe.C, Florina.R, Using K-Means Clustering Method in Determination of the Energy Losses Levels from Electric Distribution Systems, Conference Paper, Romania ,2014. pp.52-56, ISSN:1792-5967 52, ISBN: 978-960-474-238-7.

[3] Gheorghe. G, Maria. S, Catalin.D and al, Contributions to Power Grid System Analysis Based on Clustering Techniques Paper journal in Sensors , Romania , 2023. 28p, published on https://doi.org/10.3390/s23041895.

[4] Elena. B, Gheorghe. C, Gheorghe.G, Clustering Techniques in Load Profile Analysis for Distribution Stations, IEEE Conference on Advances in Electrical and Computer Engineering, Romania, 2009, pp.63-66. DOI:10.4316/aece.2009.01011.

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.

[5] Cartina.G, Gheorghe.G, Bobric.E.C, Clustering Techniques for Energy Losses Evaluation in Distribution Networks, 2009 IEEE Power Tech Conference in Bucharest, Romania,2009,05p. ISBN: 978-1-4244-2235-7
[6] Gheorghe.G, Bogdan.L, Bogdan-Constantin.N, Data Clustering-Driven Fuzzy Inference System-Based Optimal Power Flow Analysis in the Electric Networks Integrating the Wind Energy, paper journal in preprints.org, Romania, February 2025, 32p, published on https://doi:10.20944/preprints202502.0714.v1
[7] George E.O , Mohamed C.B, Chisom N.O and al, *K*-means clustering of electricity consumers using time-domain features from smart meter data, journal of Electrical Systems and Information Technology ,Nigeria,2023,18p, published on https://doi.org/10.1186/s43067-023-00068-3, Springer Access.
[8] Damayanti.R, Abdullah.A, Purnama1.W and al, Electrical Load Profile Analysis Using Clustering Techniques, IOP Conference Series: Materials Science and Engineering, Indonesia, 2017, 7p, doi:10.1088/1757-899X/180/1/012081.
[9] Betsy S, Emilio. B , Miriam.E and al, A hybrid clustering approach for electrical load profiles considering weather conditions based on matrix-tensor decomposition, Published by Elsevier Ltd, Switzerland, 2024, 13p. ISSN : 2352-4677.
[10] Dine.T, Norashikin.A, Sharifah.S and al , clustering algorithm for electrical load profiling analysis: a systematic review of machine learning approaches for improved clustering algorithms, paper journal in jatit. org, Malaysia, 2024 pp.5453-5468, ISSN: 1992-8645, E-ISSN: 1817-3195,
[11] Kalyani.K, Chakravarthy.T , Load profile clustering: an algorithmic approach with improved replacement in bee optimization algorithm, paper journal on ictact soft computing, India, 2015, ISSN: 2229-6956, DOI: 10.21917/ijsc.2015.0127.
[12] Navarro.M ,Young.M, Prasetyo1.Y and al, Customer Load Profile Clustering Using K-means Algorithm: A Case Study in an Electric Distribution Company in the Philippines Amidst the COVID-19 Pandemic, IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Philippines, 2022, pp. 0625-0629, ISBN: 978-1-6654-8687-3,published on IEEE, DOI: 10.1109/IEEM55944.2022.9989809.
[13] Brian.A, Peter. M, Analysis of Clustering Techniques on Load Profiles for Electrical Distribution, IEEE Conference Paper, United Kingdom,2014, 7p. DOI: 10.1109/POWERCON.2014.6993986.
[14] S.K.Nandi, H.R.Ghosh,"A wind–PV-battery hybrid power system at Sitakunda in Bangladesh", Bangladesh, 2009 Energy Policy, pp. 3659–3664.

Rakotomalala Lovasoa Feno Fanantenana, Andriniriniaimalaza Philibert, Totozafiny Theodore, Randrianirina Jean Marc Fabien Sitraka, Rakotoarimanana Liva Graffin, and Zely Randriamanantany, "Clustering process based on daily load profiles with 20 kV voltage level for electric distribution networks in Mahajanga City, Madagascar," *International Research Journal of Advanced Engineering and Science*, Volume 10, Issue 3, pp. 86-92, 2025.