INFLUENCE OF ENERGY USE ON POWER QUALITY

Mădălina CĂLIN-ARHIP[1], Cristian GOREA[2]

This work is based on real analysis data obtained from the measurement programs within the Faculty of Energy and aims to identify the usage behavior by determining the optimal number of clusters and by calculating the Euclidean distances between the resulting centroids.

Cuvinte cheie: centroid distances, clusters, interruption, outliers, power quality

1. Introduction

In the electricity market, minimizing energy losses in electricity networks is a constant concern for transmission and distribution operators. For the consumers, the losses may be viewed as financial losses, given the introduction of these in the energy bill. These costs significantly affect the economic balance of operators.

In order to limit forecasting errors and improve economic processes, it is essential to identify with a high degree of confidence the individual load profiles of users. Detailed information about the state of the system allows choosing an optimal structure for power transmission, thus ensuring the minimization of energy losses, even in the context of restrictions specific to the electricity market.

Outlier detection faces various challenges, but multiple techniques using varied methodologies and algorithms have been developed to address them [1]. Among the difficulties encountered are the nature of the input data, the types of outliers, the data labels, as well as issues related to the accuracy and complexity of the calculation, including CPU (Central Processing Unit) time and memory consumption [2], [3].

Researchers continue to develop more effective solutions to address the challenges associated with detecting outliers in various types of data streams.

These include:

- Distributed data streams: This involves managing and analyzing data that is generated by multiple distributed sources, which complicates anomaly detection [4].

[1] PhD Student, Electrical Power Systems Department, Faculty of Energy Engineering, National University of Science and Technology POLITEHNICA Bucharest, Romania, e-mail: madalina.arhip@upb.ro

[2] PhD Student, National University of Science and Technology POLITEHNICA Bucharest, Romania, e-mail: cristian.gorea@stud.energ.upb.ro

- RFID read streams: RFID (Radio Frequency IDentificator) technologies generate large volumes of real-time data, and identifying anomalies in these streams is essential for applications such as inventory management and logistics [5].
- Big multidimensional data: Analyzing data that has multiple dimensions requires sophisticated methods to identify anomalies that may occur in the interaction of different variables [6].
- Wireless Sensor Networks: Data from distributed sensors is often noisy and incomplete, and anomaly detection is crucial to ensure proper operation of monitored systems [7].
- Effective trajectories: Analyzing trajectories, such as the movement of vehicles or people, requires specific techniques to identify unusual behaviors [8].
- Data quality: Data quality issues, such as missing data or measurement errors, can generate outliers that need to be handled appropriately [9].

As the volume of data increases, a similar trend in the number of outliers is observed [10]. This phenomenon emphasizes the need to develop more advanced and scalable techniques to detect and manage anomalies effectively, thereby ensuring data integrity and usefulness in various applications.

An outlier is generally defined as a data point that deviates significantly from other points or does not align with the expected normal pattern. This can affect the interpretation of the data and the decisions made based on it [7].

There are several methods of detecting outliers, each with its own advantages and disadvantages. These methods can be classified as follows [11]:

- Distance-based methods: These measure distances between data points to identify outliers, usually by comparing them to other points in the set.
- Density-based methods: These techniques identify areas of low density of data points, considering them as possible outliers.
- Clustering methods: These group the data and identify the points that do not fit into the formed clusters, considering them as outliers.
- Graphical methods: Using visualizations such as boxplots or scatter plots can help quickly identify outliers by observing the behavior of the data.

Each of these methods can be adapted according to the specifics of the data and the context of the analysis, providing diverse solutions to the problem of outlier detection.

2. Preliminary data

The data used in this article were provided by the measurement system installed at the Faculty of Energy, specifically for the local energy production system. This system was developed under the auspices of the European WEDISTRICT project, which aims to improve the efficiency and sustainability of energy systems. The project aims to promote innovative solutions for energy management, and the data obtained were essential for consumption analysis and the identification of usage patterns, thus contributing to the improvement of the customer energy performance. The measurement program recorded values every 2 seconds, so the average hourly values were used in the work. The data analysis was used to also identify possible interruption patterns which may arise in the energy feeding system. This outcome may be linked at a latter stage to a specific power quality behavior of the system, providing insightful information for future automatization of power quality issues identification and classification.

3. Case study

The data analysis period extends from October 2022 to January 2023, thus covering the first semester of the courses. This analysis was conducted in two distinct scenarios:

- Scenario 1: This includes weekdays, reflecting typical energy use during educational activities.
- Scenario 2: This scenario focuses on non-working days, providing insights into energy consumption in the absence of school activities.

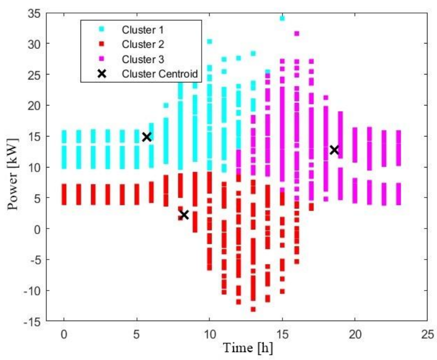
For both scenarios, a division of the consumption areas into three, four or five clusters was carried out, which allowed a detailed analysis of the energy consumption behavior. The clusters were defined in a gradual manner. For the cases with fewer clusters, we have concentrated on the smaller than average values, and morning hours, as well as afternoon hours, as exhibiting different behavior. For the cases with a higher number of clusters, we have concentrated our analysis on the upper-mentioned criteria, and then added the values that exhibited a standard deviation of more than 30%, and, respectively, 60% than the average values for the same hours. The distances between the obtained centroids were also calculated, providing valuable information about the similarities and differences between the identified clusters.

In Fig. 1, the data related to transformer 1 recorded on working days are analyzed. The three-cluster analysis provides a clear structure of energy consumption. It highlights how power is distributed across different ranges, showing that:

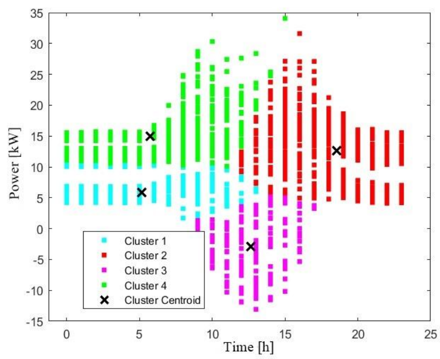
- The first cluster (00:00 - 10:00): This is characterized by average power values, suggesting moderate energy use during this morning period.
- The second cluster (10:00 - 23:00): In this window, high power values are recorded, indicating increased activity, probably due to the use of equipment and activities within the institution.
- The last cluster: It is composed of low power values, reflecting the periods of minimum energy use, as well as the fact that the photovoltaic system can support the respective load.

Comparatively, four or five cluster analysis allows a more detailed division of the data, facilitating a deeper understanding of fluctuations in consumption and their relationship to specific time intervals.

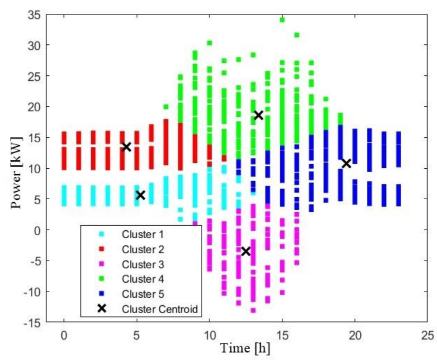
This granularity brought by additional clusters helps identify patterns and optimize power management.



a) 3 clusters



b) 4 clusters



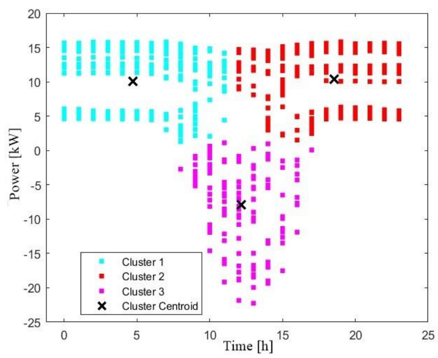
c) 5 clusters

Fig.1. Graphic representation for scenario 1-Transformer 1

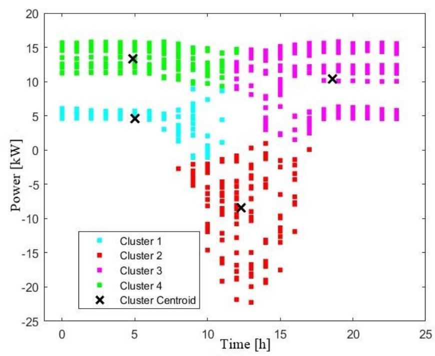
In fig 2, the period of non-working days is analyzed for the data recorded by transformer 1. From the three-cluster analysis, it can be seen that the data is divided into three distinct time intervals:

- 00:00 - 12:00: This interval is characterized by high power values, indicating a normal activity in this interval
- 12:00 - 00:00: This interval also sees high power values, suggesting continued normal energy use.
- 08:00 - 16:00: In this time window, the power values are medium and low, which reflects that the photovoltaic system supports the load, and the surplus is injected into the grid.

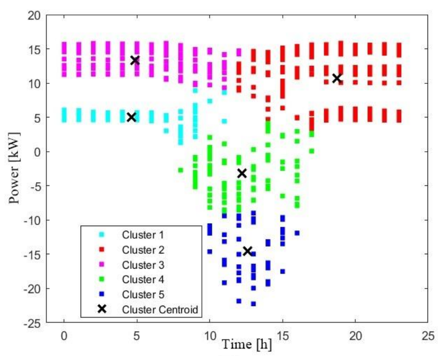
This structuring of the data allows a better understanding of energy consumption on non-working days and can contribute to more efficient energy resource management strategies.



a) 3 clusters



b) 4 clusters



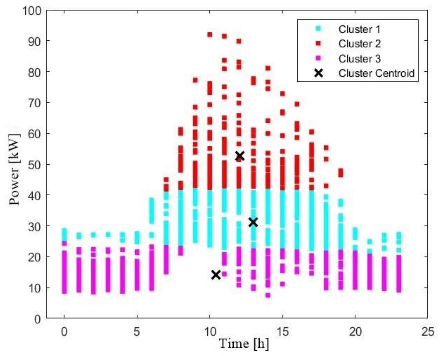
c) 5 clusters

Fig.2. Graphic representation for scenario 2-Transformer 1

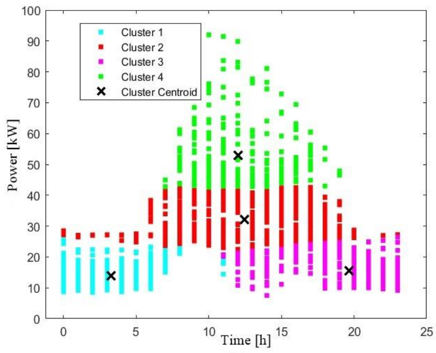
Analysis of the data in Fig. 3 and 4 suggest a clear structure in the power distribution recorded by Transformer 2. The division into five clusters provides a valuable breakdown of power variability throughout the day.

- Cluster 1 and Cluster 4: These clusters, characterized by low power values, reflect periods of inactivity or low consumption, which is expected in the mentioned time slots (00:00–08:00 and 12:00–00:00).
- Cluster 2: The extension of this cluster over a longer period and with average power values suggests constant consumption, possible during daily activities.
- Cluster 3: This is the most interesting because it contains the maximum values and the outliers. Identifying these outliers is crucial for anomaly analysis and can indicate periods of heavy usage.
- Cluster 5: Here we observe high values between 09:00-18:00, corresponding to peak hours, which suggests a higher consumption, probably due to teaching activities.

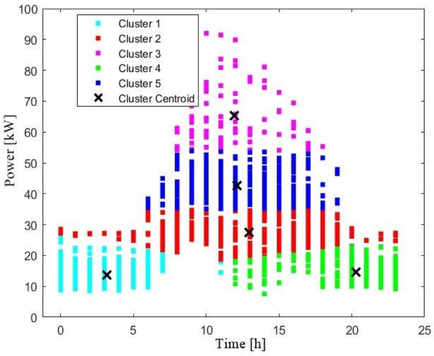
The outliers were defined by their values, with respect to the surrounding data characteristics. Therefore, if a measured value was differentiated with more than 15% of the adjacent values, it automatically was flagged and marked as an outlier.



a) 3 clusters

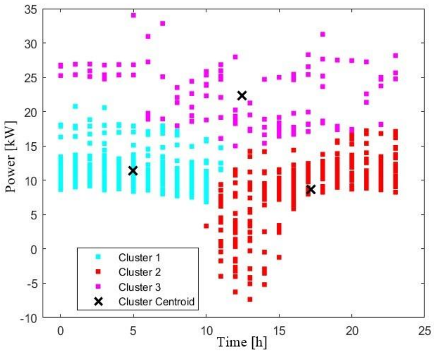


b) 4 clusters

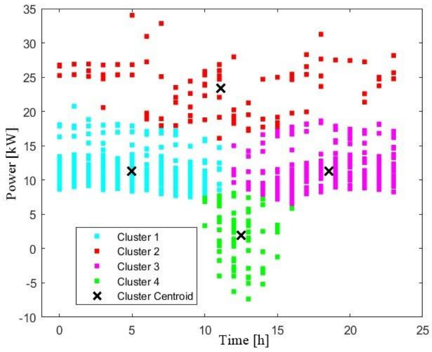


c) 5 clusters

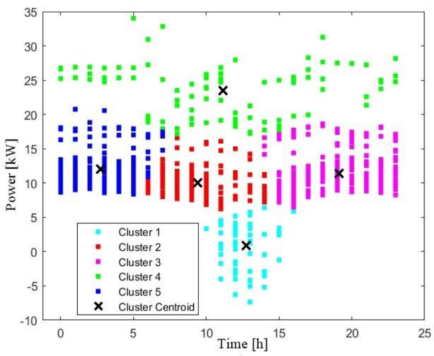
Fig.3 Graphic representation for scenario 1-Transformer 2.



a) 3 clusters



b) 4 clusters



c) 5 clusters

Fig.4. Graphic representation for scenario 2-Transformer 2

The calculation algorithm for determining the diagrams in the figures was run using the Matlab program and is presented in Fig.5.

```
dataset=xlsread ("ZL_P3f_Bus2.xls", 'Sheet1');
x=dataset(:,1);
y=dataset(:,2);
X=[x,y];
[idx,C] = kmeans(X,3);
figure
gscatter(X(:,1),X(:,2),idx,'crmgb','.',12)
hold on
plot(C(:,1),C(:,2),'kx','MarkerSize',10,'LineWidth',2)
legend('Cluster 1','Cluster 2','Cluster 3','Cluster Centroid')
%title 'Identificarea clusterelor si a centroizilor';
xlabel 'Time (h)';
ylabel 'Power (kW)';
hold off
```

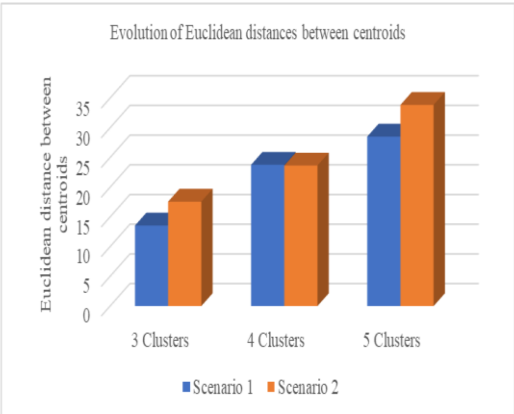
Fig.5. The calculation algorithm in Matlab

Analyzing table 1, we can see the existence of a correlation between the number of clusters and the distance between the centroids. Increasing the number of clusters can lead to a further spacing of the centroids, which can indicate a better segmentation of the data. Choosing an optimal number of clusters, in this case three, seems to balance this distance well with the efficiency of outlier analysis. The evolution of Euclidean distances is also presented in Fig. 6.

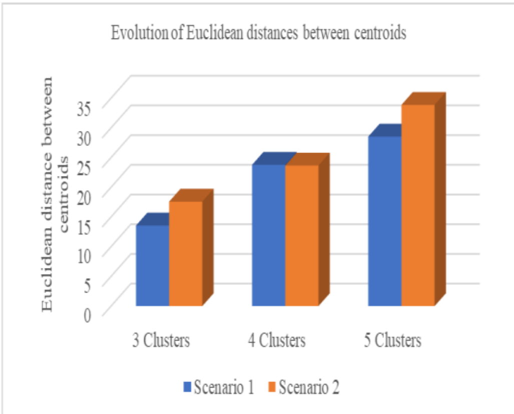
Table 1

Centroid distances – Transformer 1										
3 Clusters										
	C12	**C13**	**C23**							
Scenario 1	12,89	13,01	14,74							
Scenario 2	13,82	19,47	19,44							
4 Clusters										
	C12	**C13**	**C14**	**C23**	**C24**	**C34**				
Scenario 1	15,01	22,5	9,21	16,64	13,01	19,19				
Scenario 2	14,97	14,1	8,65	19,85	22,96	14,05				
5 Clusters										
	C12	**C13**	**C14**	**C15**	**C23**	**C24**	**C25**	**C34**	**C35**	**C45**
Scenario 1	7,87	11,66	15,34	15,05	18,84	10,49	15,39	22,18	15,84	9,96
Scenario 2	15,18	8,31	11,16	21,07	14,11	15,29	25,92	18,05	28,87	11,34

Centroid distances – Transformer 2										
3 Clusters										
	C12	C13	C23							
Scenario 1	21,52	17,12	38,45							
Scenario 2	12,54	13,26	14,42							
4 Clusters										
	C12	C13	C14	C23	C24	C34				
Scenario 1	40,23	6,59	40,14	38,11	0,44	38,19				
Scenario 2	13,55	7,21	12,09	14,22	21,56	11,17				
5 Clusters										
	C12	C13	C14	C15	C23	C24	C25	C34	C35	C45
Scenario 1	16,41	52,43	17,12	30,34	37,92	15,26	15,23	51,50	22,69	29,29
Scenario 2	9,75	12,37	22,67	15,03	9,83	13,56	6,99	14,42	16,42	14,18



a) Transformer 1



b) Transformer 2

Fig. 6. Evolution of Euclidean distances between centroids for both scenarios

5. Conclusions

The graphical representation of the data clearly highlighted the time zones of use as well as the identification of outliers. although the analysis showed that an increase in the number of clusters from three to five would provide a deeper granularity of the data, evaluation of the euclidean distances between centroids demonstrated that the optimal solution remains the three-cluster solution. this choice not only simplifies the analysis, but also minimizes possible errors, providing a clearer and more robust interpretation of the data.

In addition, the observation that the outliers identified are in fact normal behaviors in the context of laboratory use and brought a new dimension to the analysis. These fluctuations in consumption are attributed to specific activities within the faculty, such as laboratory activities or intensive courses, which may generate temporary increases in use. Thus, it is essential to consider these contextual aspects to avoid misinterpretation of outliers and to optimize resource management strategies.

This work contributes to a deeper understanding of energy consumption, depending on the type of day and time intervals, helping to optimize the management of energy resources within the institution.

REFERENCES

[1] *Hodge, V., Austin, J. A*, „A Survey of Outlier Detection Methodologies," Artificial Intelligence Review, vol. 22, pp. 85-126, 2004.

[2] *C. C. Aggarwal, P. S. Yu*, „An effective and effcient algorithm for high-dimensional outlier detection," The VLDB, vol. 14, pp. 211-221, 2005.

[3] *M. Breunig, H. Kriegel, R. T. Ng, and J. Sander*, „LOF: Identifying density-based local outliers," International Conference on Management of Data, vol. 29, pp. 93-104, 2000.

[4] *Z. Zheng, H. Y. Jeong, T. Huang, and J. Shu*, „KDE based outlier detection on distributed data streams in multimedia network," Multimedia Tools Appl, vol. 76, 2017.

[5] *Sheng, L. L*, „Fractal-based outlier detection algorithm over RFID data streams," International Journal of Online and Biomedical Engineering (iJOE), vol. 12, pp. 35-41, 2016.

[6] *D. van Hieu and P. Meesad*, „A fast outlier detection algorithm for big datasets," Advances in Intelligent Systems and Computing, vol. 463, 2015.

[7] *A. Ayadi, O. Ghorbel, A. M. Obeid, and M. Abid*, „Outlier detection approaches for wireless sensor networks: A survey," Computer Networks, vol. 129, pp. 319-333, 2017.

[8] *J. Mao, T. Wang, C. Jin and A. Zhou*, „Feature grouping-based outlier detection upon streaming trajectories," IEEE Transactions on Knowledge and Data Engineering, vol. 29, pp. 2696-2709, 2017.

[9] *C. D'Urso*, „EXPERIENCE: Glitches in databases, how to ensure data quality by outlier detection techniques," Journal of Data and Information Quality, vol. 7, 2016.

[10] *H. P. Kriegel, P. Kröger, E. Schubert, and A. Zimek*, „Outlier detection in axis-parallel subspaces of high dimensional data," Pacifc-Asia Conference on Knowledge Discovery and Data Miting, pp. 831-838, 2009.

[11] *H. Wang, M. J. Bah and M. Hammad*, „Progress in Outlier Detection Techniques: A Survey," IEEE Access, vol. 7, pp. 107964-108000, 2019.