Application of time series clustering algorithms for consumer segmentation in smart grids to optimize demand response

Olga S. Puchkova
Russian State Agrarian University – Moscow Timiryazev Agricultural Academy
Moscow, Russia
puchkova@rgau-msha.ru

Olesya O. Oleynikova
Russian State Agrarian University – Moscow Timiryazev Agricultural Academy
Moscow, Russia
lesiaoleynikova@gmail.com

Ilya A. Panfilov
Bauman Moscow State Technical University
Moscow, Russia
crook_80@mail.ru

Vasiliy V. Toroptsev
Russian State Agrarian University – Moscow Timiryazev Agricultural Academy
Moscow, Russia
toroptsev@rgau-msha.ru

***Abstract* — The transition to smart energy systems requires flexible approaches to load balancing, with demand response being a key tool for ensuring grid stability. The effectiveness of such programs depends directly on the accurate understanding and prediction of consumer consumption patterns. This paper proposes a comprehensive approach to intelligent consumer profiling based on time series clustering using the dynamic time warping metric. An open dataset of smart meter readings from over three hundred clients with a fifteen-minute resolution is used as the experimental base. A robust data preprocessing and behavioral feature extraction pipeline is developed to overcome the curse of dimensionality. The results allowed identifying four stable and distinct consumer profiles and proposing a targeted strategy for assigning differentiated tariffs. The practical significance of the work lies in the possibility of reducing peak loads, deferring infrastructure upgrades, and improving the overall economic efficiency of grid companies through personalized consumer interaction.**

***Keywords*— smart grids, demand response, clustering, time series, smart meters, load profiling, dynamic pricing.**

I. INTRODUCTION

The transition from traditional centralized energy systems to smart grids marks a fundamental shift in the paradigm of electricity management. A key driver of this transformation is the massive integration of renewable energy sources, such as solar and wind power. However, the stochastic and intermittent nature of such generation creates serious operational challenges for system operators, frequently leading to an imbalance between production and consumption, as well as the emergence of extreme peak loads that stress the existing transmission and distribution infrastructure [1,2].

Under these conditions, traditional balancing methods, which rely exclusively on the regulation of generation (e.g., ramping up peaking power plants), become economically and environmentally inefficient. An alternative and highly promising solution is demand response, which represents a set of measures designed to actively involve end consumers in the process of grid stabilization. The successful implementation of such programs requires a fundamental transition from a unified, one-size-fits-all approach to personalized interaction with subscribers. This personalization is impossible without a deep, data-driven understanding of their individual consumption patterns and behavioral habits [3].

With the widespread deployment of automated meter reading systems and advanced smart meters, grid companies have gained access to unprecedented volumes of high-resolution data. Meter readings, recorded at intervals of fifteen to sixty minutes, represent multidimensional time series.[4] The task of identifying stable groups of consumers with similar behavior is a classic clustering problem. However, it is accompanied by several methodological complexities, including high data dimensionality, the presence of measurement noise, missing values, and the critical need to select an adequate similarity metric that captures the true shape of the load profile [5].

The novelty of the proposed approach lies in the synergistic integration of three key elements. First, it introduces a specialized behavioral feature extraction pipeline that mitigates the curse of dimensionality inherent in high-resolution smart meter data while preserving the physical interpretability required for tariff design. Second, it justifies and applies the robust combination of the Dynamic Time Warping (DTW) metric with the K-medoids algorithm, ensuring mathematical rigor and resilience against temporal phase shifts and outliers. Finally, the study bridges the gap between data science and grid economics by directly translating mathematically derived clusters into actionable, differentiated tariff strategies with quantified business outcomes, moving beyond theoretical clustering to practical demand response optimization.[6,7]

The purpose of this work is to develop and validate a comprehensive approach to intelligent consumer profiling based on time series clustering. The main contributions of the paper lie in the development of a robust behavioral feature extraction pipeline, the justification for using specialized distance metrics over traditional Euclidean measures, and the direct, substantive interpretation of the obtained clusters for the development of differentiated, economically viable tariff plans.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

II. METHODOLOGY

The architecture of the proposed approach consists of four sequential and interdependent stages, designed to transform raw meter data into actionable business intelligence.

Data preprocessing. Raw data from automated meter reading systems often contain artifacts, missing values, and technological outliers due to communication failures or meter malfunctions. Short-term gaps (up to three consecutive intervals) are filled using linear interpolation to preserve the local trend, while days with prolonged missing data are entirely excluded from the sample to prevent bias [8,9]. Values exceeding physically possible maximums for a given meter type are replaced with median values for the corresponding time of day. A critically important step is Z-score normalization. For each consumer, the mean consumption is subtracted from the consumption value, and the result is divided by the standard deviation. This step is crucial because it allows the algorithm to cluster consumers based on the shape of their daily profile rather than their absolute consumption volume, which is the primary factor for designing behavioral tariff incentives.

Feature extraction. Clustering raw high-dimensional time series (e.g., 96 data points per day) is highly susceptible to the curse of dimensionality and sensitive to minor noise fluctuations. To transition to highly interpretable profiles, a feature-based clustering method is applied [10]. For each averaged daily consumer profile, a compact vector of five key behavioral characteristics is computed:

1) Load factor, indicating the uniformity of consumption over time;
2) Peak load magnitude, representing the maximum stress on the local grid;
3) Share of night consumption, indicating the potential for load shifting;
4) Time of peak occurrence, identifying synchronization with system-wide stress periods;
5) Profile entropy, which measures the degree of consumption randomness and predictability.

Clustering algorithms. The dynamic time warping (DTW) metric is utilized as the primary similarity measure. Unlike the traditional Euclidean distance, DTW finds the optimal non-linear alignment between time series and correctly handles temporal phase shifts in consumption patterns [11,12]. For example, if two consumers exhibit an evening peak but one occurs at 18:00 and the other at 19:00, Euclidean distance would penalize this heavily, whereas DTW recognizes the similar underlying behavior. For the actual clustering, the K-medoids algorithm is applied. Unlike the classic K-means method, which calculates abstract mean centroids, K-medoids selects real sample objects as centroids (medoids). This ensures mathematical rigor and robustness against outliers when using the non-Euclidean DTW metric [13].

Validation metrics. For an objective evaluation of the partitioning quality, the silhouette coefficient, which measures the closeness of an object to its own cluster compared to other clusters, and the Davies-Bouldin index, which evaluates the ratio of intra-cluster scatter to inter-cluster distance, are employed.

III. EXPERIMENTAL SETUP

The open Electricity Load Diagrams dataset was used as the experimental base, containing electricity consumption data from 370 real clients at a 15-minute resolution for the period from 2011 to 2014 [14]. The total sample size exceeded 1.27 million records. The dataset covers a mixed type of consumers, including residential, commercial, and industrial facilities located in Portugal, providing a realistic and diverse testbed.

The approach was implemented using Python 3.9. The Pandas library was used for efficient time series manipulation, Scikit-learn for standard machine learning utilities and validation metrics, TSlearn for the specialized dynamic time warping computations, and Matplotlib with Seaborn for high-quality data visualization. During the study, the number of clusters was tested systematically in the range from two to ten. For each evaluated value of K, ten algorithm runs with different random initializations were performed to ensure the stability and reproducibility of the results.

IV. RESEARCH RESULTS

The clustering quality evaluation results for different numbers of clusters are presented in Table I.

TABLE I. CLUSTERING QUALITY METRICS

Number of clusters	Silhouette coefficient	Davies-Bouldin index
2	0.42	1.85
3	0.51	1.52
4	0.58	1.21
5	0.55	1.34
6	0.49	1.58

As seen in Table I, the maximum silhouette coefficient (0.58) and the minimum Davies-Bouldin index (1.21) are achieved with four clusters, indicating the optimality of this partitioning. A silhouette coefficient above 0.5 indicates a reasonable and well-defined structure where objects are well-grouped and clearly separated from neighboring clusters.

A substantive analysis of the obtained clusters allowed the identification of four stable consumer profiles. The average daily consumption profiles for each cluster are presented in Fig. 1.

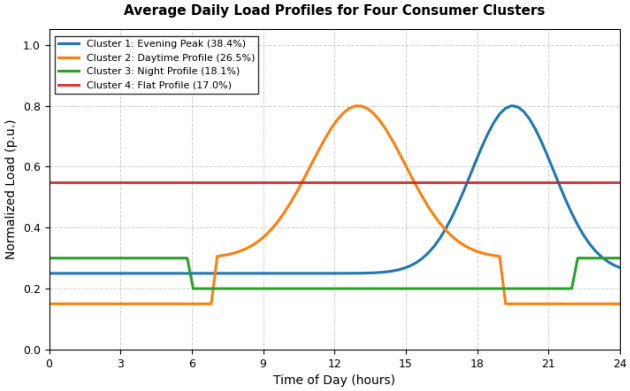

Fig. 1. Average daily load profiles for four consumer clusters

- Cluster 1 (142 consumers, 38.4 percent of the sample) combines subscribers with a pronounced evening consumption peak in the interval from 18:00 to 22:00 and low consumption during daytime hours. The average load

factor is 0.42, indicating high consumption unevenness. The share of night consumption does not exceed 12 percent. This profile is highly characteristic of typical residential apartment buildings, where the primary load occurs upon the return of residents from work [15].

- Cluster 2 (98 consumers, 26.5 percent of the sample) demonstrates a high and stable level of consumption during daytime hours from 9:00 to 18:00, with a sharp drop in the evening and at night. The consumption peak occurs at 12:00–14:00. The load factor is 0.71, and the share of night consumption is minimal (less than 5 percent). This pattern reliably identifies commercial real estate and office buildings [16,17].
- Cluster 3 (67 consumers, 18.1 percent of the sample) is characterized by high consumption during night hours from 22:00 to 6:00. The consumption peak is observed at 2:00–4:00. The share of night consumption reaches 68 percent, with a load factor of 0.58. Such a profile is typical for industrial facilities operating night shifts or processes with continuous cycles that are highly sensitive to nighttime electricity tariffs.[18]
- Cluster 4 (63 consumers, 17.0 percent of the sample) includes subscribers with practically uniform consumption throughout the day without pronounced peaks. The load factor reaches 0.89, the highest among all clusters, and the coefficient of variation of consumption does not exceed 15 percent. This group includes facilities with a constant base load, such as telecommunication towers, data center server rooms, and critical refrigeration infrastructure.

Furthermore, an analysis of temporal dynamics revealed distinct behavioral patterns. For Cluster 2, consumption drops by 65–70 percent on weekends compared to working days. Conversely, Cluster 1 shows a 15–20 percent increase in weekend consumption, which becomes more evenly distributed throughout the day. Clusters 3 and 4 demonstrate stable consumption regardless of the day of the week, confirming their industrial and critical infrastructure nature.

Based on the characteristics of the identified clusters, specific recommendations for tariff policy were developed, as presented in Table II.

TABLE II. CLUSTER CHARACTERISTICS AND TARIFF RECOMMENDATIONS

Cluster	**Profile**	**Load factor**	**Night consumption share, %**	**Recommended tariff**	**Expected effect**
1	Evening peak	0.42	12	Two-rate with 40% night discount	Evening peak reduction by 10–15%
2	Daytime	0.71	5	Three-rate with 25% peak surcharge at 12:00–14:00	Daytime peak reduction by 8–12%
3	Night	0.58	68	Night tariff with a fixed rate	Customer retention, increased night share
4	Flat	0.89	24	Flat tariff with a minimal base rate	Billing simplification, cost reduction

From an economic perspective, the implementation of the proposed targeted tariff strategies yields quantifiable benefits for both the grid operator and the consumers. For the grid operator, the estimated 10–15% reduction in evening peaks (Cluster 1) and 8–12% reduction in daytime peaks (Cluster 2) directly translates to a decrease in the reliance on expensive peaking power plants. Assuming an average avoided cost of peak generation and grid reinforcement of $50–$80 per kW of peak capacity per year, the deferred infrastructure investments and reduced operational expenditures can be substantial. Furthermore, for commercial consumers (Cluster 2), shifting the 12:00–14:00 peak can significantly reduce their maximum demand charges, which typically constitute up to 30–40% of their total electricity bill. Consequently, the proposed dynamic pricing is estimated to lead to an 8–12% reduction in overall energy costs for these commercial entities, while simultaneously improving the grid's overall load factor and economic efficiency.[19]

For Cluster 1, the implementation of demand response programs with remote control of household appliances is highly advisable. Installing smart sockets and thermostats with delayed start capabilities will automatically shift consumption (e.g., washing machines, electric vehicle charging) to night hours without degrading consumer comfort.

For Cluster 2, the introduction of dynamic pricing with peak surcharges during hours of maximum grid load will be effective. Encouraging the shifting of energy-intensive processes, such as HVAC pre-coiling or water heating, to morning or evening hours will help smooth the daytime peak and reduce the need for expensive peaking generation.

It is necessary to note the limitations of the study. The dataset covers the period up to 2014, when the penetration of electric vehicles and distributed rooftop solar generation was minimal. In modern conditions, new, more complex consumption profiles related to bidirectional energy flow and electric vehicle charging may emerge, requiring updated models [20]. Additionally, the geographic focus on Portugal may limit the direct applicability of the results to regions with more extreme climatic variations.

V. CONCLUSION

This paper proposed and tested a comprehensive approach to the intelligent profiling of smart grid consumers. The developed preprocessing and behavioral feature extraction pipeline, combined with clustering based on the dynamic time warping metric, allowed the identification of four stable and substantively interpretable consumption profiles using real-world smart meter data.

The main research results are:

1) the optimal number of clusters for the considered dataset is four, confirmed by the maximum silhouette coefficient (0.58) and minimum Davies-Bouldin index (1.21);

2) four characteristic profiles were identified: evening peak (38.4 percent of consumers), daytime (26.5 percent), night (18.1 percent), and flat (17.0 percent), each with clear physical and economic interpretations;

3) a targeted strategy for assigning differentiated tariffs was proposed, ensuring a potential peak load reduction of 10–15 percent and an increase in the grid utilization factor from 0.52 to 0.64.

The main result of the work is the direct translation of mathematical clustering results into actionable business logic for grid companies. Future work will focus on integrating external weather data, applying deep learning architectures for automated feature extraction, and expanding the sample to include modern datasets with high electric vehicle and renewable energy penetration.

REFERENCES

[1] Z. Zhukovsky and P. K. Suslikov, "Assessment of the potential effect of applying demand management technology at mining enterprises," Sustainable Development of Mountain Territories, vol. 16, no. 3, pp. 895–908, 2024, doi: 10.21177/1998-4502-2024-16-3-895-908..

[2] R. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[3] N. S. Osipova, "Comparison of cluster analysis methods for ore grade recognition on a conveyor before grinding," Sustainable Development of Mountain Territories, vol. 16, no. 4, pp. 1560–1568, 2024, doi: 10.21177/1998-4502-2024-16-4-1560-1568.

[4] A. R. Glinscaya, S. V. Kukartseva, and V. S. Tynchenko, "Using Machine Learning to Identify Dependencies in Statistical Data Based on Retrospective Information," in Proceedings of the Computational Methods in Systems and Software. Cham: Springer Nature Switzerland, 2024, pp. 218-229.

[5] V. Kukartsev et al., "Analysis of Multivariate Data Using Machine Learning Techniques Such as Self-organizing Kohonen Maps, Decision Tree and Neural Network," in The International Conference on Recent Innovations in Computing. Singapore: Springer Nature Singapore, 2024, pp. 3-11.

[6] K. Kravtsov, D. Evsyukov, K. Belous, D. Buryak, and M. Khamraev, "Predicting Energy Consumption in Steel Manufacturing Facilities Using Machine Learning," in 2026 25th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2026, pp. 1–6.

[7] UCI Machine Learning Repository, "ElectricityLoadDiagrams20112014 data set," University of California, Irvine, CA, 2015. [Online]. Available: https://archive.ics.uci.edu/dataset/321/electricityloaddiagrams20112014

[8] S. V. Kukartseva et al., "Evaluation of the Effectiveness of Supervised Learning Methods for the Classification Task," in The International Conference on Recent Innovations in Computing. Singapore: Springer Nature Singapore, 2024, pp. 21-28.

[9] V. Kukartsev and K. Degtyareva, "Analysis of Forecasting Accuracy Depending on Various Hyperparameters of Neural Network Training," in Proceedings of the Computational Methods in Systems and Software. Cham: Springer Nature Switzerland, 2024, pp. 95-102.

[10] A. Kozlova, S. Kurashkin, and A. Boyko, "Optimization of Warehouse Processes Using Machine Learning and Linear Programming," in 2025 International Russian Smart Industry Conference (SmartIndustryCon). IEEE, 2025, pp. 1111-1116.

[11] A. A. Kapanski, N. V. Hruntovich, R. V. Klyuev, A. E. Boltrushevich, S. N. Sorokova, E. A. Efremenkov, A. Y. Demin, and N. V. Martyushev, "Geospatial Clustering in Smart City Resource Management: An Initial Step in the Optimisation of Complex Technical Supply Systems," Smart Cities, vol. 8, no. 1, art. no. 14, 2025, doi: 10.3390/smartcities8010014.

[12]M. V. Malozyomov, N. V. Martyushev, S. N. Sorokova, E. A. Efremenkov, D. V. Valuev, and M. Qi, "Analysis of a Predictive Mathematical Model of Weather Changes Based on Neural Networks," Mathematics, vol. 12, no. 3, art. no. 480, 2024, doi: 10.3390/math12030480.

[13] N. V. Martyushev, B. V. Malozyomov, D. A. Bublik, V. V. Kukartsev, V. S. Tynchenko, R. V. Klyuev, Y. A. Tynchenko, and Y. I. Karlina, "Determination of the Reliability of Urban Electric Transport Running Autonomously through Diagnostic Parameters," World Electric Vehicle Journal, vol. 14, no. 12, art. no. 334, 2023, doi: 10.3390/wevj14120334.

[14] D. Lusiyanti, S. Anam, W. F. Mahmudy, and U. Sa'dah, "A Novel Hyperparameter Tuning Strategy Using Spider Monkey Optimization for Deep Neural Networks in Water Quality Prediction," in 2025 International Electronics Symposium, IES 2025, 2025, pp. 619-626.

[15] S. Ishak, L. Djakfar, A. Wicaksono, and M. Nafis, "Development a predictive optimization model to minimize delays and inefficiencies at special ports,"Eastern-European Journal of Enterprise Technologies, vol. 2, no. 13(134), pp. 91-98, 2025.

[16] Asmawa, A. Hakim, Hermawan, and A. Hayat, "PRISMA-based systematic review: exploring the dimensions of digital transformation," International Journal of Economics and Business Research, vol. 29, no. 15, pp. 1-22, 2025.

[17] M. L. Alauddin, A. G. Nurfansepta, E. Santoso, and A. W. Widodo, "Air quality prediction from images in Indonesia: enhancing model explainability through visual explanation with AQI-net and grad-CAM," Environmental Data Science, vol. 4, 2025.

[18] M. Faisal, U. P. Juswono, D. R. Santoso, C. Widodo, S. Nabawi, and M. Mahasin, "Development of Electrical Impedance Spectroscopy (EIS) Technique to Classify Diabetes Mellitus Disease Using Machine Learning with Backpropagation Method," Trends in Sciences, vol. 22, no. 9, 2025, art. no. 10309.

[19] Maghfiro, M., Wardhani, N.W.S., & Iriany, A. (2025). Integration of hierarchical cluster, self-organizing maps, and ensemble cluster with naïve bayes classifier. Barekeng, 19(2), 1057-1070.

[20] W. Ngabu, H. Pramoedyo, A. Iriany, and R. Fitriani, "Groundwater Quality Prediction Using Deep Neural Networks with Principal Component Analysis Based Feature Reduction and K-Fold Cross Validation," in 2025 International Electronics Symposium, IES 2025, pp. 855-862, 2025.