Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2025, № 4 (120) https://doi.org/10.33108/visnyk_tntu2025.04
ISSN 2522-4433. Web: ***visnyk.tntu.edu.ua***

UDC 004.62

MODELLING RESIDENTIAL ELECTRICITY CONSUMPTION

Tetyana Mamchych; Ivan Mamchych

Lesya Ukrainka Volyn National University, Lutsk, Ukraine

Abstract. *Our work is devoted to the problem of modeling residential electricity consumption based on time series formed by sequential values of metering devices related to smart-grid technologies. Residential consumption is an important component of the overall energy system. The study of this component is of particular importance in our time of epidemics, the spread of remote types of work, and increased dependence on Internet technologies. Mathematical models for the nature of consumption that allow us to identify patterns, to identify similarities in such patterns would be extremely useful for the tasks of short-term forecasting and demand prediction, and for ensuring the stability of the functioning of the energy supply system as a whole. We note that there is a special need for models that use only data from smart-grid devices, without involving other types of data, such as the number of inhabitants, income, area of the dwelling, and others, for remote monitoring based on current data. In the works [1] and [2], the coefficient of similarity and the coefficient of auto-similarity were first introduced to describe the nature of consumption, identify possible patterns and measure the stability of these patterns, to determine cases when such patterns do not exist. The cited works on real data demonstrate the effectiveness of these coefficients for monitoring consumption, and the results obtained are an achievement in the field of energy. At the same time, these works do not contain the study of these computational structures from the point of view of applied mathematics, since this is beyond the scope of energy science. This work is devoted to filling this gap. In our work, some properties of the coefficient of similarity and the coefficient of auto-similarity associated with the use of the correlation approach have been established, and the recognition ability of the coefficients for fixed data has been studied, compared with the Szekely's coefficient of the distance correlation) ([3] and [4]). The testing of this technology has been performed on data obtained by the project [5].*
Key words: *the coefficient of similarity, the coefficient of auto-similarity, modelling, residential electricity consumption, statistical analysis, correlation, smart-grid, classification.*

Submitted 02.09.2025
Revised 01.12.2025
https://doi.org/10.33108/visnyk_tntu2025.04.032 *Published 27.01.2026*

1. INTRODUCTION

The overall goal of our research is to build a mathematical model of residential electricity consumption based on time series of consumption indicators for the purpose of detailed study of individual consumption graphs, and to develop measurement technology for classification tasks.

Residential consumption is an important part of total electricity consumption. According to Eurostat, in 2023, household electricity consumption accounted for 26.2% of total consumption in the EU ([6]). In the U.S., this proportion is about a third ([7]). While industrial consumption is more predictable due to production and technological reasons, residential consumption is extremely diverse. Different users differ in their electricity consumption, and even the same household does not repeat its consumption figures exactly from one day to the next. Identifying patterns in consumption, identifying similarities, and classifying individual consumers is a challenge, and developing approaches for such modeling is an ambitious task.

Household electricity consumption depends on factors such as the nature of residents' activities, work schedules, the presence of minors, the set of electrical appliances, income, area, and even behavioral habits, etc. A new factor influencing private consumption is the charging of electric vehicles ([8]). Some of these factors are independent, some are directly or indirectly dependent. Such data are generally unavailable. Whereas data obtained from modern metering devices that provide detailed recording of consumption – the so-called smart-grid systems – are available (a review of these technologies is devoted, in particular, to the work [9]). The spread of smart metering technologies has given impetus to research into electricity consumption and stimulated the development of mathematical modeling of behavioral patterns of individual consumers. Publications [10] and [11] are examples of such research. In particular, clustering of observed series is carried out and averaged patterns are formed for these clusters. Clustering techniques, regressions, adaptive machine learning algorithms and their modifications are used for consumer grouping and demand forecasting ([12], [13]). The authors of [14] indicate that «Research on Data Analytics in the electricity sector is growing faster than related fields of research».

The input data of the research are indicators that are recorded at the same time interval (say, one hour), forming a sequence of values, which is mathematically a time series. From the point of view of energy scientists, computational technologies based exclusively on these time series are appropriate, this would make it possible to predict, classify, compare consumers, that is, to study problems in the field of marketing and energy. Such approaches are relevant for the general problem of consumption in energy, and not only for electricity. In [15], periodic empirical series in the field of gas supply are studied.

In the noted works, the problems of pattern identification and classification are considered, but these models are not developed to the level of formal metric spaces with a metric introduced on them. A certain step in this direction is the works [1] and [2], where the similarity coefficient for two different consumption series and the self-similarity coefficient for one series are introduced, which is a measure of how much this series repeats the nature of consumption over successive periods are considered. In these works, a modified correlation approach was used, since the traditional application of Pearson's pairwise correlation turned out to be insufficient.

In our problem, the input data is a set of time series, each of which represents a household. These series can be of different lengths and correspond to different time periods.

Let $X=\{x_i, i=1, 2, \ldots, 24n\}$ та $Y=\{y_i, i=1, 2, \ldots, 24m\}$ be time series of consumption readings corresponding to two households, with the first of them observed for n days, and the second for m days, 24 corresponding to the number of hours in a day. Let us assume that the values are synchronized by i, i.e. i = 1, 25, 49 and so on correspond to the first hour of the day, and in the same pattern the other values. It is necessary to compare these two time series for similarity in a certain sense. In the case of the same length of the series and complete synchronization in time, one could traditionally use the Pearson correlation coefficient, but these conditions are not met in our case. For example, we need to compare series that refer to different seasonal time intervals, here it is clear that winter and summer trends are different, and patterns during the holiday period are also special. Therefore, a direct appeal to Pearson correlation, as in autoregressive models, is unsuitable for our purpose.

The coefficient, which is a measure of similarity between two numerical sequences, is called the coefficient of distance correlation (or «energy statistics»), and was proposed by Gabor Szekely in 2008 for problems of our type ([3], [4]). This coefficient can take values in the range from 0 to 1 as the similarity measure increases. This coefficient is integrated into the «Energy» package of the R programming language ([16]).

Example 1.

The first step of our study was to test the capability of this technology for a specific set of real data. Having at our disposal real observation data of Swedish households, we selected five time series corresponding to the consumption indicators of five apartments during a month with an interval of one hour, we will conditionally label these series A, B, C, D, E. Using the «Energy» package of the R language, we calculated pairwise distance correlation coefficients, which are given in Tab. 1.

Table 1

The distance correlations of Gabor Szekely for households A, B, C, D, E

	Households				
	A	B	C	D	E
A	1	0.989	0.987	0.980	0.946
B		1	0.986	0.984	0.940
C			1	0.993	0.972
D				1	0.949
E					1

As we can see, all the obtained values are close to 1, which means strong similarity. Our attempts to detect a pair of weakly similar series on other data were also unsuccessful. In our opinion, this coefficient does not allow us to clearly demonstrate the similarity measure for our data. Perhaps the general patterns of daytime activity versus nighttime activity and the traditional daily routine provide such a «similarity» of these time series that this measure is insensitive to such data. Therefore, there is a need to develop computational recognition technologies for data of the appropriate scale.

2. DATA

The calculations were based on real-world data from a survey of 400 Swedish households in 2008 ([6]). The study involved households of various types and sizes. For some cases, the observation period was one month, for others, one year. The households are represented by time series of electricity consumption, which for our study were converted to one-hour series.

3. METHOD

In [1] and [2], it was first proposed to use the coefficient of similarity as a measure of similarity for two periodic time series, and the coefficient of auto-similarity as a measure of stability for one time series.

Let X and Y be time series with the same period p (in [1] p = 24, which corresponds to 24 hours in a day, but in this paper we will use a more general formulation of the definition with an arbitrary p). The series can be represented in the form of matrices of dimensions n x p and m x p, where n and m are the number of observation periods for the two series from X and Y, respectively.

Then each of the series can be divided into fragments of length p and presented as a matrix with p columns, placing successive fragments one under the other: $X = (x_{ij}), i = \overline{1,n}, j = \overline{1,p}$.

Let x_i, (i=1, 2, …, n) and y_j (j= 1, 2, …, m) be vectors whose elements consist of the elements of the corresponding rows of matrices X and Y, each of which has length p. Then $X = (x_1^T, x_2^T, \ldots, x_n^T)$ and $Y = (y_1^T, y_2^T, \ldots, y_m^T)$.

For the coefficients of similarity and autosimilarity, we use the following definitions from [1].

Definition 1. The coefficient of similarity for periodic time series with the same period X and Y is the value

$$T(X,Y) = \sqrt[nm]{\left|\prod_{i=1}^{n}\prod_{j=1}^{m} cor\left(x_i, y_j\right)\right|}, \qquad (1)$$

where $x_i = \left(x_{ik}, k = \overline{1,p}\right)$ and $y_j = \left(y_{jk}, k = \overline{1,p}\right)$; p is a period; n and m are the numbers is the number of fragments of length p in the series X and Y, respectively;
cor(.,.) is the Pearson correlation coefficient.

Definition 2. The coefficient of autosimilarity for a periodic time series X is the quantity

$$T(X) = T(X,X). \qquad (2)$$

The coefficient introduced by formula (1) is the geometric mean of the pairwise Pearson correlation coefficients for all possible pairs (x_i, y_j)(i=1, 2, …, n та j =1, 2, …, m). The series are assumed to be synchronized in phase by the period. It is also assumed that the series have a total length that is a multiple of p.

Theorem 1.

The coefficient defined in (1) has the following properties:

1. $\forall X, \forall Y \; 0 \leqslant T(X,Y) \leqslant 1$;
2. $\forall X$, $\forall Y$ the coefficient T(X,Y) is symmetric with respect to X and Y: T(X,Y) = T(Y,X);
3. $\forall X$, $\forall Y$ the coefficient T(X,Y) is defined up to permutations between x_i (i=1, 2, …, n) and between y_j (j=1, 2, …, m);
4. if there is a permutation between x_i (i = 1, 2, …, n) or between y_j (j = 1, 2, …, m) such that Y takes the form of multiple repetitions of a fragment that is proportional to X, then T(X,Y) = 1;
5. In the case where both matrices X and Y consist of rows that are repeated $T(X,Y) = |cor(x_1, y_1)|$;
6. $\forall X \; 0 \leqslant T(X) \leqslant 1$.

Proof.

1. By definition $T(X,Y) = \sqrt[nm]{\left|\prod_{i=1}^{n}\prod_{j=1}^{m} cor\left(x_i, y_j\right)\right|}$, therefore $T(X,Y) = \sqrt[nm]{\prod_{i=1}^{n}\prod_{j=1}^{m}\left|cor\left(x_i, y_j\right)\right|}$.

The resulting value is the geometric mean of the values that are the absolut values of the correlation coefficients. The modulus of the correlation coefficient is in the interval [0; 1]. Since the geometric mean does not exceed the arithmetic mean for an arbitrary set of numerical values, and the arithmetic mean does not exceed the largest value from this set, it follows that $0 \leqslant T(X,Y) \leqslant 1$.

2. The symmetry of the coefficient T(.,.) follows from the symmetry of the correlation coefficient.

3. The commutativity of the multiplication operation ensures the fulfillment of property 3.

4. The proof of this property follows from the fact that the coefficient of Pearson correlation is 1 for proportional vectors, and multiple repetition of the same fragment in the series Y does not change this value.

Remark. Property 4 also holds if X and Y are interchanged in the formulation.

5. In the case when both matrices X and Y consist of repeating rows, all root factors coincide with cor(x1,y1), therefore the indicated equality is fulfilled.

6. The proof of property 6 follows from property 1.

Let us consider the application of these coefficients to the time series used in Example 1.

Example 2.

For the time series A, B, C, D, E used in example 1, the values of the coefficients T(.,.) and T(.) were calculated for the specified period parameter p = 24 * 7 = 168, i.e. the component vectors were synchronized by weeks (168 hours per week). The period 168 was chosen for the following reasons: when comparing weeks, one can expect a stronger similarity in the nature of consumption than when comparing working and weekend days.

Table 2

The coefficients of similarity and the coefficients of autosimilarity (on the diagonal) for households A, B, C, D, E

	Households				
	A	B	C	D	E
A	0.671	0.459	0.563	0.496	0.302
B		0.686	0.573	0.486	0.332
C			0.796	0.603	0.424
D				0.586	0.241
E					0.604

Tab. 2 presents the calculated values of the coefficients T(.,.) (the values of T(.) are on the diagonal, which means the case X = Y). Calculations on the selected data show that the coefficients T(.,.) and T(.) are sensitive enough to distinguish similarities of different degrees. Already for these 5 households there is a spread of the values of the similarity coefficient from the smallest value T(D,E) = 0.241 to the largest T(B,C) = 0.573. The autosimilarity coefficient is also different, here it has a spread from T(D) = 0.586 to T(C)= 0.796. That is, unlike the distance correlation coefficient used in example 1, the use of the similarity coefficient (1) provides better opportunities as a measure of similarity between these time series.

4. DISCUSSION

The use of the coefficient of autosimilarity allows us to measure how stable the consumption pattern (mode) is for an individual household. The value of the coefficient reflects the extent to which it is possible to speak of the presence of a certain pattern for a fixed consumer. Small values of the coefficient of autosimilarity indicate that consecutive periods do not provide similarity in the consumption schedule, and therefore this case is poorly suited for forecasting, in contrast to cases with high coefficient values. Our study, which has a methodological focus, substantiates the suitability of the both coefficients for use in household electricity consumption analysis tasks.

5. CONCLUSIONS

The paper defines some properties of the coefficients of similarity and autosimilarity that follow from the mathematical construction of these quantities. Attention is also paid to the testing of the coefficients on real data. The results obtained indicate the effectiveness of the coefficient of similarity as a measure of similarity for two synchronized periodic series.

The coefficient of autosimilarity is used to measure the stability of one series. This coefficient is appropriate to use to find out whether an individual consumer demonstrates a sufficiently stable consumption pattern so that any pattern can be attributed to him at all.

In this work, we limited ourselves to testing the coefficients on a small number of examples, since the subject of the study is the modeling technology itself, and not the results in energy science.

This methodology contains potential opportunities for research in the energy industry. For example, by changing the period p from 24 (day) to 24*7 (week) and 24*30 (month), the range of tasks can be diversified. The stability measure actually introduces another dimension into the model, which can be taken into account in classifications.

We believe that the developed technology can have practical applications for analyzing data such as empirical periodic time series, in particular, for marketing tasks.

ACKNOWLEDGMENT

The work is a development of previous research within the project «Power of the Consumer» hosted by the Mälardalen University in Västerås, Sweden. Our data come from the project [5].

References

1. Mamchych T., Wallin F. (2014) Looking for patterns in residential electricity consumption. Energy Procedia 61, pp. 1768–1771. Doi:10.1016/j.egypro.2014.12.208.
2. Mamchych T., Wallin F. (2015) Stability of patterns in residential electricity consumption. Energy Procedia 75, pp. 2738–2744. https://doi.org/10.1016/j.egypro.2015.07.494
3. Székely, Gábor J., Rizzo, Maria L., Bakirov, Nail K. (December 2007). “Measuring and testing dependence by correlation of distances”. The Annals of Statistics. 35 (6): 2769–2794. https://doi.org/10.1214/009053607000000505
4. Szekely G. J. and Rizzo M. L. (2017) The Energy of Data, The Annual Review of Statistics and Its Applications. Extended Review, 4:447-479. https://doi.org/10.1146/annurev-statistics-060116-054026
5. Zimmermann J. P. (2009). End-use metering campaign in 400 households in Sweden. Assessment of the potential electricity savings. Enertech.
6. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_consumption_in_households.
7. Mingyue Guo, Youngsik Choi, So-Min Cheong, Zheng O’Neill (2025) Current and future residential electricity demand using large-scale smart meter data in a changing climate. Sustainable Cities and Society, 130 (15), 106623. https://doi.org/10.1016/j.scs.2025.106623
8. Mavlonov J., Ruzimov S., Tonoli A., Amati N., Mukhitdinov A. (2023) Sensitivity Analysis of Electric Energy Consumption in Battery Electric Vehicles with Different Electric Motors. World Electr. Veh. J. 14, 36. https://doi.org/10.3390/wevj14020036
9. Matsui K., Yamagata Y., Kawakubo Sh. (2019) Real-time sensing in residential area using IoT technology for finding usage patterns to suggest action plan to conserve energy. Energy Procedia 158, pp. 6438–6445. https://doi.org/10.1016/j.egypro.2019.01.171
10. Williams B., Hooper R. J., Gnoth D. and Chase J. G. (2025) Residential Electricity Demand Modelling: Validation of a Behavioural Agent-Based Approach. Energies, 18 (6), 1314. https://doi.org/10.3390/en18061314
11. Pak W., Kim I., Choi J. (2021) Proposal of the energy consumption analysis process for the residential houses using big data analytics technique. Journal of Computational Design and Engineering, 8 (6), pp. 1591–1604. https://doi.org/10.1093/jcde/qwab063
12. Softah W., Tafakori L., Song H. (2025) Analyzing and predicting residential electricity consumption using smart meter data: A copula-based approach. Energy and Buildings, 332 (1), 115432 https://doi.org/10.1016/j.enbuild.2025.115432
13. Hosseini B. (2025) Forecasting household monthly electricity consumption using the similar pattern algorithm. Academia Green Energy, 2. https://doi.org/10.20935/AcadEnergy7500
14. Scheidt F., Medinová H., Ludwig N., Richter B., Staudt Ph., Weinhardt Ch. Data analytics in the electricity sector –A quantitative and qualitative literature review. Energy and AI 1 (2020) 100009. https://doi.org/10.1016/j.egyai.2020.100009

15. Lytvynenko Ia., Lupenko S., Nazarevych O., Shymchuk H., Hotovych V. (2021) Additive mathematical model of gas consumption process. Scientific Journal of TNTU (Tern.), vol. 104, no. 4, pp. 87–97. https://doi.org/10.33108/visnyk_tntu2021.04.087
16. Maria Rizzo, Gabor Szekely. (2025) Package ‘energy’. E-Statistics: Multivariate Inference via the Energy of Data. Available at: https://cran.r-project.org/web/packages/energy/energy.pdf.

УДК 004.62

МОДЕЛЮВАННЯ ЕЛЕКТРОСПОЖИВАННЯ ДОМОГОСПОДАРСТВ

Тетяна Мамчич; Іван Мамчич

Волинський національний університет імені Лесі Українки, Луцьк, Україна

Резюме. *Роботу присвячено проблемі моделювання споживання електроенергії приватними домогосподарствами на основі часових рядів, утворених послідовними значеннями облікових приладів, які відносяться до smart-grid технологій. Електроспоживання домогосподарств є важливою складовою загальної енергетичної системи. Дослідження цієї складової набуває особливої ваги у наш час епідемій, поширення віддалених видів робіт, посилення залежності від інтернет-технологій. Математичні моделі характеру споживання, які дозволяють виявляти шаблони, виявляти подібності таких шаблонів були б надзвичайно корисними для задач короткострокового прогнозування і передбачення попиту, та для забезпечення стабільності функціонування системи енергозабезпечення в цілому. Зауважимо, що існує особлива потреба в моделях, які використовують лише дані smart-grid приладів, не залучаючи іншого типу даних, як кількість жителів, дохід, площа помешкання та інших для віддаленого моніторингу за поточними даними. У працях [1] і [2] було запропоновано коефіцієнт подібності (the coefficient of similarity) та коефіцієнт автоподібності (the coefficient of auto-similarity) для описування характеру споживання, виявлення можливих шаблонів та вимірювання стабільності шаблонів споживання, для визначення випадків, коли такі шаблони не існують. Цитовані роботи на реальних даних демонструють ефективність цих коефіцієнтів для моніторингу споживання, а отримані результати є певним досягненням у галузі енергетики. Водночас ці роботи не містять вивчення цих обчислювальних конструкцій з точки зору прикладної математики, оскільки це роботи в галузі енергетичної науки. Заповненню цієї прогалини присвячена дана робота. Встановлено деякі властивості коефіцієнтів подібності та автоподібності, пов’язані з використанням кореляційного підходу, протестовано розпізнавальну здатність коефіцієнтів для фіксованих даних, порівняно з коефіцієнтом інтеграції Шеклі (the distance correlation) [3], [4]. Апробація даної технології виконана на даних, отриманих проектом з дослідження електроспоживання для 400 шведських домогосподарств ([5]).*

Ключові слова: *коефіцієнт подібності, коефіцієнт автоподібності, моделювання, електроспоживання домогосподарств, статистичний аналіз, кореляція, smart-grid, класифікація.*