UDC 621.317.38:681.3

https://doi.org/10.33271/nvngu/2021-2/089

M. Zagirnyak[1], orcid.org/0000-0003-4700-0967, A. Perekrest[1], orcid.org/0000-0002-7728-9020, V. Ogar[1], orcid.org/0000-0003-3719-5369, Ye. Chebotarova[1], orcid.org/0000-0003-4725-7914, O. Mur[2]

1 – Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine; email: pksg13@gmail.com
2 – freelancer, Kremenchuk, Ukraine

SEGMENTATION OF HEAT ENERGY CONSUMERS BASED ON DATA ON DAILY POWER CONSUMPTION

Purpose. Improving the quality of the analysis of energy consumption modes of buildings of educational institutions by determining the typical patterns of their consumption by the "*k*-means" method based on statistics of thermal energy consumed, their area, outdoor and indoor air temperature.

Methodology. Methods of statistical and intellectual data analysis, optimization methods, *k*-means method.

Findings. It is shown that most of the researched consumers of thermal energy have similar patterns of its consumption and can be divided into separate segments according to the peculiarities of functioning. The used data set is diverse as it includes three groups of data (conditionally formed by the type of educational institutions). The buildings in groups have different areas, internal temperature to be maintained, operation modes of heating systems and so on. For each of the groups, the number of clusters corresponding to low, medium and high values of heat consumption is determined. It was found that the amount of the consumed heat and consumer behavior depends on the day of the week. The obtained results can be difficult to generalize for all existing types of buildings, but it is expected that communal buildings may have similar consumption patterns, as operating modes and norms of internal microclimate remain the same for different temperature zones and territories.

Originality. The method for dividing the consumers of thermal energy into segments on the basis of the data on daily energy consumption applying the "*k*-means" clustering algorithm, implemented in the programming language R, is further developed. For the first time, it has been proposed to consider not one, but three groups of objects of the educational institutions. It is proposed to determine the modes of energy consumption by buildings analytically by taking into account such parameters as the area of buildings, the amount of heat consumed by them, external and internal air temperature.

Practical value. The results of the research are useful for heat supply companies, municipalities and can be used to develop programs and policies for energy efficiency. The given information and analytical support create a basis for the development of software solutions with the function of integration into local energy monitoring systems of a separate building, energy management system of the district and/or city. The presented results provide an opportunity to predict the cost of energy resources used for heating buildings, which increases the efficiency of engineering systems of buildings.

Keywords: *heat consumption, k-means; segmentation, clustering, data analysis*

Introduction. Today the housing and communal service sector is the largest consumer of energy resources in Ukraine. Under the conditions of economic crisis in the state, the buildings of the budgetary sphere require special attention. This is due to the fact that they have limited funding, and sometimes underfunding, and high requirements for regulatory sanitation in the premises. Improving energy efficiency and reducing financial costs for utilities are achieved through the introduction of technical solutions, such as automated monitoring and management of energy consumption [1].

Apart from reducing overall consumption of energy, the design should help to accumulate and store large amounts of real data on the number of the used energy resources and information on the operation of building engineering systems. The analysis of these data makes it possible to forecast the energy needs of consumers [2], to identify abnormal consumption and improve the efficiency of energy resource control [3, 4]. Besides, the use of these data allows segmentation, i.e. structuring of consumers according to certain features [5]. Division of consumers into segments allows not only heating companies to provide better personalized services but also enables municipalities to develop recommendations to improve energy consumption [6].

Literature review. The problem of segmentation of energy consumers is topical today in the scientific community and is being actively researched.

In paper [7], the segmentation of consumers by two clustering methods for the identification of similar patterns of energy consumption is proposed. The first method is based on the analysis of the profiles of daily consumption, the second one – on the curves of duration of peak level of consumption. The obtained results further provide an opportunity to create demand management models and load change strategies.

The problem of consumer segmentation with a central heating system is solved in [8] using machine learning methods without a teacher. Three new methods of pattern selection on the basis of data on consumption of thermal energy are offered, taking into account characteristics (peculiarities) of its consumption.

Papers [9, 10], based on the hourly data on heat consumption of residential buildings, identify regularities in load profiles, changes in consumer behavior and variability in consumption. The research was performed using cluster (k-means) and regression analyzes. Typical consumption patterns and factors that determine the difference between clusters were obtained.

The results of the study [11] reveal that the improvement of segmentation performed by the method of k-means clustering can be achieved by normalization (purification) of data by using wavelet transforms and autocorrelation.

A separate topic of recent research is the separation of household consumption from total energy consumption by clustering methods [12]. This paper considers the case when household appliances have a similar level of consumption, which leads to the intersection of clusters. The method developed by the authors allows analyzing the unity of clusters of these devices and determining the need for their separation.

In addition to buildings, the occupants of the houses are also selected as research objects because their lifestyle, schedule and preferences have a significant impact on a building's energy consumption. For example, in paper [13] it is proposed to identify and segment patterns of residents' behavior using data on energy consumption by clustering methods. Instead of the standard ASHRAE (The American Society of Heating, Refrigerating and Air-Conditioning Engineers) energy consumption schedule, the authors use a combination of k-means method and neural network. As a result, 10 typical patterns of behavior were obtained. It was also found that the probability of forecasting by this method is higher than by ASHRAE graphs.

In paper [14] the authors used two clustering methods: hierarchical and SOM (Self-organizing map) to identify consumers who create the greatest load on power supply systems. Responsibility coefficient and consumption variance were used as input characteristics of clustering algorithms. According to the authors, the obtained results will improve the recommendations for energy reduction and ensure the development of more dynamic pricing plans depending on the consumer's impact on the peak load.

Equally important is the research on health care buildings. Thus, the authors of paper [15] segmented the buildings of medical institutions by the amount of electricity consumption into clusters that characterize the intensity of resource consumption and determined the reference energy indices for each building. The segmentation was performed by the Ward method (hierarchical clustering) and the k-means method (non-hierarchical clustering).

Despite the large number of studies that use segmentation to solve practical problems in the energy sector, the issues of management of heating systems of buildings of educational institutions are not fully disclosed. In addition, the data are mostly studied within only one group of consumers. Therefore, in contrast to other similar research studies, it is proposed to use as an object the buildings of the educational institutions, which were previously divided by the authors into three groups (formed by the type of educational institution). Based on the analysis of previous research, the k-means method was chosen to select clusters of typical consumption patterns, which was implemented by means of the R programming language. This is because this method is easy to calculate when analyzing large amounts of data.

Purpose. Improving the quality of analysis of energy consumption modes of buildings of educational institutions by singling out the typical patterns of heat consumption and obtaining information, which will allow one to better understand the peculiarities of heat consumption of individual buildings and the relationship between the studied parameters and patterns of consumers' behavior.

Methods. *Description of the research method.* The idea of the k-means method, one of the most popular clustering algorithms, was formed almost simultaneously by two scientists, Hugo Steinhaus and Stuart Lloyd, in the 1950s. [16, 17]. The purpose of the algorithm is to divide observations into k clusters, and each observation belongs to the cluster with the closest average value. The data are similar within one cluster, but differ from other clusters, according to the selected distance metric. The method is based on minimizing the sum of the quadratic deviation of the cluster points from their centers (Lindsten, F., Ohlsson, H., Ljung, L., 2011). The formula of an objective *SSE* function, which is employed in the k-means method, is given below (Reddy, C., Vinzamuri B., 2013)

$$SSE(C)=\sum_{k=1}^{K}\sum_{x_i\in C_k}\left\|x_i-c_k\right\|^2, \tag{1}$$

where $D=\{x_1, x_2, \ldots, x_N\}$ is a dataset, which consists of N points; $C=\{C_1, C_2, \ldots, C_k, \ldots, C_K\}$ is the clustering obtained after applying the k-means method; c_k is the centroid of cluster C_K.

$$c_k=\frac{\sum_{x_i\in C_k} x_i}{\left|C_k\right|}. \tag{2}$$

Euclidean distance is chosen as a distance metric for this research.

The calculations of the 26 popular indices are performed for the k-means method to determine the optimal number of clusters using the programming language R [18]. Clustering is carried out on the data of heat consumption for the heating period.

The description and analysis of the obtained results. Data on heat consumption were obtained from the open data portal of Drohobych City Council, which was implemented with the support of "TAPAS/Transparency and Accountability in Public Administration and Services", funded by the United States Agency for International Development (USAID), the UK Government and the Eurasia Foundation. The research used daily data on heat consumption in the period from January to December 2018 of 46 objects in the city of Drohobych, which can be divided into 3 groups: preschool educational institutions, secondary schools and institutions not included in the previous groups (out-of-school education institutions, libraries, and so on). Apart from the data on heat consumption and building area, specific values of heat consumption were additionally calculated, and the data on weather conditions in this area were obtained from open sources.

At the initial stage of the research, a descriptive analysis of the data based on the constructed box graphs was performed. This made it possible to reveal typical patterns of seasonal consumption related to weather conditions, operating modes, building characteristics, and others. For example, in Fig. 1 a box graph of consumption data of preschool educational institutions of the city of Drohobych is shown.

It can be seen from the graph that most of the energy is consumed in the winter period from October to April, which is logical because there is a sharp drop in outdoor temperature. It should also be noted that some buildings refuse heating on weekends, and some reduce the level of consumption at the end of the calendar months. It may indicate an attempt to artificially reduce the level of consumption to the locally approved consumption norms, without complying with the temperature norms in buildings.

In addition, the box graph (Fig. 1) shows the explicit emissions in the data caused by errors in data records. The presence of such emissions complicates the process of identifying non-

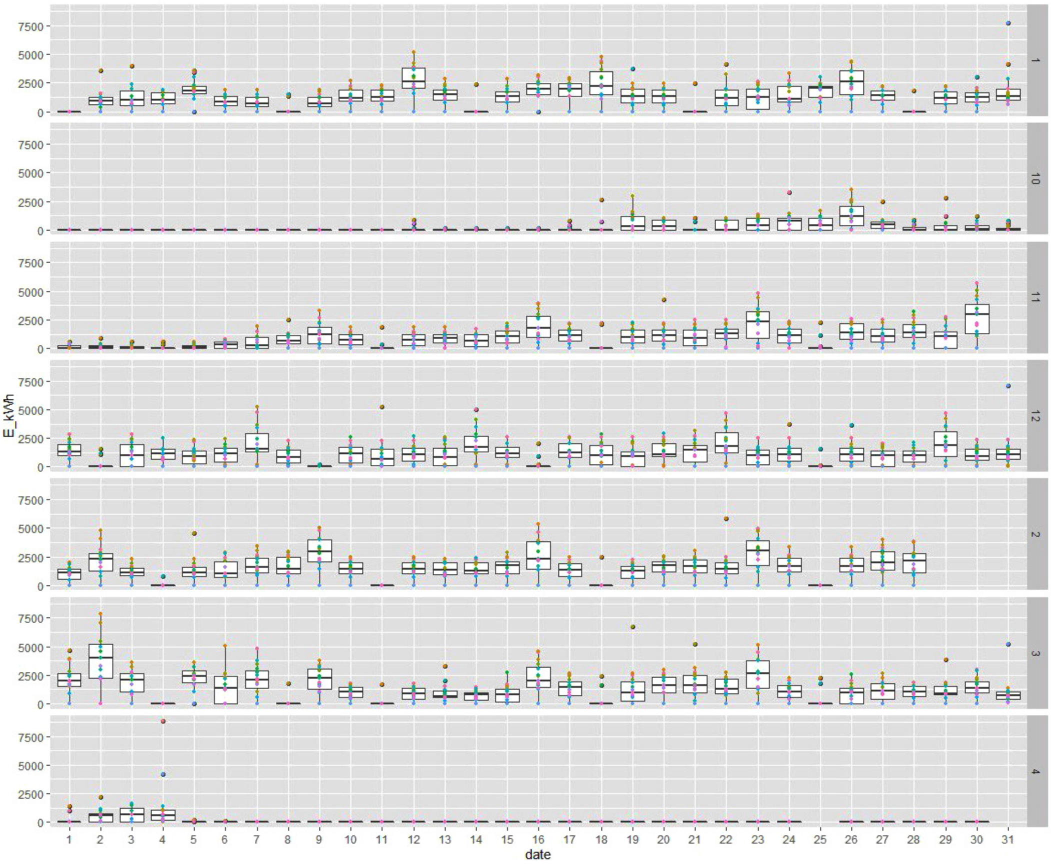

Fig. 1. Heat consumption data of preschool educational institutions of Drohobych

obvious patterns of resource consumption behavior. Therefore, in previous research [19] the data were prepared for clustering, involving the purification of data (emissions, missing values, and errors) by machine learning methods and their formatting for further processing.

The next step is clustering, which is performed separately for each previously selected data group. Each group differs in temperature and operating modes of the heating system. At first, the initial numbers of clusters were determined for each of the data groups according to the above indices. The optimal number of clusters for each index corresponds to the value that meets the conditions [18] for this index. For example, the optimal number of clusters on the index KL is the number that corresponds to the maximum calculated value [18]. According to Table 1, the maximum calculated value by index KL is equal to 4.7366, which corresponds to number 8. Therefore, the number of clusters for this data by index KL is 8. For indices whose number of clusters is found by a graphical method, the optimal number of clusters is chosen by selecting the column that corresponds to the name of the index and contains one. Similarly, the values that meet the conditions of a particular index (highlighted in bold) were selected for the other 26 indices.

The optimal number of clusters for each data group is chosen by the authors as indicated by most indices. For example, the optimal number of clusters for a group of preschool data is six, because this number is indicated by the most indices (CH, Friedman, Cindex, DB, Dunn, Hubert, SDindex, Dindex) (Table 1). The sizes of these six clusters (the number of values in the cluster) are: 134, 107, 437, 344, 485, 175. Fig. 2 shows the data in these six clusters in a 3-D space. The values of the axes correspond to value *E_kWh* of heat consumption (kW·hour), building heated areas S (m^2) and outside air temperature T_{out} (°C).

A similar sequence of determining the number of clusters is made for the data group of secondary schools and further summarized in a simplified Table 2. The following results are obtained. The number of clusters for the given data set is equal to 8 (Table 2).

The clusters sizes are 336, 348, 201, 9, 175, 319, 251, 92. Fig. 3 shows eight clusters of daily heat consumption of these institutions.

Similar results were obtained for the data on out-of-school educational institutions, libraries, and so on. According to Table 3 the number of clusters for this data set is six. The sizes of the clusters are as follows: 1059, 18, 90, 60, 169 and 322. Fig. 4 shows six clusters of daily heat consumption of out-of-school educational institutions, libraries, and others.

The clustering of energy consumption data of preschool educational institutions allowed us to allocate six clusters, i.e. to divide this group into six consumer segments (Fig. 2). Clusters *1* and *5* correspond to the data of buildings with the lowest heat consumption of large and small area, respectively.

Clusters *3* and *6* are the data on the average heat consumption of large and small buildings, respectively. Despite almost identical values of absolute heat consumption (*E_kWh*) of these clusters, the value of specific heat consumption ($q1$) differs significantly. At the same time, based on the analysis of area and temperature values, the following conclusion can be made: cluster *6* corresponds to the heat consumption of small buildings at temperatures below zero; the value of specific heat consumption is the second largest of all clusters. Two groups can be distinguished separately according to the size of the heating area (clusters *5*, *6* and *1*, *3*). Cluster *4* combines the heat consumption of most large buildings at an outside temperature below zero. Cluster *2* is the smallest (107 observations) and is of the greatest research interest, as it has the largest dispersion and the highest level of heat consumption. Each cluster has observations that go beyond the interquantum range, but it is cluster *2* that has the largest number of emissions. So we can conclude that

Table 1

The value of indices for the group of data on heat consumption of preschool educational institutions

Index name \ Number of clusters	4	5	6	7	8	9
KL	2.106	0.889	1.223	1.514	**4.737**	0.235
CH	1698.463	1676.694	**1778.127**	1738.150	1739.355	1627.157
Hartigan	329.359	358.174	320.226	242.708	**102.686**	225.597
CCC	**25.0488**	–2.9533	3.2108	9.0455	13.0830	13.0138
Scott	11171.12	**12371.79**	13202.97	13766.77	14440.36	14872.49
Marriot	2.758263e+38	**2.110773e+38**	1.854351e+38	1.805148e+38	1.579698e+38	1.546338e+38
TrCovW	2.508601e+16	1.959410e+16	1.097680e+16	6.896717e+15	**5.126481e+15**	4.430103e+15
TraceW	551719463	461195657	380029085	319066517	**278684980**	262578013
Friedman	199.173	224.491	**251.730**	266.869	283.768	293.214
Rubin	18.326	21.923	26.605	31.688	36.280	**38.505**
Cindex	0.146	0.133	**0.122**	0.141	0.129	0.127
DB	0.978	0.961	**0.862**	0.872	0.916	0.947
Silhouette	0.401	0.398	0.407	**0.412**	0.384	0.357
Duda	**1.706**	0.956	1.930	2.075	1.759	1.795
Pseudot2	**–344.801**	11.045	–85.283	–227.924	–171.281	–113.801
Beale	**–1.589**	0.178	–1.840	–1.978	–1.646	–1.689
Ratkowsky	**0.286**	0.276	0.280	0.264	0.255	0.245
Ball	137929866	**92239131**	63338181	45580931	34835622	29175335
Ptbiserial	**0.507**	0.480	0.482	0.480	0.442	0.422
Frey	0.787	0.263	0.314	**0.991**	1.943	0.773
McClain	0.959	1.190	1.249	1.287	1.568	1.737
Dunn	0.005	0.0051	**0.0067**	0.0057	0.0046	0.003
Hubert	0	0	**1**	0	0	0
SDindex	0.0043	0.0043	**0.0041**	0.0042	0.0056	0.007
Dindex	0	0	**1**	0	0	0
SDbw	0.581	0.459	0.337	0.292	**0.228**	0.289

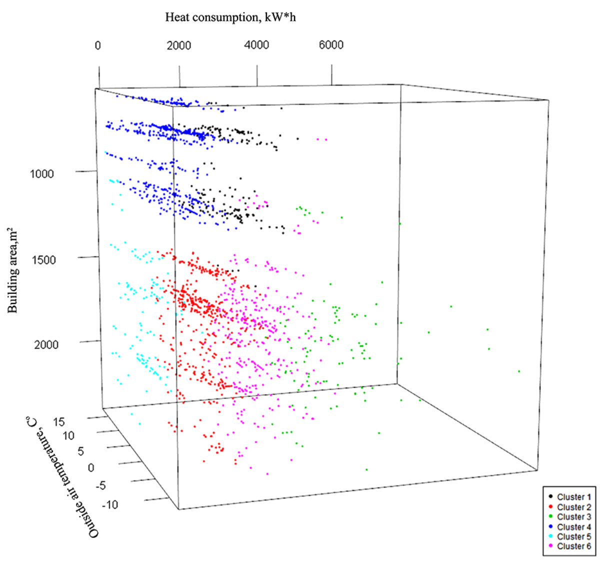

Fig. 2. Clusters of daily heat consumption of preschool educational institutions

this cluster combines all the data on overspending of this group of consumers.

Clustering of energy consumption data of secondary schools allowed us to identify eight clusters (segments) of consumers (Fig. 3). Clusters *2* and *3* correspond to small (up to 4000 m^2) and large (from 4000 m^2) buildings with the lowest consumption. In clusters *6* and *7* we observe almost the same outdoor air temperature (approximately 0.5 °C) and average specific consumption of 0.68 kW · h/m^2 (Table 4), and therefore there is a division of consumers into two groups according to the size of the heated area. Cluster *1* combines data on significant heat consumption of small buildings at outdoor temperatures below zero. Cluster *5* contains data on significant heat consumption at medium temperature of –1.49 °C. In addition to buildings with a heated area exceeding 4000 m^2, this cluster includes small buildings for which this consumption is abnormally high. This can be caused by a significant drop in outdoor temperature or incorrect decisions of the energy manager. Cluster *4* is characterized by a large specific heat consumption, whose value is 3.5 times greater than the next in terms of specific consumption cluster *8*. Cluster *8* is characterized by significant energy consumption of buildings with a heating area exceeding 4000 m^2 at an average temperature of –1.28 °C. This indicates an overuse of energy, because at a higher average air temperature, the value

Table 2

The optimal number of clusters for each index for the data group on heat consumption of secondary schools

Index name	Index value	Optimal number of clusters
KL	5.7183	8
CH	1977.0090	8
Hartigan	120.0589	8
CCC	–3.0540	4
Scott	12159.4930	8
Marriot	3.998990e+40	8
TrCovW	9.629376e+17	5
TraceW	2096555174	8
Friedman	182.2333	8
Rubin	31.6823	8
Cindex	0.0482	8
DB	0.9378	5
Silhouette	0.4412	4
Duda	1.0339	4
Pseudot2	–28.3646	4
Beale	–0.1258	4
Ratkowsky	0.2500	4
Ball	616246727	5
Ptbiserial	0.5481	4
Frey	2.0386	5
McClain	0.7940	4
Dunn	0.0027	4
Hubert	–	8
SDindex	0.0018	4
Dindex	–	8
SDbw	0.3765	8

Table 3

The value of indices for the group of data on heat consumption of out-of-school education institutions, libraries, and others

Index name	Index value	Optimal number of clusters
KL	2.9386	6
CH	3076.97	4
Hartigan	11.006	6
CCC	13.1259	4
Scott	12610.25	5
Marriot	1.059745e+41	5
TrCovW	2.099364e+17	6
TraceW	1353357119	6
Friedman	81.3752	5
Rubin	12.4435	6
Cindex	0.0299	9
DB	0.9292	6
Silhouette	0.5029	5
Duda	4.7268	4
Pseudot2	–875.1675	4
Beale	–3.0306	4
Ratkowsky	0.2806	5
Ball	325468161	5
Ptbiserial	0.5322	4
Frey	2.6209	5
McClain	0.2003	4
Dunn	0.0026	4
Hubert	–	5
SDindex	0.0058	6
Dindex	–	6
SDbw	0.3811	9

of specific heat consumption is greater in relation to the 5^{th} cluster.

Clustering of data on energy consumption of out-of-school education institutions, libraries, and others (1718 values) allowed us to allocate six clusters (segments) of consumers (Fig. 4). Cluster *1* contains the largest number of values of this group (61.64 %), because a significant number of buildings, whose data were selected for the research in this group have a small heating area. Cluster *6* involves consumption of buildings of small area at an outside temperature below zero, characterized by the second largest specific heat consumption (Table 5). Cluster *4* consists of data on the average level

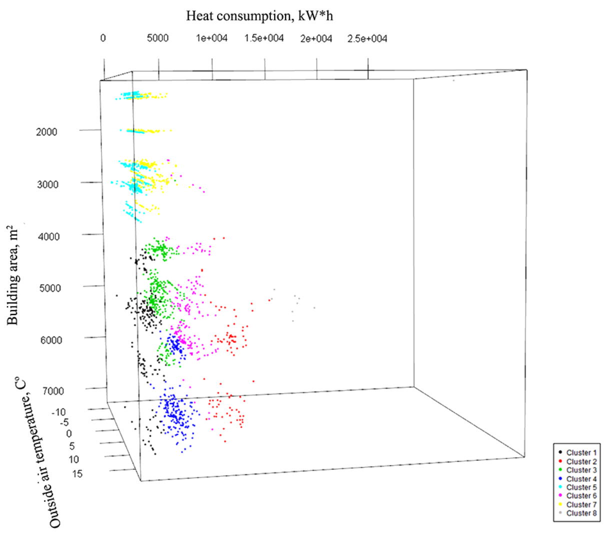

Fig. 3. Clusters of daily heat consumption of secondary schools

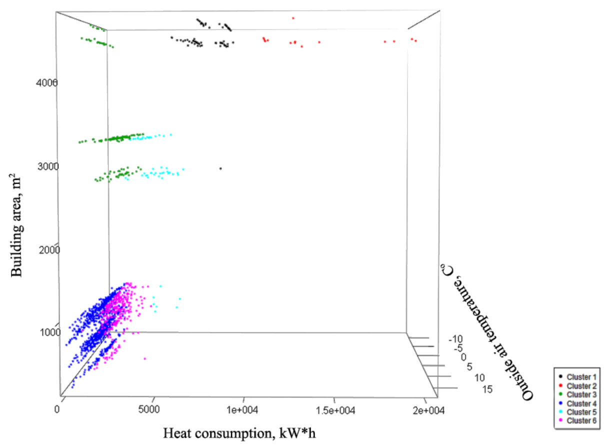

Fig. 4. Clusters of daily heat consumption of out-of-school educational institutions, libraries, and others

Table 4

Average values of clusters of comprehensive schools

Cluster number	S, m^2	T_{out}, °C	E_kWh	q_1, kWh/m^2
1	2405.85	–0.72	2404.68	1.07
2	2505.86	4.64	860.73	0.38
3	5293.86	7.74	1275.69	0.25
4	4752.44	–2.06	16 837.87	5.41
5	5208.36	–1.49	5912.41	1.17
6	4784.10	0.68	3288.73	0.69
7	6677.80	0.52	4449.17	0.67
8	6199.29	-1.28	9357.75	1.54

Table 5

Average values of clusters of out-of-school educational institutions, libraries, and others

Cluster number	S, m^2	T_{out}, °C	E_kWh	q_1, kWh/m^2
1	566.92	1.96	468.22	1.09
2	4602.11	1.44	13935.33	3.03
3	4663.32	1.09	7546.13	1.62
4	2687.22	–0.2	3760.61	1.71
5	3342.69	0.83	1346.07	0.44
6	678.58	–0.84	1344.64	2.67

of heat consumption for buildings with an area exceeding 2000 m^2, in addition, contains data on the high level of heat consumption (more than 2500 kW · h) of small buildings (Table 5). Cluster *3* combines data with an average heat consumption of 7546.1323 kWh of large buildings at an average outdoor temperature of about 1 °C. Cluster *2* is the smallest (1.05 %) and contains data on significant energy over-consumption of large buildings – on average 13935.3294 kWh at an outdoor temperature above zero. It is also characterized by the highest value of specific energy consumption among all clusters. Clusters *3* and *6* are the data on the average heat consumption of buildings of large and small area, respectively (Table. 5).

Conclusions. Clusterization of heat consumption data of three types of municipal buildings of the educational institutions of the city of Drohobych has been performed by the *k*-means method. The clustering results are as follows, preschools – six clusters, secondary comprehensive schools – eight clusters and institutions that are not included into the previous groups (out-of-school education institutions, libraries, and so on) – six clusters. Each cluster characterizes a consumer segment with the same pattern of consumption behavior. The clusters correspond to different levels of thermal energy consumption ranging from low to high.

The division into working days and weekends/holidays has shown that most buildings refuse heating at weekends or minimize it.

It has been found that behavior patterns of a large number of consumers are regular and predictable. It is likely that the building repeats the same pattern of consumption for several days, taking into account changes in outside temperature. Thus, it is possible to predict the amount of heat consumption and to segment the consumer based on clustering. The obtained results enable municipal energy managers to compile current, detailed and specialized reports, which can be further used as a basis for sustainable development plans in accordance with the Covenant of Mayors [20], community economic development programs, as well as for planning relevant budget expenditures. The developed method of division into consumer segments allows energy managers to reasonably set limits on the use of thermal energy.

The presented results demonstrate the efficiency of thermal energy consumption by its end users, which is an additional tool for suppliers to adjust the operation of their thermal networks and develop a strategy to improve the energy efficiency of the entire heat supply complex.

Since most buildings do not have smart meters or weather-dependent regulators, the obtained results will allow end users and/or energy managers to better understand the energy patterns of the building and substantiate decisions to implement measures to improve its energy efficiency.

This research is the basis for creating software to support decisions on energy modernization of buildings taking into account the needs of specific consumers and the diagnosis of abnormal levels of consumption within consumer segments.

In the future, the research will be continued concerning the time data on heat consumption and taking into account a larger number of factors that may affect the amount of the consumed heat.

References.

1. Perekrest, A., Shendryk, V., Pijarski, P., Parfenenko, Y., & Shendryk, S. (2017). Complex information and technical solutions for energy management of municipal energetics. *Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017, 10445,* 1044567. https://doi.org/10.1117/12.2280962.

2. Perekrest, A., Konokh, I., & Kushch-Zhyrko, M. (2019). Administrative Buildings Heating Automatic Control Based on Maximum Efficiency Criterion. *2019 IEEE International Conference on Modern Electrical and Energy Systems (MEES)*, 202-205. https://doi.org/10.1109/mees.2019.8896517.

3. Ramos, S., Duarte, J., Soares, J., Vale, Z., & Duarte, F. (2012). Typical load profiles in the smart grid context: A clustering methods comparison. *2012 IEEE Power and Energy Society General Meeting*. https://doi.org/10.1109/pesgm.2012.6345565.

4. Chebotarova, Y., Perekrest, A., & Ogar, V. (2019). Comparative Analysis of Efficiency Energy Saving Solutions Implemented in the Buildings. *2019 IEEE International Conference on Modern Electrical and Energy Systems (MEES)*. Kremenchuk, Ukraine, 434-437. https://doi.org/10.1109/MEES.2019.8896691.

5. Ma, Z., Xie, J., Li, H., Sun, Q., Si, Z., Zhang, J., & Guo, J. (2017). The Role of Data Analysis in the Development of Intelligent Energy Networks. *IEEE Network*, *31*(5), 88-95. https://doi.org/10.1109/mnet.2017.1600319.

6. Power institute Hrvoje Požar (2020). *Research on the current state of energy management in Ukrainian communities*. Retrieved from http://misto-em.org.ua/wp-content/uploads/2020/07/Zvit-pro-doslidzhennya-MEM-v-gromadah_ukr.pdf.

7. Du, Y., Wang, Ch., Li, H., Song, J., & Li, B. (2019). Clustering Heat Users Based on Consumption Data. *Energy Procedia*, *158*, 3196-3201. https://doi.org/10.1016/j.egypro.2019.01.1010.

8. Wang, C., Du, Y., Li, H., Wallin, F., & Min, G. (2019). New methods for clustering district heating users based on consumption patterns. *Applied Energy*, *251*, 113373. https://doi.org/10.1016/j.apenergy.2019.113373.

9. Do Carmo, C., & Christensen, T. (2016). Cluster analysis of residential heat load profiles and the role of technical and household characteristics. *Energy And Buildings*, *125*, 171-180. https://doi.org/10.1016/j.enbuild.2016.04.079.

10. Gianniou, P., Liu, X., Heller, A., Nielsen, P., & Rode, C. (2018). Clustering-based analysis for residential district heating data. *Energy Conversion and Management*, *165*, 840-850. https://doi.org/10.1016/j.enconman.2018.03.015.

11. Tureczek, A., Nielsen, P., Madsen, H., & Brun, A. (2019). Clustering district heat exchange stations using smart meter consumption data. *Energy and Buildings*, *182*, 144-158. https://doi.org/10.1016/j.enbuild.2018.10.009.

12. Aiad, M., & Lee, P. (2018). Energy disaggregation of overlapping home appliances consumptions using a cluster splitting approach. *Sustainable Cities and Society*, *43*, 487-494. https://doi.org/10.1016/j.scs.2018.08.020.
13. Diao, L., Sun, Y., Chen, Z., & Chen, J. (2017). Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings*, *147*, 47-66. https://doi.org/10.1016/j.enbuild.2017.04.072.
14. Azaza, M., & Wallin, F. (2017). Smart meter data clustering using consumption indicators: responsibility factor and consumption variability. *Energy Procedia*, *142*, 2236-2242. https://doi.org/10.1016/j.egypro.2017.12.624.
15. De la Puente-Gil, A., González-Martínez, A., Borge-Diez, D., Martínez-Cabero, M.-Á., & de Simón-Martín, M. (2019). True power consumption labeling and mapping of the health system of the Castilla y León region in Spain by clustering techniques. *Energy Procedia*, *157*, 1164-1181. https://doi.org/10.1016/j.egypro.2018.11.283.
16. Pérez-Ortega, J., Nely Almanza-Ortega, N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., & Martínez-Rebollar, A. (2019). The K-Means Algorithm Evolution. *Introduction to Data Science and Machine Learning*. https://doi.org/10.5772/intechopen.85447.
17. Koren, O., Hallin, C., Perel, N., & Bendet, D. (2019). Enhancement of the K-Means Algorithm for Mixed Data in Big Data Platforms. *Proceedings of the 2018 Intelligent Systems Conference (IntelliSys), London, UK, 1*, 1025-1040. https://doi.org/10.1007/978-3-030-01054-6_71.
18. Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2019). *NbClust Package. An examination of indices for determining the number of clusters*.
19. Perekrest, A., Chornyi, O., Mur, O., Kuznetsov, V., Kuznetsova, Y., & Nikolenko, A. (2018). Preparation and preliminary analysis of data on energy consumption by municipal buildings. *Eastern-European Journal Of Enterprise Technologies*, *6*(8(96)), 32-42. https://doi.org/10.15587/1729-4061.2018.147485.
20. *Covenant of Mayors*. Retrieved from http://com-east.eu/uk/pro-nas/ugoda-meriv/.

Сегментація споживачів теплової енергії на основі щоденних даних про енерговикористання

М. В. Загірняк[1], *А. Л. Перекрест*[1], *В. О. Огарь*[1], *Є. О. Чеботарьова*[1], *О. В. Мур*[2]

1 – Кременчуцький національний університет імені Михайла Остроградського, м. Кременчук, Україна, email: pksg13@gmail.com
2 – фрілансер, м. Кременчук, Україна

Мета. Підвищення якості аналізу режимів енергоспоживання будівель закладів освіти шляхом визначення методом «*k*-means» типових шаблонів їх споживання на основі статистичних даних використаної теплової енергії, їх площ, зовнішньої та внутрішньої температури повітря.

Методика. Методи статистичного та інтелектуального аналізу даних, методи оптимізації, метод k-means.

Результати. Показано, що більшість досліджуваних споживачів теплової енергії мають схожі шаблони її споживання й можуть бути розділені на окремі сегменти відповідно до особливостей функціонування. Використаний набір даних відрізняється різноманітністю, оскільки включає три групи даних (умовно сформованих за типами закладів освіти). Будівлі у групах мають різні площі, внутрішню температуру, що має підтримуватися, режими роботи систем опалення тощо. Для кожної із груп визначена кількість кластерів, що відповідають низьким, середнім і високим значенням теплоспоживання. З'ясовано, що кількість споживаної теплової енергії й поведінка споживачів залежить від дня тижня. Отримані результати складно узагальнити для всіх існуючих типів будівель, але очікується, що комунальні споруди можуть мати схожі шаблони споживання, оскільки режими роботи та норми внутрішнього мікроклімату залишаються однаковими для різних температурних зон і територій.

Наукова новизна. Набула подальшого розвитку методика розділення на сегменти споживачів теплової енергії на основі даних про щоденне енерговикористання за допомогою алгоритму кластеризації «*k*-means», реалізованого мовою програмування R. Уперше запропоновано розглядати не одну, а три групи об'єктів закладів освіти. Запропоновано визначати режими енерговикористання будівлями аналітично шляхом урахування таких параметрів як площа будівель, кількість спожитого ними теплового ресурсу, зовнішня та внутрішня температура повітря.

Практична значимість. Результати дослідження корисні для підприємств теплопостачання, муніципалітетів і можуть бути використані для розробки програм і політики з енергоефективності. Приведене інформаційно-аналітичне забезпечення створює підґрунтя для розробки програмних рішень із функцією інтеграції в локальні системи енергетичного моніторингу окремої будівлі, системи енергетичного менеджменту району та/або міста. Представлені результати надають можливість прогнозування витрат енергетичних ресурсів, які використовуються для обігріву будівель, що підвищує ефективність функціонування інженерних систем будівель.

Ключові слова: *теплоспоживання, k-means, сегментація, кластеризація, аналіз даних*

Сегментация потребителей тепловой энергии на основе ежедневных данных про энергоиспользование

М. В. Загирняк[1], *А. Л. Перекрест*[1], *В. А. Огарь*[1], *Е. О. Чеботарёва*[1], *А. В. Мур*[2]

1 – Кременчугский национальный университет имени Михаила Остроградского, г. Кременчуг, Украина, e-mail: pksg13@gmail.com
2 – фрилансер, г. Кременчуг, Украина

Цель. Повышение качества анализа режимов энергопотребления зданий учебных заведений путем определения методом «*k*-means» типовых шаблонов их потребления на основе статистических данных использованной тепловой энергии, их площадей, внешней и внутренней температуры воздуха.

Методика. Методы статистического и интеллектуального анализа данных, методы оптимизации, метод *k*-means.

Результаты. Показано, что большинство исследуемых потребителей тепловой энергии имеют схожие шаблоны ее потребления и могут быть разделены на отдельные сегменты в соответствии с особенностями функционирования. Использованный набор данных отличается разнообразием, поскольку включает три группы данных (условно сформированных по типам учебных заведений). Здания в группах имеют разные площади, внутреннюю температуру, которая должно поддерживаться, режимы работы систем отопления и так далее. Для каждой из групп определено количество кластеров, соответствующее низким, средним и высоким значениям теплопотребления. Выяснено, что количество потребляемой тепловой энергии и поведение по-

требителей зависит от дня недели. Полученные результаты трудно обобщить для всех существующих типов зданий, но ожидается, что коммунальные сооружения могут иметь схожие шаблоны потребления, поскольку режимы работы и нормы внутреннего микроклимата остаются одинаковыми для разных температурных зон и территорий.

Научная новизна. Получила дальнейшее развитие методика разделения на сегменты потребителей тепловой энергии на основе данных о ежедневном энергопотреблении с использованием алгоритма кластеризации «*k*-means», реализованного на языке программирования R. Впервые предложено рассматривать не одну, а три группы объектов учебных заведений. Предложено определять режимы энергопотребления зданиями аналитически путем учета таких параметров как площадь зданий, количество потребленного ими теплового ресурса, внешняя и внутренняя температура воздуха.

Практическая значимость. Результаты исследования полезны для предприятий теплоснабжения, муниципалитетов и могут быть использованы для разработки программ и политики по энергоэффективности. Приведенное информационно-аналитическое обеспечение создает основу для разработки программных решений с функцией интеграции в локальные системы энергетического мониторинга отдельного здания, системы энергетического менеджмента района и/или города. Представленые результаты дают возможность прогнозирования затрат энергетических ресурсов, которые используются для обогрева зданий различного назначения. Это позволяет повысить эффективность функционирования инженерных систем зданий.

Ключевые слова: *теплопотребление, k-means, сегментация, кластеризация. анализ данных*

Recommended for publication by O. P. Chornyi, Doctor of Technical Sciences. The manuscript was submitted 23.11.20.