International Journal for Computational Civil and Structural Engineering, 21(3) 93–104 (2025)

DOI:10.22337/2587-9618-2025-21-3-93-104

COMPREHENSIVE INTELLECTUAL AND STATISTICAL ANALYSIS OF WATER CONSUMPTION

Elena V. Ignatova, Elena Kh. Kitaytseva

National Research Moscow State University of Civil Engineering, Moscow, RUSSIA

Abstract: This article deals with the water consumption regime in a residential building. The study is based on data of cold and hot water hourly consumption in a multi-storey apartment building. The measurement period is one month. The study comprehensively uses statistical analysis of water consumption and data mining of group outliers. Statistical data analysis is designed to determine the distribution pattern of different data samplings. The analysis is carried out for three different samplings of apartment water consumption data. As a result, group outliers of hourly water consumption are identified. Machine learning methods are used to identify group outliers. The task boils down to clustering the hours of the day to find hours with reduced (nighttime) water consumption. Clustering is carried out using five methods, and clustering quality is assessed by three metrics. As a result, nighttime consumption periods are determined for different samplings of water consumption data in apartment buildings. In general, comprehensive intellectual and statistical analysis of water consumption is useful for solving the tasks of designing water supply and sanitation systems, adjusting the operating modes of engineering equipment, and clarifying the calculated parameters of water consumption in apartment buildings.

Keywords: statistical analysis, data mining, machine learning, clustering, hourly water consumption, apartment building

КОМПЛЕКСНЫЙ ИНТЕЛЛЕКТУАЛЬНО-СТАТИСТИЧЕСКИЙ АНАЛИЗ ВОДОПОТРЕБЛЕНИЯ

Е.В. Игнатова, Е.Х. Китайцева

Национальный исследовательский Московский государственный строительный университет, г. Москва, РОССИЯ

Аннотация: В статье исследуется режим водопотребления в многоквартирном доме. Исследование основано на данных о почасовом потреблении холодной и горячей воды в многоэтажном жилом здании. Период измерения – один месяц. Исследование в комплексе использует статистический анализ водопотребления и интеллектуальный анализ групповых выбросов данных. Статистический анализ данных предназначен для определения характера распределения различных выборок данных. Анализ проводится для трех различных выборок данных водопотребления квартирами. В результате выявляются групповые выбросы почасового потребления воды. Для выделения групповых выбросов используются методы машинного обучения. Задача сводится к кластеризации часов суток для поиска часов с пониженным (ночным) потреблением воды. Кластеризация проводится пятью методами, качество кластеризации оценивается тремя метриками. В результате определяются периоды ночного потребления для разных наборов данных водопотребления в доме. В целом комплексный интеллектуально-статистический анализ водопотребления полезен для решения задач проектирования систем водоснабжения и водоотведения, настройки режимов эксплуатации инженерного оборудования, уточнения расчетных параметров водопотребления в многоквартирных домах.

Ключевые слова: статистический анализ, интеллектуальный анализ данных, машинное обучение, кластеризация, почасовое водопотребление, многоквартирный дом

INTRODUCTION

The object of the study is the consumption of cold and hot water in an apartment building. The subject of the study is the mode of water consumption depending on the time of day and day of the week. The purpose of the study is to obtain reliable characteristics of water consumption modes in a house based on the integrated use of various data analysis methods.

The problem of determining the regime and volume of water consumption is relevant for all countries. Digitalization programs are being implemented in Russia, which should turn apartment buildings into high-tech and energy-efficient digital facilities. A mandatory component of such facilities is a smart resource consumption telemetry system.

It is possible to take meter readings daily (hourly, every second), store digital information in data centers, quickly compile reports and transmit data to interested authorities and relevant information systems. It should be noted that telemetry data is collected for statistics and is mainly used in financial settlement services for automatic accrual of payments for used resources [1].

An analysis of the scientific literature in the field of water consumption research shows that water consumption monitoring data can be used to solve other important tasks, including to refine engineering calculations, adjust regulatory indicators, and make recommendations on resource conservation [2].

It should be noted that when designing water supply systems, overestimated design loads lead to an increase in pipeline diameters, an increase in heat loss of hot water pipes, and an overestimation of the required characteristics of pumping equipment. In addition, overestimated values of estimated water consumption led to an unjustified increase in the diameters of drainage systems.

Despite the common goals facing the researchers, the study of the process relies on diverse information. For example, article [3] analyzes the second flow rate passing through different devices installed in the same apartment, article [4] analyzes the hourly flow rate in an 80-apartment residential building. The article [5] examines hourly expenses in an urban-type settlement, the article [6] deals with the daily water consumption of a city where about one hundred thousand inhabitants live in apartment buildings. In all these articles, the object of research is water consumption in residential apartment buildings. It is noted in [7] that there is practically no data on water consumption in public buildings.

A commonly used method of data analysis is visualization of water consumption. Graphs help to understand the process on a qualitative level, but they do not provide accurate quantitative characteristics.

In the article [5], based on graphs, the maximum water consumption in an urban-type settlement is analyzed depending on the day of the week. The nighttime water consumption period is allocated for different days of the week. As a result, the schedules of the actual daily consumption do not coincide with the schedules of the standard daily uneven water consumption.

A similar study is conducted in the article [8]. The telemetry data of water consumption in a multi-storey residential building is used. Nighttime hours of reduced consumption of hot and cold water are defined for weekdays and weekends. The graphs show a significant difference between night and day water consumption.

The main characteristics of water consumption in a house (average, median, minimum, maximum, and standard deviation) are determined using descriptive statistics.

In the article [9], based on a statistical analysis of data on water consumption in an apartment building, it is concluded that the actual average daily water consumption per inhabitant is less than recommended by standards for calculating internal water supply systems. A similar conclusion is found in the article [4].

Several papers attempt to describe empirical samplings using a theoretical distribution law. The intensity of water consumption using individual devices is described by a Gamma distribution [3]. The intensity of water consumption by a building is estimated by a normal distribution [4], the number of simultaneously operating appliances in the building is described by a binomial distribution [7].

The least squares method was used to determine the time of intensive consumption of hot water during the day in apartment buildings [10]. The total time of intensive consumption of hot water

per day is defined as 18-21 hours. However, the specific hours are not specified. From this study, it can be concluded that the period of reduced water consumption lasts 3-5 hours per day.

The tasks of identifying trends, seasonality, or cyclicity in water consumption can be solved based on time series analysis [11, 12]. Artificial intelligence methods are increasingly being used to solve classification, clustering, optimization, and forecasting problems [13]. Machine learning methods make it possible to recover missing data [14].

If we consider the volume of nighttime water consumption to be significantly different from the volume of water consumption at other hours of the day, then they can be considered as a group data outlier. Searching for group outliers involves identifying abnormal patterns or behaviors within datasets. Unlike individual outliers, group outliers exhibit anomalies common to a subset of observations within a larger population. Statistical methods or machine learning algorithms can be used to detect group outliers [15]. Data clustering methods can be considered suitable intelligent machine learning algorithms.

In the article [16], the DBSCAN machine learning method was used to identify uneven water consumption in an apartment building, designed to cluster data considering their distribution density. Unfortunately, the quality of clustering was not discussed. A variety of density-based clustering methods are described in [17]. It is important to note that the choice of clustering quality assessment metrics is of great importance. Metrics make it possible to identify the most suitable methods for the available data set [18].

Research [19] shows the complex use of various analysis methods to study water consumption patterns and search for anomalies based on payment documents of water consumers with mechanical meters. The DBSCAN clustering method and time series analysis were used together. The integrated use of various data analysis methods helps to solve tasks more effectively.

METHODS

The numerical experiment is used in the research. Water consumption data is processed and analyzed based on the integrated use of statistical analysis and data mining. The units of measurement of water consumption are liters.

At the first step of the study the data is cleaned.

At the second step of the study statistical analysis is carried out on various samplings to determine patterns of water use by apartments.

At the third step of the study the task of clustering the hours of the day is solved to determine the periods of reduced (nighttime) consumption of cold and hot water in the house. Machine learning methods are used.

Data preprocessing

The first stage of data processing is its cleaning. Each data set requires an individual approach, but the general rule of data validation remains the search for impossible or unlikely events.

Equal hourly water consumption by different apartments at different times is a regular event. The same daily water consumption by different apartments is a possible event. The same hourly water consumption during the day for the entire observation period is an unlikely event. A composite key was used to search for duplicate information for the entire observation period. This key includes:

– total water consumption for the entire observation period;
– number of hours with no consumption;
– maximum flow rate;
– average consumption;
– standard deviation;
– the number of cases with an hourly consumption of more than 100 l/h.

All the above parameters are determined for the entire observation period separately for the consumption of cold and hot water. Several apartments are excluded from the analysis as duplicate data. Apartments with zero water consumption for the entire measurement period are excluded from the analysis.

Data preprocessing is performed in an Excel environment.

Statistical data analysis

Preliminary data analysis includes the calculation of the main sampling parameters: minimum (min), maximum (max), mean (m), standard deviation (s), quartiles (Q_1, Q_2, Q_3), coefficient of skewness (A) and kurtosis (E), mode (Mo).

The consumption of cold and hot water during the entire observation period is considered as samplings:

– hourly consumption by apartments (sampling 1);

– average hourly consumption for all apartments (sampling 2);

– average daily hourly consumption of apartments (sampling 3).

The frequency distribution of cold and hot water consumption for each sampling is visualized.

Statistical data analysis is performed in the Excel environment.

Clusterization

To identify the hours of reduced (night) water consumption, the task of clustering the hours of the day is solved. Clustering is understood as an algorithm that divides data into clusters, that is, groups of elements with similar characteristics. Similarity is determined by calculating the conditional "distance" between the elements. Machine learning methods for clustering are tailored to the characteristics of the source data. The use of clustering methods is often associated with setting the required number of clusters or other necessary algorithm parameters. To identify the hours of reduced (nighttime) water consumption, 5 clustering methods are used: Affinity Propagation, DBSCAN, Agglomerative Clustering, K-means, Spectral Clustering.

At first the Affinity Propagation method is used, because it does not require specifying the number of clusters. Then the DBSCAN method is used, which also does not require specifying the number of clusters. It requires specifying the minimum number of elements in clusters (Nb_{min}) and the conditional distance between cluster elements (ε). Three options for parameter values are considered for this method. Then Agglomerative Clustering method, K-means method, and the Spectral Clustering method are used consistently. These three methods require setting the number of clusters. Several numerical experiments were conducted with different numbers of clusters. The initial approximation is the number of clusters determined by the Affinity Propagation method.

Three metrics were selected to assess the quality of clustering [11].

The Davies-Bouldin (*DB*) index is based on the distance from cluster objects to their centroids, and separability is based on the distance between the centroids:

$$DB = \frac{1}{K}\sum_{k=1}^{K} max\left\{\frac{S(c_k) + S(c_l)}{\|\overline{c_k} - \overline{c_l}\|}\right\} \quad (1)$$

Where:

$$S(c_k) = \frac{1}{n_k}\sum_{x_i \in c_k} \|x_i - \overline{c_k}\|$$

K – number of clusters,

n_k – size of cluster,

$\overline{c_k}$ - centroid of cluster *k.*

As the clustering result improves, the *DB* index should decrease.

The Calinsky-Harabasz (*CH*) index is based on the distance from cluster points to their centroids, and separability is based on the distance from cluster centroids to the global centroid.

$$CH = \frac{N-K}{K-1} \cdot \frac{\sum_{k=1}^{K} n_k \left\|\overline{c_k} - \overline{C}\right\|}{\sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \overline{c_k}\|} \quad (2)$$

Where:

K – number of clusters,

n_k – size of cluster k,

$\overline{c_k}$ - centroid of cluster *k*,
$\overline{C}$ - global centroid of sampling,
N – sample size.
The CH index should increase.
Silhouette (*Sil*) coefficient defined as the average of silhouettes across all elements.
The silhouette of an element shows the difference between the average distance with elements inside its cluster and the average distance with elements of another cluster:

$$Sil = \frac{1}{N}\sum_{k=1}^{K}\sum_{x_i \in c_k}\frac{b(x_i, c_k) - a(x_i, c_k)}{max\{a(x_i, c_k), b(x_i, c_k)\}} \quad (3)$$

Where:

$$a(x_i, c_k) = \frac{1}{n_k}\sum_{x_j \in c_k}\|x_i - x_j\|$$

- average distance from $x_i \in c_k$ to other objects in the c_k cluster,

$$b(x_i, c_k) = min_{k \neq j}\left\{\frac{1}{n_k}\sum_{x_j \in c_j}\|x_i - x_j\|\right\}$$

- distance from $x_i \in c_k$ to objects from another cluster c_j,
n_k – size of cluster *k*,
N – sample size.
The *Sil* index should aim at 1.
Clusterization is performed using Python programming language libraries.

RESULTS AND DISCUSSION

Data preprocessing

Using a composite key allows you to detect duplicates in the measurement results of both cold and hot water. The results are presented in table 1.
The first three data blocks contain information that is completely repetitive throughout the entire observation period. The first two blocks include 2 apartments each, the third block includes 3 apartments. The fourth block contains one apartment, where the information about the consumption of cold water completely coincides with the information about the consumption of hot water. The fifth block includes apartment numbers in which no consumption of cold and/or hot water was recorded during the entire observation period.

Table 1. Apartment numbers with duplicated information on cold and hot water consumption

Block	Cold water			Hot water		
	apartment	duplicate		apartment	duplicate	
1	1	2	-	4	5	-
2	4	5	-	26	28	-
3	50	51	53	50	51	53
4	62	-	-	-	62	-
5	7	-	-	7	38	61

Further analysis will provide information on the consumption of cold and hot water in these 9 apartments. (№ 2, 5, 51, 53, 62, 7, 38, 61) it is not counted. As a result, the sampling consists of the hourly water consumption for 73 apartments for 32 days.
The analysis of daily water consumption shows that out of 2336 (73*32) cases, in 26 cases (1.1%) there is no consumption of cold and hot water, in 105 cases (4.5%) there is no consumption of hot water, in 28 cases (1.2%) cold water. Hourly readings for those apartments where there is no consumption of cold and/or hot water during the day are removed from the initial sampling. The volume (n) of the final sampling of hourly cold and hot water costs is 52248.

Statistical data analysis

The results of the statistical analysis are presented in Table 2. It should be noted that the values are similar or practically identical to the average values of m for all 3 samplings for cold and hot water, while the standard deviations of s for samplings 2 and 3 are several times smaller than for sampling 1.
Averaging during the organization of samplings 2 and 3 makes it possible to offset the error in determining hourly costs, the values of which are less than 10 liters. Unlike Sampling 1, the quartiles Q_1 and Q_2 of samplings 2 and 3 have values other than 0.

The values of the skewness indicators A and kurtosis E for sampling 2 are less than the corresponding values for samplings 1 and 3. The values of A and E for this sampling indicate that the distribution tends to a normal distribution.

Table 2. Statistical analysis

Param.	Cold Water (CW)			Hot Water (HW)		
	Samplings					
	CW1	CW1	CW1	HW1	HW2	HW3
n	52248	768	2177	52248	768	2177
Min, l/h	0	0,1	0,4	0	0	0,4
Max, l/h	420	21,1	91,3	820	33,2	79,6
m, l/h	8,2	7,6	8,2	11,5	10,7	11,5
s, l/h	17,9	3,9	7,7	34,2	6,5	10,5
Q_1, l/h	0	5	3	0	6	4
Q_2, l/h	0	8	6	0	10	9
Q_3, l/h	10	10	11	10	15	16
Mo, l/h	0	7	1	0	9	3
A	5,3	0,3	3,0	6,7	0,4	1,9
E	55,7	-0,2	18,2	69,4	-0,1	5,5
m+ 3s, l/h	62	19	31	114	30	43
Out-liers %	893	2	35	1040	4	37
	1,7	0,3	1,6	2	0,5	1,7
Q_1+ 3(Q_3-Q_1), l/h	30	20,5	27,9	30	32,2	40
Out-liers %	5290	1	53	6357	1	52
	10,1	0,1	2,4	12,2	0,1	2,4

Regardless of the method of estimating anomalous values, the number of outliers in samplings 2 and 3 is lower both absolutely and relative to the size of sampling 1.

Figures 1-3 show the frequency distribution of hourly consumption of cold (CW) and hot (HW) water by apartments for 3 samplings.

The distribution pattern is the same for samplings 1 and 3 (figures 1 and 3). The distribution of hourly water consumption averaged across apartments (sampling 2) differs (figure 2).

Visualization of the average hourly consumption of all apartments (figure 2) allows us to identify two consumption modes, regardless of whether the consumption of cold or hot water is considered.

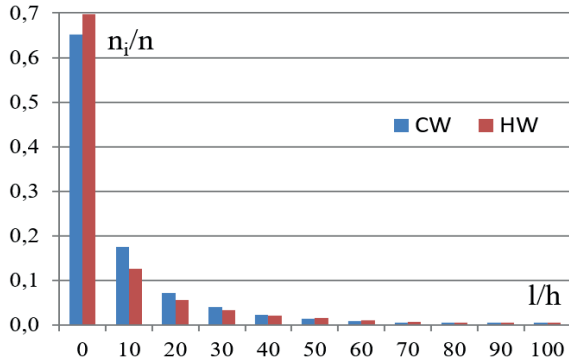


Figure 1. Frequency distribution of cold (CW) and hot (HW) water consumption - sampling 1

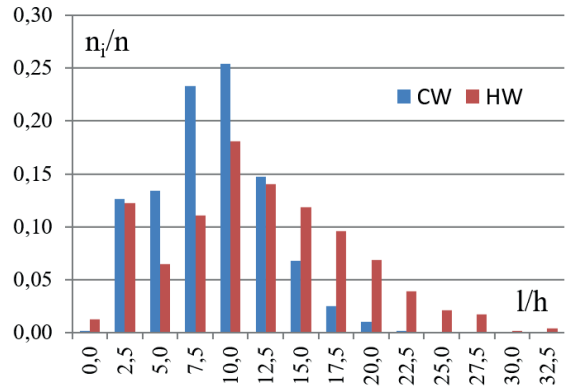


Figure 2. Frequency distribution of cold (CW) and hot (HW) water consumption - sampling 2

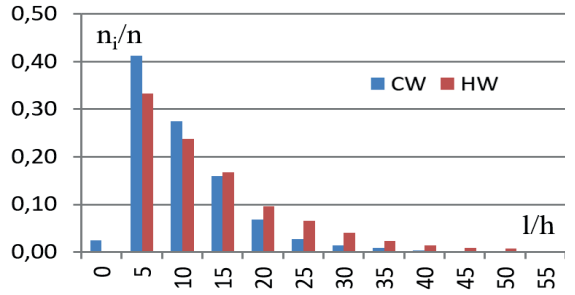


Figure 3. Frequency distribution of cold (CW) and hot (HW) water consumption - sampling 3

The first mode corresponds to the mode occurring at a flow rate of 2.5 l/h, the second

mode corresponds to the mode occurring at a flow rate of 10 l/h. The points forming the first mode are a group outlier for the sampling. Identifying a group outlier is equivalent to dividing a sampling into two subsets. Solving this problem is laborious, and clusterization methods must be used to automate this process.

Clusterization

Clustering is performed for the hours of total water consumption by the apartments in the house. The data is converted into datasets in which hourly water consumption is assigned to one of the 24 hours of the day (from 0 to 23). In addition, subsets of data are generated separately for weekends and weekdays of the measurement period.

The results of using various clustering methods, considering experiments with the number of clusters, are evaluated for quality. As a result, the number of clusters with the best clustering quality is determined for each method. Table 3 shows five clustering methods and three measures of the quality of their results.

Table 3. Results of quality assessment of clusterization of hot water consumption

Methods	Para-meter	Metrics (index)		
		DB	CH	Sil
Affinity Propagation	-	0,93	14,00	0,33
DBSCAN	ε =2000	1,12	17,15	0,33
	ε =2100	1,09	17,20	0,35
	ε =2200	1,17	11,25	0,30
Agglomerative Clustering	K =3	1,15	16,95	0,33
	K =4	0,97	13,78	0,32
	K =5	0,90	11,99	0,31
K-means	K =3	1,09	17,20	0,35
	K =4	1,10	14,10	0,33
	K =5	1,37	11,82	0,19
Spectral Clustering	K =3	1,16	16,91	0,32
	K =4	1,50	11,01	0,19
	K =5	1,33	12,01	0,21

The Affinity Propagation method allocates four clusters. Based on the metrics for the DBSCAN method, the parameters ε=2100 and Nb_{min}=2.

Four clusters are defined for the Agglomerative Clustering method, three for the K-means method, and three for the Spectral Clustering method.

As an example, Figures 4-8 show the results of clustering the hours of hot water consumption in a house over the entire measurement period. The hours of the day are shown on the horizontal axis, and the volume of hot water consumption is shown on the vertical axis. For clarity, there is one dot corresponding to each hour. The ordinate of the point is calculated as the median of a set of values for hourly hot water consumption over the entire observation period. The point color is selected automatically depending on the cluster number. Different methods give different numbering and number of clusters.

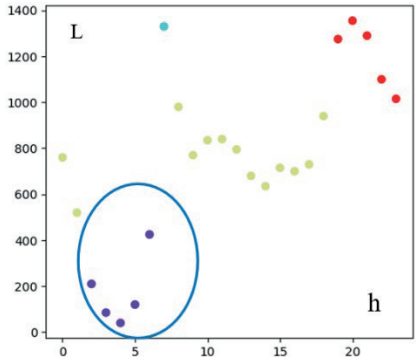


Figure 4. Affinity Propagation method result

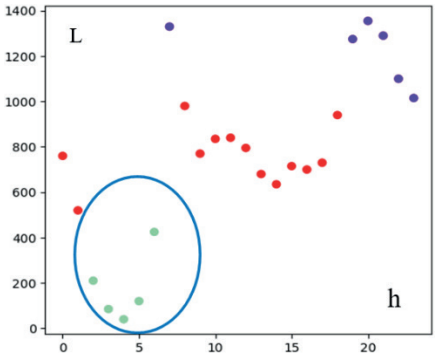


Figure 5. DBSCAN method result

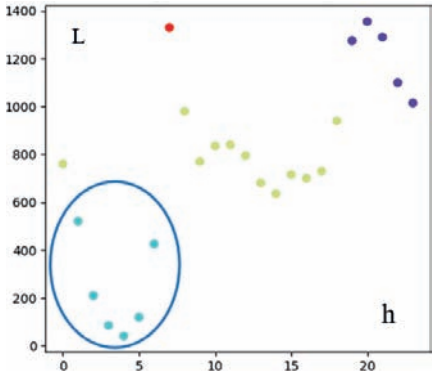


Figure 6. Agglomerative Clustering method result

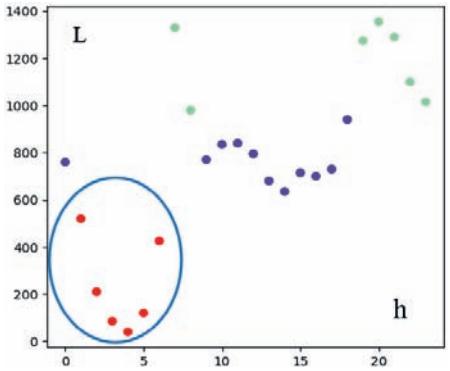


Figure 7. K-means method result

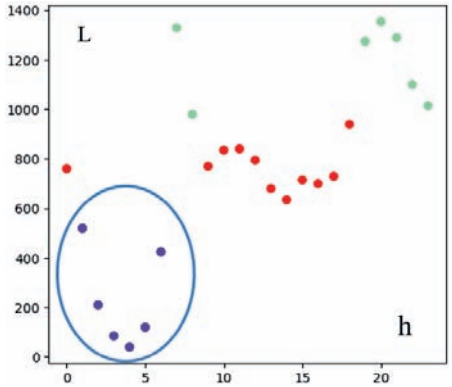


Figure 8. Spectral Clustering method result

The results of clustering using different methods are presented in Table 4. The method numbers match their enumeration above.

The table shows that the results obtained by different methods for different samplings are in most cases close or identical. The final conclusion about the period of nighttime (reduced) water consumption is based on the results obtained by most methods.

Table 4. Hours of reduced (overnight) consumption of cold and hot water: 1-Affinity Propagation, 2-DBSCAN, 3-Agglomerative Clustering, 4-K-means, 5-Spectral Clustering

Water	Days of week	Methods					Result , Hours
		1	2	3	4	5	
Cold Water	All days	1-6	1-6	2-6	1-6	1-6	1–6
	Weekdays	1-6	1-6	1-6	1-6	1-6	1–6
	Weekends	1-6	2-6	2-6	1-6	1-6	1–6
Hot Water	All days	2-7	2-7	1-7	1-7	1-7	1–7
	Weekdays	2-6	2-6	2-6	2-6	1-7	2–6
	Weekends	1-9	1-9	2-8	1-9	1-9	1–9

CONCLUSION

1. The scale of water consumption facilities, methods, and conditions for obtaining source data, as well as the structure and format of data storage, are important for analyzing water consumption.
2. To solve various problems, it is usually necessary to transform data values (normalization, averaging) and change the data structure.
3. Improving data quality involves removing duplicates, anomalies, deviations from the norm, and zeros. However, in data related to hourly water consumption, it is necessary to carefully study anomalies and outlines, including analyzing them as group outlines.
4. Data samplings for analyzing water consumption in an apartment building can be formed for various objects (the building, apartments, appliances, tenants) and for different time periods (the entire observation period, days of the week, hours of the day). Samplings for statistical analysis can be formed based on averaged data values.
5. Different data samplings may have different distribution laws. The choice of further data analysis methods depends on the identified distribution function.
6. Intelligent machine learning methods for clustering differ in cluster formation algorithms, control parameters, and the quality of the result.

During the research, it is necessary to use a set of clustering methods and evaluate their quality.
7. Most clustering methods have yielded consistent results. The Agglomerative Clustering method has shown special results in many cases and has not been considered.
8. The results of the study can be used in the design of water supply and sanitation systems, setting operating modes of engineering equipment, clarifying the regulatory parameters of water consumption in apartment buildings.

REFERENCES

1. **Ignatova E.V., Kruglova L.V.** Information systems for accounting and data analysis of housing and communal services of apartment buildings. J. Construction and Architecture Vol. 4, 2023, p. 38. DOI:10.29039/2308-0191-2023-11-4-38-37
2. **Kitaytseva E.H., Ignatova E.V.** Trends in the use of water supply systems telemetry data. J. Construction and Architecture Vol. 4, 2023, p. 22. DOI:10.29039/2308-0191-2023-11-4-22-22
3. **Polivanov D.E., Semenov A.A., Yarkova O.N.** Mathematical modeling of the intensity of water consumption by various types of water collecting devices. J. Information and mathematical technologies in science and management. 2024 Vol. 1 (33)
4. **Andreenko A.A., Sharipov T.R.** Analysis of hourly water consumption in a residential building. Modern problems of water supply and sanitation : collection of materials of the Interuniversity scientific and practical conference [December 1-3, 2021]. St. Petersburg : SPbGASU, 2022, p.3-11
5. **Ignatchik V.S., Sarkisov S.V., Obvintsev V.A.** Research of water consumption hour inequality coefficients. J. Water and ecology Vol.2 (70), 2017. p. 27-39. DOI:10.23968/2305-3488.2017.20.2.27–39
6. **Kermany E., Mazzawi H., Baras D., Naveh Y., Michaelis H.** Analysis of advanced meter infrastructure data of water consumption in apartment Buildings. 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013, p.1159-1167. DOI:10.1145/2487575.2488193
7. **Josey B.M., Buchberger S.G., Jinzhe G.J.** Comparing Actual and Designed Water Demand in Australian Multilevel Residential Buildings. Water Resour. Plann. Manage., 2023, 149(1): 05022013. DOI:10.1061/(asce)wr.1943-5452.0001625
8. **Kitaytseva E.H.** Numerical analysis of hourly consumption of hot and cold water. International Journal for Computational Civil and Structural Engineering. Vol. 8, Issue 4, 2012 - M. : DIA Publishing House, p. 78-84. Library ID: 18973119 EDN: https://www.elibrary.ru/pzexll
9. **Zhulin A.G., Aminova A.Kh., Belova L.V.** Determination of the amount of water consumed by various water users of the residential sector. Arkhitektura, stroitel'stvo, transport [Architecture, construction, transport], 2021 (1), p. 47-57. (in Russia)
10. **Nejranowski J., Szaflik W.** Hot water consumption time in multi-apartment buildings. Journal of Ecological Engineering. 2020, 21(4), p.199–202. DOI:10.12911/22998993/119906
11. **Surendra P., Deka H., Rajakumara N.** Application of Mamdani model-based fuzzy inference system in water consumption estimation using time series. J.Soft Computing - A Fusion of Foundations, Methodologies and Applications. 2022, 5. DOI:10.1007/s00500-022-06966-4
12. **Lee S.S., Lee H.H., Lee Y.J.** Prediction of Minimum Night Flow for Enhancing Leakage Detection Capabilities in Water Distribution Networks. Appl. Sci. 2022, 12, 6467. DOI:10.3390/app12136467
13. **Kim J., Lee H., Lee M., Han H., Kim D., Kim H.S.** Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level. Water 2022, 14, 1512. DOI:10.3390/w14091512

14. **Heydari Z., Cominola A., Stillwell A.S**. Is smart water meter temporal resolution a limiting factor to residential water end-use classification? A quantitative experimental analysis. Environmental Research: Infrastructure and Sustainability, 2022, Vol.2, Num.4, 045004. DOI:10.1088/2634-4505/ac8a6b
15. **Vedishcheva E.V., Kapyrin A.S., Vasilenko M.S.** Analiz i utochnenie klassifikatsii anomalij i vybrosov na economicheskih dannyh [Analysis and refinement of classification of anomalies and outliers on economic data] // Bulletin of the Altai Academy of Economics and Law. 2019, Vol. 6-1. P. 41-46 (in Russia) URL: https://vaael.ru/ru/article/view?id=589
16. **Ignatova E.V.** Water supply telemetry data processing in apartment buildings. BIO Web of Conferences Volume 107 01004 (2024) (YRC-2024). DOI:10.1051/bioconf/202410701004
17. **Bhattacharjee P., Mitra P.** A survey of density-based clustering algorithms, Frontiers of Computer Science. (2021) 15, 151308, DOI:10.1007/s11704-019-9059-3.
18. **Sivogolovko E.V.** Metody otsenki kachestva chotkoj klasterizatsii [Methods for assessing the quality of clear clusterization] // Computer tools in education, Tver', 2011, Issue 4(96), p. 14-31. (in Russia)
19. **Ghamkhar H., Ghazizadeh J.M., Mohajeri S.H., Moslehi I., Yousefi-Khoshqalb E.** An unsupervised method to exploit low-resolution water meter data for detecting end-users with abnormal consumption: Employing the DBSCAN and time series complexity. Sustainable Cities and Society, Volume 94, 2023, 104516 DOI:10.1016/j.scs.2023.104516

СПИСОК ЛИТЕРАТУРЫ

1. **Игнатова Е.В., Круглова Л.В.** Информационные системы учета и анализа данных жилищно-коммунального хозяйства многоквартирных домов/ Строительство и архитектура № 4 2023. стр. 38 DOI:10.29039/2308-0191-2023-11-4-38-37
2. **Китайцева Е.Х., Игнатова Е.В.** Тренды использования данных телеметрии систем водоснабжения. Строительство и архитектура № 4, 2023. стр. 22. DOI:10.29039/2308-0191-2023-11-4-22-22
3. **Поливанов Д.Е., Семенов А.А., Яркова О.Н.** Математическое моделирование интенсивности водопотребления различными типами водоразборных устройств / Информационные и математические технологии в науке и управлении. 2024. № 1(33). С. 78-92. DOI:10.25729/esi.2024.33.1.007
4. **Андреенко А.А., Шарипов Т.Р.** Анализ почасового водопотребления в жилом доме. Современные проблемы водоснабжения и водоотведения : сборник материалов межвузовской научно-практической конференции [1–3 декабря 2021 года]– Санкт-Петербург : СПбГАСУ, 2022, С.3-11
5. **Игнатчик В.С., Саркисов С.В., Обвинцев В.А.** Исследование коэффициентов часовой неравномерности водопотребления / Вода и экология: проблемы и решения. 2017. №2 (70). С. 27-39 DOI:10.23968/2305-3488.2017.20.2.27–39
6. **Kermany E., Mazzawi H., Baras D., Naveh Y. and Michaelis H.** Analysis of advanced meter infrastructure data of water consumption in apartment Buildings 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013, P.1159-1167 DOI:10.1145/2487575.2488193
7. **Josey B.M., Buchberger S.G. and Jinzhe G.J.** Comparing Actual and Designed Water Demand in Australian Multilevel Residential Buildings. Water Resour. Plann. Manage., 2023, 149(1): 05022013 DOI:10.1061/(asce)wr.1943-5452.0001625

8. **Китайцева Е.Х.** Численный анализ часового потребления горячей и холодной воды. International Journal for Computational Civil and Structural Engineering. Volume 8, Issue 4, 2012. M. : Издательство ACB, С. 78-84. elibrary ID: 18973119 EDN: https://www.elibrary.ru/pzexll
9. **Жулин А.Г., Аминова А.Х., Белова Л.В.** Определение количества расходуемой воды различными водопотребителями жилого сектора./ Архитектура, строительство, транспорт. № 1 (95). 2021. С. 47–57 elibrary ID: 46409679 EDN: https://www.elibrary.ru/njfcvo
10. **Nejranowski J., Szaflik W.** Hot water consumption time in multi-apartment buildings. Journal of Ecological Engineering. (2020) 21(4), pp.199–202. DOI:10.12911/22998993/119906
11. **Surendra P., Deka H., Rajakumara N.** Application of Mamdani model-based fuzzy inference system in water consumption estimation using time series. J.Soft Computing - A Fusion of Foundations, Methodologies and Applications. 2022, 5. DOI:10.1007/s00500-022-06966-4
12. **Lee S.S., Lee H.H., Lee Y.J.** Prediction of Minimum Night Flow for Enhancing Leakage Detection Capabilities in Water Distribution Networks. Appl. Sci. 2022, 12, 6467. DOI:10.3390/app12136467
13. **Kim J., Lee H., Lee M., Han H., Kim D., Kim H.S.** Development of a Deep Learning-Based Prediction Model for Water Consumption at the Household Level. Water 2022, 14, 1512. DOI:10.3390/w14091512
14. **Heydari Z., Cominola A., Stillwell A.S**. Is smart water meter temporal resolution a limiting factor to residential water end-use classification? A quantitative experimental analysis. Environmental Research: Infrastructure and Sustainability, 2022, Vol.2, Num.4, 045004. DOI:10.1088/2634-4505/ac8a6b
15. **Видищева Е.В., Копырин А.С., Василенко М.С.** Анализ и уточнение классификации аномалий и выбросов на экономических данных // Вестник Алтайской академии экономики и права. 2019. № 6-1. С.41-46 URL: https://vaael.ru/ru/article/view?id=589
16. **Ignatova E.V.** Water supply telemetry data processing in apartment buildings. BIO Web of Conferences Volume 107 01004 (2024) (YRC-2024). DOI:10.1051/bioconf/202410701004
17. **Bhattacharjee P., Mitra P.** A survey of density-based clustering algorithms, Frontiers of Computer Science. (2021) 15, 151308, DOI:10.1007/s11704-019-9059-3.
18. **Сивоголовко Е.В.** Методы оценки качества четкой кластеризации. // Компьютерные инструменты в образовании. Тверь, 2011, Вып. 4(96), С. 14-31
19. **Ghamkhar H., Ghazizadeh J.M., Mohajeri S.H., Moslehi I., Yousefi-Khoshqalb E.** An unsupervised method to exploit low-resolution water meter data for detecting end-users with abnormal consumption: Employing the DBSCAN and time series complexity. Sustainable Cities and Society, Volume 94, 2023, 104516. DOI:10.1016/j.scs.2023.104516

Elena V. Ignatova — Candidate of Technical Sciences, Associate Professor, Department of Information Systems, Technologies and Automation in Construction, National Research Moscow State University of Civil Engineering (NRU MGSU), 26 Yaroslavskoe Shosse, Moscow, 129337, Russia, Email: IgnatovaEV@mgsu.ru

Елена Валентиновна Игнатова — кандидат технических наук, доцент кафедры Информационных систем, технологий и автоматизации в строительстве, Федеральное государственное бюджетное образовательное учреждение «Национальный исследовательский Московский государственный строительный университет» (НИУ МГСУ), 26, Ярославское шоссе, Москва, 129337, Россия, E-mail: IgnatovaEV@mgsu.ru

Elena Kh. Kitaytseva — Candidate of Technical Sciences, Associate Professor, Department of Information Systems, Technologies and Automation in Construction, National Research Moscow State University of Civil Engineering (NRU MGSU), 26 Yaroslavskoe Shosse, Moscow, 129337, Russia, E-mail: KitaytsevaEH@mgsu.ru

Елена Халиловна Китайцева — кандидат технических наук, доцент, доцент кафедры Информационных систем, технологий и автоматизации в строительстве, Федеральное государственное бюджетное образовательное учреждение «Национальный исследовательский Московский государственный строительный университет» (НИУ МГСУ), 26, Ярославское шоссе, Москва, 129337, Россия, E-mail: KitaytsevaEH@mgsu.ru