A STATISTICAL APPROACH FOR SELECTING BUILDINGS FOR EXPERIMENTAL MEASUREMENT OF HVAC NEEDS

Paweł MALINOWSKI[1], Piotr ZIEMBICKI[2]
[1] Technical University of Wroclaw, Wrocław, Poland
[2] University of Zielona Gora, Zielona Góra, Poland

Abstract

This article presents a statistical methodology for selecting representative buildings for experimentally evaluating the performance of HVAC systems, especially in terms of energy consumption. The proposed approach is based on the k-means method. The algorithm for this method is conceptually simple, allowing it to be easily implemented. The method can be applied to large quantities of data with unknown distributions. The method was tested using numerical experiments to determine the hourly, daily, and yearly heat values and the domestic hot water demands of residential buildings in Poland. Due to its simplicity, the proposed approach is very promising for use in engineering applications and is applicable to testing the performance of many HVAC systems.

Keywords: k-means method, HVAC system performance, energy consumption, building characteristics

1. INTRODUCTION

In the construction industry, there is a rich variety of different types of objects and structures. Thus, there has been a demand a long standing demand for

---

[1] Corresponding author: Technical University of Wroclaw, Institute of Air-Conditioning and District Heating, Norwida st. 4/6, 50-373 Wroclaw, Poland

[2] Corresponding author: University of Zielona Gora, Institute of Environmental Engineering, Z. Szafrana st. 15, 65-516 Zielona Gora, Poland, e-mail: p.ziembicki@iis.uz.zgora.pl, tel.+48683282570

building classification systems that can be used for various research purposes [1-3]. These studies have suggested various methods for analysis and various tools for classifying buildings with respect to their use, maintenance and construction.

For researching the use and maintenance of buildings, as well as HVAC systems, there is a need for appropriate methodology for selecting representative buildings. Heating and air-conditioning are important factors for research on energy consumption. Therefore, choosing representative buildings is particularly important for generalization of the results of experimental energy consumption research.

As far as mathematical methods are concerned, principal component analysis (PCA) represents a promising approach [4]. Climatic classification and energy consumption have also been analyzed using the particle swarm algorithm (PSA), according to references [5] and [6]. Using a combination of PSA and PSC (PSA/PSC), Cohen [7] achieved better performance than was observed using the k-means method. However, the PSA/PSC algorithm is quite complicated to use. Therefore, we developed a method using the k-means algorithm, which is simple in application and operates with sufficient accuracy.

There are many empirical formulas and statistical indicators that can be used to design, operate, and maintain the equipment used in building services engineering, such as equipment involved in heating, ventilation, air conditioning, plumbing, and electricity. There are many methods and calculation indicators derived from empirical research carried out in real buildings that are related to the design and operation of internal systems. However, reference books typically lack information on the methodology for selecting objects for research. As a result, the methodology for research and calculation indicators is often inaccessible. It is impossible to analyze and refer to suggested default values, making it necessary to carry out individual measurements to optimize the design and working conditions of a particular system. We present a statistical approach for selecting buildings for use in such experiments. This method was tested using numerical experiments to determine the values for hourly, daily, and yearly heat use and domestic hot water demands of residential building in Poland.

The proposed approach consists of two phases. In the first phase (described in the second chapter), the buildings are divided into an optimal number of subgroups using the k-means procedure to organize and arrange buildings sharing similar characteristics. Object location is not considered in this division. In the second phase, a few representative buildings are randomly selected from the subgroups. The number of buildings chosen from each subgroup depends on the numerical value of the subgroup.

We present the results of numerical tests conducted for three datasets (chapter 3). The largest set of real data described 419 residential buildings from five cities: Bolesławiec (1), Legnica (2), Stargard Szczeciński (3), Szczecin (4), and Wrocław (5). Finally, we summarize the conclusions drawn from these numerical tests (chapter 4).

2. STATISTICAL APPROACH TO SELECTING BUILDINGS

Data clustering is a statistical approach that divides a set of points into smaller subsets that share similar characteristics. There are many specific data clustering algorithms that can be used to appropriately classify and divide elements into smaller groups. This article briefly discusses three of the most commonly used algorithms for data clustering.

The hierarchical method [8] requires determination of the order in a dataset, which is established based on similarities between elements. A dendrogram that describes the correlation between particular sets is created using a distance with defined metrics. There are two variants of the hierarchical method. Some variants divide sets into smaller sets. Other variants build (agglomerate) larger sets from smaller sets. The greatest disadvantage of the hierarchical algorithm is the complexity of its calculation, which is proportional to the cubed sample size. Thus, the hierarchical method is inappropriate for grouping buildings due to the need to analyze large amounts of data.

The next most frequently used method is the expectation-maximization (EM) method [9], which is used to identify the most credible estimators in statistical models. This method operates properly when the probability distributions from which the samples originate are known. However, this method can be ineffective at grouping buildings if the researcher does not have adequate information about the element distribution groups, as was the case in this study.

In this study, we turned to the k-means method [10-12]. This algorithm is conceptually simple and easy to implement. The method can also be applied to large quantities of data with unknown distributions, as described in detail in the next section.

2.1. Characteristic-based grouping of buildings using the k-means method

This analysis considers a set of p objects (buildings) $\{o_1, o_2, \ldots, o_p\}$, with n characteristics that describe each object $\{c_1, c_2, \ldots, c_n\}$. Therefore, the ith building

is described by a vector of variables

$$o_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \tag{2.1}$$

where for an optional $1 \leq j \leq n$, the variable x_{ij} determines the value of the characteristic describing object o_i. Therefore, the set of all values for n characteristics and p objects forms the following matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \ldots & x_{pn} \end{bmatrix} \tag{2.2}$$

The k-means method is used to divide a dataset into k subsets, so that elements that ultimately belong to the same subset share similar characteristics. Thus, each group is differentiated from the other groups in the greatest possible way. The similarities and differences between two optional objects with characteristics o_s and o_t are defined through the metric $d(o_s, o_t)$.

Formally, the k-means method is based on the specific division of dataset X into k disjointed, non-void $X_1, X_2, \ldots, X_k$, subsets. This division minimizes the following objective:

$$w(X) = \sum_{l=1}^{k} \sum_{o_m \in X_l} d(o_m, \mu_l) \tag{2.3}$$

where μ_l, the centroid of subset X_l, is defined as

$$\mu_l = \frac{1}{|X_l|} \sum_{o_m \in X_l} o_m = \frac{1}{|X_l|} \sum_{o_m \in X_l} (x_{m1}, x_{m2}, \ldots, x_{mn}) \tag{2.4}$$

and $|X_l|$ is the lth number of this subset.

The algorithm for the division of set X into k subsets can be performed according to the following procedure:

1. k objects are randomly chosen from set $\{o_1, o_2, \ldots, o_p\}$. It is assumed that these objects are the centroids $\{\mu_1, \mu_2, \ldots, \mu_p\}$ of the $X_1, X_2, \ldots, X_k$ subsets.
2. The distances from each centroid $\{d(o_m, \mu_1), d(o_m, \mu_2), \ldots, d(o_m, \mu_k)\}$ are determined for each object o_m. Based on the calculated distances, object o_m is subsequently assigned to subset X_l, for which the distance $d(o_m, \mu_l)$ is the smallest of all possible distances. In other words, $d(o_m, \mu_l) = min\{d(o_m, \mu_1), d(o_m, \mu_2), \ldots, d(o_m, \mu_k)\}$, where $1 \leq m \leq p$ and $1 \leq l \leq k$.

3. For subsets $X_1, X_2, ..., X_k$ (calculated in step 2), new centroids $\{\mu_1, \mu_2, ..., \mu_k\}$ are counted according to formula (2.4).
4. Steps 2 and 3 are repeated until there is convergence on a value (i.e., the assumed iteration number of the calculated centroids or change in centroid distances in the additional steps is smaller than assumed level of the importance algorithm).

The k-means method is a relatively easy to use algorithm that enables the separation of subsets with similar characteristics. However, using this method, it is impossible to determine the optimal number of subsets k_{opt} into which the set of objects is divided. Instead, this parameter must be determined a priori. In the next section, we discuss issue associated with selecting the optimal number of sets.

2.2.   Determining the optimal number of groups

To determine the optimal number of sets into which the data were divided, a method based on distance measurements was proposed. There are numerous other methods for selecting optimal numbers of sets (e.g., [13-15]); however, they are often conceptually complicated and difficult to implement. When grouping data, the groups must be as small as possible and the distances between groups must be as large as possible. In this manner, a number of cohesive groups are given such that the elements for each group are close to each other (i.e., they are similar). The groups are then separated from each other, which enables the diversification of data between different groups. The proposed method for the optimal selection of a number of groups is based on the abovementioned characteristics.

The proposed method for selecting the optimal number of groups is described below. The metric d for space R^p is introduced, where p is the number of characteristics (except for the characteristic "location") and d is typically the natural Euclidean metric. The number of groups is represented by k and μ_i is the point representing the centroid of the ith group $1 \leq i \leq k$.

The function $f(k)$ is defined as

$$f(k) := \frac{\min\{d(\mu_i, \mu_j) : 1 \leq i, j \leq k, i \neq j\}}{\max\{d(x_i, y_i) : 1 \leq i \leq k\}} \qquad (2.5)$$
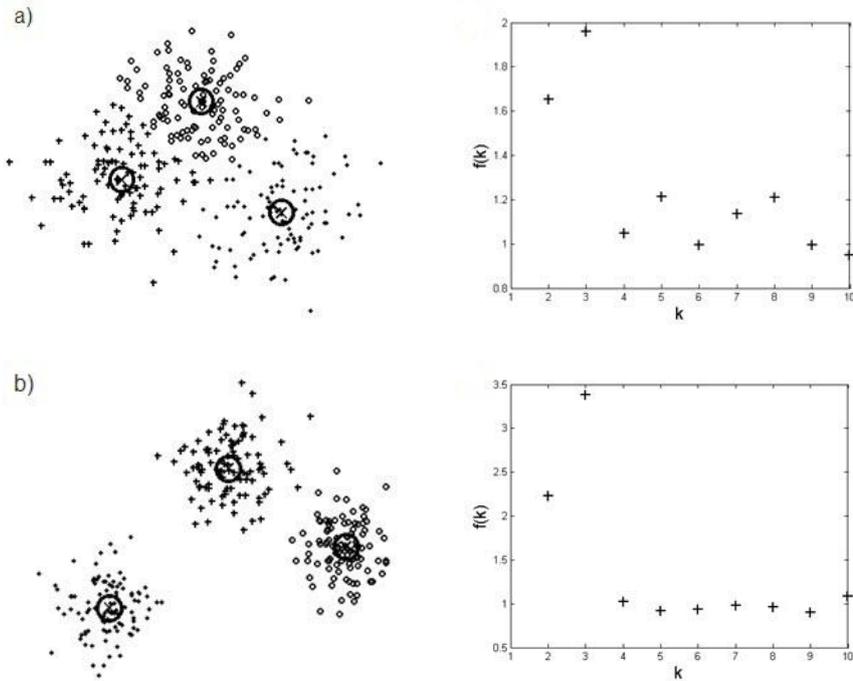
Fig. 1. An example of three datasets and their corresponding functions f

The smallest distance between the centroids of the data groups (i.e., the numerator of function f) must be maximized to distance the data groups from each other. The average distance between the points of a given group (the denominator f) must be minimized to make the subsets coherent. Therefore, to determine the optimal number of subsets k_{opt}, one should identify the value of k that maximizes the function f according to

$$k_{opt} = \arg \max f(k) \qquad (2.6)$$

Figure 1 presents a graph of the function f for three two-dimensional subsets generated as examples. The figure presents two scenarios, a and b, dependent on the mutual distance between the groups and shows that in each example, the function f is maximized in the correct $(k = 3)$ place. The crossed circles are the centroids calculated using the k-means method.

Another problem connected to the choice of an optimal number of groups is the fact that the proposed algorithm requires the value of the function f to be checked for every possible k smaller than the total number of analyzed objects. This process is numerically complex and time consuming. Therefore, the effectiveness of the proposed method depends on the size of the object set under

analysis. A small number of analyzed objects facilitates the choice of an optimal number of groups but may lead to trivial results (e.g., all objects in one group). A large number of analyzed objects require time-consuming calculations. It seems reasonable that datasets of up to 1,000 objects are appropriate when considering calculation times.

2.3.   Selecting representative objects from a group

After creating a k_{opt} group of buildings, the next step in the process is the choosing the representative buildings from each group that will later be measured and analyzed with specialized equipment. There are many studies dedicated to sampling problems (e.g., [16-18]). Our method for selecting a suitable sample is based on multi-stage sampling, which is a natural method that involves the sharing of data multiple times before drawing subsets of the samples. In the case presented here, the entire population of data is divided into k_{opt} groups using the k-means method. Each group is again divided into subgroups based on one feature (e.g., city name). Only the subgroup generated by double application of the k-means procedure is used to select the representative objects.

Suppose that N_{rep} is the sample size that should be taken from a population of size N. It is assumed that the entire population is divided into k_{opt} groups. The number of elements in the ith group is n^i, where $1 \leq i \leq k_{opt}$. An indicator w_i is assigned to each group according to

$$w_i = \frac{n_i}{N} \qquad (2.7)$$

Subsequently, the number of representative samples that will be chosen from the ith group is assigned according to the formula

$$N_i = \lfloor w_i N_{rep} \rfloor \qquad (2.8)$$

where N_i is the number of elements that will be randomly selected from the ith group. Provided that after rounding the number of representative samples does not add to N_{rep}, the excess is corrected at the expense of the largest group.

The last step is the choice of a sample. In group i, we randomly (with equal probability of choosing an element from i^{th} group) chose N_i elements. Random sample selection was used because it theoretically ensures a high diversity of objects in a particular group, meaning that the sample will statistically contain the most information. By repeating the procedure for every I, one is given an N_{rep}-element representative samples originating from different groups, which enables the choice of objects that vary globally and locally.

3. NUMERICAL RESULTS FOR CHOOSING REPRESENTATIVE OBJECTS FOR MEASUREMENT

3.1. Description of the dataset used in the numerical experiments

The proposed methods were applied to select a group of representative buildings from among 419 buildings in five different cities.

The following object characteristics were chosen for data analysis:

1. Commissioning of building (year of construction).
2. Width of building.
3. Length of building.
4. Type of building (downtown/suburbs/housing estates).
5. Thermomodernization (yes/no).
6. Thermomodernization (year of realization).
7. Thermomodernization (thickness of insulation in centimeters).
8. Orientation of world directions (north-south).
9. Number of floors.
10. Number of stairways.
11. Number of apartments.
12. Number of tenants.
13. Usable area.
14. Size of the apartments.
15. Cubature.
16. Heated area.
17. Height of the floor.
18. Building system (e.g., traditional, block of flats).
19. Roof type (sloping/flat).
20. Roof insulation (insulated/uninsulated).
21. Heating source (local/remote).
22. Type of primary energy.
23. Type of exchanger (low-temperature gas-fired boiler/condensing boiler/plate exchanger/water heat exchanger).
24. Type of heating source regulation.
25. Power commissioned for heating needs (unknown/value in kilowatts).
26. Modernization of heating system (yes/no).
27. Modernization of heating system.
28. Individual measuring system for heating needs (yes/no).
29. Modernization of domestic hot water system (yes/no).
30. Modernization of domestic hot water system.
31. Individual domestic hot water heater (none/electric/gas).
32. Energy consumer group *($Q_{el} < 40\ kW/Q_{el} \geq 40\ kW$)*.
33. Type of internal installation system.

34. Electric cookers in the building.

The following characteristics were rejected based on a lack of data, a lack of importance of the characteristic in the building groupings, a similarity of the characteristics for all objects, or a deficiency of data on the characteristics for all objects:

1. City (insignificant characteristic for the groupings).
2. Building owner (characteristic was equal for all objects: "housing association").
3. Street (insignificant characteristic for the groupings).
4. Number (insignificant characteristic for the groupings).
5. Latitude (insignificant characteristic for the groupings).
6. Longitude (insignificant characteristic for the groupings).
7. Size of the apartments (incomplete data).
8. Thermomodernization of the roof (the thickness of insulation) (incomplete data).
9. Power commissioned for heating needs (unknown/value in kilowatts) (incomplete data).
10. Power commissioned for domestic hot water (unknown/value in kilowatts) (incomplete data).
11. Measured yearly thermal energy consumption for heating needs in 2010 (kilowatts/hour) (incomplete data).
12. Measured yearly thermal energy consumption for domestic hot water in 2010 (kilowatts/hour) (incomplete data).
13. Capacity of hot water tanks (incomplete data).
14. Installed heat power (incomplete data).
15. Type of consumer (equal characteristic for all objects: residential building).
16. Type of measuring and billing system (equal characteristic for all objects).

It should be noted that the complete data for the inquiry consisted of 427 objects, but based on the lack of characteristic data, the analysis did not consider eight objects situated in Wrocław (420-427).

3.2. Results of numerical experiments – object division into subgroups and representative choices

The proposed methods were applied to select a set of representative objects from among 419 buildings in five cities. The set of 419 objects appeared to be appropriate for applying the k-means method. Analysis of an optimal number of groups for the set of objects aided us in the statistical analysis used to choose the

buildings for the measurement. A graph was prepared using the Monte Carlo method with $M = 50$ repetitions.

As illustrated in Fig. 2, the function f reaches its maximum at $k_{opt}=5$; therefore, the data were separated into five groups. It should be noted that all of the buildings characteristics except for geographical position were to separate the building into groups.
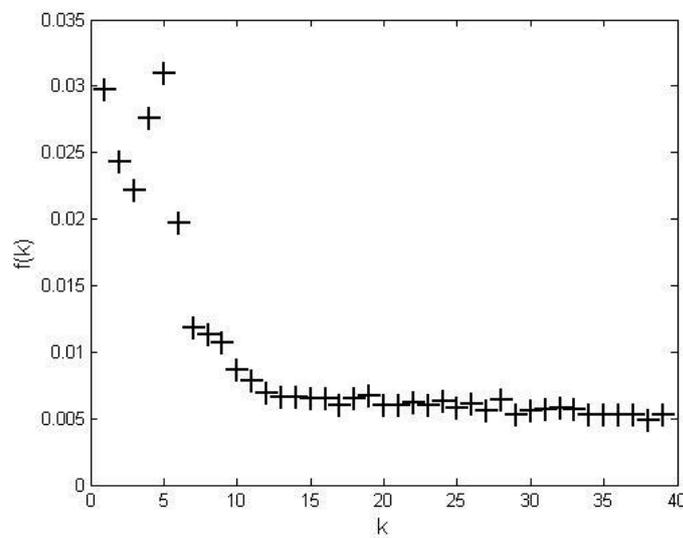
Fig. 2. Function f showing an optimal number of groups

Table 1 shows the separation of the data into the five groups. Table 2 provides a short description of each group.

Table 1. Data group divisions

Group	Object's address reference numbers	Quantity
1	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 20, 21, 22, 23, 25, 27, 28, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147,148, 149, 150, 151, 152, 153, 154, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 173, 174, 175, 176, 181, 187, 188, 189, 190, 191, 192, 208, 213, 215, 216, 217, 218, 219, 220, 221, 222, 229, 230, 233, 234, 236, 268, 269, 270, 271, 272, 273, 274, 275, 301, 302, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419	130
2	3, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 65, 66, 67,	61

	68, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114	
3	57, 58, 59, 60, 61, 62, 63, 64, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86	26
4	303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399	97
5	17, 18, 19, 24, 26, 155, 156, 157, 170, 171, 172, 177, 178, 179, 180, 182, 183, 184, 185, 186, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 209, 210, 211, 212, 214, 223, 224, 225, 226, 227, 228, 231, 232, 235, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300	105

Table 2. Characteristics of the building groups

Group	Description
1	The buildings were commissioned mainly in the 1980s (exception: objects from city 3, which were commissioned earlier). The objects were built using large-panel systems with floor heights of 2.5 m (mainly housing estates). Almost all buildings were thermomodernized (both buildings and roofs) in the last dozen or so years (approximately 2000). Most building were five-stories (exception: several 11- or 12-story objects). The roof insulation layer was approximately 9 cm thick. Most buildings had an external source of heat and plate exchangers (exception: local source of heat and condensing boiler). The heating installations in these buildings had not been modernized, but the domestic hot water installations had been modernized over the past 20 years. Approximately two-thirds of the objects were in the energy consumer group >= 40kW (the rest were < 40kW).
2	The buildings were commissioned in the 1980s and built in the traditional building system. The floor heights were 2.8 m. The buildings were located mainly in housing estates (a small portion was situated downtown). They were thermomodernized and their roofs were isolated after 2000. Solid fuel was the source of primary energy and the buildings

	had plate exchangers. The heating system and domestic hot water system had been modernized in 2005 for most of the buildings. The buildings fell in the energy consumer group < 40kW.They had single-phase internal installation systems.
3	The buildings were commissioned before 1970 and were all built in the traditional building system (floor heights of 2.8 m). Half of the buildings were thermomodernized after 2000. The buildings were low-rise with an average of three stories. Most buildings had sloping roofs. They had local sources of heat and low-temperature gas-fired boilers. Their heating installation systems had been modernized in 2002. Their domestic hot water systems had not been modernized. The buildings fell in the energy consumer group < 40kW. They had single-phase installation systems.
4	The buildings were commissioned in the late 1970s and early 1980s. The buildings were built using large-panel systems and floor heights of 2.5 m (housing estates). This group consisted mainly of five-story buildings (exception: a few 11-story objects). Half of the buildings were thermomodernized and their roofs were isolated after 2004. The thickness of the thermomodernized layer was 5 cm. All had external sources of heat and plate exchangers. The heating installation systems had been modernized in 2002-2004 (depending on the building). The domestic hot water installation system had been modernized in some buildings in 2005. Most of the buildings had individual domestic hot water heaters. The buildings were in the energy consumer group < 40kW. They had single-phase installation systems.
5	The buildings were commissioned in the late 1980s and early 1990s. The buildings were built using large-panel systems and floor heights of 2.5 m (mainly housing estates). The buildings were slightly smaller than those in groups 1, 2, and 4. The average number of staircases was two. They were not thermomodernized, but half had thermomodernized roofs. They had external sources of heat and plate exchangers. The heating installation system had been modernized. The buildings with thermomodernized domestic hot water systems were three times smaller than those in groups 1 and 2 on average. The domestic hot water systems had been thermomodernized after 2007. The buildings were in the energy consumer group > 40kW. They had three-phase systems.

Table 3. Representative object choices for different numbers of samples

	Group number				
	1	2	3	4	5
$N_{rep} = 15$	8, 28, 135, 141, 411	65, 95	80	309, 312, 378	195, 196, 265, 283
$N_{rep} = 10$	8, 141, 411	65	80	312,378	195,265,283
$N_{rep} = 5$	141, 411	65	-	312	265

Figures 3-7 depict building situations and their division by city (the figures show each city separately because to clearly display the localization data). The groups are marked. Representatives were chosen from $N_{rep} = 15$. The list of objects in each group is shown in Table 3.
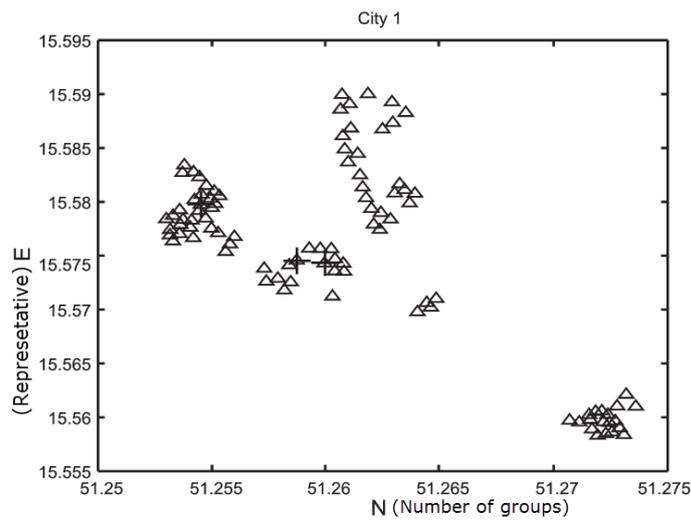
Fig. 3. Objects divided into groups and representatives. The triangles represent the fourth group's cross-chosen representatives
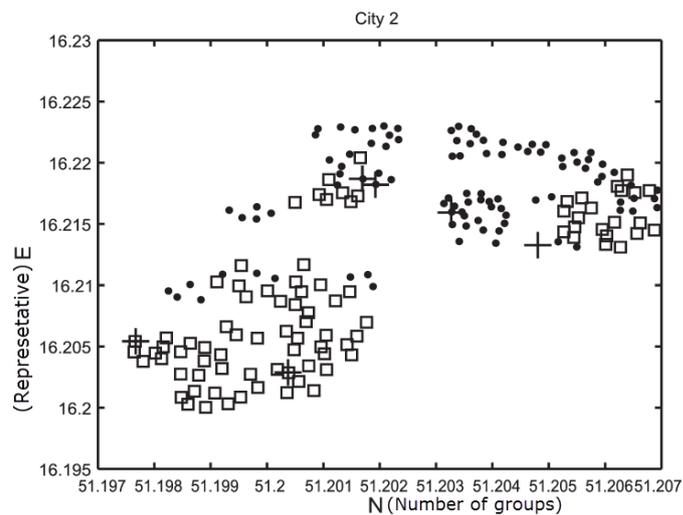
Fig. 4. Objects divided into groups and representatives. The squares represent the first group. The diamonds represent the fifth group with cross-chosen representatives
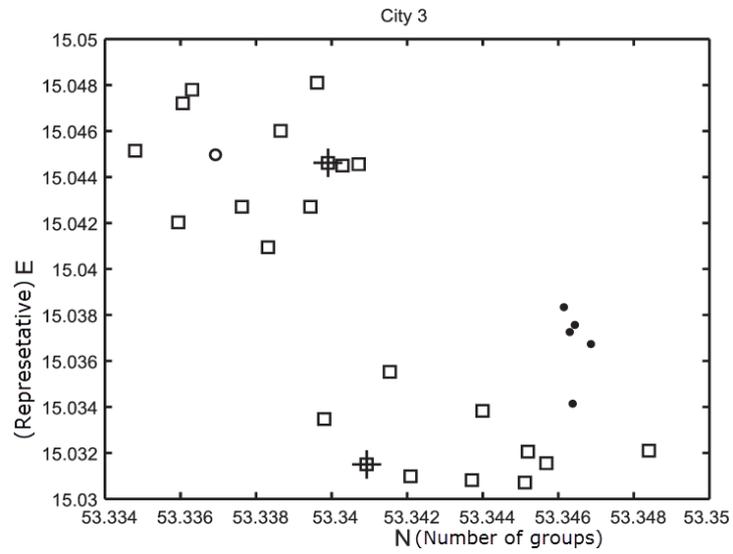
Fig. 5. Objects divided into groups and representatives. The squares represent the first group. The diamonds represent the fifth group with cross-chosen representatives
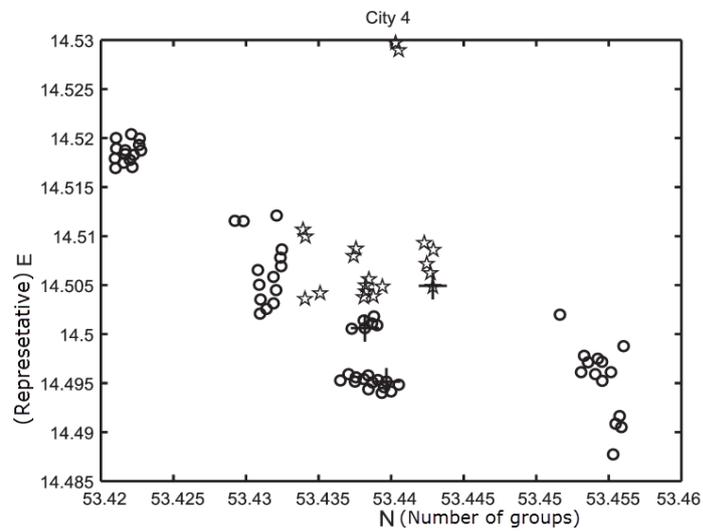
Fig. 6. Objects divided into groups and representatives. The circles represent the second group. The stars represent the third group with cross-chosen representatives
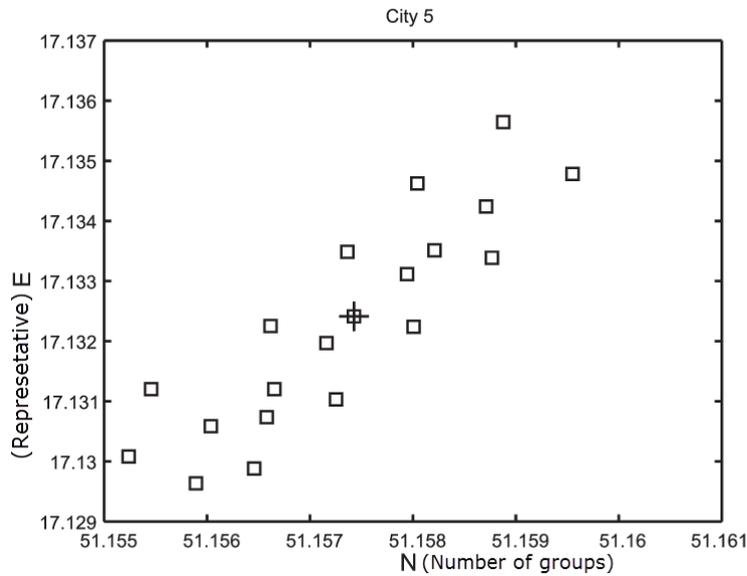
Fig. 7. Objects divided into groups and representatives. The squares represent the first
group with cross-chosen representatives

4. CONCLUSIONS

This paper presents a statistical algorithm for choosing representative residential buildings in western Poland that for detailed analysis and measurement of energy consumption needs. In the initial phase of our data analysis, the proposed algorithm enabled the identification of an optimal number (k_{opt}) of subgroups from which objects with similar characteristics could be categorized. The numerical results revealed that the most optimal division for large datasets of real objects using the k-means method was five groups (Fig. 2). Each object was assigned to the subgroups obtained from this algorithm (Table 1). An analysis of the object's characteristics in the subgroup data enabled the characterization of classes of similar buildings to be obtained. Finally, representative samples were chosen for measurement using a random choice algorithm (Table 3). The abovementioned random choice method was used for different available measurement equipment (5, 10, and 15 representatives). The results and details of the representative samples are shown in Table 4.

Table 4. Details of the representative buildings based on the number N_{rep}

Number of representatives	Representative object's address
$N_{rep} = 15$	8 – Dworcowa Street 16, Stargard Szczeciński
	28 – Bolesława Chrobrego Street 11, Stargard Szczeciński
	65 – Czorsztyńska Street 45, Szczecin
	80 – Mickiewicza Street 143-147, Szczecin
	95 – Roweckiego Street 4-4a, Szczecin
	135 – Tatrzańska Street 30, Legnica
	141 – Karkonoska Street 12-14-16, Legnica
	195 – Batalionu „Zośka" Street 2-4-6, Legnica
	196 –Tokarzewskiego Street 2-4-6, Legnica
	265 – Przybosia Street 1, Legnica
	283 – Gojawiczyńskiej Street 1, Legnica
	309 – Buczka Street 6-9, Bolesławiec
	312 – Cicha Street 9-12, Bolesławiec
	378 – Starzyńskiego Street 35-37, Bolesławiec
	411 – Hetmańska Street 2-10, Wrocław
$N_{rep} = 10$	8 – Dworcowa Street 16, Stargard Szczeciński
	65 – Czorsztyńska Street 45, Szczecin
	80 – Mickiewicza Street 143-147, Szczecin
	141 – Karkonoska Street 12-14-16, Legnica
	195 – Batalionu „Zośka" Street 2-4-6, Legnica
	265 – Przybosia Street 1, Legnica
	283 – Gojawiczyńskiej Street 1, Legnica
	312 – Cicha Street 9-12, Bolesławiec
	378 – Starzyńskiego Street 35-37, Bolesławiec
	411 – Hetmańska Street 2-10, Wrocław
$N_{rep} = 5$	65 – Czorsztyńska Street 45, Szczecin
	141 – Karkonoska Street 12-14-16, Legnica
	265 – Przybosia Street 1, Legnica
	312 – Cicha Street 9-12, Bolesławiec
	411 – Hetmańska Street 2-10, Wrocław

5. ACKNOWLEDGMENTS

REFERENCES

1. Gunel M. H., Ilgin H. E.: *A proposal for the classification of structural systems of tall buildings*, Building and Environment 42 (2007), pp. 2667-2675.
2. Theodoridou K., Papadopoulos A., Hegger M.: *A typological classification of the Greek residential building stock*, Energy and Buildings 43 (2011), pp. 2779-2787.
3. Theodoridou L., Papadopoulos A., Hegger M.: *Statistical analysis of the Greek residential building stock*, Energy and Buildings 43 (2011), pp. 2422-2428.
4. Fovell R. G., Fovell M. Y. C.: *Climate zones of the conterminous United States defined using cluster analysis*, Journal of Climate 6 (1993), pp. 2103-2135.
5. Lau C., Lam J. C., Yan L.: *Climate classification and passive solar design implications in China*, Energy Conversion and Management 48 (2007), pp. 2006-2015.
6. Wan K., Li D., Yang L., Lam J.: *Climate classifications and building energy use implications in China*, Energy and Buildings 42 (2010), pp. 1463-1471.
7. Cohen S. C. M., de Castro L. N.: *Data clustering with particle swarms*, IEEE Congress on Evolutionary Computation, 2006, pp. 1792-1798.
8. Ward J. H. Jr.: *Hierarchical grouping to optimize an objective function*, Journal of the American Statistical Association 58 (1963), pp. 236-244.
9. Dempster A. P., Laird N. M., Rubin D. B.: *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society Series B (Methodological) 39 (1977), pp. 1-38.
10. Lloyd S. P.: *Least square quantization in PCM*, IEEE Transactions on Information Theory 28 (1982), pp. 129-137.
11. MacQueen J. B.: *Some methods for classification and analysis of multivariate observations*, in: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, 1967, pp. 281-297.
12. Steinhaus H.: *Sur la division des corps materiels en parties*, Bulletin of the Polish Academy of Sciences Mathematics 4 (1956), pp. 801-804.
13. *Finding the Right Number of Clusters in k-Means and EM Clustering: v-Fold Cross-Validation*, in: Electronic Statistics Textbook, StatSoft, Tulsa, 2010.
14. Rousseu P.J.: *Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis*, Computational and Applied Mathematics 20 (1987), pp. 53-65.

15. Sugar C. A., James G.M.: *Finding the number of clusters in a data set: an information theoretic approach*, Journal of the American Statistical Association 98 (2003), pp. 750-763.
16. Bartlett J. E., Kortlik J. W., Higgins C.: *Organizational research: determining appropriate sample size for survey research*, Information Technology, Learning, and Performance Journal 19 (2001), pp. 43-50.
17. Kish L.: *Survey Sampling*, Wiley, New York, 1995.
18. Lohr S.L.: *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, 1999.

STATYSTYCZNA METODA WYBORU BUDYNKÓW BADAWCZYCH NA POTRZEBY ANALIZY SYSTEMÓW HVAC

S t r e s z c z e n i e

W artykule przedstawiono statystyczną metodę wyboru budynków reprezentatywnych pod względem charakterystyki energetycznej oraz cech wbudowanych systemów grzewczych wentylacyjnych i klimatyzacyjnych. Proponowane podejście opiera się na metodzie k-średnich. Algorytm dla tej metody jest stosunkowo prosty, co pozwala na łatwe wdrożenie i nie wymaga dużego nakładu (kosztu) obliczeniowego. Sposób ten może być stosowany dla dużych ilości danych. Metodą k-średnich badano dane pozyskane w czasie inwentaryzacji obiektów oraz w wyniku symulacji komputerowych funkcjonowania budynków, zawierające m.in. roczne wartości zapotrzebowania na ciepło (symulowane z krokiem czasowym godzinnym). Z uwagi na względną prostotę metodyki oraz uzyskane bardzo dobre wyniki, proponowane podejście jest bardzo obiecujące dla zastosowań technicznych, w tym analiz budynków pod kątem systemów HVAC.

Słowa kluczowe: metoda k-średnich, systemy HVAC, zużycie energii, charakterystyka energetyczna budynków