

Article

The Use of Cluster Analysis to Evaluate the Impact of COVID-19 Pandemic on Daily Water Demand Patterns

Paulina Dzimińska¹, Stanisław Drzewiecki¹, Marek Ruman^{2,*}, Klaudia Kosek^{3,*}, Karol Mikołajewski^{2,4} and Paweł Licznar⁵

- ¹ MWiK—The Water Supply and Sewage Company of Bydgoszcz Sp. z o.o., 85-817 Bydgoszcz, Poland; paulina.dziminska@mwik.bydgoszcz.pl (P.D.); drzewiecki@mwik.bydgoszcz.pl (S.D.)
- ² Faculty of Natural Sciences, University of Silesia in Katowice, 41-200 Sosnowiec, Poland; karol.mikolajewski@retencja.pl
- ³ Faculty of Civil and Environmental Engineering, Gdansk University of Technology, 80-233 Gdansk, Poland
- ⁴ RETENCJAPL Sp. z o.o., 80-868 Gdansk, Poland
- ⁵ Faculty of Environmental Engineering, Wrocław University of Science and Technology, 50-377 Wrocław, Poland; pawel.licznar@pwr.edu.pl
- * Correspondence: marek.ruman@us.edu.pl (M.R.); klaudia.kosek@pg.edu.pl (K.K.)

Abstract: Proper determination of unitary water demand and diurnal distribution of water consumption (water consumption histogram) provides the basis for designing, dimensioning, and all analyses of water supply networks. It is important in the case of mathematical modelling of flows in the water supply network, particularly during the determination of nodal water demands in the context of Extended Period Simulation (EPS). Considering the above, the analysis of hourly water consumption in selected apartment buildings was performed to verify the justification of the application of grouping by means of k-means clustering. The article presents a detailed description of the adopted methodology, as well as the obtained results in the form of synthetic distributions of hourly water consumption, and the effect of the COVID-19 pandemic on their change.

Keywords: COVID-19; water consumption; urban environment; residential water demand pattern; cluster analysis; statistics



Citation: Dzimińska, P.; Drzewiecki, S.; Ruman, M.; Kosek, K.; Mikołajewski, K.; Licznar, P. The Use of Cluster Analysis to Evaluate the Impact of COVID-19 Pandemic on Daily Water Demand Patterns. *Sustainability* **2021**, *13*, 5772. <https://doi.org/10.3390/su13115772>

Academic Editors:
Danuta Szumińska,
Aleksandra Jezierska-Thöle and
Marta Gwiaździska-Goraj

Received: 15 April 2021
Accepted: 17 May 2021
Published: 21 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Research and analyses of water demand for the purposes of, e.g., housing, services, or public purpose buildings has been enjoying increasing popularity in recent years. It is primarily due to the development of technology allowing for the measurement of water flow or automatic water meters readings (AMR) with a higher frequency than so far (once a month, quarter, or even year—depending on the adopted system of billing settlements), but also due to an increase in the use of computer networks in the exploitation of modern water distribution systems, allowing for the collection of any amount of data. This progress permits the measurement of water flow with any time interval [1].

The preparation of an accurate diagram of water demand is essential from the point of view of mathematical modelling of water distribution systems. It also provides the basis for appropriate designing and dimensioning of water supply networks, connections, and selection of the water meter [1,2].

Water demand is determined by a number of variables depending on the type of object to which water is supplied. In housing, the primary factors determining the course of water consumption during the day include daily behaviours of residents, their lifestyle, and routine activities involving the consumption of supplied water. Another important aspect is the day of the week, i.e., whether it is a working day or a holiday (weekend or a bank holiday). The announcement of the state of epidemic emergency in the country, caused by the virus SARS-CoV-2, could have considerably affected daily behaviours of

residents. The COVID-19 pandemic has imposed massive health and economic burdens on communities worldwide, and no sector of society is going untouched, including the water sector [3]. Moreover, while business is up for a handful of sectors, such as hospitals and some food production, other important water-using sectors have slowed or shut down entirely. The average effect of the COVID-19 pandemic on total water demand varies, depending on the relative proportion of residential water consumption [4]. Issues related to water consumption are a growing challenge in terms of sustainability, especially in developing countries, although reaching sustainable water resources development is a matter of global importance [3], especially during the pandemic period.

Due to high variability of water consumption, however, measurement of water flow and data collection on the server are insufficient for their use in, e.g., mathematical modelling. It is necessary to perform analysis with the application of a statistical model that allows for standardisation, classification, and selection of the most probable water consumption histograms [4], which is the topic of conducted research described in this article.

1.1. Background

Interest in processing databases of water meter records has been increasing in recent years due to increasingly frequent application of computer hydraulic modelling in the management and dimensioning of water supply networks. Hydraulic modelling of water supply networks is performed with the application of EPS, determining quasi-dynamic behaviour of the system in time, by calculating the state of the system for a series of determined simulations in which hydraulic demand and threshold conditions change in time. Credibility of a hydraulic model for conducting any analyses of the operation of the distribution system requires its relevant calibration [5]. This process requires accurate measurement of water consumption with the application of AMR throughout the water supply system, and then defining threshold conditions within the consumption range, not only in the form of mean diurnal consumption in computational nodes, but also their changes in time. The already developed models of water consumption can be found in the literature. For example, Obradović and Lonsdale [6] devoted a considerable part of their publication to models of water consumption for an exceptionally broad circle of recipients, including a hospital, university, school, church, bank, hotel, authority office, military barracks, prison, industrial factories by type, and many others. In the introduction, the authors emphasised, however, that an attempt to determine “typical patterns” of water consumption is doomed to be unsuccessful due to the overlapping of a number of factors, often difficult to identify, that affect water consumption [6]. Therefore, the proposed form of processing data involved averaged hourly models of water consumption with their maximum and minimum envelopes.

As mentioned above, such models can also be obtained through AMR solution and further analysis of their results. It should be emphasised that this type of solution is possible for all recipients and requires limitation to designating so-called reference recipients for whom hourly patterns of water consumptions are identified and then ascribed to particular nodes of the hydraulic model [5,6].

Although the technique of allocation of water consumption variables in time for particular computational nodes based on the selected reference recipients appears simple, unfortunately, it is not common. The conventional method described as the top-down approach is applied considerably more frequently. It assumes water consumption in nodes of the model determined only for entire zones of water supply and then ascribed to all nodes of the network with consideration of correlation coefficients [7]. This not only permits the simplification of the development of computer models of water supply networks, but also fast obtainment of convergences of computational algorithms in the calibration process. The resulting image of water flows, however, can carry a significant error.

Considering the above, Blokker et al. [7] proposed an entirely different method of estimation of water consumption in computational nodes of the network, described as the bottom-up approach. They developed a stochastic model of water consumption by the final

recipient (for a unitary water supply connection), based only on statistical information on water users such as: number of residents, their age, frequency of water consumption, duration and rate of water flow per water consumption episode, and frequency of occurrence of different types of water consumption for different purposes during the day (e.g., for flushing the toilet, washing clothes, and washing hands). The model is called SIMDEUM (Simulation of water Demand; an End-Use Model) and is based on results of earlier research by Buchberber and Wells [8]. Its authors evidenced the possibility of describing water consumption patterns for household purposes by means of the non-homogeneous Poisson process (PRP).

The PRP model in reference to modelling water consumption found application among others in countries such as the USA [9], Spain [10], or Mexico [11]. In each case, the implementation required conducting a measurement campaign, and the obtained parameters of the PRP model showed mutual differences. The SIMDEUM model appears to have more practical application, because it does not require conducting research on water flows and is limited to the identification of statistical parameters of end consumption. This offers the possibility of generating models of water consumption also for non-existing areas planned for expansion. Moreover, the model can be applied not only to water consumption for household purposes, as in the case of the PRP model, but also to recipients such as, e.g., hotels, office buildings, or social care facilities. Both models generate consumption patterns with high resolution for the individual recipient, whereas hydraulic models of water supply networks are calculated with a 1 h time step. Therefore, Alvisi et al. [12] proposed a special spatial–temporal procedure for the aggregation of synthetic water consumption originally generated at the level of a single recipient and small-time step (e.g., 1 min–SIMDEUM and PRP) to synthetic series of consumption related to a group of recipients with a temporal resolution of 1 h.

Next to the already discussed models, Aksela and Aksela [2] proposed another solution employing the probabilistic model for generating patterns of water demand in single-family and semi-detached housing, with a temporal resolution of 1 h. The computation procedure involves the following first two steps: forecasting of weekly water consumption based on the linear regression model and division of the analysed water recipients into four separate classes based on the recorded mean weekly consumption. The division was performed by means of the cluster analysis and specifically k-means clustering. Then, the probability distribution of water consumption in time was modelled with the application of mixture of Gaussian models, and eventually, final probabilistic water consumption models were developed as a result of sampling from the probability distributions determined earlier.

The aforementioned examples of publications focusing on the issue of modelling of water consumption patterns reveal the current availability of specialised computer models allowing for the determination of the course of water consumption by individual recipients, primarily in housing. This issue, however, still enjoys great interest in the scientific circles, which provided the basis for this paper. For this purpose, a water consumption measurement campaign was conducted in three apartment buildings with a temporal resolution of 1 h. The temporal resolution resulted from the possibility of data registration by the existing AMR system. The determination of patterns of hourly demand for water in these buildings employed the cluster analysis, namely k-means clustering.

1.2. Formulation of the Model

The cluster analysis is a discipline in multi-dimensional statistics that includes a group of methods for the identification of uniform subsets of elements. Based on variables characteristic of our elements, the cluster analysis finds groups (clusters) of elements that are similar to elements belonging to the same cluster and simultaneously differ from those in the remaining clusters [13]. Importantly, depending on the adopted method, the division can occur to an a priori non-defined or defined number of groups (also known as clusters). The obtained groups cover subsets of the analysed population satisfying the conditions of



their decouplability and completeness [13]. The division of the set of water consumption diagrams initially involved the application of the hierarchical agglomeration method and then the non-hierarchical method–k-means clustering.

The hierarchical agglomeration method initially assumes that each element is a separate cluster. Then, it gradually combines the mutually nearest (most similar) elements into new clusters, until a single cluster is obtained. The determination of sufficient similarity of two clusters requires defining measures of distance between the elements and developed clusters, and the rules of their combining. A review of available distance measures and their characteristics can be found in the publication by Stanisiz [13] and the cited source literature [14]. The analyses employed the most popular and considered most natural distance metrics: Euclidean metric and square Euclidean distance. The Euclidean distance metric $d(x,y)$ of elements x and y is represented by the following Formula (1) [13,15]:

$$d(x,y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (1)$$

where $x = (x_1, \dots, x_r)$ and $y = (y_1, \dots, y_r)$. Formula (1) for $p = 2$ and $p = 3$ is the equivalent to distance on a plane and in space of two points x and y .

In the case of methods of linkage, a broad range of algorithms is also available, primarily including the following methods: single linkage, complete linkage, unweighted pair-group average, weighted pair-group average, unweighted pair-group centroid, weighted pair-group centroid, and Ward's method. The analysis of water consumption employed the popular unweighted pair-group average method, where differences (distances) between all pairs of elements included in particular clusters are calculated. Averaged differences between all pairs are adopted as the measure of distance between particular clusters. The application of that computation method allows for the determination of which elements are similar to each other and can be included to the same clusters, and to what degree particular clusters are similar to each other and can be combined into structures of larger clusters. As a result of these methods, development of characteristic "chains" composed of elements combined into sequences forming clusters is observed for similar elements of the set. This permits easy identification of mutually strongly similar elements. The aforementioned chains are evidently visible in dendrograms resulting from agglomeration, constituting graphical illustration of the structure of a set of elements by decreasing similarity between its elements (and therefore increasing linkage distances).

The methods presented above lead to obtaining a dendrogram in which lower grade clusters are included in the composition of higher-grade clusters. K-means clustering is considerably different. It is a non-hierarchical method resulting in a breakdown in which no cluster is a sub-cluster of another cluster. In k-means clustering, k clusters are designated, differing from each other to the highest possible degree. It is necessary to assume a priori a specific number k of clusters into which the set of elements is partitioned. Therefore, k subsets are developed, and then objects are moved between them for the distance between them within subsets to be possibly small, and for the distance between clusters to be possibly large. The transfer procedure is repeated towards the most effective distinction of clusters. It can be traced in detail based on the calculation example presented in the publication by Larose [15]. In the study, the possibility of use of the clustering method was preceded by analysis with the application of the agglomeration method. As a result, a potential range of the number of clusters was determined, to be considered in the case of partitioning of patterns of hourly water demand in the object. Then, from that range, the optimal number of clusters was determined, conducted with the application of the analysis of total within sum of squares (wss) and Caliński and Harabasz Index values ($CHIndex$) for a different number of clusters. The total within the sum of squares wss is calculated based on the Formula (2) below:

$$wss = \sum_i^k \sum_{x \in C_i} \|x - m_i\|^2 \quad (2)$$



where k —number of clusters, x —element of cluster, C_i — i -th data cluster, m_i —centroid of the cluster i , and $\|x - m_i\|^2$ —Euclidean distance between two vectors.

The *CHIndex* is calculated based of the following Formula (3) [16]:

$$CHIndex = \frac{SS_B}{SS_W} \times \frac{N - k}{k - 1} \quad (3)$$

where k —number of clusters, N —total number of observations (elements of the set), SS_B —total variance between clusters (trace of interclass covariance matrix), and SS_W —total within cluster variation (trace of intraclass covariance matrix).

It is assumed that the optimal number of clusters equals obtaining the maximum value of the *CHIndex*. A high value of the index is related to maximising the ratio of SS_B and SS_W , which means that particular clusters show considerable differences, and elements of the set grouped in particular clusters show strong similarity (relatively weakly differentiated). The *wss* value is also important. It naturally decreases with an increase in the number of clusters, and after reaching the optimal number of clusters, the rate of the decrease substantially decreases [16].

As a result of adopting the number of clusters, after running the k -clustering algorithm, it is possible to eventually assign specific patterns to the adopted clusters but also to prepare diagrams of averaged histograms of accumulated water demand for such clusters. The clustering algorithm was run three times by means of the bootstrap method, used for estimating the distribution of estimation errors, by means of multiple random drawing with sample return. This means that the clustering algorithm was performed each time for random samples from the entire set of water demand diagrams. Results obtained in subsequent iterations are comparable, permitting the determination of the bootmean parameter values. The bootmean parameter is calculated as the average value of the Jaccard index (Jaccard similarity coefficient) for each cluster. The Jaccard coefficient itself measures similarity between two sets and is determined as a quotient of the power set of the intersection of sets and power set of the sum of these sets. It is assumed that the value of the bootmean parameter should be higher than 0.6, because it is presumed from this value that the designated clusters do not include a random cluster, i.e., a cluster that includes patterns deviating from the remaining clusters but at the same time mutually dissimilar [16].

2. Materials

2.1. Characteristics of the Studied Objects

The research on water demand covered 3 mutually similar apartment buildings constructed in the 1980s within the same housing estate in Bydgoszcz, which is located in northern Poland. With a population of approximately 348,000, Bydgoszcz is the eighth largest city in the country [17].

The buildings chosen for research are composed of 5 floors and are divided into 4 or 5 staircases, depending on the number of apartments. One of them features a total of 60 apartments, and the remaining ones feature 50 apartments each. All apartments are equipped with the basic sanitary facilities, i.e., washbasin taps, kitchen sink taps, toilet flushes, showers or bath tubs, and washing machines. Moreover, each building has a discharge valve for watering greenery. A list of the facilities with the indication of the number of apartments and water meter diameter is presented in Table 1.

The primary source of water supply in the buildings is water supply service connections with a diameter of DN50, with mounted volumetric water meters with diameters DN32 and DN40, characterised by high sensitivity to very low flows. Moreover, due to the principle of their operation, their installation does not require keeping straight sections in front of and behind the water meter. They can also be mounted in any position, because it does not affect the metrological properties of the device. Their only disadvantage is high sensitivity to solid particles occurring in water [18].

Table 1. List of sanitary facilities.

	Apartment Building 1	Apartment Building 2	Apartment Building 3
Type of Facility	Number of Sanitary Facilities		
Washbasin tap	60	50	50
Kitchen sink tap	60	50	50
Toilet flush	60	50	50
Shower/bathtub	60	50	50
Washing machine	60	50	50
Discharge valve	1	1	1

Table 2 below presents the basic parameters for each water meter. Reading of data from water meters was performed by means of transceivers or recording attachment, permitting recording hourly water consumption.

Table 2. Technical parameters of water meters.

Parameters		d_{nom} 32 mm	d_{nom} 40 mm
Nominal flow rate	Q_3 [m ³ /h]	10	16
Maximum flow rate	q_s [m ³ /h]	12.5	20
Transition flow rate	q_t [dm ³ /h]	100	160
Minimum flow rate	q_{min} [dm ³ /h]	62.5	100
Water meter start flow	[dm ³ /h]	3	3
Relative error of measurement	within $q_s \div q_t$	±2%	±2%
	below $q_t \div q_{min}$	±5%	±5%

2.2. Study Period

The study of hourly water demand in the discussed objects by means of the aforementioned devices commenced on 16 May 2019 and lasted, with intervals, until 6 October 2020, over a total of 464 days of measurement. Breaks in measurements resulted among others from the necessity to replace water meters—on 14 January a water meter with a reading panel in the form of a recorder was replaced, causing temporary problems with data reading; from the outbreak of the COVID-19 pandemic; and more specifically, from the ‘total lockdown’ from March 2020 and all the resulting restrictions that made readings temporarily impossible.

Therefore, the analysis of hourly water demand in the each of three buildings covered 464 days, whereas:

1. 331 days were business days,
2. 133 days were days free from work (weekends).

Moreover, the data set for 464 days was partitioned into two separate sets to allow for the analysis of differences in water use by recipients caused by the announcement of the state of epidemic emergency in the country. Therefore, further in the paper, the following division is applied:

3. Period before the COVID-19 pandemic covering 276 days, whereas:
 - a. 197 days were business days,
 - b. 133 days were days free from work (weekends).
4. Period during the COVID-19 pandemic covering 188 days, whereas:
 - a. 134 days were business days,
 - b. 54 days were days free from work (weekends).

It was also necessary for the data sets to take account of holidays. For the period before the COVID-19 pandemic, their number was 8, and they occurred on business days, and for the period during the pandemic, their number was 5, and 3 of them occurred on business days. After consideration of the additional division, the data sets adopted for analysis were of the following character:

1. for the period before the COVID-19 pandemic:
 - a. 189 days were business days,
 - b. 87 days were days free from work (weekends and holidays),
 - c. i.e., 68.5% and 31.5% of all days, respectively.
2. for the period during the COVID-19 pandemic:
 - a. 131 days were business days,
 - b. 57 days were days free from work (weekends and holidays),
 - c. i.e., 68.7% and 31.3% of all days, respectively.

3. Calculations

Data initially obtained from AMR system regarding hourly water consumption in the object were presented in a table with one column specifying the date and time of reading, and the other total hourly flow through the water meter in units dm^3/h . The data were partitioned with consideration of two periods: before the outbreak of the COVID-19 pandemic and during the pandemic. Then, the water flow data were processed to obtain data matrices in which rows represented subsequent days of the study, and columns represent the percent contribution of hourly water consumption (of total diurnal water demand). It should be emphasised that each row of the developed matrices formed an individual histogram of water consumption for each day. A total of 6 data matrices were prepared. They were used in subsequent stages to perform classification by means of the cluster analysis.

The first stage of the cluster analysis employed the hierarchical agglomeration method in which the determination of the measure of distance between elements and the resulting clusters employed the Euclidean distance metric, and the linkage rule was determined by means of the unweighted pair-group average. The set agglomeration resulted in a graphical illustration of the structure of the set of elements by decreasing similarity between its elements (i.e., increasing linkage distance). Dendrograms for each of the objects with division into measurement periods are presented in Figures 1–6. They suggest the occurrence of clusters, i.e., groups of demand patterns with similar histograms. Strong outlier elements (histograms) were also identified, for example, day number 124 for apartment building 1 (Figure 1).

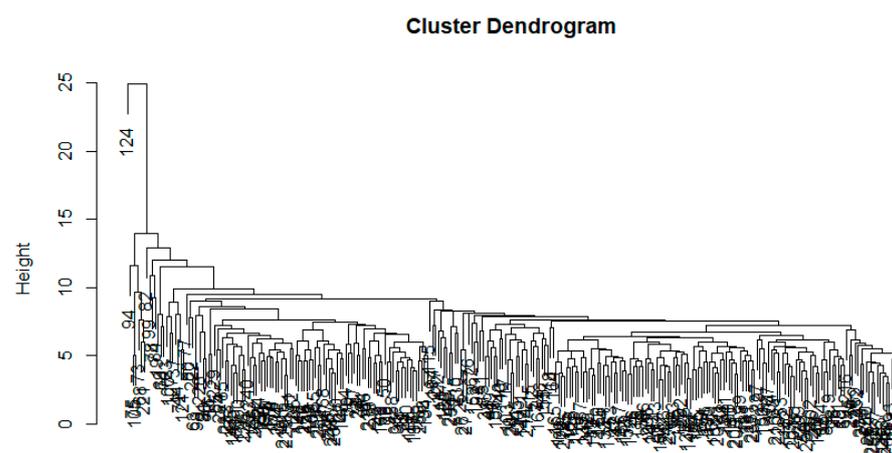


Figure 1. Dendrogram obtained for 276 days of normalised histograms of hourly flows for apartment building 1 from the period before the COVID-19 pandemic. Axis y shows linkage distances for particular flows and groups (clusters) of patterns.

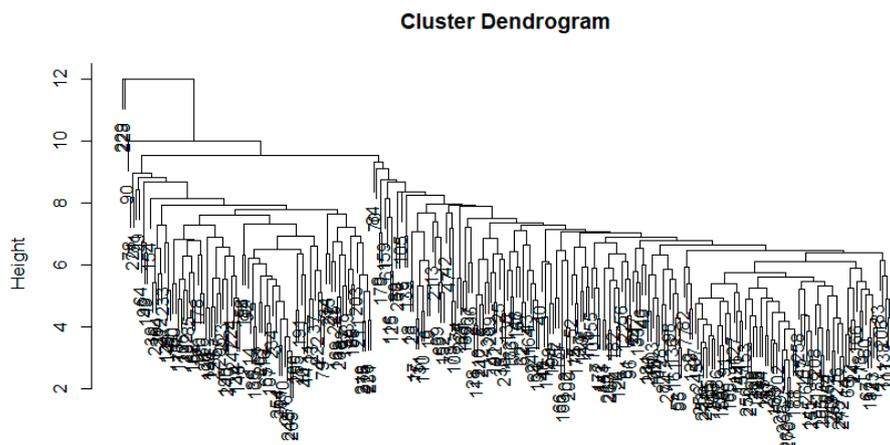


Figure 2. Dendrogram obtained for 276 days of normalised histograms of hourly flows for apartment building 2 from the period before the COVID-19 pandemic. Axis *y* shows linkage distances for particular flows and groups (clusters) of patterns.

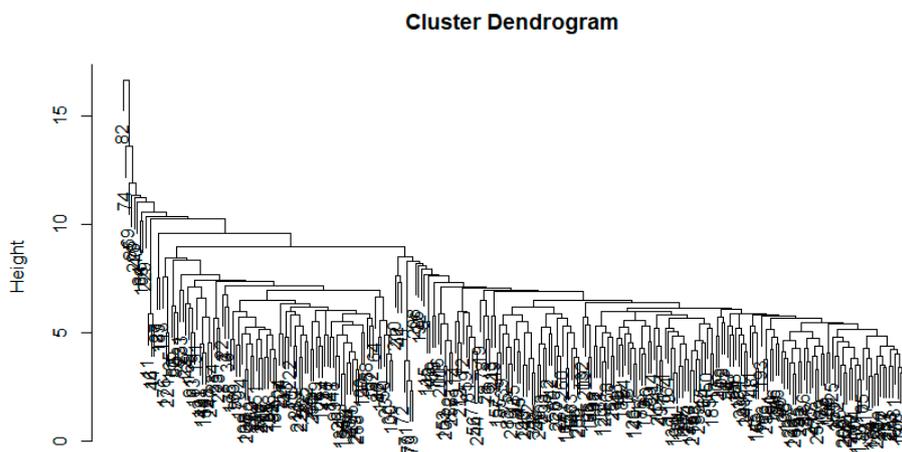


Figure 3. Dendrogram obtained for 276 days of normalised histograms of hourly flows for apartment building 3 from the period before the COVID-19 pandemic. Axis *y* shows linkage distances for particular flows and groups (clusters) of patterns.

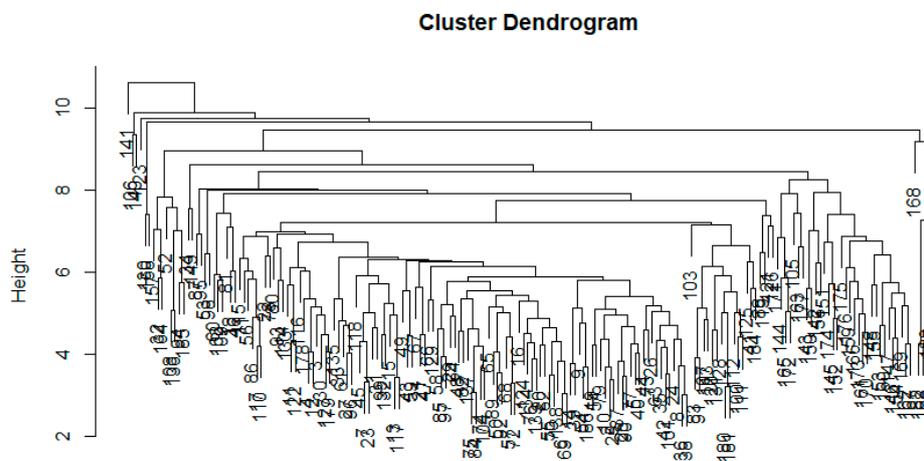


Figure 4. Dendrogram obtained for 188 days of normalised histograms of hourly flows for apartment building 1 from the period during the COVID-19 pandemic. Axis *y* shows linkage distances for particular flows and groups (clusters) of patterns.

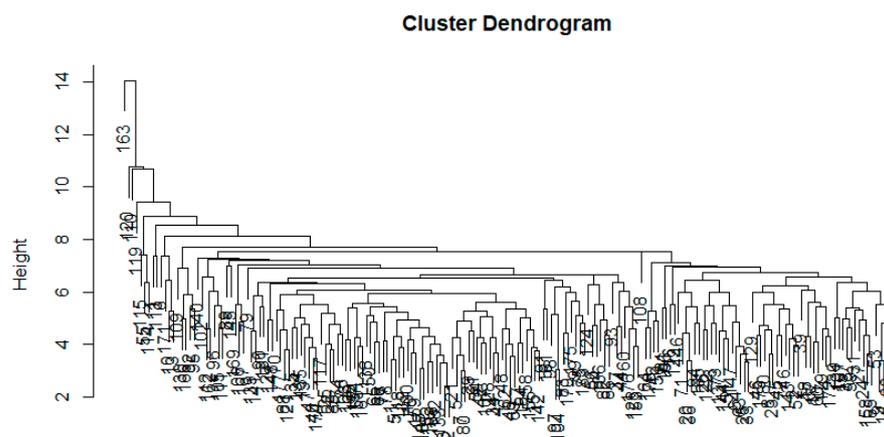


Figure 5. Dendrogram obtained for 188 days of normalised histograms of hourly flows for apartment building 2 from the period during the COVID-19 pandemic. Axis y shows linkage distances for particular flows and groups (clusters) of patterns.

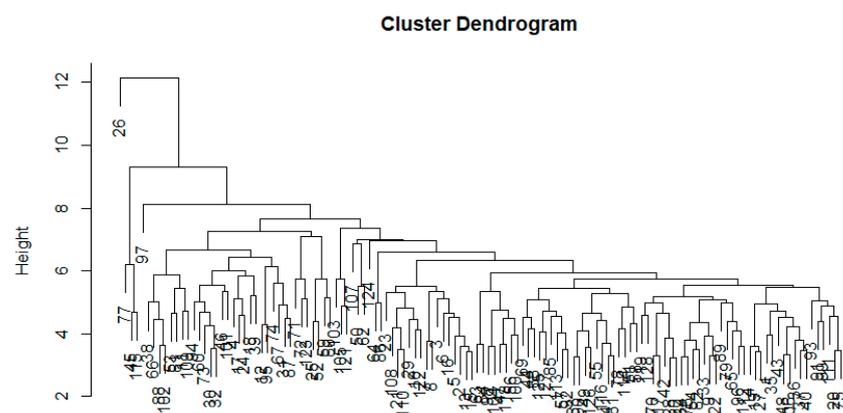


Figure 6. Dendrogram obtained for 188 days of normalised histograms of hourly flows for apartment building 3 from the period during the COVID-19 pandemic. Axis y shows linkage distances for particular discharges and groups (clusters) of patterns.

It should be emphasised that the cluster analysis is very sensitive to strong outliers. Therefore, it is recommended to remove them from the set of elements. Before further analyses based on the dendrograms (Figures 1–6), the strongest outlier histograms were eliminated from the set. Then, agglomeration of data was performed again, resulting in new structures of the set of elements. The operation of exclusion of days with anomalies in water consumption was performed for buildings 2 and 3 in the period before the pandemic and for buildings 1 and 2 for the period during the pandemic. In the case of apartment building 1 for the period before the COVID-19 pandemic, as well as apartment building 3 for the period during the COVID-19 pandemic, the current dendrogram was kept due to single outlier days that did not considerably affect the further part of computations.

Therefore, Figures 7 and 8 present dendrograms for buildings 2 and 3 from the period before the pandemic the structures of which changed considerably. The set is evidently divided into two clusters with single histograms deviating from these two clusters. Figures 9 and 10 for buildings 1 and 2 from the period during the COVID-19 pandemic also show changed structures. In their case, the partitioning into clusters is ambiguous.

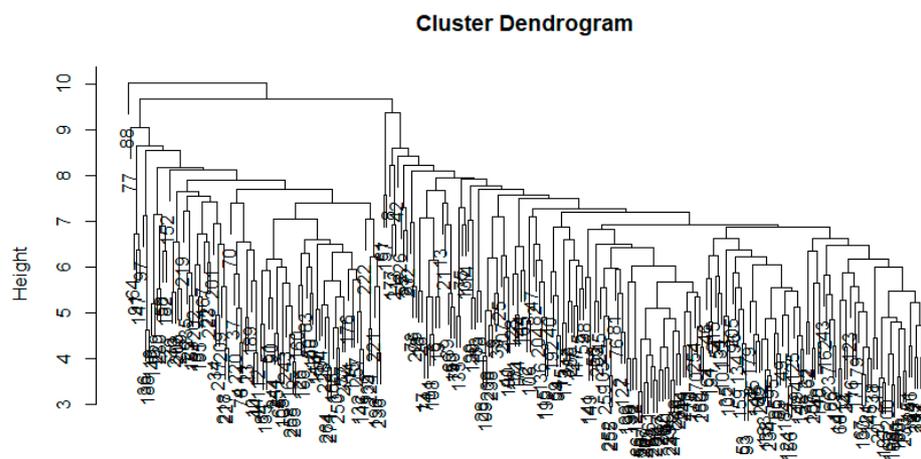


Figure 7. Dendrogram obtained for 272 days of normalised histograms of hourly flows for apartment building 2 from the period before the COVID-19 pandemic. Axis y shows linkage distances for particular discharges and groups (clusters) of patterns.

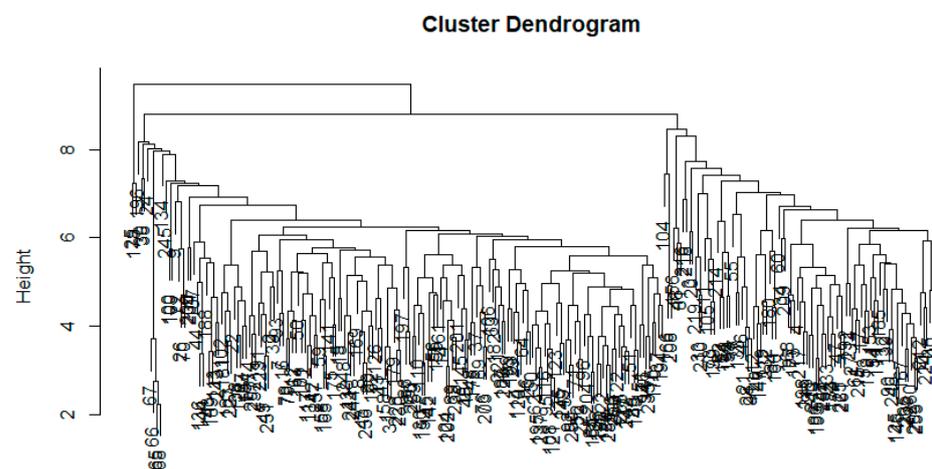


Figure 8. Dendrogram obtained for 263 days of normalised histograms of hourly flows for apartment building 3 from the period before the COVID-19 pandemic. Axis y shows linkage distances for particular flows and groups (clusters) of patterns.

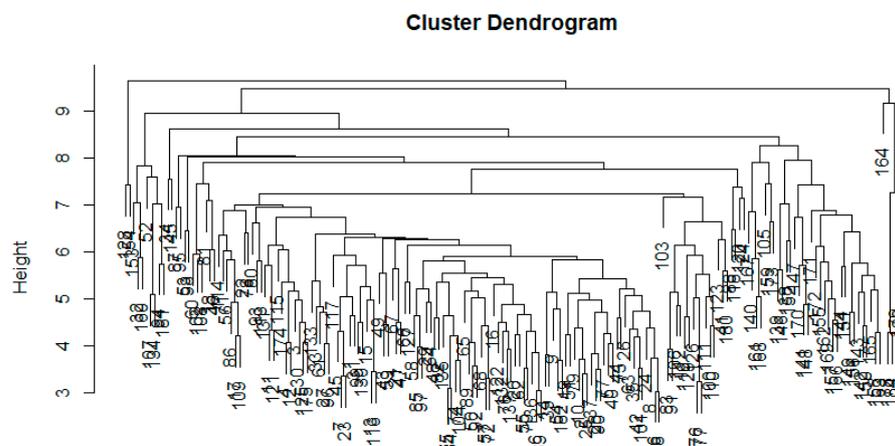


Figure 9. Dendrogram obtained for 184 days of normalised histograms of hourly flows for apartment building 1 from the period during the COVID-19 pandemic. Axis y shows linkage distances for particular flows and groups (clusters) of patterns.

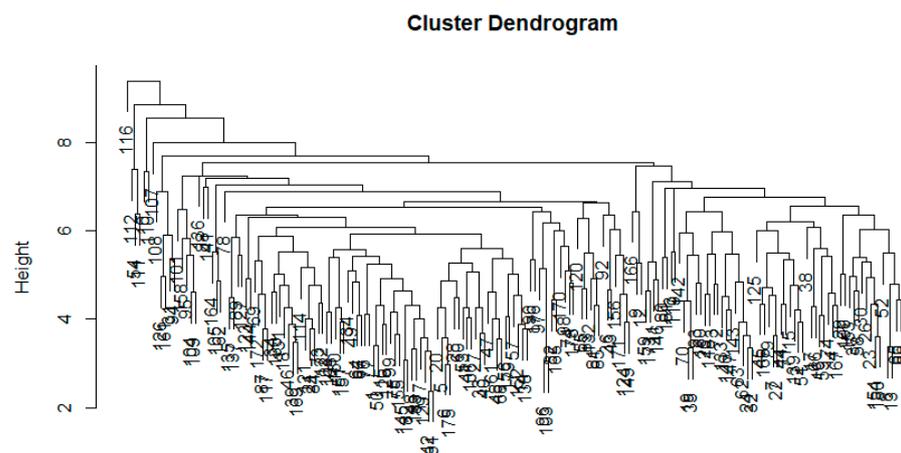


Figure 10. Dendrogram obtained for 183 days of normalised histograms of hourly flows for apartment building 2 from the period during the COVID-19 pandemic. Axis y shows linkage distances for particular discharges and groups (clusters) of patterns.

The analyses of data agglomeration resulting in dendrograms permitted the determination of the potential range of the number of clusters that should be considered in the case of partitioning of patterns of hourly water consumption in buildings. The optimal number of clusters required for k -means clustering was determined based on the analysis of the total wss and $CHIndex$ for different numbers of clusters. Both of these analyses were again performed and presented in the graphic form for each object separately with division into the period before the pandemic (Figures 11–13) and the period during the COVID-19 pandemic (Figures 14–16).

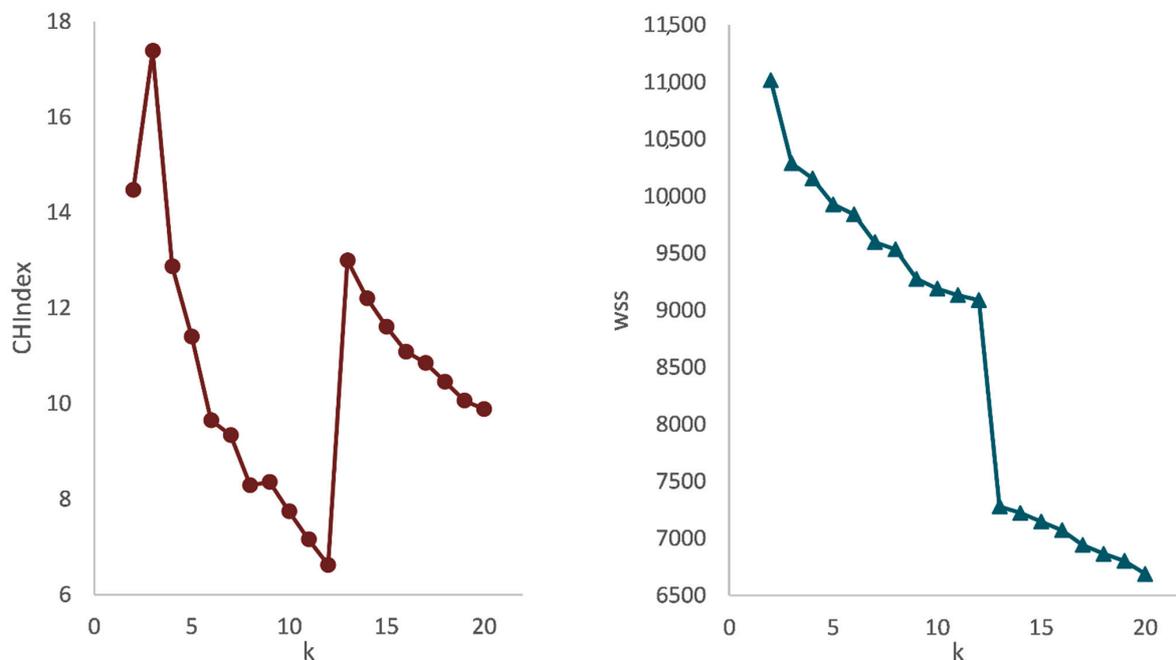


Figure 11. Value of the $CHIndex$ and total within sum of squares (wss), and for a set of 276 days of normalised histograms of hourly flows for apartment building 1 from the period before the COVID-19 pandemic, depending on the adopted number of clusters k .

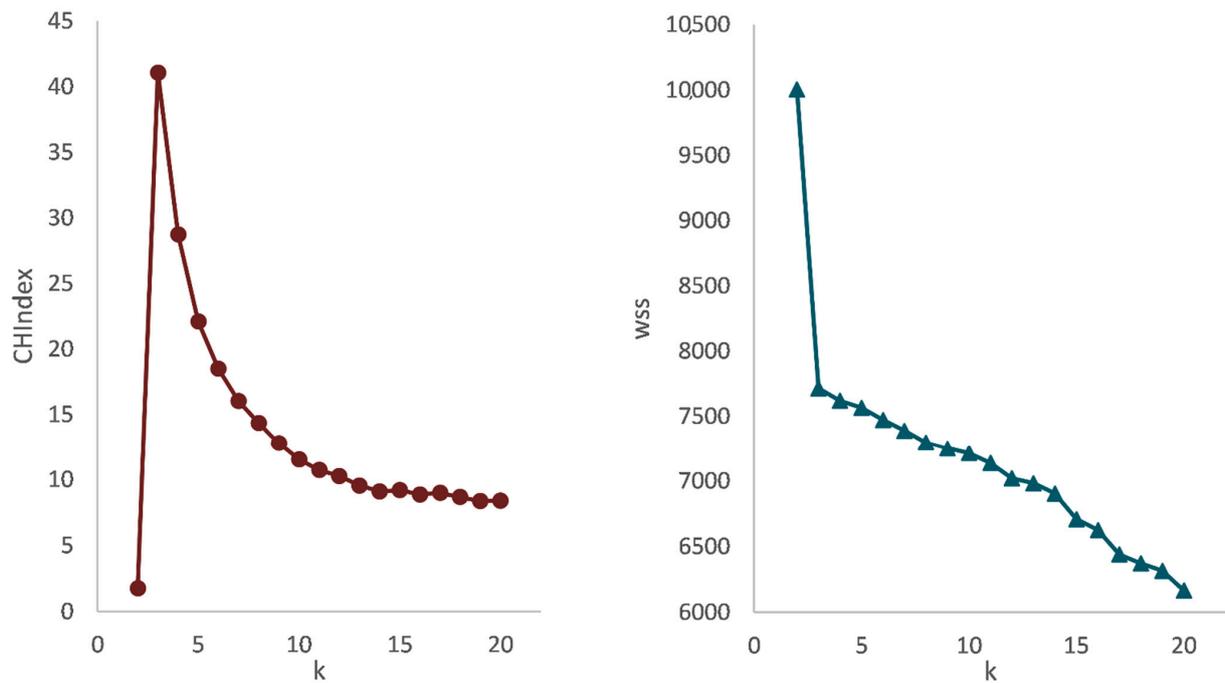


Figure 12. Value of the $CHIndex$ and total within sum of squares (wss), and for a set of 272 days of normalised histograms of hourly flows for apartment building 2 from the period before the COVID-19 pandemic, depending on the adopted number of clusters k .

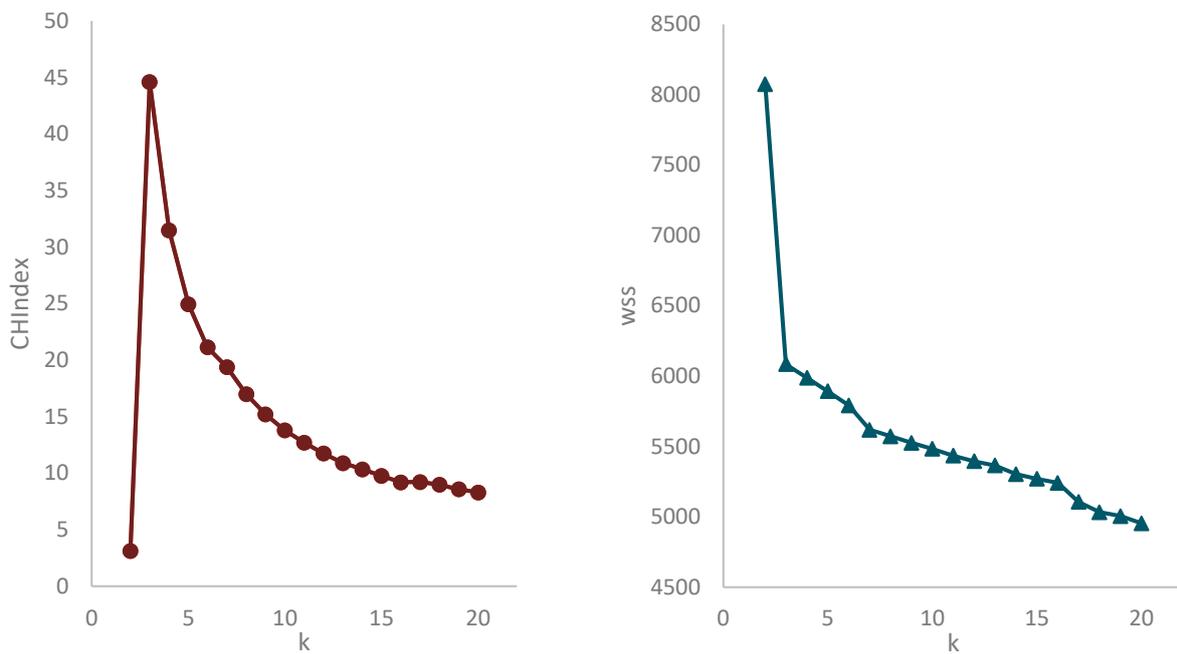


Figure 13. Value of the $CHIndex$ and total within sum of squares (wss), and for a set of 263 days of normalised histograms of hourly flows for apartment building 3 from the period before the COVID-19 pandemic, depending on the adopted number of clusters k .

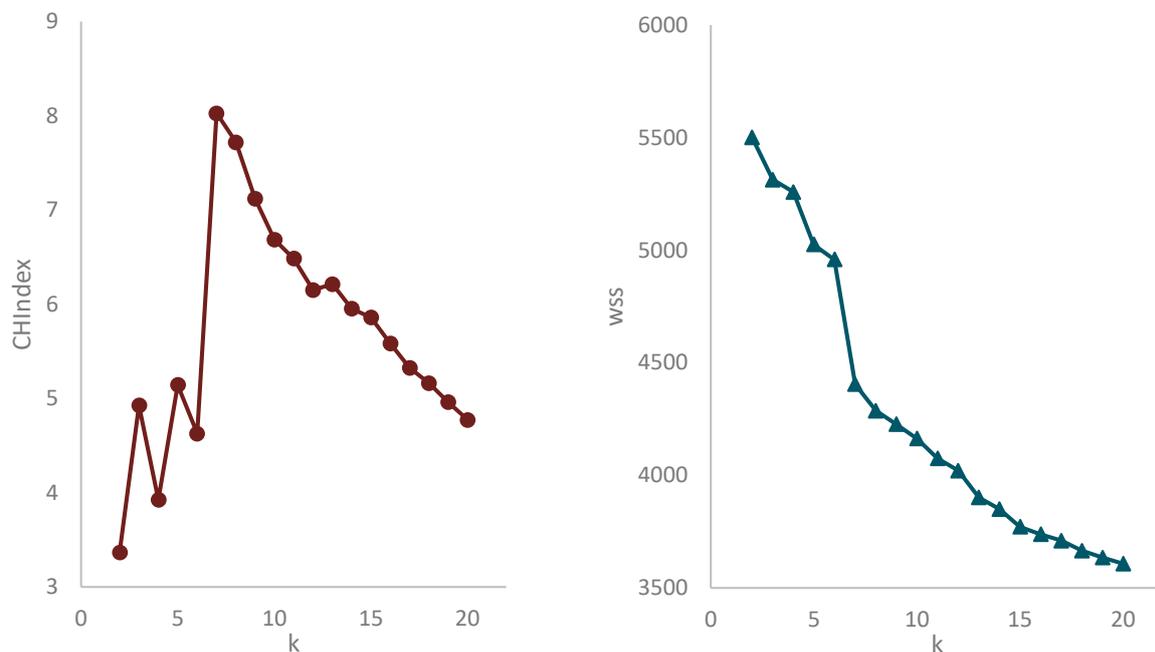


Figure 14. Values of the *CHIndex* and total within sum of squares (*wss*) and for a set of 184 days of normalised histograms of hourly flows for apartment building 1 from the period during the COVID-19 pandemic, depending on the adopted number of clusters *k*.

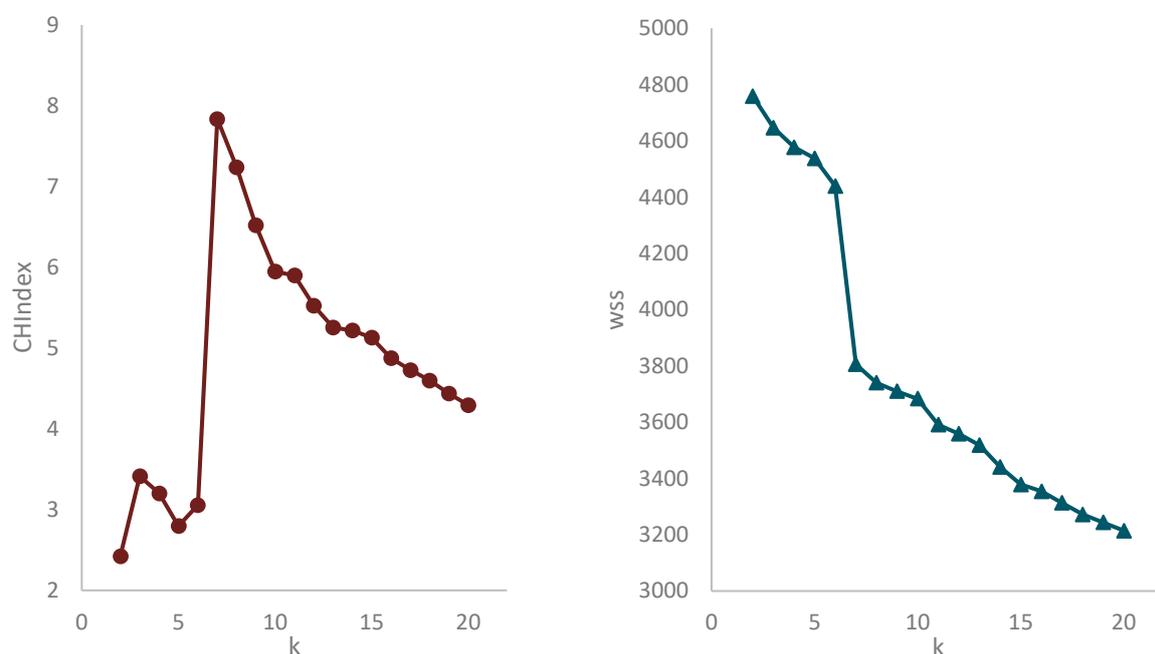


Figure 15. Values of the *CHIndex* and total within sum of squares (*wss*) and for a set of 183 days of normalised histograms of hourly flows for apartment building 2 from the period during the COVID-19 pandemic, depending on the adopted number of clusters *k*.

For all objects in the study period before the COVID-19 pandemic (Figures 11–13), the optimal number of clusters *k* should be adopted as 3. The optimal number of clusters equal to 3 results from obtaining the maximum *CHIndex* (*CHIndex* = 17.38; 41.07; 44.59, respectively) and from the course of the *wss* diagram whose rate of decrease in the parameter value from the same point evidently decreases. In the case of apartment building 1 (Figure 11), a second evident decrease in the *wss* parameter value is observed for a higher

number of clusters equal to 13. For the optimal number of clusters equal to 13, however, the $CHIndex$ value is not the highest value in the analysed range from 2 to 20 clusters.

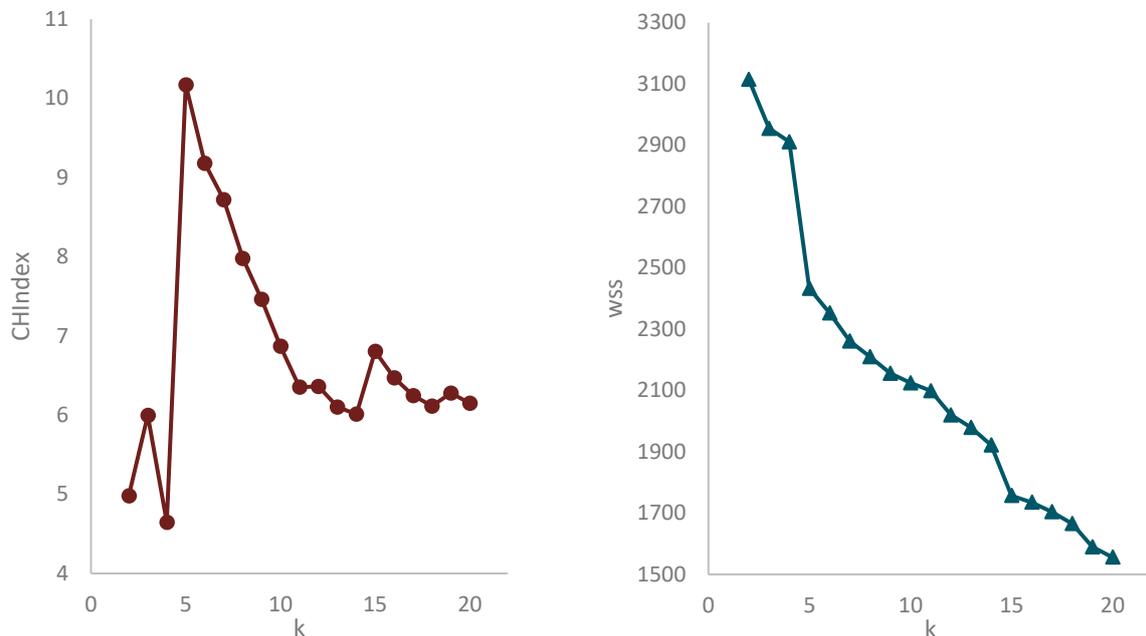


Figure 16. Values of the $CHIndex$ and total within sum of squares (wss) and for a set of 188 days of normalised histograms of hourly flows for apartment building 3 from the period during the COVID-19 pandemic, depending on the adopted number of clusters k .

For objects in the period during the COVID-19 pandemic (Figures 14–16), the optimal number of clusters k is different depending on the analysed building. For buildings 1 and 2, the optimal number of clusters k should be 7, and for apartment building 3, it should be 5. This again results from obtaining the maximum value of the $CHIndex$ and course of the wss diagram whose rate of decrease in the parameter value from the same point evidently decreases.

It should be emphasised that in the case of this study period, the $CHIndex$ did not reach a very high value:

- $CHIndex = 8.03$ for apartment building 1,
- $CHIndex = 7.83$ for apartment building 2,
- $CHIndex = 10.17$ for apartment building 3.

This means that particular clusters do not considerably differ from each other, and the elements of the set grouped in particular clusters are mutually similar (relatively weakly variable).

After the determination of the optimal number of clusters for each set of normalised histograms of accumulated water consumption in 3 objects with consideration of the study period (Table 3), k-means clustering was performed. The results and their analysis are presented in the following section.

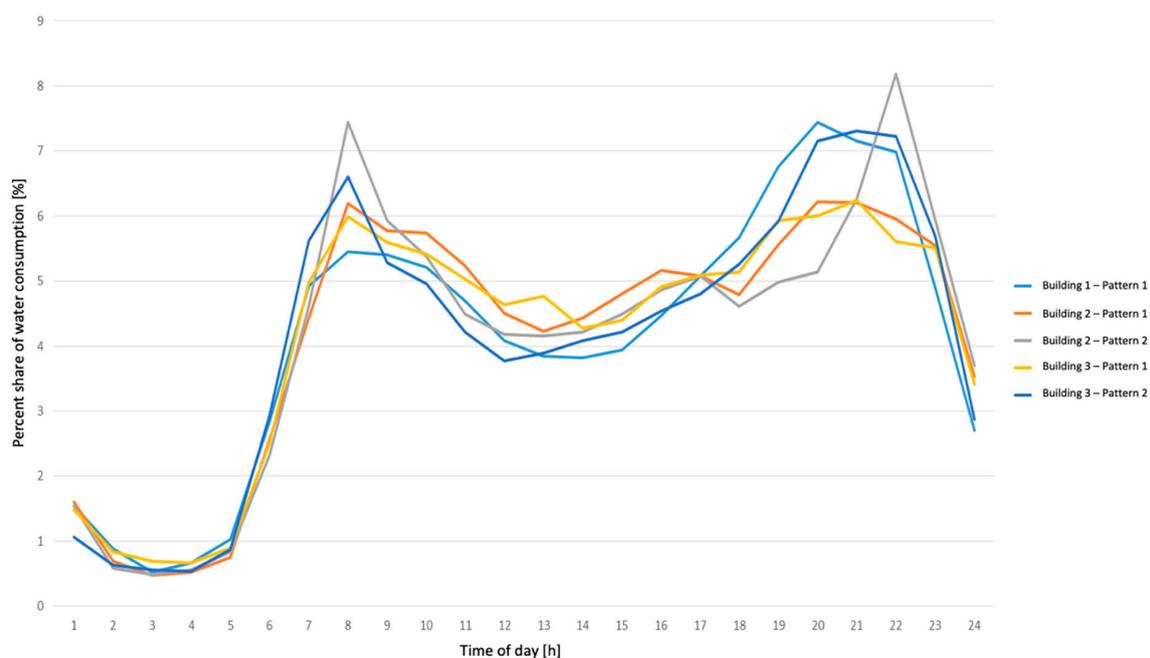
Table 3. The adopted optimal number of clusters for particular sets.

Type of Study Object	Apartment Building 1		Apartment Building 2		Apartment Building 3	
	Before COVID-19	COVID-19	Before COVID-19	COVID-19	Before COVID-19	COVID-19
Optimal number of patterns	3	7	3	7	3	5



Table 5. Numbers of days and their type assigned to particular averaged histograms of accumulated water consumption (consumption patterns) in the period before the COVID-19 pandemic.

Study Object	Consumption Pattern	Total Number of Days	Percent Share	Type of Days			
				Business Days		Days Free from Work	
				No.	%	No.	%
Apartment building 1	1	157	56.9%	154	98.09%	3	1.91%
	2	10	3.6%	7	70.00%	3	30.00%
	3	109	39.5%	35	32.11%	74	67.89%
	TOTAL:	276	100.0%	196	71.00%	80	29.00%
Apartment building 2	1	101	37.1%	99	98.02%	2	1.98%
	2	81	29.8%	81	100.00%	0	0.00%
	3	90	33.1%	13	14.44%	77	85.56%
	TOTAL:	272	100.0%	193	70.96%	79	29.04%
Apartment building 3	1	82	31.2%	76	92.68%	6	7.32%
	2	98	37.3%	98	100.00%	0	0.00%
	3	83	31.6%	14	16.87%	69	83.13%
	TOTAL:	263	100.0%	188	71.48%	75	28.52%

**Figure 17.** Averaged histograms of accumulated water consumption in buildings for business days in the study period before the COVID-19 pandemic.

Due to the characteristics of their course, histograms generated for business days (Figure 17) are distributions of diurnal water demand typical of that type of day. They reach two maximums: first in the morning hours (7:00–9:00) when people prepare for work or school, and second, which is considerably greater in the evening hours, when residents spend time preparing meals and bathing (19:00–22:00).

In the case of days free from work (Figure 18), averaged histograms of water consumption have a somewhat different course in comparison to business days. The first difference is a shift of maximum water consumption in the morning hours, starting only at 9:00 and lasting until 12:00. This suggests that recipients rest longer on these days. Moreover, before noon, maximum water consumption is also considerably greater than in the evening.

This suggests that water is consumed not only for daily hygiene activities but also for the purposes of, e.g., cleaning or preparing meals [19].

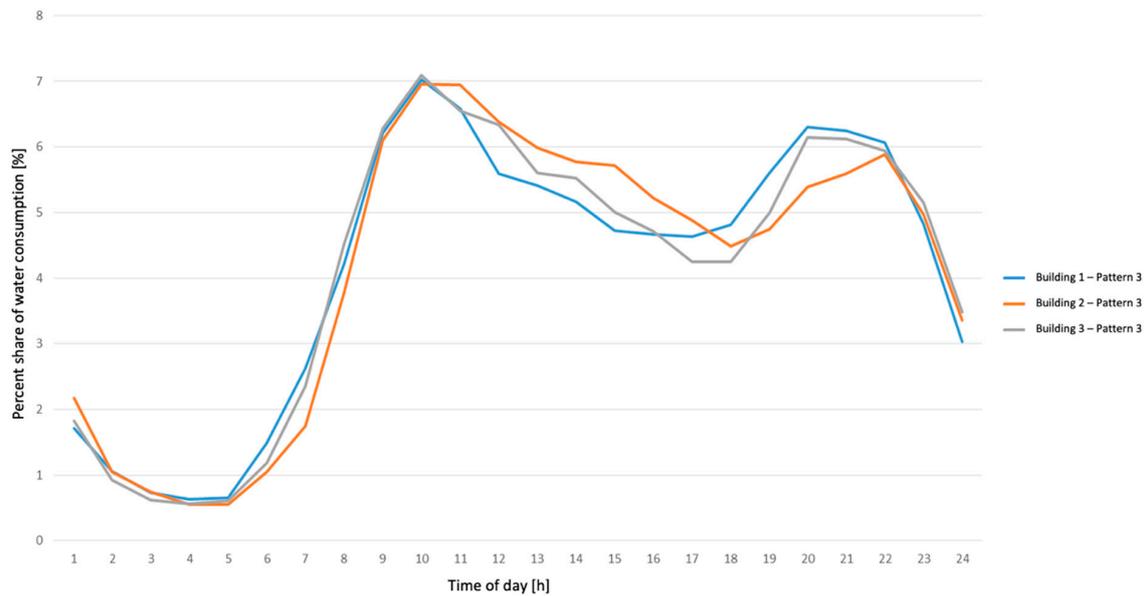


Figure 18. Averaged histograms of accumulated water consumption in buildings for days free from work in the study period before the COVID-19 pandemic.

As was mentioned above, pattern No. 2 for building 1 showed strongly untypical diagrams of diurnal water consumption by recipients. This is also confirmed by the course of the averaged histogram for 10 days classified for that cluster (Figure 19). Due to its characteristics, i.e., maximum water demand around noon (between 9:00 and 14:00), the pattern could be classified as days free from work. According to Table 5, however, the cluster primarily covers business days (70%). Moreover, due to the low number of days in the cluster and low value of the bootmean parameter (0.2810), it cannot be considered representative for analyses, e.g., in mathematical modelling of water supply networks. These histograms should be interpreted individually.

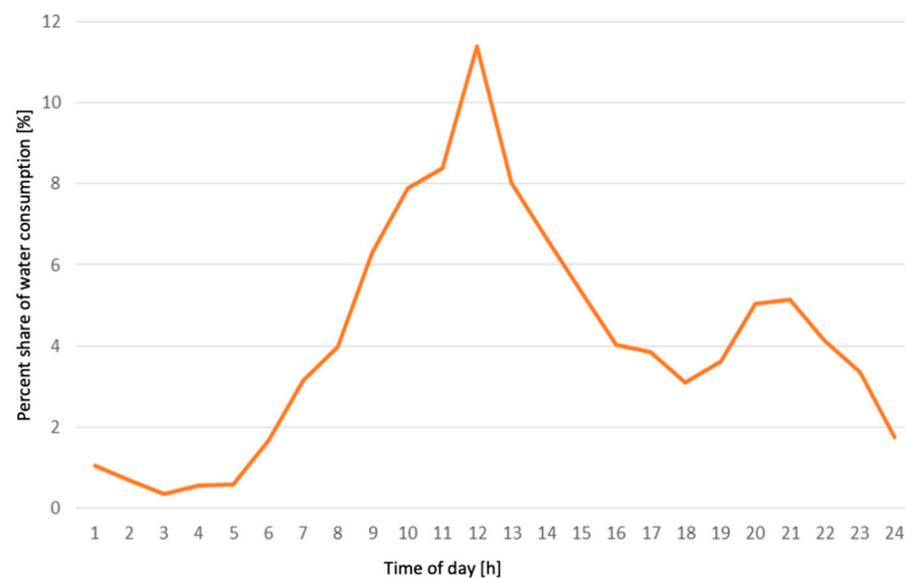


Figure 19. Averaged histogram of accumulated water consumption—pattern 2 for building 1 in the study period before the COVID-19 pandemic.

In the case of research conducted after the announcement of the state of epidemic emergency, 7 synthetic patterns of water demand were generated for buildings 1 and 2, and 5 patterns for building 3. The entire data set was divided between particular clusters in a practically even way. No cluster includes a large majority of classified days (Table 6). Moreover, bootmean parameters for the designated clusters reached the threshold of a value approximate to or higher than 0.6 only in single cases—for pattern No. 2 in building 1 and for pattern No. 1 and 3 in building 3 (Table 4). This shows a high level of uniqueness in the way of water consumption by recipients.

Table 6. Number of days and their type assigned to particular averaged histograms of accumulated water consumption (demand patterns) in the period during the COVID-19 pandemic.

Study Object	Demand Pattern	Total Number of Days	Percent Share	Type of Days			
				Business		Free from Work	
				Number	%	Number	%
Apartment building 1	1	25	13.6%	25	100.00%	0	0.00%
	2	25	13.6%	24	96.00%	1	4.00%
	3	28	15.2%	6	21.43%	22	78.57%
	4	36	19.6%	35	97.22%	1	2.78%
	5	30	16.3%	19	63.33%	11	36.67%
	6	32	17.4%	20	62.50%	12	37.50%
	7	8	4.3%	2	25.00%	6	75.00%
	TOTAL:	184	100.0%	131	71.20%	53	28.80%
Apartment building 2	1	19	10.4%	14	73.68%	5	26.32%
	2	36	19.8%	36	100.00%	0	0.00%
	3	27	14.8%	14	51.85%	13	48.15%
	4	20	11.0%	1	5.00%	19	95.00%
	5	32	17.6%	31	96.88%	1	3.13%
	6	30	16.5%	30	100.00%	0	0.00%
	7	18	9.9%	2	11.11%	16	88.89%
	TOTAL:	182	100.0%	128	70.33%	54	29.67%
Apartment building 3	1	43	22.9%	4	9.30%	39	90.70%
	2	51	27.1%	49	96.08%	2	3.92%
	3	50	26.6%	50	100.00%	0	0.00%
	4	41	21.8%	28	68.29%	13	31.71%
	5	3	1.6%	3	100.00%	0	0.00%
	TOTAL:	188	100%	134	71.28%	54	28.72%

Wanting to verify the results of classification in terms of type of days assigned to particular clusters, however, in the majority of cases, based on the number of particular days in those sets, we are not able to assign them to a specific type of day (Table 6). For example, pattern No. 5 for building 1 covers 19 business days (63.33%) and 11 days free from work (36.67%). Similar situations are observed in patterns: No. 6 for building 1; No. 1 and 3 for building 2; and No. 4 for building 3. The remaining patterns show a considerable dominance of one type of day in their set.

It was also observed that some clusters cover a number of assigned days low in comparison to the entire set. They are among others: pattern No. 7 for building 1, including 8 days (which constitutes 4.3%), and pattern No. 5 for building 3, including only 3 days (1.6%).

Presentation of the obtained synthetic diagrams of water demand during the COVID-19 pandemic in the buildings was attempted in Figures 20 and 21 with consideration of the partitioning into patterns for business days and days free from work. The obtained patterns showed different courses. Moreover, they differ from those before the pandemic among others in higher water consumption during the day and by night, between 2:00 and 4:00.

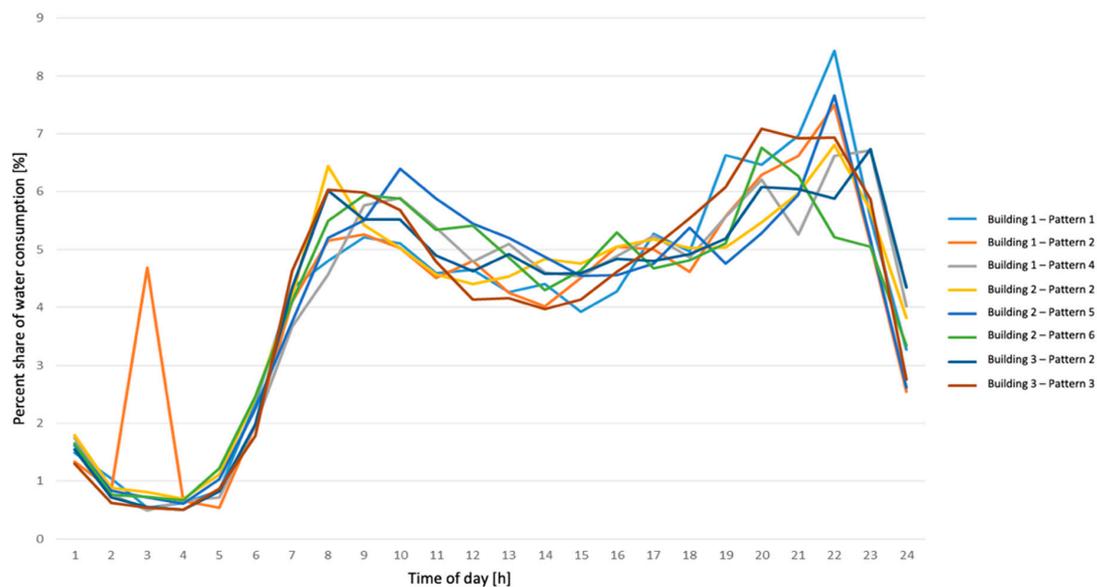


Figure 20. Averaged histograms of accumulated water consumption in buildings for business days during the COVID-19 pandemic.

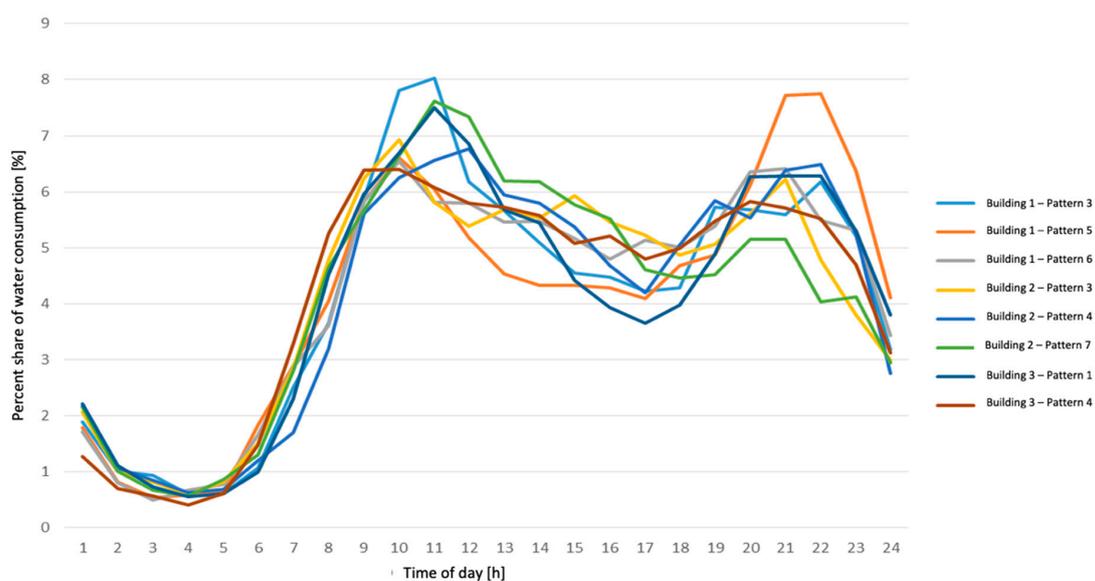


Figure 21. Averaged histograms of accumulated water consumption in buildings for days free from work during the COVID-19 pandemic.

Due to reaching two maximums (in the morning and evening), histograms generated for business days (Figure 20) constitute specific typical distributions of diurnal water demand. Only water consumption during the day is untypical. It is probably the effect of introduction of remote work in the majority of companies and remote learning at schools.

For days free from work (Figure 21), averaged histograms of water consumption are somewhat more approximate to those generated for the period before the pandemic, i.e., with a shifted maximum in water consumption in the morning to 9:00, and water consumption in the evening considerably lower than in the morning. Each synthetic pattern, however, has its own individual course, and they show no similarities. The synthetic diagrams of water demand presented in Figures 20 and 21 therefore confirm a certain type of disturbance in water use by residents in the 3 study objects.

5. Conclusions

This paper's objective was to develop a methodology supporting clustering and generation of synthetic distributions of diurnal water demand in apartment buildings for the purposes of mathematical modelling that is increasingly frequently applied in the processes of management and dimensioning of water supply networks [5]. It particularly involves hydraulic computations with the application of the EPS simulation, i.e., for longer temporal horizons, allowing for the understanding of the hydraulics of the distribution system, tracing changes in flows in time, or designation of zones of water mixing. This requires the investigation of the temporal and spatial dynamics of water consumption in particular nodes. A currently accepted simplification is the application in the mathematical model of the top-down approach in which consumption in line with that recorded in measurement points or pumping stations is imposed. A number of publications and scientific studies show that only the bottom-up approach is appropriate. It involves a determination of nodal consumption from the level of an individual recipient. This approach, however, requires knowledge of patterns of water demand in different objects [20,21].

Research of diurnal time series of water consumption was conducted in three mutually similar apartment buildings constructed in the 1980s in the same housing estate in Bydgoszcz. The time of recording of hourly water consumption covered a total of 464 days, whereas 276 days occurred before the announcement of the state of epidemic emergency in the country, and the other 188 days were recorded during the COVID-19 pandemic [22]. The results were analysed in terms of possibilities of the application of the cluster analysis for clustering and generation of synthetic distributions of daily water consumption. It should be emphasised, however, that before performing computations with the application of the k-means method, it was necessary to eliminate outlier days from the data set. This procedure was performed based on dendrograms developed by means of the hierarchical agglomeration method.

The application of the cluster analysis—k-means clustering—permitted the development of characteristic patterns of hourly water demand with division into business days and days free from work and holidays. The division was particularly unambiguous for the period before the pandemic. In the process, three synthetic patterns of hourly water consumption were generated. Two of them were patterns typical of business days, and the third one was a pattern for days free from work with a low number of business days, predominantly close to weekends (it probably results from the so-called “long weekends” in the holiday period). The exception was pattern 2 for building 1. It covered only 10 days completely deviating from the remaining ones and simultaneously dissimilar towards one another, which should be interpreted individually. The resulting averaged histograms of water consumption for particular buildings could be used for the determination of nodal water consumption in mathematical modelling of water supply networks.

A somewhat different situation was observed in the case of recording of data during the COVID-19 pandemic. Considerably more patterns of water demand were generated, and no cluster included a large majority of recorded data. Moreover, the majority of the designated clusters were too similar to each other or included histograms with a random character, as evidenced by the bootmean parameter value. A number of untypical behaviours in terms of water consumption were also observed, for example: water consumption by night, between 2:00 and 4:00, or increased varied water consumption during the day. A change in water consumption by recipients probably results from the introduction of remote work in the majority of companies and remote learning at schools. This suggests that credible classification for this data set is impossible due to the high uniqueness of water consumption by residents. Using the data in a mathematical model of water supply networks would require partitioning the set into smaller subsets and performing the clustering again. It may turn out that the generated clusters include a single or a low number of days. There is no doubt, however, that the COVID-19 pandemic has greatly influenced the daily water demand patterns worldwide [3]. Many studies [19,23–30] show that during COVID-19 pandemic, water consumption patterns have followed other notable

trends of the new normal ‘stay at home’ life. Families are getting their day started later, with peak morning consumption shifting two hours later. In addition, comparing pre-COVID-19 to current residential usage patterns, it is clear that the highest increase in water usage is happening in the afternoon as stay at home schoolers and workers are taking a break, getting up to use the restroom, washing their hands, and prepping meals. In general, people on average are using the bathroom at home three times more each day, and they are flushing the toilet five times more per day than before the pandemic [24,27,29,30]. They are also showering almost three more times per week than they did. In addition to the increase in the number of showers, the time of day when people are showering has also changed, further reflecting a later start to their day: midday and evening showers increased while morning showers shifted later in the day [29,30].

To sum up, this paper presents the method of clustering and generating synthetic diagrams of diurnal water consumption for the purposes of mathematical modelling. It also points to the variability among the generated patterns and the way the COVID-19 pandemic affected water consumption. Therefore, hydraulic analyses with the application of mathematical models require continuous records of water consumption by recipients for the purpose of updating the synthetic patterns of water consumption.

Author Contributions: Conceptualization, P.L.; Data curation, P.D. and S.D.; Formal analysis, P.D. and S.D.; Methodology, P.D., S.D. and K.M.; Visualization, P.L.; Writing—original draft, P.D., S.D., M.R. and K.K.; Writing—review & editing, P.D., S.D., M.R., K.K., K.M. and P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy restrictions.

Acknowledgments: This work was carried out as part of the fourth edition of the program implementation doctorate realized by the Ministry of Science and Higher Education, Republic of Poland.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mensah, J. Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature re-view. *Cogent Soc. Sci.* **2019**, *5*, 1653531. [[CrossRef](#)]
2. Aksela, K.; Aksela, M. Demand Estimation with Automated Meter Reading in a Distribution Network. *J. Water Resour. Plan. Manag.* **2011**, *137*, 456–467. [[CrossRef](#)]
3. Kalbusch, A.; Henning, E.; Brikalski, M.P.; de Luca, F.V.; Konrath, A.C. Impact of coronavirus (COVID-19) spread-prevention actions on urban water consumption. *Resour. Conserv. Recycl.* **2020**, *163*, 105098. [[CrossRef](#)] [[PubMed](#)]
4. Abu-Bakar, H.; Williams, L.; Hallett, S.H. Quantifying the impact of the COVID-19 lockdown on household water consumption patterns in England. *NPJ Clean Water* **2021**, *4*, 13. [[CrossRef](#)]
5. Aisopou, A.; Stoianov, I.; Graham, N.J.D. In-pipe water quality monitoring in water supply systems under steady and un-steady state flow conditions: A quantitative assessment. *Water Res.* **2012**, *46*, 235–246. [[CrossRef](#)] [[PubMed](#)]
6. Obradović, D.; Lonsdale, P. *Public Water Supply: Models, Data and Operational Management*; CRC Press: London, UK, 1998.
7. Blokker, E.J.M. *Stochastic Water Demand Modelling for a Better Understanding of Hydraulics in Water Distribution Networks*; Water Management Academic Press: Delft, The Netherlands, 2010.
8. Buchberger, S.G.; Wells, G.J. Intensity, Duration and Frequency of Residential Water Demands. *J. Water Resour. Plan. Manag.* **1996**, *122*, 11–19. [[CrossRef](#)]
9. Buchberger, S.G.; Carter, J.T.; Lee, Y.H.; Schade, T.G. *Random Demands, Travel Times and Water Quality in Dead End*; Report No. 294; American Water Works Association Research Foundation: Denver, CO, USA, 2003; pp. 1–470.
10. Garcia, V.J.; Garcia-Bartual, R.; Cabrera, E.; Arregui, F.; Garcia-Serra, J. Stochastic Model to Evaluate Residential Water Demands. *J. Water Resour. Plan. Manag.* **2004**, *130*, 386–394. [[CrossRef](#)]
11. Alcocer-Yamanaka, V.; Tzatchkov, V.; Wu, Z.Y. Demand forecasting, leakage and energy management. In *Integrating Water Systems*; Boxall, J., Maksimovic, C., Eds.; Spectral Analysis of Instantaneous Residential Water Demand Time Series; CRC Press: London, UK, 2010; pp. 503–508.

12. Alvisi, S.; Ansaloni, N.; Franchini, M. Generation of synthetic water demand time series at different temporal and spatial aggregation levels. *Urban Water J.* **2013**, *11*, 297–310. [[CrossRef](#)]
13. Stanisz, A. *Przystępny Kurs Statystyki z Zastosowaniem STATISTICA PL na Przykładach Medycyny. Tom I-III*; StatSoft Polska Sp. z o.o.: Kraków, Poland, 2007.
14. Hartigan, J.A. *Clustering Algorithms*; Wiley: New York, NY, USA, 1975.
15. Larose, D.T. *Discovering Knowledge in Data. An Introduction to Data Mining*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005.
16. Walesiak, M.; Gatnar, E. *Analiza Skupień. Statystyczna Analiza Danych z Wykorzystaniem Programu R*; Wydawnictwo Naukowe PWN: Warszawa, Poland, 2012.
17. Local Data Bank. Statistics Poland. Data for Territorial Unit 0461000. Available online: <https://bdl.stat.gov.pl/BDL/dane/teryt/jednostka> (accessed on 21 June 2020).
18. Cichoń, T.; Królikowska, J.; Nachlik, E. *Gospodarka Wodociągowa. Wyzwania i Możliwości*; Politechnika Krakowska: Kraków, Poland, 2020.
19. Renukappa, S.; Kamunda, A.; Suresh, S. Impact of COVID-19 on water sector projects and practices. *Util. Policy* **2021**, *70*, 101194. [[CrossRef](#)]
20. Blokker, E.J.M.; Vreeburg, J.H.G.; Van Dijk, J.C. Simulating Residential Water Demand with a Stochastic End-Use Model. *J. Water Resour. Plan. Manag.* **2010**, *136*, 19–26. [[CrossRef](#)]
21. Blokker, E.J.M.; Beverloo, H.; Vogelaar, A.J.; Vreeburg, J.H.G.; Van Dijk, J.C.; Blokker, M. A bottom-up approach of stochastic demand allocation in a hydraulic network model: A sensitivity study of model parameters. *J. Hydroinform.* **2011**, *13*, 714–728. [[CrossRef](#)]
22. Luczkiewicz, A.; Pierpaoli, M.; Kosek, K.; Szopińska, M.; Remiszewska-Skwarek, A.; Kotlarska, E.; Szatkowska, B.; Fudala-Książek, S. Ocena i predykcja dynamiki epidemii COVID-19 na podstawie obecności wirusa SARS-CoV-2 w ściekach. *Wodociągi Polskie* **2021**, *15*, 44–49.
23. Rizvi, S.; Rustum, R.; Deepak, M.; Wright, G.B.; Scott, A. Identifying and analyzing residential water demand profile; including the impact of COVID-19 and month of Ramadan, for selected developments in Dubai, United Arab Emirates. *Water Supply* **2021**, *16*, 13.
24. Bich-Ngoc, N.; Teller, J. Potential Effects of the COVID-19 Pandemic through Changes in Outbound Tourism on Water Demand: The Case of Liège (Belgium). *Water* **2020**, *12*, 2820. [[CrossRef](#)]
25. Cole, G.; Stewart, R.A. Smart meter enabled disaggregation of urban peak water demand: Precursor to effective urban water planning. *Urban Water J.* **2013**, *10*, 174–194. [[CrossRef](#)]
26. Antwi, S.H.; Getty, D.; Linnane, S.; Rolston, A. COVID-19 water sector responses in Europe: A scoping review of preliminary governmental interventions. *Sci. Total Environ.* **2021**, *762*, 143068. [[CrossRef](#)] [[PubMed](#)]
27. Zvobgo, L.; Do, P. COVID-19 and the call for ‘Safe Hands’: Challenges facing the under-resourced municipalities that lack potable water access—A case study of Chitungwiza municipality, Zimbabwe. *Water Res. X* **2020**, *9*, 100074. [[CrossRef](#)] [[PubMed](#)]
28. Ji, B.; Zhao, Y.; Wei, T.; Kang, P. Water science under the global epidemic of COVID-19: Bibliometric tracking on COVID-19 publication and further research needs. *J. Environ. Chem. Eng.* **2021**, *9*, 105357. [[CrossRef](#)] [[PubMed](#)]
29. Sayeed, A.; Rahman, H.; Bundschuh, J.; Herath, I.; Ahmed, F.; Bhattacharya, P.; Tariq, M.R.; Rahman, F.; Joy, T.I.; Abid, M.T.; et al. Handwashing with soap: A concern for overuse of water amidst the COVID-19 pandemic in Bangladesh. *Groundw. Sustain. Dev.* **2021**, *13*, 100561. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, N.; You, F. Food-energy-water-waste nexus systems optimization for New York State under the COVID-19 pandemic to alleviate health and environmental concerns. *Appl. Energy* **2021**, *282*, 116181. [[CrossRef](#)]