smart cities

MDPI

Article

Geospatial Clustering in Smart City Resource Management: An Initial Step in the Optimisation of Complex Technical Supply Systems

Aliaksey A. Kapanski [1,*], Roman V. Klyuev [2], Aleksandr E. Boltrushevich [3], Svetlana N. Sorokova [4], Egor A. Efremenkov [4], Anton Y. Demin [3] and Nikita V. Martyushev [3,*]

1 Department of Power Supply, Sukhoi State Technical University of Gomel, 246746 Gomel, Belarus
2 Department of Automation and Control, Moscow Polytechnic University, 107023 Moscow, Russia; kluev-roman@rambler.ru
3 Department of Information Technologies, Tomsk Polytechnic University, 634050 Tomsk, Russia; aeb20@tpu.ru (A.E.B.); ad@tpu.ru (A.Y.D.)
4 Department of Mechanical Engineering, Tomsk Polytechnic University, 634050 Tomsk, Russia; s_sorokova@tpu.ru (S.N.S.); egorefr@tpu.ru (E.A.E.)
* Correspondence: kapanski@mail.ru (A.A.K.); martjushev@tpu.ru (N.V.M.)

Highlights:

What are the main findings?

- A method for clustering urban water supply consumers based on their geographic location and water consumption levels has been developed, allowing for the identification of key areas with high water demand and consideration of spatial distribution characteristics on the map.
- The effectiveness of applying the agglomerative clustering method has been demonstrated, which, combined with the silhouette coefficient, allows for selecting the optimal data partition strategy and distributing the territorial division of consumers across service zones.

What are the implications of the main findings?

- The proposed geospatial clustering method allows for planning the placement of pressure sensors, taking into account the distance of consumers from water sources, consumption centers, and their cluster membership, which creates a foundation for solving hydraulic modeling tasks.
- The results of the study are aimed at addressing the task of determining the magnitude and influence zones of pressures generated by water sources on the key points of the urban water supply system, in order to optimize the operation modes of pump stations.

Abstract: For large cities with developing infrastructures, optimising water supply systems plays a crucial role. However, without a clear understanding of the network structure and water consumption patterns, addressing these challenges becomes significantly more complex. This paper proposes a methodology for geospatial data analysis aimed at solving two key tasks. The first is the delineation of service zones for infrastructure objects to enhance system manageability. The second involves the development of an approach for the optimal placement of devices to collect and transmit hydraulic network parameters, ensuring their alignment with both water supply sources and serviced areas. The study focuses on data from the water supply network of a city with a population exceeding half a million people, where hierarchical clustering using Ward's method was applied to analyse territorial distribution. Four territorial clusters were identified, each characterised

check for updates

Academic Editor: Pierluigi Siano

Received: 28 November 2024
Revised: 10 January 2025
Accepted: 14 January 2025
Published: 21 January 2025

Citation: Kapanski, A.A.; Klyuev, R.V.; Boltrushevich, A.E.; Sorokova, S.N.; Efremenkov, E.A.; Demin, A.Y.; Martyushev, N.V. Geospatial Clustering in Smart City Resource Management: An Initial Step in the Optimisation of Complex Technical Supply Systems. *Smart Cities* **2025**, *8*, 14. https://doi.org/10.3390/smartcities8010014

Copyright: © 2025 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/).

by unique attributes reflecting consumer concentration and water consumption volumes. The cluster boundaries were compared with the existing service scheme of the system, confirming their alignment with real infrastructure. The quality of clustering was further evaluated using the silhouette coefficient, which validated the high accuracy and reliability of the chosen approach. The paper demonstrates the effectiveness of cluster boundary visualisation for assessing the uniform distribution of pressure sensors within the urban water supply network. The results of the study show that integrating geographic data with water consumption information not only facilitates effective infrastructure planning and resource allocation but also lays the foundation for the digitalization of the hydraulic network, a critical component of sustainable development in modern smart cities.

Keywords: spatial clustering; optimisation; geographic segmentation; Ward's method; water resource management; silhouette quotient; infrastructure planning

1. Introduction

In today's context of increased demands for reliability and efficiency in resource supply systems, the implementation of advanced technologies such as the Internet of Things (IoT), machine learning, and artificial intelligence has become increasingly significant. These technologies enable the collection, processing, and analysis of real-time data, opening up new opportunities to optimise the operation of complex technical systems, including urban water supply systems. Currently, water supply systems represent a complex array of devices and infrastructure that ensure the uninterrupted extraction, delivery, and distribution of drinking water across vast territories. In large cities, these systems are multi-tiered, involving different stages of water lifting to maintain optimal pressure in various sections of the water supply network. Scientific studies indicate that these systems operate under the influence of numerous external and internal factors, which can significantly affect their efficiency and reliability [1–3]. Key challenges include optimising water resource management, reducing leakage, and minimising operational failures in technological processes [4,5].

Thus, the efficiency of water supply systems depends on numerous factors, including configuring optimal operating modes, promptly responding to emergencies, and ensuring the quality maintenance of pipelines and equipment. Achieving the stable operation of such systems requires the continuous monitoring of their condition, a detailed analysis of operational characteristics, and timely managerial decision-making. However, existing approaches to managing water supply systems often face several limitations that hinder a comprehensive understanding of the structure and functioning of the network. In particular, there is no effective way to accurately determine the concentration of key consumers on a territorial map, which creates significant challenges when conducting energy audits. This makes it difficult to analyse the distribution of objects relative to water intakes and evaluate their impact on the water supply network. Such limitations complicate not only monitoring but also the search for ways to improve the efficiency of system operations.

Most modern water supply systems have a significant amount of data, offering opportunities to develop a mechanism for quickly identifying key consumer sectors based on essential factors. This study focuses on two such factors: geospatial data (longitude and latitude) and water consumption volumes. These parameters enable the structuring of water supply system objects and their analysis in a territorial context. It is worth noting that factors such as population density and types of industrial facilities remain open topics requiring further investigation. Accounting for these characteristics could significantly expand the analysis capabilities and provide a deeper understanding of water consumption patterns using the methodology proposed in this research.

This paper examines the water supply system of a city with a population exceeding half a million people, using Gomel (Republic of Belarus) as an example. The goal of the study is to develop a methodology for clustering the territorial location of urban water supply system objects, aimed at solving the following tasks:

1. Developing a methodology for dynamic spatial analysis to delineate territorial cluster boundaries based on water consumption data and the geographic location of objects. This will account for changes in a developing infrastructure by leveraging a continuous flow of information to plan service zones for water supply networks and enhance their manageability.
2. Designing a strategy for the optimal placement of sensors (pressure, flow rate, etc.) in the water supply network, aligning their placement with territorial clusters and ensuring the visualisation of their uniform distribution on a map. This will enable effective data collection on the hydraulic parameters of the system—from water supply sources to control points within the urban infrastructure. The collected data will serve as the foundation for creating hydraulic models necessary for optimising system performance and analysing the zones of influence of water intakes on various sections of the water supply network.

The expected outcomes of the study include the development of a methodology that will not only improve the manageability and monitoring of the existing water supply system but also lay the groundwork for creating hydraulic models (digital twins) in the context of smart city development. The results of the research represent an initial yet significant step toward modelling energy-efficient modes of water supply networks and optimising pressure schedules for pumping stations. Additionally, the findings can be used for planning infrastructure modernization and developing resource management strategies in cities with similar water supply structures.

2. Related Works

2.1. Clustering Algorithms

A review of global practises highlights hierarchical clustering and the k-means algorithm as the most widely used clustering methods. An example of their application is the study by Poncelet et al. (2017), which demonstrates the use of these methods for identifying representative historical days in energy consumption and generation optimisation models [6,7]. The authors proposed new criteria and metrics for evaluating model representativeness and developed an optimisation method that more accurately models annual variations in demand and energy production. Another example is the publication by Arefifar et al., where similar approaches were applied to create virtual microgrids in smart distribution systems [8]. This improved energy system management by accounting for probabilistic characteristics of generation and consumption, optimising power distribution among consumers [9].

The combination of these clustering methods with geographic data creates a powerful synergy, enabling the effective resolution of a wide range of tasks. For instance, in the study by Zhang et al., spatial clustering methods were used to analyse the energy efficiency levels of various regions in China [10]. The research identified a trend toward grouping consumers by their energy consumption levels. To gain deeper insights into the factors affecting energy efficiency, such as industrial specialisation and economic development levels, a spatial quantile autoregression model was employed [11]. This approach not only identified key factors but also provided a foundation for developing targeted strategies to improve energy efficiency.

The application of clustering methods is also actively developing in tasks related to water supply system reliability. An example is the study by Abokifa et al., which focuses on the use of spatial data analysis combined with the predictive modelling of pipeline failures [12]. The methodology proposed by the authors included the segmentation of the pipeline network based on geographic location and pipe characteristics, allowing for the identification of areas with a high risk of failure. The use of machine learning methods to predict failure probabilities in each cluster enabled more accurate accident forecasting and planning of preventive measures [13].

As the literature review demonstrates, territorial clustering has proven effective in solving various tasks related to resource system management. It not only facilitates optimal resource allocation and improves system manageability but also provides a foundation for integration with advanced artificial intelligence and machine learning technologies.

2.2. Water Resource Management in Smart Cities

Global trends in the development of water supply systems rely on the concept of smart cities and focus on implementing advanced technologies that enable the continuous collection and processing of large volumes of data using artificial intelligence (AI) and machine learning (ML) to solve highly specialised tasks [14–16]. Until recently, the high cost of establishing connectivity among numerous water supply facilities spread across vast urban areas limited the adoption of modern solutions. However, global advancements in specialised communication standards for telemetry devices with low data transmission volumes have opened new opportunities. With the development of Industry 4.0 projects and smart city initiatives, obtaining information about the state of hydraulic networks has become significantly easier, which has profoundly impacted the improvement of many processes, including urban water supply. Internet of Things (IoT) technologies have not only made data collection and analysis more accessible but also created broad opportunities for exploring new ways to improve the efficiency of natural and energy resources [17–19].

Issues related to pipeline system management remain particularly relevant, where AI models are capable of effectively addressing tasks such as pressure prediction at control points, optimising operations, forecasting failures, and detecting leaks. In this process, information resources play a key role by providing accurate and up-to-date data necessary for model testing and training. An illustrative example is the study by Aljameel, Alomari et al. [20], which explored the application of machine learning methods in pipeline systems. The publication examined the effectiveness of recurrent neural networks (RNNs) and the k-nearest neighbours (KNN) method for leak detection. A comparative analysis of various ML models demonstrated the superiority of RNNs for such tasks, owing to their ability to self-learn and process large datasets, resulting in high prediction accuracy in dynamic informational environments.

Another study [21] presented the application of machine learning methods for modelling pipeline failure scenarios under increased pressure. Predictive algorithms based on artificial neural networks (ANN), gradient boosting (XGBoost, XGB), and categorical boosting (CatBoost, CAT) were used. These models were trained on statistics from actual pipeline inspections and results from finite element methods used for burst predictions. Among the evaluated models, the CatBoost algorithm provided the most accurate predictions with minimal errors. This highlights the effectiveness of combining AI models with classical methods, making them a powerful tool for assessing pipeline safety, particularly under changing operational conditions requiring adjustments to technological parameters.

Continuing the discussion on the application of machine learning in water supply systems, the study by Sourabh N., Timbadiya P.V., and Patel P.L. [22], published in the ISH Journal of Hydraulic Engineering in 2023, unveils new research possibilities. The publication focuses on water leak control and describes the application of reverse engineering approaches for detection. The main idea of the study was to use classification and regression tasks with artificial neural networks and support vector machines (SVM) to identify pressure or flow deviations caused by leaks. The study was conducted in two scenarios: the first involved measurements only for pressure and the second for water flow in the system. The authors developed and trained multilayer perceptron (MLP) models and multi-level classification and regression SVMs. It was established that ANN models showed superior results compared to SVMs in detecting leaks across both scenarios. The publication highlights that model performance can be further enhanced by optimising the number of input data points during training, which remains an important area of research.

In this context, it is worth mentioning the study conducted by Momeni A. and Piratla K.R. (2022) [23], which proposed using a simplified hydraulic model of water supply networks that considers the condition of pipelines based on monitoring actual flow and pressure values combined with machine learning methods. This approach, demonstrated on a benchmark water distribution network, proved effective. The study illustrates how modern machine learning algorithms can enhance the analysis of hydraulic characteristics and provide continuous real-time monitoring of emergency scenarios [24,25].

To achieve maximum efficiency in using ML models, it is crucial to obtain reliable hydraulic network information that accurately describes the behaviour of the water supply system. This necessitates a clear understanding of optimal zones for placing data-collection devices to monitor pressure, flow, and other critical parameters across urban infrastructure. Understanding how water consumption objects form territorial clusters enables the localization and minimization of sensors while ensuring comprehensive geographical coverage for information systems. The justified placement of monitoring devices in complex and extensive water supply networks is vital for accurate data collection [26–28]. For this purpose, it is necessary to develop a territorial clustering methodology that segments objects based on their geographic location and water consumption volumes, significantly simplifying data analysis. Clustering not only facilitates optimal sensor placement but also provides a deeper understanding of water consumption patterns across different city areas, identifying high-demand zones and key monitoring areas. In the context of expanding infrastructure, this approach becomes essential for improving system manageability and control, as it helps visualise data distribution density on geographic maps. This, in turn, contributes to more precise adjustments of pumping station operating modes and optimisation of pressure in the water supply network.

3. Methods

The study was structured in several successive phases. The first stage involved a preliminary analysis of urban infrastructure and water supply system operation, focusing on the geographical location of water intakes, water production volumes, and system pressure levels. The second stage involved the collection and preparation of data for the analysis, including statistics on water supply by abstraction and georeferenced water consumption data for major urban sites. Intermediate stages covered digitisation, processing, and the integration of data into a unified information system. At the final stage, water-consuming objects were grouped into clusters by geographical features using the hierarchical clustering method, forming segments with different locations and water consumption of objects. This made it possible to identify territorial groups where the concentration of water consumers had a specific geographical structure and to identify key clusters with different loads on the water supply infrastructure.

3.1. Sources of Statistical Information and Methods of Data Processing

The Gomel water supply system has several key water intakes, each of which plays an important role in providing water supply to both the population and industrial enterprises. The unique characteristics and peculiarities of each intake affect the overall efficiency of the system and require a thorough preliminary infrastructure analysis. For this analysis, statistical data compiled from hourly water supply values to the City's network, obtained from the engine room logs of the second lift pump stations for the year 2023, were used. These data were digitised to further display the loading intensity on a map, allowing the system performance to be assessed in the context of the geographical location of the water intakes. In addition to these data, pump station-operating modes were investigated by recording the output pressure in the discharge pipework on a minute-by-minute basis. This allowed a more accurate assessment of the relationship between water consumption and pressure schedules of the intakes as a function of time intervals and their spatial location. Daily statistics of water supply to the city for a long period from 2017 to 2023 were also used to analyse long-term trends.

The analysis of the water supply load distribution covered more than 250 of the largest industrial facilities with reference to their geographical location. These facilities were selected on the basis of their significant contribution to the water consumption of the city, forming 66% of the balance. Since automated metering of water consumption was not available for most of these facilities, a combined approach was used, combining data from customer service departments and digital metering devices. The data provided by the customer departments included the legal addresses of customers and consumption volumes for certain periods. The Yandex. Maps geoservice was used to translate these addresses into precise geographical coordinates.

This study used a variety of tools and libraries for data collection, processing, and visualisation. The Jupyter Notebook platform was chosen as the main analysis environment, providing the possibility of step-by-step code execution and visualisation of intermediate results. Data processing was performed using Pandas and NumPy libraries. These tools allowed us to perform the tasks of data loading, cleaning, aggregation, and normalisation for further analysis. The visualisation of temporal and spatial characteristics of water consumption and pressure at pumping stations was performed using Matplotlib and Seaborn libraries, which allowed for the visualisation of daily and minute-by-minute pressure fluctuations, as well as their georeferencing. Algorithms implemented in the scikit-learn library were used for the spatial clustering of wo-supply objects. The ConvexHull tool from the SciPy library was used to create a convex surface showing the boundaries of geographical cluster zones.

3.2. Preliminary Overview of the Water Supply System

The structure of the studied water supply system includes five main water intakes: Sozh, Iput, Korenevsky, Southwest, and Central, which supply water resources to both the population and industry of the city. The water supply sources operate in a common hydraulic system, maintaining established technological water parameters for consumers. Each of the water intakes is located in different parts of the City and has unique characteristics in terms of both capacity and pressure levels. Figure 1 shows the dynamics of water delivery volumes by day for each of the intakes in the study system. To construct the time diagram, the data were smoothed using the moving average method, which allowed for the reduction of the influence of short-term fluctuations and better reflects the main trends in the dynamics of water supply during visual analysis.

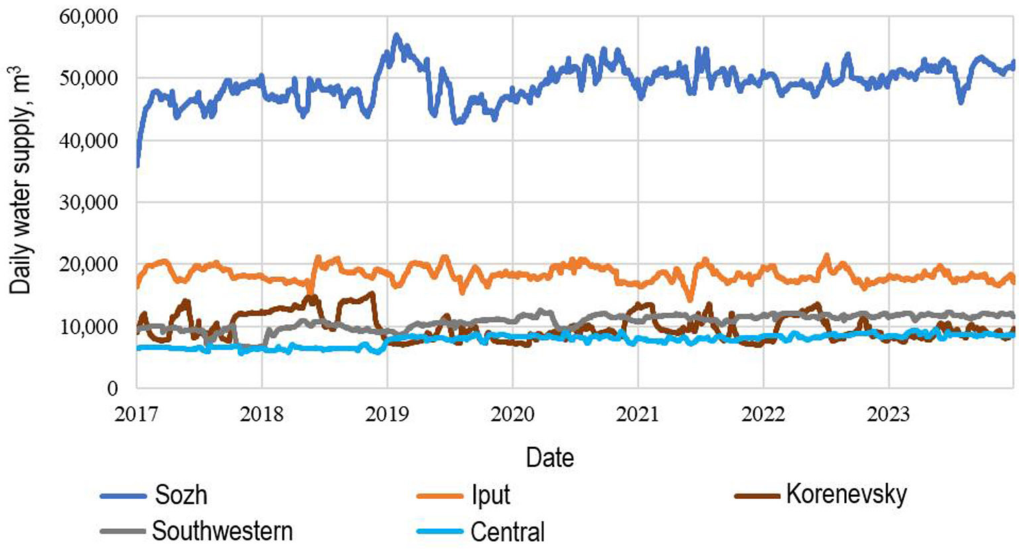

Figure 1. Smoothed data of daily water supply by water intakes for 2023 using the moving average method.

On average, the Sozh intake shows the highest average values of daily water supply (49,192 m^3/day). The coefficient of variation in water consumption is 7.4%, indicating relatively low variability of the data. The Iput, Southwest and Central intakes also contribute significantly to the total water supply with average daily supply values of 18,358, 10,713 and 7764 m^3/d respectively. The coefficients of variation for these intakes were 8.8, 13.4 and 13.3%, respectively, which also indicates relatively low variability in the city's water supply. In what sense, these intakes can be considered as sources providing the basis for the daily supply of water to different parts of the city. The remaining water intake 'Korenevsky' with an average volume of 9820 m^3/day has a high coefficient of variation (25.0%), which determines the role of this intake in smoothing peak loads, adjusting to the daily demands of consumers in the southern part of the city.

To visualise the geographical location of the intakes, the OpenStreetMap platform was used, which made it possible to display the location of each intake on a map of the city (Figure 2). Using the 3D Maps tool in Excel, an interactive map was created that shows the intensity of water supply by each intake. This framework was then used to georeference water consumers, providing a visual representation of the distribution of load on the water supply system using a heat map. Using these tools, it became possible not only to analyse the territorial distribution of water intakes, but also to effectively integrate their data with other information systems for the further analysis and clustering of water consumption by geographical zones. Thus, as a result of the visual analysis, the distance of water intakes from the centre of water consumption and their actual load have been established. For example, the actual capacity of the 'Sozh' intake, located in the north-eastern part of the city, was significantly higher than other water supply sources. However, due to its remoteness from the main water consumption facilities, its utilisation averaged slightly more than 50% of the design capacity. In turn, water intakes located in the depths of industrial districts operated with higher utilisation, reaching 74–86%.

The remoteness of water supply sources from consumers determined the existing water intake regimes, where the differences in the maintained pressure levels in the urban network were clearly observed. Figure 3 shows the daily pressure dynamics of the second lift pumping stations, which illustrates the influence of territorial location on pressure levels. The blue line shows the pressure of the 'Sozh' intake, which maintains the highest values (about 0.42–0.44 MPa) during the entire observation period. In contrast, the Korenevsky (orange line) and Southwest (yellow line) intakes operate at lower pressures: Korenevsky maintains 0.26–0.30 MPa and Southwest maintains 0.33–0.36 MPa.

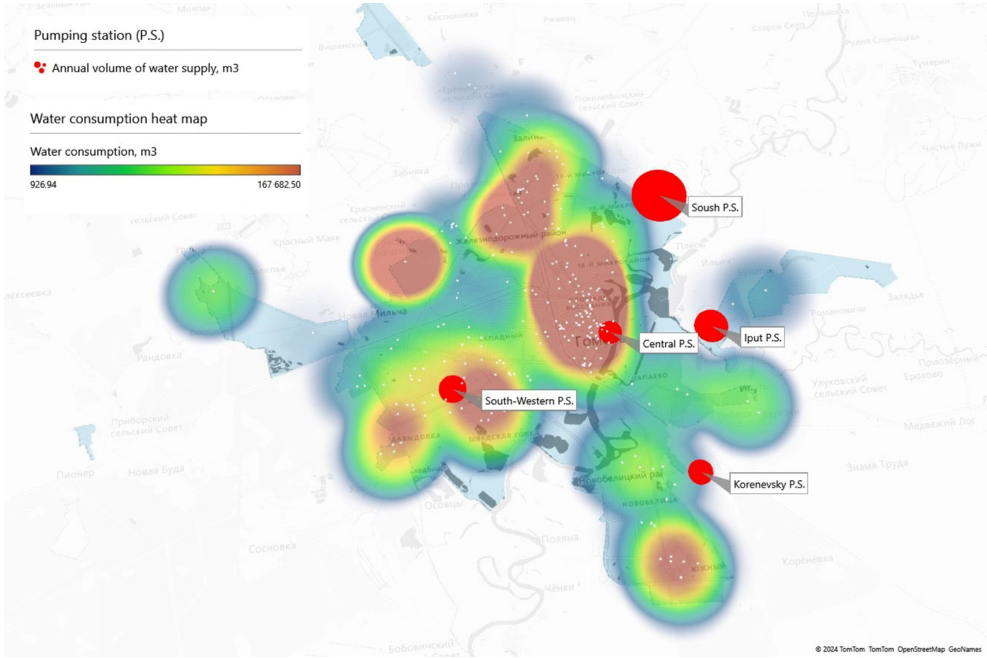

Figure 2. Location of water intakes in Gomel city with labelling of water supply intensity and heat map of water consumption.

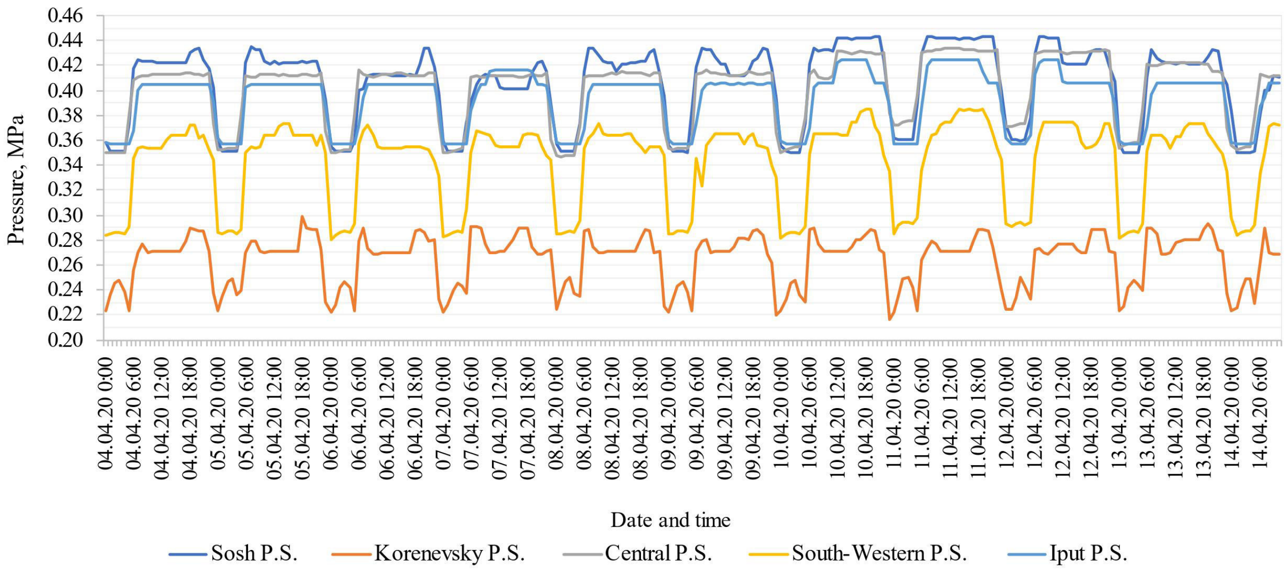

Figure 3. Pumping station (P.S.) pressure schedules of water intakes for the period from 04.04.2023 to 14.04.2023.

3.3. Methods of Spatial Clustering of Water Consumers

The preliminary analysis of the distribution of water intakes and their characteristics has shown a significant influence of geographical location on water supply load, which requires a more detailed approach to data clustering to identify key areas with high concentrations of water consumption. In order to better analyse and optimise water supply management, a spatial clustering method was applied. The study used two approaches to data clustering. The first approach relied solely on the geographical coordinates of water consumption facilities, which allowed for a detailed study of the spatial distribution of water supply and the identification of areas with high water demand. This analysis provided important information on the spatial organisation of water use, identifying key areas of concentrated pressure on the city's infrastructure. The second approach combined data on the location of facilities with their water use levels, which not only identified which neighbourhoods were the highest water users but also assessed the geographic distribution of these facilities across the city.

Both approaches used a hierarchical (agglomerative) clustering method to cluster the data. In this study, Ward's method [29] was used to cluster the data. This method clustered the sites based on their geographical proximity and similarity in the level of water consumption, which accurately compared with the established view of the water supply pattern in the city. In accordance with Ward's algorithm, at each clustering step, the two closest clusters were combined if this minimised the increase in the total sum of squares of variances. This ensured the creation of the most homogeneous groups when combining water consumption facilities [30]:

$$\Delta = \sum_{i} (x_i - \overline{x})^2 - \left(\sum_{x_i \in A} (x_i - \overline{a})^2 + \sum_{x_i \in B} \left(x_i - \overline{b} \right)^2 \right), \tag{1}$$

x_i—coordinates of the i-th object (latitude and longitude); $\overline{a}$, $\overline{b}$, $\overline{x}$—centres of clusters A, B, and their association, respectively.

The quality of clustering in the study was assessed by calculating the silhouette coefficient, which measures how well each object fits into its cluster compared to its location in other clusters [31]:

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))}, \tag{2}$$

where $a(i)$—the average distance between the ith facility and all other facilities in its cluster (intra-cluster distance); $b(i)$—the minimum average distance between the ith facility and all facilities in other clusters (inter-cluster distance).

Here, the silhouette coefficient was calculated as the difference between the intra-cluster distance and the inter-cluster distance normalised by the maximum of these two distances. High values of the coefficient, closer to 1, indicated good clustering and separation efficiency of the data. It is worth noting that understanding the most effective structure of the separation of water consumption objects on the map allows for the adaptation of the management of water withdrawals to real water demands [32,33]. At the same time, the efficiency of clustering and the possibility of changing the boundaries of clusters in real time directly depend on the automation of the water supply system in terms of obtaining operational information about the levels of water consumption in the city. The integration of spatial analysis into water consumption management requires the use of modern technologies for data collection and processing [34].

4. Results and Discussion

4.1. Integration of Spatial Analysis into Water Use Management

As a result of this research, an algorithm for territorial clustering of water supply consumers has been formalised. This algorithm includes several consecutive stages aimed at systematising data, linking them to geographical coordinates and labelling objects into clusters:

Step 1: Collection and aggregation of water consumption data. This step collects data from various sources, including automated metering systems and customer service departments, where monthly or quarterly water consumption figures are recorded for each facility in the city. All data are brought to a single form necessary for analysis and aggregated to the required level of sampling. The sample size is determined by the available resources for data processing and the ability to obtain statistical data for analysis. If capacity is limited, it is recommended to focus on the largest consumers located in different parts of the city to ensure that the data are representative.

Step 2: Geocoding of addresses. Due to the fact that in most cases, subscriber department systems do not support exporting data with latitude and longitude coordinates required for spatial clustering, limitations arise that make the operational spatial analysis of the data difficult. The geographical reference of consumers is usually defined by the legal addresses of the consumers, which requires an intermediate geocoding step. This process consists of converting addresses into geographic coordinates using external geoservices such as Google Maps or Yandex. Maps to obtain the latitude and longitude of each object. The result is a database with geographic coordinates for each consumer and water consumption volume, which allows us to link the facilities to their exact location on the map and prepares the data for the next stages of analysis and clustering.

Step 3: Aggregation of data by address. To reduce computational costs, the aggregation of water consumption data at the house and street level is performed. This step allows us to combine information about several consumers located in the same building, preserving key consumption indicators while reducing the total amount of data. If the data structure includes not only industrial users but also residential customers, it is useful to aggregate the data by bringing the indicators from apartment metering to individual houses or neighbourhoods in order to obtain a generalised picture of water consumption. Aggregated data are used for visual analysis when constructing a heat map of water consumption and assessing its distribution over the city territory, which provides a more visual representation of the distribution of load on the water supply system and remoteness from water intakes. For the system under study, the heat map (Figure 2) visually represents areas with high water demand (red areas) and areas with moderate and low consumption (green and blue areas, respectively).

Step 4: Spatial clustering. Based on the obtained coordinates and aggregated data, a hierarchical clustering method is applied to group sites based on their geographical location and water consumption level. This process results in the formation of data points labelled according to their belonging to specific clusters. These data serve as a basis for constructing convex surfaces that unite the extreme points of each cluster and form the boundaries of territories with different intensities of water consumption.

Let us consider the results of applying the above algorithm to the example of the water supply system under study.

4.2. Hierarchical Clustering and Finding the Optimal Number of Clusters

The primary visual analysis of the dendrogram (Figure 4) plotted on the basis of co-ordinates (latitude and longitude) revealed distinct gaps between the grouped datasets, indicating a natural intergroup spread. For automated cluster selection, a cut-off method was used, which involves drawing a horizontal line at the level of a significant gap. This approach allowed us to identify four groups, the reliability of whose formation was additionally checked by calculating the silhouette coefficient. In the dendrogram, clusters are marked with different colors: orange corresponds to cluster 1, green to cluster 2, red to cluster 3, and purple to cluster 4. These colors indicate groups that were subsequently visualized in the spatial distribution of the data, allowing their location relative to geographic coordinates to be seen more clearly. This approach ensures ease of interpretation of clustering results and the connection between the dendrogram and spatial data.

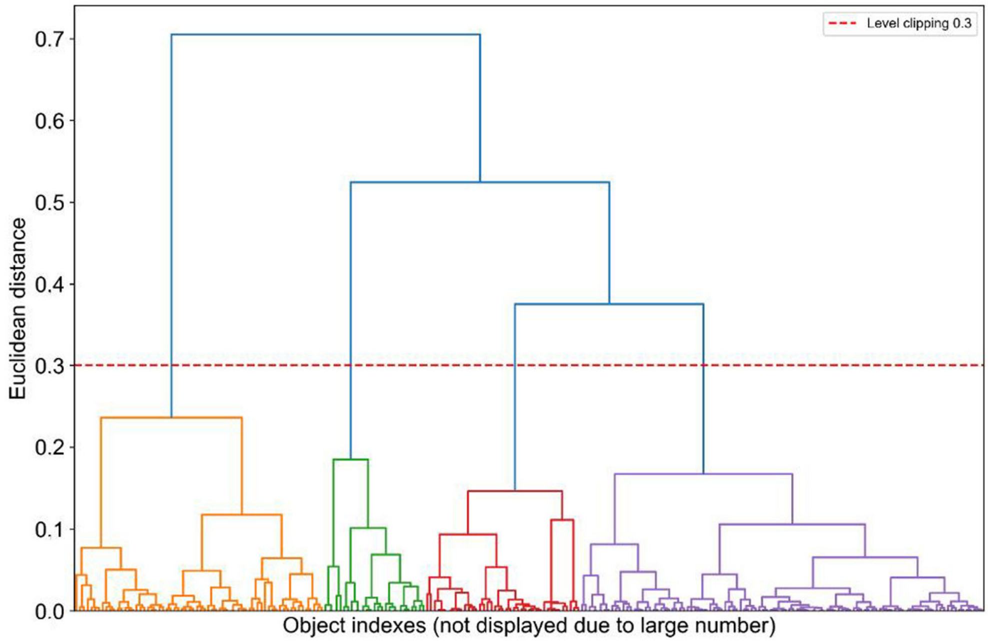

Figure 4. Dendrogram by geographical location of water consumption facilities.

The results confirmed that the highest value of the silhouette coefficient (0.499) is indeed achieved when the data are divided into four clusters. The highest values of the coefficient justify the chosen division, indicating good internal consistency of the clusters and their distinct differentiation. The addition of the water use factor to the geographical coordinates shifted the structure of the clusters from 4 to 6, with the silhouette coefficient decreasing to 0.432. This decrease indicates a more complex data structure arising when additional parameters are taken into account during the geoclustering process. Table 1 shows the results of the silhouette coefficient calculation for clustering the data by location and water consumption level.

Table 1. Results of calculating the silhouette coefficient for various clustering methods.

Cluster Index on the Map	Silhouette Coefficient (by Location of Facilities)	Silhouette Coefficient (by Location of Facilities and Level of Consumption)
2	0.497	0.354
3	0.499	0.418
4	0.485	0.425
5	0.496	0.432
6	0.414	0.400
7	0.433	0.398

Figure 5 shows the character of the silhouette coefficient change when the number of clusters increases when dividing the data by the location of objects and their water consumption level.

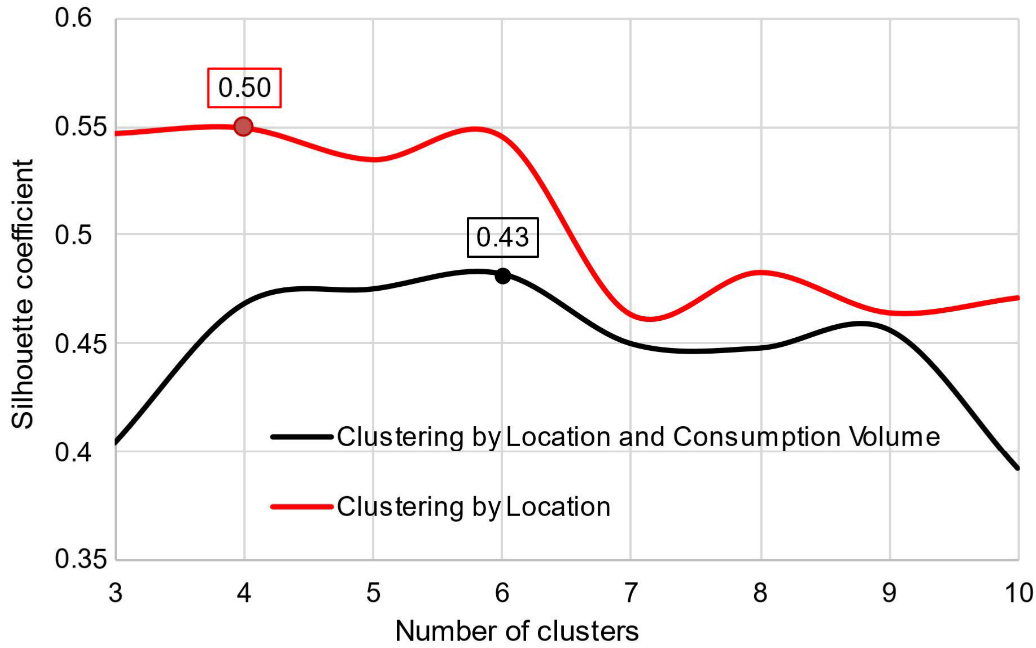

Figure 5. Change in silhouette coefficient with increasing number of clusters.

4.3. Segmentation of Water Consumption Facilities

According to the obtained clusters, the water consumption objects were plotted on a coordinate grid, which allowed visualising groups of consumers by their geographical location. To draw the boundaries, the convex hull algorithm from the ConvexHull class of the SciPy library was used to connect the points defining the outer boundaries. This approach clearly marked the territorial boundaries of the clusters (Figure 6). It should be noted that the allocation of four clusters turned out to be justified not only from the point of view of spatial optimality of their location on the map, but also from the point of view of the boundaries of the pipeline network service areas. For example, cluster No. 1, which is located near the Southwest water intake, is predominantly situated in the south-western part of the city of Gomel and covers residential and industrial areas of the Sovetsky district with socially developed infrastructure. The classification of water consumers is distinguished by an increased number of health care and educational institutions, indicating the active use of the selected zone for both residential and industrial needs. This group included 27.4% of the total number of water consumers considered.

Cluster No. 2, located in the southern part of the city and served by water intakes 'Korenevsky' and 'Iput', includes the territory of Novobelitsky district. This area is characterised by a low density of facilities (11.2% of the total number), large residential areas and suburban zones, and a small number of industrial and social facilities using water from the city network. Cluster 3, covering the north-eastern part of the city in the Zheleznodorozhny District and served by the Sozh water intake, has similar characteristics, with a share of 17.0% of the total number of facilities.

Cluster 4, located in the central part of the city and served by the Central water intake, is the most saturated in terms of density and diversity of infrastructure. Key facilities such as central universities and the city's main shopping centres are located here. With 44.4% of the total number of properties, Cluster 4 reflects the highest density of water use.

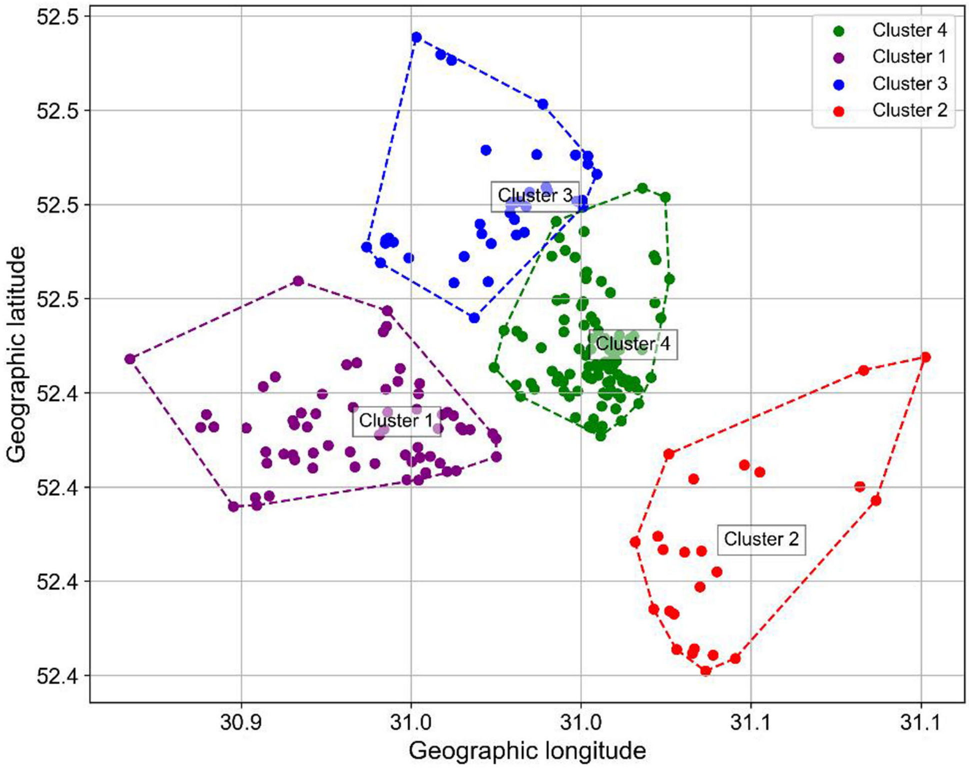

Figure 6. Clustering of water consumption objects according to geographical data.

It is worth noting that the addition of water consumption volumes to the clustering process resulted in two additional clusters compared to the variant where only geographical data were taken into account. The first cluster included the largest consumers with volumes of about 305,000 m^3/year (food factory) and 270,050 m^3/year (moulding companies), while the second cluster included facilities with volumes of about 99,563 to 142,000 m^3/year (hospitals, correctional facilities). This suggests that the approaches used to analyse territorial data have a certain universality and flexibility depending on the factors taken into account. Thus, the allocation of water consumption zones can be used for different urbanised areas with similar population sizes and water supply structures, for example, Irkutsk or Vladikavkaz. In particular, the water supply in Vladikavkaz is organised through underground sources of the Ordzhonikidzevskoye field, using water intake structures. For comparison, the water supply system of Gomel includes five water intakes, similar to Vladikavkaz.

4.4. Application of Clustering Results for Optimal Installation of Pressure Sensors

It is impossible to build mechanisms for optimising water supply without obtaining continuous information on the hydraulic regimes in the city network. For this purpose, it is important to clearly define at which points in the city it is necessary to install registration devices and which territorial zones they will cover. The division of the city of Gomel into clusters made it possible to prioritise areas for the placement of pressure sensors, which greatly simplified the planning of further development of the monitoring system. To date, more than 20 pressure sensors have been installed in Gomel, located at key control points of water supply and at the outlet of pumping stations. The figure shows the distribution of installed pressure sensors across the city territory in accordance with the preliminary clustering of legal water consumption facilities. The map (Figure 7) shows the outer boundaries of each of the four clusters based on the results of the water consumption analysis. All points of placement of pressure recording devices both at the outlet of WS pumping stations and at CP control points in the city are marked here.

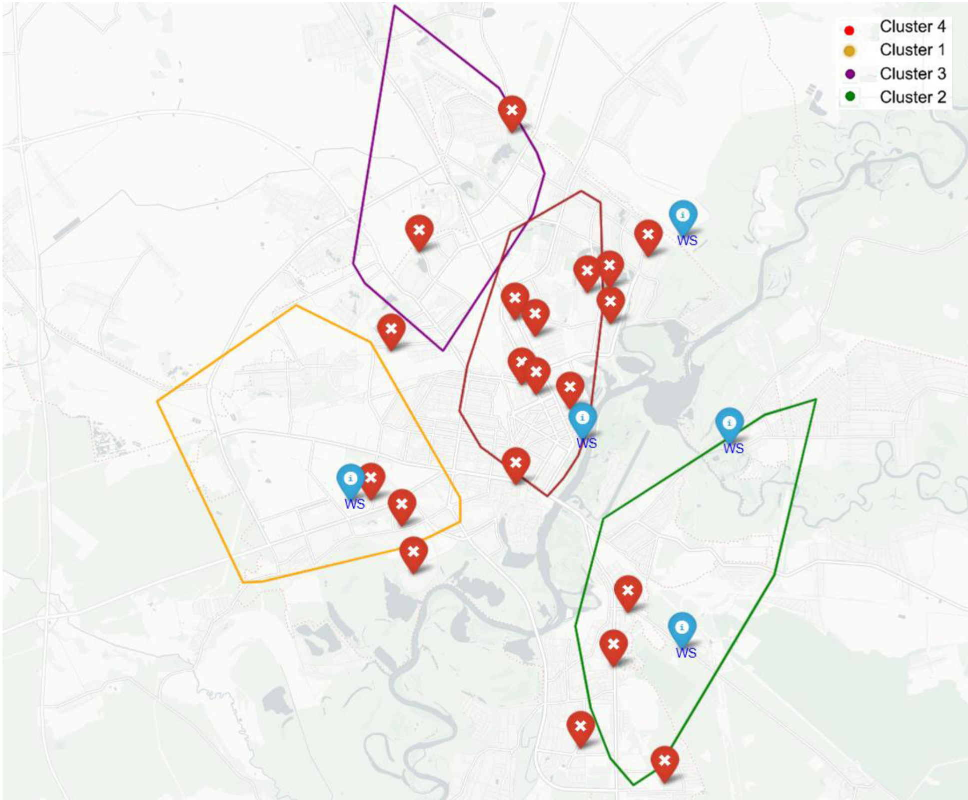

Figure 7. Construction of water consumption cluster boundaries and pressure sensor location points: WS—water station.

It can be seen from the placement scheme that some sensors extend beyond the cluster boundaries. This is due to the fact that the constructed cluster boundaries covered mainly commercial water consumption facilities, not taking into account the housing stock due to the limitation of the provided information. Nevertheless, this arrangement of sensors does not distort the validity of the overall device distribution picture and allows us to cover significant areas of the urban water supply network. It can be clearly seen here that the current distribution of sensors across clusters differs from neighbourhood to neighbourhood. For example, cluster 1 of the south-western neighbourhood requires additional sensor equipment for uniform pressure monitoring in this area. At the same time, the central part of the city (cluster 4) is already quite densely equipped with sensors, so the priority of installing additional devices in this neighbourhood is much lower than in others.

The ultimate goal of this research is to develop a hydraulic modelling system for Gomel's water supply. For this purpose, an IoT platform for pressure monitoring based on the Internet of Things technology is being developed. The platform organises data exchange using the MQTT protocol, which allows for the efficient collection and transmission of information from pressure sensors. This also opens up opportunities for research and the development of more accurate prediction models based on neural networks. Figure 8 shows the pressure model of one of the control points in the city network and its actual pressure. The data on the pressure at the pumping stations of the investigated water intakes are used as input parameters. Output parameters are presented as pressures in the control points of the system.

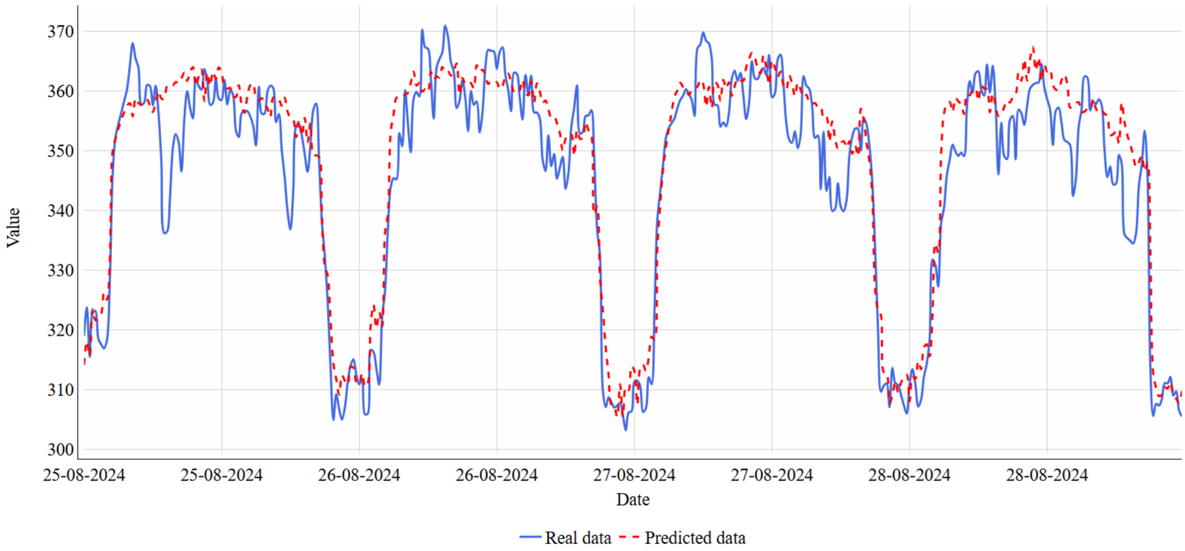

Figure 8. Pressure modelling at control points using a multilayer perceptron model.

The preliminary plot clearly shows the high accuracy of the multilayer perceptron model, which is confirmed by the following metrics calculated for one of the benchmarks: the mean absolute error (MAE) was 5.07 kPa, the root mean square error (RMSE) was 6.57 kPa, and the mean absolute percentage error (MAPE) was 1.48%. Mathematical aspects of model construction and features of hydraulic modelling are part of further research and will be discussed in detail in the following scientific publications.

5. Conclusions

The study of spatial clustering of water consumption facilities using Ward's method allowed us to identify four main clusters based on the geographical location of facilities in the example of the city of Gomel. The clustering showed high spatial optimality: 44.4% of water consumption facilities are concentrated in the central part of the city (cluster 4), while the south-western part (cluster 1) covers 27.4% of facilities, demonstrating an equal distribution of consumers between residential and industrial areas. The southern part of the city (cluster 2) is characterised by low density (11.2%) and the north-eastern part (cluster 3) covers 17% of the facilities, highlighting the different levels of water consumption and distribution of infrastructure across the city districts.

The application of the agglomerative clustering method in this study allowed for the separation of over 200 industrial facilities by geographical location and water consumption volume, and the introduction of the water consumption volume parameter into the clustering provided additional separation of the data by identifying two new clusters Some facilities, such as the confectionery factory with 305,000 m^3/year and metal casting facilities (270,050 m^3/year), were found to be the most water-intensive, indicating significant infrastructure loads in these clusters. Accounting for water consumption volumes allowed for a deeper segmentation of facilities and the identification of key areas with high resource demand, which is particularly important for optimising water and energy management. However, it is worth noting that this study, due to data limitations, did not consider factors such as population density and types of industrial facilities, which could significantly impact the analysis results. Accounting for these factors is part of future research aimed at expanding the methodology and identifying more detailed water consumption clusters for various urbanised areas.

The results of the study confirm that agglomerative clustering, enhanced by accurate geographical data, provides a powerful tool to highlight consumers by their specific characteristics, which opens up opportunities to further optimise supply systems and improve their efficiency. In the context of smart city development, such methods play a key role as they allow the integration of resource consumption data with other elements of urban infrastructure, creating more sustainable, reliable and energy-efficient systems.

Author Contributions: Conceptualisation, A.A.K. and R.V.K.; methodology, S.N.S. and E.A.E.; software, A.E.B.; validation, S.N.S. and E.A.E.; formal analysis, N.V.M.; investigation, A.Y.D.; resources, A.Y.D.; data curation, A.Y.D.; writing—original draft preparation, A.A.K. and R.V.K.; writing—review and editing, N.V.M.; visualisation, A.E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available from the corresponding authors upon reasonable request.

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

References

1. Twomey, K.M.; Webber, M.E. Evaluating the energy intensity of the US public water system. *Energy Sustain.* **2011**, *54686*, 1735–1748. [CrossRef]
2. Zapata, O. More water please, it's getting hot! The effect of climate on residential water demand. *Water Econ. Policy* **2015**, *1*, 1550007. [CrossRef]
3. Bezerra, B.G.; Lee, T.S. Crop evapotranspiration and water use efficiency. In *Irrigation Systems and practices in Challenging Environments*; IntechOpen: Rijeka, Croatia, 2012; Volume 9, pp. 57–76. [CrossRef]
4. Chupin, R.V.; Pham, M.N.; Chupin, V.R. Optimal reconstruction of water supply network under conditions of reduced water consumption. *E3S Web Conf. EDP Sci.* **2019**, *102*, 03005. [CrossRef]
5. Fracasso, P.T.; Barnes, F.S.; Costa, A.H.R. Energy cost optimization in water distribution systems using markov decision processes. In Proceedings of the 2013 International Green Computing Conference Proceedings (IEEE), Arlington, VA, USA, 27–29 June 2013; pp. 1–6.
6. Poncelet, K.; Hoschle, H.; Delarue, E.; Virag, A.; Drhaeseleer, W. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning. *IEEE Trans. Power Syst.* **2016**, *32*, 1936–1948. [CrossRef]
7. Poncelet, K.; Delarue, E.; Six, D.; Duerinck, J.; D'haeseleer, W. Impact of the level of temporal and operational detail in energy-system planning models. *Appl. Energy* **2016**, *162*, 631–643. [CrossRef]
8. Arefifar, S.A.; Mohamed, Y.A.-R.I.; El-Fouly, T.H.M. Supply-adequacy-based optimal construction of microgrids in smart distribution systems. *IEEE Trans. Smart Grid* **2012**, *3*, 1491–1502. [CrossRef]
9. Arefifar, S.A.; Mohamed, Y.A.-R.I.; El-Fouly, T.H.M. Optimum microgrid design for enhancing reliability and supply-security. *IEEE Trans. Smart Grid* **2013**, *4*, 1567–1575. [CrossRef]
10. Zhang, J.; Lu, Q.; Guan, L.; Wang, X. Analysis of factors influencing energy efficiency based on spatial quantile autoregression: Evidence from the panel data in China. *Energies* **2021**, *14*, 504. [CrossRef]
11. Zheng, J.; Mi, Z.; Coffman, D.M.; Milcheva, S.; Shan, Y.; Guan, D.; Wang, S. Regional development and carbon emissions in China. *Energy Econ.* **2019**, *81*, 25–36. [CrossRef]
12. Abokifa, A.A.; Sela, L. Integrating spatial clustering with predictive modeling of pipe failures in water distribution systems. *Urban Water J.* **2023**, *20*, 465–476. [CrossRef]
13. Abokifa, A.A.; Sela, L. Identification of spatial patterns in water distribution pipe failure data using spatial autocorrelation analysis. *J. Water Resour. Plan. Manag.* **2019**, *145*, 04019057. [CrossRef]
14. Ezechi, C.G.; Okoroafor, E.R. Integration of Artificial Intelligence with Economical Analysis on the Development of Natural Gas in Nigeria; Focusing on Mitigating Gas Pipeline Leakages. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 31 July–2 August 2023.
15. Liu, W.; Chen, Z.; Hu, Y. Failure Pressure Prediction of Defective Pipeline Using Finite Element Method and Machine Learning Models. In Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, TX, USA, 13 January 2025.
16. Fan, X.; Wang, X.; Zhang, X.; Yu, X. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliab. Eng. Syst. Saf.* **2022**, *219*, 108185. [CrossRef]
17. Gunde, S.; Chikaraddi, A.K.; Baligar, V.P. IoT based flow control system using Raspberry PI. In Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 1–2 August 2017; IEEE: Piscataway, NJ, USA; pp. 1386–1390.

18. Meena, K.B.; Dhanalakshmi, S. IoT Water Management and Distribution System for Smart City Using Artificial Intelligence. *J. Ubiquitous Comput. Commun. Technol.* **2023**, *5*, 339–351.
19. Gautam, G.; Sharma, G.; Magar, B.T.; Shrestha, B.; Cho, S.; Seo, C. Usage of IoT framework in water supply management for smart city in Nepal. *Appl. Sci.* **2021**, *11*, 5662. [CrossRef]
20. Aljameel, S.S.; Alomari, D.M.; Alismail, S.; Khawaher, F.; Alkhudhair, A.A.; Aljubran, F.; Alzannan, R.M. An Anomaly Detection Model for Oil and Gas Pipelines Using Machine Learning. *Computation* **2022**, *10*, 138. [CrossRef]
21. Yang, F.-P.; Zhang, L.; Luo, J.-H.; Zhao, X.-W.; Zhang, G.-L. Study on the yield behavior of oil and gas pipe in typical stress states. *Gongcheng Lixue/Engineering Mechanics* **2013**, *30*, 293–297. [CrossRef]
22. Sourabh, N.; Timbadiya, P.V.; Patel, P.L. Leak detection in water distribution network using machine learning techniques. *ISH J. Hydraul. Eng.* **2023**, *29*, 177–195. [CrossRef]
23. Momeni, A.; Piratla, K.R. Prediction of Water Pipeline Condition Parameters Using Artificial Neural Networks. *Pipelines* **2022**, *2022*, 21–29.
24. Bata, M.; Carriveau, R.; Ting, D.S.K. Short-term water demand forecasting using hybrid supervised and unsupervised machine learning model. *Smart Water* **2020**, *5*, 2. [CrossRef]
25. Saroja, H.; Dharshini, S. Deep learning approach for prediction and classification of potable water. *Anal. Sci.* **2023**, *39*, 1179–1189. [CrossRef] [PubMed]
26. Chang, N.B.; Prapinpongsanone, N.; Ernest, A. Optimal sensor deployment in a large-scale complex drinking water network: Comparisons between a rule-based decision support system and optimization models. *Comput. Chem. Eng.* **2012**, *43*, 191–199. [CrossRef]
27. Khorshidi, M.S.; Nikoo, M.R.; Sadegh, M. Optimal and objective placement of sensors in water distribution systems using information theory. *Water Res.* **2018**, *143*, 218–228. [CrossRef] [PubMed]
28. Zhang, W.; Yang, X.; Li, J. Sensor Placement for Leak Localization in Water Distribution Networks Based on Graph Convolutional Network. *IEEE Sens. J.* **2022**, *22*, 21093–21100. [CrossRef]
29. Eszergár-Kiss, D.; Caesar, B. Definition of user groups applying Ward's method. *Transp. Res. Procedia* **2017**, *22*, 25–34. [CrossRef]
30. Ward, J.H., Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]
31. Shahapure, K.R.; Nicholas, C. Cluster quality analysis using silhouette score. In Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), Sydney, Australia, 6–9 October 2020; IEEE: Piscataway, NJ, USA; pp. 747–748.
32. Kapanski, A.; Hruntovich, N.; Bakhur, S.; Markaryants, L.; Dolomanyak, L. Optimize the cost of paying for electricity in the water supply system by using accumulating tanks. *E3S Web Conf.* **2020**, *178*, 01065. [CrossRef]
33. Hruntovich, N.V.; Kapanski, A.A.; Baczynski, D. Optimization of a variable frequency drive pump working on a water tower. In Proceedings of the E3S Web of Conferences: 2019 International Scientific and Technical Conference Smart Energy Systems, Kazan, Russia, 18–20 September 2019; Volume 124, p. 05060. [CrossRef]
34. Sarbu, I. A study of energy optimisation of urban water distribution systems using potential elements. *Water* **2016**, *8*, 593. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.