sustainability

MDPI

Article

Comparative Analysis of Leak Detection Methods Using Hydraulic Modelling and Sensitivity Analysis in Rural and Urban–Rural Areas

Izabela Bartkowska [1], Łukasz Wysocki [1,*], Artur Zajkowski [1] and Piotr Tuz [2]

[1] Department of Water Supply and Sewerage, Bialystok University of Technology, 15-351 Bialystok, Poland; i.bartkowska@pb.edu.pl (I.B.); artur.zajkowski@sd.pb.edu.pl (A.Z.)
[2] Water Energy Systems (WES), 15-256 Bialystok, Poland; p.tuz@wes.info.pl
* Correspondence: lukasz.wysocki@sd.pb.edu.pl; Tel.: +48-798-735-257

Abstract: Water scarcity is a significant global challenge, exacerbated by leakages in water distribution networks. This paper addresses the challenge of detecting leakages in rural and urban–rural water supply systems through hydraulic modelling and a sensitivity analysis. Two distinct real-world network models were studied to assess real and simulated leakage scenarios varying in location and magnitude. A distinct leakage detection approach utilizing outflow measurements from hydrants was tested. Additionally, the effectiveness of various statistical measures—such as correlation, angular closeness, Euclidean distance, Manhattan distance, Chebyshev distance, cosine similarity, and Spearman correlation—were evaluated to determine their efficacy in leakage detection. Different methods for identifying leak candidates were explored and compared, either by selecting a single leak candidate based on similarity measures or by identifying a group of candidates to mark leak hotspots. Density-based spatial clustering of applications with noise was used to assess the number of potential leak candidate groups. The study's findings contribute to the optimization of leak detection strategies in water supply networks, particularly in rural settings, where detection is challenging due to scarce measurement datasets, budget restrictions, and operational constraints.

Keywords: hydraulic model; leak detection; sensitivity analysis; sensitivity matrix

check for updates

Citation: Bartkowska, I.; Wysocki, Ł.; Zajkowski, A.; Tuz, P. Comparative Analysis of Leak Detection Methods Using Hydraulic Modelling and Sensitivity Analysis in Rural and Urban–Rural Areas. *Sustainability* **2024**, *16*, 7405. https://doi.org/10.3390/su16177405

Academic Editor: Marc A. Rosen

Received: 19 April 2024
Revised: 30 July 2024
Accepted: 22 August 2024
Published: 28 August 2024

Copyright: © 2024 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

Water loss in water distribution systems constitutes a widespread issue globally that is routinely confronted by water utility companies. These losses can originate from apparent losses, which are tied to unauthorized consumption, measurement errors, or billing inaccuracies, or from real losses, which are primarily associated with leaks [1]. According to a report by EUREAU, the average percentage of non-revenue water in Europe is around 25%, translating to 2.69 m^3 per kilometre of networks annually [2]. Increased water losses have an economic impact, such as costs related to the extraction and treatment of raw water—which leads to the overutilization of technical equipment and the higher use of chemicals. The excessive energy used in the treatment processes and for pumping surplus water into the distribution system directly correlates with the level of losses [3], with these costs typically passed on to the end consumer in the form of higher water fees. Furthermore, excessive water losses have environmental implications, such as the depletion of water—a valuable resource, affecting aquatic ecosystems and leading to unnecessary greenhouse gas emissions, particularly if the energy is derived from the combustion of fossil fuels. The magnitude of this issue is underscored by the fact that the water and wastewater industries account for 2% of all greenhouse gas emissions in the US annually, positioning them as substantial contributors to overall GHG emissions [4]. On a global scale, the water sector's consumption accounts for about 4% of the world's total electricity use, amounting to 850 terawatt-hours annually [5]. These factors highlight the importance of leakage

reduction, water use efficiency, and service quality improvement as key considerations for sustainable water management [6]. While these challenges are global, specific issues in small utility networks present unique difficulties that require tailored solutions.

Small water utility companies face several challenges, including networks that are poorly structured and inadequately maintained; a lack of detailed information regarding the characteristics, age, and locations of valves, as well as the connections and topology of pipes; the absence of advanced tools and techniques for detecting and controlling leaks (e.g., Supervisory Control and Data Acquisition—SCADA); outdated and incomplete maps; insufficient metering; and a deficiency in flow and pressure monitoring [7,8]. In most cases, rural and mixed areas present higher water losses per unit length than their urban counterparts [9], with much fewer financial resources to carry out advanced water loss strategies [10].

To address the challenges associated with water loss in rural and urban–rural water distribution systems, one of the viable solutions is the implementation of hydraulic modelling and a sensitivity matrix-based approach. The initial concept involves the application of a fault sensitivity matrix (FSM), which stores the information about every potential leak scenario in the analysed network generated using a hydraulic model [11]. The sensitivity matrix is defined by Equation (1) as follows:

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix} = \begin{bmatrix} \partial P_1/\partial q_1^{\ell} & \partial P_2/\partial q_1^{\ell} & \vdots & \partial P_N/\partial q_1^{\ell} \\ \partial P_1/\partial q_2^{\ell} & \partial P_2/\partial q_2^{\ell} & \vdots & \partial P_N/\partial q_2^{\ell} \\ \vdots & \vdots & \ddots & \vdots \\ \partial P_1/\partial q_N^{\ell} & \partial P_1/\partial q_N^{\ell} & \cdots & \partial P_N/\partial q_N^{\ell} \end{bmatrix} \quad (1)$$

where N—the number of nodes in the network, s_N—the vector of sensitivities to the leak at the N-th node, P_N—the pressure at the nodes, and $q_N{}^{\ell}$—the leak flow rate.

Each row of the sensitivity matrix specifies how a leak at a particular point in the network influences pressure changes throughout the system compared to a normal state. To locate a leak, the sensitivity vectors s_i (the rows of the sensitivity matrix) are compared with the residual vector. The residual vector is a vector of differences between the measured pressures at each node and the nominal pressures (in the non-leak scenario), estimated based on simulations using the hydraulic model of the network. The residual vector is defined by Equation (2) as follows:

$$r = [P_1 - P_1{}^{nom}, P_2 - P_2{}^{nom}, \ldots, P_N - P_N{}^{nom}], \quad (2)$$

where P_N—the pressure measured at point N, and $P_N{}^{nom}$—the non-leak pressure at node N.

Practically, to simplify, one may only calculate sensitivities for pressure at nodes where pressure measurements are taken. Over the years, various concepts have been adopted concerning the selection of leak candidates based on the FSM. The sensitivity matrix can be binarized by setting a threshold, transforming each sensitivity vector into a binary string of ones and zeros, which act as leak signatures. These are then compared to a similarly binarized residual vector for matching [12,13]. The binarized sensitivity matrix has also been used for determining sensor placement [14]. Due to difficulties in determining the binarization threshold value, attempts have also been made to determine leak candidates by calculating the correlation between the residual vector and each sensitivity vector. The highest correlation corresponds to the highest probability of a leak occurring at that location [15–17]. Other works have also included the additional use of measures such as the angle or Euclidean distance between the residual vector and each row of the sensitivity matrix [18,19]. Attempts have also been made to increase the model's reliability by structuralizing pressure residuals, which are designed to be sensitive only to specific subsets of faults in a system, which, in real scenarios, have shown an increased accuracy compared to that of directional residuals [20]. Efforts have also been made to increase the precision of localization by employing a leak signature space (LSS) that is only minimally impacted by the size of the leak [21], considering uncertainties related to water demands [22]. In some

studies, in addition to using a static model, a dynamic model was also used, and a given statistical measure was calculated for each time step in a selected time horizon and then averaged; for instance, the calculated average angle between vectors was utilised to reduce the impact of the demand uncertainty on results [18]. Methods based on the sensitivity matrix were compared with other methods such as the inverse problem and classification methods, which, depending on the scenario, achieved comparable or weaker results [23]. In comparison to the method based on linear approximation, the sensitivity matrix was three times less precise when using a perfect hydraulic model, but when measurement data affected by errors were introduced, the sensitivity matrix was 10% more effective [24], which may mean that for rural conditions, whose data may be prone to errors, the use of a sensitivity matrix may be justified. There are also studies using flow rate measurements and a sensitivity matrix [25–27]; however, in rural water supply conditions, such data may not be available due to the relatively higher construction and installation costs compared to pressure measurement points.

The objective of this study is to explore the feasibility of applying hydraulic modelling and sensitivity matrix-based leak detection methods in rural and urban–rural water distribution systems. Furthermore, the study will examine the effectiveness of various similarity measures in detecting and locating leaks, endeavouring to identify which metrics yield the most accurate results in the context of the measurement data gathered by the water utility. Moreover, we will assess the potential of adopting an alternative approach through hydrant flow rate measurements to explore a distinct method apart from the traditionally used pressure measurements, in consideration of the fact that water supply companies are obliged to carry out periodic tests of hydrants and have this type of data at their disposal, which creates the possibility of applying them.

2. Materials and Methods

This research was conducted on two real-world water distribution systems located in rural areas, labelled as M1 and M2. For both of them, computer hydraulic models were created in a similar manner using QGIS 3.34.8 software, which facilitates the integration, transformation, and analysis of various types of spatial data. The process of model creation was as follows:

- Information regarding the layout of the water distribution network was acquired, and the necessary spatial transformations were carried out to develop a network graph structure. This graph was populated with data related to the diameters of the water pipes, their material, and roughness, utilizing available information from both digital and analogue maps, as well as through consultations with the employees of the water utility company.
- Node points were generated at pipe ends. Information about elevation was obtained from publicly accessible DEM (digital elevation model) with a resolution of 1 m, assuming the required depth of the network depending on the freezing zone.
- In both cases, the networks are solely supplied by a singular water treatment plant. Data from the monitoring systems of the water treatment facility were obtained, and the measured pressure data at the facility's output were analysed. In both cases, the pump systems are controlled to maintain a constant pressure. The operational mode of the water treatment facility was accurately replicated in the model by parametrized reservoir object.
- Data on water consumption by end-users were obtained. Address data were standardized and geolocated by integration with a publicly accessible address database and the use of a geolocation API. The average daily water consumption data were then allocated to the model using Voronoi polygons, which partition the network service area into zones centred around node locations, assigning water usage data based on proximity, allowing for accurate representation of demand within the hydraulic model.
- Once the integration and transformation of the required data were complete, the model was exported to the EPANET 2.2 input file format (inp extension).
- Models used in this study were not calibrated against measurement data before being used.

In both M1 and M2 networks, pressure measurements were taken by utilizing designated fire hydrants for this purpose. These measurements were taken using a data logger

with two programmable inputs connected to a pressure sensor. The sensor, with a measuring range of 0–10 bar and an accuracy of ±0.5% FS, possessed valid calibration certificates. For M2 model, additional flow rate measurements were conducted from a hydrant. The flow rate values were determined using a Class 2 R400 DN65 ultrasonic water meter, which also holds a valid verification mark. The flow rate information was transmitted to the logger via pulse output. The measurement setup included appropriately selected adapters and connectors to facilitate the device's connection to above-ground hydrants, and a hydrant stand was used for measurements on underground hydrants when necessary.

Table 1 presents a compilation of the analysed scenarios along with their corresponding leak magnitudes. In the case of M1 model, six simulated leak scenarios were executed in three different locations (A, B, and C) by opening a hydrant and analysing the flow. These scenarios were labelled as follows: M1-A-1, M1-A-2, M1-A-3, M1-B-1, M1-B-2, and M1-C-1. During each simulated leak scenario, readings of the measured pressure values were taken at seven different designated locations of above-ground hydrants.

Table 1. Analysed scenarios along with their corresponding leak magnitudes.

Scenario	Leak Magnitude (m^3/h)	Number of Pressure Measurements	Number of Hydrant Outflow Measurements
M1-A-1	52	7	0
M1-A-2	36	7	0
M1-A-3	28	7	0
M1-B-1	51	7	0
M1-B-2	25	7	0
M1-C-1	28	7	0
M2-A-P	25	14	12
M2-A-F	25	14	12

For M2 model, a real-case scenario was investigated, in which a leak of 25 m^3/h occurred at unknown location A within the network. Pressure measurements were conducted on 14 hydrants across the network, and in the presence of this leak, the performance of 12 hydrants was further examined through flow rate measurements. However, flow rate check could not be conducted on two hydrants due to damage that prevented device connection. Scenario based on pressure measurements was labelled as M2-A-P, and that for hydrant outflow measurements was labelled as M2-A-F.

All hydrants used for measurement purposes were dry barrel, as they are designed to prevent water from freezing inside by keeping the valve below the frost line, making them suitable for colder climates.

For the hydrants, where pressure measurements or both pressure and flow measurements were taken, an additional measurement of the sensor installation elevation was carried out. For relevant nodal points within M1 and M2 models, the elevation was accordingly modified.

After collecting measurement data, sensitivity matrices were created for M1 and M2 models by simulating various leak magnitudes at every possible point of the network using the WNTR Python package [28] and calculating sensitivity vectors. A residual vector was also computed for each analysed scenario. In the M2-A-F scenario, hydrant flow tests were simulated in the model for each potential leak location at nodes corresponding to actual hydrant performance tests. This was achieved by implementing an emitter coefficient at specific nodes, which models the flow through a nozzle or orifice discharging to the atmosphere [29], set at a value of 14 l/s/m, representing a nominal outlet size of 2–2.5 inches [30]. For M2 model, 12 hydrant tests were simulated in locations corresponding to field tests for each potential leak location. In the M2-A-F scenario, the

sensitivity matrix appropriately includes hydrant outflow residuals. The residual vector was computed according to Equation (3) as follows:

$$r = [O_1 - O_1{}^{nom}, O_2 - O_2{}^{nom}, \ldots, O_N - O_N{}^{nom}] \quad (3)$$

where O_N—hydrant outflow measured at point N, and $O_N{}^{nom}$—non-leak hydrant outflow at point N.

In each of the analysed scenarios, computations were performed for different similarity measures.

- Correlation was computed according to Equation (4).

$$corr(r, s_N) = \frac{(r - \bar{r})(S_N - \overline{S_N})}{\sqrt{(r - \bar{r})(r - \bar{r})}\sqrt{(S_N - \overline{S_N})(S_N - \overline{S_N})}} \quad (4)$$

where r—residual vector, and s_N—sensitivity vectors.

- Angular closeness is given by Equation (5).

$$\alpha_i = arccos(\frac{r\, s_N}{||r||\, ||s_N||}) \quad (5)$$

where r—residual vector, and s_N—sensitivity vectors.

- Euclidean distance is described in Equation (6).

$$d(r, s_N) = \sqrt{(r - s_N)^T (r - s_N)} \quad (6)$$

where r—residual vector, and s_N—sensitivity vectors.

- Manhattan distance was calculated according to the expression in Equation (7).

$$d(r, s_N) = \sum_{i=1}^{n} |r - s_N| \quad (7)$$

where r—residual vector, s_N—sensitivity vectors, and n—length of vectors.

- Chebyshev distance was calculated following Equation (8).

$$d(r, s_N) = max_i |r_i - s_{Ni}| \quad (8)$$

where r—residual vector, s_N—sensitivity vectors, and r_i, s_{Ni}—corresponding elements of vectors r, s_N.

- Cosine similarity was calculated based on Equation (9).

$$\cos(r, s_N) = \frac{r\, s_N}{||r||\, ||s_N||} \quad (9)$$

where r—residual vector, and s_N—sensitivity vectors.

- Spearman correlation was calculated as defined in Equation (10).

$$\rho(r, s_N) = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \quad (10)$$

where n—number of elements in vector, and d_i—difference between the ranks of the corresponding values of r and s_N.

Potential leak locations were identified using two methods for each calculated similarity measure: a single point in the network exhibiting the highest similarity relative to the residual vector and a group of points identified as the top 1% with the highest similarity to the residual vector for each of the similarity measures. For the group of points, additional

clustering was performed using the DBSCAN method in order to extract the most likely areas of leakage by grouping nearby high-probability leak points.

To assess the effectiveness of each employed method, calculations of the topological distance between the identified leak location and the actual leak location were performed. For the method based on a group of points, topological distance calculations were carried out for each of the generated areas (clusters), and the minimum, maximum, and average distances were computed to illustrate how concentrated the generated clusters are.

2.1. M1 Model

M1 model is water distribution system, situated in Poland, within the Podlaskie Voivodship. This system is characterized by network graph comprising 1754 nodes and 1768 pipes, encapsulating a total network length of 35 kilometres, with average link length around 20 m. It serves the water needs of the local population and infrastructure with the number of connections totalling 768. The system sustains an average daily water consumption of 312.72 cubic meters per day (m^3/d). The system is not divided by any pressure regulating devices, and there are no pressure management zones.

Figure 1 presents water distribution network topology of M1 with leakage scenario, and measurement point locations are marked. Measurement points in M1 model were primarily located at hydrants situated at the middle of ring network. One point was located near the entry point (J719) and another at the branch end (J37). Simulated leakage points of A, B, and C were at the ends of the network at hydrants that are used for flushing. The water treatment plant is situated near a population cluster, with its outlet splitting into a 225 mm pipe to the north and a 160 mm pipe to the south.

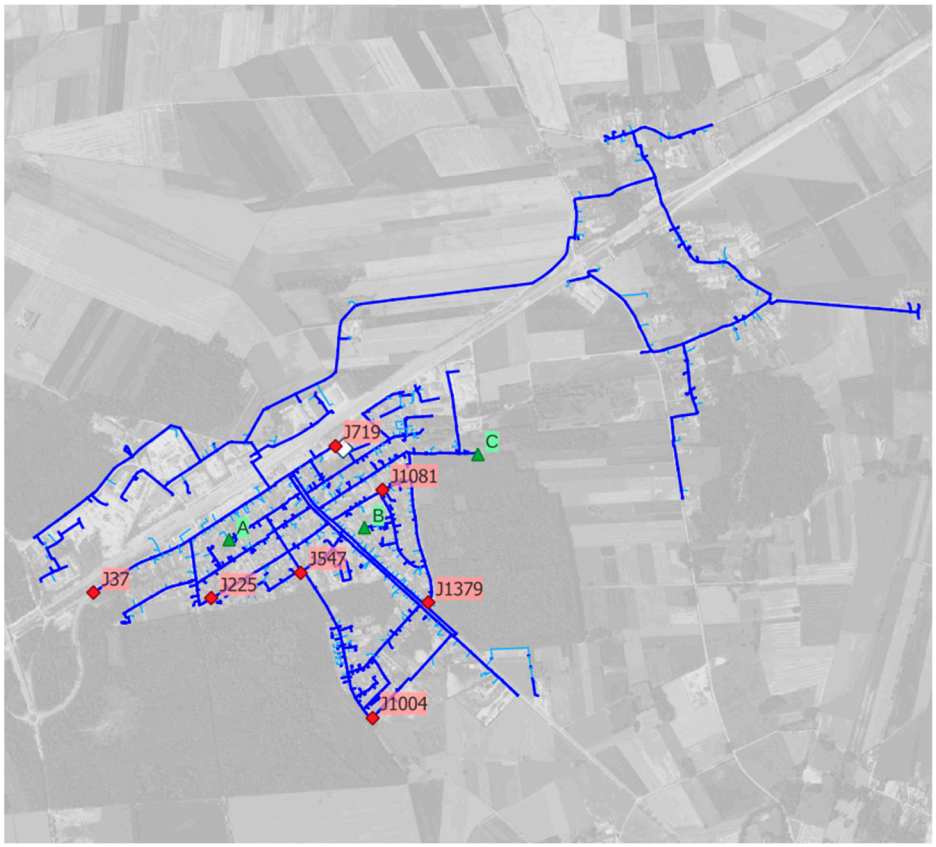


Figure 1. M1 water distribution network topology. Red points—measurement locations, green triangles with labels A–C—leak locations, white point—water treatment plant.

2.2. M2 Model

M2 model is the water distribution system located in Poland in the Kuyavian-Pomeranian Voivodeship. The network's structure is detailed with 3296 nodes and 3305 pipes, spanning a comprehensive length of 107.78 kilometres, with average model link length around 51 m. This system facilitates a total of 856 connections, serving the community's water distribution needs. The recorded average daily water consumption within this network amounts to 426 cubic meters per day (m^3/d). The system is not divided by pressure-regulating devices, and there are no pressure management zones.

Figure 2 presents water distribution network topology of M2 with leakage scenario, and measurement point locations are marked. Half of the measurement points in M2 model were situated in the middle of the branches, while the other half were located at the ends of the branches. Only one point was located in the ring network (1888). The water treatment plant is situated in the middle of the water distribution system, with its outlet splitting into two 160 mm pipes.



Figure 2. M2 water distribution network topology. Red points—measurement locations, green triangle—leak location, white point—water treatment plant.

Figure 3 describes the percentage composition of different nominal diameters in the length of a network for M1 and M2 models. The data show that the 110 mm diameter is the most common in the M1 network, while M2 has a higher occurrence of the 90 mm diameter. The 160 mm size is present in both networks but is more prevalent in M1. The 100 mm diameter is used marginally in M2 and not at all in M1. The smallest diameter, 80 mm, has very low occurrence in both networks, and the largest, 225 mm, is minimally used in M1 and absent in M2.

Figure 4 represents minimum and maximum velocity data for M1 and M2 networks. Data indicate that both systems predominantly operate at very low velocities: 97.74% of M1's velocities are within the 0–0.05 m/s range for minimum values and 79.25% for maximum values, whereas M2 operates almost entirely within this range with 99.94% for minimum values and 78.68% for maximum values. Presumably, such small values of velocity can be caused by oversized pipe diameters, which are necessary to meet fire protection requirements, and low water transit through the network due to low population density.

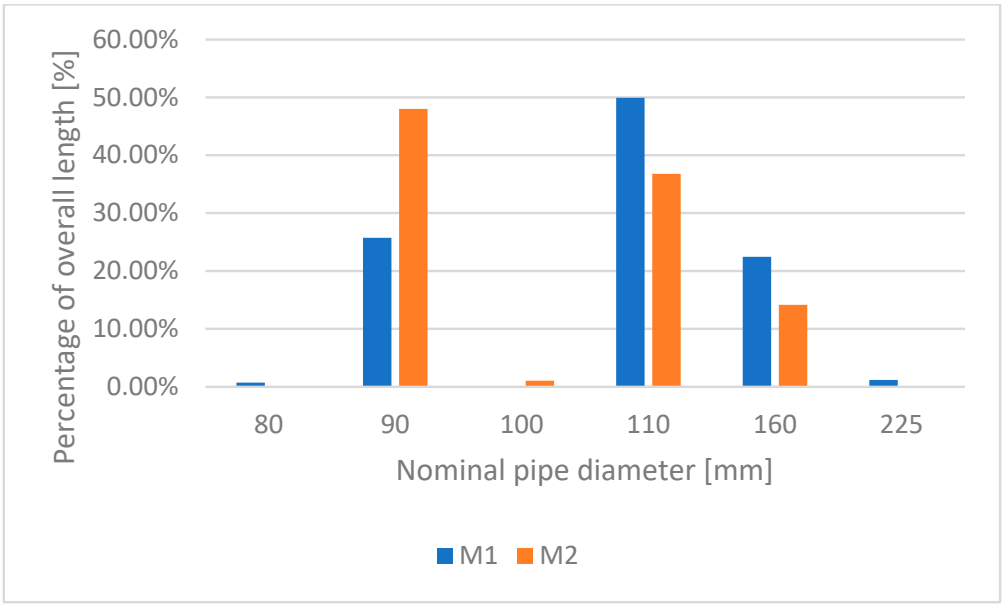
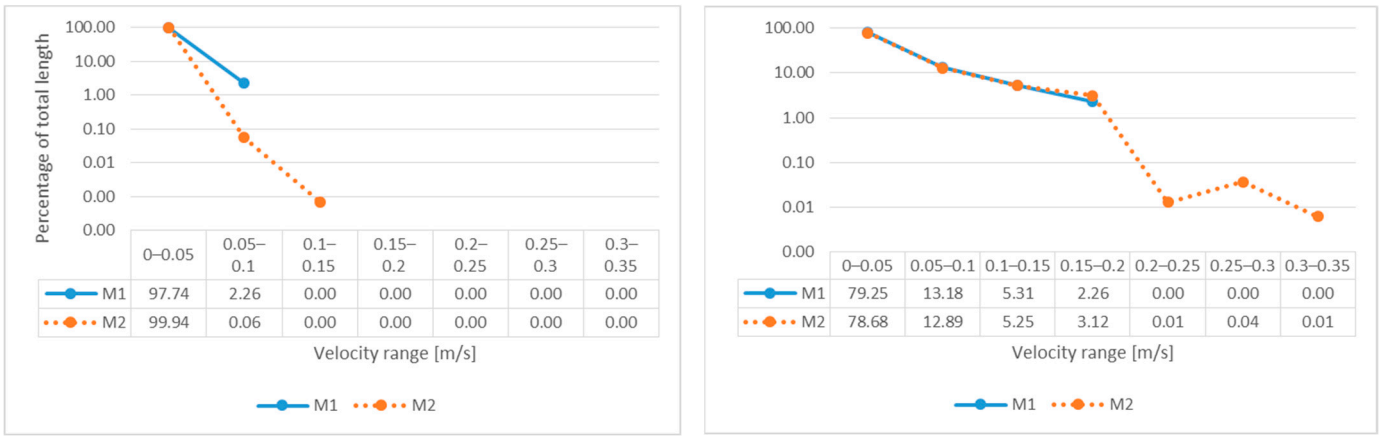


Figure 3. Structure of nominal diameters in the length of a network for M1 and M2 models.



(**a**) (**b**)

Figure 4. Distribution of velocity in M1 and M2 water distribution networks (**a**) hour of minimum water consumption; (**b**) hour of maximum water consumption.

Figure 5 represents a comparison of the average daily demand range for analysed models. The data show that the majority of the average daily water demand for both M1 and M2 networks falls within the 0–1 m^3/d range, with M1 at 94.66% and M2 slightly lower at 92.40%. There is a noticeable drop in demand as the range increases, with M1 generally having a higher percentage in lower ranges and M2 seeing a slight increase at the 1–2 m^3/d range. Demands above 2 m^3/d are rarely present for both networks, and there are no demand points in the 6–7 m^3/d range and above for M1, while M2 has some demand up to the 15–16 range, albeit it is rarely encountered. Notably, there is demand for M1 in the >20 m^3/d range, which is absent in M2. Overall, both networks show a concentrated demand in the lower ranges, significantly tapering off as the demand range increases.

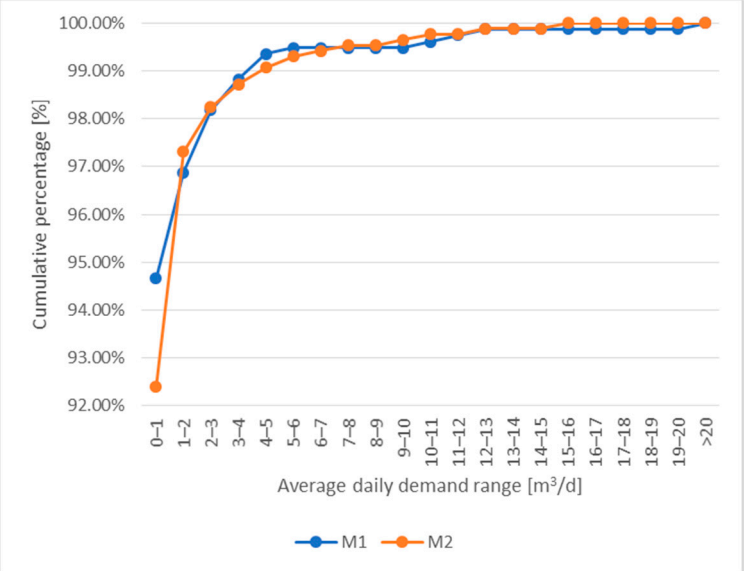


Figure 5. Cumulative distribution of the average daily demand for M1 and M2 networks.

3. Results

Table 2 presents a statistical summary of the topological distance to the actual leak location in a water network, based on the selected method for identifying leak candidates, scenario, and statistical similarity measure, thereby encapsulating the efficacy of various leak detection methodologies within the M1 water distribution network.

Table 2. Topological distance to real leak location depending on scenario, statistical measure, and method of leak candidate selection for M1 water network.

Scenario	Statistical Measure	Single Candidate	Candidate Clusters			
		Distance to Real Leak Location (m)	Number of Clusters	Minimum Distance to Real Leak Location (m)	Mean Distance to Real Leak Location (m)	Maximum Distance to Real Leak Location (m)
M1-A-1	correlation	355.67	9	250.83	354.34	430.16
M1-A-1	ang. clos.	602.98	7	478.01	587.91	723.84
M1-A-1	eucl. dist.	250.53	5	219.39	376.50	494.20
M1-A-1	manh. dist	240.39	2	207.52	223.95	240.39
M1-A-1	cheb. dist.	357.61	4	189.20	248.35	326.22
M1-A-1	cos. sim.	602.98	7	478.01	587.91	723.84
M1-A-1	sp. corr.	250.83	11	250.83	503.54	932.24
M1-A-2	correlation	923.96	7	594.48	858.00	1068.13
M1-A-2	ang. clos.	658.72	6	501.98	569.25	658.72
M1-A-2	eucl. dist.	386.28	11	250.83	462.06	893.99
M1-A-2	manh. dist	325.08	9	250.83	406.18	636.89
M1-A-2	cheb. dist.	250.83	11	250.83	400.15	533.46
M1-A-2	cos. sim.	658.72	6	501.98	569.25	658.72
M1-A-2	sp. corr.	2004.56	12	1092.09	1446.76	2004.56

Table 2. *Cont.*

Scenario	Statistical Measure	Single Candidate	Candidate Clusters			
		Distance to Real Leak Location (m)	Number of Clusters	Minimum Distance to Real Leak Location (m)	Mean Distance to Real Leak Location (m)	Maximum Distance to Real Leak Location (m)
M1-A-3	correlation	986.20	11	893.38	1052.14	1230.12
M1-A-3	ang. clos.	559.61	12	548.67	4297.19	4776.07
M1-A-3	eucl. dist.	706.72	9	583.02	1031.31	2004.56
M1-A-3	manh. dist	703.28	9	583.02	1031.31	2004.56
M1-A-3	cheb. dist.	614.23	9	583.02	1090.84	2004.56
M1-A-3	cos. sim.	559.61	12	548.67	4297.19	4776.07
M1-A-3	sp. corr.	665.00	11	665.00	966.27	1098.13
M1-B-1	correlation	277.37	9	203.34	329.62	459.53
M1-B-1	ang. clos.	404.70	9	310.33	379.52	448.31
M1-B-1	eucl. dist.	394.06	7	348.13	397.29	476.02
M1-B-1	manh. dist	394.06	7	325.62	386.03	476.02
M1-B-1	cheb. dist.	448.90	7	348.13	397.29	476.02
M1-B-1	cos. sim.	404.70	9	310.33	379.52	448.31
M1-B-1	sp. corr.	983.96	8	66.79	398.06	1024.44
M1-B-2	correlation	277.37	9	203.34	314.74	439.14
M1-B-2	ang. clos.	943.66	5	809.86	857.05	943.66
M1-B-2	eucl. dist.	423.08	10	233.50	374.14	521.55
M1-B-2	manh. dist	1332.36	8	194.78	437.05	1332.36
M1-B-2	cheb. dist.	772.09	7	751.15	850.97	990.31
M1-B-2	cos. sim.	943.66	5	809.86	857.05	943.66
M1-B-2	sp. corr.	949.93	10	40.69	702.51	1282.36
M1-C-1	correlation	1058.98	9	1027.28	1147.95	1265.72
M1-C-1	ang. clos.	1137.18	6	951.13	1023.22	1124.24
M1-C-1	eucl. dist.	833.35	6	778.96	881.38	1019.22
M1-C-1	manh. dist	833.35	9	833.35	1393.33	2470.89
M1-C-1	cheb. dist.	962.95	10	927.50	1141.78	1311.17
M1-C-1	cos. sim.	1137.18	6	951.13	1023.22	1124.24
M1-C-1	sp. corr.	1260.36	9	438.21	1138.73	1462.94

Table 3 presents a statistical summary of the topological distance to the actual leak location in a water network, based on the selected method for identifying leak candidates, scenario, and statistical similarity measure, thereby encapsulating the efficacy of various leak detection methodologies within the M2 water distribution network.

In both Tables 2 and 3, the topological distance to the actual leak location is presented for the method of selecting a single leak candidate point. For the group leak candidates obtained from the density-based spatial clustering of applications with noise algorithm, the number of clusters representing potential leak hotspots is shown, along with the calculated minimum, average, and maximum topological distance for the group of clusters generated in each scenario.

Table 3. Minimum topological distance to real leak location depending on scenario, statistical measure, and leak candidate selection for M2 water network.

Scenario	Statistical Measure	Single Candidate	Candidate Clusters			
		Distance to Real Leak Location (m)	Number of Clusters	Minimum Distance to Real Leak Location (m)	Mean Distance to Real Leak Location (m)	Maximum Distance to Real Leak Location (m)
M2-A-P	correlation	500.00	27	300.00	2956.73	5519.74
M2-A-P	ang. clos.	500.00	24	500.00	4082.40	5517.01
M2-A-P	eucl. dist.	200.01	18	100.00	1295.86	3770.37
M2-A-P	manh. dist	3770.37	18	100.00	1295.86	3770.37
M2-A-P	cheb. dist.	400.00	20	100.00	850.59	1693.94
M2-A-P	cos. sim.	500.00	24	500.00	4082.40	5517.01
M2-A-P	sp. corr.	3239.06	23	100.00	1789.66	5519.74
M2-A-F	correlation	5807.37	22	4836.80	5641.78	6644.65
M2-A-F	ang. clos.	6394.38	22	4836.80	5801.41	7710.97
M2-A-F	eucl. dist.	3350.18	17	2425.37	3198.57	3866.49
M2-A-F	manh. dist	3363.66	19	2684.81	3406.27	3914.85
M2-A-F	cheb. dist.	5026.58	18	4261.08	4902.29	5561.90
M2-A-F	cos. sim.	6394.38	22	4836.80	5801.41	7710.97
M2-A-F	sp. corr.	6394.38	30	1678.38	3120.92	6394.38

3.1. Leak Candidate Selection Based on Pressure Measurements

The subset of the dataset examining the efficacy of isolating a single candidate for leak detection in water distribution networks through pressure measurements displays a pronounced variability in topological distances to real leak locations. This variability is closely associated with the statistical measure employed and the specific scenarios under consideration. In the context of the real case scenario that occurred in the M2 water network, the Euclidean distance achieved the highest precision, approximating the leak location within 200 m. For the M1-A-1 scenario, the Manhattan distance metric showed the highest localization accuracy at approximately 240 m in the M1-A-1 scenario. This scenario also displayed closely matched precision levels for the Euclidean distance (250.53 m) and Spearman correlation (250.83 m). In the M1-A-2 scenario, a similar level of precision was achieved at 250.83 m using the Chebyshev distance, whereas the Spearman correlation had a significantly lower accuracy, indicating a distance of 2004.56 m. For M1-A-3, angular closeness provided the best accuracy. Relative to M1-A-1 and M1-A-2, this scenario exhibited a more than twofold increase in the minimum topological distance. A reduction in leak outflow intensity also correlated with a decrease in locational precision in scenario A for most of the metrics evaluated. In the scenarios of M1-B-1 and M1-B-2, the correlation measure consistently yielded the highest precision, localizing the leak with an accuracy of 277.37 m in both instances. Furthermore, the Euclidean distance consistently assured a stable precision of approximately 400 m to the actual leak source across varying outflow intensities in the detection of the source's location in M1-B. For the other metrics, the precision fluctuated between approximately 400 and 900 m. The M1-C-1 scenario presented the most significant challenge for leak detection, with the minimum distance to the actual leak being 833.35 m, as determined by both the Euclidean and Manhattan distances.

Figure 6 delineates the comparative performance of the statistical metrics analysed in the single-candidate leak isolation method for all of the scenarios assessed.

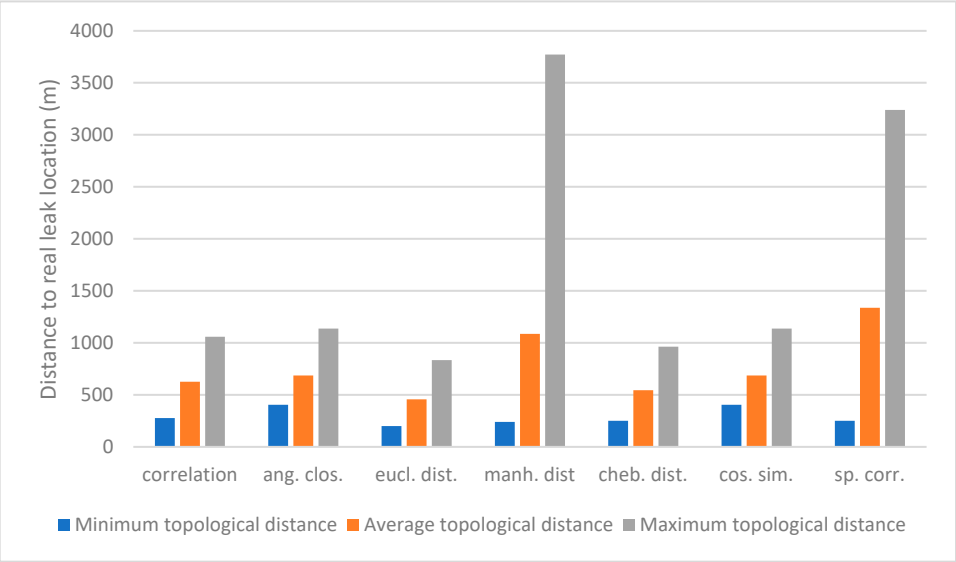


Figure 6. Performance comparison of analysed statistical metrics for all examined scenarios.

Figure 6 shows that the Euclidean distance achieved the highest level of precision with an approximate topological distance of 200 m to the actual leak site. It also had the lowest average distance to the genuine leak source, registered at 456.29 m, and the lowest maximum proximity of 833.35 m.

For the subset of the dataset concerning the delineation of groups of nodal points to identify leak hotspots, an analysis was conducted on parameters such as the number of clusters generated by the DBSCAN algorithm within a given scenario and the statistical similarity measure to determine the minimum, average, and maximum distances of the generated clusters to the actual leak location in the water network. In the scenario of M1-A-1, the highest precision was achieved using the Chebyshev distance—the nearest cluster was located at a distance of 189 m. A comparable level of precision was obtained with the Manhattan distance and the Euclidean distance, yet it is noteworthy that the Manhattan distance generated only two clusters in this scenario, representing the minimum number across all of the analysed scenarios and statistical measures. For the M1-A-2 scenario, the distance to the nearest cluster from the leak was higher, amounting to 250.83 m for the Euclidean distance, Manhattan distance, and Chebyshev distance, with the Manhattan distance again generating the smallest number of clusters among the aforementioned, totalling nine. In the scenario of M1-A-3, the lowest topological distance among the generated clusters was recorded for angular closeness, and all measures except for the correlation and Spearman correlation demonstrated similar efficacy levels at the 500–600 m range, with the Euclidean distance, Manhattan distance, and Chebyshev distance generating the smallest number of clusters at nine groups of points. In the other analysed scenarios, the Spearman correlation exhibited the highest precision, achieving an accuracy of 66.79 m in the scenario of M1-B-1, 40.69 m in M1-B-2, and 438.21 m in M1-C-1. For the M2 model in the M2-A-P scenario, similar results were obtained for the Euclidean distance, Manhattan distance, Chebyshev distance, and Spearman correlation at the 99.99 m range. Using the Spearman correlation, the lowest distance of the generated cluster to the actual leak location was observed in the scenario of M1-B-2. In the M1 model, the scenario of M1-C-1 once again proved to be the most challenging, and the distance of the nearest clusters to the real leak source exceeded 800 m.

A measure that illustrates the overall accuracy level of the group candidate identification method using clustering is the average topological distance from n number of individual clusters generated in a given scenario. Figure 7 displays the minimum of the average, the overall average, and the maximum value from the average distances of the clusters from each of the analysed scenarios, depending on the statistical measure.

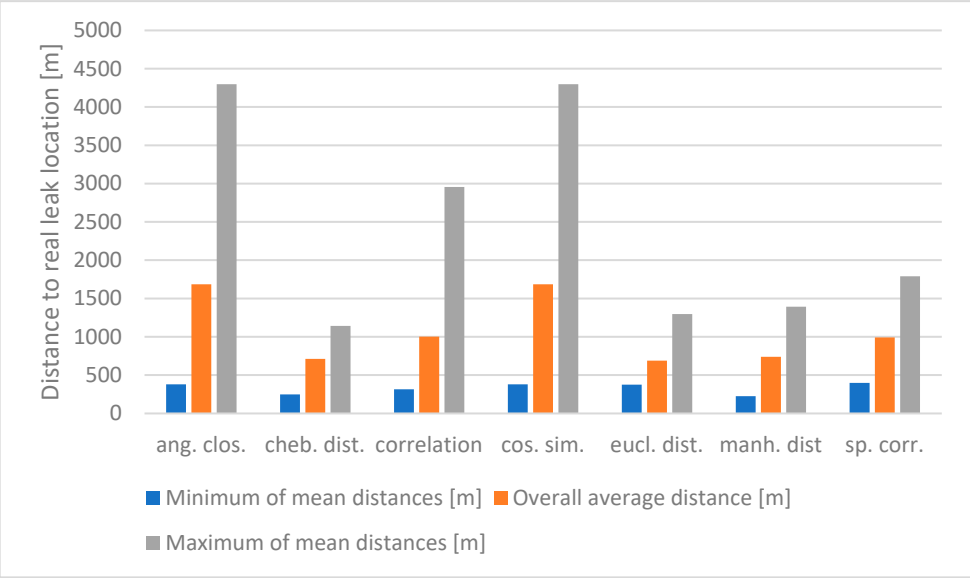


Figure 7. Statistical summary of the mean values of the topological distances to the actual leakage in the respective analysed scenarios.

Within the framework of average distances, the smallest mean value of the generated clusters was observed in the scenario of M1-A-1, with a distance of 223.96 m using the Manhattan distance. Regarding the overall average distance, the lowest value was recorded for the Euclidean distance at 688.36 m, with comparable results obtained for the Chebyshev distance at 711.42 m and the Manhattan distance at 739.10 m.

Figure 8 illustrates the minimum, maximum, and mean number of clusters generated by each statistical measure across all of the analysed scenarios. It is evident that the Manhattan distance produced the lowest number of clusters among the measures analysed, with a minimum of two clusters. On average, this measure, along with the Euclidean distance, generated the fewest clusters, averaging nine clusters. The maximum number of clusters was generated using the correlation parameter, reaching up to 27 clusters.

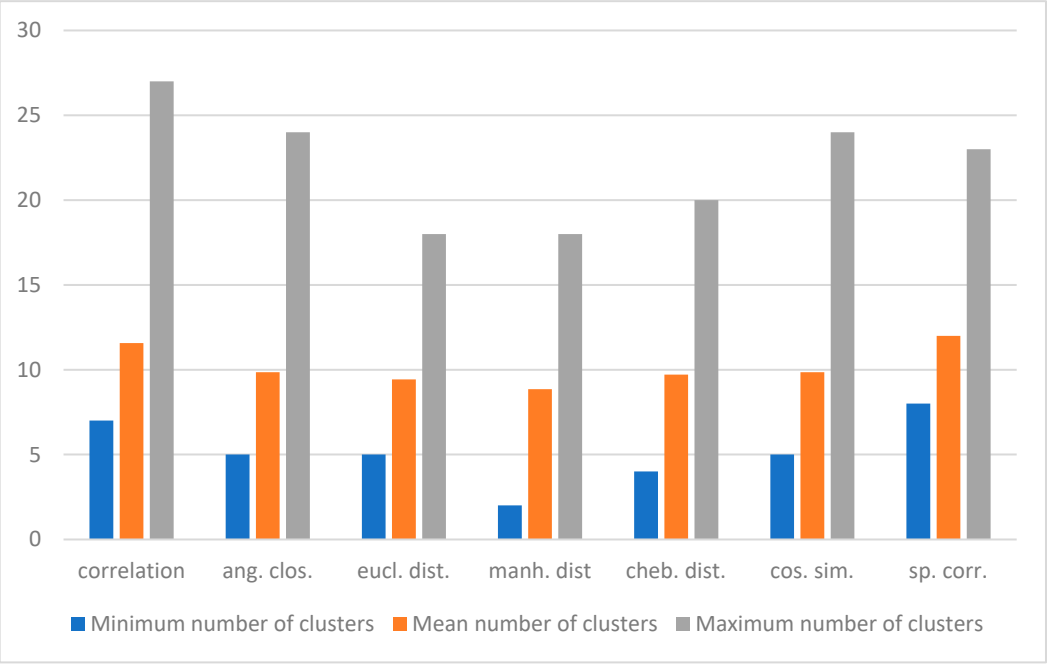


Figure 8. Number of clusters generated by each of analysed metrics.

3.2. Leak Candidate Selection Based on Hydrant Outflow Measurements

In the case of the M2 model, methods for determining leak candidates based on the measurements of two different parameters were compared—pressure measurements within the network and outflow measurements from hydrants obtained from hydrant tests to explore the potential use of data related to hydrant outflows. For the method utilizing hydrant outflows to identify a single candidate, the best result was achieved using the Euclidean distance, amounting to 3350 m from the actual leak source. On the other hand, by using the method indicating a single leak point based on pressure measurements, the minimum topological distance was 200 m for the same statistical measure. The average distance to the actual leak source for the method incorporating hydrant outflows was 5247.27 m, whereas, for the pressure measurements, the average was 1301.35 m. This average was significantly inflated by outlier values of the Manhattan distance and Spearman correlation, which exceeded 3000 m. In determining leak candidates through a group of points and identifying a leak hotspot, the method using the outflow recorded a minimum topological distance to the real leak source of 1678 m. Utilizing pressure measurements in the same scenario, the highest precision achieved among the identified clusters was 99.99 m to the actual leak point. This level was reached for most of the measures, excluding the angular closeness and cosine similarity. Regarding the number of generated clusters, the use of pressure measurements produced, for most of the statistical measures, nearly 6 to 22 percent more clusters compared to the methods based on hydrant outflows.

4. Discussion

The conducted research highlights that within the analysed scenarios, which included various leak locations and sizes, there was a diversity in the effectiveness of the similarity measures used to achieve the shortest topological distance to the actual leak location. When utilizing pressure measurements, the best results in pinpointing a single leak candidate were achieved with the Euclidean distance, as it consistently showed the closest proximity (approximately 200 m) to the actual leak site, the lowest average distance recorded at 456.29 m, and the smallest maximum distance of 833.35 m.

In the identification of potential leak areas through a group of points exhibiting the highest similarity and subsequent clustering with the DBSCAN method, both the Euclidean and Manhattan distances also demonstrated high levels of precision. These measures achieved the best accuracy at approximately 100 m, with the lowest average distance around 350 m and the smallest maximum distance around 800 m across the various analysed scenarios of leak locations and sizes.

The comparison of the methods for determining leak candidates—either by a single point or by a percentage of points with the highest similarity—revealed that on average, the DBSCAN method generated about 10 clusters, indicating areas with a heightened probability of leaks. It is important to note that the group method, when analysing the clusters closest to the actual leak source (minimum distance), an increase in the average precision for each of the statistical measure was observed ranging from 14.33% (angular closeness) to 56.66% (Spearman correlation). The average precision within the scenario also increased from 7.57% (in the scenario of M1-A-3) to 51.32% (in the M2-A-P scenario). Considering the average distance of clusters generated within the individual scenarios, there was generally a deterioration in precision from 33.79% (Chebyshev distance) to 191.84% (angular closeness) or a slight improvement—0.42% in the case of the Manhattan distance and 3.11% for the Spearman correlation. Within the scenarios, average precision improvements occurred only in the scenarios of M1-B-1 (9.41%) and M1-B-2 (14.21%).

Comparing a different approach that involved using hydrant outflow measurements instead of the traditional method of pressure measurements, the method based on hydrant outflows was significantly less precise within the analysed scenario. On average, using pressure-based measurements achieved a 10 to 22 times greater accuracy in leak detection than using hydrant outflow measurements. This comparatively lower level of effectiveness may be attributed to the method's lesser sensitivity to leaks within the investigated range.

Another potential cause is the need to use the emitter coefficient to simulate hydrant outflow. The adopted coefficient is based on literature data, which may differ from actual conditions. This discrepancy is particularly relevant in rural areas, where hydrants are often old and poorly maintained, leading to variations in their performance characteristics that are not accurately captured by standard coefficients. Additionally, the use of an uncalibrated model may further contribute to these errors.

Research has shown that the use of hydraulic modelling and sensitivity analyses provides promising results in leak detection, which may depend on factors such as the sensor placement on the network, leak magnitude, the location of the leak relative to the measurement points, the structure of the water network, and the quality of measurement data. The achieved accuracy of the pressure-based method indicates that it can be successfully implemented in real-world conditions. The entry threshold required to apply the method is relatively low—it necessitates having flow rate measurements at the outlet of all water treatment stations, data on the average daily water consumption by customers, basic information about the network layout to build a model in free software, e.g., EPANET 2.2 and WNTR, and map data, which can usually be obtained for free from water utility companies.

The scalability of methods based on hydraulic modelling and sensitivity analyses is noteworthy. When operating on a limited budget, sensors can be primarily placed in areas most prone to leaks, such as old, faulty, and problematic networks or high-pressure areas, and then expanded over time with additional measurement points. Similarly, during the initial phase of implementation, the uncalibrated or basically calibrated model can be utilized as in this study, with further opportunities to increase the model's accuracy, while also improving leak localization precision. Compared to other methods, e.g., data-driven methods, there is no need to collect historical data, and the method can be implemented faster. Referring to the DMA zoning, which is traditionally used in water utilities, hydraulic modelling and sensitivity analysis methods can be implemented at much lower costs, especially when the company can build the network model on its own. Additionally, pressure measurements are simpler to carry out, e.g., on hydrants (also as temporary points), and do not require network intervention, the installation of flow meters, digging, pavement reconstruction, formal agreements, or prior project development.

On the other hand, with more data and a greater budget, it is possible to embed the discussed methods as part of an integrated system. Hydraulic-modelling-based methods can be used with data from DMA zone points, narrowing the search area for leaks to a specific zone, or with real-time data from IoT water meter modules, which may provide more precise data on the structure and size of consumer water demand, increasing the model's accuracy and providing greater responsiveness in leak management. Using information on potential leak locations with clustering methods, e.g., DBSCAN, teams equipped with acoustic leak detection devices, such as digital multi-correlators, can be coordinated. This highlights that advanced technologies and analytical tools play a crucial role in enhancing our understanding and management of water resources. These technologies enable precise assessments and predictive simulations, which are essential for addressing contemporary water-related challenges [31].

5. Conclusions

Based on the conducted research, it can be asserted that employing leak detection methods based on hydraulic modelling and a sensitivity analysis enables the approximate identification of probable leak locations with a precision under 100 m for the M2 model and under 200 m for the M1 model. Notably, these results were achieved using temporary pressure measurements and uncalibrated mathematical models. This suggests that even with very limited financial resources, a significant narrowing of the search area for leaks is achievable in rural and urban–rural water supply systems, which can lead to reduced real losses and the enhanced reliability of water supply systems.

When utilizing the DBSCAN clustering method to find the group of nodes with the highest probability of leakage, a different form of result is obtained. While clustering improves

levels of precision by identifying a cluster close to the actual leak, it also introduces spatial variability, as some clusters may be further from the target, thereby expanding the range of potential leak locations. Such an approach can help water utility teams in developing strategies for leak localisation, prioritizing clusters that on average show the highest similarity of sensitivity vectors to residual vectors. Future research could explore experimenting with various percentage values while choosing leak group candidates or testing different maximum distance parameters for DBSCAN clustering to achieve an optimal balance between increasing precision levels and minimizing the number of generated clusters.

Regarding the results obtained, both for the single point location and group point methods, the best average results were achieved using the Euclidean distance.

Using different methods based on hydrant flow measurements is not advisable considering the much worse levels of precision achieved; in addition, they are more computationally expensive and require greater amounts of fieldwork.

The scalability of hydraulic modelling and sensitivity analysis methods is significant, especially for budget-constrained operations. Sensors can initially be placed in areas with an increased risk of leakage and expanded over time. Uncalibrated models, as used in this study, can be improved gradually, enhancing their leak detection accuracy. The described methods are characterised by relatively low costs and the possibility of rapid implementation. For water companies whose budget is not an issue, these methods can be integrated into advanced systems with real-time IoT data and clustering techniques, improving overall leak management and responsiveness.

Author Contributions: Conceptualization, Ł.W. and A.Z.; methodology, Ł.W.; software, Ł.W.; validation, I.B., P.T. and A.Z.; formal analysis, P.T.; investigation, I.B.; resources, Ł.W.; data curation, A.Z.; writing—original draft preparation, Ł.W.; writing—review and editing, A.Z.; visualization, A.Z.; supervision, I.B.; project administration, P.T.; funding acquisition, I.B. All authors have read and agreed to the published version of the manuscript.

Funding: The research was carried out as part of work No. WZ/WB-IIŚ/3/2022 at the Białystok University of Technology and financed by a research subsidy provided by the Minister of Education and Science.

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.

Conflicts of Interest: Author Piotr Tuz was employed by the company Water Energy Systems (WES). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Lambert, A.O. International Report: Water losses management and techniques. *Water Supply* **2002**, *2*, 1–20. [CrossRef]
2. EurEau. Europe's Water in Figures—An Overview of the European Drinking Water and Waste Water Sectors. Available online: https://www.eureau.org/resources/publications/eureau-publications/5824-europe-s-water-in-figures-2021/file (accessed on 5 April 2024).
3. Karney, B. Impacts of Leaks on Energy Consumption in Pumped Systems with Storage. *J. Water Resour. Plan. Manag.* **2005**, *131*, 146–155.
4. Zib, L.; Byrne, D.M.; Marston, L.T.; Chini, C.M. Operational carbon footprint of the U.S. water and wastewater sector's energy consumption. *J. Clean. Prod.* **2021**, *321*, 128815. [CrossRef]
5. International Energy Agency. *World Energy Outlook*; OECD/IEA: Paris, France, 2009.
6. Bozkurt, C.; Ates, A.; Fırat, M.; Yılmaz, S.; Özdemir, Ö. A Novel Strategic Water Loss Management Model and Its Optimization with Harris Hawk Algorithm. *Water Resour. Manag.* **2024**, *38*, 1543–1561. [CrossRef]
7. Oviedo-Ocaa, E.R.; Dominguez, I.C.; Celis, J.; Blanco, L.C.; Cotes, I.; Ward, S.; Kapelan, Z. Water loss management under data scarcity: Case study in a small municipality in a developing country. *J. Water Resour. Plan. Manag.* **2020**, *146*, 13. [CrossRef]
8. Piasecki, A. Water and Sewage Management Issues in Rural Poland. *Water* **2019**, *11*, 625. [CrossRef]
9. Meireles, I.; Sousa, V.; Matos, J.P.; Cruz, C.O. Determinants of water loss in Portuguese utilities. *Util. Policy* **2023**, *83*, 101603. [CrossRef]
10. Oberascher, M.; Möderl, M.; Sitzenfrei, R. Water Loss Management in Small Municipalities: The Situation in Tyrol. *Water* **2020**, *12*, 3446. [CrossRef]

11. Pudar, R.S.; Liggett, J.A. Leaks in Pipe Networks. *J. Hydraul. Eng.* **1992**, *118*, 1031–1046. [CrossRef]
12. Pérez, R.; Puig, V.; Pascual, J.; Peralta, A.; Landeros, E.; Jordanas, L.L. Pressure sensor distribution for leak detection in Barcelona water distribution network. *Water Supply* **2009**, *9*, 715–721. [CrossRef]
13. Pérez, R.; Puig, V.; Pascual, J.; Quevedo, J.; Landeros, E.; Peralta, A. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Eng. Pract.* **2011**, *19*, 1157–1167. [CrossRef]
14. Steffelbauer, D.; Neumayer, M.; Günther, M.; Fuchs-Hanusch, D. Sensor Placement and Leakage Localization Considering Demand Uncertainties. *Procedia Eng.* **2014**, *89*, 1160–1167. [CrossRef]
15. Perez, R.; Quevedo, J.; Puig, V.; Nejjari, F.; Cugueró, M.A.; Sanz, G.; Mirats, J.M. Leakage isolation in water distribution networks: A comparative study of two methodologies on a real case study. In Proceedings of the 2011 19th Mediterranean Conference on Control & Automation (MED), Corfu, Greece, 20–23 June 2011; pp. 138–143. [CrossRef]
16. Pérez, R.; Sanz, G.; Puig, V.; Quevedo, J.; Escofet, M.A.C.; Nejjari, F.; Meseguer, J.; Cembrano, G.; Tur, J.M.M.; Sarrate, R. Leak Localization in Water Networks: A Model-Based Methodology Using Pressure Sensors Applied to a Real Network in Barcelona [Applications of Control]. *IEEE Control Syst.* **2014**, *34*, 24–36. [CrossRef]
17. Pérez, R.; Cugueró, M.A.; Cugueró, J.; Sanz, G. Accuracy Assessment of Leak Localisation Method Depending on Available Measurements. *Procedia Eng.* **2014**, *70*, 1304–1313. [CrossRef]
18. Casillas Ponce, M.V.; Garza Castañón, L.E.; Cayuela, V.P. Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. *J. Hydroinformatics* **2014**, *16*, 649–670. [CrossRef]
19. Casillas Ponce, M.V.; Garza Castanon, L.E.; Puig, V. Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks: Application to the Barcelona network. In Proceedings of the 2013 European Control Conference (ECC), Zurich, Switzerland, 17–19 July 2013; pp. 401–409. [CrossRef]
20. Rosich, A.; Puig, V.; Casillas, M.V. Leak localization in drinking water distribution networks using structured residuals. *Adapt. Control Signal* **2015**, *29*, 991–1007. [CrossRef]
21. Casillas, M.; Garza-Castañón, L.; Puig, V.; Vargas-Martinez, A. Leak Signature Space: An Original Representation for Robust Leak Location in Water Distribution Networks. *Water* **2015**, *7*, 1129–1148. [CrossRef]
22. Blesa, J.; Pérez, R. Modelling uncertainty for leak localization in Water Networks. *IFAC-PapersOnLine* **2018**, *51*, 730–735. [CrossRef]
23. Ferreira, B.; Antunes, A.; Carriço, N.; Covas, D. Comparison of model-based techniques for pipe burst location in water distribution networks. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, *1136*, 012039. [CrossRef]
24. Marzola, I.; Alvisi, S.; Franchini, M. A Comparison of Model-Based Methods for Leakage Localization in Water Distribution Systems. *Water Resour. Manag.* **2022**, *36*, 5711–5727. [CrossRef]
25. Jiménez-Cabas, J.; Romero-Fandiño, E.; Torres, L.; Sanjuan, M.; López-Estrada, F.R. Localization of Leaks in Water Distribution Networks using Flow Readings. *IFAC-PapersOnLine* **2018**, *51*, 922–928. [CrossRef]
26. Geng, Z.; Hu, X.; Han, Y.; Zhong, Y. A Novel Leakage-Detection Method Based on Sensitivity Matrix of Pipe Flow: Case Study of Water Distribution Systems. *J. Water Resour. Plann. Manag.* **2019**, *145*, 04018094. [CrossRef]
27. Salguero, F.J.; Cobacho, R.; Pardo, M.A. Unreported leaks location using pressure and flow sensitivity in water distribution networks. *Water Supply* **2019**, *19*, 11–18. [CrossRef]
28. Klise, K.; Hart, D.; Bynum, M.; Hogge, J.; Haxton, T.; Murray, R.; Burkhardt, J. *Water Network Tool for Resilience (WNTR)*; User Manual, Version 0.2.3., 2020:SAND–2020-9301R, EPA/600/R--20/185; U.S. Department of Energy Office of Scientific and Technical Information: Washington, DC, USA, 2020; p. 1660790. [CrossRef]
29. Rossman, L.A.; Woo, H.; Tryby, M.; Shang, F.; Janke, R.; Haxton, T. *EPANET 2.2 User Manual*; U.S. Environmental Protection Agency: Washington, DC, USA, 2020; Volume EPA/600/R-20/133.
30. Haestad, M.; Walski, T.; Chase, D.; Savic, D.; Grayman, W.; Backwith, S.; Koelle, E. *Advanced Water Distribution Modeling and Management*; Haestad Press: Waterbury, CT, USA, 2003.
31. Pietrucha-Urbanik, K.; Rak, J. Water, Resources, and Resilience: Insights from Diverse Environmental Studies. *Water* **2023**, *15*, 3965. [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.