A Comparison of Model-Based Methods for Leakage Localization in Water Distribution Systems

Irene Marzola[1] · Stefano Alvisi[1] · Marco Franchini[1]

Abstract

Model-based methods for leakage localization in water distribution systems have recently been gaining more attention. These methods identify the leakage position by comparing the measured network data with the corresponding values simulated by a hydraulic model. In this study two model-based methods already proposed in literature, one based on the Sensitivity Matrix method and the other one on the Linear Approximation method, are analysed and compared to each other. The methods are applied to the same case study network, exploiting only data provided by pressure sensors. Various analyses are undertaken in order to investigate the main critical issues tied to the two methods, i.e. a) the use of different amounts of data averaged over different time windows, b) the impact of the model's accuracy in terms of water demands and pipe roughness, and c) the effect of the number of pressure measuring points. The results show that higher efficiency is obtained by considering the hourly averaged data all together. Moreover, the Linear Approximation method is on average 3 times more accurate than the Sensitivity Matrix when a perfect hydraulic model is used, even with a reduced number of pressure sensors. However, when a hydraulic model and/or measured data affected by errors are considered, the Sensitivity Matrix is more accurate, with an average error almost 10% lower than the Linear Approximation.

1 Introduction

Leakages in water distribution systems are a common phenomenon and are becoming a major concern for water utilities (Alvisi et al. 2019). The presence of leakages leads to repercussions not only for the water utilities themselves, which face economic losses, and

✉ Irene Marzola
  irene.marzola@unife.it

  Stefano Alvisi
  stefano.alvisi@unife.it

  Marco Franchini
  marco.franchini@unife.it

1   Department of Engineering, University of Ferrara, Via Saragat 1, 44122 Ferrara, Italy

for users, as they may suffer from irregularities in the water supply, but also for the environment, as water is not an inexhaustible resource (Marzola et al. 2021). Therefore, minimizing water losses is a priority in order to increase the efficiency of water distribution systems. However, leakage localization is a serious challenge. In an attempt to overcome this challenge several methods have been developed in order to determine the position of the leakage, and, within this framework, data-driven methods and model-based methods are recently gaining attention and becoming widespread, as highlighted in Hu et al. (2021). Data-driven methods are gaining popularity thanks to improvements in the field of data acquisition and the continuous advancement of technologies (Li et al. 2015; Pacchin et al. 2019). In fact, with data-driven techniques, the acquired data are processed in order to identify values that deviate from the normal pattern of flow and pressure data and may be caused by leakage formation. Simple statistical analysis can be applied on the acquired data to identify the outliers, i.e. the induced changes due to leakage presence (Lee et al. 2016; Louriero et al. 2016).

Otherwise, models such as artificial neural networks or decision trees can be trained with the historical data acquired by the water utilities to identify the influence of leakages by either classifying the actual data (Mounce and Machell 2006; Aksela et al. 2009) or predicting the values of pressure and flow that the network should display under normal conditions and comparing them with the actual values (Mounce et al. 2002). However, especially as regards the latter two cases, a long computational time and high capacity are required, as well as a huge amount of historical data relating to both standard situations and different cases of leakage; such data are not always at the water utilities' disposal (Wu and Liu 2017; Chan et al. 2018).

Alternatively, in model-based methods the leakage position is determined by using the network data measured through pressure and flow sensors with a low sampling rate (i.e. 5 minutes, 15 minutes), and the corresponding data simulated through a hydraulic model. Generally, pressure sensors are used for monitoring the network, as they are cheaper than flow meters, and can be easily installed in several points and evenly distributed.

More in detail, the measured data are compared with their corresponding values simulated by a hydraulic model, which has previously been accurately calibrated in terms of both network characteristics and demand patterns. Normally, measured and simulated values are in good agreement if the model is accurate and realistically represents the network behaviour. However, if a new leakage occurs in the network, for which neither the flow rate nor the position is known, the measured pressures will no longer match the corresponding simulated values. In fact, leakage increases the network head losses, leading to lower values of the measured pressure. However, even though the flow rate and position of the leakage are unknown, the hydraulic model can be used to simulate different leakage scenarios and the leakage will be localized in the network node or pipe that allows the difference between the simulated and corresponding measured data to be minimized.

To this end different strategies can be used such as sensitivity matrix methods (Perez et al. 2014; Ponce et al. 2014), mixed model-based/data-driven methods (Javadiha et al. 2019; Li et al. 2021, 2022), optimization-calibration methods (Wu et al. 2010; Berglund et al. 2017; Moasheri and Jalili-Ghazizadeh 2020), and error domain falsification methods (Goulet et al. 2013; Moser et al. 2018).

In any case, regardless of the strategy used, the main concept of the model-based approach lies in a comparison between measured and simulated data. However, because of this very concept, there are several critical issues that can affect the methods' performance. First of all, the available time series of data, collected at time step in the order of 5 - 15

minutes, may be quite long. In the comparison between measured and simulated data, it is still not clear how these data should be managed, i.e. whether it is necessary to use them all, to use only the data from particular periods, such as the periods of lowest or highest consumption, or to use averaged data (Berglund et al. 2017). Another issue is related to the level of accuracy of both the hydraulic model and the information used as inputs — the former in terms of demand patterns and the network topology and the latter in terms of the measured data — as the efficiency of model-based methods is highly sensitive to modelling and measurement errors (Perez et al. 2014; Adedeji et al. 2017; Hu et al. 2021). Lastly, network monitoring is a fundamental requirement, but the budget of water utilities generally does not allow for the installation of pressure sensors at every potential measurement location (Soroush and Abedini 2019), and a low amount of available data can influence the results obtained (Perez et al. 2009).

Moreover, as reported in the scientific literature, the various proposed model-based methods have been applied to different case studies, some synthetically generated with a hydraulic model (Kang and Lansey 2014), some using tests based on engineered events, i.e. by considering the real network and artificially creating leakages by opening fire hydrants (Farley et al. 2013; Perez et al. 2014), and others still based on actual historical data (Wu et al. 2010; Sophocleous et al. 2019), making it difficult to fairly compare the various methods.

To the best of the authors' knowledge, no prior studies have compared model-based methods by taking into consideration all the aforementioned critical points and examining their effects on localization accuracy. Therefore, this paper aims to highlight pro and cons of the model-based methods for leakage localization according to the different characteristic of the data and model available to the water utility. For this purpose, two model-based methods already proposed in the literature are considered and applied to analyse all the previously highlighted issues and carry out a fair comparison. Indeed, these two methods have been selected for their effectiveness on leakage localization as highlighted within the framework of the Battle of the Leakage Detection and Isolation Methods (Daniel et al. 2022; Steffelbauer et al. 2022).

In particular, the first method, originally proposed by Perez et al. (2014), belongs to the category of sensitivity matrix methods according to the classification described above, and is hereinafter named Sensitivity Matrix method. The second method is a variant of the Linear Approximation method proposed by Berglund et al. (2017), belonging to the category of optimization-calibration methods. Both methods are used to localize the leakage, but the Sensitivity Matrix method requires information about the leakage flow rate, whereas the Linear Approximation method provides an estimate of the latter.

Here they are applied to the same water distribution network, the subject of a case study proposed within the framework of the Battle of the Leakage Detection and Isolation Methods (BattLeDIM), i.e. a network monitored with pressure sensors and for which a perfect hydraulic model is available. Their efficiency and performance are firstly analysed considering four different approaches to the use of pressure datasets. Then, with the approach that provided the best results, further analyses are carried out by varying the accuracy of the hydraulic model in terms of demand patterns and pipe characteristics, the accuracy of the input data, and the number of pressure monitoring points.

The following section describes the two methods, the case study network and the application of the methods. The results obtained are then examined and some conclusions are drawn.

2 Materials and Methods

2.1 Model-Based Methods

In order to describe the two model-based methods compared in this work, some preliminary assumptions are required.

Let us consider a hydraulic network, composed of a total of n_{tot} pipes, which is monitored by n_{sen} pressure sensors and let i indicate the generic node at which the pressure sensor is positioned (with $i = 1 \ldots n_{sen}$). For the same system a perfect hydraulic model is available, i.e. a model that perfectly represents the behaviour of the network, with accurate pipe characteristics and water demand patterns. Let $p_{i,measured}$ be the pressure measured at the generic node i at a generic time instant. Let $p_{i,baseline}$ be the corresponding pressure at the same node i and at the same time instant, simulated by the hydraulic model and not including the leakage to be localized. Clearly, as the position is unknown, the leakage cannot be allocated in the model in the correct position.

2.1.1 Sensitivity Matrix

The Sensitivity Matrix (SM) method, based on the approach proposed by Perez et al. (2014) and applied by Steffelbauer et al. (2022) in the framework of the BattLeDIM, enables a leakage of a known flow rate f to be localized. In general, several scenarios are considered with the hydraulic model by simulating the leakage in a different position each time. The pressures obtained are then compared with the corresponding measured pressures, and the scenario that allows the differences between these values to be minimized is identified.

In practice, as schematized in Fig. 1a, first the residual vector r is determined by the differences between the measured pressure $p_{i,measured}$ and the simulated pressure $p_{i,baseline}$ for each node i:

$$r = \begin{bmatrix} p_{1,measured} - p_{1,baseline} \\ \ldots \\ p_{i,measured} - p_{i,baseline} \\ \ldots \\ p_{n_{sen},measured} - p_{n_{sen},baseline} \end{bmatrix} \tag{1}$$

Next, all the pipes j (with $j = 1 \ldots n_{tot}$) of the network are considered individually one by one. Given that the leakage flow rate f must be known in advance, a hydraulic simulation is performed by locating the leakage in the pipe j, i.e. by distributing the flow rate equally in the two corresponding adjacent nodes. The pressure $p_{i,perturbed_j}$, for each node i with the leakage associated with the pipe j is thus obtained, and is defined as *perturbed* because it results from the addition of the leakage. Finally, a sensitivity matrix S is evaluated. This matrix has as many rows as the number of pressure sensors (as the residual vector r) and n_{tot} columns, each corresponding to a different pipe j. Each element S_{ij} measures the effect that the leakage f positioned in the pipe j has on the pressure at the nodes i, as defined below:
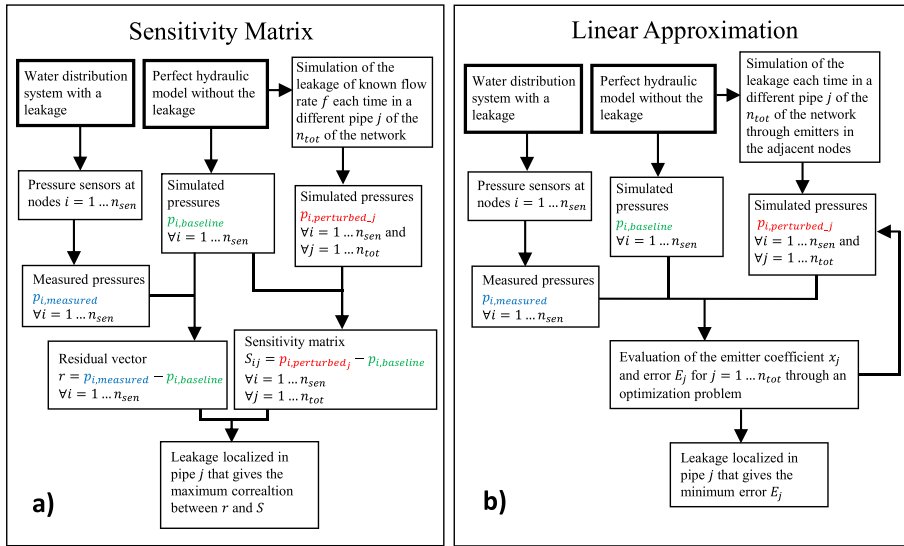
Fig. 1 Diagrams of **a** the Sensitivity Matrix and **b** the Linear Approximation methods

$$S = \begin{bmatrix} \dfrac{p_{1,perturbed_1} - p_{1,baseline}}{f} & \cdots & \dfrac{p_{1,perturbed_{n_{tot}}} - p_{1,baseline}}{f} \\ \cdots & \cdots & \cdots \\ \dfrac{p_{n_{sen},perturbed_1} - p_{n_{sen},baseline}}{f} & \cdots & \dfrac{p_{n_{sen},perturbed_{n_{tot}}} - p_{n_{sen},baseline}}{f} \end{bmatrix} \quad (2)$$

The correlation coefficients between the residual vector r and each column of the matrix S is then calculated. The leakage will be localized in the pipe j relative to the column of S that gives the highest correlation value, namely, the pipe that allows the differences between measured and simulated pressures (with the leakage positioned) to be minimized.

2.1.2 Linear Approximation

The second leakage localization method is based on the approach called Linear Approximation (LA), first proposed by Berglund et al. (2017) and then applied by Daniel et al. (2022) in the framework of the BattLeDIM. Indeed, this method was originally aimed only at evaluating the leakage flow rate, but was modified in order also to determine the position of the leakage in addition to the flow rate. In general, the leakage is simulated each time in a different position within the hydraulic model, and the corresponding simulated pressures are obtained. As the flow rate is unknown, it is initially chosen randomly. Then, for each scenario, a new flow rate is considered with the aim of minimizing the difference between the measured and simulated pressures. The scenario that, with the new assumed flow rate, shows the lowest error of all provides the leakage position.

More in detail (see also Fig. 1b), all the pipes j (with $j = 1 \dots n_{tot}$) of the network are considered individually one by one. Then, the leakage is positioned in a pipe j by associating an emitter at each of the two corresponding adjacent nodes (j_1 and j_2), i.e. by defining the leakage flow rate f as the sum of the square root of the two nodal pressures (p_{j_1} and p_{j_2}) multiplied by an emitter coefficient c_j, as showed in the following equation:

$$f = c_j \cdot (p_{j_1}^{0.5} + p_{j_2}^{0.5})$$ (3)

The real leakage flow rate f being unknown, the initial value of c_j can be chosen randomly or made equal to 1. The pressure $p_{i,perturbed_j}$ for each node i is calculated. Then, for each pipe j, a new value x_j for the emitter coefficient is calculated by assuming a linear relationship between the pressure change and leakage flow rate. Where the differences between the measured and baseline pressures are defined as the measured pressure changes, and the differences between the perturbed and baseline pressures are defined as the perturbed pressure changes, the new emitter coefficient x_j will be the value that modulates the perturbed pressure changes so that they are as close as possible to the corresponding measured changes, and is determined as expressed in the following equation:

$$minimize \quad E_j = \sum_{i=1}^{n_{sen}} |(p_{i,measured} - p_{i,baseline})$$
$$- \frac{(p_{i,perturbed_j} - p_{i,baseline})}{c_j} \cdot x_j| \qquad \forall j = 1 \ldots n_{tot}$$ (4)

The emitter coefficient x_j obtained is passed back to the hydraulic model in order to update the assumed flow rate of the simulated leakage and the simulated pressures. This iterative process continues until the solution no longer changes significantly. In this particular case the method was iterated until the emitter coefficient providing the best solution differed by no more than 5% from the corresponding value of the previous simulation.

The leakage is localized in the pipe where the error E_j is the lowest compared to the rest of the network and the flow rate is given by the corresponding emitter x_j.

Mathematically, the determination of the new emitter coefficient x_j for each pipe could be modelled as a linear optimization problem (Berglund et al. 2017).

2.2 Case Study

The specific case considered is the water distribution system proposed within the framework of the Battle of the Leakage Detection and Isolation methods (BattLeDIM 2020), which supplies a small town with a population of around 10,000 people through a network of pipes with a total length of 42.6 km (Fig. 2). More in detail, the main zone of the system, called Area A, was considered in this study. It receives water from two reservoirs and is composed of 660 nodes and 765 pipelines.

Area A is monitored by $n_{sen} = 29$ pressure sensors, from which time series of pressure values at a 5-min time step for a two-year period were obtained. Moreover, an EPANET hydraulic model containing accurate pipe roughness values and the exact water consumption patterns at a 5-min time step for the same two-year period is available. Furthermore, 26 leakages that occurred during the same two-year period are known, and, for each of them, the position in terms of the pipe, the starting time step, the possible repair time step, and the flow rate at a 5-min time step are likewise known (Vrachimis et al. 2020). In particular, 15 were bursts, characterized by a steady flow rate ranging between 5 m^3/h and 35 m^3/h. The remaining 11 were incipient leaks, starting from 0 m^3/h, growing at different rates up to values of 30 m^3/h, and then becoming steady (Marzola et al. 2022). By adding the time series of the leakage flow rate as water demand at the two nodes adjacent to the corresponding leaking pipe, a hereinafter referred to as a *perfect* hydraulic model was
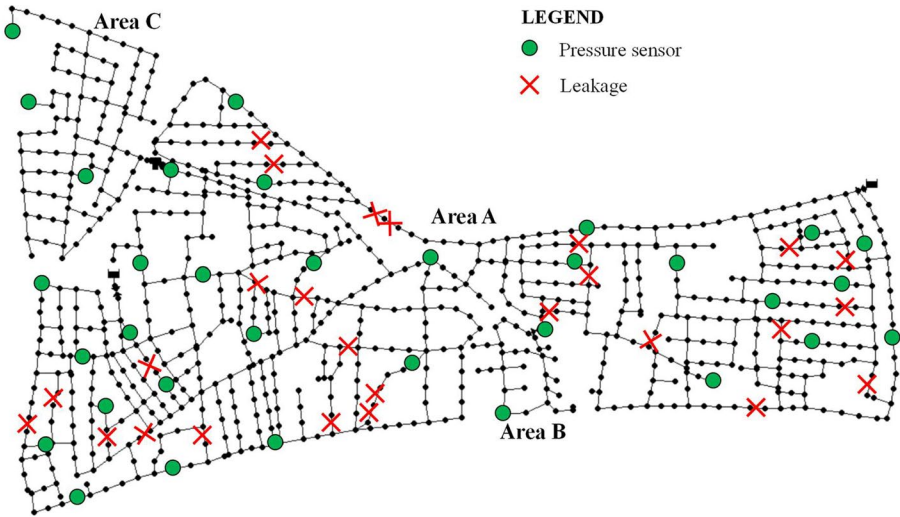
Fig. 2 The network considered in the case study. The pressure sensors are highlighted with green dots and the leaking pipes with red crosses

obtained, that is, a model that could almost perfectly describe the behaviour of the network. In fact, for this theoretical case study, both the pipe properties, such as diameter, roughness and length, and the node properties, such as elevation and demand, perfectly reflect the real ones. Indeed, the residual error between the simulated and measured pressure was on average 3 *mm*. All the simulations were performed by using EPANET 2.2 software for hydraulic modelling.

3 Comparison of the Methods

The methods were applied to locate the 26 leakages occurring in Area A. To ensure that the type of leakage did not affect the results obtained, all the leakages were considered as bursts, regardless of how they evolved. This was done by applying the SM and LA methods considering the first day in which the leakage flow rate was steady, i.e. the first day of occurrence in the case of bursts and the first day in which the flow rate remained steady in the case of incipient leakages.

Under this assumption, several analyses were carried in order to address the critical factors that affect the methods' performance in relation to:

- the type of data to be used (analysis A);
- the level of accuracy of the hydraulic model and the data given as inputs (analysis B1-B6);
- the number of pressure sensors monitoring the network (analysis C1-C2).

All these analyses are described in detail below and summarized in Table 1.

Table 1 Summary of the analyses carried out

Analysis	Model accuracy	Data accuracy	Number of sensors	Data exploitation
A	Perfect	Perfect	29	Snapshot
				Averaging
				Stacked
				Hourly stacked
B1	20% error in water demand	Perfect	29	Hourly stacked
B2	5% error in pipe roughness	Perfect		
B3	20% error in water demand and 5% error in pipe roughness	Perfect		
B4	Perfect	0.5% error in pressure sensor		
B5	Daily average consumption pattern and 5% error in pipe roughness	0.5% error in pressure sensor		
B6	Perfect	+20% in input leakage flow rate		
		-20% in input leakage flow rate		
C1	Perfect	Perfect	25 (uniform and random)	Hourly stacked
			20 (uniform and random)	
			15 (uniform and random)	
			10 (uniform and random)	
C2	20% error in water demand and 5% error in pipe roughness	Perfect	25 (uniform)	Hourly stacked
			20 (uniform)	
			15 (uniform)	
			10 (uniform)	

3.1 Analysis A

As previously highlighted, one critical aspect of these methods is related to the way in which the data are exploited. Indeed, considering a period of one day and an acquisition time step of 5 minutes, a time series of 288 pressure values is available. Conceptually, several approaches could be applied by exploiting the data in different ways, for example by considering the data of each time step independently, considering all the data simultaneously, considering the data averaged over the whole one day period, or mixing the previous approaches, e.g. by averaging the data over a shorter time interval and using the averaged values together. Based on these considerations, analysis A was carried out to determine which dataset should be used in order to obtain the best results, i.e. the number of data to be considered and the averaging time window. Four different approaches were examined.

In the first approach, defined as snapshot, the methods are applied 288 times using the data of each time step separately. In the second approach, referred to as stacked, all the values are exploited together by stacking them into a long column vector. In this case, when the SM method is used the number of rows of the residual vector and sensitivity matrix becomes 29 multiplied by 288, i.e. the number of the pressure sensors ($n_{sen} = 29$) multiplied by the number of time steps, as shown in the following equations, in which $t = 1, \ldots, 288$ represents the time step of 5 min during the day and $p_{i,measured}(t)$, $p_{i,baseline}(t)$ and $p_{i,perturbed_j}(t)$ are the pressure values corresponding to the time step t:

$$
r = \begin{bmatrix}
p_{1,measured}(1) - p_{1,baseline}(1) \\
\ldots \\
p_{1,measured}(t) - p_{1,baseline}(t) \\
\ldots \\
p_{1,measured}(288) - p_{1,baseline}(288) \\
\ldots \\
\ldots \\
p_{n_{sen},measured}(1) - p_{n_{sen},baseline}(1) \\
\ldots \\
p_{n_{sen},measured}(288) - p_{n_{sen},baseline}(288)
\end{bmatrix} \tag{5}
$$

$$
S = \begin{bmatrix}
\dfrac{p_{1,perturbed_1}(1) - p_{1,baseline}(1)}{f} & \cdots & \dfrac{p_{1,perturbed_{n_{tot}}}(1) - p_{1,baseline}(1)}{f} \\
\ldots & \cdots & \ldots \\
\dfrac{p_{1,perturbed_1}(t) - p_{1,baseline}(t)}{f} & \cdots & \dfrac{p_{1,perturbed_{n_{tot}}}(t) - p_{1,baseline}(t)}{f} \\
\ldots & \cdots & \ldots \\
\dfrac{p_{1,perturbed_1}(288) - p_{1,baseline}(288)}{f} & \cdots & \dfrac{p_{1,perturbed_{n_{tot}}}(288) - p_{1,baseline}(288)}{f} \\
\ldots & \cdots & \ldots \\
\ldots & \cdots & \ldots \\
\dfrac{p_{n_{sen},perturbed_1}(1) - p_{n_{sen},baseline}(1)}{f} & \cdots & \dfrac{p_{n_{sen},perturbed_{n_{tot}}}(1) - p_{n_{sen},baseline}(1)}{f} \\
\ldots & \cdots & \ldots \\
\dfrac{p_{n_{sen},perturbed_1}(288) - p_{n_{sen},baseline}(288)}{f} & \cdots & \dfrac{p_{n_{sen},perturbed_{n_{tot}}}(288) - p_{n_{sen},baseline}(288)}{f}
\end{bmatrix} \tag{6}
$$

When the LA method is applied, the number of constraints becomes 29 multiplied by 288, as expressed in the following equation:

$$
minimize \quad E_j = \sum_{i=1}^{n_{sen}} \sum_{t=1}^{288} |(p_{i,measured}(t) - p_{i,baseline}(t))
$$
$$
- \frac{\left(p_{i,perturbed_j}(t) - p_{i,baseline}(t)\right)}{c_j} \cdot x_j| \qquad \forall j = 1 \ldots n_{tot} \tag{7}
$$

In the third case, defined as the averaging approach, the values are averaged on a daily basis, thus removing any fluctuations occurring during the day. In the fourth case, an hourly stacked approach is considered. It can be defined as a combination of the previous three, as the data are first averaged on a hourly basis, and then the hourly average values exploited together using the stacked approach, meaning that the number of rows of the residual vector and sensitivity matrix becomes 24 (the hours of the day) multiplied by 29,

as shown below, where $h = 1, \ldots, 24$ defines the hour of the day, and $\bar{p}_{i,measured}(h)$, $\bar{p}_{i,baseline}(h)$ and $\bar{p}_{i,perturbed_j}(h)$ are the hourly averaged pressures corresponding to the hour h of the day:

$$r = \begin{bmatrix} \bar{p}_{1,measured}(1) - \bar{p}_{1,baseline}(1) \\ \ldots \\ \bar{p}_{1,measured}(h) - \bar{p}_{1,baseline}(h) \\ \ldots \\ \bar{p}_{1,measured}(24) - \bar{p}_{1,baseline}(24) \\ \ldots \\ \ldots \\ \bar{p}_{n_{sen},measured}(1) - \bar{p}_{n_{sen},baseline}(1) \\ \ldots \\ \bar{p}_{n_{sen},measured}(24) - \bar{p}_{n_{sen},baseline}(24) \end{bmatrix} \tag{8}$$

$$S = \begin{bmatrix} \dfrac{\bar{p}_{1,perturbed_1}(1) - \bar{p}_{1,baseline}(1)}{f} & \ldots & \dfrac{\bar{p}_{1,perturbed_{n_{tot}}}(1) - \bar{p}_{1,baseline}(1)}{f} \\ \ldots & \ldots & \ldots \\ \dfrac{\bar{p}_{1,perturbed_1}(h) - \bar{p}_{1,baseline}(h)}{f} & \ldots & \dfrac{\bar{p}_{1,perturbed_{n_{tot}}}(h) - \bar{p}_{1,baseline}(h)}{f} \\ \ldots & \ldots & \ldots \\ \dfrac{\bar{p}_{1,perturbed_1}(24) - \bar{p}_{1,baseline}(24)}{f} & \ldots & \dfrac{\bar{p}_{1,perturbed_{n_{tot}}}(24) - \bar{p}_{1,baseline}(24)}{f} \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ \dfrac{\bar{p}_{n_{sen},perturbed_1}(1) - \bar{p}_{n_{sen},baseline}(1)}{f} & \ldots & \dfrac{\bar{p}_{n_{sen},perturbed_{n_{tot}}}(1) - \bar{p}_{n_{sen},baseline}(1)}{f} \\ \ldots & \ldots & \ldots \\ \dfrac{\bar{p}_{n_{sen},perturbed_1}(24) - \bar{p}_{n_{sen},baseline}(24)}{f} & \ldots & \dfrac{\bar{p}_{n_{sen},perturbed_{n_{tot}}}(24) - \bar{p}_{n_{sen},baseline}(24)}{f} \end{bmatrix} \tag{9}$$

With the LA method, the number of constraints becomes 24 multiplied by 29, as shown in the following equation:

$$minimize \quad E_j = \sum_{i=1}^{n_{sen}} \sum_{h=1}^{24} | (\bar{p}_{i,measured}(h) - \bar{p}_{i,baseline}(h)) \\ - \frac{(\bar{p}_{i,perturbed_j}(h) - \bar{p}_{i,baseline}(h))}{c_j} \cdot x_j | \qquad \forall j = 1 \ldots n_{tot} \tag{10}$$

Solely in the case of the snapshot approach, 288 solutions are provided for each of the two model-based methods, whereas with the other three approaches, only one solution for each method is obtained.

3.2 Analysis B

Exploiting the approach that provided the best results in analysis A (i.e. the hourly stacked approach as will be explained in the Sect. 4), further analyses were carried out to address the other critical aspects highlighted before, namely, the effects of an incorrectly calibrated model, inaccurate input information, and a reduced number of pressure sensors.

In fact, water demand patterns are not always at the disposal of water utilities. Generally, the available data derive from meter readings, collected every four or six months for billing purposes, and the patterns used are typically derived from the literature, but they can differ from the real patterns. Moreover, information on the pipe characteristics, in terms of age and material, is not always available in water utilities' records. In order to take into account errors in the hydraulic model, in analysis B1 the water demands of the perfect hydraulic model were modified. In particular, all the values of the water demand patterns were perturbed by adding a noise randomly generated from a uniform distribution with boundaries of -40% and +40% of the original value considered. This led to an average absolute error of 20% over all the water demand data. Then, in analysis B2, the pipe roughness values of the perfect hydraulic model were modified. In particular, all the roughness values were perturbed by adding a noise randomly generated from a uniform distribution with boundaries of -10% and +10% of the value considered. This led to an average absolute error of 5% over all the pipe roughness data. Moreover, in analysis B3, these two considerations were combined, i.e. a model was obtained by merging the modified water demands as in B1 and the modified pipe roughness values as in B2, resulting in an average error of 20% over all the water demand data and of 5% over the pipe roughness data. However, the pressure sensors used to obtain the field measurements may themselves be subject to faults and performance impairment over time and thus be inaccurate. Therefore, in analysis B4, considering the perfect model, all the measured pressure data were perturbed by adding a noise randomly generated from a uniform distribution with boundaries of -1% and +1% of the value considered, which led to an average absolute error of 0.5%. Moreover, in analysis B5, a model with a daily average consumption pattern was considered, that is a pattern typically used by the water utilities, and the 5% error in roughness and the average absolute error of 0.5% on pressure sensor data previously described were also added.

A further analysis, analysis B6, was carried out only for the SM method, as it is the one that requires information about the leakage flow rate in order to proceed with the localization. The efficiency of the method was evaluated by modifying the required leakage flow rate f by plus and minus 20%, using the perfect model.

3.3 Analysis C

Analyses C1 and C2 were carried out to assess the effect of different numbers and positions of the pressure sensors, as only a limited number of sensors might be available due to water utilities' budget constraints. In particular, in analysis C1 the perfect model was considered, and the impact of a reduced number of pressure sensors was investigated by considering sets of 25, 20, 15 and 10 sensors. Two different approaches were used to determine the pressure sensors for each set:

- Uniform approach: this approach was used to obtain sets of sensors with a uniform coverage and layout. In practical terms, the two sensors nearest to each other are identified, i.e. the ones providing redundant information. Then, for each of these two sensors, the sum of the distances from all the other sensors is calculated. The sensor for which this value is the lowest is eliminated, as it is the one most closely surrounded by other sensors. This operation is repeated until the desired number of sensors is obtained.
- Random approach: the sensors are randomly selected, and this operation is repeated five time for each set considered.

Then, in analysis C2, the effect of the reduction of the number of pressure sensors were combined with an inaccurate model. In fact, the model used in analyses B3, i.e., with an average error of 20% over all the water demand data and of 5% over the pipe roughness data was considered with the sets of pressure sensor obtained by the uniform approach.

For each analysis and each method, a total localization error was calculated by summing 26 values, one for each leakage, corresponding to the distances between the centre of the actual leaking pipe and the centre of the identified pipe (or the average distance between the centre of the actual leaking pipe and the centres of the 288 identified pipes only in the case of the snapshot approach in analysis A).

4 Results and Discussions

As regards analysis A, the total localization errors obtained for each method using the four different approaches are shown in Table 2. The snapshot approach led to the highest errors in the case of the Linear Approximation method and to a similarly high error in the case of the Sensitivity Matrix method. This is probably due to the use of data corresponding to a small time window (5 minutes), which may therefore be influenced by consumption and not be representative of the network behaviour. It is also a disadvantageous approach in that it must be applied several times. The stacked approach led to the highest errors with the SM method, probably because of conflicting data in the determination of the correlation coefficients, while it led to good results with the LA method. However, it is worth highlighting that, in the case of the LA method, the stacked approach requires the longest computational time, due to the high number of constraints considered in the optimization problem. The averaging approach led to high errors with both methods. In this case, averaging the pressure values over the whole day may dampen the signals, resulting in the loss of very important information regarding daily pressure fluctuations. Overall, the best results were obtained with the hourly stacked approach. Indeed, this approach can be considered as a combination of the others, as it uses a time window of one hour, in which the data are averaged and then stacked. This leads to a reduction in flaws compared to the other approaches.

Overall, considering the application of the best approach, i.e. the hourly stacked approach, for both model-based methods, it is worth noting that the method providing the best results was not the SM, as might have been expected given the use of the exact leakage flow rate, but rather the LA, with an error of 51 *m* versus 619 *m*, even though, in order to obtain this result, many iterations were needed with the LA method. Based on the results of analysis A, the hourly stacked approach was selected and used for all the following analyses; the respective results were taken as a reference, since the perfect model and the totality of available pressure sensors were used.

Analyses B1-B5 were carried out to evaluate the effects of an inaccurate hydraulic model and inaccurate measured data on leakage localization. The localization errors obtained are showed in Table 2.

In particular, similar results were obtained with the two methods in analysis B1, i.e. assuming a model with an absolute average error of 20% in water demand patterns. As expected, the localization errors increased compared to when the perfect model was used, due to the reduction in the hydraulic model's accuracy. However, it can be noted that the

Table 2 Localization errors obtained for each method in analysis A, with the four different approaches to data usage, in analyses B1-B6, with the different model inaccuracies, and in analyses C1-C2 with the uniform approach and, in parentheses, the random approach

Analysis	Model and data accuracy	Number of sensors	Data exploitation	Localization error (m)	
				LA	SM
A	Perfect	29	Snapshot	911	807
			Averaging	716	635
			Stacked	51	1147
			Hourly stacked	51	619
B1	20% error in water demand	29	Hourly stacked	1495	1732
B2	5% error in pipe roughness			3418	3340
B3	20% error in water demand and 5% error in pipe roughness			4502	3076
B4	0.5% error in pressure sensor			4711	3027
B5	Daily average consumption pattern, 5% error in pipe roughness and 0.5% error in pressure sensor			14680	14540
B6	+20% in input leakage flow rate			-	719
	-20% in input leakage flow rate			-	421
C1	Perfect	25	Hourly stacked	(51 - 226) 51	(708 - 1020) 619
		20		(51 - 802) 51	(884 - 2050) 708
		15		(275 - 774) 275	(1385 - 2580) 1082
		10		(716 - 2500) 461	(3440 - 6190) 1281
C2	20% error in water demand and 5% error in pipe roughness	25	Hourly stacked	2699	2964
		20		3845	3161
		15		3007	2627
		10		3093	3496

perturbation introduced into the water demands had a higher impact on the LA method. In fact, the error of 1495 *m* obtained in this analysis is almost 30 times higher than the corresponding error of 51 *m* found with the perfect model in Analysis A. The two methods showed similar results once again in analysis B2, where a model with an absolute average error in pipe roughness of 5% was considered. However, the localization errors nearly doubled compared to the previous analysis, thus highlighting the significance of the effect of incorrect pipe characteristics on the methods' accuracy.

In analysis B3 the errors of the two previous models were considered simultaneously, thus creating a situation of model inaccuracy that water utilities may actually face. The SM method resulted in a total error of 3076 m, proving to be more robust to model inaccuracies than the LA method, which gave a total error of 4502 m, 46% higher than SM. More in detail, it is worth highlighting that, in the case of both methods, approximately 75% of the error is tied to the localization of leakages with a lower flow rate, i.e. below 15 m^3/h. Moreover, while for leakage flow rates above 15 m^3/h, the methods perform similarly well, with an average error of 45 m in the case of the SM and of 61 m in the case of the LA, when the leakage flow rates are below 15 m^3/h the SM shows an average error of 256 m and the LA an average error of 384 m, i.e. 50% higher. Therefore, the difference in the results of the two methods is due to the fact that the LA has more difficulty in identifying leakages with a low flow rate when model inaccuracies are introduced.

The SM method also proved to be robust to inaccuracies in the measured data, obtaining the best results in analysis B4, where an average absolute error of 0.5% was added to the pressure sensor data. Combining, in Analysis B5, the errors in the properties of the model with the inaccuracies in the pressure sensor data, i.e. considering a hydraulic model that is typically available to the water utilities, it can be noted how the errors increase significantly for both the methods, up to an average value of about 560 m per leakage. In any case, once again, the SM proved to be more robust, providing overall slightly more accurate results than the LA method.

However, since the SM requires knowledge of the leakage flow rate, the impact of the use of an incorrect leakage flow rate was then considered in analysis B6 by modifying the value by plus and minus 20%. The localization errors obtained were close to the value found when the exact flow rate was used (619 m, obtained in Analysis A), thus confirming once again the robustness of the SM method with regard to the inaccuracy of the data used.

Analysis C1 and C2 were carried out to assess the impact of using a reduced number of sensors. Regarding analysis C1, Table 2 shows the localization errors obtained with the uniform approach when considering sets of 25, 20, 15 and 10 pressure sensors. It can be noted that, with both methods, the reduction to 25 sensors does not lead to a decrease in accuracy, thus indicating that the 4 sensors removed were in excess and their information redundant. However, while in the case of the SM method the error increases greatly with a further removal of sensors, the LA method maintains the same accuracy with a reduction to 20 sensors. Moreover, even when only 10 sensors are considered, the LA method can be still considered efficient, as the errors it gives are still lower than the ones obtained with the SM considering all 29 sensors. As regards the random approach, the localization error ranges obtained for all sets with the 5 different random choices are shown in Table 2 in parenthesis. It is worth noting that these errors are always equal to or higher than the corresponding ones obtained with the uniform approach, thus highlighting the importance of sensor positioning for the purpose of evenly covering the network.

These results are due to the fact that, with the perfect model, the LA method is more efficient, as also seen in analysis A. However, in analysis C2, i.e. considering the model with inaccuracies in both water demand patterns and pipe roughness, the errors are higher — between 2500 m and 4500 m — and the LA method gives an average error about 3% higher than the one obtained with the SM method, thus in accordance with the previous results.

Thus, in summary, the analysis carried out showed that, using a perfect model, the LA method is the most efficient, even with a limited number of pressure sensors, while the SM method is the more robust when there are inaccuracies in both the model and the data.

5 Conclusions

In this paper two model-based methods for leakage localization exploiting only pressure data, i.e. the Linear Approximation method and the Sensitivity Matrix method, were analysed and applied to a case study. In particular, several analyses were carried out in order to evaluate the critical aspects tied to their application, i.e. how to exploit the available data, the effects of hydraulic model and data inaccuracies, and the impact of a reduced number of pressure sensors.

The results showed that the best approach in terms of data usage is the hourly stacked approach, which uses the hourly averaged data all together. The Linear Approximation method showed to be on average 3 times more accurate when the perfect hydraulic model was considered, even with a reduced number of sensors. However, the Sensitivity Matrix method is more robust and in fact it provides better results than the Linear Approximation method when the hydraulic model and data are affected by inaccuracies, leading to an error on average 10% lower that the Linear Approximation method. It is also worth highlighting that the Linear Approximation method provided an accurate estimation of the leakage flow rate in each analysis, but it required more iterations and computational time, while the Sensitivity Matrix method makes an instantaneous comparison between pressures, but needs prior knowledge of the flow rate value. The results obtained from the comparison of these two methods are related to this particular case study network and, therefore, further studies are needed to analyse the behaviour of the methods with other case studies. Nevertheless, the analyses conducted in this study could be of help to water utilities, as they may enable the latter to choose the most efficient method according to their situations. Moreover, they could be useful for the purpose of adapting existing methods so that they can become more robust and efficient in the presence of model or data errors.

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Adedeji KB, Hamam Y, Abe BT, Abu-Mahfouz AM (2017) Towards achieving a reliable leakage detection and localization algorithm for application in water piping networks: An overview. IEEE Access 5:20272–20285. https://doi.org/10.1109/ACCESS.2017.2752802

Aksela K, Aksela M, Vahala R (2009) Leakage detection in a real distribution network using a SOM. Urban Water J 6(4):279–289. https://doi.org/10.1080/15730620802673079

Alvisi S, Luciani C, Franchini M (2019) Using water consumption smart metering for water loss assessment in a DMA: a case study. Urban Water J 16(1):77–83. https://doi.org/10.1080/1573062X.2019.1633675

Berglund A, Areti VS, Brill D, Mahinthakumar G (2017) Successive linear approximation methods for leak detection in water distribution systems. J Water Resour Plann Manage 143(8):04017042. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000784

Chan TK, Chin CS, Zhong X (2018) Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. IEEE Access 6:78846–78867. https://doi.org/10.1109/ACCESS.2018.2885444

Daniel I, Pesantez J, Letzgus S, Fasaee MAK, Alghamdi F, Berglund E, Mahinthakumar G, Cominola A (2022) A sequential pressure-based algorithm for data-driven leakage identification and model-based localization in water distribution networks. J Water Resour Plann Manage 148(6):04022025. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001535

Farley B, Mounce SR, Boxall JB (2013) Development and field validation of a burst localization methodology. J Water Resour Plann Manage 139(6):604–613. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000290

Goulet JA, Coutu S, Smith IFC (2013) Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Adv Eng Inform 27:261–269. https://doi.org/10.1016/j.aei.2013.01.001

Hu Z, Chen B, Chen W, Tan D, Shen D (2021) Review of model-based and data-driven approaches for leak detection and location in water distribution systems. Water Supp 21(7):3282–3306. https://doi.org/10.2166/ws.2021.101

Javadiha M, Blesa J, Soldevila A, Puig V (2019) Leak localization in water distribution networks using deep learning. Paper presented at the 6th International Conference on Control, Decision and Information Technologies (CoDIT), Paris, France, 23–26 April 2019. https://doi.org/10.1109/CoDIT.2019.8820627

Kang D, Lansey K (2014) Novel approach to detecting pipe bursts in water distribution networks. J Water Resour Plann Manage 140(1):121–127. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000264

Lee SJ, Lee G, Suh JC, Lee JM (2016) Online burst detection and location of water distribution systems and its practical applications. J Water Resour Plann Manage 142(1):04015033. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000545

Li J, Wu Y, Lu C (2022) An accurate leakage localization method for water supply network based on deep learning network. Water Resour Manag 36:2309–2325. https://doi.org/10.1007/s11269-022-03144-x

Li J, Wu Y, Zheng W, Lu C (2021) A model-based bayesian framework for pipeline leakage enumeration and location estimation. Water Resour Manag 35:4381–4397. https://doi.org/10.1007/s11269-021-02955-8

Li R, Huang H, Xin K, Tao T (2015) A review of methods for burst/leakage detection and location in water distribution systems. Water Sci Tech - W Sup 15(3):429–441. https://doi.org/10.2166/ws.2014.131

Louriero D, Amado C, Martins A, Vitorino D, Mamade A, Coelho S (2016) Water distribution systems flow monitoring and anomalous event detection: A practical approach. Urban Water J 13(3):242–252. https://doi.org/10.1080/1573062X.2014.988733

Marzola I, Alvisi S, Franchini M (2021) Analysis of MNF and FAVAD model for leakage characterization by exploiting smart-metered data: The case of the Gorino Ferrarese (FE-Italy) district. Water 13(5):643. https://doi.org/10.3390/w13050643

Marzola I, Mazzoni F, Alvisi S, Franchini M (2022) Leakage detection and localization in a water distribution network through comparison of observed and simulated pressure data. J Water Resour Plann Manage 148(1):04017077. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001503

Moasheri R, Jalili-Ghazizadeh M (2020) Locating of probabilistic leakage areas in water distribution networks by a calibration method using the imperialist competitive algorithm. Water Resour Manag 34:35–49. https://doi.org/10.1007/s11269-019-02388-4

Moser G, Paal SG, Smith IFC (2018) Leak detection of water supply networks using error-domain model falsification. J Water Resour Plann Manage 32(2):04017077. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000729

Mounce SR, Day AJ, Wood AS, Khan A, Widdop PD, Machell J (2002) A neural network approach to burst detection. Water Sci Technol 45(4–5):237–246. https://doi.org/10.2166/wst.2002.0595

Mounce SR, Machell J (2006) Burst detection using hydraulic data from water distribution systems with artificial neural networks. Urban Water J 3(1):21–31. https://doi.org/10.1080/15730620600578538

Pacchin E, Gagliardi F, Alvisi S, Franchini M (2019) A comparison of short-term water demand forecasting models. Water Resour Manag 33:1481–1497. https://doi.org/10.1007/s11269-019-02213-y

Perez R, Puig V, Pascual J, Peralta A, Landeros E, Jordanas L (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Sci Tech - W Sup 9(6):715–721. https://doi.org/10.2166/ws.2009.372

Perez R, Sanz G, Puig V, Quevedo J, Cuguero-Escofet MA, Nejjari F, Meseguer J, Cembrano G, Mirats Tur JM, Sarrate R (2014) Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona. IEEE Control Syst 34(3):24–36. https://doi.org/10.1109/MCS.2014.2320336

Ponce MVC, Castanon LEG, Cayuela VP (2014) Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. J Hydroinform 16(3):649–670. https://doi.org/10.2166/hydro.2013.019

Sophocleous S, Savic D, Kapelan Z (2019) Leak localization in a real water distribution network based on search-space reduction. J Water Resour Plann Manage 145(7):04019024. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001079

Soroush F, Abedini MJ (2019) Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm. J Hydroinform 21(6):1030–1047. https://doi.org/10.2166/hydro.2019.023

Steffelbauer DB, Deuerlein J, Gilbert D, Abraham E, Piller O (2022) Pressure-leak duality for leak detection and localization in water distribution systems. J Water Resour Plann Manage 148(3):04021106. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001515

Vrachimis SG, Eliades DG, Taormina R, Ostfeld A, Kapelan Z, Liu S (2020) Battle of the leakage detection and isolation methods dataset. https://zenodo.org/record/4017659#.YgUsUZbSK3A

Wu Y, Liu S (2017) A review of data-driven approaches for burst detection in water distribution systems. Urban Water J 14(9):972–983. https://doi.org/10.1080/1573062X.2017.1279191

Wu ZY, Sage P, Turtle D (2010) Pressure-dependent leak detection model and its application to a district water system. J Water Resour Plann Manage 136(1):116–128. https://doi.org/10.1061/(ASCE)0733-9496(2010)136:1(116)