Leak Localization in Water Distribution Networks Using Data-Driven and Model-Based Approaches

Luis Romero-Ben[1], Débora Alves[2], Joaquim Blesa[3], Gabriela Cembrano[4], Vicenç Puig[5], and Eric Duviella[6]

[1]Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain. Email: luis.romero.ben@upc.edu (corresponding author)

[2]Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, Rambla Sant Nebridi, 22, 08222 Terrassa, Barcelona, Spain. Email: adeboracris@gmail.com

[3]Dr.Eng. Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain; Serra Húnter fellow, Automatic Control Department of the Universitat Politècnica de Catalunya, Avinguda Eduard Maristany, 16, 08019, Barcelona, Spain. Email: joaquim.blesa@upc.edu

[4]Dr.Eng. Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain. Email: gabriela.cembrano@upc.edu

[5]Dr.Eng. Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain; Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, Rambla Sant Nebridi, 22, 08222 Terrassa, Barcelona, Spain. Email: vicenc.puig@upc.edu

[5]Dr.Eng. Informatics and Automatics Department, IMT Lille Douai, Lille, France. Email: eric.duviella@imt-lille-douai.fr

ABSTRACT

The detection and localization of leaks in water distribution networks (WDNs) is one of the major concerns of water utilities, due to the necessity of an efficient operation that satisfies the

worldwide growing demand of water. There exists a wide range of methods, from equipment-based techniques that rely only on hardware devices to software-based methods that exploit models and algorithms as well. Model-based approaches provide an effective performance but they rely on the availability of an hydraulic model of the WDN, while data-driven techniques only require measurements from the network operation although they may produce less accurate results. This paper proposes two methodologies: a model-based approach that uses the hydraulic model of the network, as well as pressure and demand information; and a fully data-driven method based on graph interpolation and a new candidate selection criteria. Their complementary application was successfully applied to the Battle of the Leakage Detection and Isolation Methods (BattLeDIM) 2020 challenge, and the achieved results are presented in this paper to demonstrate the suitability of the methods.

INTRODUCTION

A recent study, developed in Liemberger and Wyatt (2019), estimated that leaks account for up to 126 billion cubic meters of water per year worldwide (expressed as non-revenue water), which represents a remarkably significant quantity considering the worldwide growing demand, supposed to increase by 55% between 2000 and 2050 according to Leflaive (2012). Furthermore, apart from the associated economical and operational costs, water leaks increase the risk of contamination (Xu et al. 2014) and health problems (LeChevallier et al. 2003). Multitude of solutions have been proposed during the years to address the leak detection and localization problem (see Chan et al. (2018) for an extensive review). They are typically classified into two categories: hardware-based and software-based methods (Li et al. 2015).

On the one hand, hardware-based methods use hardware devices to detect the existence and position of bursts. They are usually divided into acoustic methods, including listening rods, leak correlators and leak noise loggers (Mutikanga et al. 2013); and non-acoustic approaches, like gas injection, ground penetrating radar technology or thermal infrared imaging among others (Fanner et al. 2007). Even if these methods provide a high degree of accuracy, their usage is usually prohibitive for large pipe networks due to its high associated costs and reduced detection range, and hence their application is limited to small zones of the WDNs (Rajeswaran et al. 2018).

On the other hand, software-based techniques rely on models or algorithms that exploit additional information from metering devices (pressure meters, flow sensors, etc.) to perform the detection/localization of the leaks. These methods can be split into three main categories: transient-based, model-based and data-driven methods.

Transient-based approaches analyse the transients induced by leaks using signal processing techniques. A leak detection method that exploits transient-based information and genetic algorithms is described in Vítkovskỳ et al. (2000). The detection task is tackled in Kapelan et al. (2003) by a hybrid inverse transient procedure, formulated as a constrained optimization problem of a weighted least-squares cost function. Besides, Wang et al. (2020) proposes a method that uses matched-field processing to locate leaks by incorporating prior information of the modelling error. Techniques based on leak transients are also used for leak diagnosis in pipelines, as shown in Pérez-Pérez et al. (2021) and Torres et al. (2021).

Model-based methodologies use hydraulic models and simulators, calibrating both the network characteristics and the demands, to compare simulated hydraulic information with actual measurements from the WDN (see the introduction in Sanz et al. (2016) for a review). A leak detection and localization technique using flow velocities is presented in Goulet et al. (2013), built as an error-domain model falsification methodology. A widely-used localization approach that matches pressure residuals to a fault signature matrix obtained by means of hydraulic simulations is proposed in Pérez et al. (2014). A similar consideration is used in Sophocleous et al. (2019), performing a sensitivity analysis along with a search-space reduction approach to find the leakage location.

While model-based approaches have been widely researched and exploited due to their efficiency and effectiveness (Duan et al. 2011), the associated performance is limited by the difficulty in the selection and calibration of the corresponding mathematical models (Menapace et al. 2018), the diversity and complexity of WDNs (Kim et al. 2016), and the presence of modelling errors like nodal demand uncertainties and measurement noise (Blesa and Pérez 2018). Most of these disadvantages may be gradually overcome using data-driven and machine learning techniques (Ferrandez-Gamot et al. 2015) and their reduced or non-existent dependency on an hydraulic model.

Data-driven strategies analyse the measurements from monitoring devices, mining knowledge to detect leaks and identify their location (see Wu and Liu (2017) for an extensive review). Information from accelerometers is exploited in Kang et al. (2017) to feed a two-phase method supported by a convolutional neural network (CNN) - support vector machine (SVM) architecture that detect leaks and a graph method based on virtual nodes that locates them. The concept of Kantorovich distance is applied in Arifin et al. (2018) to detect and locate leaks by exploiting the pipeline leak signature and identifying possible changes in the pipeline status using mass flow rates and pressure data. An efficient multistage method is presented in Huang et al. (2020), which uses valve operations (VOs) to split the demand metering area (DMA) into two zones and identify the leak location within these regions by means of a water balance analysis based on smart demand meters.

Recently proposed data-driven approaches use pressure sensors due to their the lower cost and easier installation in comparison with other kind of meters, resulting in an attractive option for water utilities (Soldevila et al. 2021). In Han et al. (2018), a two-stage strategy is used to estimate the WDN state by means of a Gauss-Newton Belief Propagation inference scheme applied to hydraulic heads at certain nodes of the network, and a clustering method to decompose the WDN and isolate the leak. A deep-learning scheme is proposed in Zhou et al. (2019) to locate leaks using pressure meters that are placed at limited, optimised places for a short period. A data-driven approach is developed in Soldevila et al. (2021), interpolating the pressure at every node of the network from certain measured values by means of the Kriging interpolation technique, comparing leak and leak-free scenarios to locate the leak and using Dempster-Shafer reasoning to deal with uncertainty.

This article proposes two different and complementary techniques to locate leaks: a model-based method that uses a hydraulic model of the WDN and pressure and demand measurements and/or well-calibrated demand estimations; and a data-driven approach that only requires pressure information from some inner nodes and the topology of the network. Both of them are applied to the Battle of the Leakage Detection and Isolation Methods (BattLeDIM) 2020 challenge (Vrachimis et al. 2020), using them in a complementary manner to improve the performance.

Both methodologies present several contributions with respect to previous approaches:

- They handle the multi-leak problem, overcoming the classical hypothesis of the appearance of a single leak at a time (Goulet et al. 2013; Pérez et al. 2014), which is assumed in most state-of-the-art techniques (Soldevila et al. 2016).
- Regarding the data-driven method, it reduces the complexity of the interpolation stage by using a quadratic programming approach that exploits the topology of the network, concretely the length of the pipes. Other interpolation-based methods like Soldevila et al. (2021) require extra information to be known (like diameters) or estimated during the interpolation (distribution of flow in pipes, pipe roughness, etc.). Moreover, in the referred work, a basic residual computation is employed to select the candidate to be the origin of the leak. This approach is completed in the method proposed here by combining the information of both the basic residuals and the relation among the hydraulic heads of all the nodes of the network (interpolated and measured) in a single metric.

Furthermore, a leak detection and estimation technique is proposed to complete the solution of the challenge and feed the localization strategies with the leak appearance time instants. Then, the proposed localization methods depend on the proper operation of the detection stage.

METHODOLOGY

Water distribution networks across the world present different characteristics in structure, size, demand patterns, components, etc. Moreover, the distribution, amount and properties of the measuring devices installed throughout the networks varies from one site to another. This fact indicates the necessity of adapting the selection and usage of software-based techniques.

Thus, two complementary approaches are proposed in the following to locate leaks in WDNs:

- A model-based methodology that exploits the existence of a well-calibrated hydraulic model of the WDN, as well as the availability of reliable pressure and demand information.
- A fully data-driven technique that only requires minimal topological knowledge of the network and measurements from pressure sensors distributed at a set of inner nodes.

The model-based method faces difficulties when there is a lack of demand measurements or estimations, while the data-driven strategy requires a minimum pressure sensor density in order to operate, as it is its only source of hydraulic information. The WDNs can be divided into areas with different features (like in the presented case study), e.g., sensorization properties (amount, placement, precision and type of installed sensors), existence of weirs, tanks, valves... Therefore, the best option between these two methods can be selected depending on the availability of information and/or model of the WDN, and they can be even used in a complementary manner, so that the weaknesses of one method are compensated by the strengths of the other.

Leak detection and estimation

The proposed method uses sensor fusion calculations to analyse the flow of water supplied to the DMA at all hours of the day, not only during the night hours. Initially, it is assumed that the demand forecasting method is calibrated with historical data from the DMAs (Donkor et al. 2014), which will provide a good approximation of the current inflow y at time k:

$$y(k) = \hat{y}(k) + e(k) \tag{1}$$

where $k = 0, 1, 2, ...$, denotes the discrete time corresponding to time $0, T_s, 2T_s, ...$ (T_s is the sample time of the demand forecasting model), $\hat{y}(k)$ is the demand forecast and $e(k)$ is the error that, for this study, is considered to be adjusted by a normal distribution (Malik 2016). As the incoming demand is more accurate in some periods of the day, a periodic variation in time T is considered:

$$e(k) \sim N(0, \sigma^2(k)) \qquad \text{with} \qquad \sigma^2(k) = \sigma^2(k+T) \tag{2}$$

In the case of the presence of a leak, i.e., $f(k) > 0$, Equation (1) leads to:

$$y(k) = \hat{y}(k) + e(k) + f(k) \quad \longrightarrow \quad \hat{f}(k) = y(k) - \hat{y}(k) = f(k) + e(k) \tag{3}$$

where $\hat{f}(k)$ is an approximation of the leak size given by the difference between the actual and the estimated inlet flow, with a leak estimation error equal to the demand forecasting error.

Given the current inlet flow value and previous values in a time window with N_W samples, Equation (3) can be applied considering N_W different leak estimations, approximating the leak as:

$$f(k) \approx \overline{f}(k) = \sum_{i=0}^{N_W-1} \frac{f(k-i)}{N_W} \tag{4}$$

An average leak estimation $\hat{\overline{f}}(k)$ can be computed at instant k applying the maximum likelihood estimation method to the joint probability distribution of the N_W estimations fused in $\overline{f}(k)$:

$$\hat{\overline{f}}(k) = \frac{\sum_{i=0}^{i=N_W-1} \frac{\hat{f}(k-i)}{\sigma^2(k-i)}}{\sum_{i=0}^{i=N_W-1} \frac{1}{\sigma^2(k-i)}} \tag{5}$$

In a non-leak scenario, $\hat{\overline{f}}(k)$ will lead to small values (but different from zero) due to demand estimation errors, whereas its value will increase in a leak scenario. Thus, the leak detection is formulated as a change detection problem that can be solved by computing a threshold λ that will determine the value of $\hat{\overline{f}}(k)$ above which it can be assumed that a leak exists. This value can be computed applying Equation (5) to historical leak-free data, considering the worst-case scenario λ to be equal to the maximum value of $\hat{\overline{f}}(k)$ computed for the whole historical non-leak data.

Then, the leak detection is triggered as stated by the following expression:

$$\begin{cases} \hat{f}(k) > \lambda \Rightarrow \text{Leak} \\ \text{Otherwise} \Rightarrow \text{No Leak} \end{cases} \tag{6}$$

Data-driven methodology for leak localization

The proposed fully data-driven approach is based on two phases: a graph-based interpolation that approximates the complete hydraulic state of the network and a geometric comparison between the states in leak and leak-free scenarios. This approach removes the need for a hydraulic model, as it only uses measured pressure values and the WDN topology. It presents the following features:

- It is applicable to the measurements of a single time-instant, but it can be easily extended to integrate temporal information from time-series.

- The leak-free data can be obtained from a historical dataset of hydraulic measurements, provided by the water utility. In order to deal with possible differences in the demand conditions between the leak and leak-free scenarios, updated nominal information can be obtained from the previous days to the appearance of the leak. Moreover, these demand discrepancies can be reduced even more by selecting nominal scenarios with an analogous consumption pattern, e.g., the same day of the week. In this way, an hydraulic model is not required to obtain the nominal data.
- It provides a set of leak candidate inner nodes. The size of the set depends on the sensors limitations and network structure. However, it offers the possibility of approximating the exact leak location with sufficient reliability.
- It is flexible regarding the network structure, so that modifications of the WDN only imply the update of the topological information. Besides, it does not require demand information, which is one of the critical points of most of the model-based strategies.

This leak localization method must be used together with the presented detection technique.

Graph-based state interpolation

In this work, the location of leaks is determined through the comparison between the actual hydraulic state of the network and a leak-free reference. The hydraulic heads at the nodes are chosen as a representative state variables due to the advantages of pressure sensors and the usage of this information by water utilities to determine the availability of water service.

In order to minimize the installation of sensors, the complete state is interpolated from a reduced set of measurements. The non-linear relation among the hydraulic heads of neighbouring nodes, typically described by the Hazen-Williams equation, is approximated considering the state of a node to be computed as the weighted linear contribution of its neighbours, hence simplifying the interpolation procedure and expanding the set of tools for the derivation of the network state.

Let us model the network structure by means of a simple directed graph referred to as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V}$ is representing the set of junctions of the WDN and the edge set $\mathcal{E}$ stands for the set of pipes. The total number of junctions of the network, i.e. the cardinality $|\mathcal{V}|$ of the node set, is referred to as n, as well as the number of pipes, i.e., the cardinality $|\mathcal{E}|$ of the edge set, is represented as m. A node is referred to as $v_i \in \mathcal{V}$, and an edge $e_{ij} = (v_i, v_j) \in \mathcal{E}$ connects source node v_i with sink node v_j, so that they are its endpoints. A set of important matrices can be extracted from the graph characteristics:

- The connectivity among nodes is used to derive the adjacency matrix $A(\mathcal{G}) \in \mathbb{R}^{n \times n}$:

$$a_{ij} = \begin{cases} 1, & \text{if } e_{ij} \in \mathcal{E} \;\; or \;\; e_{ji} \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

- The directionality of the graph edges represents the direction of the water flows through the network pipes. Considering that v_i is the source of an edge e_o (or e_{ij}), while v_j is its sink, the directionality is represented by the incidence matrix $B(\mathcal{G}) \in \mathbb{R}^{m \times n}$ as follows:

$$b_{oj} = \begin{cases} 1, & \text{if } e_o = (v_i, v_j) \in \mathcal{E} \\ -1, & \text{if } e_o = (v_j, v_i) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

  with o from 1 to m indexing the edges of the graph. An approximated $B(\mathcal{G})$ can be estimated considering the source-sink relation among the reservoirs and all the inner nodes, as the exact flow direction of each pipe may be unknown and/or it may vary.

- An edge e_{ij} is associated to a cost value, related in this case to the pipe length l_{ij}. The edge costs are exploited to generate a weighted adjacency matrix $W(\mathcal{G}) \in \mathbb{R}^{n \times n}$:

$$w_{ij} = \begin{cases} \frac{1}{l_{ij}}, & \text{if } l_{ij} \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

  where the definition of the weight w_{ij} is designed so that closer neighbours affect the value of the considered node in a higher degree.

- The degree matrix $D(\mathcal{G}) \in \mathbb{R}^{n\times n}$ is derived from $W(\mathcal{G})$, i.e., $d_{ij} = \sum_{h=1}^{n} w_{ih}$ only if $i = j$; being zero otherwise (h plays the role of the columns index in w_{ih} because j is used in d_{ij}).

With the aid of the these matrices, the state x_i of a certain node v_i can be expressed as:

$$x_i = \frac{1}{d_i}\mathbf{w}_i\mathbf{x} \tag{10}$$

where $d_i = d_{ii} = d_{ij}$, $\mathbf{w}_i \in \mathbb{R}^{1\times n}$ denotes the i-th row of $W(\mathcal{G})$, and $\mathbf{x} \in \mathbb{R}^n$ represents the complete state vector. The estimation of $\mathbf{x}$ can be achieved through the minimization of the sum of the quadratic difference between each node actual value and the one computed by Equation (10):

$$\begin{aligned}\sum_{i=1}^{n}\left[x_i - \frac{1}{d_i}\mathbf{w}_i\mathbf{x}\right]^2 = \left(\mathbf{x} - D^{-1}W\mathbf{x}\right)^T\left(\mathbf{x} - D^{-1}W\mathbf{x}\right) = \mathbf{x}^T\left(I_n - D^{-1}W\right)^T\left(I_n - D^{-1}W\right)\mathbf{x} = \\ \mathbf{x}^T\left(D^{-1}(D-W)\right)^T\left(D^{-1}(D-W)\right)\mathbf{x} = \mathbf{x}^T\left(D^{-1}L\right)^T\left(D^{-1}L\right)\mathbf{x} = \mathbf{x}^T L D^{-2} L\mathbf{x}\end{aligned} \tag{11}$$

where $D = D(\mathcal{G})$ and $W = W(\mathcal{G})$, and $L = L^T = L(\mathcal{G}) = D(\mathcal{G}) - W(\mathcal{G})$ is the unnormalized Laplacian matrix of graph $\mathcal{G}$ (Mohar et al. 1991). I_n stands for the identity matrix of size n.

In order to provide the available hydraulic information to the minimization process, the state of the metered nodes is filled using the measured hydraulic heads through an equality constraint:

$$S\mathbf{x} = \mathbf{x}^s \tag{12}$$

where $S \in \mathbb{R}^{n\times n}$ is a diagonal matrix which has a value of 1 at its $i - i$ component if there exists a sensor at v_i, and a value of 0 otherwise. Vector $\mathbf{x}^s \in \mathbb{R}^n$ contains the head values of the measured nodes at the corresponding components, while the rest of values are 0. The pressure at the water inputs of the network is supposed to be known, as it is commonly available in most of the WDNs.

The approximation in Equation (10) denotes the harmonic property of functions in graphs, explained in Zhu et al. (2003) to interpolate the value of a function to unknown vertices following a smoothing strategy. In that work, the harmonic property is implicitly pursued by minimizing a weighted quadratic energy function over the function values. However, the harmonic property is explicitly pursued in Equation (11) by directly minimizing the quadratic difference between the actual node state and the value estimated by computing the average of the state in neighbouring nodes. This cost function does not impose a smoothing objective like the one in Zhu et al. (2003), so all the possible combinations of neighbouring states that produce the same state for a certain node are equally considered.

This provides a higher degree of freedom in the solution to attain a second objective, related to the directionality of the flows. In WDNs, the direction of water flow through a pipe is determined by the sign of the difference in hydraulic head between its corresponding junctions, taking into account that water flows in the direction of decreasing hydraulic head. This fact can be translated to the relation among the states of the nodes of graph $\mathcal{G}$ by means of the following inequality:

$$B\mathbf{x} \leq \boldsymbol{\gamma} \tag{13}$$

where $B = B(\mathcal{G})$ is defined as stated in Equation (8) and $\boldsymbol{\gamma} \in \mathbb{R}^n$ stands for a vector with a value of γ for all its components. This expression would imply a requirement about the difference of state (that is, approximated hydraulic head) between adjacent nodes, so that it must be lower or equal to a certain threshold γ. This value is introduced because $B(\mathcal{G})$ is estimated from the WDN topology and hence some slack is desirable in the fulfilment of the directionality constraints. However, this threshold is included in a second objective of the optimization cost function to compute its lowest positive feasible value. Thus, the complete optimization problem can be arranged as:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\left[\mathbf{x}^T L D^{-2} L \mathbf{x} + \alpha\gamma^2\right] \\ \text{s.t.} \quad & B\mathbf{x} \leq \boldsymbol{\gamma} \\ & \gamma > 0 \\ & S\mathbf{x} = \mathbf{x}^s \end{aligned} \tag{14}$$

where α allows to settle the importance of the directionality objective.

Leak candidate selection method

As leak and leak-free scenarios must be compared to locate the leak, the graph-based state interpolation is applied to both the actual and nominal hydraulic information, i.e., the measured hydraulic heads at a certain time instant of the detected leaky event and the ones at a leak-free time instant (with similar boundary conditions (Pérez et al. 2011)). The former are stored in a vector denoted as $\mathbf{x}_{leak} \in \mathbb{R}^n$, while the latter are collected to form a vector referred to as $\mathbf{x}_{nom} \in \mathbb{R}^n$.

The proposed leak candidate selection method considers the components of those vectors to represent the coordinates of n points in $\mathbb{R}^2$ ($\mathbf{x}_{nom}$ provides the x-coordinates and $\mathbf{x}_{leak}$ provides the y-coordinates). In the case of comparing two healthy scenarios, these points should be rather aligned because they would be only affected by the boundary conditions. However, the presence of a leak would alter the $\mathbb{R}^2$ location of the points representing the affected nodes, considering that a leak produces a reduction of the expected hydraulic head (Adedeji et al. 2017).

Thus, the objective of this stage consists of providing the set of most distant points from their predicted position, given by the best fitting line of the complete set of points. As this line is computed considering all of them, the distance between a certain node and the line not only encodes information about the change of state in that node, but also the relation among the state of all the nodes of the network. The best-fitting line can be expressed as $\mathbf{x}_{leak} = X_{nom}\boldsymbol{\phi}$, with $X_{nom} = [\mathbf{x}_{nom}\ \mathbf{1}_n] \in \mathbb{R}^{n\times 2}$ ($\mathbf{1}_n$ is a column vector whose components are all 1) and $\boldsymbol{\phi} = [\phi_1\ \phi_2]^T \in \mathbb{R}^2$. The latter is the vector containing the line parameters, which are computed by solving the least-squares problem, i.e., $\boldsymbol{\phi} = \left(X_{nom}^T X_{nom}\right)^{-1} X_{nom}^T \mathbf{x}_{leak}$.

The perpendicular distance from the set of points to the best-fitting line can be computed as:

$$\boldsymbol{\delta} = [\mathbf{x}_{nom}\ \mathbf{x}_{leak}\ \mathbf{1}_n] \left[\frac{\phi_1}{\sqrt{\phi_1^2+1}}\ \frac{-1}{\sqrt{\phi_1^2+1}}\ \frac{\phi_2}{\sqrt{\phi_1^2+1}} \right]^T \tag{15}$$

so $\boldsymbol{\delta} \in \mathbb{R}^n$ stands for the vector of computed distances. The sign is kept in the calculation since only points with a positive distance value can be leak candidates, i.e., points that are located below the best-fitting line because their y-coordinate value, that is the leaky one, is lower than expected.

A criterion must be selected to pick a set of nodes from $\mathcal{V}$ depending on the values at $\boldsymbol{\delta}$. In this case, the standard deviation $\sigma_{\boldsymbol{\delta}}$ of the distance vector is computed to play the role of a threshold, i.e., a node v_i must have an associated distance value δ_i that exceeds $\sigma_{\boldsymbol{\delta}}$ in order to consider this node as a candidate to be the leak origin. The final candidates can be ordered from most to least probable by means of their associated distance δ_i, and therefore the node that corresponds to the highest distance value is considered as the best candidate.

Model-based methodology for leak localization

The proposed model-based leak localization method uses a hydraulic simulator to simulate theoretical pressure values caused by all potential leaks once a leak has been detected and its magnitude has been estimated. Simulated pressure values at different leak locations (hypothesis) are compared with the DMA measured pressure values to determine the most probable leak location. After the leak localization procedure, the hydraulic simulator is updated with the new extra demand, whose magnitude is the leak estimation value at the leak localization. Thus, the proposed method can tackle the problem of multi-leaks $f_{j_1}^1, f_{j_2}^2, \ldots, f_{j_{N_f}}^{N_f}$ with the constraint that the leaks should appear at sequential time instants $k_1 < k_2 < .. < k_{N_f}$ that allow the leak detection, estimation and localization method to sequentially update the hydraulic simulator with extra demands at leak localization nodes $\hat{j}_1, \hat{j}_2, \ldots, \hat{j}_{N_f}$. Once a leak has been fixed, the simulator is updated, eliminating the extra demand related to the fixed leak. This process can be done manually or automatically detecting negative values in the leak estimation, computed by Equation (5).

The model-based leak localization method performance depends on the accuracy of the hydraulic model, sensor noises and the availability of reliable users demand information (Blesa and Pérez 2018). This third factor is potentially the most critical one because it is not easy to estimate user demands with high accuracy if there are no AMRs installed in some network nodes.

CASE STUDY AND RESULTS

The presented methodologies are tested by means of their application to achieve a solution for the BattLeDIM 2020 challenge. This competition aims at objectively comparing the performance of leak detection and isolation approaches, and hence a common benchmark was prepared for all the competitors: L-Town. It is a small and hypothetical town whose WDN is composed by 782 inner nodes, 2 reservoirs, 1 tank and 909 pipes. It is represented in Fig. 1. The inner nodes are coloured to indicate their elevation (which is displayed by means of a colour bar). Pressure sensors, reservoirs and tanks are indicated with special markers. The leaks at the provided datasets are highlighted at the corresponding pipes.

The network is composed of three different zones:

- *Area A*: it is the larger area, composed of the nodes with an elevation (see Fig. 1) between 16 and 48 m (655 in total). There is a high density of pressure sensors (29 in total) distributed through the area. Besides, the two water inlets to the network are located within this zone, as well as a tank that is filled from this area to provide water to *Area C*.
- *Area B*: it comprises the nodes that are elevated less than 16 m (31 in total). This zone is connected to *Area A* by a Pressure Reduction Valve or PRV (there are also installed PRVs downstream of the reservoirs of *Area A*) and there is only one pressure meter in the area.
- *Area C*: it is composed of the nodes elevated over 48 m (92 in total), and it is supplied with water by the previously mentioned tank. There are 3 pressure sensors installed through the zone, as well as 82 Automated Metered Readings (AMRs) that provide demand information.

The competition consists of detecting and localizing the maximum number of leaks from the ones occurred during 2019, as well as the ones remaining from 2018 (represented in Fig. 1), by means of readings of pressure sensors, AMRs of *Area C* and flow meters at the output of the reservoirs and tanks. Moreover, a nominal hydraulic model of the WDN is given, although it is affected by inaccuracies at demand patterns, pipe characteristics, valves status, etc.

The different features of the three areas composing the network motivated the usage of the proposed methodologies due to their complementary nature. In this case, the model-based approach is utilized over *Area B* and *Area C*: the former has an unique pressure sensor, so the data-driven cannot perform adequately; while the latter has AMRs installed at a high percentage of the nodes, and hence there is accurate demand information that allows to obtain precise results with the model-based technique. Meanwhile, the data-driven method is exploited in *Area A*, due to the high density of installed pressure sensors, as well as the lack of accurate demand information.

A schematic flow diagram of the application of each methodology to the benchmark is showed in Fig. 2.

To remark that, for their usage in a real-world application, the leak detection and localization stages would be separated, i.e., first the leak would be detected, and then the corresponding localization algorithm would be applied, finding the leak and fixing it. The presented diagrams reflect the necessity of dealing with multiple and simultaneous leaks at the BattLeDIM2020 case.

Besides, the localization methods yield a node/group of nodes as the leak candidate/s, so the best candidates, regarding the corresponding criteria for each method, were used to compute the leaky pipes, considering that the leak must be located at the pipe that connects the best two candidates.

Leak detection and estimation

In order to apply the proposed leak detection and estimated method, the network was split into two distinct areas where the demand forecast could be adapted: *Area A* and *Area B*, containing flow meters at the outlet of reservoirs and tanks; and *Area C*, containing AMR devices.

Area A and Area B

Before applying the leak localization methodology, certain data analyses have to be performed to create leak-free historical data using the information from the flow meters during 2018. Measured data are available every 10 minutes, but it was filtered every hour in order to obtain an hourly demand prediction (T_s = 1 hour). In addition, daily periodicity has been considered, i.e., $T = 24$. A polynomial has calibrated every hourly flow prediction (feature) variation throughout the year to obtain the demand forecast with a higher accuracy, considering the variation in the behaviour during the year, which could reach more than 5 l/s on the time of the day analysing:

$$\hat{y}_{h,q_{day}} = \sum_{i=0}^{n_p} c_i^h {q_{day}}^i \tag{16}$$

where $\hat{y}_{h,q_{day}}$ is the demand estimation at hour $h = 1, ..., 24$ (first 1 am and the last 00 am) of day

q_{day}; c_i^h are the coefficients of polynomial at hour h and n_p is the order of the polynomial.

The leak detection was carried out according to the proposed leak detection and estimation method, considering $N_W = 24$ (i.e. one day), using the history of free leak-data to calculate the maximum possible error and creating a threshold $\lambda = 2.5\ l/s$.

As the inlets of Area A and B are the same, when a leak is detected considering demand estimation (16), if a drop of pressure is observed in the inner sensor of Area B, it is assumed that the leak is in Area B and otherwise it is assumed that the leak is in Area A.

The result obtained in 2018 with all reported leaks that were detected is shown in Fig. 3(a).

Area C

For this area, due to the AMRs measurement corresponding to 89% of the residences, a more accurate forecast demand was calibrated. Using the first week of the inlet flow and the measurements of the AMRs, a constant K was computed, being the percentage value between both flows. The following equation shows the demand forecast for this area:

$$\hat{y}(k) = K \sum_{i=1}^{n_m} \mathbf{q}_{AMR_i}(k) \tag{17}$$

where $\mathbf{q}_{AMR_i}(k)$ $n = 1, ..., n_m$ are the flow measurements at instant k of the n_m AMRs installed in Area C. The leak detection method was applied considering $T_s = 1$ h, $T = 24$ and $N_W = 24$, as in Areas A and B, and a estimated threshold $\lambda = 0.3\ l/s$, that is much lower than the leaks present in Area C. Using the proposed method furnished the results shown in Fig. 3(b): a leak of $2\ l/s$ at the beginning of the year, which was not reported by the water utility; and another leak at the beginning of July with a magnitude of $4.5\ l/s$ and fixed in August.

Application of the data-driven methodology

The data-driven approach is divided into its two composing stages, in order to properly expose both their separated and complementary operation in the BattLeDIM case study. To remark that, to face the challenge, the method is applied individually to each time-instant, obtaining a localization result for each one of them. The leak detection method provides the leaks starting time, so that the localization can be applied at the moment of their appearance.

Graph-based state interpolation

The graph-based state interpolation is applied first to recover the complete state of the WDN for both nominal and leaky situations. The supplied EPANET (Rossman 2000) model of the WDN is used to extract the network topology, as well as the node elevations to compute hydraulic heads from the pressure measurements. To perform all the necessary EPANET operations, the EPANET-Matlab-Toolkit (Eliades et al. 2016) was employed. Interpolation results are compared in Fig. 4 for the case of the leak at pipe *p158*, considering three possible scenarios regarding the availability and nature of the represented data: (a) simulated states from the available hydraulic model for a nominal situation; (b) interpolated states from the available measurements for a nominal case; (c) interpolated states from the available measurements during a leak event. The hydraulic head is represented by the node colour: the darkest the node, the lowest the value (all the interpolations are achieved settling the α parameter of Equation (14) to 1000).

The great similarity between the results of a leak-free EPANET simulation (Fig. 4a) and the estimation of the complete state of the network using nominal data (Fig. 4b) confirms the suitability of the interpolation method and the introduced relaxation of the relation among neighbouring nodes. Besides, an important reduction in the state value can be found if comparing the area that is circled in red at the figure associated to the leak event (Fig. 4c) with the same area at the other two subfigures. This indicates the presence of the leak at the highlighted area.

Leak candidate selection method

Once the complete network state has been computed for both the nominal ($\mathbf{x}_{nom}$) and leak ($\mathbf{x}_{leak}$) scenarios, the leak candidate selection method can be used. These vectors are arranged to generate a set of n two-dimensional points, represented as showed in Fig. 5a-b. In this case, Fig. 5a is produced using two (different) nominal vectors, in order to compare two healthy-situations, while Fig. 5b considers $\mathbf{x}_{leak}$ to contain the interpolated state in Fig. 4c, which is caused by a leak.

On the one hand, Fig. 5a confirms that the cloud of points for two healthy events is mostly aligned, as the slight differences between expected (by the best-fitting line) and actual coordinates of some points are due to the boundary conditions. On the other hand, Fig. 5b shows the leak effect: the cloud of points is no longer aligned, and a set of them is substantially further to the line in comparison with the rest. The points below the line are the most interesting ones to be considered as candidates by the method. Considering the standard deviation of the distance vector to be a threshold for the selection of the candidates, the result depicted at Fig. 5c can be obtained. It includes a representation of the set of candidates in comparison with the complete network (left), and a zoomed subplot highlighting them in colours depending on the probability of being the leak origin (right). The localization is successful due to the reduced pipe distance between the real leak and the most probable pipe, as highlighted at zoomed subplot of the figure through the colour map.

Complete data-driven strategy

Both the graph-based state interpolation technique and the leak candidate selection method are encapsulated into a complete methodology that receives hydraulic measurements from a single time instant and uses them, together with the network topology, to yield a set of candidates to be the leak location. As the datasets of measurements for the complete years are available, the method can be sequentially applied to every time instant (or a subset of them), so that the evolution of the set of candidates can be analysed to assess the existence of leaks.

Besides, as there is not a provided nominal dataset, two options are available to compute this information: to use the hydraulic model and to use data from the provided dataset. The former is discarded due to the data-driven nature of the methodology. Therefore, a certain time window must be selected as the nominal reference for the leak localization process. This characteristic can be exploited to obtain more precise localizations in the presence of old leaks, as their effect will be present at both the nominal reference and the leaky data.

Application of the model-based methodology

The model-based method was applied in two different parts of the network comparing the hourly average value of inner pressure sensor measurements with pressure estimations computed every hour by the hydraulic model. Firstly, it was used in *Area C*, using the measurements of the AMRs in the respective nodes on the demands, and an average was implemented in the nodes without

AMRs. Secondly, the method was used in *Area B*, using the first-week node demands collected from the provided EPANET model.

Area C

The proposed leak detection and estimation technique is first applied in order to detect leaks and their magnitudes. As explained in the methodology, individual analysis of each leak is needed to discover the probability of the location in the network, always starting with the first leak detected.

Fig. 6a-b shows the result of the leakage location in 2018, with the red line indicating the pipe containing the leak. The leaks were analyzed separately, starting with the first leak, and later it was added to the system to study the second leak.

Following the analysis of 2019, the leak of Fig. 6a must be added, as it was not fixed in the previous year. Fig. 6c-d shows the result for the localization leaks, remark that the area in yellow is the nodes with more leak potential, all close to the faulty pipe.

Area B

Additionally, the leak detection and estimation method has used the information of the inner pressure sensor installed in this area with the hydraulic model to determine what is the most probable leak node location (see Fig. 3a).

A leak was detected each year so the model-based method was applied. Fig. 6e-f show the leak localization results for years 2018 and 2019 respectively.

BATTLEDIM2020 RESULTS

The application of the detection and localization methods to the BattLeDIM 2019 dataset yielded the results at Table 1. They translated into a third place in the competition, with a true positive rate of 43.47% and only 1 false positive, producing savings of 210,772 €.

Several facts about these results must be highlighted to explain the methods performance:

- The model-based approach is applied to *Area B* and *Area C*, that is, to leaks *p280*, *p277* and *p680*. The localization is accurate, with only 1-2 pipes between the real leak and the detected one (the leak from 2018 affecting *Area C*, i.e., *p257*, was detected and localized, but this was not submitted as a result due to a misunderstanding).

- Regarding *Area A*, the data-driven method localized the detected leaks with a significant accuracy too, except in the case of the leak at *p142*, due to the reduced density of sensors in the area of this localization. Moreover, the application of the data-driven methodology to single time instants at a time allows to obtain incredibly fast localization results, with only a few minutes or hours of difference in some cases, e.g., *p523, p827, p331* and *p142*.
- Consulting Table 1, there were several leaks that were not detected, and hence, they were not submitted. However, the localization methodologies were indeed able to locate them correctly, as it can be appreciated at Fig. 7 (to remark that the figures for leaks at *p426, p455* and *p879* were obtained by means of custom experiments, that is, using as nominal data a week that allow to get rid of remanent leaks). The detection of these leaks was hindered because they were partially masked by other leaks during the detection phase, and their influence in the localization stage was not considered important enough to be submitted, due to the hard penalty on false positives in the BattLeDIM2020 challenge.

CONCLUSIONS

This article presents the application of two complementary strategies to solve the leak detection and localization problem for the case study proposed by the BattLeDIM challenge. On the one hand, a data-driven technique is developed to locate leaks by means of the available pressure information. This method is applied to areas with a high density of pressure meters, in order to attain the best possible results. On the other hand, a model-based approach is applied at areas where the existence of demand meters allows to use the available measured demand information to highly increase the localization precision, as well as to zones where the pressure meters density is too low to properly utilize the data-driven approach. Therefore, a maximum reduction of the dependency of the model is achieved while yielding sufficiently accurate localization results.

The results of the application of the proposed methodologies are presented for the dataset of 2019, used to evaluate the localization approaches in the competition. The performance of the methods was evaluated with the third place at the challenge, demonstrating the suitability of the usage of the proposed approaches, due to their natural synergy when exploited in a complementary fashion: adapting their utilization to the studied network characteristics and the information availability (hydraulic model, sensors...), the advantages of each technique are boosted and their drawbacks are diminished, yielding a flexible and powerful leak localization solution.

Finally, it is important to highlight the fact that there are differences between the application of the presented methods to the BattLeDIM2020 case and most real-world cases, i.e., the datasets of the former case are provided to be analysed in an offline manner, while the latter case would be based on an online detection/localization scheme. This fact implies some extra differences that should be pinpointed:

- The multi-leak problem would exist in both cases, although the importance of a solution would be radically higher in the offline case, due to the accumulation of leaks that may not be repaired. In a real exploitation, the localized leaks are repaired to avoid the water loss.
- The influence of uncertainties, noise and sensor precision would be higher in the real-world case. This could imply the necessity of performing averaging and noise/deletion pre-processes, as well as increasing the sampling rate and/or number of gathered samples.

Considering these facts, several future worklines remain open: the application of the methodologies to real cases, in order to face real-world conditions; a deep analysis of the previously commented differences, focusing on aspects like the effect of noise, sampling rates, sensorization properties (amount, placement and precision); as well as the improvement of the techniques to solve known and new problems regarding the leak detection and localization field.

DATA AVAILABILITY STATEMENT

- Some or all data, models, or code generated or used during the study are available in a repository online in accordance with funder data retention policies: hydraulic data and model provided by Vrachimis et al. (2020) (they can be downloaded from https://

battledim.ucy.ac.cy/?page_id=33).

- Some or all data, models, or code generated or used during the study is proprietary or confidential in nature and may only be provided with restrictions (e.g. anonymized data): leak detection and localization algorithms codes.

ACKNOWLEDGEMENTS

The authors want to thank the Spanish national project DEOCS (DPI2016-76493-C3-3-R) and L-BEST (Ref. PID2020-115905RB-C21), as well as the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656).

REFERENCES

Adedeji, K. B., Hamam, Y., Abe, B. T., and Abu-Mahfouz, A. M. (2017). "Leakage detection and estimation algorithm for loss reduction in water piping networks." *Water*, 9(10), 773.

Arifin, B. M. S., Li, Z., Shaha, S. L., Meyer, G. A., and Colin, A. (2018). "A novel data-driven leak detection and localization algorithm using the Kantorovich distance." *Comput. Chem. Eng.*, 108, 300–313.

Blesa, J. and Pérez, R. (2018). "Modelling uncertainty for leak localization in water networks." *IFAC-PapersOnLine*, 51(24), 730–735.

Chan, T., Chin, C., , and Zhong, X. (2018). "Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection." *IEEE Access*, 6, 78846–78867.

Donkor, E. A., Mazzuchi, T. A., Soyer, R., and Alan Roberson, J. (2014). "Urban water demand forecasting: review of methods and models." *Journal of Water Resources Planning and Management*, 140(2), 146–159.

Duan, H., Lee, P., Ghidaoui, M., and Tung, Y. (2011). "Leak detection in complex series pipelines by using the system frequency response method." *J. Hydraul. Res.*, 49(2), 213–221.

Eliades, D. G., Kyriakou, M., Vrachimis, S., and Polycarpou, M. M. (2016). "EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB." *Proceedings of the 14th International Conference on Computing and Control for the Water Industry, CCWI*.

Fanner, P., Davis, S., Hoogerwerf, T., Liemberger, R., Sturm, R., and Thornton, J. (2007). *Leakage management technologies*. Water Environ. Research Foundation.

Ferrandez-Gamot, L., Busson, P., Blesa, J., Tornil-Sin, S., Puig, V., Duviella, E., and Soldevila, A. (2015). "Leak localization in water distribution networks using pressure residuals and classifiers." *IFAC-PapersOnLine*, 48(21), 220–225.

Goulet, J., Coutu, S., and Smith, I. (2013). "Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks." *Adv. Eng. Inf.*, 27(2), 261–269.

Han, Q., Eguchi, R., Mehrotra, S., , and Venkatasubramanian, N. (2018). "Enabling state estimation for fault identification in water distribution systems under large disasters." *2018 IEEE 37th Symposium on Reliable Distributed Systems (SRDS)*, IEEE, 161–170.

Huang, Y., Zheng, F., Kapelan, Z., Savić, D., Duan, H., and Zhang, Q. (2020). "Efficient leak localization in water distribution systems using multistage optimal valve operations and smart demand metering." *Water Resour. Res.*, e2020WR028285.

Kang, J., Park, Y., Lee, J., Wang, S., and Eom, D. (2017). "Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems." *IEEE Trans. Ind. Electron.*, 65(5), 4279–4289.

Kapelan, Z., Savić, D., and Walters, G. (2003). "A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks." *J. Hydraul. Res.*, 41(5), 481–492.

Kim, Y., Lee, S., Park, T., Lee, G., Suh, J., and Lee, J. (2016). "Robust leak detection and its localization using interval estimation for water distribution network." *Comput. Chem. Eng.*, 92, 1–17.

LeChevallier, M., Gullick, R., Karim, M., Friedman, M., and Funk, J. (2003). "The potential for health risks from intrusion of contaminants into the distribution system from pressure transients." *J. Water Health*, 1(1), 3–14.

Leflaive, X. (2012). "Water Outlook to 2050: The OECD calls for early and strategic action." *Global Water Forum*.

Li, R., Huang, H., Xin, K., and Tao, T. (2015). "A review of methods for burst/leakage detection and location in water distribution systems." *Water Sci. Tech.: Water Supply*, 15(3), 429–441.

Liemberger, R. and Wyatt, A. (2019). "Quantifying the global non-revenue water problem." *Water Supply*, 19(3), 831–837.

Malik, O. (2016). "Probabilistic leak detection and quantification using multi-output Gaussian processes." Ph.D. thesis, University of Southampton,

Menapace, A., Avesani, D., Righetti, M., Bellin, A., and Pisaturo, G. (2018). "Uniformly distributed demand EPANET extension." *Water Resour. Manag.*, 32(6), 2165–2180.

Mohar, B., Alavi, Y., Chartrand, G., and Oellermann, O. (1991). "The Laplacian spectrum of graphs." *Graph theory, combinatorics, and applications*, 2(12), 871–898.

Mutikanga, H., Sharma, S., and Vairavamoorthy, K. (2013). "Methods and tools for managing losses in water distribution systems." *J. Water Resour. Plan. Manag.*, 139(2), 166–174.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., and Peralta, A. (2011). "Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks." *Control Eng. Pract.*, 19(10), 1157–1167.

Pérez, R., Sanz, G., Puig, V., Quevedo, J., Cugueró-Escofet, M., Nejjari, F., Meseguer, J., Cembrano, G., Mirats-Tur, J., and 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(4), 24–36.

Pérez-Pérez, E., López-Estrada, F. R., Valencia-Palomo, G., Torres, L., Puig, V., and Mina-Antonio, J. (2021). "Leak diagnosis in pipelines using a combined artificial neural network approach." *Control Engineering Practice*, 107, 104677.

Rajeswaran, A., Narasimhan, S., and Narasimhan, S. (2018). "A graph partitioning algorithm for leak detection in water distribution networks." *Comp. & Chem. Eng.*, 108, 11–23.

Rossman, L. A. (2000). "EPANET 2: Users Manual.

Sanz, G., Pérez, R., Kapelan, Z., and Savić, D. (2016). "Leak detection and localization through demand components calibration." *J. Water Resour. Plan. Manag.*, 142(2), 04015057.

Soldevila, A., Blesa, J., Jensen, T., Tornil-Sin, S., Fernández-Cantí, R., , and Puig, V. (2021).

"Leak Localization Method for Water-Distribution Networks Using a Data-Driven Model and Dempster-Shafer Reasoning." *IEEE Trans. Control Syst. Technol.*, 29(3), 937–948.

Soldevila, A., Blesa, J., Tornil-Sin, S., Duviella, E., Fernandez-Canti, R. M., and Puig, V. (2016). "Leak localization in water distribution networks using a mixed model-based/data-driven approach." *Control Engineering Practice*, 55, 162–173.

Sophocleous, S., Savić, D., and Kapelan, Z. (2019). "Leak localization in a real water distribution network based on search-space reduction." *J. Water Resour. Plan. and Manag.*, 145(7), 04019024.

Torres, L., Verde, C., and Molina, L. (2021). "Leak diagnosis for pipelines with multiple branches based on model similarity." *Journal of Process Control*, 99, 41–53.

Vítkovskỳ, J., Simpson, A., and Lambert, M. (2000). "Leak detection and calibration using transients and genetic algorithms." *J. Water Resour. Plan. Manag.*, 126(4), 262–265.

Vrachimis, S. G., Eliades, D. G., Taormina, R., Ostfeld, A., Kapelan, Z., Liu, S., Kyriakou, M., Pavlou, P., Qiu, M., and Polycarpou, M. M. (2020). "BattLeDIM: Battle of the Leakage Detection and Isolation Methods.

Wang, X., Waqar, M., Yan, H., Louati, M., Ghidaoui, M., Lee, P., Meniconi, S., Brunone, B., and Karney, B. (2020). "Pipeline leak localization using matched-field processing incorporating prior information of modeling error." *Mech. Syst. Signal Process.*, 143, 106849.

Wu, Y. and Liu, S. (2017). "A review of data-driven approaches for burst detection in water distribution systems." *Urban Water J.*, 14(9), 972–983.

Xu, Q., Liu, R., Chen, Q., and Li, R. (2014). "Review on water leakage control in distribution networks and the associated environmental benefits." *J. Environ. Sci.*, 26(5), 955–961.

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K., and Fu, G. (2019). "Deep learning identifies accurate burst locations in water distribution networks." *Water Res.*, 166, 115058.

Zhu, X., Ghahramani, Z., and Lafferty, J. (2003). "Semi-supervised learning using gaussian fields and harmonic functions." *Proceedings of the 20th International conference on Machine learning (ICML-03)*, 912–919.

List of Tables

TABLE 1. Results of the BattLeDIM 2020 for the *IRI* team

Real leak	Start time	Detected leak	Detection time	Distance (m)	Distance (nº pipes)
p257	2019-01-01 00:05	-	-	-	-
p427	2019-01-01 00:05	-	-	-	-
p810	2019-01-01 00:05	p798	2019-10-25 06:00	237.48	4
p654	2019-01-01 00:05	p662	2019-02-22 07:00	299.33	5
p523	2019-01-15 23:00	p500	2019-01-15 23:10	192.88	3
p827	2019-01-24 18:30	p64*	2019-01-24 18:50	335.93	6
p280	2019-02-10 13:05	p278	2019-02-10 20:00	98.41	1
p653	2019-03-03 13:10	-	-	-	-
p710	2019-03-24 14:15	-	-	-	-
p514	2019-04-02 20:40	p91	2019-04-02 21:00	249.28	5
p331	2019-04-20 10:10	p360	2019-04-20 10:15	278.91	5
p193	2019-05-19 10:40	-	-	-	-
p277	2019-05-30 21:55	p249	2019-06-26 20:00	119.49	1
p142	2019-06-12 19:55	p650*	2019-06-12 19:55	435.27	9
p680	2019-07-10 08:45	p207	2019-07-10 09:00	113.87	2
p586	2019-07-26 14:40	p563	2019-07-31 19:15	224.31	3
p721	2019-08-02 03:00	-	-	-	-
p800	2019-08-16 14:00	p179	2019-08-25 13:25	74.28	1
p123	2019-09-13 20:05	-	-	-	-
p455	2019-10-03 14:00	-	-	-	-
p762	2019-10-09 10:15	-	-	-	-
p426	2019-10-25 13:25	-	-	-	-
p879	2019-11-20 11:55	-	-	-	-

$*$ These leaks are not considered to be detected due to the detection range of 300 m

List of Figures

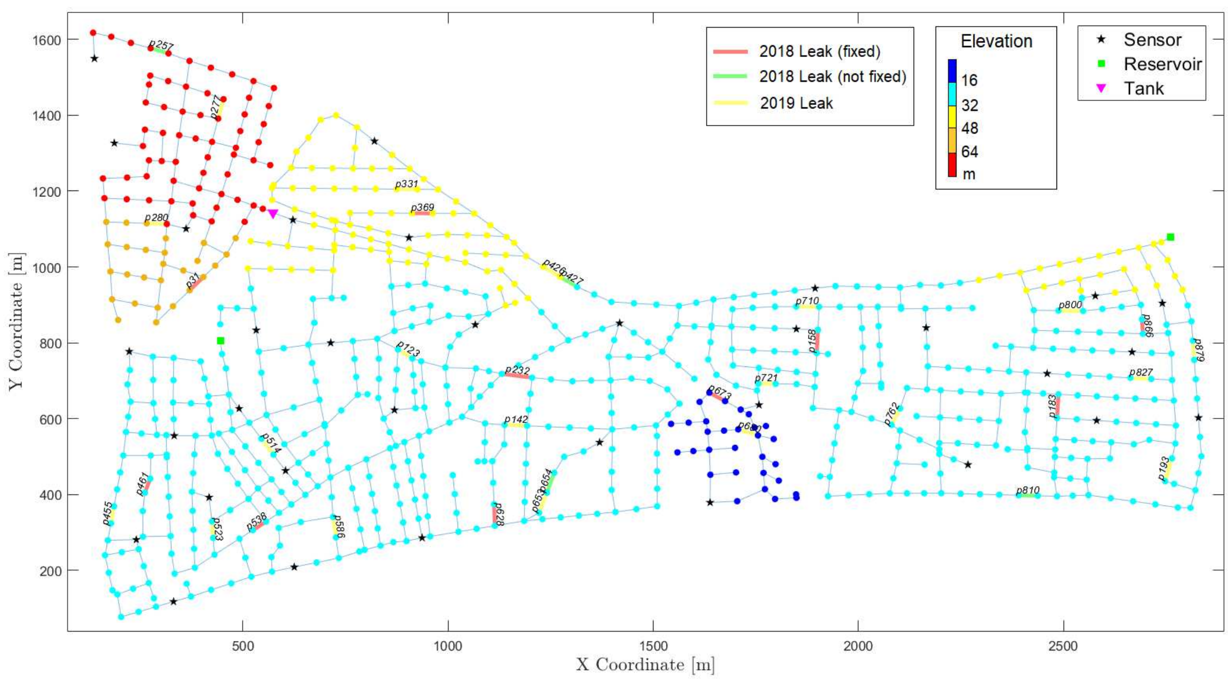


Fig. 1. Schematic representation of L-Town.

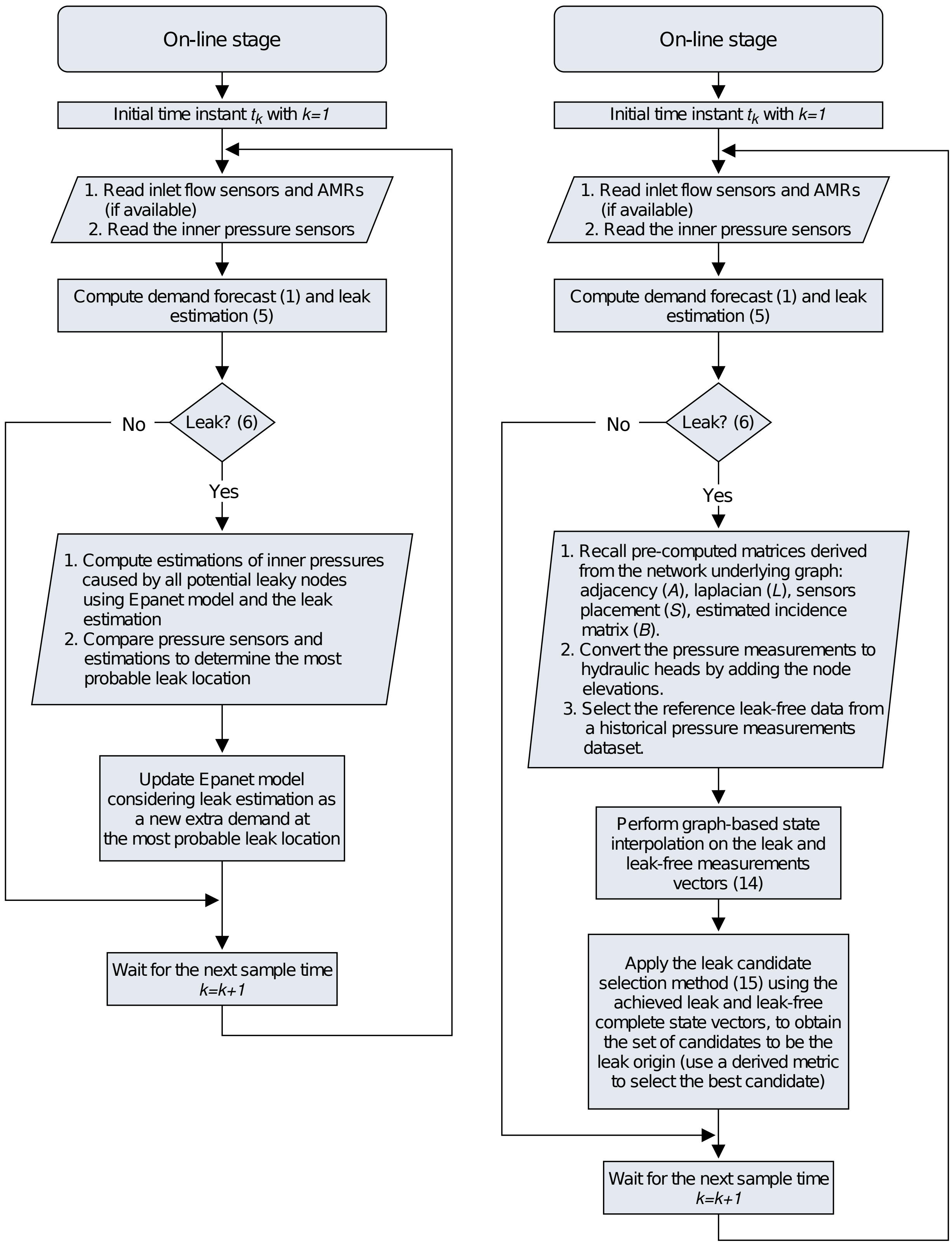
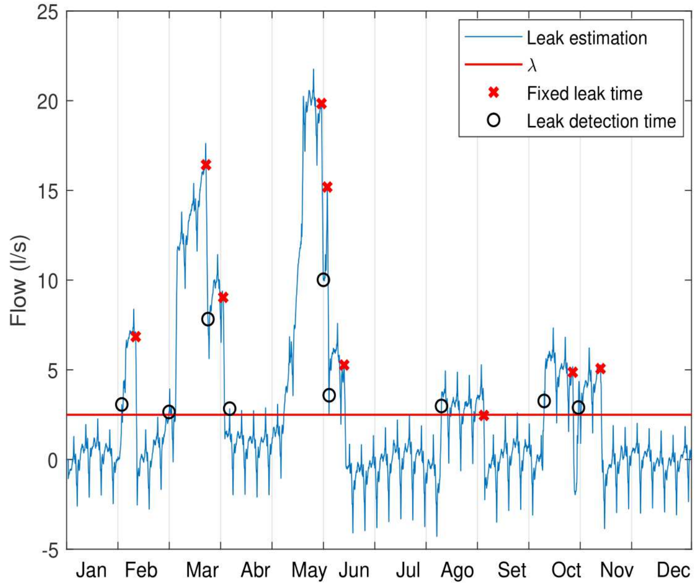
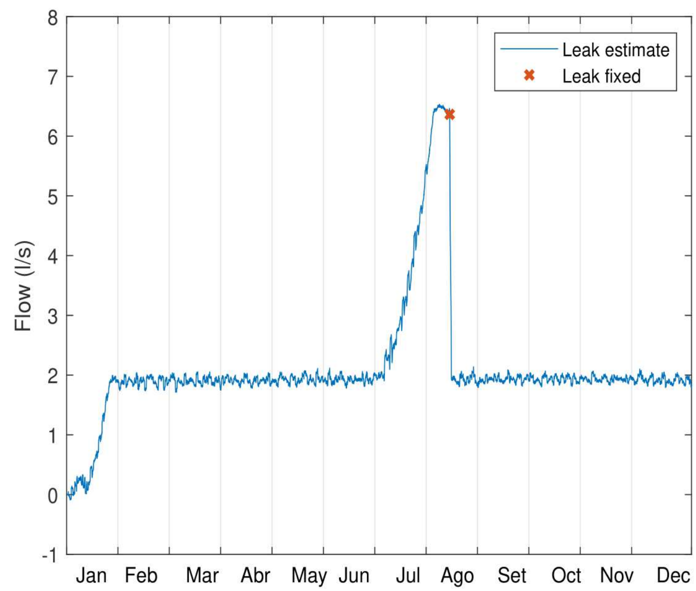


Fig. 2. Schematic flowcharts of the application of the leak detection and localization methodologies: (left) model-based; (right) data-driven.



(a)



(b)

Fig. 3. Graph of leak detection results in 2018 (a) Analysis of the input flow corresponding to Area A and B, with the black markers “o" being the time of detection and the red markers “x" are the leak fix; (b) Analysis of the input flow corresponding to Area C.

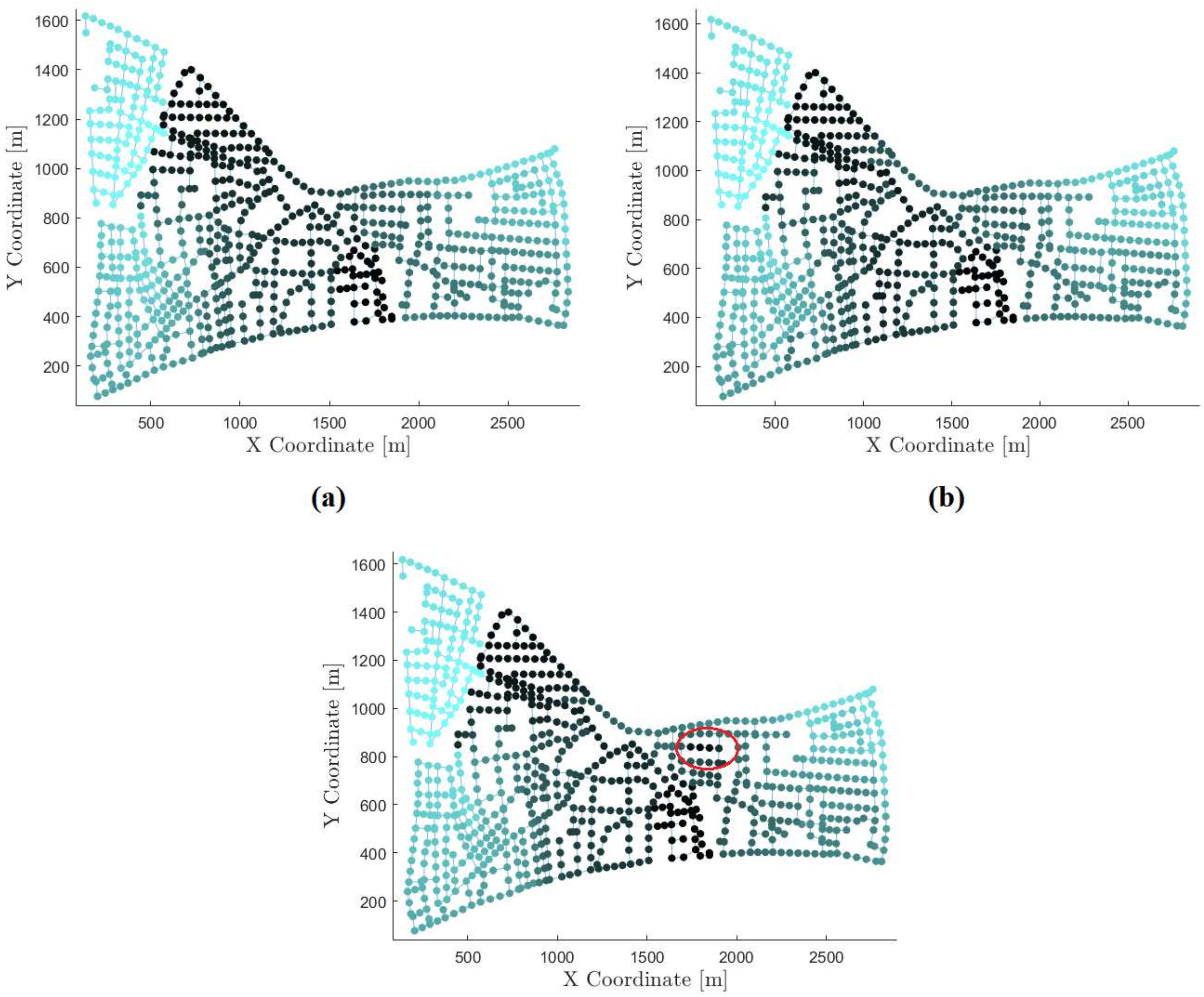
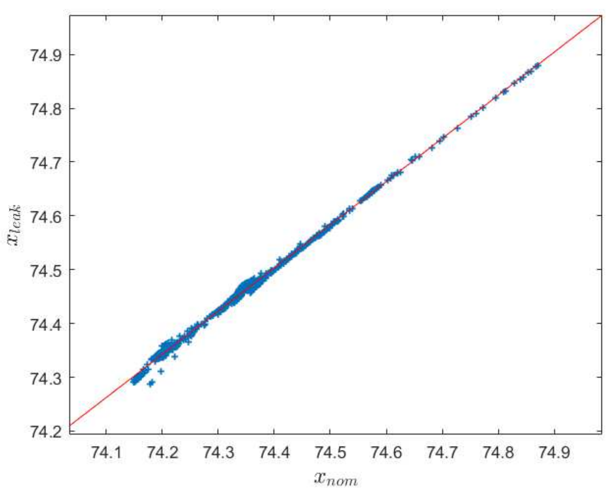
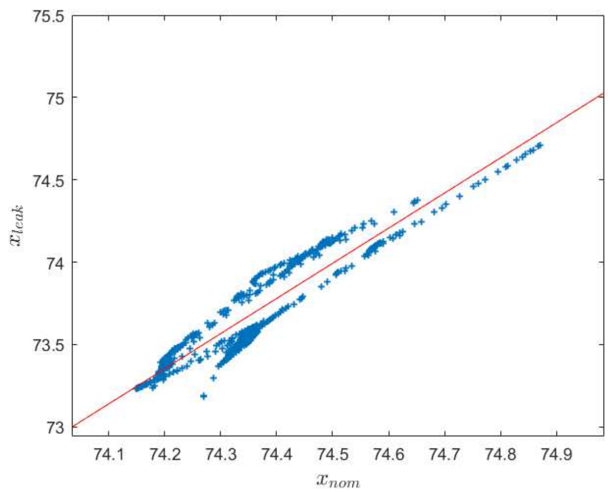


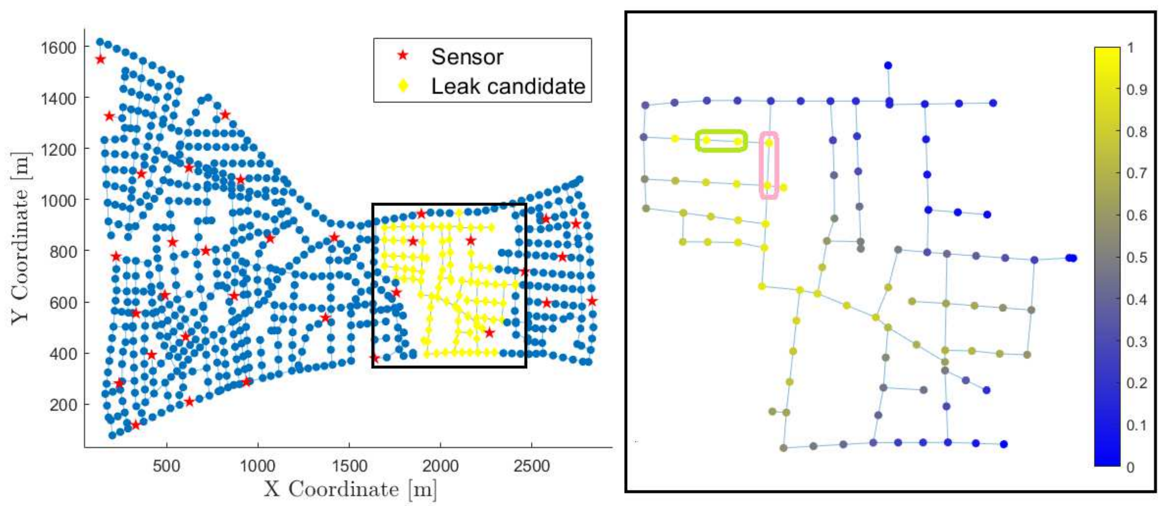
Fig. 4. Graphical comparison of the interpolated states for the case of a leak at pipe *p158* among the three possible scenarios regarding the availability and nature of the represented data: (a) Nominal EPANET data; (b) Nominal interpolated data; (c) Interpolated data for leak *p158*



(a)



(b)



(c)

Fig. 5. Graphical results of the leak candidate selection method: (a-b) Representation of the generated clouds of points for the leak-free (a) and leak (b) scenarios (blue markers) together with the best fitting line (red line); (c) Global localization result showing the complete network and highlighting the leak candidate nodes in yellow (left), and local localization result, illustrated by a colour map with blue representing the least probable candidates, and yellow indicating the most probable ones (right).

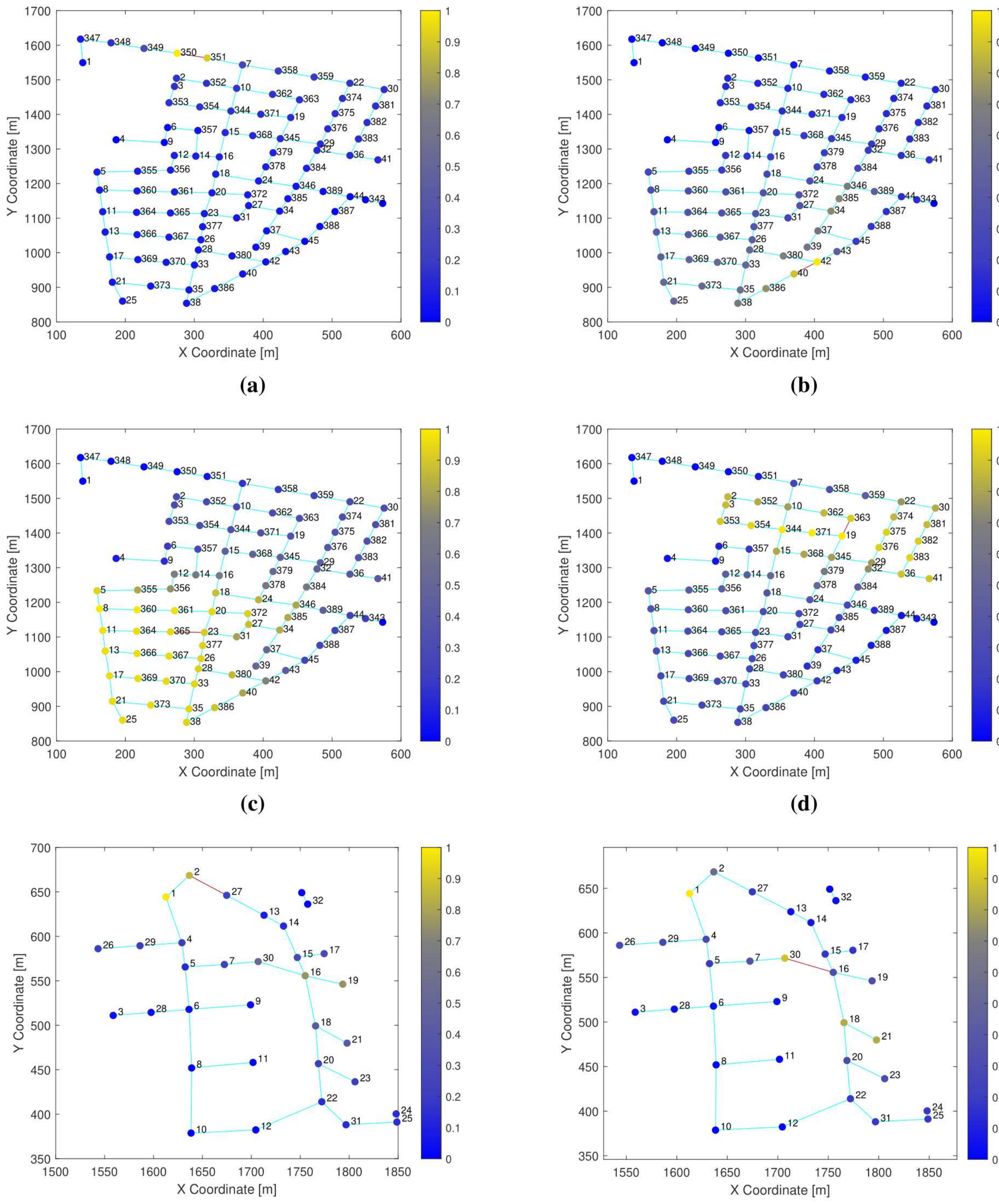


Fig. 6. Graphical representation of the localization result for the following leaks: (a) *p257*; (b) *p31*; (c) *p280*; (d) *p277*; (e) *p673*; (f) *p680*.

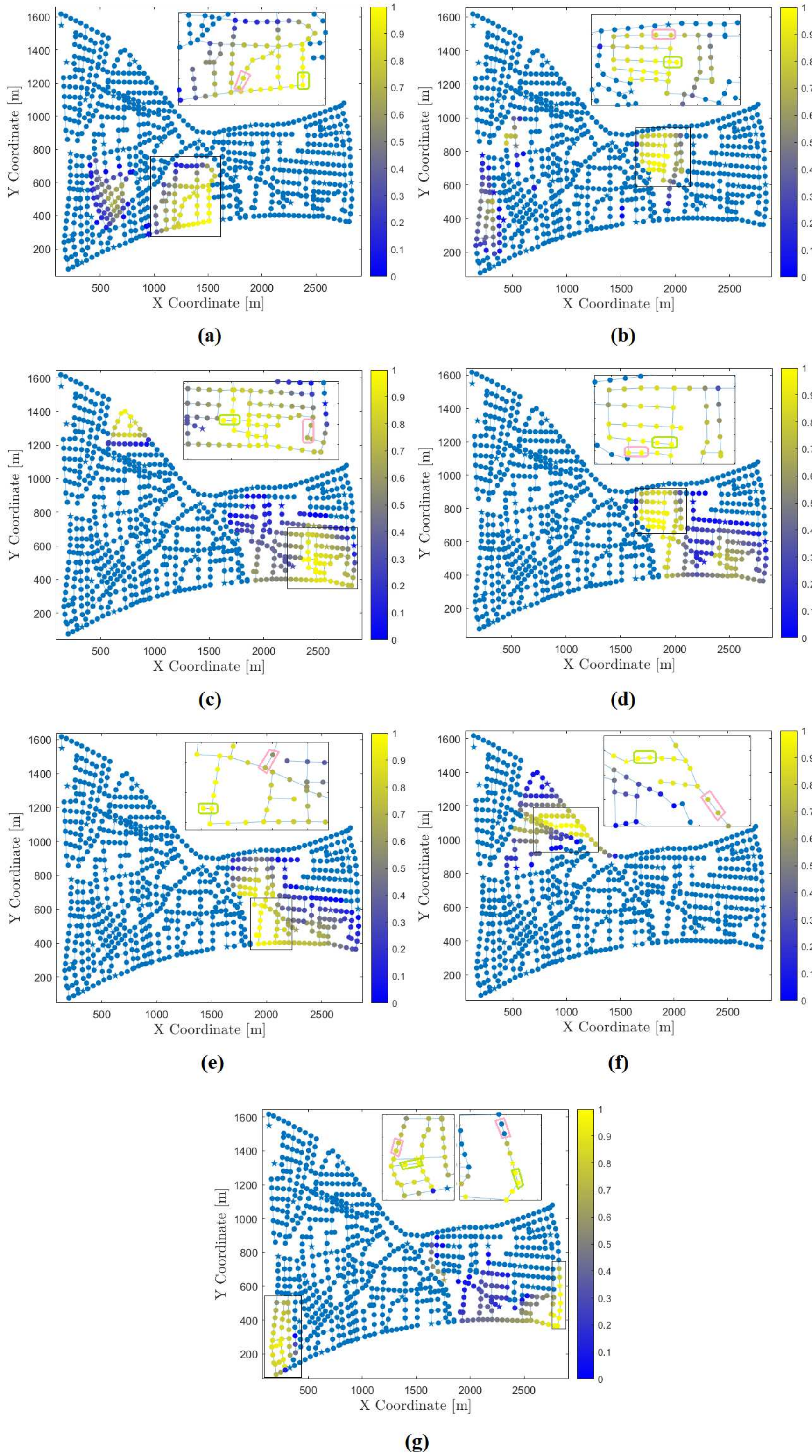


Fig. 7. Graphical representation of the localization result for the following undetected leaks: (a) *p653*; (b) *p710*; (c) *p193*; (d) *p721*; (e) *p762*; (f) *p426*; (g) *p455* & *p879*.