Highlights

Leak detection and localization in water distribution networks: review and perspective

Luis Romero-Ben,Débora Alves,Joaquim Blesa,Gabriela Cembrano,Vicenç Puig,Eric Duviella

- A complete review of the most important leak detection/localization methodologies in water distribution networks is presented.
- State-of-the-art model-based and data-driven approaches are applied to the BattLeDIM2020 benchmark.
- Several general conclusions about the current state of the leak management field are extracted.
- Various guidelines are provided, considering the practical applicability of the methods and the interest of water utilities.

Leak detection and localization in water distribution networks: review and perspective★

Luis Romero-Ben[a,*], Débora Alves[b,d], Joaquim Blesa[a,b,1], Gabriela Cembrano[a,b,c], Vicenç Puig[a,b] and Eric Duviella[d]

[a] Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer Llorens Artigas, 4-6, Barcelona, 08028, Spain
[b] Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politècnica de Catalunya, Campus de Terrassa, Gaia Building, Rambla Sant Nebridi, 22, Terrassa, 08222, Barcelona, Spain
[c] CETaqua, Water Technology Centre, Carretera de Esplugues, 75, Cornellà de Llobregat, 08940, Barcelona, Spain
[d] Informatics and Automatics Department, IMT Lille Douai, Lille, F-59508, France

ARTICLE INFO

Keywords:
leak detection
leak localization
water distribution networks
model-based
data-driven

ABSTRACT

In this paper, leak detection and localization in water distribution networks will be reviewed. In particular, the paper presents the evolution of the methods from model-based towards data-based approaches, listing, describing and comparing the main and/or most recent methods of both categories. Besides, the practical applicability in real water utilities of different existing methods is discussed, outlining the advantages and limitations of model-based and data-driven methods for this task. A well-known case study is used to compare some of the more promising methods and illustrate their performances. Perspectives of the future evolution of the current existing methods are also provided.

1. Introduction

Clean water is one of the most vital resources for humans to survive. Its supply is essential for the daily life of the world-wide population, as well as for the operation of the most important sectors of the economy: agriculture, industry, services, etc.

Nevertheless, the availability of water is reaching an alarming situation:

- Nowadays, around 4 billion people experience severe water scarcity during at least one month of the year, according to Mekonnen and Hoekstra (2016).
- On top of that, the world-wide population has been continuously growing, augmenting from 1 billion in 1800 to 7.9 billion today (Roser et al., 2013).
- This development will be mostly focused on cities, considering that a 83% of the developed world and a 53% of the developing countries inhabitants would be living in urban areas by 2030, according to Cohen (2004).
- Therefore, it is not surprising that the worldwide demand of water is estimated to grow by a 55% between 2000 and 2050 (Leflaive, 2012).

Thus, the efficient and effective management of water distribution networks (WDNs), which are in charge of conveying water to the final consumption points, is crucial for the development and sustainability of urban areas in modern society. However, from the total amount of distributed water, a 30% is estimated to be lost due to the appearance of leaks (Puust et al., 2010). This estimation has been refined to 126 billion cubic meters of water per year worldwide (expressed as non-revenue water) in Liemberger and Wyatt (2019). These estimations illustrate the

★ The authors want to thank the Spanish national project 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).

* Corresponding author

luis.romero.ben@upc.edu (L. Romero-Ben); adeboracris@gmail.com (D. Alves); joaquim.blesa@upc.edu (J. Blesa); gabriela.cembrano@upc.edu (G. Cembrano); vicenc.puig@upc.edu (V. Puig); eric.duviella@imt-lille-douai.fr (E. Duviella)

ORCID(s): 0000-0002-4790-2031 (L. Romero-Ben); 0000-0003-3207-4189 (D. Alves); 0000-0002-5626-3753 (J. Blesa); 0000-0003-1436-6022 (G. Cembrano); 0000-0002-6364-6429 (V. Puig); 0000-0002-1622-0994 (E. Duviella)

1 Serra Húnter fellow.

high economical, operational and social costs of water leaks (Gupta et al., 2017), which are even more aggravated by the associated increase of the risk of contamination (Xu et al., 2014) and health problems (LeChevallier et al., 2003). Therefore, water utilities pursue the aim of minimizing the impact of these adverse effects of leaks. This goal justifies their interest in the development of methodologies for the detection and/or localization of leaks in WDNs. Moreover, this motivates the studies on the field, so that several research groups continuously work on the design of new leak management approaches.

Recently, important reviews about this topic have been published, with Wan et al. (2022) and Islam et al. (2022) standing out. Our review is a valuable addition to the state of the art of the leak management field and completes the previous works due to several contributions. First, we include in our review a complete scope of the literature in the leak management topic with over 90 articles, widening the perspective about the state of the art. These articles are exhaustively categorized, dividing the different families of methods by considering their core algorithm. Moreover, the included tables are designed to be useful to water companies when selecting a leak management strategy, comparing a set of crucial articles from a practical and implementation-based perspective, complementing the design and performance information detailed through the text. Additionally, our review contributes with something new in comparison to previous similar works, which is the inclusion of a performance comparison of two opposite types of leak localization strategies, applying them to a common benchmark from the BattLeDIM2020 competition, in order to justify and support the conclusions extracted during the review of the state of the art. See Vrachimis et al. (2022) for all the details about the competition preparation, participants and results.

The rest of the article is structured as follows. Section 2 introduces the main characteristics and key parts of the leak management problem, as well as providing some clues about the mathematical modelling of the involved water system. Section 3 clearly defines what leak detection and localization imply, limiting the kind of methodologies that are considered, depending on the nature of the method and the ease of application; proceeding to classify the considered methods into categories depending on the philosophy behind them, their requirements, etc. A discussion about the evolution, characteristics and advantages/limitations of each category is included, and a comparison table regarding the point of view of the interests of water utilities is presented. Section 4 provides the application of two methods to a well-known case study, allowing to extract more solid conclusions about how this families of methods operate. Finally, Section 5 exposes several conclusions about the review and some future work lines that can be interesting to pursue.

2. Problem statement

The general scheme of the journey of water from the available sources to the end consumer points/water receiving bodies is known as the urban water cycle (Ocampo-Martinez et al., 2013). Among its phases, let us highlight the delivery of clean water to consumers and industries through pressurized pipe networks, which is performed by water distribution networks (WDNs) through the exploitation of intermediate or booster tanks, pressure and flow control elements such as pumps and valves, etc. Typically, these WDNs are organized in different Pressure Management Zones (PMZs), i.e., areas that maintain a common pressure level, benefiting both water utilities and consumers by guaranteeing an adequate service to the customers during the day and avoiding background leaks during the night (Vicente et al., 2016). Additionally, these PMZs are normally segmented using boundary valves, yielding a set of smaller, loosely connected areas with flow meters at their water inlet, called district metered areas (DMAs) (Water Authorities Association, 1980). This division has been reported to help in pressure control and flow monitoring of tasks in the DMA, increasing the accuracy in the consumption estimation and aiding with the leak management tasks (Charalambous, 2008; Scarpa et al., 2016).

The water loss regulation accomplished by this stratification of the network is complemented by means of leak management methodologies. Specifically, we will consider approaches defined by the following characteristics:

- Nowadays, the technological development has led to the operation of leak management methods automatically from one or more control centres using a network of remote sensors (flow, pressure, etc) and actuators (pumps, valves,etc) and a supervisory control and data acquisition system (SCADA). Thus, our work focuses on software-based methods using models or algorithms to process the information received online at the control room (extensive information about hardware-based methods can be consulted in Hamilton and Charalambous (2013); Mutikanga et al. (2013)).

- The pressure/flow sensors typically deployed by water utilities for network control purposes are not prepared to capture transient behaviours, due to the time scale of this kind of events in comparison to the usual sampling rates. Furthermore, transient-based techniques are normally costly in terms of data requirements (for calibration), computation, sensors and IT systems. Therefore, our work will consider methods dealing with a steady-state characterization of the network dynamics (see Colombo et al. (2009); Ayati et al. (2019) for extensive reviews about transient-based leak management methods).

2.1. Mathematical modelling

Let us recall the steady-state behaviour governing equations for a water distribution system. In order to express the necessary physical laws, several notation elements must be defined. An arbitrary node or junction of the network would be denoted as v_i, with $\mathcal{V}$ standing for the set of junctions; whereas the pipe connecting junctions v_i and v_i would be expressed as $e_k = e_{ij}$, where $\mathcal{E}$ is the set of pipes. Reservoirs, tanks, pumps and valves are included in the sets $\mathcal{R}$, $\mathcal{T}$, $\mathcal{P}$ and $\mathcal{B}$ respectively.

The steady-state behaviour of water flowing through pipes is governed by the conservation of mass and the energy balance. In pressurized systems, where the junction nodes can have an associated consumption, the mass conservation implies that the inflow and outflow in a pipe or junction must balance:

$$c_i = \sum_{j \in \mathcal{N}_i} q_{ij} \tag{1}$$

where c_i stands for the consumption or demand of node v_i, $\mathcal{N}_i$ denotes the set of junctions connected to v_i though pipes, and q_{ij} is the water flow through pipe e_{ij}. Note that $q_{ij} = -q_{ji}$, as inflows must be positive and outflows must be negative for the mass conservation computation.

The nodal demands may be measured using Automated Metered Readings (AMRs) if available, as they would provide this information with a high degree of accuracy. If not, they are typically estimated using a flow measurement at the network or DMA inlet and a nodal proportion (based on billing metering):

$$\tilde{c}_i = \alpha_i q^{in} \tag{2}$$

where $\tilde{c}_i$ is the approximated demand at node v_i, α_i is the normalized proportional consumption of node v_i, with $\sum_{i=1}^{|\mathcal{V}|} \alpha_i = 1$; and q^{in} is the total inflow to the network.

About the energy conservation, the difference in energy between two points is given by the energy provided to the flow in the network components between these two points, subtracting the friction-related and minor losses. Thus, the energy equilibrium can be computed for paths between all the different types of elements composing the network (namely junctions, pipes, tanks, reservoirs, pumps and valves)

$$\Delta E = \sum_{p \in \mathcal{P}} h_p^{\mathcal{P}} - \sum_{k \in \mathcal{K}_{path}} h_k^L \tag{3}$$

where ΔE is the energy difference, $h_p^{\mathcal{P}}$ denotes the hydraulic head (pressure plus geographical elevation) added by pump p and h_k^L is the head loss produced across the k-th component of the path $\mathcal{K}_{path}$.

To complete the set of equations, the relation between the flow through a determined pipe and the head loss between its endpoints can be expressed in general form as:

$$h_k^L = h_i - h_j = \tau_{ij} q_{ij}^{\alpha} \tag{4}$$

where h_i, h_j are the hydraulic heads of nodes v_i, v_j respectively, τ_{ij} denotes the pipe resistance coefficient of pipe e_{ij} that depends on the pipe length, diameter and roughness, and α is the flow exponent. There exists several different experimentally-obtained sets (τ, α), depending on the specific pressure drop model in the pipes (i.e. Darcy-Weisbach, Hazen-Williams or Chezy-Manning). Let us consider the Hazen-Williams model, due to its widespread use. In this case, the resistance coefficient τ_{ij} would be given by:

$$\tau_{ij} = \gamma \frac{l_{ij}}{\kappa_{ij}^{\alpha} d_{ij}^{4.87}} \tag{5}$$

where γ is a parameter whose value depends on the units of the subsequent variables; l_{ij}, κ_{ij} and d_{ij} are the length, Hazen-Williams roughness coefficient and diameter of pipe e_{ij}. The flow exponent α in this case has a value of 1.852.

Additionally, let us define the characteristics of a leak and how they can be mathematically modeled. Normally, leaks appear at pipes of the WDN due to malfunctions in the supplying service or aging of infrastructure. Nevertheless, most methodologies have considered the approximation of leaks appearing in the nodes of the network, due to the nature of the developed approaches and for computational advantages (virtual nodes can be defined at the pipes to merge both approaches). Under this assumption, a leak in node v_j can be regarded as an orifice that allows a leakage flow rate q_j^{leak} to be lost from the water network system. Thus, the flow leaving the system through the leak can be formulated as:

$$q_j^{leak} = \lambda p_j^{\gamma} \tag{6}$$

where λ is the discharge coefficient, γ is the pressure exponent and p_j stands for the pressure at node v_j.

This leak rate is usually modeled into the previously presented steady-state equations by considering the water loss as an extra demand. Note that the effect of this outflow must be considered over the total inflow to the WDN. On the one hand, if the leak is considered as an extra amount of water entering the network, its flow can be directly added to the leaky node demand. On the other hand, if it is considered that the total inflow does not increase, i.e., the leak flow is obtained by subtracting a certain amount of demand from the rest of nodes. This can be achieved by modifying the demand pattern distribution (Soldevila, 2018):

$$\alpha_i^{leak} = \begin{cases} \alpha_i - \alpha_i \frac{q_j^{leak}}{q^{in}}, & i \neq j; \\ \alpha_i - \alpha_i \frac{q_j^{leak}}{q^{in}} + \frac{q_j^{leak}}{q^{in}}, & i = j \end{cases} \tag{7}$$

Considering that this work refers to the problems of automatic leak detection and localization within WDNs using the information received online at the control room, let us formally define both problems with regard to the presented formulation (graphical support is provided in Fig. 1-2 for both leak management problems):

Leak detection The detection of leaks consists of the process of assessment of the existence of leaks in the network/DMA using the inlet sensors (typically flow and/or pressure meters). Considering the exposed mathematical modeling, let us consider an arbitrary WDN whose hydraulic state (in steady-state) is defined as $\boldsymbol{x}(t) = [\boldsymbol{h}(t), \boldsymbol{q}(t), \boldsymbol{\alpha}(t)] \in \mathbb{R}^{n_x}$, where t corresponds to the time instant and $n_x = 2n + m$, with $n = |\mathcal{V}|$ and $m = |\mathcal{E}|$.

A leak occurs in node v_l at time t_0. Before the leak appearance ($t < t_0$), the network is considered to be leak-free or working under nominal conditions. Then, the leak detection process would determine the appearance of the leak by means of the evaluation of the available measurements from the current state $\boldsymbol{x}(t)$, i.e., $\hat{\boldsymbol{x}}(t) = \boldsymbol{D}\boldsymbol{x}(t)$, where $\boldsymbol{D} \in \mathbb{R}^{n_s \times n_x}$ is the measurements matrix, that encodes which are the known variables (measured pressures, flows and demand patterns). Thus, the detection operation can be expressed by $\delta(t) = g(\hat{\boldsymbol{x}}(t))$, with $g : \mathbb{R}^{n_s} \rightarrow \mathbb{B}$ denoting the detection function, and $\delta(t)$ stands for a Boolean variable that indicates if the leak is detected or not. Once the leak has appeared in the WDN ($t \geq t_0$), the detection operation should yield $\delta(t) = 1$, whereas it must produce $\delta(t) = 0$ if $t < t_0$. However, note that in real applications, a certain delay would exist between t_0 and the actual detection of the leak, which would be produced at $t = t_0 + t_d$. Thus, the aim of the detection stage would be to minimize t_d. Moreover, once the leak management operations have finished and the WDN state is considered to be nominal again ($t \geq t_0 + t_R$), we require $\delta(t) = 0$ from the methodology.

Leak localization The leak localization problem is defined as the problem of pinpointing the location of the leak over the WDN using information coming from installed sensors. Let us consider that the leak detection process operates effectively and the leak alarm has been raised, because the leak localization problem would only be solved once the

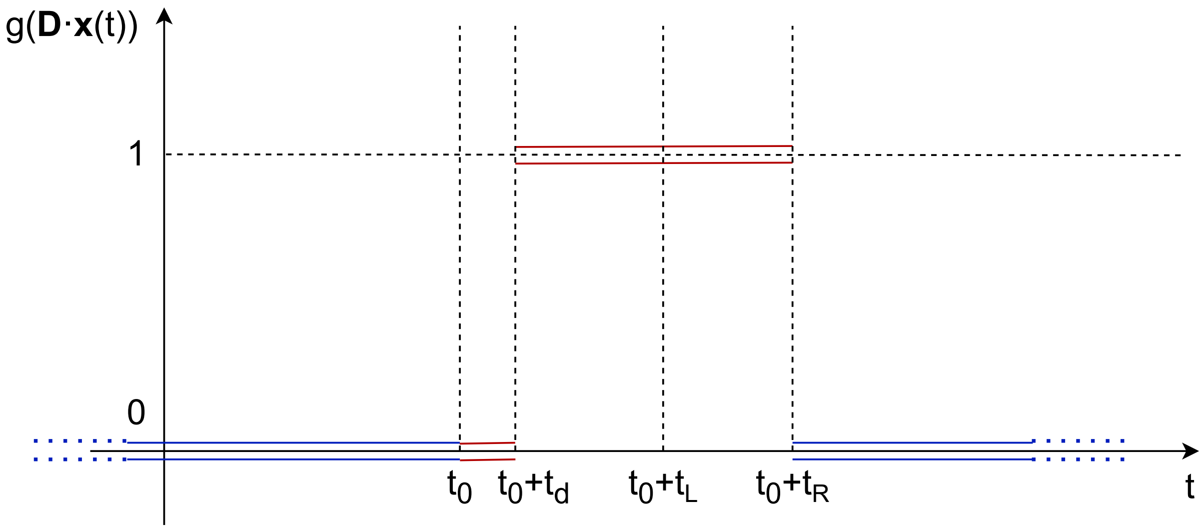

Figure 1: Graphical representation of the timeline of the leak management operation, using the value of the leak detection function to show the workflow of the methodologies: the leak appears at t_0, , it is detected at $t_0 + t_d$, located at $t_0 + t_L$ and repaired at $t_0 + t_R$.

presence of a leak is confirmed. The localization operation can be designed as a function $f : \mathbb{R}^{n_s} \rightarrow \mathcal{V}$ that receives the data from the installed sensors and return a node/set of nodes from the network (note that the isolation methodology can be designed to provide localization areas instead of single-nodes). Thus, the objective of the localization problem would be the selection of the closest node/area to v_l as possible.

Note that both the detection and localization processes can be designed to require several data instances over time, i.e., $\hat{\mathbf{x}}(t), \forall t = t_k, t_{k+1}, ..., t_{k+T}$, with T denoting the required period of time. However, it is of upmost interest for water utilities to reduce the time that leaks are losing water, so leak managenement methods are usually designed to minimize T.

3. Classification of leak detection and localization methods

During the years, the monitoring of water distribution networks has been studied with the intention of minimizing the effects of leakage. Early works were carried out regarding pipe breakage in WDNs. An analysis of the relation between leakage appearance and pipe materials was proposed in Kettler and Goulter (1985). The reduction of leaks by optimizing valve control was studied in Sterling and Bargiela (1984). A method to evaluate and determine water losses is introduced in Arreguín-Cortes and Ochoa-Alejo (1997).

From these initial works, the leak management research field has grown and it currently covers a wide range of problems.

3.1. Leak detection and localization

The methods that tackle the detection and localization problems can be classified as model-based, data-driven or mixed model-based/data-driven, according to their dependence on a hydraulic model of the network:

- Model-based: these methodologies use a hydraulic model, implemented in a simulation software, as a high-fidelity representative of the WDN hydraulic behaviour, basing the localization task in the comparison between incoming real-world hydraulic data and simulated information. During the years, the development of the modeling and simulation software tools (Sonaje and Joshi, 2015) increased the interest in this kind of methodologies, leading to a wide range of works throughout the literature.

- Mixed model-based/data-driven: these approaches intend to mitigate the disadvantages of the inclusion of the hydraulic model in the application of the methodology, i.e., the difficulty in the selection and calibration of the

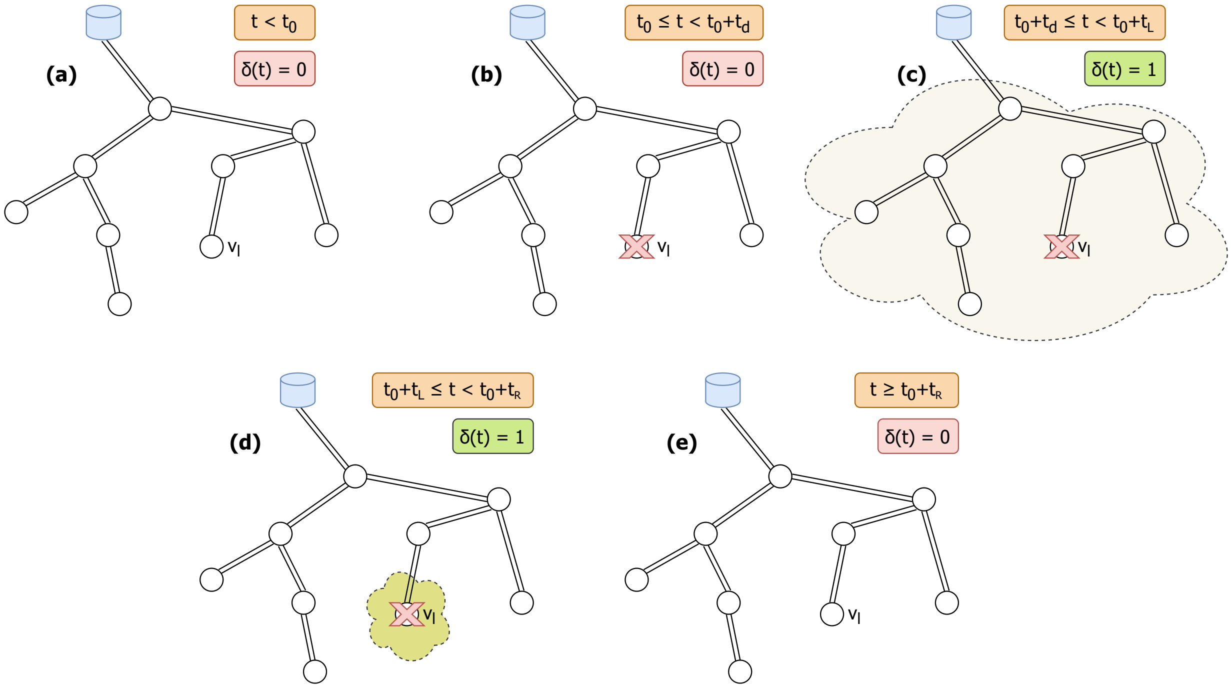

Figure 2: Graphical representation of the evolution of the network status during the complete leak management cycle: the leak appears at t_0, , it is detected at $t_0 + t_d$, located at $t_0 + t_L$ and repaired at $t_0 + t_R$.

corresponding mathematical models (Menapace et al., 2018), the diversity and complexity of WDNs (Kim et al., 2016), and the presence of modeling errors such as nodal demand uncertainties and measurement noise (Blesa and Pérez, 2018). To this end, the usage of the model is reduced to offline processes, e.g., training; additionally leading to the application of machine learning techniques (Ferrandez-Gamot et al., 2015).

- Data-driven: these approaches process the measurements from monitoring devices placed within the network, mining knowledge to address the leak detection/localization problem. Therefore, the hydraulic model is not required, i.e., hydraulic information from simulations is not necessary to apply the methods. Moreover, some methods from this category remove the importance of the availability of individual nodal demand data online, as is difficult to obtain it in most water utilities.

Then, the exposition of the considered methodologies is structured and organized considering this division. To complete the exposition of the different articles and categories, a hierarchical tree summarizing the classification of the methods is presented in Fig. 3.

3.2. Model-based leak detection and localization methods

If the leak management literature is traversed, it is noticeable how there is a majority of model-based leak detection/localization methods in comparison to the rest of the categories (see Fig. 3). During the years, a variety of methods have been proposed, and they can be classified depending on how they use the hydraulic model to address the localization problem.

3.2.1. Inverse problem approaches

The network hydraulic model was initially conceived to solve the direct problem of computing the flows and pressures in all the element of the network, given a set of network parameters, demands and initial conditions. This involves the solution of a set of linear and nonlinear equations. Conversely, the concept of inverse hydraulic problem refers to deriving the network parameters, demands or initial conditions from the values of flows and pressure in the network elements. Clearly, the direct hydraulic problem does not have, in general, an explicit inverse and the inverse

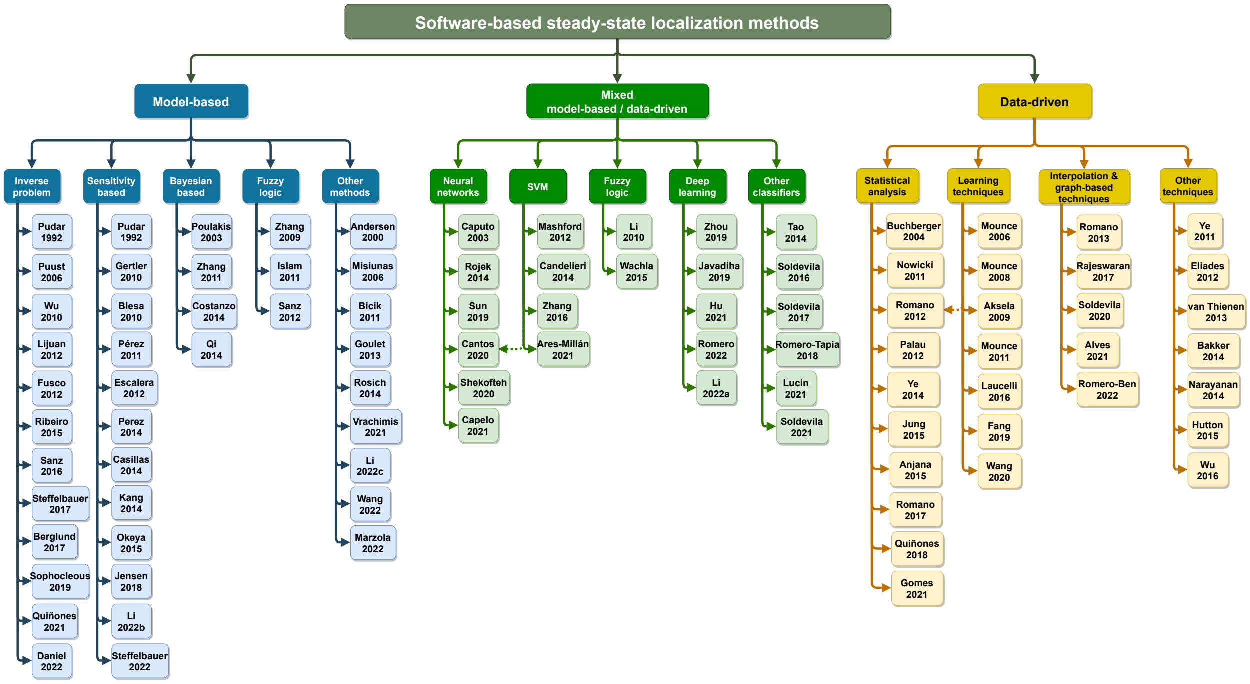

Figure 3: Hierarchical tree of the classification of the leak localization methodologies reviewed in this article.

problem may not have unique solutions. Therefore, finding a solution of the inverse hydraulic problem usually requires the use of optimization techniques as in a parameter calibration process.

One of the first works considering the management of leaks in WDNs was Pudar and Liggett (1992). They posed the leak localization task as an inverse problem, considering the features of the network and the node demands to be given, whereas other variables, i.e., leaks; are unknown. The problem is studied analytically, considering the case of having more measurements and equations than suspected leak points (overdetermined) as well as the opposite (underdetermined). In the former, the primary solution criteria is the minimization of the quadratic difference between measured and calculated hydraulic heads solved (using the Levenberg-Marquardt (LM) method (Moré, 1978)), whereas the latter is solved by means of a minimization problem of the L_2-norm of a vector of orifice areas (one element per leak). Two examples are used to show the capabilities of the methodology, leading to conclude that the underdetermined problem is difficult to solve and provides little help, while the overdetermined problem can be posed to accurately pinpoint the leaks. Thus, the number of measurements is crucial for the performance of the method.

Years later, Puust et al. (2006) presented a leak localization method that is defined as an optimization-based inverse problem seeking the minimization of the total calibration error, which is derived from the quadratic sum of the residuals between measured and modeled hydraulic variables. In this approach, the leak areas are assumed to be random variables following a joint posterior probability density function (PDF). The prior PDFs capture probabilistic beliefs about these areas before solving the inverse problem, while the posterior PDFs encode the beliefs about parameter values when the observed data are known. The problem is solved by means of the SCEM-UA - Shuffled Complex Evolution algorithm (Vrugt et al., 2003), which implements a global optimization scheme to determine the values of calibration parameters. Two case studies are used to assess the performance of the method, exploring single-leak and multi-leak cases. First, the classical aforementioned benchmark is used, and the localization results are comparable to those of the original article. Then, another artificial network is employed, considering noisy measurements. In this case, the results are interesting but deteriorated due to the presence of errors and the high sensitivity of certain nodes to the leaks, which cause the method to select the wrong nodes as candidates.

Similarly, Wu et al. (2010) used inverse analysis to detect and localize leaks, including the leakage locations and flows in an optimization that solves the model calibration problem. A dimensionality reduction process is performed, in which the maximum number of possible leaks is specified, considering that the nodes of the network are divided into several demand groups. The optimization is carried out by means of competent genetic algorithms (GAs), together with a parameter optimization tool. The method is applied in two case studies. The first is based on the typical benchmark from Pudar and Liggett (1992), and the results are excellent for perfect and noisy pipe roughness.The second case study is based on a real network from UK, in which field experiments were performed to emulate leaks, as well as testing with historical leak data. Solid results were achieved, showing the method to be promising. Note that a conceptually similar approach is proposed in Lijuan et al. (2012), so that GAs are used to obtain the optimal leak-related parameters, i.e., leak location and magnitude, for a hydraulic model that is calibrated. The results for a simple benchmark show that the method works for small leaks but struggles when the leak size increases.

In Fusco and Ba (2012), the nodal demands are considered to represent the steady-state of the network, and they are estimated using the hydraulic model and the available measurements, solving an implicit inverse problem. The fault detection and isolation is attained by performing a Z-test over those demand estimations, using historical demand values, billing information, domain expertise, etc., to build a statistical expectation of the demand of each user category. A threshold is used to decide if a demand in a node is large enough to indicate a leak. An example case study is used to assess the method's performance, achieving good results for the tested case with three simultaneous leaks, and analyse the effects of parametric variations with respect to the model. Note that several simplifications are considered, like full availability of nodal pressures, a simple network and noise-free measurements.

Years later, Ribeiro et al. (2015) presented a methodology solving the inverse problem by means of Simulated Annealing (SA) algorithm (Bertsimas and Tsitsiklis, 1993). This method is based on the annealing process, i.e., process of heating up a solid and then slowly cooling it down to reduce its defects. From this analogy, a mechanism to prevent local optimum traps is derived. The information of all possible leaks is gathered through a hydraulic simulator, and the SA scheme attempts to minimize the difference between measured and simulated pressures while deciding the leak size and location. A set of 240 case studies is generated with the aim to illustrate the performance of the method, all of them using the same WDN but using two different pipe lengths sets, two leakage sizes, ten random leak locations and six possible sets of sensor places. The results showed a leeway for action to improve the method, although sufficiently good results could be obtained for most of the leaks.

In Sanz et al. (2016), the leak detection and localization problems were combined with the network model calibration. To achieve this calibration, node demands are composed of geographically distributed demand components, which are used to check the existence and location of leaks. First, anomalies are detected by comparing the values of a set of indicators with corresponding thresholds, derived from the normal functioning of the network (Pearson correlation, conditional overlapping, unit norm, relative increment in mean component values, relative increment in mean consumption and relative residual coefficient). The localization is achieved by two possible methods: the direct one, which assigns the leak regarding the membership of each node to the abnormal demand component; and the leak membership method, which considers the Pearson correlation between the pressure sensitivities to changes in demands for all nodes and the projection of the residuals onto them. The methodology is applied to a real WDN using synthetic data, correctly localizing the leak at almost all the considered scenarios.

Furthermore, Steffelbauer et al. (2017) revisited the inverse problem for leak localization, using the differential evolution (Price, 2013) algorithm in its standard implementation. Several distance metrics are used to calculate the objective function of the optimization problem: Minkowski distance metrics, cosine distance, Pearson's correlation coefficient, the Sorensen metric and the Canberra metric; as well as different ordering schemes of the nodes (rearranging of the search space): alphabetical, random, Cuthill-Mckee algorithm (Cuthill, 1972), and depth-first search (DFS). A case study based on a real-world network from Austria is used to assess the different configurations, regarding the topological distance from leak to candidate and the computational cost. The results shown that the correlation metrics and the Cuthill-Mckee algorithm provided the best results for the considered example leak.

The same year, Berglund et al. (2017) proposed a mathematical programming-based methodology using a simulation-optimization scheme to locate and estimate leaks, solving the multi-leak problem. First, the approach optimizes a linear programming (LP) model to find the linear combination of single simulated leaks that best explain the pressure drop between the leaky and nominal cases, pursuing the minimization of the sum of absolute differences between simulated and measured pressures. The solution also yields the leak estimations, which are included in the network model to perform new simulations. Then, a mixed integer linear programming (MILP) model is used to control the solutions by constraining the total number of searched leaks. To evaluate the methodology, four case studies among three networks are considered: two synthetic benchmarks, Hanoi and Net3 (Diao et al., 2016), and a real network from North Carolina, with simulated data. The synthetic benchmarks yielded acceptable results, which deteriorated for several simultaneous leaks, while the realistic case was successful for up to 4 leaks (note that all nodal pressures are measured in most scenarios).

Moreover, Sophocleous et al. (2019) presented a leak detection/localization approach based on search space reduction and a simulation-based technique similar to Pérez et al. (2011). First, the search space is reduced in three stages: a deletion of nodes from the leak candidate list when they are not demand ones, a minimal detectable node leakage (MDNL) analysis and a search space optimization. The latter aims first to detect the total water losses by minimizing the difference between observed and simulated flows through GAs, to then use several emitter coefficients to find the fittest leak size, and finally execute a simulation per possible leak scenario to find the maximum number of leaks through the minimization of the difference between simulated and observed flows/pressures. The leak location is achieved by solving another inverse problem that searches for the maximum of number of possible leaks based on the fittest scenario from the previous stage. The method is applied in two real-world case studies: the first uses simulated observations while the second employs field data. In the former, the results showed that it was possible to locate the leaks if the pressure drop was larger than the sensor noise. The latter case provided acceptable results considering the size of the network the promising search space reduction and the final localization result.

Quiñones-Grueiro et al. (2021) proposed a model-based leak localization approach based on the solution of an inverse problem using the leak magnitude range derived from the detection/estimation stage. These phases are respectively performed by a deep neural network and a Gaussian process regression, using residuals generated from the comparison between measured and simulated pressures. Localization is achieved through a variant of the differential evolution algorithm, considering the topological information of the network to modify the search space and including a temporal analysis. The Modena benchmark (Bragalli et al., 2012) is used to assess the performance of the methodology, which is satisfactory when considering several possible leak candidates as solution, but reduced when considering node-level localization and for small sets of observations.

One of the most recent works of this kind is presented in Daniel et al. (2022), which tackles the leak detection and localization problems. The former is solved through a semi-supervised linear regression process, operating in a model-free way, while the latter uses a calibrated hydraulic model and a simulation-optimization scheme based on the ideas of Berglund et al. (2017), referred to as Succesive Linear Approximation (SLA). The localization process starts calibrating the node demands by comparing simulated and observed values, to then compute pressure residuals in order to find the most affected sensor for each leak. Finally, SLA estimates the leak and find its location. The methodology was tested over the L-TOWN benchmark from BattLeDIM2020, presenting an improved version with respect to the one whose results were presented to actual competition. The aim of the contest was the detection and localization of a set of unknown leaks from a provided measurements dataset (the network model was also provided) from the year 2019 (a dataset from 2018 was provided for calibration purposes, with several known leaks). The detection and localization performances were excellent, outperforming the BattLeDIM2020 result.

3.2.2. Sensitivity-based methods

Recalling the method in Pudar and Liggett (1992), an early application of sensitivity theory for leak management is provided. In that work, the concept of fault sensitivity matrix (FSM) was introduced as a tool to decide the measurements sites. The FSM stores the effect of each possible leak on every node of the network, and it is obtained by means of hydraulic simulations. This concept evolved, appearing as a fundamental part of the leak localization scheme in coming years, leading to a variety of model-based approaches based on pressure sensitivity analysis, i.e., they study the effect of the possible leaks, that may appear across the network, over gathered pressure measurements.

In Gertler et al. (2010), a two-phase method based on principal components analysis - PCA (Abdi and Williams, 2010), is proposed. First, this technique is used to develop a fault-free model of the system, whose behaviour is compared to fault data from a sensitivity analysis that generates a FSM, used to compute a set of structured residuals from the PCA model. The localization online phase implies the computation of residuals and their comparison against a threshold to indicate the existence of a leak. A simple example was used to show the promising performance of the method, although some simplifications were considered, e.g., all pressures are measured. They conclude that PCA shows its suitability for the task, although the problem characteristics hinder its exploitation.

The use of the FSM in the leak localization process is also presented in Blesa et al. (2010), considering a different approach. First, a detection scheme is proposed, based on the comparison between the behaviour of a Linear Varying Parameter (LPV) model (Bokor et al., 2002), whose structure is obtained from the non-linear mathematical network model, with respect to the behaviour of the real network (considering the sensor measurements); with a zonotope-based (Puig et al., 2001) approach to deal with uncertainty. Then, for the localization, a sensitivity analysis of pressure residuals is performed, computing a FSM that stores the effect of every leak on the generated pressure residuals. The method considers single-leak events, and pursues the argument node of the minimization of the difference between the actual leaky residual and the columns of the fault sensitivity matrix. A simple case study is employed to test the method, yielding satisfactory results.

This FSM strategy is considered again in Pérez et al. (2011). To this end, a binary signature matrix is generated from the FSM, approximated by simulation (using increments of pressure and keeping a constant leak magnitude) due to the difficulty of an analytic calculation (this implies that the FSM would depend on the network demand and boundary conditions). The binarization is performed by assigning a value of 1 to the entries whose associated leakage affects their associated node, and 0 otherwise. Thus, a definition of a binarization threshold is crucial for the success of the localization process. Several real case studies are considered to assess the performance, working in both simulated and real scenarios. The results were interesting and good under ideal conditions, but further work was necessary to enhance the methodology, as the performance decreased due to nodal demand uncertainty and measurement noise.

The usage of sensitivity matrices is considered in Escalera et al. (2012) too. In this work, once the sensitivity matrices are derived, a wavelet analysis method and a phase demodulation process are applied to binarize them. The former provides a more accurate local description and separation of signal features, while the latter is used to extract the information phase of the complex-valued binarized residues (after the wavelet analysis). Finally, the location of the leak is obtained by performing the Boolean Exclusive-OR (XOR) operator to the binarized residues, obtaining a binary matrix whose columns are summed (each row represents a node), to then apply a voting method so that the less-voted node is selected as the leak origin. The methodology is further extended to allow multi-leak localization. Two case studies are used to assess the method: Hanoi and the Quebra network. In both cases, the method is compared with the angle-between-vectors approach (Casillas et al., 2014), and the wavelet analysis method mostly outperforms the other. However, the results are largely degraded when considering measurement and/or demand noise.

The concept of binary fault sensitivity analysis was dismissed by Perez et al. (2014), enhancing the leak isolation process by considering residual fault sensitivity analysis. Moreover, the requirement of defining a threshold to achieve a correct location process is removed. To this end, the residual vector is compared to the theoretical fault signatures of all potential leaks, i.e., columns of the residual-based FSM. This comparison is performed through a correlation function, so that the node associated to the highest correlation value is chosen as the leak origin. The methodology is tested by means of a real leak scenario at a case study DMA from the Barcelona WDN. The results were sound, showing the improvement of the methodology.

In Casillas et al. (2014), five leak localization strategies based on the fault sensitivity matrix are compiled, i.e., sensitivity matrix binarization, correlation comparison, Euclidean distance comparison, angle-between-vectors test and least-squares optimization; and combined with an extended-horizon analysis procedure that makes the localization process less sensitive to demand uncertainty and measurement noise. Their performances are compared for two academic case studies, Hanoi (Fujiwara and Khang, 1990) and Quebra (from EPANET (Rossman, 2000) examples); to then apply the three best methods (discarding binarization and Euclidean distance) to the previously mentioned Barcelona DMA case study. Experiments for different number of sensors are performed with synthetic data, achieving the best performance with the angle-based method, that is then tested over a real leak scenario from the same network.

A similar approach is proposed in Kang and Lansey (2014), i.e., a burst sensitivity matrix is derived from a hydraulic model/simulator of the WDN, using it to place meters and locate bursts. Additionally, this work analyses the control limits of the hydraulic variables, using Monte Carlo (MC) simulations to quantify the variations in pressures and flows depending on demand and roughness uncertainties. These control limits are used during the FSM generation to discard outliers, as well as during the leak location operation. A small example was used to apply the methodology and show is capabilities. The network is interesting due to the existence of multiple tanks, whose filling/emptying varies the flow directions over time. Little results were provided from the leak localization experiments, although they were promising.

In Okeya et al. (2015), sensitivity is used to identify the most appropriate hydraulic locations for pressure sensors. Then, once a leak is detected, the obtained sensorized points are ranked in descending order based on a burst detection metric obtained by using the online burst detection model proposed in their previous work (Okeya et al., 2014). The high-ranked sensors constitute a list of likely leak locations. Then, the DMA demand and leak size are estimated through multiple hydraulic simulations and data analysis techniques, and this information is used to calibrate the hydraulic model, inducing at each run a leak in a high-ranked node, deriving the leak probability through the difference between simulated and measured flow and pressure data. The case study is based on a real-life DMA from England, using real data from the network to generate realistic simulated leak scenarios. The localization results are satisfactory considering the size of the network, and the method performs better at night than during daytime.

Years later, Jensen et al. (2018) extended the method presented in Perez et al. (2014) by considering a preliminary stage to derive a reduced model of the network. To this end, the underlying graph of the network is divided into inlet and inner nodes to derive a expression of the pressure in a certain node of the network that depends on the total flow inlet to the network and a time-varying parameter related to the head at the inlets. This model generates pressure predictions, exploited to derive pressure residuals which are then fed to the FSM generation process. Note that detection is achieved from the computed residuals, by means of a statistical analysis with a test random variable Anderson (2003). A real network is employed, with actual hydraulic data, to assess the performance of the methodology. Promising results are obtained, although some variability of the solution appears when considering a wide range of time slots to apply the method. Also, better results were obtained for larger flows in the system.

Li et al. (2022b) proposed a model-based leak localization methodology based on three steps (detection is handled by an external algorithm from Shao et al. (2019)). A leak simulation stage exploits the concept of dominant sensor sequence, developed from sensitivity analysis, to only consider highly-correlated sensors to the leak. Then, leak scenarios are simulated under varied conditions, calculating the similarity between measured and simulated variables to locate the leak, using metrics like cosine distance, Pearson/Spearman/Kendall correlation coefficient, Euclidean distance, Manhattan distance and Chebysev distance. The localization analysis allows ranking the probability of being the leak origin of a set of network areas, as well as ranking the nodes within those areas. The method is applied to a realistic WDN to evalute its performance. Interesting results were obtained regarding the network area localization, although the performance got deteriorated when considering the node-level uncertainty.

Recently, Steffelbauer et al. (2022) presented an approach combining model calibration, detection and localization of leaks. The iterative calibration of demands and pipe roughness is performed using network measurements, including automatic meter readings (AMRs). The leaks are detected through the CUmulative SUM control chart - CUSUM algorithm and the likelihood ratio test (Basseville et al., 1993), to monitor the virtual leak flow values from a dual model, generated from the network model by including additional virtual reservoirs/valves to transform the hydraulic heads at the sensors into virtual leakage flows. The localization is achieved by a residual projection strategy, similar to the scheme presented in Perez et al. (2014), that computes the Pearson correlation for pressure and flow residuals and the first-order sensitivities. The derived model is updated with the information of managed leaks to allow multi-leak solutions. The method was tested on the L-TOWN benchmark during the BattLeDIM2020 competition, demonstrating its suitability by leading the competition in terms of true positives.

3.2.3. Bayesian-based strategies

Bayesian-based approaches have been derived and implemented during years too. First, Poulakis et al. (2003) proposed a leak localization and estimation scheme based on a Bayesian methodology, coupled with an hydraulic simulator, for the update of hydraulic models whose parameters are the leakage locations and sizes. To this end, flow and pressure measurements are compared with model estimations. The uncertainty of the parameter set is quantified using PDFs measuring the plausibility of the different models. The leak locations are obtained by a set of optimizations that seek the maximization of the corresponding PDFs with respect to the leak size parameters. These optimizations are solved by GAs. The methodology is applied to an example network, considering several scenarios regarding uncertainty, sensor locations and type, single or multiple leaks, etc. The result is sound and the location of the leaks is found when the errors are not excessively high. Besides, the methodology is limited to a certain threshold of leak flow.

Years later, Zhang and Wang (2011) presented a model-based leak detection/localization approach based on Bayesian theory and Fisher's law. The former updates the randomness of the values of leak-related parameters (location and amount) from a hydraulic model (these parameters are quantified through PDFs), whereas the latter is used to estimate the parameter values. Finally, the amount of leaks is estimated by a back propagation neural network, while their location is estimated from the flow characteristics of the WDN. Thus, the methodology works with pressure and flow measurements. A case study is presented to show the performance of the method. A single-leak scenario is considered, successfully locating the leak, despite the existence of large errors when comparing simulated and measured data. Thus, further work is required to improve the methodology.

A Bayesian-statistical approach is also exploited in Costanzo et al. (2014) with the application of a model calibration algorithm known as UNINET (using the SCEM-UA methodology) to solve the optimization problem. The method estimates the roughness in pipes and demands at nodes while considering the parameters as random variables. An index measures the difference between the calibrated pipe roughness before and after the leak, using its information to identify where the leak is located. A real case study from the city of Amantea is used. A set of roughness classes was defined, so the index is computed for all the classes, and the correct roughness class has the maximum index when the number of measurements is high.

In the same year, Qi et al. (2014) proposed a leakage localization method that addresses the problem by means of a combination of GAs and Bayesian Decision Theory (BDT). It pursues the high efficiency and strong robustness of GAs, while using the Bayesian theory to consider both hydraulic model and measurement errors. The BDT model uses a method of exhaustion to gain probabilities of all leakage accidents, and the optimizations are solved by the GA scheme. A sample network is used to evaluate the method, and the results show that despite the node-level localization is difficult, a localization area can be successfully derived.

3.2.4. Fuzzy logic model-based methods

The concepts of fuzzy logic have also been exploited to derive leak management strategies. A methodology is proposed in Zhang et al. (2009) to locate leaks through hydraulic simulations, as well as the generation of a matrix of burst features and the method of fuzzy similarity ratio based on the Hamming distance as a measuring expression. This ratio is compared for the different simulated scenarios and several fuzzy similarity matrices are obtained. These matrices yield a similarity serial number that can be compared to a given threshold. When a leak is detected, the pipe with the smallest serial number is the burst location. The method was applied to a real network, achieving good performance results for different leak sizes.

Fuzzy set theory was used in Islam et al. (2011) as part of a leak detection/localization technique.This approach simulates the minimum and maximum values of dependent parameters (flows and pressures) for an extended period, using this information to fuzzify independent parameters (pipe coefficients, nodal demands, water levels, etc.). They are used to calibrate a hydraulic model and estimate an array of dependent parameters, finally achieving three sets of solutions. Then, monitoring data is evaluated with respect to the fuzzy numbers from these solutions, raising a detection alarm if the pressure is lower than the most likely value. The location process is based on the concept of index of leakage propensity (ILP): if the monitored values are beyond the derived extreme values, then the ILP will be higher than 1, and the highest ILP would indicate the leaky node/pipe. A simple case study is used to implement and demonstrate the methodology performance. Several results are provided, showing promising performance, even though several issues are left to be tackled in the future.

A year later, Sanz et al. (2012) proposed a leak localization method based on Fuzzy Inductive Reasoning - FIR (Nebot and Mugica, 2012). It finds relations between qualitative and causal variables of the system by observing their behaviour for a certain period, using these relations to predict future behaviours. First, leaks are detected by comparing real measurements with fuzzified predicted values, checking if the former lies within the envelope of the latter, and raising an alarm if this is not the case. To locate the leaks, a fuzzy model for all possible combinations of simulated leaks and available sensors is computed, as well as a fuzzy leak-free model for each sensor. These models are used to perform leakage classification, identifying which predicted pressure fits best with the real one. A real-based network from the Barcelona WDN is used to evaluate the performance of the methodology. It is compared with the correlation-based sensitivity method, and the results show that while both method are able to correctly locate leaks, the FIR-based approach is more precise, as the localization areas are smaller.

3.2.5. Other model-based methods

Apart from the mentioned families of methods, approaches which do not fit any of the above-mentioned model-based categories have been developed over the years. In Andersen and Powell (2000), a new implicit formulation of the standard weighted least squares (WLS) state-estimation problem is derived for WDNs. Graph theory is exploited by means of the consideration of loop equations, and the state variables are chosen to be the nodal demands, which are modeled by demand pattern factors that divide the total water consumption among the network nodes. A Lagrangian approach is considered to formulate the optimization problem, and the Netwton-Raphson method is applied to solve it. This formulation is modified to solve the leak localization problem: a test node is selected and the state estimator is altered to ignore the cost of accepting possible excessive demand for that node. This idea was tested over a case study network, and promising results were obtained, even if the scenario is idealized with noise-free conditions.

Years later, Misiunas et al. (2006) proposed a leak localization algorithm based on the comparison of measured and simulated pressure values, after a detection phase based on the CUSUM algorithm is performed and a leak is detected. The comparison is used to evaluate an objective function, for every possible leak location in the network, so that the node with the smallest objective function value is selected as the burst position. An example network is used to verify the methodology. Two kinds of experiments were performed: location of leaks that occur at nodes that are considered as candidate locations, and the opposite case. The former case yielded satisfactory node-level accuracy, whereas the latter required two or more nodes to clearly identified the leak location area, normally selecting adjacent locations to the leaky nodes.

Another different leak localization technique was presented in Bicik et al. (2011), considering the information fusion of the output of several models, i.e., a Pipe Burst Prediction Model (PBPM), a Hydraulic Model (HM) and a Customer Contacts Model (CCM). PBPM computes expected burst frequencies for every pipe during the current month, CCM uses a weighted linear criterion to assign a leak probability to each pipe based on the confidence on the reports of every customer, and HM simulates the effects of leaks over the different pipes, comparing the estimated and measured pressures. The information fusion is carried out using the Dempster-Shafer Theory of Evidence (Gordon and Shortliffe, 1984), which generates a result that encompasses the varying credibility of the individual models, and derives the spatial distribution of Belief and Plausibility of leakage of any pipe in the network. The method was applied using data from a real system of North Yorkshire, UK. The obtained results are not satisfying for the individual models, but improvements can be found when fusing their information. However, the authors find limitations due to the lack of information of the state of the network.

A methodology based on the concept of model falsification is presented in Goulet et al. (2013). Leak scenarios are generated by simulating extra demands at every node of the network alternatively. The method exploits flow velocity information. Thus, the scenario is considered falsified (and discarded) when the difference between simulated and measured values is outside computed bounds, which depend on the model and measurement errors, and the proposed method employs the Šidák correction technique to ensure that the methodology does not wrongly discard scenarios. A case study based on a network from Lausanne is presented, showing results for different leak events, and considering different levels of uncertainty. Its advantages are demonstrated mostly for large-size leaks.

In Rosich et al. (2014), a methodology based on the generation of structured residuals is presented. This class of residuals presents suitable structural properties for leak localization, avoiding the dependence on the leak magnitude that normally affects residual sensitivity analysis through a leak decoupling technique. Additionally, these residuals are numerically computed, so no explicit solution is required to compute them. A case study from the Limassol water network is used to assess the methodology. It is compared to the classic directional residuals method and the results show an improvement in consistency. Additionally, the combination of both residuals types yielded better results.

Recently, a model invalidation approach was presented by Vrachimis et al. (2021). Leak detection is handlded via identification of inconsistencies between a healthy interval model and sensor measurements by iteratively checking the feasibility of an LP problem. The leak is localized by identifying the locations that accomplish the constraints of a re-formulated interval model and the measurements during multiple time instants, computing bounds for unknown leak emitters, so that those with a non-zero upper bound become feasible leak candidates. These candidates are ranked by a localization priority index (LPI), indicating the minimum percentage of nodes that are excluded from the node search space before the leak is found. The method is tested first in the classical Hanoi benchmark considering two leaks and achieving excellent performance. Second, the Leakage Diagnosis Benchmark (Vrachimis et al., 2018) is used to compare the proposed method with a previous work, yielding an improvement regarding the true positive rate.

The BattLeDIM2020 challenge was faced by Li et al. (2022c), presenting a leak management framework dealing with three stages: model calibration, leak detection and leak localization. The first is performed to get a nominal hydraulic model, which is required to estimate the overall yearly leak flows and approximate the pressure at the nodes in nominal conditions. Then, the Loess decomposition (STL) (Cleveland et al., 1990) method and k-means clustering are applied to the pressure residuals from the comparison between nominal (simulated) and leaky (measured) data to identify the existence of leaks. By updating the model at each detection of non-repaired leaks, the localization stage can be performed, selecting the pipe with the highest probability based on the simulated and actual leaks. The application of the method to the BattLeDIM2020 challenge was successful, obtaining the first place in the competition.

Another BattLeDIM2020 approach was proposed by Wang et al. (2022), which exploits both statistical methods and hydraulic modeling to detect and localize leaks. First, the expected flows and pressures in leak-free conditions are retrieved by means of empirical model decomposition (EMD) and vector autoregressive models. From these values and the measured ones, residuals can be computed and used to identify leak appearances and sizes. To locate them, a comparison between observed and simulated pressures is performed at both the week of the detected leak and the week before. The method performed well in terms of true positives, although it also caused several false positives. It finally obtained the fourth position in terms of economic score.

Another BattLeDIM2020 method was presented in Marzola et al. (2022). This approach considers model calibration to adjust the hydraulic model to realistically behave as the network (considering demands to have similar profiles in the complete WDN, using the measurements from the area with AMRs), to then identify leaks through an engineering judgment process of the measured inflows and water demands. The localization is performed by carrying out simulations of every possible leak, and computing the error between each obtained pressures set ans the actual ones from the network, selecting the pipe that yields the lowest error. This method was among the top-performers, achieving the fifth place in the competition.

3.3. Mixed model-based/data-driven leak detection and localization methods

In recent years, the development of machine learning and data analysis methods helped increasing the usage of these kind of techniques to develop new leak localization schemes, as well as complement previously existing ones.

3.3.1. Neural networks approaches

As shown previously, the use of artificial neural networks (ANN) has been widely considered to address the leak detection task, due to the possibility of training the ANN to discern between leak and leak-free states. Nevertheless, this kind of techniques have also been applied to solve the leak localization problem.

In Caputo and Pelagagge (2003), ANNs are used to monitor the states of pipe networks in order to locate leakage. Specifically, a multilayer perceptron ANN - MLP-ANN (Delashmit et al., 2005) with sigmoidal activation function (Rasamoelina et al., 2020) is exploited and trained with pressure and flow data from a set of sensors scattered through the network. The neural network is trained on different sets of leak and leak-free scenarios from hydraulic simulations, with a wide range of operating conditions. To this end, a two-level architecture is considered: a main ANN identifies the branch that contains the leak, whereas the second level is formed by several ANNs, in cascade to the main one, that estimate the leak size and location in the selected branch. An example is considered first to compare the two-level approach with a single ANN scheme, obtaining results that illustrate the advantages of the former. Then, a realistic case study is used to test the performance of the two-level strategy. The trained ANN was able to identify the correct branch in all test cases, and satisfactory performance was reached for the second-level ANNs.

Years later, Rojek and Studziński (2014) followed a similar scheme to test different kinds of neural network structures, trained with datasets of leak- and leak-free scenarios for different operating conditions from a a hydraulic model. In this case, apart from the MLP architecture, Kohonen networks (Kohonen, 2013) were also tested. The latter are self-organizing nets with the capability of adapting to input data that was not previously known. A water network was considered to test the performance of both ANN schemes. Their comparison shown that the MLP outperformed the Kohonen net, although no sources of noise or uncertainty were considered.

Another work dealing with neural networks and classification is presented in Sun et al. (2019). Pressure data from sensors is interpolated using the Kriging spatial interpolation approach (Oliver and Webster, 1990), that allows to obtain approximate values at not measured nodes from the measured ones, additionally considering the network topology. Then, two different types of classifiers are considered: ANNs and linear discriminant analysis or LDA. The former fits a single hidden-layer neural network, while the latter is a method to find a linear combination of features that divides two or more classes of data. They are trained to be ready to be applied to locate leaks from incoming data, using hydraulic information from the possible leak scenarios. The case study of Hanoi is considered to evaluate the performance of both classifiers, considering different sensor configurations. A ANN-scheme with Bayes reasoning led to the best results. However, the sensorization analysis shows that the Kriging method can introduce errors for some sensor configurations, worsening the localization.

In Shekofteh et al. (2020), ANNs techniques and graph theory, combined with pressure sensor measurements were applied for the leak detection/localization problem. A network division/clustering algorithm is used to iteratively split the network, to then identify where the leak is located using information from pressure loggers and ANNs. A hydraulic model is required to obtain the pressure data for the ANN training. The Balerma benchmark (Reca and Martínez, 2006) is used as a case study to test the method. Several scenarios are considered regarding uncertainty sources and number of simultaneous leaks.The results show a perfect functioning for the cases without uncertainty, and a degradation of performance in the presence of uncertainty.

In the same year, Cantos et al. (2020) proposed a machine-learning-based methodology for leak detection and localization based on risk assessment. The approach makes profit of a hydraulic database generated through hydraulic simulations, to then proceed with a risk assessment strategy that establishes a spatial time series of leak likelihood indicators through a statistical data analysis of the measured flows. The localization is performed by both ANNs and SVMs, which are trained to geolocate the leak in the source with a higher risk of occurrence. The campus of the Lille University was used as case study. In this case, ANNs provided a significantly higher accuracy than SVMs for detection, whereas they yielded similar high accuracies regarding localization.

Capelo et al. (2021) recalled again the multilayer perceptron ANN strategy, using the explained scheme to generate hydraulic data for a wide range of leak scenarios, train and test the decided ANN architecture (it depends on a sensitivity analysis) and apply it to a realistic scenario. The LM algorithm is considered as the core optimization mechanism for the ANN. A case study about a real WDN from Portugal is used in order to assess the performance of the method. The ANN is structured to have as input the pressure heads of the sensorized nodes, and as output the leaky node location in Cartesian coordinates. Several possible configurations within the ANN are considered, as well as different analysis regarding number and location of sensors, and number of considered leak scenarios for training. All the results for the different studies are satisfactory and logical, despite the distance uncertainty between the selected leak locations and the real ones may be high.

3.3.2. Support vector machine methods

Support vector machine or SVM is a technique that has been applied to solve a wide range of problems since its development (Pisner and Schnyer, 2020), including leak management tasks. In Mashford et al. (2012), the data gathered from sensors is used to feed SVMs, trained to predict the leak magnitude and location. In this case, EPANET is used to generate a large number of leak scenarios for the SVM training. A WDN from Melbourne is selected as case study to assess the performance of the method. First, SVMs were used as regressors to predict the leak emitter coefficient values when a predefined node is leaking. In this experiment, the results were notably good, with minimal error. Second, SVMs were used as classifiers to locate leaks. The results showed that a satisfactory node-level accuracy, although it decreased when the number of possible leaky nodes increased. However, the area-level localization was quite successful. Finally, advice is provided for the application of the method with low leak rates, although the accuracy results were not satisfactory.

A SVM-based approach is combined with a clustering strategy in Candelieri et al. (2014) to handle the leak localization operation. Hydraulic data is generated using a simulator, achieving information for all possible leaks (in pipes) in the network. Then, the leak-scenario data are clustered in order to divide the possible leaks into different groups, so that the leaks within a cluster produce a similar effect over the network. Specifically, five clustering methods are tested: simple K-means or SKM (Shukla and Naganna, 2014), Farthest-First or FF (Kumar et al., 2013), the Spectral Clustering (Von Luxburg, 2007) versions of both previous methods, and Partitioning Around Medoids or PAM (Van der Laan et al., 2003). Moreover, two metrics are derived and calculated for each cluster to assess their validity, so that a global clustering index can be achieved by their multiplication. Additionally, SVM is utilized to enhance the reliability of the localization operation, due to the non-linear transformation derived from the Spectral Clustering approaches. The methodology was tested by means of two case studies, comparing the different clustering options through the mentioned indices, and Spectral Clustering SKM ended up being the best. Regarding leak localization, it achieved notably high accuracy, although there is not information about the inclusion of noise and uncertainty.

A similar scheme is exploited in Zhang et al. (2016), considering both SKM and SVM to perform leak localization. First SKM is used to divide the network into areas with a similar leak behaviour; then a multiclass SVM (M-SVM) is trained to discern the cluster that best fits the incoming leak data in. The generation of training samples is performed by means of a Monte Carlo strategy, so that this algorithm creates leakage events for each leakage zone, to then run the hydraulic simulator and compute the difference between leak and leak-free scenarios. Two case studies are considered to evaluate the performance of the method. The first example showed that the classification accuracy decreases as the number of zones increases. Similarly, a better performance of radial basis function (RBF) kernel in comparison to polynomial and sigmoid ones is reported. In the second realistic case study, a high accuracy was achieved, although the localization target was a reduced set of WDN areas, and no noise or uncertainty are included during the data generation.

Recently, Ares-Milián et al. (2021) presented a leak localization scheme combining model-based and data-driven methods. First, the network is partitioned using agglomerative clustering (Kaufman and Rousseeuw, 2009), considering its topological characteristics instead of hydraulic data. The localization starts through a leak zone selection performed by means of a trained SVM classifier combined with Bayes temporal reasoning. Then, the inverse problem is formulated and solved using topological differential evolution (TDE), constrained to consider the leak zone selected in the previous step. The Modena benchmark is considered, deriving several metrics to compare the methodology with other techniques, i.e., the SVM and TDE techniques applied standalone. The results show that the hybrid strategy outperforms the other two when the demand uncertainty increases.Additionally, the computational cost of TDE and the hybrid schemes are compared, demonstrating that the search space reduction of the latter speeds up its operation.

3.3.3. Fuzzy logic mixed model-based/data-driven strategies

Fuzzy logic theory has also been used to develop novel neural network schemes, which have been tested to play the role of pattern recognition agents for the leak management tasks. An early work in this line is presented in Li and Li (2010). Cluster analysis is used to determine virtual partitions, and then fuzzy recognition is used to define the leak that is attributable to each specific virtual partition. Once hydraulic data from nominal and abnormal behaviour (for all possible leaks) is generated, the cluster analysis is performed over this data, by means of the system clustering method. Then, the process starts to merge classes depending on their inter-similarity, until there is only one class, hierarchically organizing the clusters, generating a standard model library. The fuzzy recognition process then calculates the closeness of an incoming sample and the ones of the library to determine to which class the sample belongs. A water network from China is used to assess the performance of the methodology. The presented leak localization result is correct, showing the feasibility of the method, despite it may be insufficient to demonstrate its full performance.

Years later, another fuzzy-based method was proposed in Wachla et al. (2015), which is composed of two stages. First, residue signals are generated by means of water flow models, and then neuro-fuzzy classifiers are used to locate the leak at predefined network areas. SVM was selected to implement these models, which were trained by means of flow data collected at the real network that was used as case study. These models learn the nominal behaviour of the network and aid to generate residuals from incoming leaky data. Regarding the training of the neuro-fuzzy classifiers, an EPANET model is utilized to derive hydraulic data for their training. Specifically, the neuro-fuzzy system ANFIS (Jang, 1993) is used for data classification. The previously-mentioned case study was used to assess the method and its capabilities. The results show that the method is able to locate the leaks in the correct WDN partitions in most cases, although some specific pipes are found to be sensitive to leaks at other pipes.

3.3.4. Deep learning approaches

Since its appearance, deep learning (DL) has been successfully applied to numerous pattern recognition problems in a wide range of fields. In the specific case of leak management, several recent works have considered this technique. In Zhou et al. (2019), a hydraulic simulator is necessary to emulate the behaviour of the network in the presence of leaks to obtain a training dataset for the DL scheme, which is based on a Fully-linear DenseNet (FL-DenseNet). Two case studies are used to show the reliability of the approach. First, the method is tested over the Anytown network (Walski et al., 1987), yielding excellent performance for both node-level and 5-top-candidates accuracy for low uncertainty values on the roughness and demands, while a decrease of performance is observed for higher values, although the results are still satisfactory. Then, a real-based network is used to test the performance of the approach. In this case, the method is compared to other three strategies, with the proposed FL-DenseNet yielding the best results.

In the same year, Javadiha et al. (2019) proposed a leak localization method based on DL for image classification. A hydraulic model is required to generate hydraulic information for every possible leak scenario, to then compute pressure residuals. However, only information at the sensors is considered, and the Kriging interpolation method is used to retrieve the complete vector of residuals, achieving a residual map of the WDN, which can be converted into 2-D images by considering a specific image resolution. Finally, a DL Convolutional Neural Network - CNN (Albawi et al., 2017) is trained with the images and the labels of the leaks, preparing it to be used for incoming leak scenarios. A Bayesian reasoning strategy is used to enhance the results by integrating consecutive time instants. The case study of Hanoi is considered to assess the performance, considering uncertainty in sensor measurements and nodal demands. The testing results showed that the performance improved when the number of sensors increases.

A novel leak localization scheme was presented in Hu et al. (2021), based two complementary stages. First, the network pipes are divided into groups by means of density-based spatial clustering of applications with noise - DBSCAN (Ester et al., 1996). The clusters are continuously expanded due to the continuous sampling, until final clustering results are obtained. Then, each zone is used as a label for multiscale fully convolutional networks (MFCN), which introduces the idea of multiscale decomposition based on Fourier transform (MDFT) to generate a new layer of a CNN. An example network is used to test the methodology, generating hydraulic data for all possible leaks in order to train the MFCN. The leak localization is compared to other three algorithms, achieving the proposed method the best results, i.e., minimum mean per-class error.

Recently, Romero et al. (2022) proposed a clustering-learning approach to solve the leak localization problem. Pressure vectors for all possible leaks is generated using a hydraulic simulator, and converted into 2-D images (of the size of the number of sensors) by Gramian Angular Field - GAF (Wang and Oates, 2015). Then, a recursive clustering-learning scheme is used to divide the network (or subnetwork if it is not the first iteration) into two clusters by means of Graph Agglomerative Clustering - GAC (Zhang et al., 2013), training a DL net to discern the leak location between the generated clusters. The process is repeated until the desired level of division, and the trained DL networks are hierarchically organized, yielding a classification tree. A case study of a real network is used to assess the method performance, feeding actual data to a trained classification tree, leading to a localization result that qualitatively and quantitatively outperforms a compared state-of-the-art model-based approach.

Furthermore, Li et al. (2022a) proposed a ResNet-based (He et al., 2016) leak localization scheme based on four stages: dataset generation, training, classification and regression. A hydraulic model generates hydraulic information for every possible leak scenario, providing data of the selected sensors as input to the ResNet network. Once trained, the classification process provides as output the probability of each pipe to be the origin of the leak, using the regression process to determine the exact location of the leak over the pipe with maximum probability. Two benchmarks are used to demonstrate the capabilities of the presented framework, namely the Anytown example and the Net3 network. The accuracy results showed a good performance in both cases,with the regression stage demonstrating reliability for the precise leak location over the specific pipes in Net3, although uncertainty was not considered for this benchmark.

3.3.5. *Other classifiers for mixed model-based/data-driven methods*

During the years, other techniques and algorithms have been exploited to solve leak management problems. In Tao et al. (2014), an artificial immune system - AIS (Galeano et al., 2005) network is used to identify the position of bursts in WDNs. This method, inspired by the immune system, is a stochastic optimization algorithm, similar to evolutionary algorithms. Data for the considered leak scenarios is obtained through EPANET to train and calibrate the AIS for pattern recognition, finally locating the leak by means of a nearest-neighbour approach. Three case studies of growing complexity are used to test the methodology. Excellent accuracy was achieved for an example network considering no uncertainty, as well as for a bigger network where the leak localization result worked at area-level. Finally, data from a real leak on a real-world network was analysed by the method, yielding satisfactory results although the localization failed for certain data sets.

The classical mixed model-based/data-driven scheme is used in Soldevila et al. (2016), using a hydraulic model to generate pressure residuals for the possible leak scenarios and conditions (magnitudes, noise...) in order to then supply them to a classifier, based on k-Nearest Neighbours - k-NN (Alpaydin, 2010). Nodes whose leak induce a similar WDN behaviour are grouped before training, reducing the number of labels. A temporal reasoning scheme is also included to reduce the effect of uncertainty and noise. Three case studies are considered to show the capabilities of the method. The proposed approach is compared with the angle method from Casillas et al. (2014) in the Hanoi example, showing the superiority of the former despite the performance degradation caused by noise. A larger network from Limassol is used to show the importance of temporal reasoning to achieve a sound accuracy result. Finally, the method is applied over a real-based network, training with artificial data and testing with measurements from two real leaks. The results showed promising performance although the leaks were not exactly located.

Subsequently, Soldevila et al. (2017) proposed a leak localization scheme based on the exploitation of Bayesian classifiers. They are fed with residuals from the difference between measured and estimated pressures, computed by means of a hydraulic simulator. With every sample, the probability of each leak can be estimated by applying the Bayes theorem. The prior probabilities are estimated to be equal for every leak at first, and the likelihood of a residual, considering that a certain leak is obtained through a calibration process. The method performance is tested over two case studies. First, it is compared to the angle method (Casillas et al., 2014) and k-NN (Soldevila et al., 2016) in Hanoi, considering several sources of uncertainty. The Bayesian approach outperformed the other methodologies. Then, the comparison with k-NN and the correlation method from Casillas et al. (2014) was performed over a real pilot from Barcelona considering a real leak. Again, the Bayesian approach yielded a better performance.

Sensitivity analysis and classifiers are combined in Romero-Tapia et al. (2018) to perform leak localization. Pressure data is obtained from simulations for the different leak scenarios, obtaining pressure residuals and then the sensitivity matrices. From these matrices, the within-class-scatter and the between-class-scatter matrices are computed. This works uses a Fisher Discriminant Analysis - FDA (Mai, 2013) classifier, whose aim is to maximize the scatter between classes and minimize the scatter within classes. To this end, perpendicular FDA vectors achieving these goals are iteratively computed to derive the eigenvectors of a discriminant matrix, which allows to generate a discriminant function for each class. This function can be evaluated to locate the leak of incoming data. The FDA method is compared to the angle strategy in Casillas et al. (2014) over the Hanoi example, leading to a smaller localization error for two and three sensors (no uncertainty is considered).

Some years later, Lučin et al. (2021) presented another localization scheme based on classifiers. As usual, a hydraulic simulator was used to generate leak scenarios, randomly selecting the locations, magnitude and demands.

The classifier considered in this work is the random forest algorithm (Breiman, 2001). It consists of multiple decision trees where each tree is trained independently on a random subset of the dataset. During the application stage, the classification with more occurrences is chosen by the random forest and considered as the class prediction. The Hanoi and Net3 benchmarks are used to evaluate the methodology, analysing several factors that may affect the performance. The results showed that the demand variation was the responsible of the larger errors, greatly reducing the localization accuracy. Overall, the method works well for small and medium sized networks, but starts to fail for large-scale ones.

In (Soldevila et al., 2021) a novel integrated solution is presented to address both leak detection and localization. Detection is handled by a sequential monitoring algorithm that analyzes the inlet flow, to then validate each detection through an ad hoc statistical test. Localization is tackled by means of a classification problem, training the classifiers through leak data obtained by means of a hydraulic model. A customized clustering scheme links areas of the WDN where accurate localization is not possible due to the lack of sensors. The integrated solution is evaluated over synthetic data from the Limassol DMA and real data from a DMA in Barcelona, by means of a comparison to other state-of-the-art methods. The proposed approach outperforms the others for the synthetic case in detection and localization using flow sensors, but the localization results are poor for pressure sensors. About the real case, the detection by the proposed method is in a middle tier among the compared ones, but it outperforms them when considering localization.

3.4. Data-driven detection and localization methods

The development of purely data-driven methods started recently, in comparison to the previous two families of methods. This was caused by the boom in the usage of calibrated models. Thus, more and more data-driven schemes are derived, effectively reducing the dependence on the hydraulic model that not always available or well-calibrated.

3.4.1. Statistical analysis approaches

During the years, many leak detection techniques have made the most of statistical-based techniques to analyse the network behaviour in abnormal scenarios.

In Buchberger and Nadimpalli (2004), the historic data of the flow measurements at the DMA inlet is analyzed; the largest value of the hourly mean and standard deviation is searched. Then, it is checked if the new measurements exceed those two indicators to detect a leak. The leak size is estimated when the leak is detected by computing the difference in both means. Several examples based on simulated and measured flows were used to test the method, considering constant and variable leaks.

Years later, a detection method based on Kernel PCA (KPCA) was presented in Nowicki and Grochowski (2011), which extends the linear PCA method and extends it to the non-linear case. The main idea behind the method is to construct the KPCA model by means of data from nominal operation of the network, and then use the model to determine the state of the network by means of real-time measurements. The method performance was assessed with simulations from the Chojnice town case study. The experiments showed a different quality of solution depending on factors like the relative position of sensors and leaks, their size and time of appearance.

The next year, Palau et al. (2012) proposed a PCA-based method to detect leaks in WDNs. The approach uses flow measurements, which are mean centered and scaled, to train a nominal PCA model that will be then used to detect outliers by means of statistical metrics like T^2-Hotelling and Distance to Model (DMOD) (Eriksson et al., 2001). The model training is iterative, as outliers are extracted from the input data when the matrix updates the PCA model. Data from a real Spanish WDN was used to evaluate the suitability of the methodology. The results showed that DMOD is the best out of the two metrics, although its results were satisfactory at most.

An abnormal-event detection technique was presented in Romano et al. (2014). An ANN is created using Normal Operating Patterns - NOP data for training and testing. From the discrepancies between the training and testing data, the mean and standard deviation are computed to evaluate statistical tests. The outputs of three test indicators feed two inference systems based on Bayesian Networks (BNs). One assesses the probability of an event occurrence by means of these outputs and the daily average difference between the NOP values and the current ones (which can also be used to estimate the leak size). The other BN is used for event detection by taking into account the outputs of the subsystems of the DMA, and also the signals from others DMAs in the same WDN. Finally, the BN result is compared with a threshold to decide whether or not to raise an alarm. The method was tested using real data from pressure/flow sensors deployed in a UK network, using both synthetic and real leak events. The method reported satisfactory results, despite long delays occurred at some engineered tests, and several false alarms were raised during the real events testing.

In Ye and Fenner (2014), polynomial models were proposed to predict the total weekly water consumption for each measurement using the latest week data. The parameter estimation for these models uses an EM algorithm and weighed least squares. The residual obtained using the real measurement is compared with a threshold calculated using the standard deviation of the leak-free case. Once the leak is detected, the difference between prediction and measurement is computed to estimate the leak size. The method was evaluated using data from several actual water networks, leading to a high rate of true positives and a reduced ratio of false positives, although the uncertainty in the demand dynamics may hinder the operation of the method.

In Jung and Lansey (2015), a Non-linear Kalman Filter (NKF) was used in combination with pressure and flow measurements inside the network to estimate the nodal demands, which are compared with historic data. The difference is analyzed using Statistical Process Control (SPC) techniques like CUSUM and Hotelling T^2 (Aparisi, 1996) technique and a threshold for leak detection purposes. The study uses the Austin benchmark with synthetic simulated scenarios, considering flow and pressure measurements, as well as several simplifications like perfectly known pipe roughness coefficients and other parameters or a perfect hydraulic model representation of the reality for the benchmark. The detection results were acceptable, with high detectability and reduced detection times.

A similar approach was presented in Anjana et al. (2015), where a Particle Filter - PF (Djuric et al., 2003) is used to model the dynamics of the network and retrieve the WDN state, given by the flows and pressures, and then the SPC-CUSUM method is applied to detect abnormal consumption. The strategy was tested in a reduced version of a real-world network in Mandya, India. The method was able to detect the simulated leaks, but the pressure estimation by the PF was oscillating and hence not satisfactory, as well as the case study size was quite small and the flow sensors density was high.

In Romano et al. (2017), three SPC control charts were used to compare, on a daily basis, descriptive statistics gathered from pressure meters for the analysed night with the corresponding values computed for the previous nights, in order to find unexpected variations that can be caused by a leak. The outputs of the three tests are then unified in one indicator used to rank the sensors from the most affected to the least affected. The area surrounding the most affected sensor is the leak area candidate. The method was tested using data from a set of real DMAs, selecting both pressure-manasged and gravity-fed ones, through historical datasets of real-life burst events. The tests showed a satisfactory performance for pressure-managed DMAs, but the performance for gravity-fed ones was poor. Additionally, note that the sensorization density was high and large leak sizes were considered.

In Quiñones et al. (2018), an unsupervised leak detection and localization approach was presented, based on the combination of a pre-processing operation and PCA. The pre-processing operation is used to avoid the difficulty of segmenting the demand temporally and the lack of consideration of dynamic network characteristics. This process is based on determining the periodic expected value and standard deviation of each measured variable. Then, PCA can be used to estimate the contribution of each variable in the localization of the potential leak zone, deriving a reconstruction-based contribution index. The approach was evaluated over the Hanoi case study, performinf localization over pre-defined network areas. The achieved results are satisfactory and superior to the comparison methods, which are the Bayesian and k-NN methods from Soldevila et al. (2017, 2016) respectively.

A correlation-based methodology was proposed in Gomes et al. (2021) for leak detection purposes. The method tries to describe the network dynamics through the spatiotemporal correlation of pressure and flow measurements, analysing outliers appearing in the expected correlation due to leaks. The well-known Pearson Cross-Correlation coefficient - PCC and the Detrended Cross-Correlation Analysis - DCCA (Podobnik and Stanley, 2008) are considered to perform the correlation-based feature space construction. The features of the identified leaks are used to provide assist to leak localization tasks. The method is evaluated over the WDN of Infraquinta, Portugal, considering both simulated scenarios and real data from the network. As expected, the detection was performing better for the synthetic scenarios, and DCCA generally performed better in both types of events (synthetic and real).

3.4.2. Learning techniques

Most of learning-based methods fit in the mixed model-based/data-driven category due to the necessity of labelled data from a hydraulic model, covering the complete set of possible scenarios. However, additional application methods or learning philosophies helped to derive purely data-driven detection approaches, e.g., online learning (the samples are gathered from historical datasets or the method learns while an additional leak detection approach operates), unsupervised learning...

Burst detection using pressure and flow measurements data from DMAs with artificial neural networks was proposed in Mounce and Machell (2006). Results were presented using data from an experimental setup in a real WDN from UK, where leaks were forced by means of hydrant flushing. Despite the satisfactory performance, the study revealed the difficulty of detecting leaks when other abnormal events occur at the network, e.g., unusual demands or system changes. The difficulty of obtaining labelled samples for training is also discussed.

This method was improved in Mounce et al. (2008) by combining the ANN scheme with a Fuzzy Inference System (FIS), as well as reducing the required data to only flow measurements. The ANN model is trained by means of a continuously updated historical datasets that construct a probability density model, which the FIS compares with observed flows to link confidence intervals to alerts, helping to even supply a precise estimation of the leak size. The method was tested in a real WDN from UK, running together with an existing leak detection method. Despite the performance was satisfactory, improving the overall detection solution, several false alarms were raised at events that were not correlated to bursts repairs or reports.

A leak detection scheme based on an unsupervised learning technique was presented in Aksela et al. (2009). Specifically, Self Organizing Maps or SOMs (Kohonen, 1990) were exploited. This kind of techniques have the advantage of not needing information about all the possible leaks, as the SOMs are trained so that the vectors of the model represent training flows so that they are independent on the detected leaks, i.e., similar flows would be mapped into the same or close model vectors. The method is evaluated through a realistic case study, using actual flow data from the studied WDN. The performance on the tested events was excellent, highly reducing the appearance of false positives, although more events would be required to fully assess the method's performance.

A typically exploited technique in mixed model-based/data-driven methods like SVM was used in Mounce et al. (2011) in an online training fashion, complemented with the necessity of historical datasets. Specifically, Support Vectors Regression (SVR) is considered, and the method is tested over a real WDN in UK, working in parallel to the method explained in Mounce et al. (2008). Although the SVM-based technique seemed to perform faster than the AI/FIS one, several non-leak abnormal events (like unusual demands) raised the alarm of the SVM method, increasing the ratio of false notifications.

Years later, Laucelli et al. (2016) proposed an Evolutionary Polynomial Regression (EPR) paradigm to model the behaviour of WDNs for leak detection purposes. EPR is based on a multi-objective optimization problem, solved using GAs, that aims to maximize the fitness of prediction to data and minimize the number of required explanatory variables and polynomial terms. It is trained using pressure and flow data from datasets covering a week, to then test the obtained model with another set of datasets. The achieved nominal EPR model would be used for comparison with the entering hydraulic measurements, raising alarms when thresholds are surpassed. A case study with available real data was proposed to test the method. In average, the results were acceptable, although smaller leaks became undetectable when another large leak was occurring.

A CNN-based method was proposed in Fang et al. (2019), with the aim to solve the multi-leak detection problem. The CNN-scheme is able to extract features from a historic leak dataset and then predict if there exist a new leak in the network by considering the learned features. An experimental setup was constructed to test the method, leading to high true positive rates even for three leaks. However, several simplifications are considered, like the simplicity of the experimental setup (only 21 nodes), a high density of sensors (all the nodes in one case and 8 in the other)... Moreover, the data gathering stage is possible due to the size of the network, but in larger and real-based ones, the methodology may be need to be conceived as mixed model-based/data-driven, using a model to generate the samples/labels.

In Wang et al. (2020), a burst detection algorithm based on deep learning was presented. Its operation starts by predicting the inlet flow through a Recurrent Neural Network - RNN (Medsker and Jain, 1999) trained to learn the nominal behaviour of the network. Residuals are computed then from these predictions and the observed values, using a dynamic multithreshold strategy to identify outliers. A final phase of outlier feedback correction is applied to transform the leak-influenced flow into its corresponding leak-free flow, in order to be able to detect new outliers. The methodology was tested with both simulated scenarios and engineered experiments over a real DMA from China, yielding high true positive rates and low false positives ones, although the burst sizes were large for several scenarios.

3.4.3. Interpolation & graph-based techniques

Several fully data-driven methods are based on geostatistical techniques, which focus on the relationship between the values of variables at certain geographical locations and those at othe locations at some distance. Therefore, considering the availability of measured values in the network, the aim of these methods consists of interpolating the values at locations that are not measured.

An early work dealing with this kind of techniques is Romano et al. (2013). Detection is handled by multivariate Gaussian mixtures-based graphical models, which learn to distinguish between two classes (alarm on, alarm off).Leak localization is performed by four different geostatistical techniques, which interpolate the probability values of a burst event at every node of the network.These techniques are Inverse Distance Weighted interpolation - IDW (Shepard, 1968), Local Polynomial interpolation - LP (Cleveland and Devlin, 1988), Ordinary Kriging - OK (Krige, 1951) and Ordinary Co-Kriging (OC) (Myers, 1982). On the one hand, IDW and LP are deterministic methods. The former estimates the value of a variable at unmeasured locations by using a linear combination of the measured values surrounding these unmeasured locations, while the latter assumes that each measured value is the sum of a polynomial function of the coordinates and a random error. On the contrary, OK and OC are stochastic methods that quantify the spatial correlation structure of measured variable's values to estimate the unmeasured values, wit OC being an update of OK that can work and benefit over several variables. A real network was employed to generate real fault data by the opening of hydrants. The detection performance was quite successful and fast, with no false alarms. The localization of the four geostatistical methods was compared, with OC being the best one.

Years later, in Rajeswaran et al. (2017), a graph partitioning algorithm is proposed to locate leaks in WDNs. The method uses a repeated water balance strategy and minimal additional flow measurements. The water balance considers an envelope encompassing a set of nodes and edges, so that in steady state and under leak-free conditions, the balance must be satisfied around the envelope. The additional flow measuring sites are obtained by the multi-stage graph partitioning algorithm, posing the problem as a multi-objective mixed integer linear program (MILP) trying to minimize the size disparity of the clusters and the cut-cost.To solve the multi-objective minimization problem, a lexicographic solution and goal programming are considered. The methodology was evaluated over five case study networks, obtaining results that show that a small fraction of pipes needs to be queried for measurements in order to correctly locate the leaks.

Later, Soldevila et al. (2020) proposed a leak localization method based on Kriging interpolation. First, a data-driven adjusted model is generated to estimate the leak-free expected pressure according to the operating conditions of the network. Kriging is then used to estimate the pressure values at the network nodes that are not equipped with sensors. The leaky node is selected as the one whose residual value is the maximum. To improve this approach, the Dempster-Shafer - DS theory (Sentz and Ferson, 2002) for reasoning under uncertainty is used to analyze the differences between the leak and leak-free scenarios, fusing information from various sources, as well as from a single source at different time instants. The method is tested using data from two real-world networks from Madrid, and compared with two approaches, i.e., the same approach but using Bayesian reasoning instead of DS, and the angle method of Casillas et al. (2014). For both DMAs, the DS approach clearly outperformed the other two methods, yielding reduced results of linear and topological distance from the candidate to the leak.

In Alves et al. (2021), a leak localization method based on hydraulic measurements and the network topology was presented. A reduced-order model structure estimates leak-free pressures at sensorized inner nodes, generating residuals from the comparison with leak pressure values. Leaks can be located in area-level by determining the most affected sensor in terms of residual value. To perform node-level localization, the relative incidence of the leak in a node is determined by means of the network underlying graph and the correlation between the probable leaky nodes and the residuals. The shortest path from the inlets to a node is regarded as the most probable path for extra flows induced by a leak at this node, so the effect in a sensor of a certain leak would depend on the path intersection from the inlets to the node and the sensor respectively. Thus, each node-sensor pair yields an incidence coefficient, which combined with the residual information, allows to derive a likelihood index to determine the most probable leak location.The performance of the method is compared to that of a Kriging-based one in Hanoi, with the first outperforming the second. Additionally, the Modena benchmark is considered, showing satisfactory results.

Recently, Romero-Ben et al. (2022) presented several approaches to solve the BattLeDIM2020 problem. Detection was handled through a leak estimation algorithm, based on applying maximum likelihood estimation to the joint probability distribution of a set of leak estimates over a time window, raising an alarm if a limit is surpassed. The localization is handled by two methods, due to the different features of the network areas. On the one hand, localization in Areas B and C is operated through a model-based method, based on the comparison between actual pressure residuals and simulated ones, updating the model when a leak is found/fixed. On the other hand, the localization problem at Area A is solved by a fully data-driven approach (the article is located in this category due to the novelty of this method), based on two stages. First, the complete set of WDN hydraulic head values is interpolated by means of measured ones and the network topology through a quadratic programming problem posing the state of a node as a weighted linear function of the states of neighbouring nodes.Additionally, the preservation of the directionality of the flows in the network is pursued. The second stage consists of a geometry-based comparison between leak and leak-free state vectors, allowing to automatically select a set of nodes as leak candidates. As previously mentioned, the method was evaluated in the BattLeDIM2020 competition, achieving satisfactory results for most of the detected leaks (it was applied to undetected leaks too, showing that it was capable of finding those leaks).

3.4.4. Other data-driven techniques

During the years, several other methods have been derived and tested, which may not fit the categories presented above. In Ye and Fenner (2011), a Kalman Filter (KF) was used to model the nominal behaviour of the network and estimate the normal flow or pressure in the system, generating residuals when comparing with flow or pressure measurements. If the residual exceeds a threshold, then a leak is detected. The methodology was tested by means of data obtained at both caused leaks at engineered tests and natural leaks that were considered from records of the water utlity (from users complains and/or repair works). This method was able to detect new leaks even if there were already other leaks in the network. Also, this work concluded that measuring flow is better variable to detect leaks as compared to the measuring pressure.

Historical flow measurements at the inlet were used in Eliades and Polycarpou (2012) to perform leak detection. The methodology adapts to the unknown and time-varying inflow dynamics through an approximation technique based on the update of the coefficients of a Fourier series. Then, the bias term of the generated series is considered, analysing changes in its value through the CUSUM technique. Finally, the features are compared with a threshold for the leak detection purposes. The leak size estimation is done by comparing the actual flow consumption and the Fourier series prediction. The methodology is compared to a night-flow analysis method, applying both to historical data from a DMA in Limassol, Cyprus. The comparison showed the superiority of the new method for detecting small leaks, leading to less false negatives, although additional work was necessary to reduce the false positives and the detection time.

In van Thienen (2013), a method to compare the current inlet flow measurements with the historical ones, called Comparison of Flow Pattern Distribution (CFPD), is used to create a plot with the aim of facilitating the decision for an expert about the detection of a leak and their size estimation. Although it is not an automated method, it is not computationally expensive, easily implemented and independent on a hydraulic model. The method is tested in datasets provided by actual water companies, analysing several cases of leakage, unusual demand events, as well as theoretical experiments. The results show the usefulness of the methodology, despite the necessity of the decision from a human operator reduces the impact of its advantages.

A heuristic burst detection method was proposed in Bakker et al. (2014). Using flow and pressure measurements as input information, the approach continuously compares observed and predicted values for nodal demands and pressures. The prediction for the demands is obtained through an adaptive water demand forecasting model, whereas the expected pressures are achieved by means of a dynamic pressure drop - demand relation estimator. The former method learns the nominal demand patterns, considering week characteristics and other possible sources of outliers related to special dates. The pressure estimator analyses the relation between the pressure at the inlet and the pressures at other locations. Leak detection is achieved through the deviation with respect to a threshold. To evaluate the method, data from three supply areas at Netherlands were collected from a historic dataset. The method performed adequately for one of the areas, but poorly of the other two, as well as it yielded a low false positive ratio.

In Narayanan et al. (2014), the current flow betweeness centrality measure concept from electric networks is adapted to water networks and used to estimate burst flows. The adaptations are related to the non-linear relation between flows and pressures in WDNs. The leak candidates are obtained by computing the difference between the excess of leak flow from actual measurement at the water inlet and the one obtained as leak signature from the current centrality measure. Note that detection is achieved by calibrating ARIMA models of the inlet flow, and checking changes in its behaviour (the leak size can be estimated too). This methodology was tested using a network from the Battle of Water Calibration Networks (BWCN). The virtual tests, with data for all the possible leaks, showed that for most leak scenarios the correct node was included in the high probability set. Besides, data from actual leaks from a real-world network was considered, and the method was able to include the correct node in the high probability set for 2 out of 3 leaks. Thus, the method showed satisfactory results considering that it only uses information from the inlet of the network.

In Hutton and Kapelan (2015), a polynomial model is calibrated with past measurements of the network water consumption and is used to predict the actual water consumption. Also, the mismatches from the past data (residuals) without leaks are used to create two probabilistic models (one Gaussian, the other heavy tailed, heteroscedastic quantification) for checking (by means of thresholds) if the actual residuals fall into these two models without leaks. The case study is a DMA from a UK WDN. The results showed that both probabilistic models perform better at night times, increasing the rate of true positives and diminishing the false positive ratio, and for bursts greater to a 5% of the average daily inflow.

A clustering approach is used in Wu et al. (2016) to detect outliers in a set of transformed data matrices, obtained by arranging flow measurement vectors by flow-meter and day (each matrix correspond to a 5-minutal window of the day). The clustering is based on the idea of detecting low density points that are separated from high density points, by means of vector distance computations. Then, the leak events are extracted from the complete set of outliers by considering the *x-mean* of the temporal flow information at the outlier data vectors, so that only "large" outliers are regarded as leaks. A DMA of China is used as case study, using real data from the network, gathered at engineered experiments. The method detected all leaks in short time, although the leak sized were quite high, considering that the minimum one was around a 13.3% of the inflow.

3.5. Discussion

An extensive set of articles dealing with leak detection and localization has been presented in the previous subsections. From this information, a discussion about leak management strategies is carried out with the aim of establishing solid conclusions about this field. Table 1 summarizes the characteristics of these approaches with a selection of representative examples of each category of methods, in order to make it easier to compare them. They are chronologically ordered so that the evolution of the trends in each of those categories can be reviewed. The table has been organized to focus on issues that may be relevant for water utilities. Specifically:

- Problem: this entry indicates the leak management problem or problems that are faced, distinguishing between localization (L), detection (D), and the methods solving both problems. The latter can be divided into approaches that solve both problems using the same technique (D+L) and methods that exploit a different technique for each problem (D-L). Additionally, the consideration of the leak size estimation (SE) problem is also indicated.
- M-Leak?: several works in the literature consider the assumption of single leaks occurring in the network, i.e., there are not simultaneous leaks. However, this feature is not assured in real-world networks, where leaks can appear while other bursts are already occurring. Thus, we highlight those methods that consider the multi-leak problem in this table entry.
- Sensorization properties: this category encompasses two of the main sensor-related characteristics of interest for water companies when analysing leak management methods:
  - Type: the physical magnitudes measured by the deployed are stated here for each method, considering the information provided about sensorization requirements in their methodology explanation, as well as the actual usage at the presented case studies. The considered sensor types are demand (D), flow (F) and pressure (P). Note that in real networks, the inlet flow and pressure are almost always measured for DMA control purposes, so they are not explicitly indicated in the table, except for detection-only methods which only use them as data sources).
  - Density: the sensorization density is of great importance, as it needs to be considered along the method performance to understand if realistic conditions were applied in its evaluation. This density is indicated as x/X, with x denoting the number of sensors and X indicating the number of network elements (nodes for P and D; pipes for F). The notation $x_1/X_1;x_2/X_2;...$ is used if there are multiple case studies with different sensor configurations. Moreover, an "A" symbol is used to indicate works that analyse the effect of different sensor configurations.
- Validated in: this entry analyses the properties of the case studies that have been used to assess the performance of each methodology:
  - Case: indicates the type of case study that was used to test the method (if several types are used, only the most realistic one is indicated). We distinguish between synthethic networks (SN), which are mostly academic examples of reduced complexity; real-based networks with synthetic data (RN(SD)), which are more complex networks or based on real-life WDNs, where artificial data are generated through a simulator; and real networks with real data (RN(RD)), when the method was applied to actual data gathered at a real-world WDN (if both synthetic and real data is applied, it is marked as RN(SD+RD)).
  - U: in the case of the usage of artificial data (SN or RN(SD)), the inclusion of uncertainty is crucial to perform realistic tests that are really useful to water utilities. Here we indicate the uncertainty sources that

are considered in each article dealing with synthetic data (note that all of them are already present in real data): model (M), sensor (S), demand (D) and leak characteristics, including variable/varios sizes (V).

– Benchmark?: finally, if the article includes a benchmark that allows to reproduce the results, it is indicated in this entry.

Table 1: Summarizing table of model-based methods

Article	Problem	M-Leak?	Sensorization properties		Validated in		
			Type	Density	Case	U	Benchmark?
Model-based							
Pudar and Liggett (1992)	L (SE)	✓	P	A	SN	—	*Pudar*
Andersen and Powell (2000)	L	✗	(P,D)	(4/36,36/36)	SN	—	—
Poulakis et al. (2003)	L(SE)	✓	(P,F)	A (7/31,7/50)	SN	M, S, D	—
Misiunas et al. (2006)	D-L(SE)	✗	P	3/79	SN	D, V	—
Wu et al. (2010)	L(SE)	✓	P	28/841	RN(RD)	M	*Pudar*
Islam et al. (2011)	D+L(SE)	✗	(P,F)	(5/27, ?/40)	SN	M, D	—
Casillas et al. (2014)	D-L(SE)	✗	P	A (6&15/3377)	RN(SD)	S, D, V	*Hanoi; Quebra*
Perez et al. (2014)	L	✗	P	6/3377	RN(RD)	—	—
Sanz et al. (2016)	D-L	✗	P	5/3377	RN(SD)	S, D, V	—
Berglund et al. (2017)	L(SE)	✓	P	45/26986	RN(SD)	—	*Hanoi*; *Net3*
Sophocleous et al. (2019)	D+L(SE)	✓	P	8/202; 10/1000	RN(SD+RD)	S, V	—
Vrachimis et al. (2021)	D-L	✗	P	A (4/31)	SN	M, S, D, V	*Hanoi (LeakDB)*
Li et al. (2022b)	L	✗	P	20/491	RN(SD)	M, S, V	—
Steffelbauer et al. (2022)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M, S, D, V	*BattleDIM2020*
Daniel et al. (2022)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M, S, D, V	*BattleDIM2020*
Li et al. (2022c)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M, S, D, V	*BattleDIM2020*
Marzola et al. (2022)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M, S, D, V	*BattleDIM2020*
Wang et al. (2022)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M, S, D, V	*BattleDIM2020*
Mixed model-based/data-driven							
Caputo and Pelagagge (2003)	L (SE)	✗	(P,F)	(28/29, 1/29)	RN(SD)	V	—
Mashford et al. (2012)	L(SE)	✗	P	6/73	RN(SD)	V	—
Candelieri et al. (2014)	L	✗	(P,F)	A (7-16/1212,3-6/1385)	RN(SD)	V	—
Tao et al. (2014)	L	✗	P	26/600	RN(RD)	D,V	—
Wachla et al. (2015)	L	✗	F	6/+2600	RN(SD+RD)	—	—
Soldevila et al. (2016)	L	✗	P	5/1520	RN(RD)	S,D,V	*Hanoi*
Zhang et al. (2016)	L	✓	P	30/2189	RN(RD)	S,V	—
Soldevila et al. (2017)	L	✗	P	5/1520	RN(RD)	S,D,V	*Hanoi*
Sun et al. (2019)	L	✗	P	A	SN	—	*Hanoi*
Zhou et al. (2019)	L	✗	P	4/∼50	RN(SD)	M,D,V	*Anytown*
Shekofteh et al. (2020)	L(SE)	✓	P	A (6&19/443)	RN(SD)	S, D	*Balerma*
Cantos et al. (2020)	D-L	✗	(F,D)	(93/398, 80/412)	RN(SD)	V	—
Capelo et al. (2021)	L(SE)	✗	P	A (7&14&21/4448)	RN(SD)	V	—
Lučin et al. (2021)	L	✗	P	2/31; 4/92	SN	D,V	*Hanoi; Net3*
Hu et al. (2021)	L	✗	P	14/49	SN	V	—
Soldevila et al. (2021)	D-L(SE)	✗	(P,F)	(5/1520,2/1664)	RN(RD)	S,D,V	*Hanoi*
Romero et al. (2022)	L	✗	P	8/120	RN(SD)	M,S,D	—
Li et al. (2022a)	L	✗	P	A (2-4/19); 4/92	SN	S,V	*Anytown; Net3*
Data-driven							
Buchberger and Nadimpalli (2004)	D(SE)	✗	F	1/+21	RN(SD)	V	—
Aksela et al. (2009)	D	✗	F	3/+660	RN(RD)	—	—
Ye and Fenner (2011)	D(SE)	✓	(P,F)	(?/925,1/+925)	RN(RD)	—	—
Mounce et al. (2011)	D(SE)	✓	(P,F)	(5/?,4/?)	RN(RD)	—	—
Eliades and Polycarpou (2012)	D(SE)	✗	F	(1/236)	RN(SD)	M,S,D,V	—
Palau et al. (2012)	D	✗	F	1/+2900	RN(SD+RD)	—	—
Romano et al. (2013)	D-L	✗	P	13/925	RN(RD)	—	—
Ye and Fenner (2014)	D(SE)	✓	F	1/∼1000	RN(RD)	—	—
Romano et al. (2014)	D(SE)	✗	(P,F)	A (2/2940,1/+2940)	RN(RD)	—	—
Wu et al. (2016)	D	✗	F	5/∼60	RN(RD)	—	—
Laucelli et al. (2016)	D(SE)	✓	(P,F)	(3/2940,2/2940)	RN(RD)	—	—

Rajeswaran et al. (2017)	L	✓	(F,D)	A	RN(SD)		*EXNET, DTown...*
Quiñones et al. (2018)	D-L(SE)	✗	P	3/31	SN	D,V	*Hanoi*
Soldevila et al. (2020)	L	✗	P	10/169; 10/1031	RN(RD)	—	—
Wang et al. (2020)	D	✓	F	2/∼60	RN(SD+RD)	M,S,D,V	—
Alves et al. (2021)	L	✗	P	5&10/268	RN(SD)	S,D,V	*Hanoi; Modena*
Gomes et al. (2021)	D	✗	(P,F)	(21&6/4448,7/4494)	RN(SD+RD)	V	—
Romero-Ben et al. (2022)	D-L(SE)	✓	(P,D)	(33,82)/785	RN(SD)	M,S,D,V	*BattLeDIM2020*

3.5.1. Considered leak management problem

Regarding the families of methods, we can conclude that methodologies dealing with both detection and localization are mainly model-based, whereas mixed model-based/data-driven methods are mostly focused on leak localization tasks, and data-driven approaches normally examine one of the two problems.

This situation is explained by the availability and dependence on the hydraulic model. Model-based schemes fully exploit the advantages of the network model, having access to an incredibly useful source of information about the network dynamics. This facilitates the integration of both detection and localization into a common method (Islam et al., 2011; Sophocleous et al., 2019), as well as the design of independent detection and localization schemes that profit from the model-generated information (Sanz et al., 2016; Steffelbauer et al., 2022).

On the contrary, mixed model-based/data-driven methods focus on the localization problem because they try to reduce the dependence on the hydraulic model, which is mainly used to simulate all the leak scenarios to train a set of classifiers (Zhou et al., 2019; Romero et al., 2022). These classifiers need to separate as much as possible the different classes, with each class corresponding to a different leak. Therefore, the inclusion of detection in this kind of approaches would imply the generation of a new class that labels leak-free data. This could cause problems in the separation of low leak rates.

Data-driven approaches deal with the detection and localization problems separately because they normally use different sensors for each task. Moreover, it can be highlighted that pure leak detection methods, i.e., without considering localization; are mainly considered from a data-driven perspective (Aksela et al., 2009; Wu et al., 2016). This is justified by the different natures of leak detection and localization. The former implies the assessment of the behaviour of the network in order to distinguish between leak and leak-free states, so this task is perfectly feasible when only leak-free data is available, which is the case in most real-world networks. On the contrary, data from every possible fault has been historically needed by leak localization methods, because this tasks implies the differentiation among leaks. Thus, the relative importance of the hydraulic model is higher for leak localization methods than for detection strategies.

Finally, note that leak size estimation is usually linked to methods solving the detection problem (Mounce et al., 2011; Casillas et al., 2014). This is caused by the extended use of inflow information to perform leak detection, which makes it easier to estimate the leak by comparing the inlet flow at leak and leak-free conditions.

3.5.2. Multi-leak problem

The solution to the leak detection and/or localization of multiple simultaneous leaks in WDNs is not trivial. The items in Table 1 show that a majority of methods consider the assumption of single leaks, which is not realistic from the point of view of the actual operation of water utilities.

Note that the proportion of single-leak methodologies in the mixed model-based/data-driven case is higher. This is caused by the nature of these methods, which are usually learning-based. Considering single leaks, a label must be included for every possible leak, which would imply a maximum number of labels that is equal to the number of possible leak locations (nodes or pipes, depending on the article). However, the consideration of multiple leaks would imply a number of labels corresponding to all possible combinations of multiple-leak locations.

This problem may affect model-based strategies too, although their exploitation of the network model allows them to study how the effects of leaks are added, searching for the best combination of leaks to explain the network behaviour, leading to feasible multi-leak solutions (Berglund et al., 2017; Sophocleous et al., 2019). Data-driven strategies may be affected by this issue, although normally they can be easily extended to solve multi-leak cases, considering that they do not depend on the possible leak scenarios for their operation (Rajeswaran et al., 2017; Wang et al., 2020).

Some relevant examples of multi-leak methods are those taking part in the BattLeDIM2020 competition (Vrachimis et al., 2022), which were required to solve the multi-leak problem due to the characteristics of the proposed benchmark and the provided datasets, containing several overlapping leaks of different nature and magnitude (Daniel et al., 2022; Steffelbauer et al., 2022; Romero et al., 2022; Marzola et al., 2022; Wang et al., 2022; Li et al., 2022c).

3.5.3. Sensorization properties

The type of sensors required for implementing a certain methodology can be of great importance for water utilities, as they have normally sensor nets deployed over the WDN for control and monitoring purposes, and hence they may be interested on exploiting techniques that fit their actual sensorization characteristics.

A clear division between detection and localization schemes can be appreciated in Table 1. Most model-based and mixed model-based/data-driven methods require pressure information from a set of sensors scattered through the inner nodes of the network. This is justified by the consideration of the localization problem in all the presented works of these categories (Perez et al., 2014; Soldevila et al., 2017), which is normally tackled by means of this kind of sensors due to their lower cost in terms of price and installation with regards to the flow sensors. Note that leak localization methods normally need multiple sensors to be distributed over the network, to be able to capture the leak effect (pressure drop) independently on its location. This can be appreciated in the density of these pressure sensor networks. Let us highlight several points regarding this property:

- Methods dealing with real-world networks normally present low sensorization densities (considering the large scale of those systems in terms of junctions and pipes), so their performance, which is provided through the text, must be analysed considering this realistic property.
- Additionally, it is interesting to appreciate how early works of each category considered small networks, leading to high sensorization densities although the number of sensors may be small.
- Several methodologies include analysis about the number of sensors, which can be useful to demonstrate the necessity of a higher degree of sensorization in actual water networks to effectively solve the leak detection and localization problems.

Flow sensors are typically used in detection schemes, mostly installed in the DMA inlets in order to detect changes on the overall consumption (Buchberger and Nadimpalli, 2004; Palau et al., 2012). Until recently, online demand data using AMR has not usuallly been considered in the literature, as their installation on real WDNs is not currently extended. To remark the usage of this kind of meters in the BattLeDIM2020 articles, as the facilitated benchmark included them in one area of the WDN. It is interesting to remark that the BattleDIM2020 case study did involve a section with AMRs, so that the methods in this contest also use this information.

3.6. Methodology validation

Finally, it is interesting to study how the different methods are evaluated, considering that water companies would have great interest on approaches which have been tested on realistic scenarios.

The performance assessment of model-based approaches has evolved during the years, as indicated by Table 1. Early methods were tested over synthetic and example networks, leading to solutions that behaved satisfactorily from an academic point of view, but which were not checked to be usable for real scenarios (Pudar and Liggett, 1992; Andersen and Powell, 2000). The trend started changing and real-based case studies were considered, with a majority of works using synthetic data generated from a hydraulic simulator. The inclusion of different sources of uncertainty was also considered, with a special focus on measurements noise and demand uncertainty, replicating more realistic conditions on the data (Sanz et al., 2016; Berglund et al., 2017). Additionally, analysis about the effect of different leak sized and profiles were conducted in most works, leading to a richer study of the limits of the methods (Casillas et al., 2014; Vrachimis et al., 2021).

Mixed model-based/data-driven methods have been mostly tested on real-based networks, although several evaluations over real-world data can be found (Zhang et al., 2016; Soldevila et al., 2017). The focus on synthetic data over real-based WDNs is explained by the common usage of hydraulic models to generate training data, so that the additional generation of testing data was easy (Capelo et al., 2021; Romero et al., 2022). A higher proportion of methods tested with real-world data is found within the data-driven category. This is caused by the abundance of data-driven leak detection strategies (Ye and Fenner, 2011; Wu et al., 2016), which mainly require data from the DMA inflows. This data is normally already gathered by water utilities for control purposes, so the availability of historical datasets is more common. Besides, these techniques are usually trained and calibrated with leak-free data, which is easier to obtain.

Finally, let us highlight the existence of several benchmarks that have been used through the literature. They allow to reproduce the results and compare previous methodologies with new ones. In this category, it is necessary to highlight the recent L-TOWN benchmark from the BattLeDIM2020 competition. The top-performing methods from this contest have been described through the text and table.

4. Case study: a comparison between model-based and data-driven methods

To complete the discussion, a practical comparison between a model-based and a data-driven technique is presented is presented, using the BattleDIM2020 benchmark (see (Vrachimis et al., 2022) to complete the information about the BattLeDIM2020 characteristics). The network is represented in Fig. 4. The dataset of 2018 is used for the comparison of both methodologies, because the 2019 dataset is composed of numerous overlapping leaks that would hinder the extraction of conclusions. The features of the leaks appeared during 2018 are summarised in Table 2.

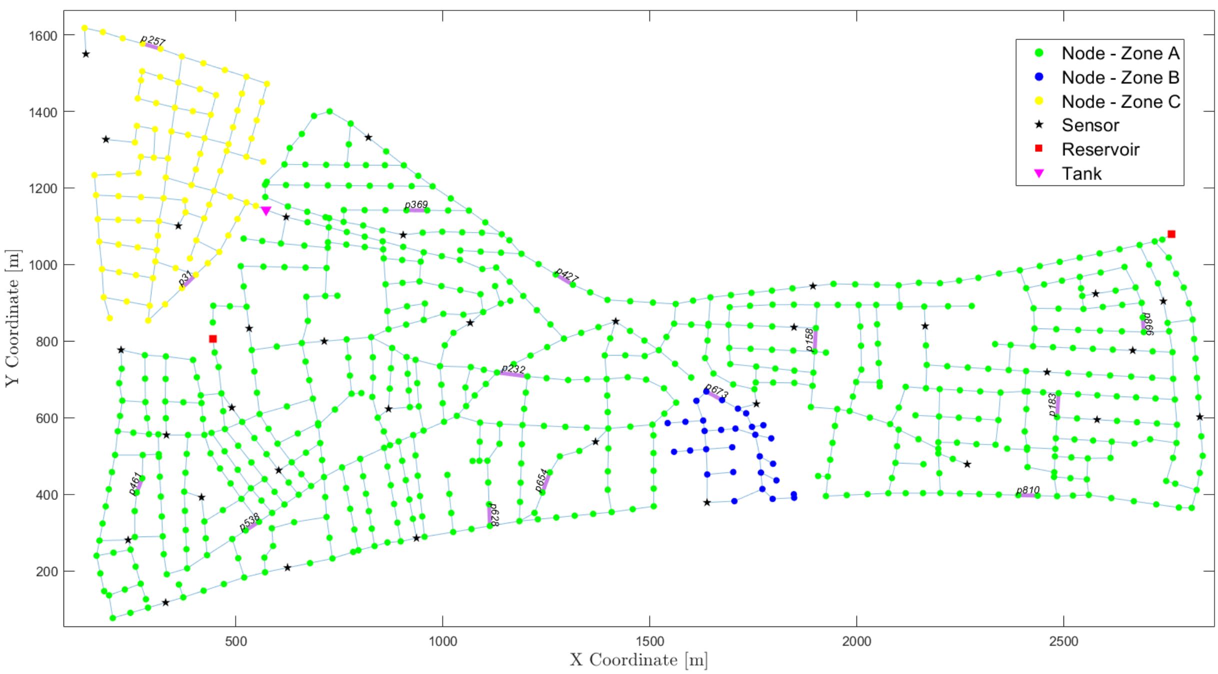

Figure 4: Graphical representation of the BattLeDIM2020 L-TOWN network. The nodes are divided into Areas (A,B and C) as shown in the legend. Reservoirs, sensors and tank are also marked. The location of the different leaks is indicated with a wider edge line, as well as a label.

The compared methodologies have different requirements and philosophies:

- The model-based approach requires a calibrated model of the WDN to be implemented into a hydraulic simulator, as well as pressure and demand measurements from a set of distributed sensors. The hydraulic simulator computes head estimations at sensor nodes for all the possible leak scenarios, in order to compare them to the heads in the nodes where the pressure is measured, obtaining the most probable leak location as a result.
- The data-driven strategy uses topological information of the network and pressure data from a set of sensors installed over inner nodes of the WDN. It is based on two stages: first the complete network state (hydraulic head) is estimated through a graph-based interpolation technique, to then compare leak and leak-free data to obtain a set of possible leak candidates.

Both methods were proposed in Romero-Ben et al. (2022), in which the model-based approach was applied to Area B and Area C, whereas the data-driven approach was applied to Area A. This was justified on the assumption that these methods were more appropriate for specific areas, considering their specific features:

Table 2
BattLeDIM2020 2018 leaks - Characteristics

Area	Link ID	Start time	End time	Leak diameter (m)	Type
C	p257	2018-01-08 10:30	2018-12-31 23:55	0.0118	Incipient
A	p461	2018-01-23 04:25	2018-04-02 11:40	0.0213	Incipient
A	p232	2018-01-31 02:35	2018-02-10 09:20	0.0201	Incipient
A	p427	2018-02-13 08:25	2018-12-31 23:55	0.0090	Incipient
B	p673	2018-03-05 15:45	2018-03-23 10:25	0.0229	Abrupt
A	p810	2018-07-28 03:05	2018-12-31 23:55	0.0100	Incipient
A	p628	2018-05-02 14:55	2018-05-29 21:20	0.0223	Incipient
A	p538	2018-05-18 08:35	2018-06-02 06:05	0.0217	Abrupt
A	p866	2018-06-01 09:05	2018-06-12 03:00	0.0181	Abrupt
C	p31	2018-06-28 10:35	2018-08-12 17:30	0.0163	Incipient
A	p654	2018-07-05 03:40	2018-12-31 23:55	0.0087	Incipient
A	p183	2018-08-07 02:35	2018-09-01 17:10	0.0158	Abrupt
A	p158	2018-10-06 02:35	2018-10-23 13:35	0.0193	Abrupt
A	p369	2018-10-26 02:05	2018-11-08 20:25	0.0193	Abrupt

(A) This zone is characterized by the presence of numerous pressure sensors (29), which, in comparison with the total number of nodes in the area (around 650), implies a high sensorization density.

(B) This area is separated from the rest of the network (Area A specifically) through a Pressure Reducing Valve (PRV), so its hydraulic behaviour is relatively independent in terms of leak effects on sensors. Therefore, the presence of only one pressure sensor largely hinders the data-driven localization.

(C) This zone is equipped with a set of AMRs, that allow estimating near real-time demand in most of the nodes, highly benefiting model-based approaches. Additionally, the pressure of only three pressure sensors reduces the possibilities of the data-driven approach.

In this work, the application of the model-based and data-driven approaches to the complete network is presented. To this end, the results of both methodologies at Areas A, B and C are shown. The presented results are provided in a graphical format, i.e., for both methodologies, the leak location is indicated on the network, and the localization results are illustrated by means of a color map over the network nodes:

- For the model-based approach, the lighter shades of green indicate lower probability, while the higher probabilities are indicated by increasingly dark shades of red.
- The data-driven approach is divided into two plots: the first represents the complete area of the network under consideration, highlighting the candidate-to-leak nodes in light blue. The second plot is a zoomed version of the first one, showing only the set of candidate nodes, and using a color map (again, green for low probability and red for high probability). The leak is marked with a cross (if it is included in the set), and the top 5 candidates are indicated with stars.

4.1. Area A

A total of 11 leaks occurred at Area A during 2018. Most of them were large leaks, with an outflow above 15 m^3/h, although there were 3 background leaks, i.e., a leak rate below 8 m^3/h. The comparison will be performed for the large leaks, selecting a set of relevant leaks because their results are more instructive for the sake of extracting solid conclusions about the advantages/limitations of both types of methods (leaving the analysis of the background leaks for future works). The rest of results from the application of both methodologies to the remaining large leaks are provided in Appendix A. Due to the large size of Area A, the leak location in the general plots (subfigures (a) and (b) of each plot) is indicated by means of a coloured square (blue for the model-based and red for the data-driven, for a better readability of the graph).

The first considered leak occurred at pipe 461. It was an incipient leak which did not reach its peak until the 27$^{\text{th}}$ of March, despite starting the 23$^{\text{rd}}$ of January. Its localization result is presented in Fig. 5. In this case, both methodologies are capable of locating the leak by narrowing its possible area to a reduced zone. However, it can be appreciated how the data-driven result is more consistent in topological terms, i.e., the transition between nodes with high probability to nodes with low probability is smooth.

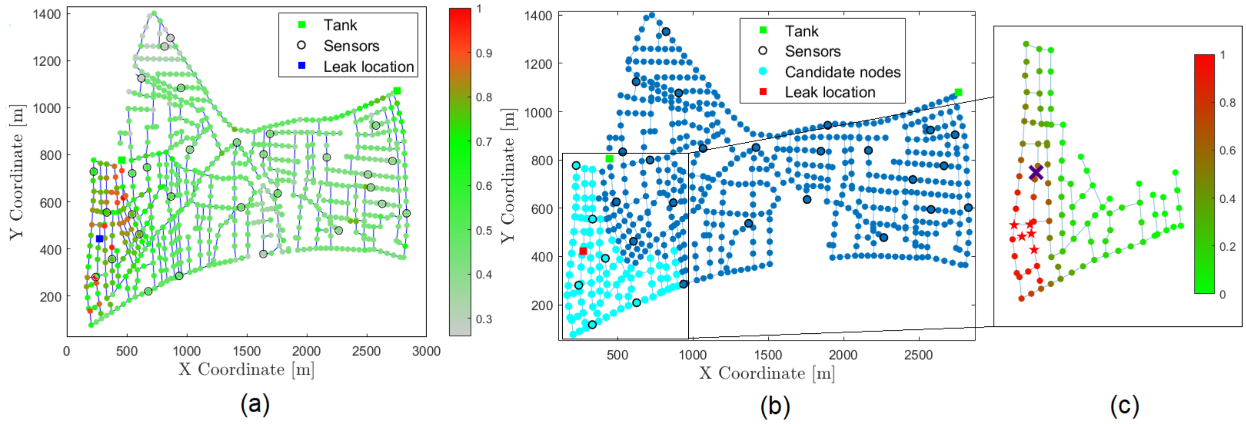

Figure 5: Localization result for leak 461: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

The second and third considered leaks occurred at pipes 628 and 538 respectively. The former is an incipient leak with a relatively high growth rate (starts the 2nd of May and reaches its peak the 16th of May). The latter is an abrupt leak that appears only two days after leak 628 reaches its peak, i.e., the 18th of May. Therefore, both leaks are almost simultaneous. Regarding the first leak, the graphical result of applying the methodologies is represented in Fig. 6. In this case, the effect of leak 538 is already affecting the localization results. The model-based approach is not capable of finding the correct leak area, pinpointing the nodes near pipe 538 as the ones with the highest probability. On the contrary, the data-driven approach selects both leaks as candidates, giving a higher probability to nodes near pipe 628 (this is caused by its higher flow rate, which affects its close sensors in a higher degree).

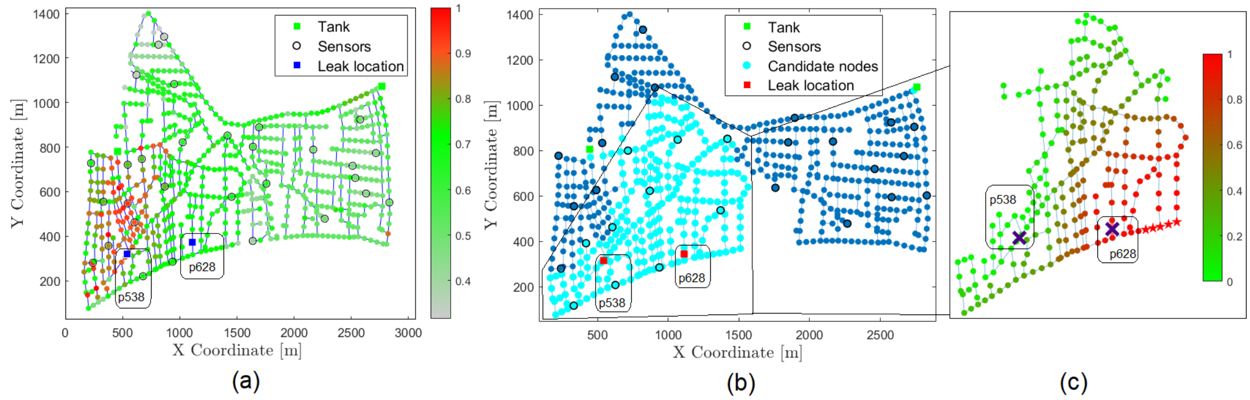

Figure 6: Localization result for leak 628 (leak 538 already appeared): (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak locations are marked by blue crosses, together with the leak ID).

Considering that leak 628 ends before leak 538, the individual effect of the latter can also be analysed, achieving the results presented in Fig. 7. In this case, both methodologies are able to successfully locate the leak in a reduced area. The data-driven result is smoother, while the model-based approach highlights various points that can be considered outliers.

To end with this couple of leaks, Fig. 8 shows the data-driven localization result for leak 628 when leak 538 has not appeared yet. Comparing it to the data-driven results in Fig. 6, it can be appreciated how the zone near leak 538 (with the leak included) is not selected in the candidates set.

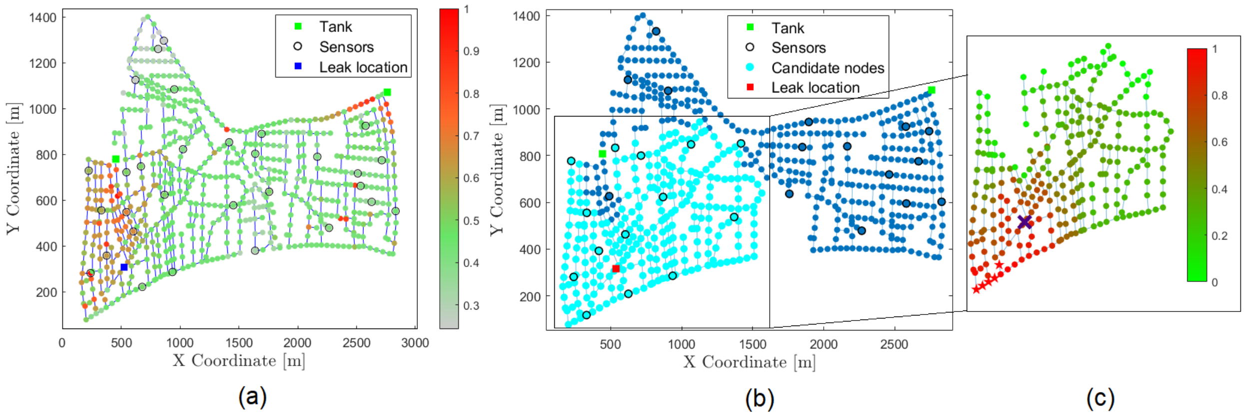

Figure 7: Localization result for leak 538: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

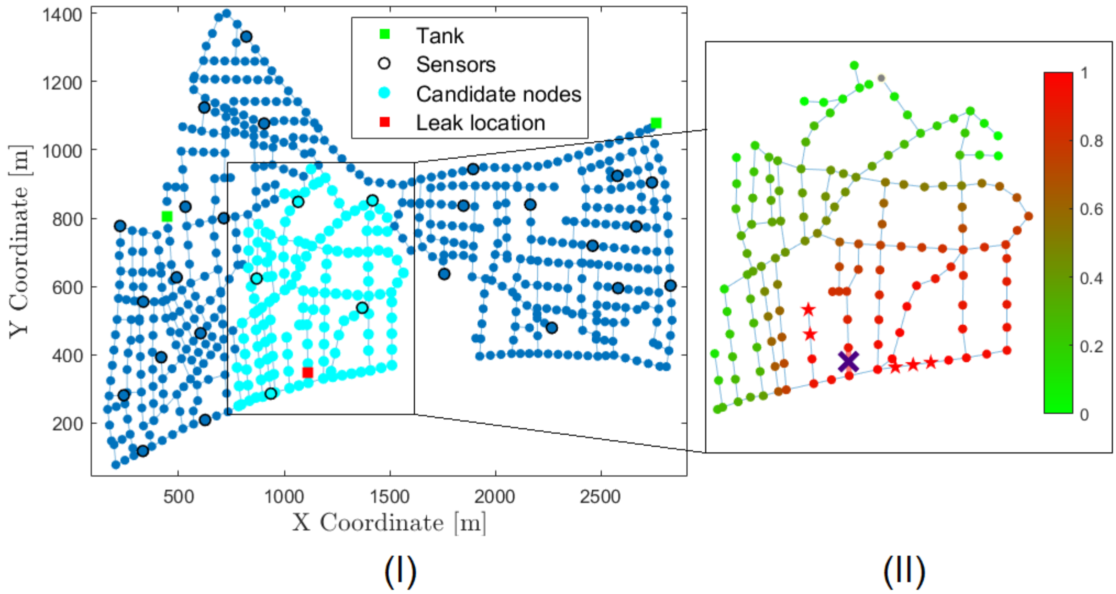

Figure 8: Localization result for leak 628 (leak 538 has not appeared yet): (I) Data-driven - candidate selection; (II) Data-driven - node-level (the exact leak location is marked by a blue cross).

Finally, the last considered leak that occurred in Area A is located in pipe 369. It is an abrupt leak, and again there are not other main leaks occurring at the same time. However, the background leaks are already stable/reaching their peak. The localization result is presented in Fig. 9. On the one hand, the model-based method provides a satisfactory result, as the nodes with the highest probability are close to the leak. On the other hand, the data-driven approach fails to select the leak in the candidate set, despite the area with the highest probability (subfigure (c)) is close to the actual leak location.

4.2. Area B

A single leak occurred at Area B during 2018. Its localization results for both methodologies are illustrated in Fig. 10. With the model-based approach, the localization result is almost perfect, with a difference of only one pipe from the obtained result to the real leak. On the contrary, the data-driven method is not even capable of selecting the proper nodes in the leak candidates. Note that the nodes with a higher probability (red stars) are scattered at a similar distance from the sensor (n215). This occurs due to the lack of measurements and the availability of a model. Therefore,

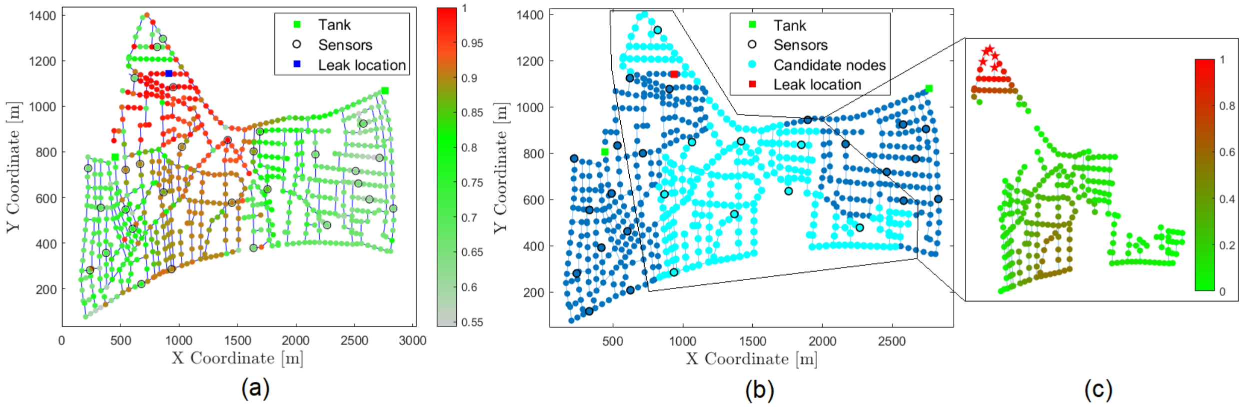

Figure 9: Localization result for leak 369: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

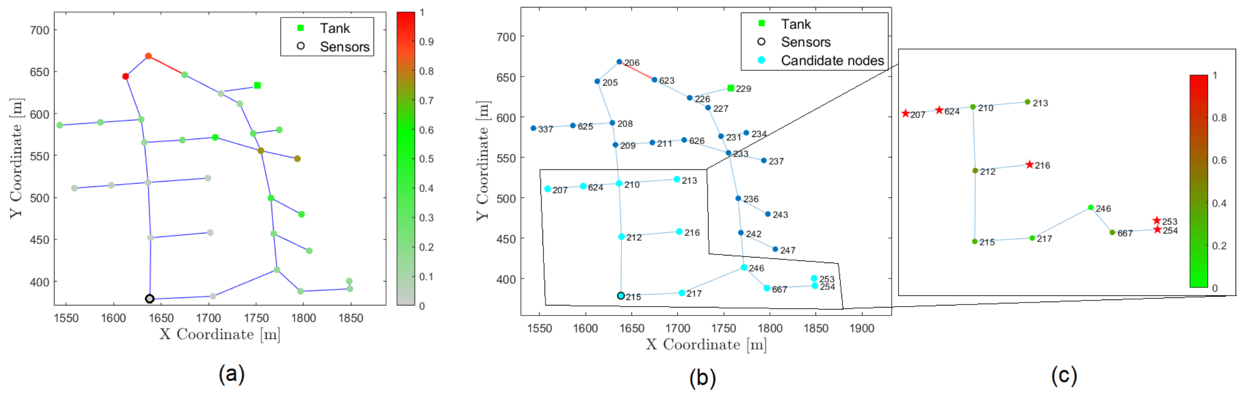

Figure 10: Localization result for leak 673: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

the superiority of the model-based approach with respect to the data-driven one is demonstrated in the case of Area B, justifying the selection at Romero et al. (2022).

4.3. Area C

Finally, a couple of leaks appeared at Area C during 2018, at p257 and p31 specifically. Note that, as stated in Table 2, the leak at p257 occurred almost at the beginning of the year, and it was not fixed. The localization results for both methodologies and both leaks are illustrated in Fig. 11 and Fig. 12 respectively.

First, regarding the leak at pipe 257, it can be appreciated how both methodologies are capable of finding it. In the model-based case, the result is perfect, whereas in the data-driven case, the best candidates (red stars) are located extremely close to the real leak.

For the localization of the leak at pipe 31, considering the model-based approach, the hydraulic model is updated to include the leak at pipe 257 as an actual demand, considering that is was not fixed. The result is perfect again. Regarding the data-driven approach, to take into account the presence of leak 257, the leak-free data for the comparison stage is selected to be gathered from the previous week, so that leak 257 was already present. The result shows that the method is not capable of completely selecting the leaky pipe in the set of candidates, although one of its endpoints is indeed in the candidates set.

The results for the 2018 Area C leaks show that:

- The model-based approach, fed with the demand information from the AMRs, is capable of perfectly locating the leaks.

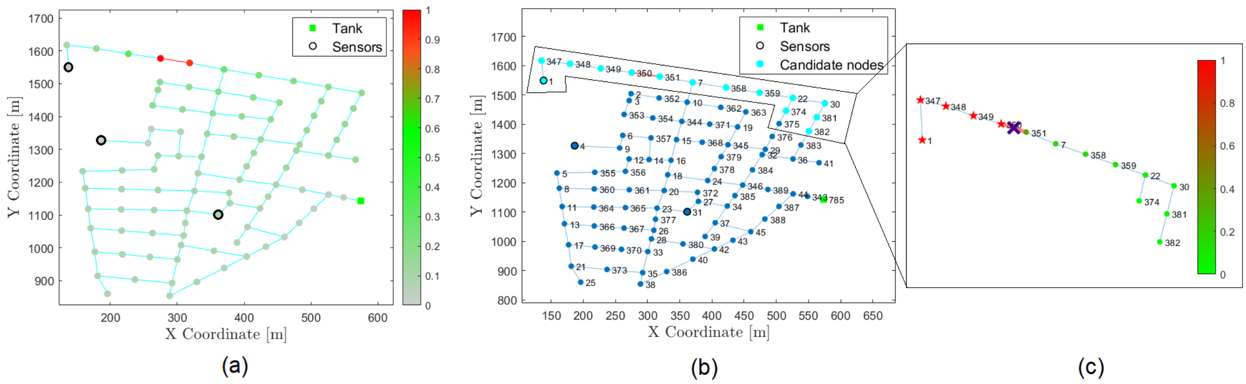

Figure 11: Localization result for leak 257: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

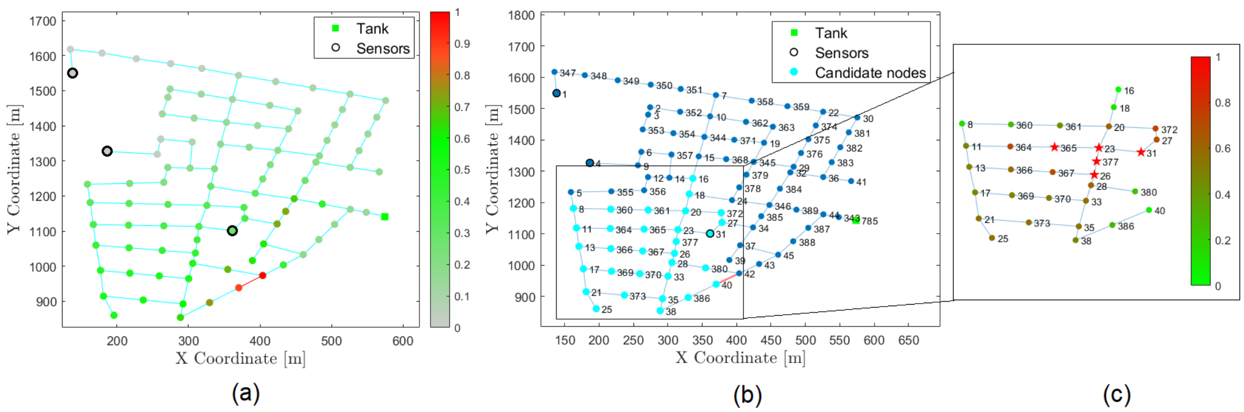

Figure 12: Localization result for leak 31: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

- The data-driven approach is capable of finding the first leak despite the low sensorization density, but it fails to locate the second. This occurs because the first leak was properly placed (close to sensor 1 and far from the rest), and no other leak was present in the network (ideal conditions). Regarding the second leak, it was placed at a difficult location (far from the sensors), and leak 257 was already present.

All these facts allow us to confirm the selection of the model-based approach for its application in Area C.

4.4. Discussion

The presented results allow us to extract several conclusions about how the model-based and data-driven methods perform the leak localization task in the BattLeDIM2020 benchmark:

- The presented results for Area A confirm the suitability of both methodologies. Nevertheless, several differences must be highlighted in their performance:
  - Regarding the model-based approach, the availability of a hydraulic model and the existence of numerous pressure sensors generally yield a satisfactory performance. However, the differences between the provided model and the one used to generate the leak information (explained in Vrachimis et al. (2022)), as well as the lack of demand measurements in a large-scale subnetwork, cause the appearance of outliers and exceptional unsatisfactory results.
  - The data-driven approach is able to locate most of the leaks, except for leak 369. This is caused by several background leaks hindering the localization, as demonstrated by the localization analysis in Fig. 13: the

nominal leak is selected to be the week after the leak disappears, so that only the background leaks are present. The successful localization confirms the effect of the background leaks over the results in Fig. 9. Therefore, in general, the data-driven methods greatly reduces the searching area, and the distance-to-leak from the top candidates is generally satisfactory.

– Additionally, the multi-leak problem is demonstrated to be solved by the data-driven approach, as shown in Fig. 6.

- The presented results in Area B confirm that the lack of sensorization greatly affects the localization performance of the data-driven approach, which yields unsatisfactory results. On the contrary, the model-based approach produces an almost perfect result, demonstrating its capability to work with insufficient measurements when considering small-sized networks (only with demand estimations).
- Finally, the results in Area C also corroborate the hypothesis of the availability of demand measurements boosting the model-based performance, which locate the leaks perfectly. On the contrary, the data-driven approach effectively locates one of the leaks, but mostly because it is quite close to one of the sensors and far from the rest; and yields an unsatisfactory result on the other leak.

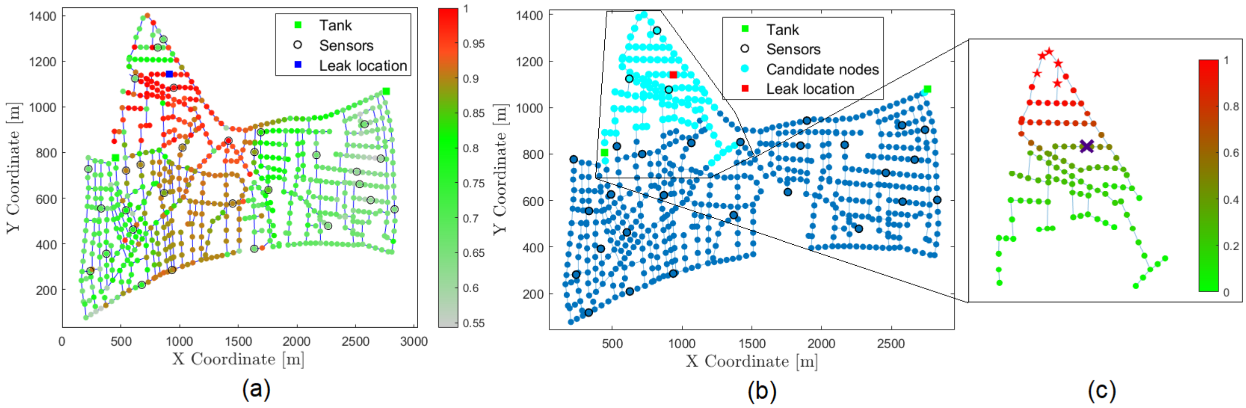

Figure 13: Localization result for leak 369 (the nominal week is selected to get rid of the background leakage): (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

All these observations help draw a few conclusions about the performance and practical utility of model-based and data-driven methods in general:

1. On the one hand, the application of model-based approaches has yielded sound results during the years, but their applicability was narrowed to the portion of WDN where a well-calibrated hydraulic model is available. Moreover, the need for precise demand estimations and/or actual demand measurements is a hard constraint to the water utilities, which may not have access to this kind of data. However, when all this information is available, their performance can hardly be improved by that of data-driven methods.
2. On the other hand, the application of data-driven methods demonstrates to be promising due to the satisfactory results that they produce when the networks are sufficiently metered. Moreover, the removal of the need for a well-calibrated hydraulic model is of great interest for modern water utilities, due to the associated costs of generation and/or continuous calibration of these models. However, the installation and maintenance of sensors can be costly for some water utilities.

To sum up, the selection of one family of methods over the rest must be carried out considering a number of factors: the availability of a hydraulic model, its degree of calibration, the physical characteristics of the network and the precise knowledge about them, the number, type and placement of existing sensors, and the the investment constraints at each water utility.

5. Conclusions

This article presents a review of the main research items in the steady-state software-based leak detection/localization field. An introduction about the motivation behind the study of these methods is provided, and the main characteristics of water distribution networks, DMAs and the posed leak management problems are described. A complete review of the majority of existing methodologies is provided, including a categorization into the main families: model-based, mixed model-based/data-driven and data-driven. A discussion from the point of view of the interests of water companies is presented and a comparative performance analysis of a model-based and a data-driven approach is developed, based on the recent and well-known benchmark from the BattLeDIM2020 competition. This study allowed to derive some conclusions about the implementation, advantages and limitations of a representative of each family.

About the classification of the reviewed works, a trend can be found in the evolution of the leak detection/localization strategies: from model-based schemes that typically require the hydraulic model of the network, the development of machine learning and data mining algorithms motivated the interest on mixed model-based/data-driven strategies, which learn the behavior of the network from different leak scenarios. However, the requirement of the availability of a hydraulic model that allows to generate those leak scenarios motivated the research of purely data-driven schemes that only require leak-free data and some topological information of the WDN.

Considering the performance yielded by the methods of the different families, larger differences start to appear when considering leak localization methods, as the availability of an hydraulic model provides extremely useful information about the effect of leaks and the behaviour of the WDN. Therefore, water utilities that have access to a well-calibrated hydraulic model may select this kind of methods before others, due to the usually better performance. Nevertheless, many utilities do not have well-calibrated models that can be used on-line for leak localization. Moreover, they do not have data from all possible leak scenarios. Thus, data-driven approaches are very appealing to utilities and have motivated new research efforts in the development of methods following this pure data-based paradigm. Moreover, the continuous development of technology leads to a reduction in the costs associated to the deployment of sensor networks over WDNs. Note that this may not only affect the amount of sensors, but the usage of new types of metering devices. Therefore, leak management methods should be designed to be capable of exploiting all the available sources of hydraulic measurements. Information fusion techniques that combine information from different sensors (pressure, flow, AMRs, velocity...) can be coupled with existing techniques in order to solve this problem. Additionally, this sensorization cost reduction can benefit data-driven methods over model-based ones, considering that the current limitations of model-free strategies lie in the scarcity of hydraulic information from the network state.

New leak management techniques need to be developed to fit the requirements of a real-world application, and therefore future research should avoid classical assumptions, which apply to academic works but hinder the implementation and successful operation of the approach. Specifically, the multi-leak problem must be considered and solved by the newly proposed methodologies, as the application of single-leak methods in a real water network could lead to undesirable errors and false alarms, causing extra and unnecessary costs related to water losses and intervention-crew organization. Additionally, the usage of realistic data is of upmost importance, and hence the different sources of uncertainty must be considered if the method needs to be trained and tested.

A. Case study results

The presented case study is based on the L-TOWN benchmark of the BattLeDIM2020 competition. Despite the results are exposed and commented in the main document, only a subset of the total set of leaks occurring in Area A are reviewed. Thus, we present here the remaining results (see Table 2 to see the most important characteristics of the leaks).

An incipient leak appeared at pipe 232 with a fast growth rate, reaching its peak the 3rd of March (only three days after its start). Its localization result is illustrated in Fig. A.1. Both methodologies produce a satisfactory but improvable performance. The model-based approach assigns a high probability to the correct nodes, although several areas of the network are highlighted, hindering the selection of the correct leak area. The data-driven method includes the leak in the candidates set, and although the leaky pipe is assigned a medium probability, the top candidates are far from the leak. Nevertheless, the data-driven approach achieves a more reduced search area.

Another leak that appeared at Area A was located in pipe 866. It was an abrupt leak that started the day before leak 538 finished. However, leak 866 lasted long enough to be able to perform a single-leak localization. The associated results are depicted in Fig. A.2. The localization performance is satisfactory with both methodologies, with the data-driven one providing a smoother result.

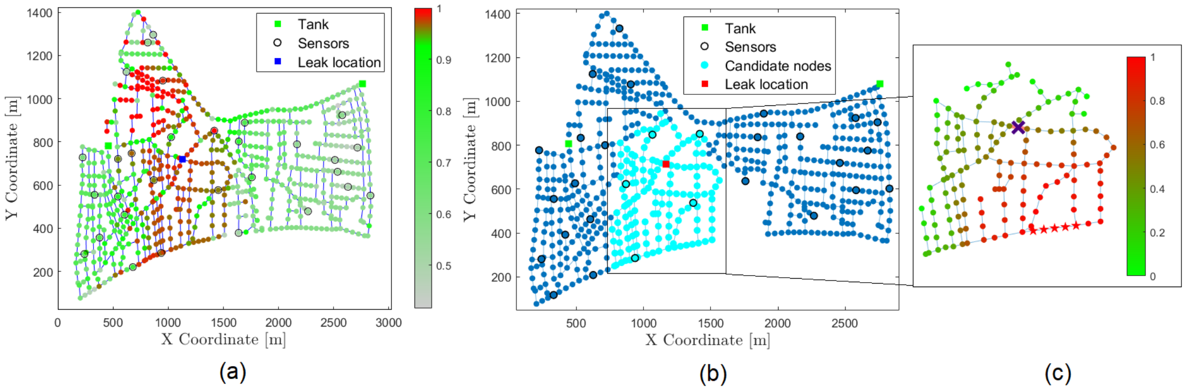

Figure A.1: Localization result for leak 232: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

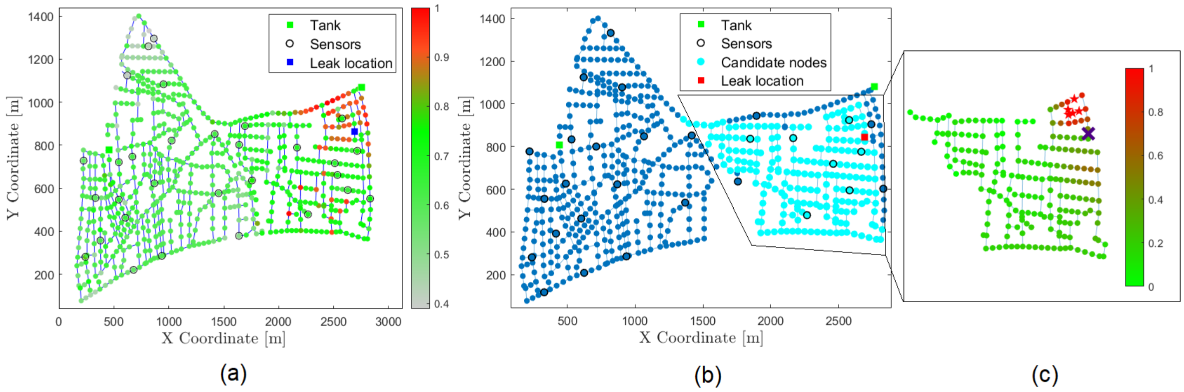

Figure A.2: Localization result for leak 866: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

An abrupt leak with no other main leaks occurring close in time (despite some background leaks start to appear/are already stable) occurred at pipe 183. The localization results for this leak are shown in Fig. A.3. On the one hand, the model-based approach highlights the leaky pipe as a high probability one, but on the other hand several areas of the network are indicated to have an equal or higher probability. On the contrary, the data-driven method properly narrows the localization area, considering the selected candidates and the reduced distance from the leak to the top candidates.

Another abrupt burst, with other main simultaneous leaks, appeared in pipe 158. The localization result for this leak event can be found in Fig. A.4. Both approaches are able to locate the area where the leak occurs, so the localization result is satisfactory. About the model-based methodology, it highlights with high probability a medium-sized zone of the network. Regarding the data-driven strategy, in this case it is capable of perfectly locating the leak, reducing the searching area to the top candidates (considering the difference in probability between this candidates and the rest of the candidate nodes).

References

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley interdisciplinary reviews: computational statistics 2, 433–459. https://doi.org/10.1002/wics.101.

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

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network, in: 2017 international conference on engineering and technology (ICET), IEEE. pp. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186.

Alpaydin, E., 2010. Introduction to machine learning. MIT press. https://dl.acm.org/doi/book/10.5555/1734076.

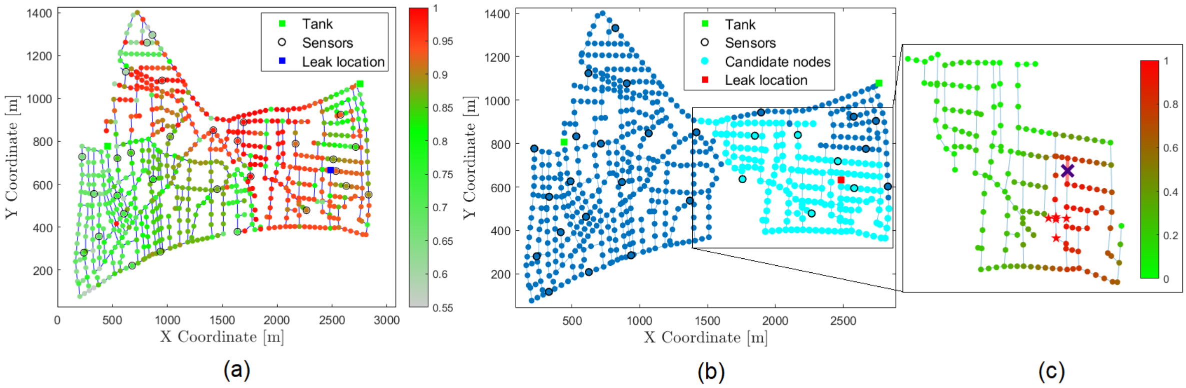

Figure A.3: Localization result for leak 183: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

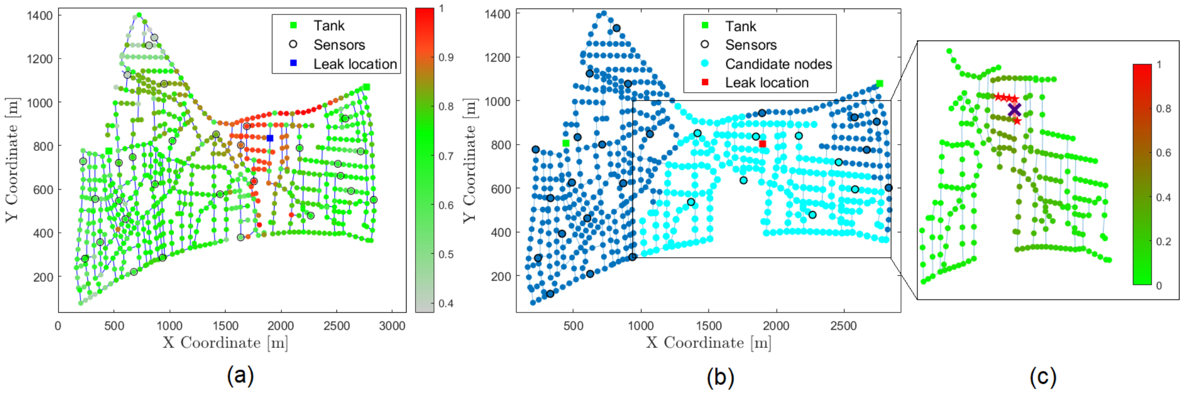

Figure A.4: Localization result for leak 158: (a) Model-based; (b) Data-driven - candidate selection; (c) Data-driven - node-level (the exact leak location is marked by a blue cross).

Alves, D., Blesa, J., Duviella, E., Rajaoarisoa, L., 2021. Robust data-driven leak localization in water distribution networks using pressure measurements and topological information. Sensors 21, 7551. https://doi.org/10.3390/s21227551.

Andersen, J.H., Powell, R.S., 2000. Implicit state-estimation technique for water network monitoring. Urban Water 2, 123–130. https://doi.org/10.1016/S1462-0758(00)00050-9.

Anderson, T.W., 2003. An introduction to multivariate statistical analysis. Technical Report. Wiley New York. http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/301_03_Anderson_An-Introduction-to-Multivariate-Statistical-Analysis-2003.pdf.

Anjana, G., Sheetal-Kumar, K., Mohan-Kumar, M., Amrutur, B., 2015. A Particle Filter Based Leak Detection Technique for Water Distribution Systems. Procedia Engineering 119, 28–34. https://doi.org/10.1016/j.proeng.2015.08.849.

Aparisi, F., 1996. Hotelling's t2 control chart with adaptive sample sizes. International Journal of Production Research 34, 2853–2862. https://doi.org/10.1080/00207549608905062.

Ares-Milián, M.J., Quiñones-Grueiro, M., Verde, C., Llanes-Santiago, O., 2021. A leak zone location approach in water distribution networks combining data-driven and model-based methods. Water 13, 2924. https://doi.org/10.3390/w13202924.

Arreguín-Cortes, F.I., Ochoa-Alejo, L.H., 1997. Evaluation of water losses in distribution networks. Journal of water resources planning and management 123, 284–291. https://doi.org/10.1061/(ASCE)0733-9496(1997)123:5(284).

Ayati, A.H., Haghighi, A., Lee, P., 2019. Statistical review of major standpoints in hydraulic transient-based leak detection. Journal of Hydraulic Structures 5, 1–26. https://doi.org/10.22055/JHS.2019.27926.1095.

Bakker, M., Vreeburg, J., Van De Roer, M., Rietveld, L., 2014. Heuristic burst detection method using flow and pressure measurements. Journal of Hydroinformatics 16, 1194. https://doi.org/10.2166/hydro.2014.120.

Basseville, M., Nikiforov, I.V., et al., 1993. Detection of abrupt changes: theory and application. volume 104. prentice Hall Englewood Cliffs. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.4795&rep=rep1&type=pdf.

Berglund, A., Areti, V.S., Brill, D., Mahinthakumar, G., 2017. Successive linear approximation methods for leak detection in water distribution systems. Journal of Water Resources Planning and Management 143, 04017042. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000784.

Bertsimas, D., Tsitsiklis, J., 1993. Simulated annealing. Statistical science 8, 10–15. https://doi.org/10.1214/ss/1177011077.

Bicik, J., Kapelan, Z., Makropoulos, C., Savić, D.A., 2011. Pipe burst diagnostics using evidence theory. Journal of Hydroinformatics 13, 596–608. https://doi.org/10.2166/hydro.2010.201.

Blesa, J., Pérez, R., 2018. Modelling uncertainty for leak localization in water networks. IFAC-PapersOnLine 51, 730–735. https://doi.org/10.1016/j.ifacol.2018.09.656.

Blesa, J., Puig, V., Saludes, J., Vento, J., 2010. Leak detection, isolation and estimation in pressurized water pipe networks using lpv models and zonotopes. IFAC Proceedings Volumes 43, 36–41. https://doi.org/10.3182/20100901-3-IT-2016.00054.

Bokor, J., Szabó, Z., Stikkel, G., 2002. Failure detection for quasi lpv systems, in: Proceedings of the 41st IEEE Conference on Decision and Control, 2002., IEEE. pp. 3318–3323. https://doi.org/10.1109/CDC.2002.1184386.

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A., Toth, P., 2012. On the optimal design of water distribution networks: a practical minlp approach. Optimization and Engineering 13, 219–246. https://doi.org/10.1007/s11081-011-9141-7.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Buchberger, S., Nadimpalli, G., 2004. Leak Estimation in Water Distribution Systems by Statistical Analysis of Flow Readings. Journal of water resources planning and management 130, 321–329. https://doi.org/10.1061/(ASCE)0733-9496(2004)130:4(321).

Candelieri, A., Soldi, D., Conti, D., Archetti, F., 2014. Analytical leakages localization in water distribution networks through spectral clustering and support vector machines. the icewater approach. Procedia Engineering 89, 1080–1088. https://doi.org/10.1016/j.proeng.2014.11.228.

Cantos, W.P., Juran, I., Tinelli, S., 2020. Machine-learning–based risk assessment method for leak detection and geolocation in a water distribution system. Journal of Infrastructure Systems 26, 04019039. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000517.

Capelo, M., Brentan, B., Monteiro, L., Covas, D., 2021. Near–real time burst location and sizing in water distribution systems using artificial neural networks. Water 13, 1841. https://doi.org/10.3390/w13131841.

Caputo, A.C., Pelagagge, P.M., 2003. Using neural networks to monitor piping systems. Process Safety Progress 22, 119–127. https://doi.org/10.1002/prs.680220208.

Casillas, M.V., Garza Castañón, L.E., Puig, V., 2014. Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities. Journal of Hydroinformatics 16, 649–670. https://doi.org/10.2166/hydro.2013.019.

Charalambous, B., 2008. Use of district metered areas coupled with pressure optimisation to reduce leakage. Water Science and Technology: Water Supply 8, 57–62. https://doi.org/10.2166/ws.2008.030.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. Stl: A seasonal-trend decomposition. J. Off. Stat 6, 3–73. http://www.nniiem.ru/file/news/2016/stl-statistical-model.pdf.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. Journal of the American statistical association 83, 596–610. https://doi.org/10.1080/01621459.1988.10478639.

Cohen, B., 2004. Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. World development 32, 23–51. https://doi.org/10.1016/j.worlddev.2003.04.008.

Colombo, A.F., Lee, P., Karney, B.W., 2009. A selective literature review of transient-based leak detection methods. Journal of hydro-environment research 2, 212–227. https://doi.org/10.1016/j.jher.2009.02.003.

Costanzo, F., Morosini, A.F., Veltri, P., Savić, D., 2014. Model calibration as a tool for leakage identification in wds: A real case study. Procedia Engineering 89, 672–678. https://doi.org/10.1016/j.proeng.2014.11.493.

Cuthill, E., 1972. Several strategies for reducing the bandwidth of matrices, in: Sparse matrices and their applications. Springer, pp. 157–166. https://doi.org/10.1007/978-1-4615-8675-3_14.

Daniel, I., Pesantez, J., Letzgus, S., Khaksar Fasaee, M.A., 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. Journal of Water Resources Planning and Management 148, 04022025. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001535.

Delashmit, W.H., Manry, M.T., et al., 2005. Recent developments in multilayer perceptron neural networks, in: Proceedings of the 7th Annual Memphis Area Engineering and Science Conference, MAESC. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.621.1785.

Diao, K., Sweetapple, C., Farmani, R., Fu, G., Ward, S., Butler, D., 2016. Global resilience analysis of water distribution systems. Water research 106, 383–393. https://doi.org/10.1016/j.watres.2016.10.011.

Djuric, P.M., Kotecha, J.H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M.F., Miguez, J., 2003. Particle filtering. IEEE signal processing magazine 20, 19–38. https://doi.org/10.1109/MSP.2003.1236770.

Eliades, D., Polycarpou, M., 2012. Leakage fault detection in district metered areas of water distribution systems. Journal of Hydroinformatics 14, 992. https://doi.org/10.2166/hydro.2012.109.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi-and megavariate data analysis. principles and applications. Umetrics Academy, Umeå , 43https://doi.org/10.1002/cem.713.

Escalera, A., Garza-Castañón, L., Vargas-Martínez, A., 2012. Multi-leak detection with wavelet analysis in water distribution networks, in: 20th Mediterranean Conference on Control & Automation (MED), IEEE. pp. 1155–1160. https://doi.org/10.1109/MED.2012.6265794.

Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise., in: kdd, pp. 226–231. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf.

Fang, Q., Zhang, J., Xie, C., Yang, Y., 2019. Detection of multiple leakage points in water distribution networks based on convolutional neural networks. Water Supply 19, 2231–2239. https://doi.org/10.2166/ws.2019.105.

Ferrandez-Gamot, L., Busson, P., Blesa, J., Tornil-Sin, S., Puig, V., Duviella, E., Soldevila, A., 2015. Leak Localization in Water Distribution Networks using Pressure Residuals and Classifiers. IFAC-PapersOnLine 48, 220–225. https://doi.org/10.1016/j.ifacol.2015.09.531.

Fujiwara, O., Khang, D.B., 1990. A two-phase decomposition method for optimal design of looped water distribution networks. Water resources research 26, 539–549. https://doi.org/10.1029/WR026i004p00539.

Fusco, F., Ba, A., 2012. Fault diagnosis of water distribution networks based on state-estimation and hypothesis testing. 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton) , 886–892https://doi.org/10.1109/Allerton.2012.6483312.

Galeano, J.C., Veloza-Suan, A., González, F.A., 2005. A comparative analysis of artificial immune network models, in: Proceedings of the 7th annual conference on Genetic and evolutionary computation, pp. 361–368. https://doi.org/10.1145/1068009.1068066.

Gertler, J., Romera, J., Puig, V., Quevedo, J., 2010. Leak detection and isolation in water distribution networks using principal component analysis and structured residuals, in: 2010 Conference on Control and Fault-Tolerant Systems (SysTol), IEEE. pp. 191–196. https://doi.org/10.1109/SYSTOL.2010.5676043.

Gomes, S.C., Vinga, S., Henriques, R., 2021. Spatiotemporal correlation feature spaces to support anomaly detection in water distribution networks. Water 13, 2551. https://doi.org/10.3390/w13182551.

Gordon, J., Shortliffe, E.H., 1984. The dempster-shafer theory of evidence. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project 3, 832–838. https://www.aaai.org/Papers/Buchanan/Buchanan15.pdf.

Goulet, J.A., Coutu, S., Smith, I.F., 2013. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. Advanced Engineering Informatics 27, 261–269. https://doi.org/10.1016/j.aei.2013.01.001.

Gupta, A., Bokde, N., Marathe, D., Kulat, K., 2017. Leakage reduction in water distribution systems with efficient placement and control of pressure reducing valves using soft computing techniques. Engineering, Technology & Applied Science Research 7, 1528–1534. https://doi.org/10.48084/etasr.1032.

Hamilton, S., Charalambous, B., 2013. Leak detection: technology and implementation. IWA Publishing. https://doi.org/10.26530/OAPEN_578133.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https:/doi.org/10.1109/CVPR.2016.90.

Hu, X., Han, Y., Yu, B., Geng, Z., Fan, J., 2021. Novel leakage detection and water loss management of urban water supply network using multiscale neural networks. Journal of Cleaner Production 278, 123611. https://doi.org/10.1016/j.jclepro.2020.123611.

Hutton, C., Kapelan, Z., 2015. Real-time burst detection in Water Distribution Systems using a Bayesian demand forecasting methodology. Procedia Engineering 119, 13–18. https://doi.org/10.1016/j.proeng.2015.08.847.

Islam, M.R., Azam, S., Shanmugam, B., Mathur, D., 2022. A review on current technologies and future direction of water leakage detection in water distribution network. IEEE Access https://doi.org/10.1109/ACCESS.2022.3212769.

Islam, M.S., Sadiq, R., Rodriguez, M.J., Francisque, A., Najjaran, H., Hoorfar, M., 2011. Leakage detection and location in water distribution systems using a fuzzy-based methodology. Urban water journal 8, 351–365. https://doi.org/10.1080/1573062X.2011.617829.

Jang, J.S., 1993. Anfis: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics 23, 665–685. https://doi.org/10.1109/21.256541.

Javadiha, M., Blesa, J., Soldevila, A., Puig, V., 2019. Leak localization in water distribution networks using deep learning, in: 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE. pp. 1426–1431. https://doi.org/10.1109/CoDIT.2019.8820627.

Jensen, T.N., Puig, V., Romera, J., Kallesøe, C.S., Wisniewski, R., Bendtsen, J.D., 2018. Leakage localization in water distribution using data-driven models and sensitivity analysis. Ifac-papersonline 51, 736–741. https://doi.org/10.1016/j.ifacol.2018.09.657.

Jung, D., Lansey, K., 2015. Water distribution system burst detection using a nonlinear kalman filter. Journal of Water Resources Planning and Management 141, 04014070. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000464.

Kang, D., Lansey, K., 2014. Novel approach to detecting pipe bursts in water distribution networks. Journal of Water Resources Planning and Management 140, 121–127. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000264.

Kaufman, L., Rousseeuw, P.J., 2009. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons. https://doi.org/10.1002/9780470316801.

Kettler, A., Goulter, I., 1985. An analysis of pipe breakage in urban water distribution networks. Canadian journal of civil engineering 12, 286–293. https://doi.org/10.1139/l85-030.

Kim, Y., Lee, S.J., Park, T., Lee, G., Suh, J.C., Lee, J.M., 2016. Robust leak detection and its localization using interval estimation for water distribution network. Computers & Chemical Engineering 92, 1–17. https://doi.org/10.1016/j.compchemeng.2016.04.027.

Kohonen, T., 1990. The self-organizing map. Proceedings of the IEEE 78, 1464–1480. https://doi.org/10.1109/5.58325.

Kohonen, T., 2013. Essentials of the self-organizing map. Neural networks 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the witwatersrand. Journal of the Southern African Institute of Mining and Metallurgy 52, 119–139. https://hdl.handle.net/10520/AJA0038223X_4792.

Kumar, M., et al., 2013. An optimized farthest first clustering algorithm, in: 2013 Nirma University International Conference on Engineering (NUiCONE), IEEE. pp. 1–5. https://doi.org/10.1109/NUiCONE.2013.6780070.

Van der Laan, M., Pollard, K., Bryan, J., 2003. A new partitioning around medoids algorithm. Journal of Statistical Computation and Simulation 73, 575–584. https://doi.org/10.1080/0094965031000136012.

Laucelli, D., Romano, M., Savić, D., Giustolisi, O., 2016. Detecting anomalies in water distribution networks using EPR modelling paradigm. Journal of Hydroinformatics 18, 409–427. https://doi.org/10.2166/hydro.2015.113.

LeChevallier, M., Gullick, R., Karim, M., Friedman, M., Funk, J., 2003. The potential for health risks from intrusion of contaminants into the distribution system from pressure transients. J. Water Health 1, 3–14. https://doi.org/10.2166/wh.2003.0002.

Leflaive, X., 2012. Water Outlook to 2050: The OECD calls for early and strategic action, in: Global Water Forum. https://globalwaterforum.org/2012/05/21/water-outlook-to-2050-the-oecd-calls-for-early-and-strategic-action/.

Li, J., Zheng, W., Lu, C., 2022a. An accurate leakage localization method for water supply network based on deep learning network. Water Resources Management , 1–17https://doi.org/10.1007/s11269-022-03144-x.

Li, X., Chu, S., Zhang, T., Yu, T., Shao, Y., 2022b. Leakage localization using pressure sensors and spatial clustering in water distribution systems. Water Supply 22, 1020–1034. https://doi.org/10.2166/ws.2021.219.

Li, X., Li, G.J., 2010. Leak detection of municipal water supply network based on the cluster-analysis and fuzzy pattern recognition, in: 2010 International Conference on E-Product E-Service and E-Entertainment, IEEE. pp. 1–5. https://doi.org/10.1109/ICEEE.2010.5660550.

Li, Z., Wang, J., Yan, H., Li, S., Tao, T., Xin, K., 2022c. Fast detection and localization of multiple leaks in water distribution network jointly driven by simulation and machine learning. Journal of Water Resources Planning and Management 148, 05022005. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001574.

Liemberger, R., Wyatt, A., 2019. Quantifying the global non-revenue water problem. Water Supply 19, 831–837. https://doi.org/10.2166/ws.2018.129.

Lijuan, W., Hongwei, Z., Hui, J., 2012. A leak detection method based on epanet and genetic algorithm in water distribution systems, in: Software Engineering and Knowledge Engineering: Theory and Practice. Springer, pp. 459–465. https://doi.org/10.1007/978-3-642-03718-4_57.

Lučin, I., Lučin, B., Čarija, Z., Sikirica, A., 2021. Data-driven leak localization in urban water distribution networks using big data for random forest classifier. Mathematics 9, 672. https://doi.org/10.3390/math9060672.

Mai, Q., 2013. A review of discriminant analysis in high dimensions. Wiley Interdisciplinary Reviews: Computational Statistics 5, 190–197. https://doi.org/10.1002/wics.1257.

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. Journal of Water Resources Planning and Management 148, 04021096. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001503.

Mashford, J., De Silva, D., Burn, S., Marney, D., 2012. Leak detection in simulated water pipe networks using SVM. Applied Artificial Intelligence 26, 429–444. https://doi.org/10.1080/08839514.2012.670974.

Medsker, L., Jain, L.C., 1999. Recurrent neural networks: design and applications. CRC press. https://dl.acm.org/doi/book/10.5555/553011.

Mekonnen, M.M., Hoekstra, A.Y., 2016. Four billion people facing severe water scarcity. Science advances 2, e1500323. https://doi.org/10.1126/sciadv.1500323.

Menapace, A., Avesani, D., Righetti, M., Bellin, A., Pisaturo, G., 2018. Uniformly distributed demand EPANET extension. Water Resour. Manag. 32, 2165–2180. https://doi.org/10.1007/s11269-018-1924-6.

Misiunas, D., Vítkovský, J., Olsson, G., Lambert, M., Simpson, A., 2006. Failure monitoring in water distribution networks. Water science and technology 53, 503–511. https://doi.org/10.2166/wst.2006.154.

Moré, J.J., 1978. The levenberg-marquardt algorithm: implementation and theory, in: Numerical analysis. Springer, pp. 105–116. https://doi.org/10.1007/BFb0067700.

Mounce, S., Boxall, J., Machell, J., 2008. Online application of ann and fuzzy logic system for burst detection, in: Water Distribution Systems Analysis 2008, pp. 1–12. https://doi.org/10.1061/41024(340)62.

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

Mounce, S.R., Mounce, R.B., Boxall, J.B., 2011. Novelty detection for time series data analysis in water distribution systems using support vector machines. Journal of hydroinformatics 13, 672–686. https://doi.org/10.2166/hydro.2010.144.

Mutikanga, H., Sharma, S., Vairavamoorthy, K., 2013. Methods and tools for managing losses in water distribution systems. Journal of Water Resources Planning and Management 139, 166–174. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000245.

Myers, D.E., 1982. Matrix formulation of co-kriging. Journal of the International Association for Mathematical Geology 14, 249–257. https://doi.org/10.1007/BF01032887.

Narayanan, I., Vasan, A., Sarangan, V., Sivasubramaniam, A., 2014. One meter to find them all: water network leak localization using a single flow meter. Proc. IPSN , 47–58https://doi.org/10.1109/IPSN.2014.6846740.

Nebot, À., Mugica, F., 2012. Fuzzy inductive reasoning: a consolidated approach to data-driven construction of complex dynamical systems. International Journal of General Systems 41, 645–665. https://doi.org/10.1080/03081079.2012.691203.

Nowicki, A., Grochowski, M., 2011. Kernel pca in application to leakage detection in drinking water distribution system, in: International Conference on Computational Collective Intelligence, Springer. pp. 497–506. https://doi.org/10.1007/978-3-642-23935-9_49.

Ocampo-Martinez, C., Puig, V., Cembrano, G., Quevedo, J., 2013. Application of predictive control strategies to the management of complex networks in the urban water cycle [applications of control]. IEEE Control Systems Magazine 33, 15–41. https://doi.org/10.1109/MCS.2012.2225919.

Okeya, I., Hutton, C., Kapelan, Z., 2015. Locating pipe bursts in a district metered area via online hydraulic modelling. Procedia engineering 119, 101–110. https://doi.org/10.1016/j.proeng.2015.08.859.

Okeya, I., Kapelan, Z., Hutton, C., Naga, D., 2014. Online burst detection in a water distribution system using the kalman filter and hydraulic modelling. Procedia Engineering 89, 418–427. https://doi.org/10.1016/j.proeng.2014.11.207.

Oliver, M.A., Webster, R., 1990. Kriging: a method of interpolation for geographical information systems. International Journal of Geographical Information System 4, 313–332. https://doi.org/10.1080/02693799008941549.

Palau, C., Arregui, F., Carlos, M., 2012. Burst Detection in Water Networks Using Principal Component Analysis. Journal of Water Resources Planning and Management 138, 47–54. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000147.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., Peralta, A., 2011. Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Engineering Practice 19, 1157–1167. https://doi.org/10.1016/j.conengprac.2011.06.004.

Perez, R., Sanz, G., Puig, V., Quevedo, J., Escofet, M.A.C., Nejjari, F., Meseguer, J., Cembrano, G., Tur, J.M.M., Sarrate, R., 2014. Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in barcelona [applications of control]. IEEE control systems magazine 34, 24–36. https://doi.org/10.1109/MCS.2014.2320336.

Pisner, D.A., Schnyer, D.M., 2020. Support vector machine, in: Machine learning. Elsevier, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Podobnik, B., Stanley, H.E., 2008. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. Physical review letters 100, 084102. https://doi.org/10.1103/PhysRevLett.100.084102.

Poulakis, Z., Valougeorgis, D., Papadimitriou, C., 2003. Leakage detection in water pipe networks using a bayesian probabilistic framework. Probabilistic Engineering Mechanics 18, 315–327. https://doi.org/10.1016/S0266-8920(03)00045-6.

Price, K.V., 2013. Differential evolution. Handbook of Optimization: From Classical to Modern Approach , 187–214https://doi.org/10.1007/978-3-642-30504-7_8.

Pudar, R.S., Liggett, J.A., 1992. Leaks in pipe networks. Journal of Hydraulic Engineering 118, 1031–1046. https://doi.org/10.1061/(ASCE)0733-9429(1992)118:7(1031).

Puig, V., Cugueró, P., Quevedo, J., 2001. Worst-case state estimation and simulation of uncertain discrete-time systems using zonotopes, in: 2001 european control conference (ECC), IEEE. pp. 1691–1697. https://doi.org/10.23919/ECC.2001.7076164.

Puust, R., Kapelan, Z., Savic, D., Koppel, T., 2006. Probabilistic leak detection in pipe networks using the scem-ua algorithm, in: Water Distribution Systems Analysis Symposium 2006, pp. 1–12. https://doi.org/10.1061/40941(247)15.

Puust, R., Kapelan, Z., Savić, D., Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water J. 7, 25–45. https://doi.org/10.1080/15730621003610878.

Qi, S., Gao, J., Wu, W., Qiao, Y., Tu, M., Wang, J., 2014. Research on an optimized leakage locating model in water distribution system. Procedia Engineering 89, 1569–1576. https://doi.org/10.1016/j.proeng.2014.11.457.

Quiñones, M., Verde, C., Prieto-Moreno, A., Llanes-Santiago, O., 2018. An unsupervised approach to leak detection and location in water distribution networks. International Journal of Applied Mathematics and Computer Science 28. https://doi.org/10.2478/amcs-2018-0020.

Quiñones-Grueiro, M., Milián, M.A., Rivero, M.S., Neto, A.J.S., Llanes-Santiago, O., 2021. Robust leak localization in water distribution networks using computational intelligence. Neurocomputing 438, 195–208. https://doi.org/10.1016/j.neucom.2020.04.159.

Rajeswaran, A., Narasimhan, S., Narasimhan, S., 2017. A graph partitioning algorithm for leak detection in water distribution networks. Computers & Chemical Engineering https://doi.org/10.1016/j.compchemeng.2017.08.007.

Rasamoelina, A.D., Adjailia, F., Sinčák, P., 2020. A review of activation function for artificial neural network, in: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE. pp. 281–286. https://doi.org/10.1109/SAMI48414.2020.9108717.

Reca, J., Martínez, J., 2006. Genetic algorithms for the design of looped irrigation water distribution networks. Water resources research 42. https://doi.org/10.1029/2005WR004383.

Ribeiro, L., Sousa, J., Marques, A.S., Simões, N.E., 2015. Locating leaks with trustrank algorithm support. Water 7, 1378–1401. https://doi.org/10.3390/w7041378.

Rojek, I., Studziński, J., 2014. Comparision of different types of neuronal nets for failures locaion within water-supply networks. Eksploatacja i Niezawodność 16, 42–47. https://bibliotekanauki.pl/articles/1365982.pdf.

Romano, M., Kapelan, Z., Savić, D., 2013. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. Journal of Hydroinformatics 15, 634–651. https://doi.org/10.2166/hydro.2013.094.

Romano, M., Kapelan, Z., Savić, D., 2014. Automated detection of pipe bursts and other events in water distribution systems. Journal of Water Resources Planning and Management 140, 457–467. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000339.

Romano, M., Woodward, K., Kapelan, Z., 2017. Statistical Process Control Based System for Approximate Location of Pipe Bursts and Leaks in Water Distribution Systems. Procedia Engineering 186, 236–243. https://doi.org/10.1016/j.proeng.2017.03.235.

Romero, L., Blesa, J., Puig, V., Cembrano, G., 2022. Clustering-learning approach to the localization of leaks in water distribution networks. Journal of Water Resources Planning and Management 148, 04022003. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001527.

Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., Duviella, E., 2022. Leak localization in water distribution networks using data-driven and model-based approaches. Journal of Water Resources Planning and Management 148, 04022016. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001542.

Romero-Tapia, G., Fuente, M., Puig, V., 2018. Leak localization in water distribution networks using fisher discriminant analysis. IFAC-PapersOnLine 51, 929–934. https://doi.org/10.1016/j.ifacol.2018.09.686.

Roser, M., Ritchie, H., Ortiz-Ospina, E., 2013. World population growth. Our World in Data https://ourworldindata.org/world-population-growth.

Rosich, A., Puig, V., Casillas, M., 2014. Leak localization in drinking water distribution networks using structured residuals. International Journal of Adaptive Control and Signal Processing 28, 991–1007. https://doi.org/10.1002/acs.2515.

Rossman, L.A., 2000. EPANET 2: Users Manual. Cincinnati US Environmental Protection Agency National Risk Management Research Laboratory 38. https://epanet.es/wp-content/uploads/2012/10/EPANET_User_Guide.pdf.

Sanz, G., Pérez, R., Escobet, A., 2012. Leakage localization in water networks using fuzzy logic. 2012 20th Mediterranean Conference on Control & Automation (MED) , 646–651https://doi.org/10.1109/MED.2012.6265711.

Sanz, G., Pérez, R., Kapelan, Z., Savic, D., 2016. Leak detection and localization through demand components calibration. Journal of Water Resources Planning and Management 142, 04015057. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000592.

Scarpa, F., Lobba, A., Becciu, G., 2016. Elementary dma design of looped water distribution networks with multiple sources. Journal of Water Resources Planning and Management 142, 04016011. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000639.

Sentz, K., Ferson, S., 2002. Combination of evidence in dempster-shafer theory https://doi.org/10.2172/800792.

Shao, Y., Li, X., Zhang, T., Chu, S., Liu, X., 2019. Time-series-based leakage detection using multiple pressure sensors in water distribution systems. Sensors 19, 3070. https://doi.org/10.3390/s19143070.

Shekofteh, M., Jalili Ghazizadeh, M., Yazdi, J., 2020. A methodology for leak detection in water distribution networks using graph theory and artificial neural network. Urban Water Journal 17, 525–533. https://doi.org/10.1080/1573062X.2020.1797832.

Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM national conference, pp. 517–524. https://doi.org/10.1145/800186.810616.

Shukla, S., Naganna, S., 2014. A review on k-means data clustering approach. International Journal of Information and Computation Technology 4, 1847–1860. https://www.ripublication.com/irph/ijict_spl/ijictv4n17spl_15.pdf.

Soldevila, A., 2018. Robust leak localization in water distribution networks using machine learning techniques. Ph.D. thesis. Universitat Politècnica de Catalunya. http://hdl.handle.net/10803/668645.

Soldevila, A., Blesa, J., Jensen, T.N., Tornil-Sin, S., Fernández-Cantí, R.M., Puig, V., 2020. Leak localization method for water-distribution networks using a data-driven model and dempster–shafer reasoning. IEEE Transactions on Control Systems Technology 29, 937–948. https://doi.org/10.1109/TCST.2020.2982349.

Soldevila, A., Blesa, J., Tornil-Sin, S., Duviella, E., Fernandez-Canti, R.M., Puig, V., 2016. Leak localization in water distribution networks using a mixed model-based/data-driven approach. Control Engineering Practice 55, 162–173. https://doi.org/10.1016/j.conengprac.2016.07.006.

Soldevila, A., Boracchi, G., Roveri, M., Tornil-Sin, S., Puig, V., 2021. Leak detection and localization in water distribution networks by combining expert knowledge and data-driven models. Neural Computing and Applications , 1–21.

Soldevila, A., Fernandez-Canti, R.M., Blesa, J., Tornil-Sin, S., Puig, V., 2017. Leak localization in water distribution networks using bayesian classifiers. Journal of Process Control 55, 1–9. https://doi.org/10.1016/j.jprocont.2017.03.015.

Sonaje, N.P., Joshi, M.G., 2015. A review of modeling and application of water distribution networks (wdn) softwares. International Journal of Technical Research and Applications 3, 174–178. https://www.ijtra.com/view/a-review-of-modeling-and-application-of-water-distribution-networks-wdn-softwares.pdf.

Sophocleous, S., Savić, D., Kapelan, Z., 2019. Leak localization in a real water distribution network based on search-space reduction. Journal of Water Resources Planning and Management 145, 04019024. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001079.

Steffelbauer, D., Günther, M., Fuchs-Hanusch, D., 2017. Leakage localization with differential evolution: a closer look on distance metrics. Procedia engineering 186, 444–451. https://doi.org/10.1016/j.proeng.2017.03.251.

Steffelbauer, D.B., Deuerlein, J., Gilbert, D., Abraham, E., Piller, O., 2022. Pressure-leak duality for leak detection and localization in water distribution systems. Journal of Water Resources Planning and Management 148, 04021106. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001515.

Sterling, M., Bargiela, A., 1984. Leakage reduction by optimised control of valves in water networks. Transactions of the Institute of Measurement and Control 6, 293–298. https://doi.org/10.1177/014233128400600603.

Sun, C., Parellada, B., Puig, V., Cembrano, G., 2019. Leak localization in water distribution networks using pressure and data-driven classifier approach. Water 12, 54. https://doi.org/10.3390/w12010054.

Tao, T., Huang, H., Li, F., Xin, K., 2014. Burst detection using an artificial immune network in water-distribution systems. Journal of Water Resources Planning and Management 140, 04014027. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000405.

van Thienen, P., 2013. A method for quantitative discrimination in flow pattern evolution of water distribution supply areas with interpretation in terms of demand and leakage. Journal of Hydroinformatics 15, 86. https://doi.org/10.2166/hydro.2012.171.

Vicente, D., Garrote, L., Sánchez, R., Santillán, D., 2016. Pressure management in water distribution systems: Current status, proposals, and future trends. Journal of Water Resources Planning and Management 142, 04015061. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000589.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Statistics and computing 17, 395–416. https://doi.org/10.1007/s11222-007-9033-z.

Vrachimis, S.G., Eliades, D.G., Taormina, R., Kapelan, Z., Ostfeld, A., Liu, S., Kyriakou, M., Pavlou, P., Qiu, M., Polycarpou, M.M., 2022. Battle of the leakage detection and isolation methods. Journal of Water Resources Planning and Management 148, 04022068. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001601.

Vrachimis, S.G., Kyriakou, M.S., et al., 2018. Leakdb: A benchmark dataset for leakage diagnosis in water distribution networks, in: WDSA/CCWI Joint Conference Proceedings. https://doi.org/10.5281/zenodo.1313116.

Vrachimis, S.G., Timotheou, S., Eliades, D.G., Polycarpou, M.M., 2021. Leakage detection and localization in water distribution systems: A model invalidation approach. Control Engineering Practice 110, 104755. https://doi.org/10.1016/j.proeng.2014.11.457.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water resources research 39. https://doi.org/10.1029/2002WR001642.

Wachla, D., Przystalka, P., Moczulski, W., 2015. A method of leakage location in water distribution networks using artificial neuro-fuzzy system. IFAC-PapersOnLine 48, 1216–1223. https://doi.org/10.1016/j.ifacol.2015.09.692.

Walski, T.M., Brill Jr, E.D., Gessler, J., Goulter, I.C., Jeppson, R.M., Lansey, K., Lee, H.L., Liebman, J.C., Mays, L., Morgan, D.R., et al., 1987. Battle of the network models: Epilogue. Journal of Water Resources Planning and Management 113, 191–203. https://doi.org/10.1061/(ASCE)0733-9496(1987)113:2(191).

Wan, X., Kuhanestani, P.K., Farmani, R., Keedwell, E., 2022. Literature review of data analytics for leak detection in water distribution networks: A focus on pressure and flow smart sensors. Journal of Water Resources Planning and Management 148, 03122002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001597.

Wang, X., Guo, G., Liu, S., Wu, Y., Xu, X., Smith, K., 2020. Burst detection in district metering areas using deep learning method. Journal of Water Resources Planning and Management 146, 04020031. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001223.

Wang, X., Li, J., Liu, S., Yu, X., Ma, Z., 2022. Multiple leakage detection and isolation in district metering areas using a multistage approach. Journal of Water Resources Planning and Management 148, 04022021. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001558.

Wang, Z., Oates, T., 2015. Imaging time-series to improve classification and imputation, in: Twenty-Fourth International Joint Conference on Artificial Intelligence. https://www.ijcai.org/Proceedings/15/Papers/553.pdf.

Water Authorities Association, 1980. Water Authorities Association: Leakage Control Policy and Practice. Technical Report 26. https://dwi-content.s3.eu-west-2.amazonaws.com/wp-content/uploads/2020/10/27105844/dwi0190.pdf.

Wu, Y., Liu, S., Wu, X., Liu, Y., Guan, Y., 2016. Burst detection in district metering areas using a data driven clustering algorithm. Water research 100, 28–37. https://doi.org/10.1016/j.watres.2016.05.016.

Wu, Z.Y., Sage, P., Turtle, D., 2010. Pressure-dependent leak detection model and its application to a district water system. Journal of Water Resources Planning and Management 136, 116–128. https://doi.org/10.1061/(ASCE)0733-9496(2010)136:1(116).

Xu, Q., Liu, R., Chen, Q., Li, R., 2014. Review on water leakage control in distribution networks and the associated environmental benefits. J. Environ. Sci. 26, 955–961. https://doi.org/10.1016/S1001-0742(13)60569-0.

Ye, G., Fenner, R., 2014. Weighted Least Squares with Expectation-Maximization Algorithm for Burst Detection in U . K . Water Distribution Systems. Journal of Water Resources Planning and Management 140, 417–424. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000344.

Ye, G., Fenner, R.A., 2011. Kalman filtering of hydraulic measurements for burst detection in water distribution systems. Journal of pipeline systems engineering and practice 2, 14–22. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000070.

Zhang, H., Huang, T., Cao, M., He, W., 2009. Study on real-time detection of pipe bursts with simulation and management system on water distribution networks, in: ICPTT 2009: Advances and Experiences with Pipelines and Trenchless Technology for Water, Sewer, Gas, and Oil Applications, pp. 217–223. https://doi.org/10.1061/41073(361)23.

Zhang, H., Wang, L., 2011. Leak detection in water distribution systems using Bayesian theory and Fisher's law. Transactions of Tianjin University 17, 181–186. https://doi.org/10.1007/s12209-011-1594-4.

Zhang, Q., Wu, Z.Y., Zhao, M., Qi, J., Huang, Y., Zhao, H., 2016. Leakage zone identification in large-scale water distribution systems using multiclass support vector machines. Journal of Water Resources Planning and Management 142, 04016042. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000661.

Zhang, W., Zhao, D., Wang, X., 2013. Agglomerative clustering via maximum incremental path integral. Pattern Recognition 46, 3056–3065. https://doi.org/10.1016/j.patcog.2013.04.013.

Zhou, X., Tang, Z., Xu, W., Meng, F., Chu, X., Xin, K., Fu, G., 2019. Deep learning identifies accurate burst locations in water distribution networks. Water research 166, 115058. https://doi.org/10.1016/j.watres.2019.115058.