Leakage detection and localization in water distribution systems: A model invalidation approach

Stelios G. Vrachimis [a,b,*], Stelios Timotheou [a,b], Demetrios G. Eliades [a], Marios M. Polycarpou [a,b]

[a] *KIOS Research and Innovation Center of Excellence, University of Cyprus, 1 Panepistimiou Avenue, 2109 Aglantzia, Nicosia, Cyprus*
[b] *Department of Electrical and Computer Engineering, University of Cyprus, 1 Panepistimiou Avenue, 2109 Aglantzia, Nicosia, Cyprus*

A B S T R A C T

Model-based methodologies can assist in addressing the challenging problem of leakage detection and localization in water distribution systems. However, this is not trivial due to inherent non-linearities and parametric uncertainties. Most importantly, due to the small number of available sensor measurements compared to the number of system states, the inverse problem for estimating leakages is highly under-determined. In this work, the utilization of *a priori* available information about the system is proposed to formulate a hydraulic model of the system in its non-linear form in which uncertainties are modeled by intervals defined by a lower and upper bound. A novel optimization-based methodology then utilizes pressure and flow measurements to perform leakage detection through model-invalidation. A modification of the optimization algorithm is activated in the case of a detection to refine possible leak locations and retain only the ones that can be explained by the interval model and available measurements from multiple time-steps. The proposed methodology is demonstrated on a benchmark network and evaluated using a leakage diagnosis benchmark dataset.

1. Introduction

Water scarcity due to increasing urban population and climate change is urging water utilities to reduce the losses in Water Distribution Systems (WDS), which worldwide are estimated on average to be 30% of the provided water. An undesirable consequence of water loss, besides the loss of this precious resource, is the increase of service costs and degradation of service quality. The main reason behind water losses are leakages from broken pipes or pipe joints as a result of aging and deteriorating WDS infrastructures (Puust, Kapelan, Savic, & Koppel, 2010).

Recent developments in information infrastructures and hydraulic sensor technology have enabled water companies to supplement these large scale civil engineering systems with appropriate sensors, actuators and a communication network; it is thus feasible to use systems engineering ideas to monitor the system in real-time and to develop methodologies for detecting and providing approximate locations of leaks (Mareels et al., 2005). Leakage diagnosis in water distribution systems has attracted a great deal of attention from both practitioners and researchers over the past years (Chan, Chin, & Zhong, 2018), while it is also the topic of the recent *Battle of the Leakage Detection and Isolation Methods* (BattLeDIM) (Vrachimis et al., 2020). It can be separated into *leakage detection* methodologies, which focus on identifying the existence of a leakage in the network, and *leakage localization*

which aims to provide an approximate location of leakages, given the available measurements.

The challenge in leakage detection and localization in WDS is the fact that these large scale systems are associated with high uncertainty in modeling of network parameters, such as the pipe roughness co-efficients, and most importantly, high uncertainty in consumer water demands which are modeled as an uncontrolled input to the system. Moreover, typically only a small number of sensors are installed in the networks compared to the number of unknown states, especially in residential areas where the majority of leakages occur, resulting in a non-observable system when performing state estimation. Model-based fault-diagnosis methodologies face challenges due to modeling uncertainty and lack of observability, while data-driven methodologies may not be effective due to insufficient data. In this respect, the leakage detection and localization problem is rather challenging and it requires the design of novel methodologies to specifically address it.

1.1. Model-based leakage diagnosis

The term *model-based leakage diagnosis* is used to describe methodologies that utilize a model of the WDS (also referred to as numerical model) to estimate the steady-state hydraulic conditions in the network. This steady-state model neglects transients (Colombo, Lee, & Karney,

2009) and it can be obtained using known physical properties of the system. As a result, it is the most widely used model in water distribution systems analysis (Rossman, 2000). Typically, model-based methods assume the existence of a larger number of pressure sensors than flow sensors in WDS because they are cheaper and easier to install and maintain (Pérez et al., 2011). However, WDS are large-scale systems and the number of sensors is still limited compared to the system size. This is why methodologies for optimal placement of pressure sensors based on the available model are used to facilitate leakage diagnosis (Casillas, Puig, Garza-Castañón, & Rosich, 2013; Cugueró-Escofet, Puig, & Quevedo, 2017; Farley, Mounce, & Boxall, 2010).

The operating principle behind model-based leakage detection, as first suggested by Pudar and Liggett (1992), is to find discrepancies of measurements to their estimates obtained by the network model, which would indicate the existence of a leakage. For localization, some approaches try to relate acquired measurements with the simulated output from many leakage simulation scenarios on different locations of the network (Farley et al., 2010; Goulet, Coutu, & Smith, 2013). The geographical mapping of each model component can then be used to signal the probability of a zone to contain a leakage (Perez et al., 2014). Researchers have also used pressure residual analysis, by creating a system pressure sensitivity matrix, to identify the location of leaks with the assumption of a single leakage occurring in the system (Cugueró-Escofet et al., 2017; Pérez et al., 2011).

1.2. Demand and parameter uncertainty

The aforementioned methodologies consider the availability of consumer demand estimates in the network model. The estimates, or pseudo-measurements, of these demands can be derived by utilizing historical billing data or data from smart meters. These data are essentially low resolution measurements which can only be used to compute an average base demand at nodes. The variation of base demands at different times of the day can be approximated by multiplying them with a pattern derived from the measured system inflow (Sanz, Pérez, Kapelan, & Savic, 2016). However, the resulting demand pseudo-measurements may contain significant errors with no information on their statistical characterization, which is why some researchers have argued that the most accurate depiction of demand pseudo-measurement error is in the form of error bounds (Andersen, Powell, & Marsh, 2001). In practice, consumer demands are the greatest source of uncertainty when using model-based methodologies for leakage diagnosis (Goulet et al., 2013). This is because demand variations can cause significant pressure variations, which may either hide the effect of a leakage (unless the leak size is large relative to the system inflow), or cause false positives (Moors, Scholten, van der Hoek, & den Besten, 2018).

The uncertainty in water demands makes *demand calibration* using sensor measurements an important component of model-based leakage diagnosis. Demand calibration refers to the use of available historical sensor measurements and the system model to approximate using optimization based methodologies the demand magnitude and behavior at various nodes of the network. It can be viewed as the inverse problem to hydraulic-state estimation (Kun, Tian-Yu, Jun-Hui, & Jin-Song, 2015). Similarly to hydraulic-state estimation in which the scarcity of measurements does not guarantee observability (Díaz, González, & Mínguez, 2016), the inverse problem of demand calibration is ill-posed, and has attracted significant attention by researchers and practitioners (Cheng & He, 2011; Savic, Kapelan, & Jonkergouw, 2009). Typically, the under-determined problem of demand calibration is dealt by clustering several demand node parameters to make the problem over-determined. This produces an accurate estimate of the grouped node demand, however the individual node estimates may have significant errors (Cheng & He, 2011). The accurate estimation of group demands was exploited in Sanz et al. (2016) where the authors propose a leakage-detection and localization approach coupled with

a calibration methodology that identifies geographically distributed parameters. The geographical distribution of the calibrated parameters enables localization, as abrupt changes to these parameters are associated with a specific zone in the network. Leakages are in essence an unknown or unaccounted demand of the system and this is why some research approaches have considered leakage detection as part of calibration (Wu, Sage, & Turtle, 2010). However, the joint demand and leakage estimation problem increases the number of tuning parameters and the heuristic optimization techniques used to solve this problem may not produce reliable results (Do, Simpson, Deuerlein, & Piller, 2016; Ribeiro, Sousa, Marques, & Simões, 2015).

In addition to demand uncertainty, model parameters such as pipe roughness coefficients, introduce additional uncertainty into the calibration and leakage diagnosis problems. Due to the relatively small number of measurements, demand and parameter calibration may be jointly performed (Kang & Lansey, 2011). An estimate of pipe roughness parameters can be calculated using the pipe specifications given by the water utility. Since these estimates contain errors more accurately depicted by error bounds, a recent approach considers the bounded model parameter and demand uncertainties to calculate bounds on state estimates (Vrachimis, Timotheou, Eliades, & Polycarpou, 2018).

1.3. Dealing with uncertainty for fault-diagnosis

In the fault diagnosis literature, modeling uncertainty is a key component in calculating the threshold on the residual; i.e., the difference between the measured and estimated states. When the threshold is violated by the residual, the occurrence of a fault is declared (Rosa, Silvestre, Shamma, & Athans, 2010). This concept is applied to leakage detection in Blesa, Puig, Saludes, and Vento (2010) where a Linear Parameter Varying model is used to compute a set-valued state estimate. The method allows the creation of zonotopes, which are compared with the pressure residuals and are used as thresholds for detection. However, the uncertainty on water demands is not considered, while the use of a linearized model may introduce additional modeling errors. In Vrachimis, Eliades, and Polycarpou (2018) the authors investigated the leakage detection problem by comparing sensor measurements to thresholds derived from hydraulic interval-state estimation on a nonlinear model using demand pseudo-measurements. Although the particular algorithm is designed to achieve zero false-positives, in practice it has low leakage detection rate on a related benchmark data set (Vrachimis, Kyriakou, Eliades, & Polycarpou, 2018). The reason is that the algorithm obtains hydraulic state bounds without incorporating into the model available information such as state measurements and aggregate demand bounds.

Recent works on fault diagnosis have proposed the concept of set-membership model invalidation, in which data are used to create a set-valued non-faulty model of the system, and check its validity against future data using an online optimization formulation (Harirchi & Ozay, 2018). This formulation keeps implicit constraints to represent sets and the problem of fault-detection is reduced to the feasibility of a convex optimization problem (Ozay, Sznaier, & Lagoa, 2014), which can be checked using state-of-the-art solvers (Gurobi Optimization, 2019). This method guarantees that there will be no false alarms, as long as the model of the non-faulty system remains valid. Moreover, there is no need to compute a decision threshold, while the flexibility in which constraints can be added into the formulation facilitates the utilization of additional system information to increase fault detectability.

1.4. Summary and contributions of this work

In this work, the utilization of demand and model parameter bounds is proposed to formulate a "healthy" interval-model of the system in its nonlinear form, which also incorporates current measurements and estimated demands of node clusters derived from a demand calibration process. In summary, the *a priori* available information used in this work is:

- Available network model.
- Bounds on model parameters (e.g. pipe roughness coefficients).
- Bounds on consumer water demands.
- Bounded water demand estimates of consumer groups, derived by a demand calibration process.

A novel optimization-based methodology for leakage detection through model invalidation is then proposed which detects inconsistencies between the healthy interval-model and sensor measurements. It is shown that the leakage detection problem in WDS can be reduced to iteratively examining the feasibility of a Linear Program (LP). In case of a leakage detection, a novel localization algorithm is activated to refine the possible leak locations and retain the ones that satisfy the constraints imposed by the interval-model and the available measurements from multiple time-steps. The overall contributions of this work are summarized as follows:

- The extension of the interval-model formulation proposed in Vrachimis, Timotheou, et al. (2018) to include current measurements and estimated node group demands derived from a demand calibration process.
- The introduction of a novel methodology for leakage detection through interval-model invalidation.
- The development of an optimization-based methodology for leakage localization using the proposed interval-model.
- The definition of a novel localization index for evaluating leakage localization methodologies which prioritize possible leakage nodes.

The novel approach for leakage detection and localization in this work offers some advantages compared to other approaches in the literature. The use of a non-linear model which includes the uncertainties in the form of intervals avoids the linearization errors exhibited by methodologies which use a pressure sensitivity matrix to perform leakage detection (Pérez et al., 2011). Moreover, there is no need to calculate an optimal threshold for detection, as this is implicitly defined by the bounds on the uncertain parameters. Previously proposed set-based techniques do not consider all the uncertainties in the model while they also use a linearized model (Blesa et al., 2010). Techniques which do not linearize the system model require performing multiple simulations of potential leakages and attempt to minimize the error between prediction and measurements given the model uncertainty (Goulet et al., 2013). Techniques which combine demand calibration in the leakage diagnosis problem, require the continuous update of demand estimates (Sanz et al., 2016), however, demand estimation is an under-determined problem and these estimates may not be accurate. This work uses known bounds on these demands instead. Moreover, the demand estimates from node groups used by the methodology are derived from the solution of an over-determined system of equations. A drawback of this approach is that it may result in lower detection rates if the demand bounds are conservative.

A disadvantage of the proposed methodology is that the current design only considers a single leakage in the system. However, even methodologies which compare pressure residuals against multiple simulated faults, do not consider simultaneous faults in the system, but rather give candidate locations given the signature of individual faults (Perez et al., 2014). Another disadvantage is the computational complexity of the proposed algorithm which requires the solution of several optimization problems, whereas other optimization based approaches try to reduce the number of variables by considering only the leakages in the optimization problem (Wu et al., 2010). However, we have investigated the computational time of the methodology in Section 6.4 and we have concluded that it can be applied in real-time for real-life networks.

The remainder of this paper is organized as follows: In Section 2 the problem of formulating an interval-hydraulic-model using *a-priori* available information is described. Then, in Section 3, the methodology

for leakage detection using interval-model invalidation is introduced. A novel leakage localization procedure is then described in Section 4 and the Localization Priority Index for evaluating this procedure is proposed in Section 5. The proposed methodologies are demonstrated using two illustrative examples in Section 6, and moreover they are extensively tested using a leakage benchmark dataset in the same Section. Finally, conclusions are drawn in Section 7.

Notation. Matrices are denoted with capital letters, vectors with lower bold letters, and scalars by italic letters. Sets and graphs are denoted by calligraphic letters. Uncertain parameters are represented by a continuous interval of values defined by a lower and upper bound. Intervals are accompanied by a tilde and defined as follows: $\tilde{\boldsymbol{v}} = [\boldsymbol{v}^l, \boldsymbol{v}^u]$ is a closed interval vector, where $\boldsymbol{v}^l$ is the lower bound vector and $\boldsymbol{v}^u$ is the upper bound vector, such that: $\tilde{\boldsymbol{v}} = \{\boldsymbol{v} \in \mathbb{R}^n : v_i^l \leq v_i \leq v_i^u, \forall\, i = \{1, \ldots, n\}\}$, and n is the size of the vector.

2. Problem formulation

The topology of a WDN is modeled by a directed graph denoted as $\mathcal{G} = (\mathcal{N}, \mathcal{L})$. Let $\mathcal{N} = \{1, \ldots, n_n\}$ be the set of all nodes, where $|\mathcal{N}| = n_n$ is the total number of nodes. These represent junctions of pipes, consumer water demand locations, reservoirs and tanks. Moreover, let $\mathcal{N}_t \subset \mathcal{N}$ represent the subset of reservoir and tank nodes. The hydraulic state associated with nodes is the *hydraulic head*, indicated by h_j, $j \in \mathcal{N}$. The hydraulic head consists of a component analogous to the pressure p_j at node j, and of the node elevation z_j in respect to a geodesic reference. Each node j is also associated with a water consumer demand at the node location, denoted by $q_{\text{ext},j}$. Water demands drive the dynamics of a WDN and are typically an unknown input to the system. Leakages in this work are modeled at network nodes, such that $q_{\text{leak},j}$ represents a leakage occurring at node j. Let $\mathcal{L} = \{1, \ldots, n_l\}$ be the set of links, where $|\mathcal{L}| = n_l$ is the total number of links. These represent network pipes, water pumps and pipe valves. The hydraulic quantity associated with a link $i \in \mathcal{L}$ is the *water flow*, indicated by q_i. The hydraulic equations which describe the behavior of the system are the conservation of energy equations and the conservation of mass equations.

Energy in WDS is associated with the head at nodes and when water flows through a network link i which connects two nodes, a flow dependent, typically nonlinear head function $f_i(q_i)$ describes the change in head. The *energy equations* for all the network links, can be written as follows:

$$\mathbf{f}(\boldsymbol{q}) + B\boldsymbol{h} = \mathbf{0}, \tag{1}$$

where $B \in \mathbb{R}^{n_l \times n_n}$ is the incidence flow matrix, indicating the connectivity of nodes with links such that element $B_{ij} = +1$ if the conventional direction of link i enters node j; element $B_{ij} = -1$ if the conventional direction of link i leaves from node j; otherwise $B_{ij} = 0$. Moreover, the nonlinear mapping $\mathbf{f}(\boldsymbol{q}) : \mathbb{R}^{n_l} \mapsto \mathbb{R}^{n_l}$ represents the head-loss at links. Each function $f_i(q_i)$ represents the head-loss at link i, which is a measure of the energy dissipated due to friction of water flowing through the link. In the case of simple pipes, head-loss depends on the water flow through the pipe but also on pipe parameters, such as pipe length, diameter and pipe roughness coefficient. In this work, the Hazen–Williams (H–W) formula (Boulos, Lansey, & Karney, 2006) is used to calculate head-loss in pipes, in which the effect of pipe parameters is aggregated in the H–W resistance coefficient r_i of each pipe. The head-loss across pipe $i \in \mathcal{L}_p$ is then calculated using the H–W formula as follows:

$$f_i(q_i) = r_i \left| q_i \right|^{(\nu-1)} q_i\,, \tag{2}$$

where ν is a constant exponent associated with the H–W formula and q_i is the water flow in pipe i. The head-loss function $f_i(q_i)$ can be adjusted to represent other system components modeled at links, such as pumps and valves (Vrachimis, Timotheou, et al., 2018).

The conservation of mass law for each node $j \in \mathcal{N}$ dictates that the sum of branch water flows from pipes incident to a node j must be equal to the node's external water demand plus any leakages associated with that node. The *conservation of mass equations*, considering all the nodes of the network, can be written using the incidence flow matrix as follows:

$$B^{\top} q = q_{\text{ext}} + q_{\text{leak}}(h) \tag{3}$$

The underlying modeling assumptions in (3) are:

1. The hydraulic heads at reservoir and tank nodes are measured and known with uncertainty which is bounded by known bounds, such that:

   $$\tilde{h}_j(k) \in \left[h_j^l(k), h_j^u(k) \right], j \in \mathcal{N}_t, \tag{4}$$

   where $h_j^l(k)$ and $h_j^u(k)$ are known lower and upper bounds respectively for the head at node j.

2. The network is operating in pressure sufficient conditions, such that the demand at each node is independent of the node pressure.

3. Leakage magnitude is pressure dependent. In this work, the pressure dependent leakage function is given by:

   $$q_{\text{leak},j}(h_j) = c_j(h_j - z_j)^{\alpha} \tag{5}$$

   where $c_j \geq 0$ is the leak emitter coefficient, $z_j \geq 0$ is the elevation at node j and $\alpha > 0$ is the leak emitter exponent.

The complete hydraulic-state of a WDS, comprised of the water flows in pipes and hydraulic heads at nodes, is indicated by $x = [q^{\top} \ h^{\top}]^{\top} \in \mathbb{R}^{n_l+n_n}$. The state can be estimated using the hydraulic equations of (1) and (3) and available measurements from the system. In the case of a typical District Metered Area (DMA) of a WDS, flow $\hat{q}_j(k), j \in \mathcal{L}_{in} \subset \mathcal{L}$ measurements are available at the inlets at each discrete measurement time-step k, where $|\mathcal{L}_{in}| = n_i$ is the number of inlets. Additionally, a number of pressure sensors are installed inside the DMA, giving pressure measurements indicated by $\hat{p}_i(k), i \in \mathcal{N}_s \subset \mathcal{N}$, where $|\mathcal{N}_s| = n_s$ is the number of sensors. Sensor measurements include noise, which is assumed bounded, such that the true value of measured state relates to the measurements as follows:

$$C_q q(k) = \hat{q}(k) + \tilde{v}_q(k) \tag{6a}$$
$$C_h h(k) = \hat{p}(k) + z + \tilde{v}_p(k) \tag{6b}$$

where $C_q \in \{0,1\}^{n_i \times n_l}$ and $C_h \in \{0,1\}^{n_s \times n_l}$ are matrices used to identify measured flow and head states respectively, and $\tilde{v}_q, \tilde{v}_p$ are the noise bounds for flow and pressure devices respectively.

WDS have typically a small number of sensors installed, i.e. $(n_s + n_i) \ll n_n$, which makes the system of equations defined by (1) and (3) under-determined. According to the analysis in Díaz, Mínguez, and González (2017), at least n_n sensors in an observable configuration are needed to guarantee the topological observability of the system. A possible observable configuration according to Díaz et al. (2017), is when water demands are measured at all nodes and at least one additional head measurement is available. For this reason demand estimates, called *pseudo-measurements*, are used to complement missing measurements and make the problem over-determined by utilizing in total more than n_n measurements (Hutton, Kapelan, Vamvakeridou-Lyroudia, & Savic, 2014). Demand estimates can be derived using off-line information such as population densities, building areas and the consumption patterns of typical consumers (Blokker, Vreeburg, & van Dijk, 2010). These estimates are then further improved by combining them with available low-resolution measurements, such as quarterly billing data. They can also be a function of the measured DMA inflow, as shown in Sanz et al. (2016). Since it is impossible to obtain exact estimates, a typical depiction of demand pseudo-measurements is in the

form of error bounds (Andersen et al., 2001), given for each discrete time-step k as follows:

$$\tilde{q}_{\text{ext}}(k) = \left[q_{\text{ext}}^l(k), \ q_{\text{ext}}^u(k) \right]. \tag{7}$$

Demand calibration is the use of available historical sensor measurements, hydraulic equations and optimization methodologies to approximate the demand magnitude and behavior at each node as close as possible. Due to the small number of measurements, nodal demands with similar user characteristics are grouped (i.e., aggregated) to make the problem of demand calibration over-determined (Kun et al., 2015). The procedure results in a demand estimate of n_g node groups, where $n_g \leq n_s$, which is less uncertain related to demands pseudo-measurements at nodes. The group estimate is written as:

$$M q_{\text{ext}}(k) = g(k) \tag{8}$$

where $M \in \{0,1\}^{n_g \times n_n}$ identifies demand groups, $g \in \mathbb{R}^{n_g}$ is the uncertain group demand estimate. Notice that overlapping groups are allowed. The uncertainty on group estimates is considered by defining the bounds on these estimates, such that $\tilde{g} \in [g^l, g^u]$. Moreover, considering a fault-free situation when $q_{\text{leak},j} = 0, \forall j \in \mathcal{N}$, (3) can be used to re-write (8) in terms of the system flow-states, as follows:

$$M B^{\top} q(k) = \tilde{g}(k) \tag{9}$$

Uncertainty in model parameters is modeled as an uncertain head-loss $\tilde{f}(q)$, where the tilde indicates that this function contains uncertain parameters. Specifically for pipes, uncertainty is included in the uncertain H–W coefficients r_i. The pipe head-loss function for a pipe i, given by (2), when this contains uncertainty becomes:

$$\tilde{f}_i(q_i) = \tilde{r}_i |q_i|^{\nu-1} q_i, \quad i \in \mathcal{L}_p. \tag{10}$$

Model parameter bounds are assumed known and constant, because these parameters vary slowly over time. Note that this approach of modeling uncertainties can be applied to the head-loss function of any element of the system modeled by a link, as explained in Vrachimis, Timotheou, et al. (2018).

3. Leakage detection by interval-model invalidation

The bounded uncertainties on demand and parameter estimates for every time-step k are derived from historical data of the system in healthy operation. During healthy operation, it is assumed that $q_{\text{leak},j} = 0, \forall j \in \mathcal{N}$ and measurements are included in the feasible solution set when solving the hydraulic equations considering bounded uncertainties and sensor noise. The fault-free interval-model for each discrete time-step k, is given by:

$$\tilde{f}(q(k)) + B h(k) = 0 \tag{11a}$$
$$B^{\top} q(k) = \tilde{q}_{\text{ext}}(k) \tag{11b}$$
$$M B^{\top} q(k) = \tilde{g}(k) \tag{11c}$$
$$C_q q(k) = \hat{q}(k) + \tilde{v}_q(k) \tag{11d}$$
$$C_h h(k) = \hat{p}(k) + z + \tilde{v}_p(k) \tag{11e}$$

The interval hydraulic model of (11) offers a framework for the design of an event detection methodology using model-invalidation. Unlike many set-membership methods that compute explicit set representations for the system states, a model-invalidation method uses an online optimization formulation that *keeps implicit constraints to represent sets*; it has been shown that model invalidation problems can be reduced to the feasibility of a linear optimization problem, which can be checked efficiently using state-of-the-art solvers (Harirchi & Ozay, 2018).

In previous work, the authors have proposed a way to deal with hydraulic models which contain uncertainties in the form of intervals, by converting the nonlinear model into a set of linear inequalities and then using an iterative optimization procedure to calculate tight
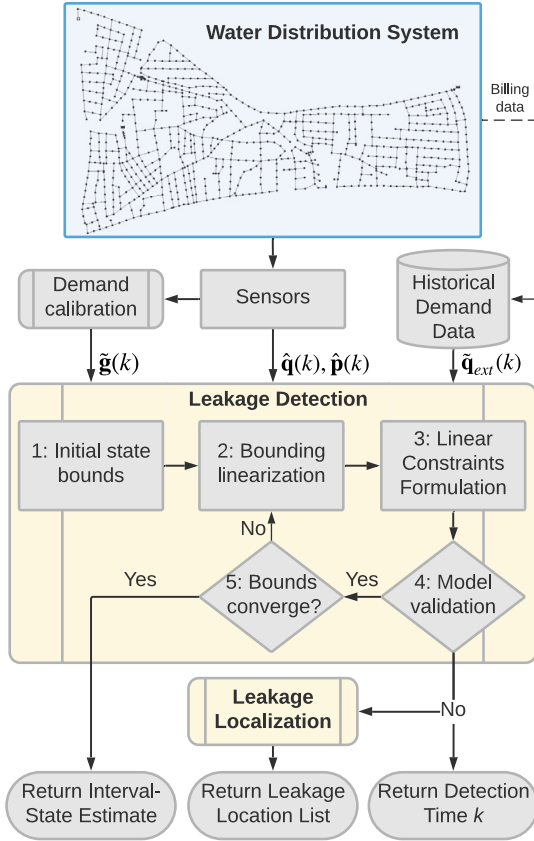
Fig. 1. Illustration of the proposed leakage detection procedure.

bounds on the system states (Vrachimis, Timotheou, et al., 2018). However, the methodology could not deal with inconsistencies between the model and sensor measurements occurring during the presence of a fault; consequently, when applying the aforementioned methodology for leakage detection in Vrachimis, Eliades, and Polycarpou (2018), the state bounds are calculated using offline demand pseudo-measurements during healthy operation and then compared to online sensor measurements to detect the presence of a fault.

In this work, building on the previous work of Vrachimis, Timotheou, et al. (2018) and Vrachimis, Eliades, and Polycarpou (2018), a novel optimization procedure is designed which includes both offline and online measurements and incorporates additional constraints imposed by demand calibration. The proposed model-based methodology for leakage detection through model invalidation exhibits a much higher detection rate than the one described in Vrachimis, Eliades, and Polycarpou (2018), as shown in the results of Section 6.3. Moreover, during healthy operation, the optimization procedure yields tighter bounds on state variables.

The improved performance can be explained by the fact that using model-invalidation the linear constraints formulated during the leakage detection procedure are part of the state-set representation, along with the bounds on states; in comparison, the state-set in Vrachimis, Eliades, and Polycarpou (2018) is represented only by the bounds on states.

The proposed leakage detection methodology is formulated in a series of five steps which are illustrated in Fig. 1 and explained below.

3.1. Step 1: Initial bounds on state variables

The iterative leakage detection algorithm requires initial bounds on the state $\boldsymbol{x} = [\boldsymbol{q}^\top \; \boldsymbol{h}^\top]^\top$ as a starting point. This iterative procedure is applied to single time-step thus, for notational convenience, the use of the time-step notation k is omitted. The initial bounds on the

unknown head vector are chosen by considering physical properties of the network (Vrachimis, Timotheou, et al., 2018) and denoted by

$$\tilde{\boldsymbol{h}}^{(m)} = \left[\boldsymbol{h}^{l\,(m)}, \; \boldsymbol{h}^{u\,(m)} \right] \tag{12}$$

where m indicates the iteration counter which for the initial bounds $m = 0$. Note that reservoir and tank node heads are represented here by their known bounds $\tilde{h}_j(k), j \in \mathcal{N}_t$. The initial bounds on the flow state vector, denoted by

$$\tilde{\boldsymbol{q}}^{(m)} = \left[\boldsymbol{q}^{l\,(m)}, \boldsymbol{q}^{u\,(m)} \right], \tag{13}$$

are calculated using $\tilde{\boldsymbol{h}}^{(m)}$ in Eq. (11a) and solved with respect to flow states using interval arithmetic (Moore, Kearfott, & Cloud, 2009). The initial bounds are intentionally conservative in order to ensure the inclusion of the true system state.

3.2. Step 2: Bounding linearization

Bounding linearization is a procedure which aims at enclosing in a convex set the non-linear uncertain functions $\tilde{f}_i(q_i)$ given an interval for $\tilde{q}_i \in \left[q_i^l, q_i^u \right]$, as described in Vrachimis, Timotheou, et al. (2018). One way to do this is to enclose the uncertain function $\tilde{f}_i(q_i)$ between two lines, the lower line $f_{L,i}^l(q_i)$ and the upper line $f_{L,i}^u(q_i)$, for different values of the flow interval $\tilde{q}_i$ as follows:

$$f_{L,i}^l(q_i) \le \tilde{f}_i(q_i) \le f_{L,i}^u(q_i), \; \forall q_i \in \left[q_i^l, q_i^u \right], \; i \in \mathcal{L} \tag{14}$$

with the two lines defined as $f_{L,i}^l(q_i) = \lambda_i^l q_i + \beta_i^l$ and $f_{L,i}^u(q_i) = \lambda_i^u q_i + \beta_i^u$ The goal is to define the line parameters such as to minimize the area of the resulting convex set. In the case of pipes, $\tilde{f}_i(q_i)$ is *inclusion isotonic* (meaning that $A \subseteq B \implies f(A) \subseteq f(B)$, where A and B are arbitrary sets of continuous numbers) (Moore et al., 2009), and the minimization of the convex set area can be formulated and solved using the Karush–Kuhn–Tucker optimality conditions. Finally, by noting that the function $\tilde{f}_i(q_i)$ depends on only one $q_i(k)$, and given that this is *inclusion isotonic*, it can be guaranteed that at every subsequent iteration the convex set defined will be smaller or equal than the previous one, resulting in a bounded linearization error (Vrachimis, Timotheou, et al., 2018).

The nonlinear uncertain functions contained in $\tilde{\boldsymbol{f}}(\boldsymbol{q}(k))$ of (11a) are then eliminated through the use of the auxiliary variables ζ_i at each diagonal position i. The convex set for each auxiliary variable ζ_i is defined by the flow bounds $\tilde{\boldsymbol{q}}^{(m)}$, thus it changes at each iteration m. The resulting linear inequality derived from bounding linearization, without denoting the iteration number m, is given by:

$$\lambda_i^l q_i + \beta_i^l \le \zeta_i \le \lambda_i^u q_i + \beta_i^u \tag{15}$$

where $\lambda_i^l, \beta_i^l, \lambda_i^u$ and β_i^u are line parameters defined during bounding linearization. This procedure introduces uncertainty into the problem, as the intervals in (13) and the inequalities in (15) define a convex superset of the set defined by the function $\tilde{f}_i(q_i)$.

3.3. Step 3: Formulation of a relaxed problem

The bounding linearization procedure in Section 3.2 enables us to rewrite problem (11) as a set of linear inequalities and eliminates the interval parameters through the use of their lower and upper bounds:

$$\boldsymbol{\zeta} + \boldsymbol{B}\boldsymbol{h} = 0 \tag{16a}$$

$$\Lambda^l \boldsymbol{q} + \boldsymbol{b}^l \le \boldsymbol{\zeta} \le \Lambda^u \boldsymbol{q} + \boldsymbol{b}^u \tag{16b}$$

$$\boldsymbol{q}_{\text{ext}}^l \le \boldsymbol{B}^\top \boldsymbol{q} \le \boldsymbol{q}_{\text{ext}}^u \tag{16c}$$

$$\boldsymbol{g}^l \le \boldsymbol{M} \boldsymbol{B}^\top \boldsymbol{q} \le \boldsymbol{g}^u \tag{16d}$$

$$\boldsymbol{v}_q^l \le \boldsymbol{C}_q \boldsymbol{q} - \hat{\boldsymbol{q}} \le \boldsymbol{v}_q^u \tag{16e}$$

$$\boldsymbol{v}_p^l \le \boldsymbol{C}_h \boldsymbol{h} - \hat{\boldsymbol{p}} - \boldsymbol{z} \le \boldsymbol{v}_p^u \tag{16f}$$

$$\boldsymbol{q}^l \le \boldsymbol{q} \le \boldsymbol{q}^u \tag{16g}$$

$$\boldsymbol{h}^l \le \boldsymbol{h} \le \boldsymbol{h}^u \tag{16h}$$

The linear constraints in (16) can be used to formulate two Linear Programs (LP) for obtaining lower (LPmin) and upper (LPmax) bounds on each state variable x_i, as follows:

LPmin :

$$\left\{ \begin{array}{cc} \min_{\boldsymbol{x}} & x_i \\ s.t. & (16) \end{array} \right\}$$

LPmax :

$$\left\{ \begin{array}{cc} \max_{\boldsymbol{x}} & x_i \\ s.t. & (16) \end{array} \right\}$$

Problems **LPmin** and **LPmax** are a relaxed version of the original problem in (11), due to the bounding linearization of non-linearities. However, it is ensured that the feasible set which satisfies the original problem is a subset of the feasible set satisfying the relaxed problem.

3.4. Step 4: Model validation

The feasible set for the states $\boldsymbol{x}$ which satisfies the relaxed problem is found by solving a total of $2(n_l + n_n)$ LPs, with each problem calculating either the lower (**LPmin**) or upper (**LPmax**) bound of an individual state variable x_i. Assuming no faults in the system and valid uncertainty bounds, the constraints of (16) define a feasible set which contains the true system state and problems **LPmin** and **LPmax** can be solved for all state variables x_i. However, if a fault exists in the system, its effect will be present in the measurements and a feasible set may not exist, unless the fault effect is "concealed" by the combined demand, parameter and bounding linearization uncertainty. The decision that a feasible set does not exist is taken when any of the $2(n_l + n_n)$ LPs gives an infeasible solution. Problems **LPmin** and **LPmax** are in linear form, i.e., a linear objective function with linear inequalities and these type of optimization problems can be solved very effectively using Linear Programming. A large range of solvers can be used to solve LP problems. In this work the Gurobi mathematical programming solver was used (Gurobi Optimization, 2019).

3.5. Step 5: Iterative model validation

To find the smallest possible interval $\tilde{x} = \begin{bmatrix} x^l, & x^u \end{bmatrix}$ that satisfies (11), an iterative method is used which reduces the bounding linearization uncertainty. At each iteration m, the tightest possible state vector bounds $\tilde{x}^{(m)}$ given current constraints are calculated. At the next iteration $m + 1$, these bounds are used in the bounding linearization procedure to calculate new constraints on the variable $\boldsymbol{\zeta}$. The new constraints are then in turn used to calculate new state vector bounds $\tilde{x}^{(m+1)}$. As long as the linear program of Step 4 is able to be solved, the model is validated and iterations continue. The iterations stop when the bounds on the state vector remain relatively unchanged, in which case no fault is detected given time-step k measurements. The algorithm is considered to have converged when the L^1-norm of the change in bounds at iteration m, denoted by $e^{(m)}$ and defined in (17), is smaller than a small number ϵ.

$$e^{(m)} \triangleq \left| \left(\boldsymbol{x}^{u(m)} - \boldsymbol{x}^{l(m)} \right) - \left(\boldsymbol{x}^{u(m-1)} - \boldsymbol{x}^{l(m-1)} \right) \right|_1. \tag{17}$$

The complete iterative algorithm for model-validation for each time-step k containing Steps 1–5 is outlined in Algorithm 1.

4. Leakage localization using an interval-model

When a leakage is detected using Algorithm 1, the interval-model is re-formulated to consider the leakage localization problem. The interval-model of the system for each time-step k, considering the effect of leakages, is given by:

$$\tilde{\boldsymbol{f}}(\boldsymbol{q}(k)) + B\boldsymbol{h}(k) = 0 \tag{18a}$$

Algorithm 1 Leakage detection algorithm

Input: time-step k, measurements $\hat{\boldsymbol{q}}(k)$, $\hat{\boldsymbol{p}}(k)$, demand bounds $\tilde{\boldsymbol{q}}_{\text{ext}}(k)$, $\tilde{\boldsymbol{g}}(k)$, parameter bounds $\tilde{r}$
Output: $\tilde{\boldsymbol{x}}(k)$, Detection
begin
1: Calculate initial state bounds $\tilde{\boldsymbol{x}}^{(0)}$
2: Iteration: $m = 0$
3: **while** $e^{(m)} > \epsilon$ **do**
4:     Bounding linearization using $\tilde{\boldsymbol{x}}^{(m)}$
5:     Formulate problems **LPmin** and **LPmax**
6:     **for** $i = 1$ **to** $n_l + n_n$ **do**
7:         Solve **LPmin** and **LPmax** to find x_i^l and x_i^u
8:         **if** infeasible **then**
9:             Detection = True
10:            Go to **Algorithm 2** (Leakage Localization)
11:        **else**
12:            $\tilde{x}_i^{(m)} = \begin{bmatrix} x_i^l, & x_i^u \end{bmatrix}$
13:        **end if**
14:    **end for**
15:    $m = m + 1$
16: **end while**
17: $\tilde{\boldsymbol{x}}(k) = \tilde{\boldsymbol{x}}^{(m)}$
return

$$B^\top \boldsymbol{q}(k) = \tilde{\boldsymbol{q}}_{\text{ext}}(k) + \boldsymbol{q}_{\text{leak}}(\boldsymbol{h}) \tag{18b}$$

$$M B^\top \boldsymbol{q}(k) = \tilde{\boldsymbol{g}}(k) + M \boldsymbol{q}_{\text{leak}}(\boldsymbol{h}) \tag{18c}$$

$$C_q \boldsymbol{q}(k) = \hat{\boldsymbol{q}}(k) + \tilde{\boldsymbol{v}}_q(k) \tag{18d}$$

$$C_h \boldsymbol{h}(k) = \hat{\boldsymbol{p}}(k) + \boldsymbol{z} + \tilde{\boldsymbol{v}}_p(k) \tag{18e}$$

The leakage localization problem reduces to the determination of a feasible set for the augmented vector of state variables $\boldsymbol{\chi} = \begin{bmatrix} \boldsymbol{x}^\top \boldsymbol{c}^\top \end{bmatrix}^\top$ given the interval model in (18) and assuming that only one leakage is present in the system. The new state variables $\boldsymbol{c}$ are the unknown leak emitter coefficients in the nonlinear functions $\boldsymbol{q}_{\text{leak}}(\boldsymbol{h})$ given in (5), and the feasible set for these variables indicates possible leakage locations. The leakage localization procedure is divided into the five steps described below.

4.1. Step 1: Initial bounds on states and leakage functions

Initial bounds on the states $\boldsymbol{x}$ are calculated for iteration $m = 0$ as described in Section 3.1. The initial bounds on the states $\boldsymbol{c}$ are calculated implicitly, by defining bounds on the leakage magnitudes $q_{\text{leak},j} = \left\| q_{\text{leak},j}(h_j) \right\| \forall j \in \mathcal{N}$. The existence of a leak at any node is assumed possible, thus, all nodes have a non-zero upper bound for the leakage flow equal to $q_{\text{leak},j}^u = q_{\text{leak}}^{max} \forall j \in \mathcal{N}$. Similarly, the lower bound of all leakages is zero, such that leakage magnitude bounds become $\tilde{q}_{\text{leak}} = \begin{bmatrix} 0, q_{\text{leak}}^{max} \end{bmatrix}$. The maximum leakage magnitude can be roughly determined by methodologies that monitor and detect abnormalities of the system inflow (Eliades & Polycarpou, 2012). Given leakage and state $\boldsymbol{x}$ bounds, bounds on the emitter coefficients $\boldsymbol{c}$ in (5) can be calculated using interval arithmetic, as follows:

$$\tilde{c}_j^{(m)} = \frac{\tilde{q}_{\text{leak},j}^{(m)}}{\left(\tilde{h}_j^{(m)} - z_j \right)^\alpha}, \, j \in \mathcal{N} \tag{19}$$

This work assumes pressure sufficient conditions in the network, thus $h_j > z_j$. The emitter exponent is assumed known and equal to $\alpha = 0.5$.

4.2. Step 2: Bounding linearization

The bounding linearization procedure is generalized to enclose in a convex set not only the uncertain head-loss functions $\tilde{f}_i(q_i)$ as

described in Section 3.2, but also the uncertain non-linear leakage functions $q_{\text{leak},j}(h_j)$. Each leakage function $q_{\text{leak},j}(h_j)$ is approximated through the use of the auxiliary scalar variable $q_{\text{leak},j}$ which corresponds to the leakage magnitude at node j. The additional linear inequality for each node j derived from bounding linearization is calculated using emitter coefficient bounds $\tilde{c}_j$ and head bounds $\tilde{h}_j$ and given by:

$$\lambda_{q,j}^l h_j + \beta_{q,j}^l \leq q_{\text{leak}_j} \leq \lambda_{q,j}^u h_j + \beta_{q,j}^u, \; j \in \mathcal{N} \tag{20}$$

where $\lambda_{q,j}^l$, $\beta_{q,j}^l$, $\lambda_{q,j}^u$ and $\beta_{q,j}^u$ are line parameters defined during bounding linearization.

4.3. Step 3: Formulation of the relaxed problem and calculation of leakage bounds

The elimination of nonlinearities through bounding linearization and the use of bounds enables the reformulation of Problem (18) as a set of linear inequalities, similar to Section 3.3. The decision variables will now include the auxiliary leakage variables q_{leak}, such that the complete decision variables vector will be given by

$$z = [x^\top \; q_{\text{leak}}^\top]^\top. \tag{21}$$

Here, however, the assumption of a single leakage existing in the system must be exploited, by considering that only one element $q_{\text{leak},j} j \in \mathcal{N}$ is non-zero, which should satisfy the constraint $q_{\text{leak},j}^l \leq q_{\text{leak},j} \leq q_{\text{leak},j}^u$. The set of linear inequalities are formulated as follows:

$$\zeta + Bh = 0 \tag{22a}$$

$$\Lambda_\zeta^l q + b_\zeta^l \leq \zeta \leq \Lambda_\zeta^u q + b_\zeta^u \tag{22b}$$

$$q_{\text{ext}}^l \leq B^\top q - q_{\text{leak}} \leq q_{\text{ext}}^u \tag{22c}$$

$$\Lambda_q^l h + b_q^l \leq q_{\text{leak}} \leq \Lambda_q^u h + b_q^u \tag{22d}$$

$$g^l \leq MB^\top q - Mq_{\text{leak}} \leq g^u \tag{22e}$$

$$v_q^l \leq C_q q - \hat{q} \leq v_q^u \tag{22f}$$

$$v_p^l \leq C_h h - \hat{p} - z \leq v_p^u \tag{22g}$$

$$q^l \leq q \leq q^u \tag{22h}$$

$$h^l \leq h \leq h^u \tag{22i}$$

$$q_{\text{leak},j}^l y_j \leq q_{\text{leak},j} \leq q_{\text{leak},j}^u y_j \tag{22j}$$

$$\sum y_j = 1, j \in \mathcal{N} \tag{22k}$$

$$y_j \in \{0,1\} \tag{22l}$$

where constraints (22j)–(22l) define the single leakage assumption. The linear constraints in (22) can be used to calculate a feasible set for the system states x using Linear Programming.

Because of the single leakage assumption, binary variables are introduced, which is undesirable considering the added computational complexity of solving a Mixed-Integer Linear Program (MILP), especially when dealing with large systems. In this work, the set defined in (22j)–(22l) is approximated by the convex hull of this set, which is defined in the following lemma:

Lemma 1. *The convex hull of the feasible set of q_{leak} given by (22j)–(22l), is given by the following linear constraints:*

$$\sum_{j \in \mathcal{N}_0^u} \frac{q_{\text{leak},j}}{q_{\text{leak},j}^u} \leq 1 \tag{23a}$$

$$q_{\text{leak},j} \geq 0 \; \forall j \in \mathcal{N} \tag{23b}$$

$$q_{\text{leak},j} = 0 \; \forall j \in \mathcal{N} - \mathcal{N}_0^u \tag{23c}$$

where $\mathcal{N}_0^u$ is the set of all nodes with a non-zero leakage upper bound, i.e. $\mathcal{N}_0^u = \{j : q_{\text{leak},j}^u > 0, \forall j \in \mathcal{N}\}$.

Proof of Lemma 1 is provided in the Appendix.

Two LPs for obtaining lower (**LPmin2**) and upper (**LPmax2**) bounds on each state variable $x_i \in z$ are then formulated, as follows:

LPmin2 :

$$\left\{ \begin{array}{ll} \min_z & x_i \\ s.t. & (22a) - (22i), \\ & (23) \end{array} \right\}$$

LPmax2 :

$$\left\{ \begin{array}{ll} \max_z & x_i \\ s.t. & (22a) - (22i), \\ & (23) \end{array} \right\}$$

Similar to problems **LPmin** and **LPmax** in the detection phase, **LPmin2** and **LPmax2**, are a relaxed version of the original problem in (18), due to the bounding linearization of non-linearities, and moreover, the consideration in this formulation of every possible leakage location in the network. Using this approach it is ensured that the calculated state bounds for $x_i \in [x_i^l, x_i^u]$ include the true state value irrespective of where the leakage is. Problems **LPmin** and **LPmax** are solved for each state x_i, $i \in \mathcal{N}$, deriving bounds on the state vector x. Note that state bounds are considerably more conservative compared to Algorithm 1 due to the current constraints considering a leakage at every node.

In contrast, when calculating bounds on an individual leakage $q_{\text{leak},i}$, only the existence of the specific leakage should be considered. Leakage constraints are again modified as follows:

$$q_{\text{leak},i}^l \leq q_{\text{leak},i} \leq q_{\text{leak},i}^u, i \in \mathcal{N} \tag{24a}$$

$$q_{\text{leak},l} = 0, \forall l \neq i \in \mathcal{N} \tag{24b}$$

Due to the modified constraints, two separate LPs are formulated for obtaining lower (**LPmin2L**) and upper (**LPmax2L**) bounds on each leakage variable $q_{\text{leak},i} \in z$, as follows:

LPmin2L :

$$\left\{ \begin{array}{ll} \min_z & q_{\text{leak},i} \\ s.t. & (22a) - (22i) \\ & (24) \end{array} \right\}$$

LPmax2L :

$$\left\{ \begin{array}{ll} \max_z & q_{\text{leak},i} \\ s.t. & (22a) - (22i) \\ & (24) \end{array} \right\}$$

Leakage bounds are reduced by solving problems **LPmin2L** and **LPmax2L** for each leakage state $q_{\text{leak},i}$, $i \in \mathcal{N}$. **LPmin2L** or **LPmax2L** may be infeasible when it is impossible to find a solution under the consideration that all nodes are leakage-free except from node i. This signifies the impossibility of a leakage at node i; thus the leakage bounds are set to zero, such that $\tilde{q}_{\text{leak},i} = [0,0]$.

The end of this procedure yields new bounds for the decision variable vector z. New bounds for the emitter coefficients $\tilde{c}$ are then calculated using (19).

4.4. Step 4: Iterative reduction of leakage bounds

In this step, the calculated bounds on states and leakage are used to improve the problem constraints and iteratively reduce state and leakage bounds. This is achieved by reducing the feasible set at the next iteration through the procedure of bounding linearization. The iterations stop when the bounds on the vector of decision variables

remain relatively unchanged. The convergence criterion used is that the change in bounds at iteration m, denoted by $e^{(m)}$ and defined as

$$e^{(m)} \triangleq \left| \left(z^{u(m)} - z^{l(m)} \right) - \left(z^{u(m-1)} - z^{l(m-1)} \right) \right|_1, \tag{25}$$

should be smaller than a small number ϵ, i.e., $e^{(m)} < \epsilon$.

4.5. Step 5: Considering multiple time-steps to improve localization

Assuming that the leakage persists until a known repair event, localization can be improved by exploiting a leakage characteristic that persists over multiple time-steps. Leakage magnitude q_{leak} varies between time-steps because of the varying pressure conditions in the network. However, the leak hole area, a quantity analogous to the emitter coefficient c_j in (5), frequently remains constant during leakage events. Thus, information from previous time-steps can be passed onto the problem solved at time-step k in the form of the previously calculated bounds on emitter coefficients. The initial bounds (iteration $m = 0$) on emitter coefficients at time-step k will then be $\tilde{c}^{(0)}(k) = \tilde{c}(k-1)$.

Algorithm 2 presents the complete procedure for leakage localization using an interval model over multiple time-steps.

Algorithm 2 Leakage localization algorithm

Input: time-step k, measurements $\hat{q}(k)$, $\hat{p}(k)$, demand bounds $\tilde{q}_{\text{ext}}(k)$, $\tilde{g}(k)$, parameter bounds $\tilde{r}$, emitter bounds $\tilde{c}(k-1)$, max leak q_{leak}^{max}

Output: $\tilde{q}_{\text{leak}}(k)$, $\tilde{c}(k)$

begin

1: Define leakage bounds $\tilde{q}_{\text{leak}}^{(0)} = [0, q_{\text{leak}}^{max}]$
2: Calculate state bounds $\tilde{x}^{(0)}$
3: Define emitter bounds $\tilde{c}^{(0)} = \tilde{c}(k-1)$
4: Iteration: $m = 0$
5: **while** $e^{(m)} > \epsilon$ **do**
6:     Bounding linearization using $\tilde{z}^{(m)}$, $\tilde{c}^{(m)}$
7:     **for** $i = 1$ to $n_l + n_n$ **do**
8:         Solve **LPmin2** and **LPmax2** to find x_i^l and x_i^u
9:     **end for**
10:    $\tilde{x}^{(m+1)} = \left[x^l, \ x^u \right]$
11:    **for** $i = 1$ to n_n **do**
12:        Solve **LPmin2L** and **LPmax2L** to find $q_{\text{leak},i}^l$ and $q_{\text{leak},i}^u$
13:        **if** infeasible **then** $\tilde{q}_{\text{leak},i} = [0, 0]$
14:        **end if**
15:    **end for**
16:    $\tilde{q}_{\text{leak}}^{(m+1)} = \left[q_{\text{leak}}^l, \ q_{\text{leak}}^u \right]$
17:    Find $\tilde{c}^{(m+1)}$ using $\tilde{q}_{\text{leak}}^{(m+1)}$ and $\tilde{h}^{(m+1)}$ in (19)
18:    $m = m + 1$
19: **end while**
20: $\tilde{q}_{\text{leak}}(k) = \tilde{q}_{\text{leak}}^{(m)}$, $\tilde{c}(k) = \tilde{c}^{(m)}$

return

---

5. Localization priority index

The proposed leakage localization procedure calculates bounds on node emitter coefficients, given by $\tilde{c}$. The nodes associated with a non-zero emitter coefficient upper bound, are the locations where the existence of a leak is feasible. The leakage-feasible node set $\mathcal{N}_l \subseteq \mathcal{N}$ is defined as follows:

$$\{ j \in \mathcal{N} : c_j^u > 0 \} \in \mathcal{N}_l. \tag{26}$$

When all upper bounds c_j^u are zero except one, a single leak node is accurately identified. The leak node can be determined accurately when the leakage is large, in which case the leak node will have a non-zero lower bound, which confirms the leak location given that only one leakage exists in the system.

In the case when a non-zero emitter upper bound c_j^u is calculated for a node j, it is concluded that a leak is feasible to exist there. However, additional information is included in the localization results in the

form of the magnitude of the derived emitter coefficient bounds $\tilde{c}$. The comparison of node upper bounds can be used to prioritize possible leak locations, which can be a useful tool for water utility repair crews. In simple terms, *priority is given to nodes with larger emitter upper bounds*. The logic behind this approach is that given the model and sensor data which contain the leakage effect, the leak node will typically be one of the most suitable locations for the leakage to exist.

Previously proposed leakage localization assessment metrics in the literature typically depend on the localization results; e.g., when a single candidate leak node is identified, the topological distance from the actual leak is considered as the evaluation metric (Soldevila et al., 2020). In general, most methodologies provide leakage hotspots, i.e., multiple candidate nodes, and the evaluation metric can be the average topological distance from the actual leak node (Soldevila et al., 2016), as well as the number of nodes contained in the leakage hotspot (Blocher, Pecci, & Stoianov, 2020). These evaluation approaches, in general, are not suitable for evaluating localization results in the form of a priority list of possible leak node locations, which may be useful for water utility repair crews.

In this work, in order to quantify how "good" the localization result is, the *Localization Priority Index* (LPI) is defined as follows:

$$\text{LPI} \ (\%) = 100 \cdot \left(\sum_{i \in \mathcal{N} \neq l : c_i^u < c_l^u} 1 \right) / \left(n_n - 1 \right), \tag{27}$$

where l is the index of the leak node. The LPI is used to rank the feasible leak locations by their emitter upper bounds value and is a confidence index of the leakage location. It essentially indicates the minimum percentage of nodes that are excluded from the node search space before the leakage is found by a leakage repair crew, given the priority defined by $\tilde{c}$. This is computed by finding the percentage of nodes with lower priority than the actual leak node l. The LPI = 100% when perfect localization is achieved, i.e. the leak node has the highest priority, while LPI = 0% when the leak node has the same or lower priority than all the other nodes.

6. Case studies

In this section the proposed leakage detection and localization methodology is demonstrated using benchmark case studies. In the first case study, a benchmark transport network is used to illustrate the results obtained by the proposed methodology when two specific leakages occur in the network. The second case study, uses a leakage diagnosis benchmark dataset containing sensor measurements from multiple operating scenarios of a benchmark network. Each scenario varies with respect to pipe parameters, water demands, leakage location and magnitude. The code, models and data generated or used during these case studies are available in an online repository (Vrachimis, 2021).

6.1. Example application on a transport network

The benchmark network 'Hanoi', shown in Fig. 2(a), is used to examine the performance of the proposed leakage diagnosis methodology. The 'Hanoi' network is a transport network with large demands at nodes of up to 300 m³/h. A version of the network, which is considered the "real" system, is simulated for 48 h from time 00:00 h, with a hydraulic time-step of $\Delta t = 30$ min. It contains unique realistic water demand patterns for each node with a 24 h periodicity. A leakage occurs sometime in the first 24 h, ensuring in this way that the performance of the algorithm is assessed under the daily varying network conditions.

A nominal hydraulic model of the network is given which is assumed topologically identical to the real network. The given pipe parameters are accurate within ±5% of the true parameter values. Moreover, demand pseudo-measurements are available for all nodes which are accurate within ±10% of the true demand values.

The real system is measured at the inlet of the network, i.e. the head at the reservoir node 1 and the flow at link 1. Additionally, 4
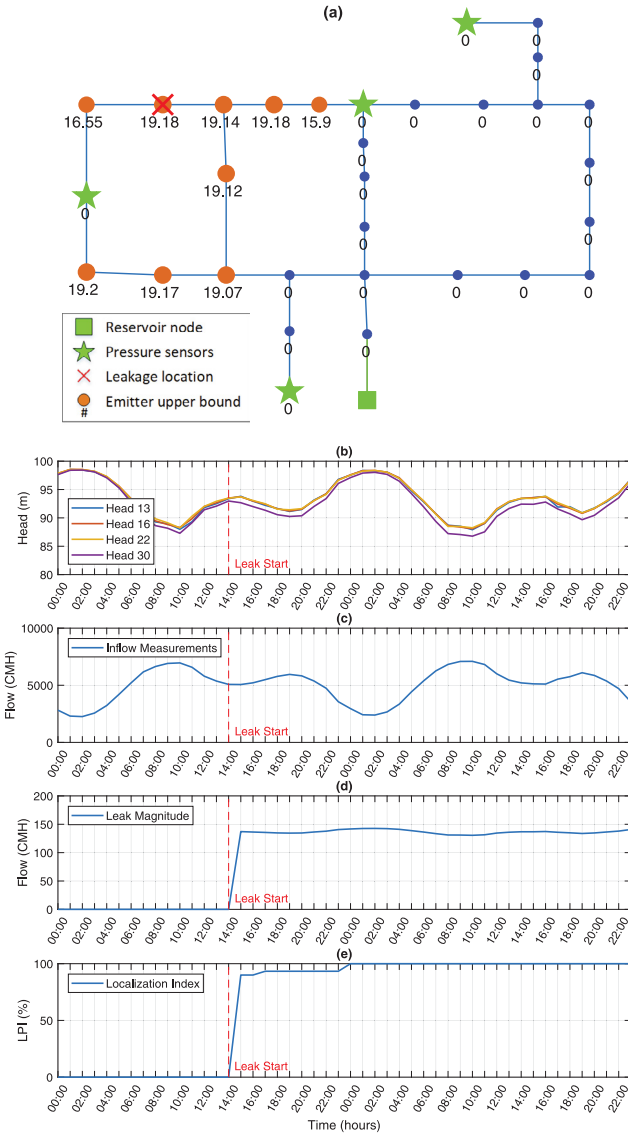
Fig. 2. Example of leakage diagnosis on the Hanoi network: Leakage magnitude of ∼ 2.73% of average system inflow.

Fig. 3. Example of leakage diagnosis on the Hanoi network: Leakage magnitude of ∼ 1.53% of average system inflow.

pressure sensors are installed at nodes $\{13, 16, 22, 30\}$ using a sensor placement procedure which maximizes the minimum sensitivity of all sensors to all possible leakages (Casillas et al., 2013). Pressure and flow sensor noise bounds $\bar{v}_p$ and $\bar{v}_q$ respectively are defined relative to the measurement value, such that $\bar{v}_p(k) = \pm 2\% \, \hat{p}(k)$ and $\bar{v}_q(k) = \pm 2\% \, \hat{q}(k)$.

A demand calibration methodology is considered to actively use the sensor measurements to give demand estimates for groups of nodes. To make the problem of calibration over-determined, the methodology divides the network into as many groups as the number of pressure sensors. Network partitioning is performed in this case by minimizing the flow exchange between node groups. Moreover, the following two constraints are imposed: i) one pressure sensor should belong in each group, ii) node groups should have the same number of nodes plus/minus two nodes. Algorithms that perform this kind of partitioning are available not only for water systems (Rajeswaran, Narasimhan, & Narasimhan, 2018), but also for power systems (Kyriacou, Demetriou, Panayiotou, & Kyriakides, 2018) and smart buildings (Kyriacou, Timotheou, Michaelides, Panayiotou, & Polycarpou, 2017). The calibration methodology is assumed to be able to give an estimate of a group demand within $\pm 5\%$ of the true group demand values (Kun et al., 2015).
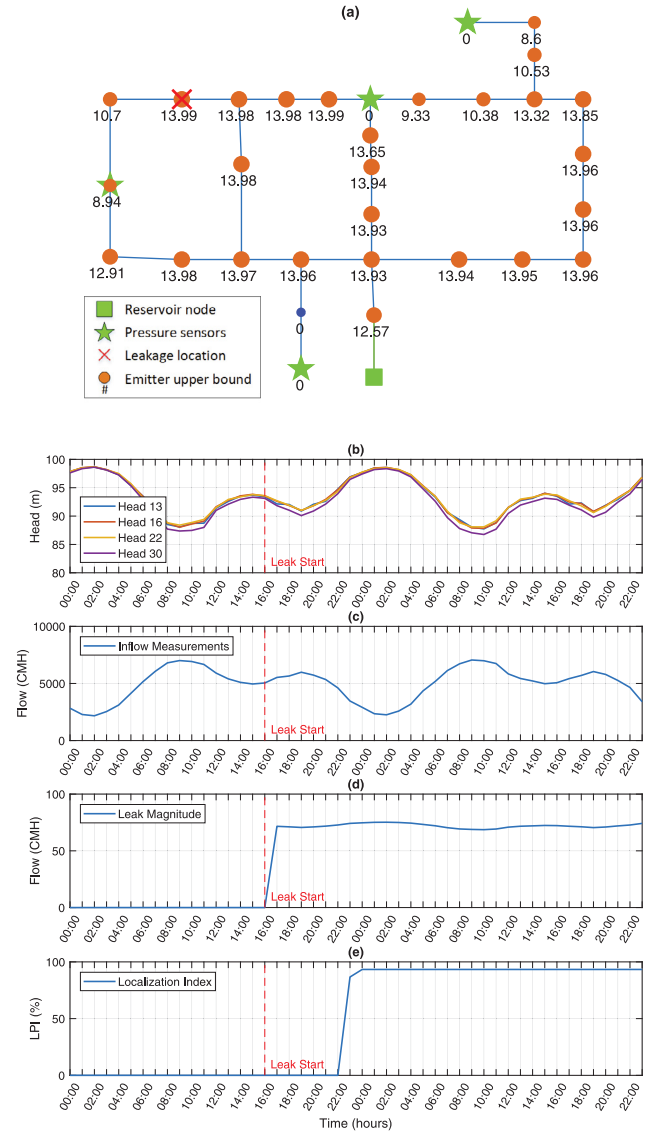
In the first illustrative example of Fig. 2, an abrupt pressure-dependent leakage occurs at node $j = 32$ at 14:00 h, with an emitter coefficient of $c_j = 17.3$ and an emitter exponent of $\alpha = 0.5$. This results in a leakage with magnitude of approximately 137 m³/h, which is 2.73% of the average system inflow for the given network. The effect of the leakage is unnoticeable by just observing the measurements, as illustrated in Fig. 2(b)–(c). In this example, the proposed leakage detection methodology detects the occurrence of a leakage immediately, i.e. given the first measurements that contain the effect of the fault.

After leakage detection, the localization procedure begins by calculating initial bounds on the leakage. These are assumed to be given by a methodology which monitors the inflow of the system. Here, conservative bounds of $\tilde{q}_{\text{leak},j} = [50, 200], \, \forall j \in \mathcal{N}$ are assigned. The localization procedure yields bounds on node emitter coefficients $\tilde{c}_j, \, \forall j \in \mathcal{N}$. The most useful information is provided by emitter upper bounds, as it can be used to define the priority of each node in the localization procedure. The emitter upper bounds derived in this example can be seen in Fig. 2(a). It is illustrated that nodes close to the leakage have a non-zero upper bound, indicating the feasibility of these nodes having a leakage given the uncertain model and measurements.

The priority given to nodes by their upper bounds is evaluated using the Localization Priority Index (LPI). The higher the LPI, the higher the priority given by the methodology to the node containing the leak. In Fig. 2(d), the LPI for the given example begins at 90% when the leakage is detected at 14:00 h and increases to 100% by utilizing measurements from subsequent time-steps.

In the second illustrative example of Fig. 3, a smaller leakage occurs at the same node ($j = 32$) as in the first example at 16:00 h, with an emitter coefficient of $c_j = 9.09$. This results in a leakage with average magnitude of 74.9 m^3/h, which is approximately 1.53% of the average system inflow. In this example, the leakage is detected 6 h after it occurs, at 22:00 h. The initial value of LPI is 86.67% and gradually increases to 93.33% with the accumulation of new measurements, as seen in Fig. 3(d). Due to the smaller leakage magnitude, it is feasible for the leak to exist in more nodes than in the previous example, as observed in Fig. 3(a). Interestingly though, the leak node is given top priority with only one other node having the same emitter value, a result which translates in LPI = 93.33%.

6.2. Sensitivity analysis of the number of sensors

In this Section a sensitivity analysis is provided which shows the relationship of the detection delay and Leak Priority Index (LPI) to the number of pressure sensors on the transport network example application.

The benchmark network 'Hanoi' has been used to perform a sensitivity analysis on the performance of the proposed methodology as a function of the pressure sensors installed in the network. The pressure sensors in the network were varied between one (1) to sixteen (16), which correspond to a coverage of 3.1% to 50% of the network nodes. A sensor placement procedure which optimizes sensitivity of all sensors to all possible leakages was used (Casillas et al., 2013). For every sensor configuration, extensive simulations with a single leakage occurring at a random network node were performed. The leakage emitter coefficient for a leakage occurring at a node j was set to $c_j = 12, \forall j \in \mathcal{N}$. This configuration results in leakages with magnitude of approximately 2% of the system inflow (which is a relatively small leakage), depending on the pressure of the corresponding leakage node. In every simulation, demand patterns and pipe parameters were randomized within ±10% and ±5% respectively of their estimated values.

The proposed leakage detection and localization methodology was applied on every simulation, and the leakage detection delay, the final Localization Priority Index (LPI) as well as the leakage-feasible node set $\mathcal{N}_l$ were recorded. The results from approximately 500 simulations are aggregated in Fig. 4, where the median of the aforementioned metrics for each sensor configuration (given as a percentage of the sensor nodes) are plotted. The median for each metric and sensor configuration is derived using at least 30 simulations with the leakage existing at a different location in each simulation. It is observed that the median detection delay starts from 2 h when 3% of the nodes are monitored and it converges to 30 min when approximately 25% of the nodes are monitored. Note that 30 min is the fastest the leakage can be detected in these scenarios, because this is the measurement time-step used. The median LPI begins at approximately 56% and converges to 100% (perfect localization) when approximately 25% of the nodes are monitored. The median percentage of nodes on which a leakage is feasible according to the proposed methodology, begins at 90.6% and converges to 12.5% (which is equivalent to just four nodes in the network under examination) when approximately 25% of the nodes are monitored. As a remark, all leakages were successfully detected in these scenarios even with the minimum number of pressure sensors (one). Even though the value of uncertainty bounds and leakage magnitude chosen for these simulations can be considered representative of real networks, it is expected that the performance would degrade and some leakages will not be detected if larger uncertainty bounds and smaller leakages are considered.
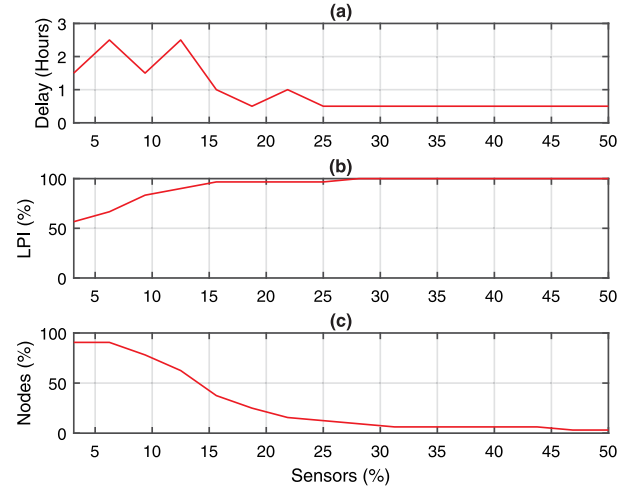

Fig. 4. Sensitivity analysis of the proposed methodology performance with respect to the percentage of nodes in the network with pressure sensors installed: (a) Median Leakage Detection Delay in hours, (b) Median Localization Priority Index (LPI), (c) Median percentage of nodes in the leakage-feasible set $\mathcal{N}_l$.

6.3. Methodology evaluation using LeakDB

In this subsection the proposed methodology is extensively tested using a benchmark dataset of leakages. The Leakage Diagnosis Benchmark (LeakDB) (Vrachimis, Kyriakou, et al., 2018) is a realistic leakage dataset comprised of a large number of realistic leakage scenarios which occur randomly at different water distribution benchmark networks, of different size and topology. For each benchmark network and for each leakage scenario, the leakage parameters (e.g., leak location and size), the structural parameters (e.g., length, pipe roughness) and realistic consumer pressure-driven demands are varied. For this case study, the dataset of the Hanoi benchmark network was selected.

In total, 500 different scenarios are used in which the parameters and demands are varied randomly inside known uncertainty bounds. The demand uncertainty bounds $\tilde{q}_{\text{ext}}(k)$ and parameter uncertainty upper bounds $\tilde{r}$ are varied between ±10% and ±5% respectively of the given model nominal values. Moreover, a calibration methodology gives an estimate of group demands within ±5% of the true group demand values, given the sensor set.

It is assumed that flow and pressure are measured at the inlet of the network, i.e. the head at the reservoir node 1 and the flow at link 1. Additionally, 4 pressure sensors are installed at nodes $\{13, 16, 22, 30\}$ as in Examples 1 and 2. Pressure and flow sensor noise bounds $\tilde{v}_p$ and $\tilde{v}_q$ respectively are defined relative to the measurement value, such that $\tilde{v}_p(k) = ±2\% \, \hat{p}(k)$ and $\tilde{v}_q(k) = ±2\% \, \hat{q}(k)$.

In each scenario, a leakage is induced on a node, which is random in location and magnitude. It is possible for a scenario not to include a leakage. The leakages have a different profile, by varying the leak emitter coefficient. The emitter coefficients were calculated so that the resulting leakage magnitudes $q_{\text{leak}}(k)$ (m^3/h) are within $0 - 4\%$ of the average system inflow.

The benchmark also provides a scoring algorithm to evaluate the ability of an algorithm to detect leakages. Here the following standard metrics were extracted for the application of the proposed algorithm for leakage detection on all scenarios: (1) A fault existed and was detected (True Positive TP), (2) A fault existed but was not detected (False Negative FN), (3) A fault did not exist but a fault was detected (False Positive FP), (4) A fault did not exist and was not detected (True Negative TN). The methodology in Vrachimis, Eliades, and Polycarpou (2018), where interval-state estimation is used to create thresholds for leakage detection, is also applied on the same benchmark in order to compare the performance with the proposed model-invalidation
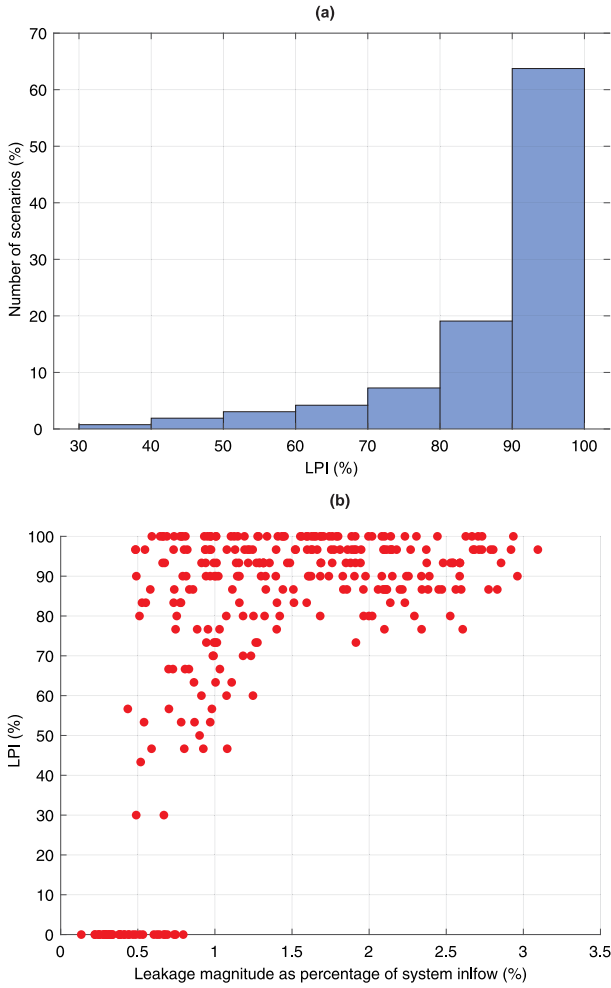
Fig. 5. Application of the proposed Leakage Localization methodology on 500 scenarios: (a) Percentage of scenarios with specific range of Localization Priority Index value, (b) Localization Priority Index with respect to leakage magnitude.

Table 1
Detection statistics from applying the proposed fault detection algorithm (Alg.1) to a total of 500 leakage scenarios and comparison with the methodology (Alg.2) proposed in Vrachimis, Eliades, and Polycarpou (2018).

Method	TP %	FN %	FP %	TN %
Alg.1	87.33%	12.66%	0	100
Alg.2	56.28%	43.72%	0	100

methodology. The percentages for each case with 4 pressure sensors in the network are presented in Table 1.

The median and average detection delay for the scenarios where a detection occurs for the proposed methodology was 0 and 4.28 time-steps respectively. Both methodologies achieve zero false alarms when a fault is not present (0% False Positives and 100% True Negatives). This is due to the accurate knowledge of uncertainty bounds, which are provided by the LeakDB As a result, the interval model is always valid when a fault does not exist.

It is worth noticing in Table 1 the high detection rate of the proposed methodology (Alg.1) compared to the work in Vrachimis, Eliades, and Polycarpou (2018) (Alg.2). The improved performance is explained by the fact that using model-invalidation the linear constraints formulated during the leakage detection procedure are part of the state-set representation, along with the bounds on states; in comparison, the state-set in Vrachimis, Eliades, and Polycarpou (2018) is represented only by the bounds on states. This indicates that model invalidation

is a more powerful tool than creating detection thresholds when using models containing bounded uncertainties.

To evaluate the ability of the proposed methodology for leakage localization, the LPI in all scenarios was examined . In Fig. 5(a), scenarios with similar LPI are binned together, and the percentage of scenarios in a given LPI range is calculated. It is observed that the majority of scenarios have an LPI between $80 - 100\%$. Moreover, it is interesting to investigate the relationship between leakage magnitude and LPI. As seen in Fig. 5(a), the leakage scenarios with leakage magnitude below 1.5% of the average system inflow have a LPI which varies between $0 - 100\%$, depending on the leakage location. However, for leakages above 1.5%, the LPI remains above 70% in all cases.

6.4. Scalability

In this Section the complexity of the proposed methodology and the scalability to larger networks is discussed. In practice, the ability of the proposed algorithm to scale well for increasing network sizes can be assessed by analyzing its execution time on standard processing units. Fig. 6 depicts the execution time of the algorithm on benchmark networks of different sizes. Note that the execution time of interest is that of a single time-step, at which point the algorithm receives new measurements and needs to re-assess the existence of leakages in the network. The non-parallelized algorithm execution times were recorded on a personal computer with an Intel Core i5-2400 CPU at 3.10 GHz. The networks used range from small benchmark networks representing transport networks of under 100 states ('Net1', 'Hanoi','Net2') to medium size networks representing typical District Metered Areas ('CY-DMA','Net3'), to large DMAs of a city in Cyprus ('CY-DMA-2','CY-DMA-3').

It can be observed in Fig. 6 that the execution time exhibits polynomial growth with respect to the number of states. The straight line shows the nonlinear least-squares fit of the polynomial model $f(x) = ax^b + c$, with $a = 2.397^{-5}, b = 1.995, c = 0.03846$, indicating that the execution time scales almost quadratically to the number of states. The exponent obtained from this model $b = 1.995$ implies the empirical computational complexity of executing Algorithm 1 in the manuscript to be $O(N^2)$, where N is the number of states.

The demonstrated scalability results in Fig. 6 indicate that the proposed algorithm scales reasonably well to realistic problem sizes and also maintains the online nature of the proposed technique. This is because the typical sensor sampling times in water distribution systems is in the range of $5 - 30$ min, due to relatively slow dynamics driven by water demands in these systems. Note that the algorithm execution times can be significantly improved by utilizing more advanced computational techniques, such as code parallelization, on a server which typically hosts a water utility's data acquisition system. Using these improvements, the proposed algorithm can be utilized for real-time leakage detection on real-size DMA networks.

7. Conclusions and future work

In this work, a methodology for leakage detection and localization in WDS using an interval-model is proposed, i.e. a hydraulic model in which uncertainties are modeled by intervals defined by a lower and upper bound. The interval-model is created using historical data of water demands by which uncertainty bounds are derived. Demand calibration data in the form of group demand estimates are also incorporated into the problem. Leakage detection is performed by incorporating real-time measurements from the network and checking the validity of the healthy interval-model. The results show increased detectability rates compared to similar methodologies which work using bounded uncertainties. Moreover, a methodology for leakage localization is proposed using this interval-model. The localization procedure yields a node priority list which can be used by water utilities to search for the exact leakage location using equipment based methods.
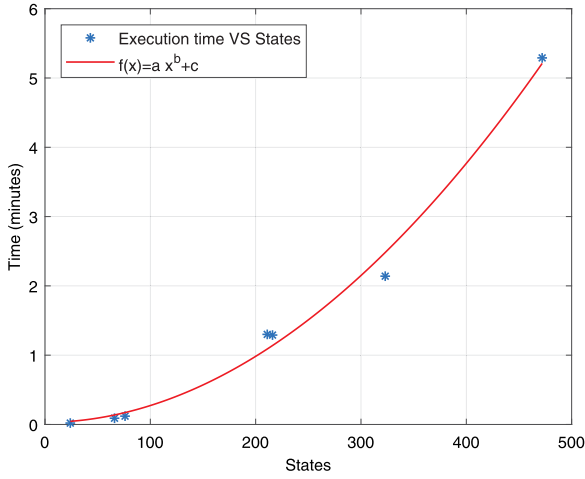
Fig. 6. Execution times for a single time step of the proposed algorithm as a function of the number of states in the network.

Fig. A.1. Convex hull of S_q.

Localization results on a leakage diagnosis benchmark dataset show that the leakage search space can be reduced to less than 20% of the network for leakage magnitudes above 1.5% of the system inflow.

The proposed methodology was designed specifically to deal with the non-linearities present in WDS, as described in Section 3.2. However, this procedure can be generalized so that the proposed leakage diagnosis methodology may be applied to a range of networked systems with the same characteristics, such as industrial processes with hydraulic dynamics (including desalination plants, irrigation, wastewater), as well as oil transport networks. Another example are DC power networks, of which the structure and mathematical representation closely resembles that of water systems, e.g. considering the equivalence of water flow to power flow and of head-loss in pipes to power-loss in wires. Another direction could be the modification of this methodology to capture different features from the available model and data, such as the probability of a fault existing at a node location. However, this would require to change the assumption that only bounds on the uncertain parameters are available, which is a key principle in the formulation of this work, and instead apply probabilistic rather than deterministic methods in deriving the leak location probability. The ability of the methodology to improve localization results using the available control inputs in the system can also be investigated in the future, while considering a tracking error constraint (Sun, Qiu, Karimi, & Gao, 2019) to conserve system performance. Finally, future work will investigate the case when multiple leakages exist simultaneously in the network.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments
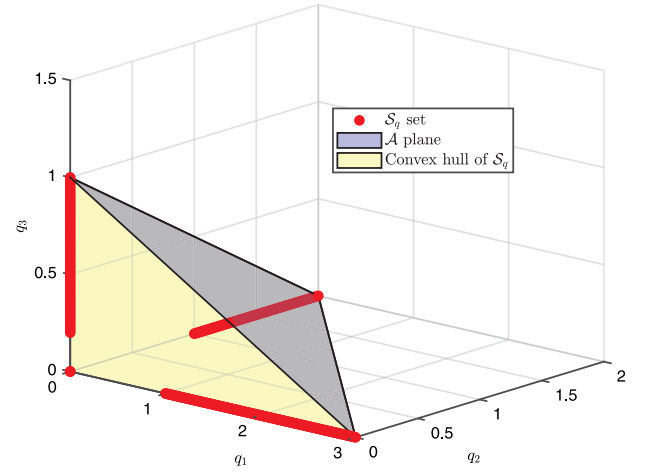
Appendix. Proof of Lemma 1

Consider three variables, indicated by $q_i, i \in \{1, 2, 3\}$, out of which only one is non-zero, while if q_i is non-zero then $q_i \in \tilde{q}_i$ holds, such that $q_i^l \leq q_i \leq q_i^u$, $0 \leq q_i^l \leq q_i^u$, $\forall i$. The feasible set for the vector $q = [q_1, q_2, q_3]^\top$ is then given by S_q as follows:

$$S_q = \{q_1^l \leq q_1 \leq q_1^u, q_2 = 0, q_3 = 0\} \cup \tag{A.1}$$

$$\{q_1 = 0, q_2^l \leq q_2 \leq q_2^u, q_3 = 0\} \cup \tag{A.2}$$

$$\{q_1 = 0, q_2 = 0, q_3^l \leq q_3 \leq q_3^u\} \tag{A.3}$$

The convex hull of S_q can be defined as the intersection of the following half-spaces, as illustrated in Fig. A.1:

$$\{q_1 \geq 0\} \cap \{q_2 \geq 0\} \cap \{q_3 \geq 0\} \cap \{\mathcal{A} \leq 0\} \tag{A.4}$$

where $\mathcal{A}$ is the plane passing through points $p_i = [q_{1i}, q_{2i}, q_{3i}]$, $i \in \{1, 2, 3\}$ explicitly defined as $p_1 = [q_1^u, 0, 0]$, $p_2 = [0, q_2^u, 0]$, $p_3 = [0, 0, q_3^u]$.

Let plane $\mathcal{A}$ be described by equation

$$aq_1 + bq_2 + cq_3 + d = 0 \tag{A.5}$$

where d is any non-zero number. Coefficients a, b, and c can be found by solving the following system of equations:

$$aq_{11} + bq_{21} + cq_{31} + d = 0 \tag{A.6a}$$

$$aq_{12} + bq_{22} + cq_{32} + d = 0 \tag{A.6b}$$

$$aq_{13} + bq_{23} + cq_{33} + d = 0 \tag{A.6c}$$

The solution to the system of equations (A.6) for the given points p_i yields:

$$a = \frac{-d}{q_1^u}, b = \frac{-d}{q_2^u}, c = \frac{-d}{q_3^u}. \tag{A.7}$$

Notice that the coefficients a, b, c depend only on their corresponding variable q_1, q_2, q_3 upper bound. Substituting the coefficients in (A.5), the equation describing plane $\mathcal{A}$ is given by

$$\frac{q_1}{q_1^u} + \frac{q_2}{q_2^u} + \frac{q_3}{q_3^u} = 1 \tag{A.8}$$

The half-plane $\mathcal{A} \leq 0$ is then given by:

$$\frac{q_1}{q_1^u} + \frac{q_2}{q_2^u} + \frac{q_3}{q_3^u} \leq 1 \tag{A.9}$$

It is then easy to generalize the constraints defining the convex hull of S_q for n dimensions, yielding:

$$q_i \geq 0, \forall i \in \{1, \dots, n\} \tag{A.10}$$

$$\left(\sum_{i \in \{1,\dots,n\}} \frac{q_i}{q_i^u} \right) \leq 1 \qquad\qquad (A.11)$$

References

Andersen, J. H., Powell, R. S., & Marsh, J. F. (2001). Constrained state estimation with applications in water distribution network monitoring. *International Journal of Systems Science*, *32*, 807–816. http://dx.doi.org/10.1080/002077201750282007.

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*(14), 36–41. http://dx.doi.org/10.3182/20100901-3-IT-2016.00054.

Blocher, C., Pecci, F., & Stoianov, I. (2020). Localizing leakage hotspots in water distribution networks via the regularization of an inverse problem. *Journal of Hydraulic Engineering*, *146*(4), Article 04020025. http://dx.doi.org/10.1061/(ASCE)HY.1943-7900.0001721.

Blokker, E. J. M., Vreeburg, J. H. G., & van Dijk, J. C. (2010). Simulating residential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management*, *136*(1), 19–26. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000002.

Boulos, P. F., Lansey, K. E., & Karney, B. W. (2006). *Comprehensive water distribution systems analysis handbook for engineers and planners*. American Water Works Association.

Casillas, M., Puig, V., Garza-Castañón, L., & Rosich, A. (2013). Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors*, *13*(11), 14984–15005. http://dx.doi.org/10.3390/s131114984.

Chan, T. K., Chin, C. S., & Zhong, X. (2018). Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*, *6*, 78846–78867. http://dx.doi.org/10.1109/ACCESS.2018.2885444.

Cheng, W., & He, Z. (2011). Calibration of nodal demand in water distribution systems. *Journal of Water Resources Planning and Management*, *137*(1), 31–40. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000093.

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*(4), 212–227. http://dx.doi.org/10.1016/j.jher.2009.02.003.

Cugueró-Escofet, M. À., Puig, V., & Quevedo, J. (2017). Optimal pressure sensor placement and assessment for leak location using a relaxed isolation index: Application to the barcelona water network. *Control Engineering Practice*, *63*, 1–12. http://dx.doi.org/10.1016/j.conengprac.2017.03.003.

Díaz, S., González, J., & Mínguez, R. (2016). Observability analysis in water transport networks: Algebraic approach. *Journal of Water Resources Planning and Management*, *142*(4), Article 04015071. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000621.

Díaz, S., Mínguez, R., & González, J. (2017). Topological observability analysis in water distribution systems. *Journal of Water Resources Planning and Management*, *143*(5), Article 06017001. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000762.

Do, N. C., Simpson, A. R., Deuerlein, J. W., & Piller, O. (2016). Calibration of water demand multipliers in water distribution systems using genetic algorithms. *Journal of Water Resources Planning and Management*, *142*(11), Article 04016044. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000691.

Eliades, D. G., & Polycarpou, M. M. (2012). Leakage fault detection in district metered areas of water distribution systems. *Journal of Hydroinformatics*, *14*(4), 992–1005. http://dx.doi.org/10.2166/hydro.2012.109.

Farley, B., Mounce, S. R., & Boxall, J. B. (2010). Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal*, *7*(6), 345–356. http://dx.doi.org/10.1080/1573062X.2010.526230.

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*(2), 261–269. http://dx.doi.org/10.1016/J.AEI.2013.01.001.

Gurobi Optimization, L. (2019). Gurobi optimizer reference manual. URL https://www.gurobi.com/documentation.

Harirchi, F., & Ozay, N. (2018). Guaranteed model-based fault detection in cyber-physical systems: A model invalidation approach. *Automatica*, *93*, 476–488. http://dx.doi.org/10.1016/J.AUTOMATICA.2018.03.040.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., & Savic, D. A. (2014). Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation. *Journal of Water Resources Planning and Management*, *140*, 169–183. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000325.

Kang, D., & Lansey, K. (2011). Demand and roughness estimation in water distribution systems. *Journal of Water Resources Planning and Management*, *137*(1), 20–30. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000086.

Kun, D., Tian-Yu, L., Jun-Hui, W., & Jin-Song, G. (2015). Inversion model of water distribution systems for nodal demand calibration. *Journal of Water Resources Planning and Management*, *141*(9), Article 04015002. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000506.

Kyriacou, A., Demetriou, P., Panayiotou, C., & Kyriakides, E. (2018). Controlled islanding solution for large-scale power systems. *IEEE Transactions on Power Systems*, *33*(2), 1591–1602. http://dx.doi.org/10.1109/TPWRS.2017.2738326.

Kyriacou, A., Timotheou, S., Michaelides, M. P., Panayiotou, C., & Polycarpou, M. (2017). Partitioning of intelligent buildings for distributed contaminant detection and isolation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *1*(2), 72–86. http://dx.doi.org/10.1109/tetci.2017.2665119.

Mareels, I., Weyer, E., Ooi, S. K., Cantoni, M., Li, Y., & Nair, G. (2005). Systems engineering for irrigation systems: Successes and challenges. *IFAC Proceedings Volumes*, *38*(1), 1–16. http://dx.doi.org/10.3182/20050703-6-CZ-1902.02092, 16th IFAC World Congress.

Moore, R. E., Kearfott, R. B., & Cloud, M. J. (2009). *Introduction to interval analysis*. Society for Industrial and Applied Mathematics.

Moors, J., Scholten, L., van der Hoek, J. P., & den Besten, J. (2018). Automated leak localization performance without detailed demand distribution data. *Urban Water Journal*, *15*(2), 116–123. http://dx.doi.org/10.1080/1573062X.2017.1414272.

Ozay, N., Sznaier, M., & Lagoa, C. (2014). Convex certificates for model (in)validation of switched affine systems with unknown switches. *IEEE Transactions on Automatic Control*, *59*(11), 2921–2932. http://dx.doi.org/10.1109/TAC.2014.2351714.

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*(10), 1157–1167. http://dx.doi.org/10.1016/j.conengprac.2011.06.004.

Perez, R., Sanz, G., Puig, V., Quevedo, J., Cuguero Escofet, M. A., Nejjari, F., et al. (2014). Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in barcelona. *IEEE Control Systems*, *34*(4), 24–36. http://dx.doi.org/10.1109/MCS.2014.2320336.

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

Puust, R., Kapelan, Z., Savic, D. A., & Koppel, T. (2010). A review of methods for leakage management in pipe networks. *Urban Water Journal*, *7*(1), 25–45. http://dx.doi.org/10.1080/15730621003610878.

Rajeswaran, A., Narasimhan, S., & Narasimhan, S. (2018). A graph partitioning algorithm for leak detection in water distribution networks. *Computers & Chemical Engineering*, *108*, 11–23. http://dx.doi.org/10.1016/j.compchemeng.2017.08.007.

Ribeiro, L., Sousa, J., Marques, A., & Simões, N. (2015). Locating leaks with TrustRank algorithm support. *Water*, *7*(4), 1378–1401. http://dx.doi.org/10.3390/w7041378.

Rosa, P., Silvestre, C., Shamma, J. S., & Athans, M. (2010). Fault detection and isolation of LTV systems using set-valued observers. In *Proceedings of 49th IEEE conference on decision and control* (pp. 768–773). IEEE, http://dx.doi.org/10.1109/CDC.2010.5717634.

Rossman, L. A. (2000). *EPANET 2: Users manual*. US Environmental Protection Agency. Office of Research and Development. National Risk Management Research Laboratory.

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*(2), Article 04015057. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000592.

Savic, D. A., Kapelan, Z. S., & Jonkergouw, P. M. (2009). Quo vadis water distribution model calibration? *Urban Water Journal*, *6*, 3–22. http://dx.doi.org/10.1080/15730620802613380.

Soldevila, A., Blesa, J., Jensen, T. N., Tornil-Sin, S., Fernandez-Canti, 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*, 1–12. http://dx.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. http://dx.doi.org/10.1016/j.conengprac.2016.07.006.

Sun, K., Qiu, J., Karimi, H. R., & Gao, H. (2019). A novel finite-time control for nonstrict feedback saturated nonlinear systems with tracking error constraint. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1–12. http://dx.doi.org/10.1109/TSMC.2019.2958072.

Vrachimis, S. G. (2021). KIOS-Research/Leakage-Diagnosis: Model Invalidation. Zenodo, v1.21, https://doi.org/10.5281/zenodo.4555904.

Vrachimis, S. G., Eliades, D. G., & Polycarpou, M. M. (2018). Leak detection in water distribution systems using hydraulic interval state estimation. In *Proceedings of IEEE Conference on Control Technology and Applications (CCTA)* (pp. 565–570). IEEE, http://dx.doi.org/10.1109/CCTA.2018.8511516.

Vrachimis, S. G., Eliades, D. G., Taormina, R., Ostfeld, A., Kapelan, Z., Liu, S., et al. (2020). BattLeDIM 2020: Battle of the leakage detection and isolation methods. URL http://battledim.ucy.ac.cy/.

Vrachimis, S. G., Kyriakou, M. S., Eliades, D. G., & Polycarpou, M. M. (2018). LeakDB : A benchmark dataset for leakage diagnosis in water distribution networks description of benchmark. In *WDSA/CCWI joint conference proceedings, vol. 1*. http://dx.doi.org/10.5281/zenodo.1313116.

Vrachimis, S. G., Timotheou, S., Eliades, D. G., & Polycarpou, M. M. (2018). Iterative hydraulic interval state estimation for water distribution networks. *Journal of Water Resources Planning and Management*, *145*(1), Article 04018087. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0001011.

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*(1), 116–128. http://dx.doi.org/10.1061/(ASCE)0733-9496(2010)136:1(116).