

Localizing Leakage Hotspots in Water Distribution Networks via the Regularization of an Inverse Problem

Caroline Blocher¹, Filippo Pecci², and Ivan Stoianov³

¹InfraSense Labs, Department of Civil and Environmental Engineering, Imperial College,
London, UK. Email: c.blocher16@imperial.ac.uk

²InfraSense Labs, Department of Civil and Environmental Engineering, Imperial College,
London, UK.

³InfraSense Labs, Department of Civil and Environmental Engineering, Imperial College,
London, UK.

ABSTRACT

The ill-posed inverse problem for detecting and localizing leakage hotspots is solved using a novel optimization-based method which aims to minimize the difference between hydraulic measurement data and simulated steady states of a water distribution network. Regularization constrains the set of leak candidate nodes obtained from a solution to the optimization problem. Hydraulic conservation laws are enforced as nonlinear constraints. The resulting nonconvex optimization problem is solved using smooth mathematical optimization techniques. The solution identifies leakage hotspot areas, which can then be further investigated with alternative methods for precise leak localization. A metric is proposed to quantitatively assess the performance of the developed leak localization approach in comparison with a method that uses the sensitivity matrix. In addition, we propose a strategy to select the regularization parameter when large-scale operational networks are considered. Using two numerical case studies, we demonstrate that the proposed approach outperforms the sensitivity matrix method, with regards to leak isolation, in most single leak scenarios. Moreover, the developed method enables the localization of multiple simultaneous leaks.

INTRODUCTION

The continuous detection and localization of leaks has become a critical operational priority for UK water utilities as the government regulator, Ofwat, imposed stricter service requirements with significant financial penalties and rewards (Ofwat 2019). Similar regulatory arrangements are currently adopted in other countries around the world (e.g. Italy, China) where aging infrastructure, financial constraints and population growth are enforcing a more proactive approach towards the optimization of the hydraulic performance of water supply networks, risk, investment and levels of service.

Cost-effective techniques are required to continuously monitor and identify leakage hotspots, i.e. areas of a water supply network where further leak localization resources should be prioritized. Water utilities are becoming more proactive in the use of hydraulic models for near real-time operational management and incident response, and gradually increase the level of hydraulic monitoring in their networks. Consequently, it is hugely beneficial to be able to utilize hydraulic models for the continuous detection and localization of bursts/leaks (and leakage hotspots); and also be able to quantitatively assess the limitations of hydraulic model based leak detection methods for a specific sensor placement in a water distribution network.

A major challenge for using hydraulic model based leak detection and localization is the small number of available sensor measurements in comparison to the large number of possible leak locations. Pudar and Liggett (1992) formulated and solved an inverse problem for leak localization based on a steady state hydraulic model, by minimizing the differences (residuals) between sensor measurements and simulated states. A solution to the inverse problem determines the relevant parameters of a water distribution network model, such as leak flow and location, from sensor measurements. In an operational water network, the inverse problem is in general under-determined because of the small number of available sensor measurements and hence ill-posed.

To solve the under-determined inverse problem for leak localization, where multiple equivalent leak candidates exist, Pudar and Liggett (1992) minimized the ℓ_2 -norm of the leak parameters. In comparison, Berglund et al. (2017) reduced the number of unknowns to obtain a well-posed

inverse problem by selecting candidate nodes prior to solving the optimization problem. Another approach consists in multiple runs of a heuristic localization algorithm to obtain different solutions and hence a set of possible leak locations, e.g. Steffelbauer et al. (2014). Sanz and Pérez (2015) and Sanz et al. (2016) formulated the inverse problem for the calibration of demand, which is under-determined as well, in order to solve the leak localization problem. To solve the ill-posed problem, Sanz and Pérez (2015) grouped demand nodes based on the sensitivity of the measured head to a change in demand at a node. The deviation of the calibrated demand from the expected demand for a given group indicates the presence and an approximate location of a leak (Sanz et al. 2016). Prior assumptions about the leak candidates and node groups are required and these affect the search for the leak location.

An alternative approach, that avoids solving the ill-posed inverse problem, is to directly evaluate the residuals in relation to the sensitivities of pressure measurements for different leak locations, e.g. Perez et al. (2014), Casillas et al. (2013). However, this approach based on the sensitivity matrix is limited to single leak scenarios (Perez et al. 2014), while the inverse problem does not have these constraints and can be formulated for the detection and localization of multiple simultaneous leaks.

Solving the inverse problem, however, requires strategies to deal with the ill-posedness. Ill-posed inverse problems arise in many engineering applications where regularization methods are often applied to mitigate the effect of ill-posedness (Aster et al. 2005). To the best of the authors' knowledge, regularization methods for the solution of an inverse problem have never been applied to the problem of leak detection and localization in water distribution networks. In this article, we formulate a regularized inverse problem for localizing leaks (leakage hotspots) by minimizing the sum of squares of the residuals, which are the difference between hydraulic sensor measurements and steady-state model simulations. The head losses due to friction are modelled using a quadratic approximation in order to apply smooth mathematical optimization solution methods. The resulting optimization problem for leak localization is then solved without requiring prior assumptions on leak candidates or node grouping. As the performance of the leak localization depends on the

choice of a regularization parameter, an approach to select this parameter is investigated.

In addition to the localization problem being under-determined, model and data uncertainty also present challenges for the hydraulic model based leak localization methods (Pudar and Liggett 1992). Some leak localization methods address this explicitly. In Goulet et al. (2013) and Poulakis et al. (2003) for example, the residuals are considered as random variables. The focus of the present article is on addressing the under-determined inverse problem. The effect of uncertainties is not investigated and it is a subject for further work.

When evaluating the success of a leak localization method, most studies rely on showing example scenarios, e.g. Pudar and Liggett (1992), Wu et al. (2010), Berglund et al. (2017). For an extensive evaluation of many different leak scenarios, quantitative metrics are required that reflect the overall performance of the leak localization method and consider the size of the identified leakage hotspot area. In Soldevila et al. (2016) and Soldevila et al. (2017), two separate metrics are used for training a leak classifier and for evaluating the success of the leak localization. The two separate metrics account for either the number of leak candidates and correct classification of the true leak node or for the average distance of candidate nodes to the true leak node. In this article, a single metric is formulated that combines these properties and enables a quantitative comparison of the performance of leak localization methods.

To the best of the authors' knowledge, previous literature on leak localization in water distribution networks has not compared the performance of methods based on an inverse problem formulation and methods using a sensitivity matrix. The localization performance metric is therefore applied to evaluate both the proposed method based on the inverse problem and the sensitivity matrix method in an extensive benchmarking study, which considers different single-leak scenarios. In addition, the proposed inverse problem with regularization is investigated for scenarios with multiple simultaneous leaks. The leak localization methods are tested on two case study networks, which include a published benchmark network and an operational water network from the UK.

The sections in the remainder of this article, which also outline the main contributions, are organized as follows:

- We propose a formulation of the inverse problem for leak localization that includes a regularization term to deal with the under-determined problem and that is solved using smooth mathematical optimization methods.
- We outline the leak localization method based on the sensitivity matrix.
- We propose a novel quantitative metric to evaluate the performance of leak localization methods.
- The proposed inverse problem with regularization is extensively evaluated using two different case studies with a single leak and two simultaneous leaks. An approach is described and assessed for the selection of the regularization parameter in a large scale operational network. Furthermore, the localization performance of the proposed inverse method with regularization is benchmarked against the sensitivity matrix method.

INVERSE PROBLEM FORMULATION

In this manuscript, the problem of identifying leak candidates (leakage hotspot area/s) in a water distribution network is formulated as an inverse problem, where the unknown leak parameters are estimated from pressure and flow measurements. A major advantage of this formulation compared to other methods is that the leak parameters can be estimated for multiple simultaneously occurring leaks. The problem formulation requires a calibrated extended-period-simulation model of the network, with known pipe properties, known demand profiles, and known settings of pressure control valves.

The water distribution network is modeled as a directed graph with n_p edges (pipes), n_n demand nodes (junctions) and n_0 reservoir nodes which are the sources to the network. The edge-demand node incidence matrix of the graph is $\mathbf{A}_{12} \in \mathbb{R}^{n_p \times n_n}$. Analogously, we define the edge - source node incidence matrix $\mathbf{A}_{10} \in \mathbb{R}^{n_p \times n_0}$.

Leak flow depends on pressure, and their relation has been extensively studied in previous literature, resulting in a modified orifice equation, e.g. (Clayton and Van Zyl 2007; Greyvenstein and Van Zyl 2007; Van Zyl et al. 2017). To model the pressure-dependent leak flow $d_{L,i}(h_i)$ at a

node $i \in \{1, \dots, n_n\}$, this is written as

$$d_{L,i}(h_i) = c_i(h_i - z_i)^{0.5} + c_{var,i}(h_i - z_i)^{1.5}, \quad (1)$$

where $0 \leq c_i \leq c_{\max}$ is an unknown coefficient depending on the leak orifice area and shape and $0 \leq c_{var,i} \leq c_{var,\max}$ is an unknown coefficient depending on the pressure sensitivity of the orifice area. The hydraulic head at node i is h_i , and $z_i \geq 0$ is the node elevation. In the present work, we use the orifice equation with $c_{var,i} = 0$ for all $i \in \{1, \dots, n_n\}$. However, the problem formulation and solution methods discussed here can be extended to different leakage models.

The head losses due to friction across a pipe $j \in \{1, \dots, n_p\}$ depend on the flow rate q_j in the pipe and are often modeled using a Hazen-Williams or a Darcy-Weisbach formula. The expression for the Hazen-Williams model contains a fractional exponent and is non-smooth. In case of the Darcy-Weisbach model, the head losses are given by implicit equations which involve non-smooth terms. Both expressions hinder the application of smooth optimization methods. Therefore, a quadratic approximation is used to model the head losses, as suggested in Pecci et al. (2017). Using the quadratic approximation, the head loss due to friction across pipe j is

$$\Phi_j(q_j) = q_j(a_j|q_j| + b_j), \quad (2)$$

where $a_j, b_j \geq 0$ are computed for all $j = \{1 \dots n_p\}$ as outlined in Eck and Mevissen (2015). The quadratic approximation introduces inaccuracies with respect to the traditional head loss models. However, the accuracy can be controlled by tuning the approximation domain for each pipe, where the expected flow range is determined from hydraulic simulations. The resulting approximation errors are comparable to the level of uncertainty usually observed in hydraulic models of operational water distribution systems (Pecci et al. 2017).

Let $\mathbf{d}_L(\mathbf{c}, \mathbf{h}) \in \mathbb{R}^{n_n}$ be the vector of pressure-dependent leakage at each node in the network, whose components are defined in Eq. (1). Moreover, $\Phi(\mathbf{q}) \in \mathbb{R}^{n_p}$ is a vector whose elements represent the head losses due to friction across network pipes. Pipe flows $\mathbf{q} \in \mathbb{R}^{n_p}$, node heads

$\mathbf{h} \in \mathbb{R}^{n_n}$ and leak coefficient vectors $\mathbf{c} \in \mathbb{R}^{n_n}$ are subject to the steady-state hydraulic network equations:

$$\mathbf{A}_{12}^T \mathbf{q} - \mathbf{d} - \mathbf{d}_L(\mathbf{c}, \mathbf{h}) = \mathbf{0} \quad (3)$$

$$\mathbf{A}_{12} \mathbf{h} + \mathbf{A}_{10} \mathbf{h}_0 + \Phi(\mathbf{q}) = \mathbf{0}, \quad (4)$$

with $\mathbf{h} \geq \mathbf{z}$ where $\mathbf{z} \in \mathbb{R}^{n_n}$ is the vector of node elevations. In the present work, the customer demand is modelled as non pressure-dependent, denoted by vector $\mathbf{d} \in \mathbb{R}^{n_n}$. The vector $\mathbf{h}_0 \in \mathbb{R}^{n_0}$ contains the fixed head values at the source nodes. Eq. (3) corresponds to the mass conservation law, while the energy conservation law is presented in Eq. (4).

Denote by M the set of node indices where the pressure head is measured, and by O the set of pipe indices where flow is measured. The head measurements are then $\bar{h}_m$ for all $m \in M$ and the flow measurements are $\bar{q}_o$ for all $o \in O$. A residual is the difference between sensor measurements and simulated heads and flows.

The aim is to formulate an inverse problem where the weighted sum of squared residuals is minimized over $(\mathbf{h}, \mathbf{q}, \mathbf{c})$ satisfying Eqs. (3) and (4). Since $|M| + |O| \ll n_n$, the inverse problem is ill-posed and some leak locations may not be distinguishable from other possible leak locations as discussed in Pudar and Liggett (1992). This means that several possible leak nodes could result in similar measurements and are not observable individually.

Numerical schemes to solve nonlinear ill-posed inverse problems significantly benefit from the inclusion of a regularization term within the objective function (Aster et al. 2005, Chapter 10). In this manuscript, we consider a ℓ_2 -regularization, which is often applied when fitting data with a large number of model parameters (James et al. 2013, Chapter 6.2). The problem of estimating the unknown leak coefficient vector $\mathbf{c} \in \mathbb{R}^{n_n}$ is hence formulated as an ℓ_2 -regularized optimization problem:

$$\begin{aligned}
& \underset{\mathbf{h}, \mathbf{q}, \mathbf{c}}{\text{minimize}} && w_h \sum_{m \in M} (h_m - \bar{h}_m)^2 + w_q \sum_{o \in O} (q_o - \bar{q}_o)^2 + \rho \|\mathbf{c}\|_2^2 \\
& \text{subject to} && \mathbf{A}_{12}^T \mathbf{q} - \mathbf{d} - \mathbf{d}_L(\mathbf{c}, \mathbf{h}) = \mathbf{0} \\
& && \mathbf{A}_{12} \mathbf{h} + \mathbf{A}_{10} \mathbf{h}_0 + \Phi(\mathbf{q}) = \mathbf{0} \\
& && \mathbf{h} \geq \mathbf{z} \\
& && \mathbf{0} \leq \mathbf{c} \leq \mathbf{c}_{\max}.
\end{aligned} \tag{5}$$

The weights w_h, w_q ensure that flow and head residuals are scaled similarly. The objective function in Problem (5) includes a regularization term $\rho \|\mathbf{c}\|_2^2$. The regularization term penalizes the ℓ_2 -norm of the vector of leak coefficients, to limit the number of non-zero coefficients. We refer the reader to Case study I in the remainder of this article, for a discussion of the benefits of the ℓ_2 -regularization in the case of leak localization.

The solution of Problem (5) depends on the regularization parameter. To select the regularization parameter $\rho \geq 0$, cross-validation is commonly applied (James et al. 2013, Chapter 6.2.3). In this work, Problem (5) is solved for different values of ρ and using data generated by simulating a small subset of all possible single leak scenarios that represent different leak locations in a water supply network. The value of ρ that yields the best results is then selected and cross-validated on a different set of leak scenarios. In this article, the proposed strategy for choosing ρ is investigated for a large operational water network as a case study.

In the following, a solution to Problem (5) is denoted by $(\mathbf{h}^*, \mathbf{q}^*, \mathbf{c}^*)$. Problem (5) has $2 \cdot n_n + n_p$ variables and $n_p + n_n$ equality constraints. The problem is non-convex due to constraints (3) and (4). Since the size of the problem grows linearly with the number of pipes and nodes, Problem (5) results in a large scale nonconvex optimization problem when large operational water networks are considered. However, water distribution networks have a sparse structure which is retained by the hydraulic network equations, Eqs. (3) and (4). Therefore, interior point methods offer a scalable approach for solving Problem (5) for large scale water distribution networks (Nocedal and Wright 2006, Chapter 19).

THE SENSITIVITY MATRIX METHOD VIA THE IMPLICIT FUNCTION THEOREM

The leak localization approach based on the sensitivity matrix, described in Casillas et al. (2013), Perez et al. (2014), is outlined in the following, to enable a comparison between the approach based on the inverse problem developed in the previous section and the sensitivity matrix method. While the sensitivity matrix was previously derived by numerically approximating the derivatives, (Casillas et al. 2013; Perez et al. 2014), it is computed here using the implicit function theorem. This results in exact derivatives.

Denote a constant leak flow from node k by ϵ_k . The set of node indices $M = \{m_1 \dots m_M\}$ indicates where the hydraulic head is measured. The sensitivity matrix $\mathbf{S} \in \mathbb{R}^{|M| \times n_n}$ is defined as in Perez et al. (2014):

$$\mathbf{S} = \begin{bmatrix} \frac{\partial h_{m_1}}{\partial \epsilon_1} & \frac{\partial h_{m_1}}{\partial \epsilon_2} & \dots & \frac{\partial h_{m_1}}{\partial \epsilon_{n_n}} \\ \vdots & & \ddots & \vdots \\ \frac{\partial h_{m_M}}{\partial \epsilon_1} & \dots & & \frac{\partial h_{m_M}}{\partial \epsilon_{n_n}} \end{bmatrix}. \quad (6)$$

The i th column of $\mathbf{S}$ consists of the elements $\frac{\partial h_m}{\partial \epsilon_i}$ for all $m \in M$, i.e. the sensitivities of the measurements to a leak at node i .

Denote by $\mathbf{x} = [\mathbf{q}^T \mathbf{h}^T]^T \in \mathbb{R}^{n_p + n_n}$ the state variables of the water network. A non-pressure-dependent leak flow vector $\boldsymbol{\epsilon} \in \mathbb{R}^{n_n}$ is introduced and the mass conservation Eq. (3) is re-written as

$$\mathbf{A}_{12}^T \mathbf{q} - \mathbf{d} - \boldsymbol{\epsilon} = \mathbf{0}. \quad (7)$$

Eqs. (4) and (7) is written in a more compact form as $\mathbf{g}(\boldsymbol{\epsilon}, \mathbf{x}) = \mathbf{0}$, where $\mathbf{g} : \mathbb{R}^{n_n + n_p + n_n} \rightarrow \mathbb{R}^{n_p + n_n}$ is opportunely defined. The Jacobian of the hydraulic network equations with respect to $\mathbf{x}$ is

$$\mathbf{J}_g(\mathbf{x}) = \frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\boldsymbol{\epsilon}, \mathbf{x}) = \begin{bmatrix} \mathbf{D}(\mathbf{q}) & \mathbf{A}_{12} \\ \mathbf{A}_{12}^T & \mathbf{0} \end{bmatrix}. \quad (8)$$

Matrix $\mathbf{D}(\mathbf{q})$ is diagonal and its elements are the derivatives of the head losses due to friction,

$\Phi(\mathbf{q})$, with respect to q ,

$$D_{jj}(q_j) = \frac{\partial \Phi_j}{\partial q_j} = 2a_j|q_j| + b_j. \quad (9)$$

As $a_j, b_j \geq 0$, $\mathbf{D}(\mathbf{q})$ is positive semi-definite. Furthermore, the edge-demand node incidence matrix $\mathbf{A}_{12}$ has full column rank as proven in Elhay et al. (2014). The matrix $\mathbf{J}_g$ is then nonsingular if and only if $\ker(\mathbf{D}(\mathbf{q})) \cap \ker(\mathbf{A}_{12}^T) = \{0\}$ (Benzi et al. 2005, Theorem 3.2). This condition is verified if there is no loop with all flow rates equal zero (Nielsen 1989).

If $\mathbf{J}_g$ is nonsingular, the implicit function theorem holds (Dontchev and Tyrrell Rockafellar 2009, Theorem 1B.1). There exists a unique, continuously differentiable function $\mathbf{y}(\cdot)$ mapping the leak flow to the state variables $\mathbf{x}$. The derivatives of $\mathbf{y}(\cdot)$ with respect to the leak flow ϵ , i.e. the sensitivities, can be obtained by solving

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\epsilon, \mathbf{x}) \frac{\partial \mathbf{y}}{\partial \epsilon}(\epsilon) = -\frac{\partial \mathbf{g}}{\partial \epsilon}(\epsilon, \mathbf{x}) \quad (10)$$

for $\frac{\partial \mathbf{y}}{\partial \epsilon}(\epsilon) \in \mathbb{R}^{(n_p+n_n) \times n_n}$. The elements of the sensitivity matrix can then be extracted from $\frac{\partial \mathbf{y}}{\partial \epsilon}(\epsilon)$.

Note that $\frac{\partial \mathbf{g}}{\partial \epsilon}^T(\epsilon, \mathbf{x}) = [\mathbf{0} - \mathbf{I}] \in \mathbb{R}^{n_n \times (n_p+n_n)}$ where $\mathbf{I} \in \mathbb{R}^{n_n \times n_n}$ is the identity matrix.

Let $\mathbf{p}^T = [(h_{m_1} - \bar{h}_{m_1}) \dots (h_{m_M} - \bar{h}_{m_M})] \in \mathbb{R}^{|M|}$ be the vector of residuals where $\bar{h}_m$ are the hydraulic head measurements for all $m \in M$. The corresponding hydraulic heads in a no-leak scenario, denoted by h_m for all $m \in M$, are computed by solving the hydraulic network equations, Eqs. (4) and (7) with $\epsilon = \mathbf{0}$.

In the sensitivity matrix method, a column i of the sensitivity matrix $\mathbf{S}$ is then compared with the residual vector $\mathbf{p}$. In the literature, different similarity measures for the comparison have been discussed, e.g. correlation in Perez et al. (2014). According to Casillas et al. (2013), using the angle α_i between the two vectors yields the best localization results. The node with index l such that $l = \arg \min_{i \in \{1 \dots n_n\}} \alpha_i$ is the main leak candidate. Since the sensitivity matrix method relies on the comparison of a residual vector and the sensitivities with respect to one leak only, it is not able to locate leaks that occur simultaneously at different locations (Perez et al. 2014).

PERFORMANCE METRIC FOR LEAK LOCALIZATION

The problem for leak localization is under-determined, and consequently, localization methods may fail to isolate the true leak node. However, they can identify a set of candidate leak nodes, which include the true leak node and delimit a leakage hotspot area. The nodes within a successfully identified leakage hotspot area are close to the true leak node and include the true leak node. In order to be able to investigate the effectiveness of a specific leak localization method and benchmark its performance against other methods, a single metric is required to take into account the size of the leakage hotspot area (e.g. the distance of the candidate leak nodes from the true leak node) and the successful inclusion of the true leak node within the leakage hotspot area (e.g. the correct classification of the true leak node). Metrics, that consider the correct classification of the true leak node and distance of leak node candidates separately were described in previous literature (Soldevila et al. 2016; Soldevila et al. 2017). However, to quantitatively assess the overall success of a leak localization method, these aspects need to be jointly evaluated. We propose a new leak localization performance metric that takes into account the size of the leakage hotspot area and the successful identification of the true leak node within the leakage hotspot area.

The proposed localization performance metric can be applied to evaluate the performance of any leak localization method that attributes a value u_i to each node, as an indicator for leakage. This is the case for both leak localization methods considered here. The performance metric can also be applied to evaluate the localization success in scenarios with multiple leaks.

The attributes u_i classify all nodes into non-leak nodes and potential leak candidates. To define the performance metric the values u_i are normalized such that $0 \leq \hat{u}_i \leq 1$ for all $i \in \{1 \dots n_n\}$ where a large $\hat{u}_i$ indicates leakage. As an example, if Problem (5) is solved, the estimated leak coefficients c_i^* are selected as attributes u_i . The normalized attribute at node l is then $\hat{u}_l = (\max_{i \in \{1 \dots n_n\}} c_i^*)^{-1} c_l^*$.

Let $\hat{\mathbf{u}} \in \mathbb{R}^{n_n}$ be the vector of normalized attributes. The proposed performance metric β is defined as the sum of a reward component γ and a penalty component λ :

$$\beta = \gamma - \lambda. \quad (11)$$

The reward is defined as

$$\gamma = \frac{\sum_{i \in K} \hat{u}_i}{|K|}, \quad (12)$$

where K denotes the set of node indices that correspond to the true leak nodes and $|K|$ denotes the number of leak nodes. It holds that $0 \leq \gamma \leq 1$, where $\gamma = 1$ if the leak localization method assigns $\hat{u}_i = 1$ for all $i \in K$, and $\gamma = 0$ if $\hat{u}_i = 0$ for all $i \in K$. The penalty is defined as

$$\lambda = \frac{\mathbf{r}^T \hat{\mathbf{u}}}{\sum_{i=1}^{n_n} r_i}, \quad (13)$$

where the element r_i of the distance vector $\mathbf{r} \in \mathbb{R}^{n_n}$ is the length of the shortest path from node i to the nearest leak node. The length of the shortest path is the minimal sum of pipe lengths between two nodes, measured in meters. For all $i \in K$, it holds that $r_i = 0$. Therefore, when $\hat{u}_i = 0$ for all $i \notin K$, the penalty is $\lambda = 0$. In comparison, when $\hat{u}_i = 1$ for all $i \notin K$, the penalty is $\lambda = 1$. Hence, it holds that $0 \leq \lambda \leq 1$.

Fig. 1 illustrates the developed performance metric on a simplified example. The metric can result in any value within $-1 \leq \beta \leq 1$. If $\beta = 1$, we have the maximum reward and no penalty, as the true leak is identified. If $0 < \beta < 1$, the normalized attribute at the leak nodes is large, but other nodes are leak candidates as well. While $\beta = 0$ can be interpreted as the output of the leak localization method not giving useful information on the leak location, $\beta < 0$ means that the method is guiding the leak search towards the wrong direction. Note that candidate nodes far away from the true leak node result in greater penalty than candidate nodes of equal $\hat{u}$ that are close to it. If the values for $\hat{u}_i$ are selected such that they may take any value between 0 and 1, the metric additionally takes into account a ranking of the leak candidates. Given a positive reward $\gamma > 0$, the true leak node is within the leak candidate set and a small penalty λ then indicates a small leakage hotspot area.

NUMERICAL INVESTIGATION

The proposed leak localization method based on the solution of the inverse problem, formulated in Problem (5), is numerically evaluated using two case studies and considering different leak

scenarios, under the assumption of exact measurements and known model. In Case study I, the effect of the regularization in Problem (5) is investigated. In both case studies, we compare the proposed inverse problem approach with the sensitivity matrix method (Casillas et al. 2013; Perez et al. 2014). A strategy for choosing the regularization parameter ρ for a large operational water network is presented in Case study II. In addition, the performance of the developed inverse problem approach in scenarios with multiple leaks is investigated - see Case study II.

In both case studies, leaks are simulated using Eq. (1) with a leak coefficient of $c_i = 0.6 * 10^{-3} [\text{m}^{2.5}\text{s}^{-1}]$ ($c_{var,i} = 0 [\text{m}^{1.5}\text{s}^{-1}]$) for $i \in K$. The simulated leak scenarios consider different leak locations that are selected as evenly distributed across the networks. Hydraulic head and flow measurements are then generated by solving Eqs. (3) and (4) as in Elhay et al. (2016). Since this article focuses on the solution of the under-determined inverse problem formulated in Problem (5), no uncertainty was included in the simulations. The measurements are assumed to be exact and nodal demands and pipes properties are known. Investigating the effect of uncertainty at different times during a day remains a subject for further work. For convenience, all simulations consider only one time step of the extended period simulation. The demand and reservoir levels correspond to those at 12 pm.

Let V be the set of investigated leak scenarios. For each scenario, the performance of the implemented methods is evaluated using the performance metric proposed in the previous section. Let β_v be the leak localization performance when scenario v is considered, for all $v \in V$. The percentage of scenarios, $\mathcal{P}(\tau)$, with a localization performance β greater or equal a threshold value τ is defined as

$$\mathcal{P}(\tau) = 100 \frac{|\{v \in V : \beta_v \geq \tau\}|}{|V|}. \quad (14)$$

The overall performance for set V is then derived from $\mathcal{P}(\tau)$ and is illustrated as in the example in Fig. 2. We will refer to this figure as performance profile in the following. In this example, approximately 90% of the considered scenarios have a value of β greater than $\tau = 0$, while 50% have a value greater than $\tau = 0.5$.

Problem (5) is solved using the interior point solver for large scale nonlinear optimization,

IPOPT (v3.12.9), which is implemented in MATLAB through the interface provided by the OPTI Toolbox (Currie and Wilson 2012). The analysis is carried out using a workstation with a 2.4GHz Intel(R) Xeon(R) CPU E5-2665 with 8 cores.

The sensitivity matrix is derived by applying the implicit function theorem. The angle α between residual vector and sensitivity matrix columns is used to select leak candidates (Casillas et al. 2013). To derive the localization performance β in case of the sensitivity matrix method, the normalized attributes are computed as $\hat{u}_l = 1 - (\max_{i \in \{1 \dots n_n\}} |\alpha_i|)^{-1} |\alpha_l|$ for all $l \in \{1 \dots n_n\}$ if not otherwise noted. If the regularized inverse problem is solved, the normalized attributes are $\hat{u}_l = (\max_{i \in \{1 \dots n_n\}} c_i^*)^{-1} c_l^*$ for all $l \in \{1 \dots n_n\}$.

Case study I

Problem (5) is solved for leak localization in Net25, a published benchmark water network model whose layout is presented in Fig. 3. The network consists of 22 demand nodes, 3 reservoirs nodes and 37 pipes. Pipe and node properties are as in Vairavamoorthy and Lumbers (1998) and Dai and Li (2014), respectively. Exact hydraulic head measurements are assumed to be available from 4 nodes - see Fig. 3. The total demand at 12 pm is 175.15 [l/s] and the resulting pressure distribution is shown in Fig. 3 .

Firstly, a leak is simulated at node 5, with a corresponding leak flow equal to 3.6 [l/s]. Fig. 4 shows the leak localization results when Problem (5) is solved for different values of ρ . If $\rho = 0$, the inverse problem is not regularized. In Fig. 4a, node 8 is the main leak candidate whereas the normalized attribute assigned to the true leak node is small, but not zero. This yields a localization performance $\beta = 0.1$. If $\rho = 100$, the leak localization performance improves as the true leak node 5 is included in the set of nodes having the largest estimated leak coefficient, resulting in $\beta = 0.8$. If the regularization parameter ρ is too large, the leak localization method highlights a group of nodes far away from the true leak node 5 as can be seen in Fig. 4c. Consequently, the localization performance metric results in $\beta < 0$. Such behavior can be explained by investigating the objective function in Problem (5), see Fig. 5. The objective function represents a trade-off between the ℓ_2 -norm of the residual vector and that of the estimated coefficient vector. If $\rho = 10^2$, the solution

of the optimization problem minimizes the residual and selects a leak coefficient vector that is small in the sense of the ℓ_2 -norm and that still fits the measured data. If $\rho > 10^5$, the ℓ_2 -norm of the residual increases. The optimization problem prioritizes minimizing the leak coefficients over fitting the model to the measurements.

In Fig. 6, the localization performance given different choices of the regularization parameter ρ is investigated for different leak locations. Both β and ℓ_2 -norm of the estimated coefficient vector are shown for leaks simulated at node 5 (3.6 [l/s]), node 14 (3.5 [l/s]) and node 19 (3.6 [l/s]). In case of node 5 and 19, the best performance β is achieved for values of $\rho = 10^2$. In case of the leak node 14, the performance gets worse as soon as $\rho > 0$, but the value of β remains positive for $\rho \leq 10^4$, suggesting that the leak node is still among the leak candidate nodes. Note that, in case of both node 5 and 14, increasing ρ to larger values results in the norm of the estimated leak coefficients approaching zero, and β becoming negative. In case of node 19, the performance does not seem to be affected by a large regularization parameter. Fig. 6 suggests that both the localization performance and the optimal choice of ρ are sensitive to the leak location.

As the leak location is unknown, the regularization parameter needs to be chosen such that an acceptable localization performance is achieved for most leak locations. To select a regularization parameter ρ suitable for localizing leaks in Net25, we simulate a single leak at each of the 22 demand nodes. This results in 22 different leak scenarios. Depending on the pressure at the leak node, the simulated leak flows vary between 3.3 and 3.8 [l/s] with an average flow of 3.7 [l/s]. For each of the scenarios, Problem (5) is solved with different values of ρ , and the leak localization performance β is computed. Fig. 7 reports performance profiles corresponding to different options of ρ , showing that the performance is significantly improved when introducing regularization ($\rho > 0$). A value of $10 \leq \rho \leq 10^4$ results in $\beta > 0.5$ in more than 90% of the scenarios. For $10^2 \leq \rho \leq 10^4$, the performances are comparable. Based on the previous discussion (see Fig. 5), $\rho = 10^2$ represents a good compromise between minimizing the differences between model and data, and the size of the leak coefficients. Therefore, in the remainder of this section, we investigate the leak localization performance of the proposed method when solving Problem (5) with $\rho = 10^2$.

For each of the 22 considered leak scenarios, Problem (5) is solved with different values of the regularization parameter. The solver IPOPT required on average 26 iterations and a CPU time of 0.2s to compute a solution. Since the optimization problem can be solved efficiently, choosing a regularization parameter by testing all leak locations in this network requires a small computational effort.

The sensitivity matrix method is applied for leak localization in all 22 leak scenarios defined above. To compute the performance β in a leak scenario, see Eqs. (11)-(13), we first set the elements of the vector $\hat{\mathbf{u}}$ such that $\hat{u}_l = 1$ only for the node l with the maximum normalized attribute, $l = \operatorname{argmax}_{i \in \{1 \dots n_n\}} 1 - (\max_{t \in \{1 \dots n_n\}} |\alpha_t|)^{-1} |\alpha_i|$, and $\hat{u}_i = 0$ for $i \in \{1 \dots n_n\}$, $i \neq l$. The main leak candidate is then the only candidate (Casillas et al. 2013). In Fig. 8, the resulting localization performance profile is given. The leak is correctly isolated in 60% of the scenarios, resulting in $\beta = 1$. However, the true leak node is discarded in the remaining 40% of scenarios, yielding $\beta < 0$. Since it is preferable that the leak node is among the leak candidates, the performance β of the sensitivity matrix method is evaluated considering the normalized attributes at all nodes in the remainder of this paper. In Fig. 8, the resulting performance profile is shown, in addition to the profile obtained when solving Problem (5) with $\rho = 10^2$.

Both methods, the sensitivity matrix method evaluated based on all node attributes and the method based on the regularized inverse problem, result in large, positive performance values and hence identify the leak node within the candidate set. The method based on the regularized inverse problem has a higher performance on average than the sensitivity matrix method, but the variance of its performance is greater. The different performances of the two methods can be further assessed by separating β into the reward and the penalty component. In all tested scenarios, the sensitivity matrix method attributes the largest values to a set of nodes containing the leak node. This results in a very high reward, greater than 0.997, in all scenarios. However, in most cases, the method also assigns a high attribute to a large portion of non-leak nodes, yielding an average penalty of 0.45. The penalty and reward component result in an average performance of 0.55, as shown in in Fig. 8. In comparison, the method based on the regularized inverse problem selects few nodes as candidate

nodes with positive attributes. Hence, the penalty component is small, with an average value of 0.08 and a maximum of 0.33. In some cases, the largest attributes do not correspond to the leak node and the reward experiences a large variation across the different scenarios, with values between 0.24 and 1. This results in the large variation of the overall performance as shown in Fig. 8. In conclusion, the method based on the regularized inverse problem achieves smaller average reward values than the sensitivity matrix method, but it results in better overall performance in the majority of the scenarios (e.g. a smaller leakage hotspot area with correctly included true leak node) - see Fig. 8. The small penalty for the inverse leak localization method with regularization results in a smaller leakage hotspot area, which is particularly beneficial for water operators.

These observations are illustrated in the following example scenarios. The worst performance in case of the sensitivity matrix method occurs for a leak simulated at node 12. Fig. 9 shows the localization results obtained by both methods given a leak at node 12. The sensitivity matrix method assigned the largest attribute to the true leak node. However, large attribute values have been given also to various other nodes. As previously discussed, this results in a large penalty component when computing the performance criterion as in Eq. (11). In comparison, the solution of Problem (5) yields large attributes only for two nodes where the main candidate is not the leak node but its neighbor (Fig. 9a). This results in a slightly better localization performance β and in a smaller penalty component λ .

The best localization performance for the sensitivity matrix method is achieved when a leak is simulated at node 1. As shown in Fig. 10, both approaches clearly identify only nodes 1 and 2 as potential leak candidates, and have similar localization performances.

When solving the regularized inverse problem, the lowest localization performance is experienced by the leak scenario simulated at node 14 (Fig. 11). Fig. 11a shows that the attribute assigned to the true leak node is positive, but close to zero, and the main leak candidate is a neighboring node. On the other hand, the main candidates identified by the sensitivity matrix method are both the true leak node and its neighbor (Fig. 11b). The solution of Problem (5) successfully classifies the majority of nodes as non-leak nodes, resulting in a small set of candidate leak nodes. In com-

parison, the sensitivity matrix method results in large attribute values assigned to many non-leak nodes, and a better isolation can only be achieved by selecting the nodes with the largest attributes.

The solution of Problem (5) results in the best localization performance when a leak is simulated at node 13, see Fig. 12a. The leak node is identified as the single main leak candidate. The sensitivity matrix method in case of a leak at node 13 (Fig. 12b) yields a similar attribute pattern as in the case of a leak at node 14 (Fig. 11b).

The sensitivity matrix method results in positive attributes assigned to most nodes, while the leak node is consistently among the nodes with the highest values. The method based on the regularized inverse problem identifies the true leak node as one of the main leak candidates in the majority of scenarios, but, in few instances, it does not include the true leak node within the set of nodes with the largest attributes. This results in low, but positive, values of β in some scenarios, as shown in Fig. 8. However, the method based on the regularized inverse problem identifies a smaller set of candidate nodes still including the true leak node which yields a more precise leakage hotspot localization. This results in a better overall performance according to the metric β , for the considered case study, see Fig. 8.

Case study II

In the second case study, the leak localization methods are tested using a large operational network, which is jointly operated by Bristol Water, Imperial College London and Cla-Val (BWFLnet, shown in Fig. 13). Details of the network are presented in Wright et al. (2014). It consists of 2310 nodes, 2369 pipes and two fixed head nodes at the inlets. There are three pressure reducing valves (PRV) and two dynamic boundary valves (DBV) which are open at the considered time step, 12 pm. The flow is measured at the 5 valve locations. The hydraulic head is measured at 35 locations (physically installed pressure monitoring devices): inlets and outlets of each valve and 25 further nodes - see Fig. 13. The total demand at 12 pm is 50.1 [l/s] and the pressure distribution is shown in Fig. 13.

In this section, we investigate a strategy to select and validate the value of the regularization parameter ρ when a large water network is considered. Furthermore, we compare the leak lo-

calization performance of the sensitivity matrix method and the method based on the regularized inverse problem which confirms the observations from Case study I. We then further investigate the method based on the regularized inverse problem. By solving Problem (5) for a set of different leak scenarios and then evaluating the localization performance, we identify areas in the network that are challenging for leak localization, given the current sensor placement and the selected regularization parameter. In addition, we study the localization performance in scenarios with two simultaneously occurring leaks.

We define two sets of different leak scenarios. In order to select an appropriate value for the regularization parameter ρ for the formulation of Problem (5), 116 single leaks are generated assuming 116 different leak locations that are evenly distributed across the network. This corresponds to 5% of the nodes. We refer to this set of leak scenarios as (Ia). The resulting leak flows vary between 2.9 and 6.1 [l/s] with an average flow of 4.4 [l/s]. The choice of the regularization parameter is then cross-validated on a set of single leak scenarios (Ib) where 116 leak locations, different from the previous ones, are evenly selected across the network. The average leak flow for scenarios (Ib) is 4.2 [l/s], the minimum leak flow is 2 [l/s] and the maximum 5.9 [l/s].

Fig. 14 reports the localization performance achieved solving Problem (5) for different values of ρ on set (Ia). Similarly to Net25, the localization performance improves with the regularization where $\rho > 0$. Fig. 14 indicates that the best choice is $\rho = 10^9$. The value is much larger than $\rho = 10^2$ which was selected for Net25. We conjecture that this is due to the increased number of possible leak locations. The regularization parameter needs to be selected for each network and sensor configuration individually. However, this selection process requires only a small computational effort.

Solving Problem (5) with $\rho = 10^9$ to locate leaks in set (Ib) yields a leak localization performance profile similar to the profile based on scenarios (Ia), see Fig. 15. This suggests that 5% of all possible leak scenarios are sufficient to choose a good value for ρ .

In addition, Fig. 15 shows the localization performance profile on scenarios (Ib) when using the sensitivity matrix method. Similarly to what was reported for Net25, the method based on the

regularized inverse problem achieves a better localization performance β than the sensitivity matrix method in the majority of scenarios, but also has a lower worst performance. This is in accordance with the results in the case study on Net25, where the sensitivity matrix was shown to yield a very high reward in all scenarios, as well as a large penalty in most scenarios. The method based on the regularized inverse problem on the other hand is able to discard the majority of nodes from the leak candidate set, resulting in a smaller penalty and more precise leakage hotspots. However, in some scenarios, the attribute at the true leak node does not yield a high reward. An example scenario from set (Ib) is shown in Fig. 16. Solving Problem (5) yields a small set of candidate nodes that is close to the leak node, but the maximum attribute is not assigned to the true leak node, see Fig. 16a. In case of the sensitivity matrix method (Fig. 16b), the true leak node is within the set of main candidates, but we also have positive attributes assigned to a large portion of network nodes.

In Fig. 17, it is illustrated that the leak location affects the localization performance for a given regularization parameter in case of the method based on the regularized inverse problem, similar to what was reported for Net25. Note that the area at the bottom right corner results in high localization performance, while experiencing lower pressure (Fig. 13), and hence smaller leak flows. Under the assumption of exact data and exact hydraulic model, and within the range of simulated leak flows, the leak size does not affect the localization performance.

The sensitivity matrix method is limited to the localization of one leak (Perez et al. 2014). For the localization of multiple simultaneous leaks, we only investigate the solution of the inverse problem with regularization (Problem (5)).

In order to generate a set of two-leak scenarios (II), 24 nodes across the network are selected and all possible combinations of two leaks are simulated. This yields 276 scenarios. Then, Problem (5) is formulated and solved for each of these scenarios, using different values of ρ . As shown in Fig. 18, setting $\rho = 10^9$ is still the best choice for the regularization parameter in the multiple-leak case.

The leak location and localization results in case of the two-leak scenario with highest, median and worst localization performance are presented. In the case of the highest performance, both

leaks are in the central area of the network and a set of nodes containing the true leaks has obtained large leak attributes - see Fig. 19. As both leaks in Fig. 19 are in an area that results in high performance values in single leak scenarios, see Fig. 17, the good performance reported for this two-leak scenario is predictable. The median localization performance corresponds to a scenario with one leak at the center of the network and one in the lower left corner - see Fig. 20. Note that one of the two leaks is in an area with medium performance in the single leak case - see Fig. 17. A set of nodes containing the leak node is clearly identified, however the nodes with the highest attributes do not correspond to the true leak nodes. Fig. 21 corresponds to the scenario with low localization performance. In this case, both leaks are in an area corresponding to low localization performance in the single leak case, see Fig. 17. While the true leak nodes are among the candidate nodes with positive attributes, these values are very small and the distinction between candidate nodes and non-leak nodes is not as clear as in the previous examples. However, in all three cases, the majority of nodes with no leakage are correctly assigned attributes equal to zero and the search area for locating the two leaks is significantly reduced.

Solving Problem (5) required on average 41 IPOPT iterations and 9.2s CPU time for a single leak scenario (Ia and Ib). For a two leak scenario (II), we report an average of 40 IPOPT iterations and 8.9s CPU time. Similarly to Net25, Problem (5) is solved efficiently. This suggests that solving the optimization problem several times with different parameters is feasible within a short computational time. Efficiently solving the optimization problem is necessary to choose a regularization parameter, or to consider different uncertainty scenarios for example.

CONCLUSION

Accurate and near real time leak detection and localization is a critical problem for water companies to meet serviceability requirements. In order to solve the ill-posed inverse problem for leak localization, we present a novel optimization formulation, whose main properties are summarized as follows. A regularization term is included in the objective function to address the under-determined inverse problem. The regularization parameter affects the localization performance and needs to be selected via cross-validation, but the method does not require selecting candidate nodes or group-

ing nodes beforehand. The resulting optimization problem is solved efficiently using mathematical optimization and state-of-the-art sparse nonlinear programming solvers.

A localization performance metric is proposed to quantify the success of a leak localization method for different scenarios. The metric returns a scalar value between -1 and 1 to account jointly for the number of leak candidates, their distance to the true leak node and the inclusion of the true leak node within a leakage hotspot area.

The metric is used to select a good regularization parameter via cross-validation. Furthermore, the proposed localization performance metric has enabled an extensive numerical investigation of the proposed method on two case study networks considering a representative set of leak scenarios. The outcome of this investigation is summarized as follows:

- A comparison of sensitivity matrix method and the proposed method based on the regularized inverse problem showed that while the sensitivity matrix method always identifies the true leak as one of the main candidates, the strength of the proposed method lies in discarding the majority of non-leak nodes. In contrast to the sensitivity matrix method, this yields a good isolation of leakage hotspots, which is particularly beneficial for water utilities.
- The method based on the regularized inverse problem is able to localize multiple leaks, similar to other optimization-based methods previously proposed in the literature.
- The performance metric enables the identification of areas within the network where the localization method is most/least successful.
- The proposed approach can be applied to large operational water networks since the problem is solved efficiently and the results presented in this article suggest that only a small subset of possible leak locations (5% of the demand nodes) has to be tested to identify a suitable regularization parameter.

The work presented in this article does not consider demand uncertainty or model errors. Moreover, it was observed that the location of the leaks is an important factor for a successful leak

558 localization. Further work aims therefore at improving and testing the performance of model-based
559 methods for leak localization and includes:

- 560 • the investigation of the performance of the proposed approach given uncertain data and
561 errors as well as the development of strategies to mitigate the effects of uncertainty;
- 562 • the study of the optimal sensor placement with respect to model-based leak localization
563 methods; and
- 564 • the experimental investigation of the proposed leak localization method using hydraulic
565 data acquired from the BWFLnet operational network.

566 DATA AVAILABILITY STATEMENT

567 Code and models generated or used during the study are available from the corresponding
568 author by request.

569 ACKNOWLEDGEMENTS

570 This work has been supported by EPSRC (EP/P004229/1, Dynamically Adaptive and Resilient
571 Water Supply Networks for a Sustainable Future; and, also EP/L016826/1 EPSRC Centre for
572 Doctoral Training in Sustainable Civil Engineering) and Cla-Val UK Ltd.

573 REFERENCES

- 574 Aster, R. C., Borchers, B., and Clifford, H. (2005). *Parameter Estimation and Inverse Problems*.
575 Elsevier, Inc., New York.
- 576 Benzi, M., Golub, G. H., and Liesen, J. (2005). “Numerical solution of saddle point problems.”
577 *Acta Numerica*, 14, 1–137.
- 578 Berglund, A., Areti, V. S., Brill, D., and Mahinthakumar, G. K. (2017). “Successive linear approx-
579 imation methods for leak detection in water distribution systems.” *Journal of Water Resources*
580 *Planning and Management*, 143(8), 04017042.
- 581 Casillas, M. V., Garza-Castanon, L. E., and Puig, V. (2013). “Extended-horizon analysis of pressure
582 sensitivities for leak detection in water distribution networks: Application to the Barcelona
583 network.” *Control Conference (ECC), 2013 European*, IEEE, 401–409.

- Clayton, C. R. I. and Van Zyl, J. E. (2007). "The effect of pressure on leakage in water distribution systems." *Water Management*, 160(June), 109–114.
- Currie, J. and Wilson, D. (2012). "OPTI: Lowering the barrier between open source optimizers and the industrial MATLAB user." *Foundations of Computer-Aided Process Operations*, 24, 32.
- Dai, P. D. and Li, P. (2014). "Optimal localization of pressure reducing valves in water distribution systems by a reformulation approach." *Water Resources Management*, 28(10), 3057–3074.
- Dontchev, A. L. and Tyrrell Rockafellar, R. (2009). *Implicit Functions and Solution Mappings*. Springer Monogr. Math.
- Eck, B. J. and Mevissen, M. (2015). "Quadratic approximations for pipe friction." *Journal of Hydroinformatics*, 17(3), 462–472.
- Elhay, S., Piller, O., Deuerlein, J., and Simpson, A. R. (2016). "A robust, rapidly convergent method that solves the water distribution equations for pressure-dependent models." *Journal of Water Resources Planning and Management*, 142(2), 1–12.
- Elhay, S., Simpson, A. R., Deuerlein, J., Alexander, B., and Schilders, W. H. A. (2014). "Reformulated co-tree flows method competitive with the global gradient algorithm for solving water distribution system equations." *Journal of Water Resources Planning and Management*, 140(12), 04014040.
- Goulet, J. A., Coutu, S., and Smith, I. (2013). "Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks." *Advanced Engineering Informatics*, 27(2), 261–269.
- Greyvenstein, B. and Van Zyl, J. E. (2007). "An experimental investigation into the pressure-leakage relationship of some failed water pipes." *Journal of Water Supply: Research and Technology - AQUA*, 56(2), 117–124.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.
- Nielsen, H. B. (1989). "Methods for analyzing pipe networks." *Journal of Hydraulic Engineering*, 115(2), 139–157.

- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer.
- Ofwat (2019). “Technical Appendix 1 - Delivering outcomes for customers.” *PR19 - Ofwat*, <<https://www.ofwat.gov.uk/publication/technical-appendix-1-delivering-outcomes-for-customers-final/>> (July 15, 2019).
- Pecci, F., Abraham, E., and Stoianov, I. (2017). “Quadratic head loss approximations for optimisation problems in water supply networks.” *Journal of Hydroinformatics*, 19(4), 493–506.
- Perez, R., Sanz, G., Puig, V., Quevedo, J., Cuguero Escofet, M. A., Nejari, F., Meseguer, J., Cembrano, G., Mirats Tur, J., and Sarrate, R. (2014). “Leak localization in water networks: A model-based methodology using pressure sensors applied to a real network in Barcelona [applications of control].” *IEEE Control Systems*, 34(4), 24–36.
- Poulakis, Z., Valougeorgis, D., and Papadimitriou, C. (2003). “Leakage detection in water pipe networks using a Bayesian probabilistic framework.” *Probabilistic Engineering Mechanics*, 18(4), 315–327.
- Pudar, R. S. and Liggett, J. A. (1992). “Leaks in pipe networks.” *Journal of Hydraulic Engineering*, 118(7), 1031–1046.
- Sanz, G. and Pérez, R. (2015). “Sensitivity Analysis for Sampling Design and Demand Calibration in Water Distribution Networks Using the Singular Value Decomposition.” *Journal of Water Resources Planning and Management*, 141(1977), 1–9.
- Sanz, G., Pérez, R., Kapelan, Z., and Savic, D. (2016). “Leak Detection and Localization through Demand Components Calibration.” *Journal of Water Resources Planning and Management*, 142(2), 04015057.
- Soldevila, A., Blesa, J., Tornil-Sin, S., Duviella, E., Fernandez-Canti, R. M., and Puig, V. (2016). “Leak localization in water distribution networks using a mixed model-based/data-driven approach.” *Control Engineering Practice*, 55, 162–173.
- Soldevila, A., Fernandez-Canti, R. M., Blesa, J., Tornil-Sin, S., and Puig, V. (2017). “Leak localization in water distribution networks using Bayesian classifiers.” *Journal of Process Control*, 55, 1–9.

- 638 Steffelbauer, D., Neumayer, M., Günther, M., and Fuchs-Hanusch, D. (2014). “Sensor Placement
639 and Leakage Localization considering Demand Uncertainties.” *Procedia Engineering*, 89(0),
640 1160–1167.
- 641 Vairavamoorthy, K. and Lumbers, J. (1998). “Leakage Reduction in Water Distribution Systems:
642 Optimal Valve Control.” *Journal of Hydraulic Engineering*, 124(11), 1146–1154.
- 643 Van Zyl, J. E., Lambert, A. O., and Collins, R. (2017). “Realistic Modeling of Leakage and Intrusion
644 Flows through Leak Openings in Pipes.” *Journal of Hydraulic Engineering*, 143(9), 3–9.
- 645 Wright, R., Stoianov, I., Pappas, P., Henderson, K., and King, J. (2014). “Adaptive water distribution
646 networks with dynamically reconfigurable topology.” *Journal of Hydroinformatics*, 16(6), 1280–
647 1301.
- 648 Wu, Z. Y., Sage, P., and Turtle, D. (2010). “Pressure-Dependent Leak Detection Model and Its
649 Application to a District Water System.” *Journal of Water Resources Planning and Management*,
650 136(1), 116–128.

List of Figures

- 1 The performance metric β and the penalty λ are illustrated on a small example. . . 30
- 2 Example of a localization performance profile to evaluate different leak scenarios. . 31
- 3 Layout of Net25, with fixed head at nodes 23, 24, 25 (squares) and hydraulic head measurements at nodes 1,8,13,21 (stars). The node colors illustrate the pressure at demand nodes at 12 pm. 32
- 4 Leak localization results and performance β when solving the regularized inverse problem using different values of ρ . Nodes with a large attribute are the main candidates. The true leak was simulated at node 5 (diamond shaped). 33
- 5 Net25: ℓ_2 -norm of the model residual and ℓ_2 -norm of the coefficient vector for different values of ρ . The true leak is simulated at node 5. 34
- 6 Net25: Localization performance β and ℓ_2 -norm of the estimated leak coefficient vector vary for different values of ρ and different leak scenarios. 35
- 7 Net25: Localization performance profiles of the method based on the regularized inverse problem for different choices of ρ . A leak has been simulated at each demand node in order to obtain 22 different scenarios. 36
- 8 Net25: Localization performance profile using the regularized inverse problem formulation (IP-R) with $\rho = 10^2$ and the sensitivity matrix method, considering all attributes (SMM) and only considering the maximum attribute (SMM max). A leak has been simulated at each demand node in order to obtain 22 different scenarios. 37
- 9 Leak at node 12 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates. 38
- 10 Leak at node 1 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates. 39

677	11	Leak at node 14 (diamond shaped): Localization results of the method based on	
678		the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM).	
679		Large attributes indicate the main leak candidates.	40
680	12	Leak at node 13 (diamond shaped): Localization results of the method based on	
681		the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM).	
682		Large attributes indicate the main leak candidates.	41
683	13	BWFLnet: Pressure measurements (stars), fixed head nodes at the two inlets and	
684		flow measurements at the valve locations. The node colors illustrate the pressure	
685		distribution at 12 pm.	42
686	14	BWFLnet: Performance profiles for leak scenarios (Ia) in case of the method based	
687		on the regularized inverse problem with different values of ρ	43
688	15	BWFLnet: The performance profiles for leak scenarios (Ia) and (Ib) in case of	
689		the method based on the regularized inverse problem (IP-R) with $\rho = 10^9$ and the	
690		performance profile for the sensitivity matrix method (SMM) for leak scenarios (Ib). 44	
691	16	BWFLnet: Leak location and localization results of the method based on the	
692		regularized inverse problem (IP-R) and the sensitivity matrix method (SMM).	
693		Large attributes indicate the main leak candidates.	45
694	17	BWFLnet: Leak locations (Ia and Ib) and the localization performance β , derived	
695		from the results of the method based on the regularized inverse problem with	
696		$\rho = 10^9$. Sensor nodes are star-shaped.	46
697	18	BWFLnet: Performance profiles for a set of two-leak scenarios (II) in case of the	
698		method based on the regularized inverse problem formulation with different values	
699		of ρ	47
700	19	BWFLnet: Localization results based on the regularized inverse problem (IP-R,	
701		$\rho = 10^9$) in case of the two-leak scenario with the highest localization performance:	
702		$\beta = 0.9$ ($\lambda = 0.02$). Large attributes indicate the main leak candidates.	48

703	20	BWFLnet: Localization results based on the regularized inverse problem (IP-R,	
704		$\rho = 10^9$) in case of the two-leak scenario with median localization performance:	
705		$\beta = 0.52$ ($\lambda = 0.06$). Large attributes indicate the main leak candidates.	49
706	21	BWFLnet: Localization results based on the regularized inverse problem (IP-R,	
707		$\rho = 10^9$) in case of the two-leak scenario with the worst localization performance:	
708		$\beta = 0.16$ ($\lambda = 0.02$). Large attributes indicate the main leak candidates.	50

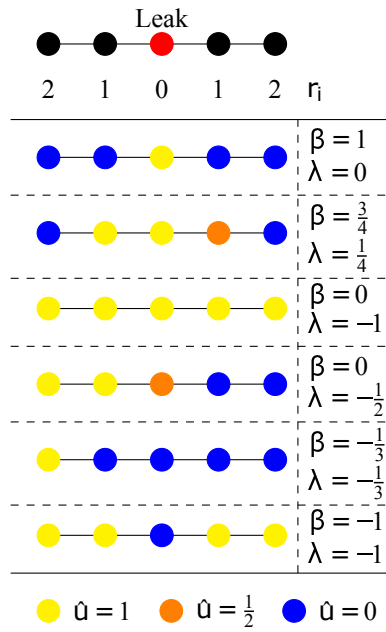


Fig. 1. The performance metric β and the penalty λ are illustrated on a small example.

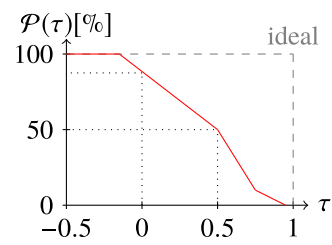


Fig. 2. Example of a localization performance profile to evaluate different leak scenarios.

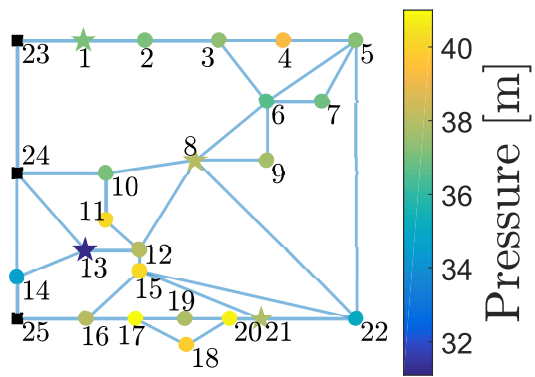
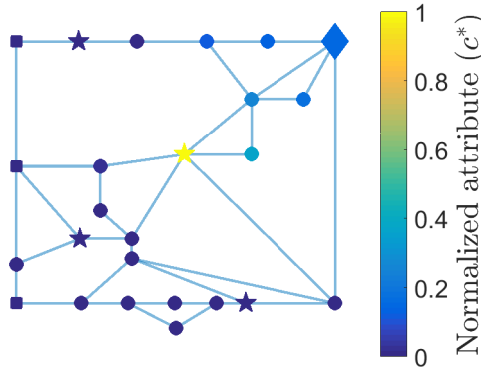
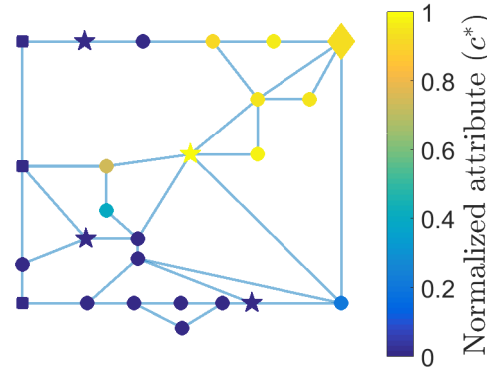


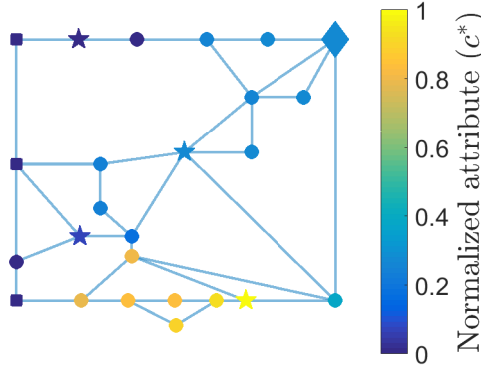
Fig. 3. Layout of Net25, with fixed head at nodes 23, 24, 25 (squares) and hydraulic head measurements at nodes 1,8,13,21 (stars). The node colors illustrate the pressure at demand nodes at 12 pm.



(a) $\rho = 0$: $\beta = 0.1$



(b) $\rho = 10^2$: $\beta = 0.8$



(c) $\rho = 10^9$: $\beta = -0.19$

Fig. 4. Leak localization results and performance β when solving the regularized inverse problem using different values of ρ . Nodes with a large attribute are the main candidates. The true leak was simulated at node 5 (diamond shaped).

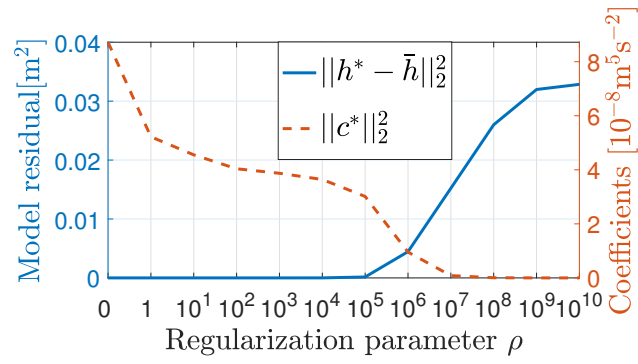
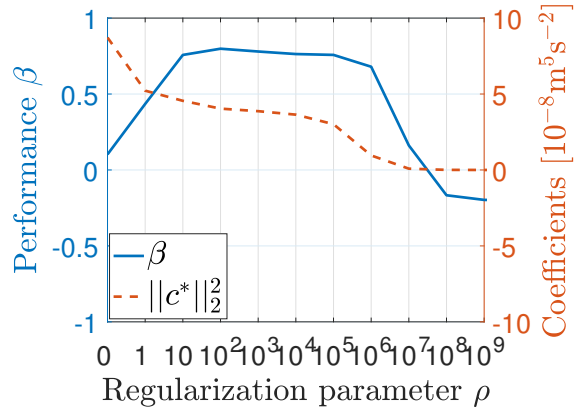
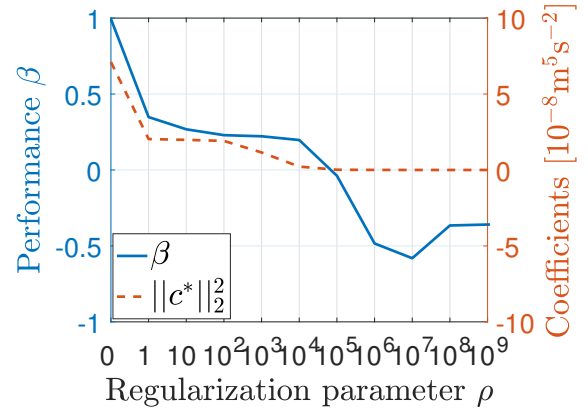


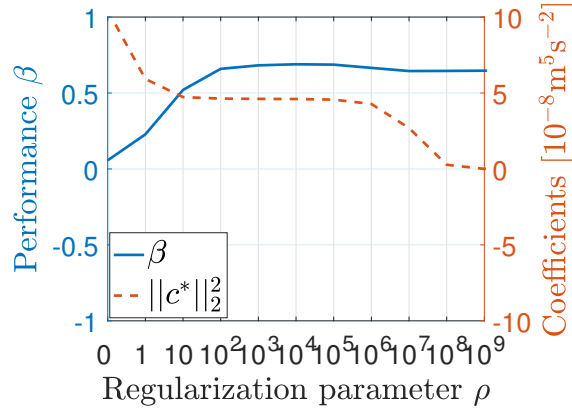
Fig. 5. Net25: ℓ_2 -norm of the model residual and ℓ_2 -norm of the coefficient vector for different values of ρ . The true leak is simulated at node 5.



(a) True leak at node 5



(b) True leak at node 14



(c) True leak at node 19

Fig. 6. Net25: Localization performance β and ℓ_2 -norm of the estimated leak coefficient vector vary for different values of ρ and different leak scenarios.

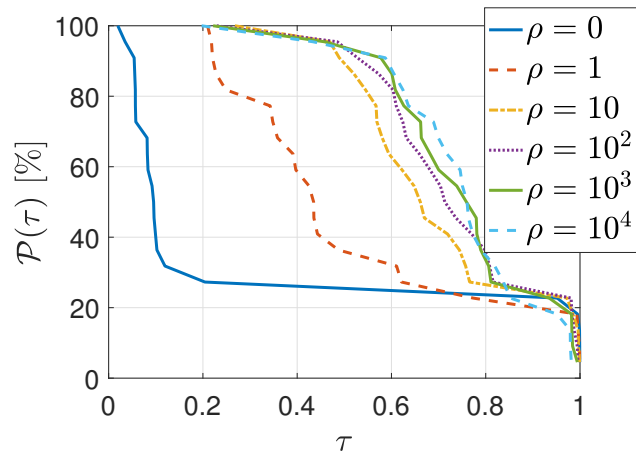


Fig. 7. Net25: Localization performance profiles of the method based on the regularized inverse problem for different choices of ρ . A leak has been simulated at each demand node in order to obtain 22 different scenarios.

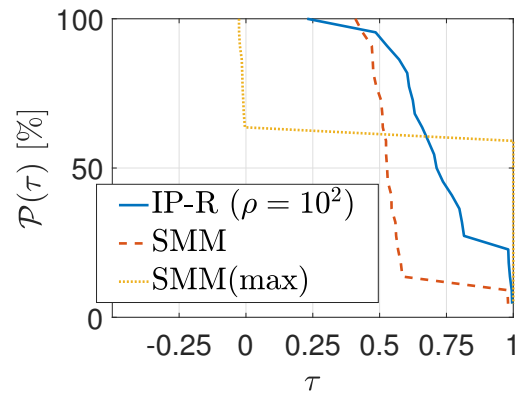
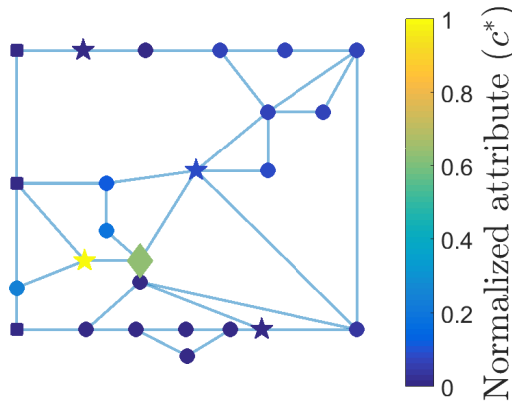
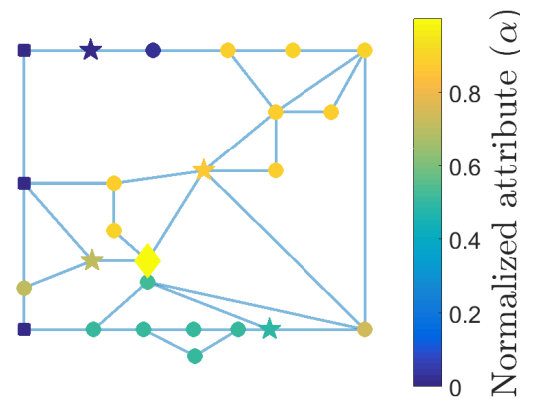


Fig. 8. Net25: Localization performance profile using the regularized inverse problem formulation (IP-R) with $\rho = 10^2$ and the sensitivity matrix method, considering all attributes (SMM) and only considering the maximum attribute (SMM max). A leak has been simulated at each demand node in order to obtain 22 different scenarios.

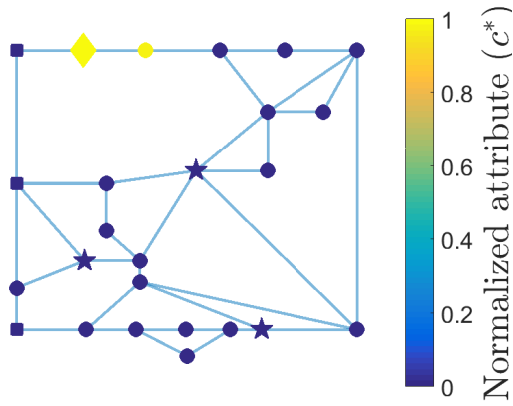


(a) IP-R ($\rho = 10^2$): $\beta = 0.57$ ($\lambda = 0.05$)

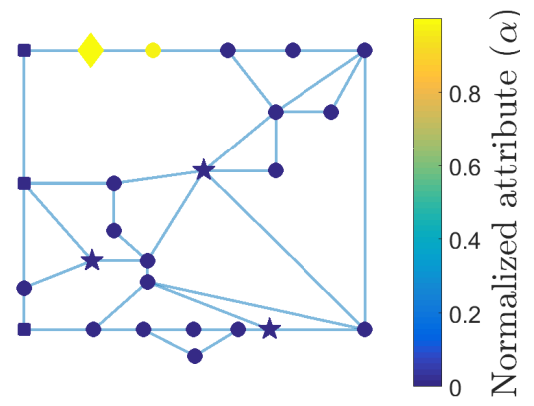


(b) SMM: $\beta = 0.41$ ($\lambda = 0.59$)

Fig. 9. Leak at node 12 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates.

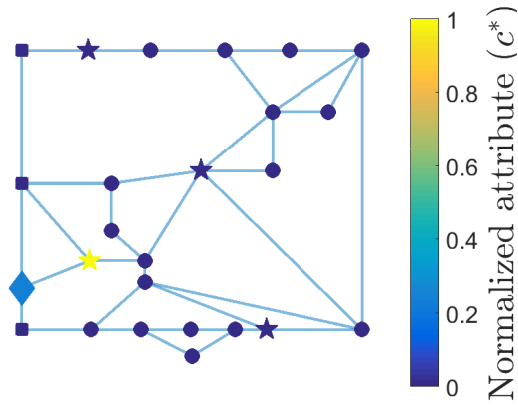


(a) IP-R ($\rho = 10^2$): $\beta = 0.98$ ($\lambda = 0.02$)

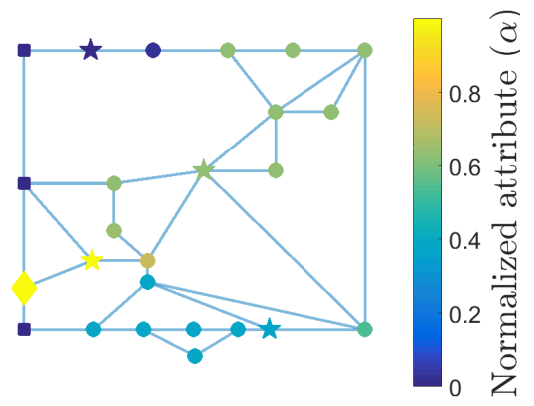


(b) SMM: $\beta = 0.98$ ($\lambda = 0.02$)

Fig. 10. Leak at node 1 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates.

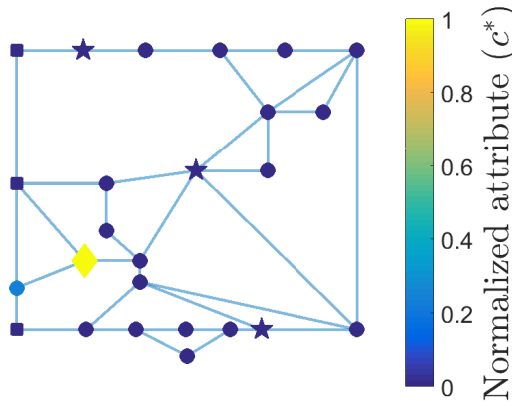


(a) IP-R ($\rho = 10^2$): $\beta = 0.23$ ($\lambda = 0.02$)

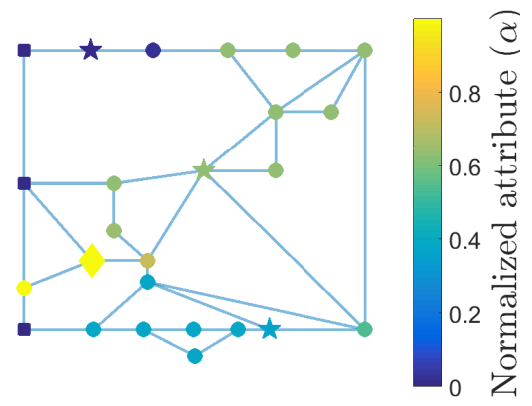


(b) SMM: $\beta = 0.51$ ($\lambda = 0.49$)

Fig. 11. Leak at node 14 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates.



(a) IP-R ($\rho = 10^2$): $\beta = 0.99$ ($\lambda = 0.01$)



(b) SMM: $\beta = 0.53$ ($\lambda = 0.47$)

Fig. 12. Leak at node 13 (diamond shaped): Localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates.

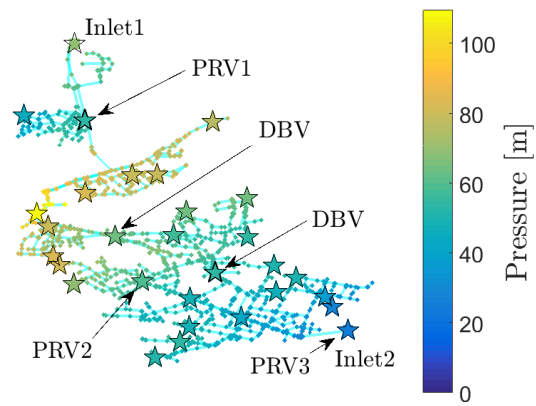


Fig. 13. BWFLnet: Pressure measurements (stars), fixed head nodes at the two inlets and flow measurements at the valve locations. The node colors illustrate the pressure distribution at 12 pm.

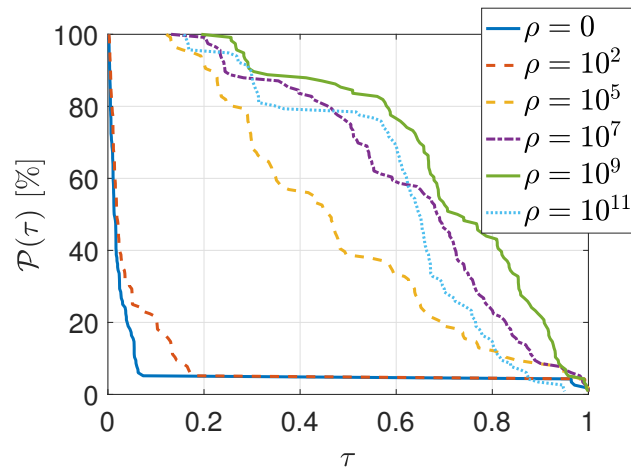


Fig. 14. BWFLnet: Performance profiles for leak scenarios (Ia) in case of the method based on the regularized inverse problem with different values of ρ .

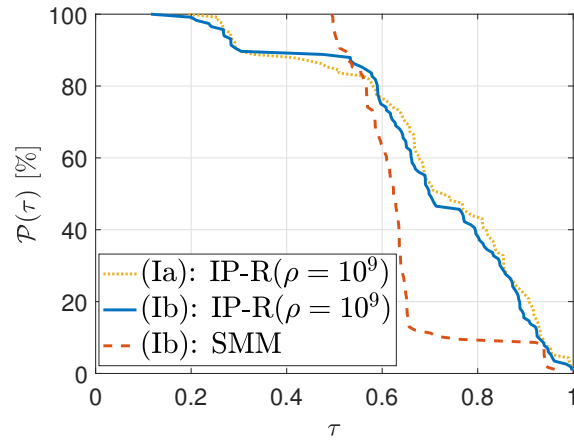


Fig. 15. BWFLnet: The performance profiles for leak scenarios (Ia) and (Ib) in case of the method based on the regularized inverse problem (IP-R) with $\rho = 10^9$ and the performance profile for the sensitivity matrix method (SMM) for leak scenarios (Ib).

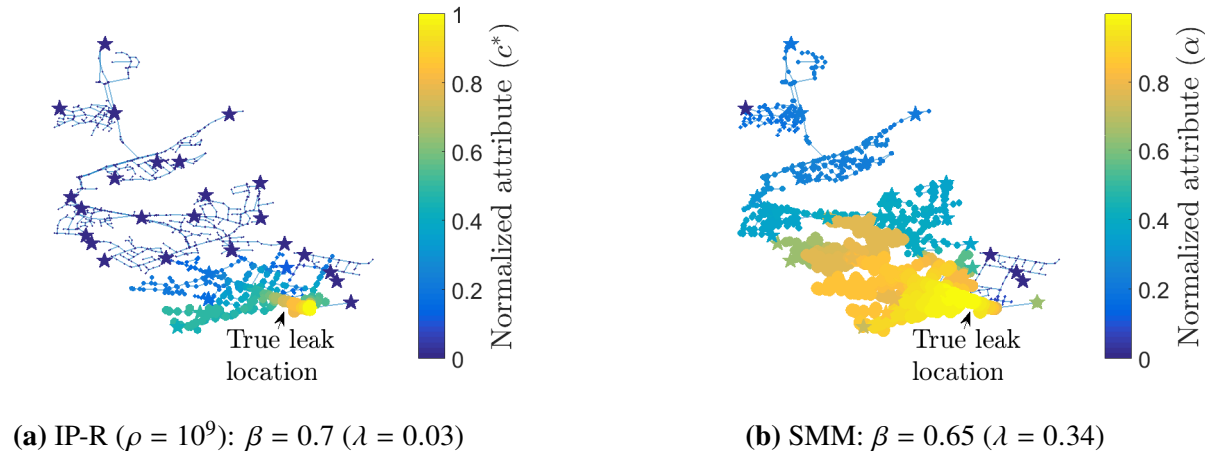


Fig. 16. BWFLnet: Leak location and localization results of the method based on the regularized inverse problem (IP-R) and the sensitivity matrix method (SMM). Large attributes indicate the main leak candidates.

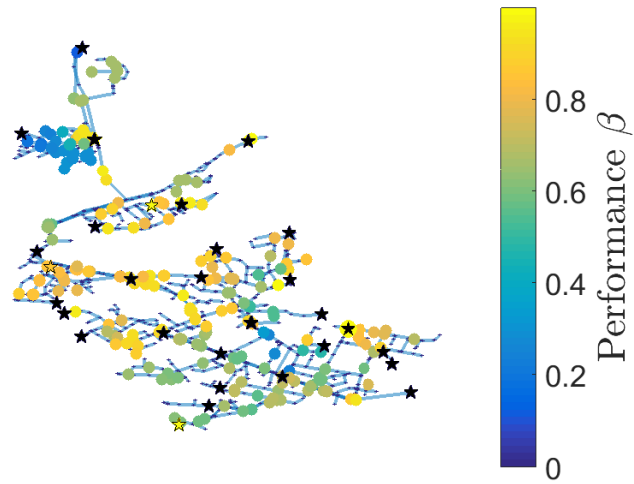


Fig. 17. BWFLnet: Leak locations (Ia and Ib) and the localization performance β , derived from the results of the method based on the regularized inverse problem with $\rho = 10^9$. Sensor nodes are star-shaped.

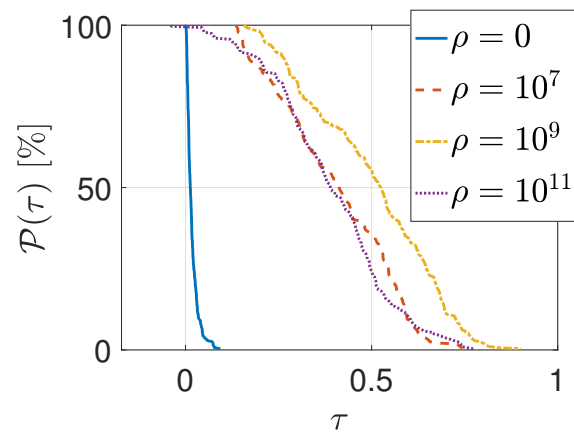


Fig. 18. BWFLnet: Performance profiles for a set of two-leak scenarios (II) in case of the method based on the regularized inverse problem formulation with different values of ρ .

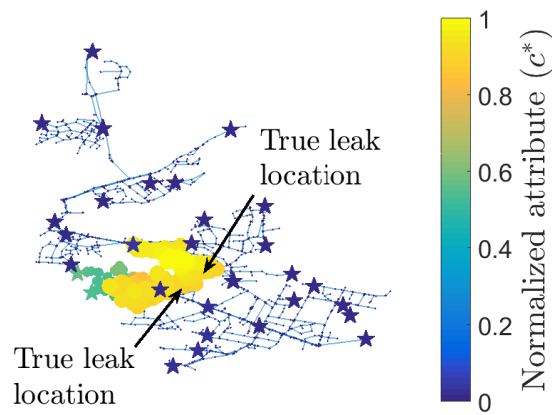


Fig. 19. BWFLnet: Localization results based on the regularized inverse problem (IP-R, $\rho = 10^9$) in case of the two-leak scenario with the highest localization performance: $\beta = 0.9$ ($\lambda = 0.02$). Large attributes indicate the main leak candidates.

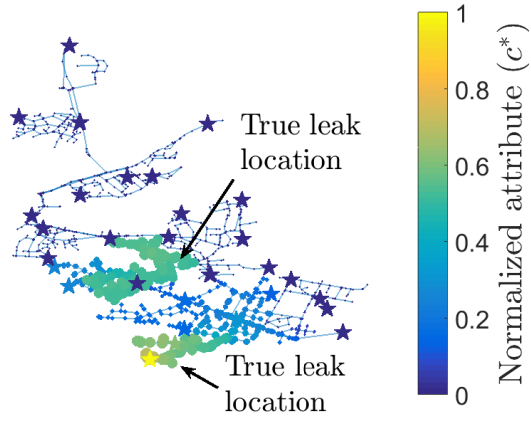


Fig. 20. BWFLnet: Localization results based on the regularized inverse problem (IP-R, $\rho = 10^9$) in case of the two-leak scenario with median localization performance: $\beta = 0.52$ ($\lambda = 0.06$). Large attributes indicate the main leak candidates.

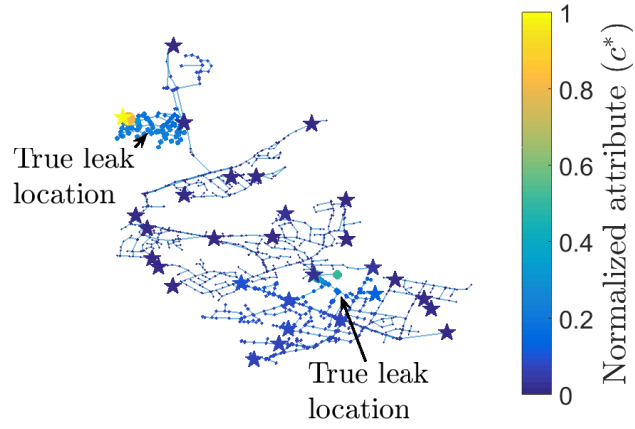


Fig. 21. BWFLnet: Localization results based on the regularized inverse problem (IP-R, $\rho = 10^9$) in case of the two-leak scenario with the worst localization performance: $\beta = 0.16$ ($\lambda = 0.02$). Large attributes indicate the main leak candidates.