Noname manuscript No.
(will be inserted by the editor)

Probabilistic leak detectability assessment via state estimation in water transport networks

Sarai Díaz · Roberto Mínguez · Javier González

This is a post-peer-review, pre-copyedit version of an article published in Stochastic Environmental Research and Risk Assessment.
The final version is available online at Springer via https://doi.org/10.1007/s00477-018-1515-3

Abstract Leak detectability or leakage awareness refers to the capability of sensing losses from a water supply system. Several methods exist in the technical literature to tackle this problem, but only few address it with a state estimation approach. The aim of this paper is to present a new methodology that enables probabilistic assessment of the extent to which water loss could be detected using state estimation by only analysing a single hydraulic state, i.e. one time period. Significant leaks are sensed by identifying unusually high normalised state estimation residuals, which can be identified based on the largest normalised residual test. More specifically, the probability of detecting leaks is computed here by working with the multivariate distribution among measurements and estimates to take into account the noisy nature of measurements with an analytical approach rather than with sampling experiments, which are time-consuming. The methodology set out herein also provides a procedure to systematically assess the minimum leak that could be detected in different parts of the network for a specific measurement setting and operating condition. The method has been applied to a water transport network case study to show its potential and to highlight the usefulness of such a tool for practitioners. The limitations of such a

Sarai Díaz
Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain).
Tel.: +34 926 295 300 (Ext. 96635)
ORCID: orcid.org/0000-0002-5478-1768
E-mail: Sarai.Diaz@uclm.es

Roberto Mínguez
HIDRALAB INGENIERÍA Y DESARROLLOS, S.L., Spin-Off UCLM, Hydraulics Laboratory Univ. of Castilla-La Mancha, Av. Pedriza, Camino Moledores s/n, 13071 Ciudad Real (Spain)
E-mail: roberto.minguez@hidralab.com

Javier González
Dept. of Civil Eng., Univ. of Castilla-La Mancha, Av. Camilo José Cela s/n, 13071 Ciudad Real (Spain). // HIDRALAB INGENIERÍA Y DESARROLLOS, S.L., Spin-Off UCLM, Hydraulics Laboratory Univ. of Castilla-La Mancha, Av. Pedriza, Camino Moledores s/n, 13071 Ciudad Real (Spain)
E-mail: Javier.Gonzalez@uclm.es

methodology are also discussed, including its possible use for on-line leak detection strategies.

Keywords Leakage awareness · State estimation · Bad data analysis

1 Introduction

Leakage in water systems has been extensively discussed by both practitioners and researchers in recent years so as to better characterise and reduce water loss [5]. According to [30], leak management models can be broadly classified as: (1) leakage assessment or water audit methods, which aim to quantify the amount of water lost, (2) leakage detection methods, which intend to detect if and/or where leaks are taking place, and (3) leakage control models, which focus on effectively controlling and forecasting leakage. Leakage detection methods are outstanding amongst these, as they constitute one of the most sophisticated and active topics of ongoing research and it is common practice to use them in conjunction with other methods [30]. Within the leakage detection field, two types of analysis can be distinguished: (1) leakage awareness, which refers to the capability of sensing loss of water in a supply system, without giving any information about its precise location, and (2) leakage localisation, which focuses on identifying and prioritising leaking areas to accurately locate the source of leakage. Note that leakage awareness aims to identify if water is being lost in the system, i.e. it is a prior analysis that should be undertaken before running leak localisation algorithms to accurately pinpoint the leak. In this study, the problem of leakage awareness or leak detectability is tackled through state estimation techniques as an alternative to other existing methods, many of which are based on artificial intelligence [27,31].

State estimation techniques have been used in the power supply field for decades [36] and were introduced on an academic level in the water industry shortly afterwards [9]. The state estimation problem is normally set out as a weighted least squares (WLS) problem where the difference between measurements and estimated variables is minimised. Therefore, such techniques provide the most likely hydraulic state of a water system based on readings from available metering devices and flow governing equations [13]. Note that the state estimation problem enables the hydraulic state (i.e. flows, demands and pressures) of the system to be determined based on the available measurements, and this problem has traditionally been considered independent from the so called parameter estimation or calibration problem [34,14], where model parameters are inferred from existing measurements. There is significant literature on off-line state estimation approaches [1], and these techniques are also regarded as effective tools that should be relied upon to take full advantage of the huge amount of real-time data provided by telemetry systems [2]. Hence, they have the potential to identify if water is being lost in the system. This possibility has been explored before by [8] and [37], but the truth is that online state estimation has hardly been implemented in water networks to date [15]. However, due to the recent surge in the installation of telemetry systems worldwide, it is worth exploring the prospects state estimation has for leak detection.

There are many issues and sources of uncertainty that must be taken into account in order to build a consistent on-line leakage awareness algorithm based on state estimation results. For this reason, we believe that before proposing a method for real-time detection, it is necessary to assess the viability of detecting leaks according to the existing metering devices in the network. This is of utmost importance in water systems, which traditionally have been poorly metered [22] and subjected to a great deal of uncertainty [3], which has thus limited the usefulness of real-time monitoring. In this paper, we develop a method that enables the probability of detecting a leak in real time via state estimation to be evaluated by only analysing a snapshot of the state of the system. Note that this is in keeping with the same line of work of other state estimation related techniques, as assessing the viability of undertaking state estimation on-line has become as important as developing algorithms for state estimation itself. For example, prior analyses have been developed to assess if sufficient measurements exist to undertake state estimation (observability analysis, e.g. [28], [11]), if there are sufficient measurements to infer pump and valve settings from real-time data (topological observability analysis, e.g. [15]), if the uncertainty of the measurement devices enables good estimates to be obtained (uncertainty evaluation, e.g. [3], [12]), or whether additional metering devices must be included at strategic locations (optimal meter placement, e.g. [39], [22]), among others. A probabilistic leak detectability assessment is set out here to evaluate beforehand to what extent the presence of leakage could be sensed based on state estimation results in water supply systems subjected to measurement noise.

As will be presented in detail further on, in this paper detecting leaks is based on the fact that when a noticeable amount of water is lost throughout the network, the metered values and the variables obtained from the state estimation process differ considerably. In such cases, state estimation normalised residuals are high, and statistical tests can be undertaken to systematically identify the presence of unexpected leakage. Such a procedure is analogous to what is currently known as "bad data analysis" or "bad data processing" in the power supply field, where the method is standard practice for online state estimation evaluation. In the power industry, bad data analysis, and more specifically bad data identification, is normally carried out using the traditional largest normalised residual test [35], which gives a positive result if a specific normalised residual (computed from the measured value and the estimated variable) is above a chosen threshold [7]. This analysis has also been adapted to evaluate state estimation errors in water systems [2], but in this paper the method is used to systematically assess the existence of leakage for the first time. Note that in order to compute the probability of leak detection, sampling experiments (i.e. Monte Carlo simulations) on the normalised residual test are required to assess the effect of measurement noise. However, this type of experiments is known to be time-consuming [23] and can become tiresome when used to assess the behaviour of the overall system.

The aim of this paper is to present a new probabilistic methodology to previously assess leak detectability via state estimation in water transport networks. Water transport networks are pipeline systems that provide water to larger communities, e.g. District Metered Areas (DMA), where incoming flows are normally monitored. Therefore, they represent the "main arteries" that enable large urban areas to be supplied with water, and they are better metered than conventional water distribution systems. This explains why other state estimation related applications have focused on water transport networks [11, 37], which is the same area explored in this study. The method represents a step forward in terms of bad data analysis, as it propagates measurement noise to normalised residuals without the need for sampling experiments. Here, obtaining the probability of leak detection is based on the measurement-estimate joint bivariate covariance matrix. This probability can then be used to identify the minimum leak that could be detected at different positions within the system for a specific measurement setting and operating scenario, providing a useful tool to better understand the behaviour of the network in terms of leakage.

The organisation of this paper is as follows: firstly, the state estimation problem and the two commonly adopted procedures for bad data analysis are presented to then explain the principle assumed here for leakage awareness. Secondly, the probabilistic methodology for leak detectability assessment is set out, and then this analysis is built into a strategy for minimum leak assessment in order to estimate the minimum leak that could be detected according to the available metering devices. Afterwards, the method is applied to a case study, and the applicability of the algorithm is discussed, highlighting its potential and limitations. Finally, concise conclusions are drawn.

2 State estimation, bad data analysis and leakage awareness

2.1 The state estimation problem

State estimation can be formulated as a constrained WLS mathematical programming problem as follows:

$$\operatorname*{Min}_{\boldsymbol{x}} J(\boldsymbol{x}, \boldsymbol{z}) = \frac{1}{2} \left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})\right]^T \boldsymbol{C}_{\boldsymbol{Z}}^{-1} \left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})\right] \tag{1}$$

subject to

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{K}\boldsymbol{Q}|\boldsymbol{Q}|^{b-1}, \tag{2}$$

$$\boldsymbol{B}\,\boldsymbol{Q} = -\boldsymbol{q}, \tag{3}$$

$$q_i = 0; \forall i \in \mathcal{V}^{\mathrm{T}}, \tag{4}$$

$$-\boldsymbol{Q}^{max} \leq \boldsymbol{Q} \leq \boldsymbol{Q}^{max}, \tag{5}$$

$$\boldsymbol{x}^{min} \leq \boldsymbol{x} \leq \boldsymbol{x}^{max}, \tag{6}$$

where $\hat{\boldsymbol{x}}$ corresponds to the optimal solution. Note that the objective function J in Eq. (1) involves the $\boldsymbol{z} \in \mathbb{R}^m$ vector of available measurements, the $\boldsymbol{x} \in \mathbb{R}^n$ state variable vector (constituted by head levels in this paper), the $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ non-linear relationship between $\boldsymbol{x}$ and $\boldsymbol{z}$, and the $\boldsymbol{C_Z}$ matrix, which is the $m \times m$ measurement variance-covariance matrix, i.e. the inverse of the traditional weight matrix $\boldsymbol{W} = \boldsymbol{C}_{\boldsymbol{Z}}^{-1}$. At the same time, the $\boldsymbol{h}$ relationship is defined by hydraulic constraints (2)-(6). Eq. (2) represents the headloss equation, where $\boldsymbol{A}$ is the connectivity matrix (+1 at initial node and -1 at final node), $\boldsymbol{K}$ is the flow resistance pipe coefficient, $\boldsymbol{Q}$

represents pipe flow, and b is the equation exponent, which is $b = 1.852$ for Hazen-Williams headloss equation. On the other hand, Eq. (3) is the continuity equation that enables water demand $\boldsymbol{q}$ to be computed, where $\boldsymbol{B}$ is a topological incidence sub-matrix that contains +1 values when water enters the node and -1 when it flows out. Note that an additional constraint (4) exists to simulate the existence of nodes known to have null demand (i.e. transit nodes). Such nodes constitute a subset $\forall i \in \mathcal{V}^{\mathrm{T}}$ of the set of nodes in the system $\forall i \in \mathcal{V}$. Additionally, Eqs. (5)-(6) impose physical limits on water flows and head levels, respectively. Note that the state estimation problem (1)-(6) traditionally assumes that model parameters (e.g. roughness parameters, pumps and valve settings) are known beforehand [11], i.e. the system has been previously calibrated and the network topology is known, thus only measurement errors are taken into account.

2.2 Bad data analysis

Bad data analysis is essential for assessing the result of any state estimator [7]. Bad data analysis has traditionally been used to detect any erroneous measurements in the system, i.e. to identify if there are significant deviations between the metered and the estimated values. Bad measurement treatment typically consists of two phases: (1) bad data detection, and (2) bad data identification.

2.2.1 Bad data detection

Bad data detection is typically formulated as a hypothesis testing problem. The null hypothesis $\mathcal{H}_0$ corresponds to a scenario in which no bad data are present, whereas the alternative hypothesis $\mathcal{H}_1$ considers that bad data exist. According to [2] and [7], the Chi-square test is normally applied for bad data detection:

$$\hat{J} \begin{cases} \leq \chi^2_{m+p-n,\alpha}, \text{ accept } \mathcal{H}_0 \\ > \chi^2_{m+p-n,\alpha}, \text{ reject } \mathcal{H}_0 \end{cases}, \tag{7}$$

where $\hat{J}$ refers to the value of the objective function at the estimated state, and $\chi^2_{m+p-n,\alpha}$ is the Chi-square distribution function corresponding to $m + p - n$ degrees of freedom and a $1 - \alpha$ confidence level (typical values for α are 0.1, 0.05, or 0.01). Note that p refers to the number of equality and binding inequality constraints provided by Eqs. (2)-(6).

2.2.2 Bad data identification

Bad data identification is typically undertaken when bad measurements have been detected. It traditionally consists of applying the largest normalised residual test [35]. According to [7], this procedure requires the computation of the normalised residual vector $\boldsymbol{r}^N$ as follows:

$$r_i^N = \frac{|z_i - \hat{z}_i|}{\sqrt{\Omega_{ii}}}, \tag{8}$$

where $\hat{\boldsymbol{z}} = \boldsymbol{h}(\hat{\boldsymbol{x}})$ refers to the estimated variables depending on the optimal state variables $\hat{\boldsymbol{x}}$, and $\boldsymbol{\Omega}$ is the residual covariance matrix. The residual covariance matrix can be computed as shown by [6] in power systems, or by [12] in water systems:

$$\boldsymbol{\Omega} = (\boldsymbol{I} - \boldsymbol{H}\boldsymbol{S_{xz}})\boldsymbol{C_Z}(\boldsymbol{I} - \boldsymbol{H}\boldsymbol{S_{xz}})^T, \tag{9}$$

where $\boldsymbol{I}$ is the $m \times m$ identity matrix, $\boldsymbol{H}$ represents the $m \times n$ measurement Jacobian matrix [11], and $\boldsymbol{S_{xz}}$ is the $n \times m$ sensitivity matrix of the state variables (i.e. nodal heads) with respect to the available measurements [12]. Once $\boldsymbol{\Omega}$ has been computed, normalised residuals for each i measurement can be calculated according to Eq. (8) and compared with a chosen identification threshold, e.g. $\Phi^{-1}(1-\alpha/2)$, which refers to the inverse of the normal distribution function for a given confidence level $1 - \alpha$.

2.3 Leakage awareness

In this paper, we assume that leakage can be detected thanks to the same principle that has been traditionally used for bad data analysis in both the power supply field [7], and the water industry [2]. Note that the presence of a leak acts as an additional demand node in the system, leading to inconsistencies among measurements and estimated variables: readings from metering devices are altered by the loss of water, whereas the estimated variables are the result of minimising the difference between measurements and estimates while also taking into account the flow governing equations specified in (2)-(6). Therefore, when there is a loss of water, the resulting objective function is abnormally large, and so are the normalised state estimation residuals. In other words, a leak can be sensed by either: (1) subjecting the estimated objective function $\hat{J}$ to a chi-squared statistic test (Eq. (7)), which is analogous to bad data detection, or (2) assessing the deviation of normalised residuals obtained with Eq. (8), which is analogous to bad data identification.

As mentioned before, the detection phase is normally carried out prior to the identification process. The purpose of this strategy is to skip the second step according to the results of the first one, thus minimising the computational expense. However, as both analyses are independent, they may not always be consistent with each other: for example, the test described in Eq. (7) can indicate that leakage does not exist while the second analysis shows there is a loss of water due to the existence of abnormally large residuals. This is because the detection phase is normally a poorer indicator due to the reliance on an aggregate value of the WLS, i.e. the objective function. For this reason, in this paper residual analysis is used straight away to detect leakage, i.e. each measurement normalised residual is subjected to a test rather than the objective function on its own. This is sustained by the fact that leaks affect the estimation of not only the flow measurements nearby, but also the remaining hydraulic variables (i.e. head levels) in the system. Note that we assume that the hydraulic problem set out in Eqs. (2)-(6) has been previously calibrated, i.e. it includes background leakage. Hence, only pipe bursts or other events that induce abnormal pressure or flow variations with respect to the calibrated model can be detected with the methodology presented hereafter. This brings outs the importance of periodic calibration of hydraulic models [14].

3 Probabilistic leak detectability assessment via state estimation

In this section, a new probabilistic methodology for prior leak detectability assessment is provided. This approach works with mean metered values and the measurement-estimate joint covariance matrix to cover all likely perturbations of state estimation normalised residuals. On this basis the probability of leak detection according to the limit state equations can be calculated. Hence, there are three main blocks in this section: firstly, the construction process of the measurement-estimate joint covariance matrix is presented. Secondly, the limit state equations inferred from the largest normalised residual test to detect leakage are put together. Finally, the probability of leak detection is computed based on the previous information.

3.1 Computation of the joint bivariate covariance matrix

The traditional largest normalised residual test can be used to identify leaks by relying on computation of normalised residuals according to Eq. (8), which considers isolated values of measurements z_i and their corresponding estimated variables $\hat{z}_i$. However, readings from metering devices are subject to noise. Note that so far, we have been talking about realizations of such variables, i.e. particular sets of readings from a measurement device i. Nevertheless, the objective of this paper is to compute the probability of leak detection by analysing a single hydraulic state, i.e. one time period, rather than undertaking sampling experiments of a large number of realizations. Therefore, from now on it is necessary to work with random variables Z_i and $\hat{Z}_i$ rather than with individual variable realizations z_i and $\hat{z}_i$. Thus, a 2×2 variance-covariance matrix $\boldsymbol{C}^{(i)}_{\boldsymbol{Z},\hat{\boldsymbol{Z}}}$ can be built for each measurement i as follows:

$$\boldsymbol{C}^{(i)}_{\boldsymbol{Z},\hat{\boldsymbol{Z}}} = \begin{bmatrix} C^{(i)}_{Z} & C^{(i)}_{Z,\hat{Z}} \\ C^{(i)}_{\hat{Z},Z} & C^{(i)}_{\hat{Z}} \end{bmatrix}; \; \forall i = 1, \ldots, m. \quad (10)$$

In this expression, $C^{(i)}_{Z}$ refers to the variance of measurement i, i.e. diagonal component of matrix $\boldsymbol{C_Z}$. Component $C^{(i)}_{\hat{Z}}$ represents the variance of the estimated variable, which can be obtained by propagating measurement uncertainty with the First-Order Second-Moment method as shown by [12]. The crossed component or covariance $C^{(i)}_{\hat{Z},Z}$ can similarly be obtained from the corresponding matrix among:

$$\boldsymbol{C}_{\hat{\boldsymbol{Z}}_{\boldsymbol{x}},\boldsymbol{Z}_{\boldsymbol{x}}} = \boldsymbol{S}_{\boldsymbol{xz}}\boldsymbol{C}_{Z}, \quad (11)$$

$$\boldsymbol{C}_{\hat{\boldsymbol{Z}}_{\boldsymbol{Q}},\boldsymbol{Z}_{\boldsymbol{Q}}} = \boldsymbol{H}_{\boldsymbol{Q}}\boldsymbol{C}_{\hat{\boldsymbol{Z}}_{\boldsymbol{x}},\boldsymbol{Z}_{\boldsymbol{x}}}, \quad (12)$$

$$\boldsymbol{C}_{\hat{\boldsymbol{Z}}_{\boldsymbol{q}},\boldsymbol{Z}_{\boldsymbol{q}}} = \boldsymbol{H}_{\boldsymbol{q}}\boldsymbol{C}_{\hat{\boldsymbol{Z}}_{\boldsymbol{x}},\boldsymbol{Z}_{\boldsymbol{x}}}, \quad (13)$$

where $\boldsymbol{H_Q}$ and $\boldsymbol{H_q}$ correspond to the rows related to flows and demand values in the measurement Jacobian matrix $\boldsymbol{H}$, respectively.

Once all required matrices have been computed, it is straightforward to extract the components that correspond to measurement and estimated variable i. Hence, its variance-covariance matrix (10) can be built immediately. It must be noted that each of these variance-covariance matrices contains all the possible perturbations that can be induced to the state estimate $\hat{Z}_i$ as a result of measurement noise in Z_i. This approach represents a step forward with respect to the traditional largest normalised residual test, which only evaluates the normalised residual for the measurement value rather than considering all likely perturbations.

3.2 Limit state equations

As commented before, the largest normalised residual test is based on comparing the normalised residual of each measurement with a chosen threshold. According to Eq. (8), z_i, $\hat{z}_i$ and Ω_{ii} must be known to compute residuals, but an implicit relationship (9) exists between $\boldsymbol{\Omega}$ and the state estimation solution. As readings from metering devices are not expected to produce gross errors, it has been tested numerically that $\boldsymbol{\Omega}$ does not vary considerably with noise (see Hanoi case study for details). Thus, this matrix can be computed only for the mean value of the measurements and be considered constant here. Thanks to this assumption, we can derive from Eq. (8) that an estimated variable $\hat{z}_i$ indicates the existence of leakage when it falls out of the confidence intervals around the metered variable z_i:

$$\hat{z}_i \geq z_i + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}};\ \forall i = 1, \ldots, m, \tag{14}$$

$$\hat{z}_i \leq z_i - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}};\ \forall i = 1, \ldots, m. \tag{15}$$

3.3 Probability of leak detection

According to limit state equations (14) and (15), leaks can be detected by comparing the result of an estimated variable $\hat{z}_i$ and its measured value $z_i; \forall i = 1, \ldots, m$. Note that both measurements Z_i and estimations $\hat{Z}_i$ are random variables, and therefore the probability of leak detection is as follows:

$$\text{Prob}\left\{\left[\hat{Z}_i \geq Z_i + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right] \cup \left[\hat{Z}_i \leq Z_i - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right]\right\} \tag{16}$$

Means (μ_{Z_i} and $\mu_{\hat{Z}_i}$) and variance-covariance matrix ($\boldsymbol{C}^{(i)}_{\boldsymbol{Z},\hat{\boldsymbol{Z}}}$) of these random variables can be easily determined. On the one hand, as this methodology evaluates the viability of detecting leaks according to the available metering devices by analysing a snapshot of the system, the mean value of measurements μ_{Z_i} is considered to be equal to the value of such variables in the flow network solution associated with the selected operating condition and including the artificial presence of the leak. Then, the associated $\mu_{\hat{Z}_i}$ can be obtained by solving the state estimation problem for the mean value of measurements μ_{Z_i}. On the other hand, the variance-covariance matrix for each measurement i can be obtained from Eq. (10).

This information enables the associated measurement-estimate probability density function to be plotted in a z_i-$\hat{z}_i$ space, whose contours are shown in Figure 1a.

Additionally, limit state equations (14) and (15) are represented in this figure: Eq. (14) corresponds to the shaded grey area above line LSE$_{i1}$, and Eq. (15) is related to the shaded grey area below line LSE$_{i2}$. Therefore, the probability of detecting a leak for a measurement i (P_{detect_i}; $\forall i = 1, \ldots, m,$) is equal to the probability that random variables Z_i and $\hat{Z}_i$ hold limit state equations (14)-(15), as described in Eq. (16). Thus, this probability can be computed by integrating the joint probability density function of both random variables over the grey region.

Starting with LSE$_{i1}$, the probability of random variables $(Z_i, \hat{Z}_i)$ being above that line (i.e. holding to Eq. (14)) can be calculated with well-known First-Order Second-Moment structural reliability methods. According to [20], the probability of being above that line can be obtained by using expression $\Phi(\beta_{i_1})$, where Φ represents the normal distribution function and β_{i_1} is the so called reliability index, which corresponds to the minimum distance between the centre of the joint bivariate probability density function $(\mu_{Z_i}, \mu_{\hat{Z}_i})$ and line LSE$_{i1}$ [25]. Note that line LSE$_{i1}$ is equal to constraint (14) but uses an equality. Similarly, the probability that Eq. (15) holds can be calculated from the distance between LSE$_{i2}$ (β_{i_2}) and the normal distribution function. Therefore, the problem of calculating the probability of detection reduces to calculating distances β_{i_1} and β_{i_2}. According to [20], $\beta_{i_j}; \forall i = 1, \ldots, m; \forall j = 1, 2$ can be invariantly defined for each limit state equation as:

$$\left.\begin{array}{l} \beta_{i_j} = \underset{z_i, \hat{z}_i}{\text{Min}} \sqrt{\begin{pmatrix} z_i - \mu_{Z_i} \\ \hat{z}_i - \mu_{\hat{Z}_i} \end{pmatrix}^T \boldsymbol{C}^{(i)-1}_{\boldsymbol{Z},\hat{\boldsymbol{Z}}} \begin{pmatrix} z_i - \mu_{Z_i} \\ \hat{z}_i - \mu_{\hat{Z}_i} \end{pmatrix}} \\ \text{subject to} \\ \quad \hat{z}_i = z_i + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}} \quad \text{if } j = 1 \\ \quad \hat{z}_i = z_i - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}} \quad \text{if } j = 2 \end{array}\right\} \forall i = 1, \ldots, m; \forall j = 1, 2. \tag{17}$$

This system of equations could be solved more straightforwardly if the original z_i-$\hat{z}_i$ space was converted to a standard normal space y_{i_1}, y_{i_2} [26] through the orthogonal transformation:

$$\begin{bmatrix} z_i \\ \hat{z}_i \end{bmatrix} = \begin{bmatrix} \mu_{Z_i} \\ \mu_{\hat{Z}_i} \end{bmatrix} + \boldsymbol{L} \begin{bmatrix} y_{i_1} \\ y_{i_2} \end{bmatrix}, \tag{18}$$

with $\boldsymbol{L}$ being the lower triangular matrix from Cholesky decomposition of the measurement-estimate variance-covariance matrix $\boldsymbol{C}^{(i)}_{\boldsymbol{Z},\hat{\boldsymbol{Z}}}; \forall i = 1, \ldots, m$. Using (18), problem (17) can be rewritten as follows:

$$\left.\begin{array}{l} \beta_{i_j} = \underset{y_{i_1}, y_{i_2}}{\text{Min}} \sqrt{\begin{pmatrix} y_{i_1} \\ y_{i_2} \end{pmatrix}^T \begin{pmatrix} y_{i_1} \\ y_{i_2} \end{pmatrix}} \\ \text{subject to} \\ y_{i_2} - \left(\frac{l_{22}}{l_{11}} + \frac{l_{21}}{l_{11}}\right) y_{i_1} = l_{22}\left[\mu_{Z_i} - \mu_{\hat{Z}_i} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right] \quad \text{if } j = 1 \\ y_{i_2} - \left(\frac{l_{22}}{l_{11}} + \frac{l_{21}}{l_{11}}\right) y_{i_1} = l_{22}\left[\mu_{Z_i} - \mu_{\hat{Z}_i} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right] \quad \text{if } j = 2 \end{array}\right\} \forall i = 1, \ldots, m; \forall j = 1, 2, \tag{19}$$

where l_{ij} refers to $\boldsymbol{L}^{-1}$ matrix components. Note that in the standard normal random space, the reliability index corresponds to the minimum Euclidean distance between the origin and the limit state equation, as shown in Figure 1b.

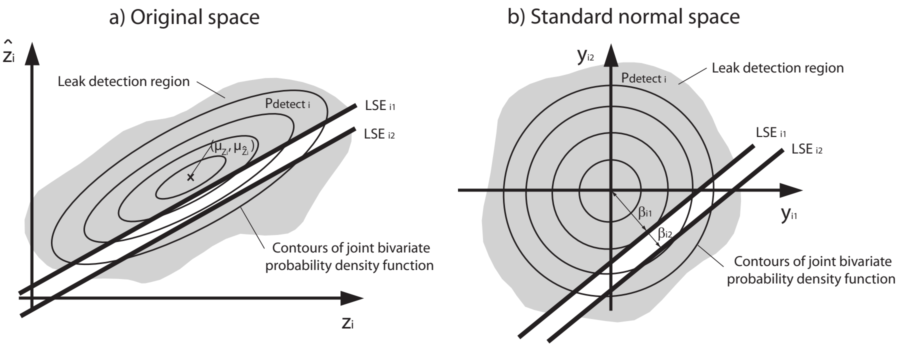

Fig. 1 Graphic illustration of measurement-estimate joint probability density function and limit state equations (LSE) for a measurement i: a) original space, b) standard normal space

Problem (19) can be solved analytically, with β_{i_1} and β_{i_2} being the respective distances:

$$\left.\begin{array}{l} \beta_{i_j} = \sqrt{\frac{a_0^2}{a_1^2+a_2^2}}, \\ \text{with} \\ \quad a_0 = -l_{22}\left[\mu_{Z_i} - \mu_{\hat{Z}_i} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right]; \quad \text{if } j=1 \\ \quad a_0 = -l_{22}\left[\mu_{Z_i} - \mu_{\hat{Z}_i} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\sqrt{\Omega_{ii}}\right]; \quad \text{if } j=2 \\ \quad a_1 = -\left(\frac{l_{22}}{l_{11}} + \frac{l_{21}}{l_{11}}\right) \\ \quad a_2 = 1; \end{array}\right\} \forall i = 1,\ldots,m; \forall j = 1,2. \tag{20}$$

Once the distances to both LSE_{i1} and LSE_{i2} have been computed, it is only a matter of combining them with the normal distribution function to obtain the probability of detection. In order to compute P_{detect_i}; $\forall i = 1,\ldots,m$, three scenarios must be distinguished to cover all the possible relative positions of the limit state equations with respect to the joint bivariate distribution (see Figure 2). For each of these cases, the probability of leak detection (Eq. (16)) can be computed as:

$$P_{detect_i} = \left\{\begin{array}{ll} 1-[\Phi(\beta_{i_2})-\Phi(\beta_{i_1})] = 1-\Phi(\beta_{i_2})+\Phi(\beta_{i_1}) & \text{if Case a} \\ 1-[\Phi(-\beta_{i_2})-\Phi(-\beta_{i_1})] = 1+\Phi(\beta_{i_2})-\Phi(\beta_{i_1}) & \text{if Case b} \\ 1-[\Phi(\beta_{i_2})-\Phi(-\beta_{i_1})] = 2-\Phi(\beta_{i_2})-\Phi(\beta_{i_1}) & \text{if Case c} \end{array}\right\}; \forall i = 1,\ldots,m, \tag{21}$$

Note that the joint bivariate distribution can be seen as a normal distribution whose axis is perpendicular to the limit state equations and passes through the origin. It must be noted that this procedure enables the probability of detecting a given leak located at a specific node to be computed according to each of the m available measurements i ($P_{detect_i}; \forall i = 1,\ldots,m$). Hence, the maximum value of this probability determines the overall network probability of detection (P_{detect}) for a specific magnitude of leak at that location considering the existing measurement setting and the selected flow scenario, i.e. $P_{detect} = \max(P_{detect_i}; \forall i = 1,\ldots,m)$.

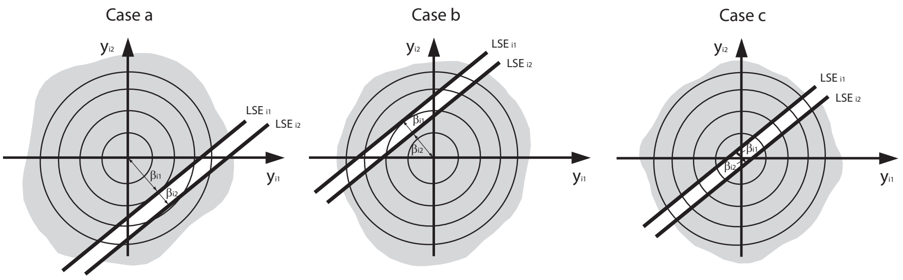

Fig. 2 Relative positions of the limit state equations (LSE) with respect to the centre of the joint bivariate distribution in the standard normal space: cases a, b and c

Finally, note that once P_{detect} is obtained, the user should define the admissible threshold for leakage detection (P_{lim}), which is the likelihood above which it can be said that the leak is noticed. For example, there could be considered to be a leak when there is a probability of detection greater than 80%. We believe that this is a reasonable value, as in water transport networks demand values are expected to be measured reasonably well and the likelihood of positive or false scenarios occurring is not that high. In any case, if this limit is increased, only greater values of leak will be detected at the leaking position; however, if this limit is reduced, lower values of leak could be identified but there will be a higher risk of false positive or false negative scenarios occurring. This demonstrates the importance of selecting a reasonable detection threshold for the water transport network under consideration.

4 Minimum leak assessment

The aforementioned procedure for leak detectability assessment could also be used to test the overall response of a water system to leak detection, i.e. so as to plot the minimum leak that could be detected in different parts of the network with the available measurement setting and for a given operating condition. Note that the aim of the approach presented in this paper is to assess the leak detection possibilities of the system on the basis of the largest normalised residual test beforehand, and estimating the minimum leak that could be sensed in different parts of the network gives an idea of how well-prepared the system is for on-line detection according to the available instrumentation and detection strategy assumed.

As schematically described in Figure 3, this application requires evaluation of leak detectability at each of the nodes in the system, one at a time. Once a node has been selected, the leak value at the node has to be initialised in order to artificially simulate the occurrence of a leak. This can be done by increasing its demand. Leaks could alternatively have been divided at the two end nodes of a pipe, but this would still be a simplification that would not preserve the energy balance equation of pipes and could lead to head loss errors [17]. For this reason, we have simplified the problem by concentrating water loss at one node of the system at a time. Then, the flow network solution is computed. Note that, as mentioned before, the solution of the flow network represents the mean value of the measurements when the system is leaking

(μ_{Z_i}). Therefore, state estimation can be undertaken (e.g. via mathematical programming) considering the values obtained from the flow network solution as metered variables. Subsequently, the probability of leak detection P_{detect} is obtained from the limit state equations and the measurement-estimate joint bivariate distribution at the standard normal space, as commented before. If this probability is above the assumed P_{lim}, the leak value has to be reduced in order to approach the minimum leak value. However, if the probability is below P_{lim} the leak value has to be increased in order to better adjust the minimum value. In this paper, the leak value is updated according to the bisection method [4] up to $P_{detect} = P_{lim}$ with a given tolerance. When this probability of detection is obtained, the leak is considered to be the minimum leak value for the node under study, and we can proceed with the next junction.

Implementation of this process enables a map to be plotted that shows leak detectability within the system, thus intuitively providing an insight of where to locate additional metering devices to enhance network performance in terms of leak detection, as will be shown in the case study. Note that this approach provides probabilistic information about how the different parts of the network behave in terms of leakage detection, but only mean values are assessed because metered values correspond to the flow network solution. As commented before, this analysis allows planners to evaluate the capability of the largest normalised residual test to detect leaks in real-time, i.e. considering the state estimation solution when online readings from metering devices are gathered.

5 Case study: Hanoi Network

The Hanoi network presented by [16] has been used in this paper as a case study. More specifically, the modified version presented by [12], which considers nodes 3, 16, 23 and 25 as transit nodes or nodes with null consumption, has been adopted here (see Figure 4) in order to introduce some hydraulic constraints for the computation of $\boldsymbol{\Omega}$. Therefore, in this particular case study we work with a head level vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_{32})$, a water demand vector $\boldsymbol{q} = (q_1, q_2, \ldots, q_{32})$ and a flow vector $\boldsymbol{Q} = Q_{1-2}, Q_{2-3}, \ldots, Q_{25-32}$. Appendix S1 gathers the specific components of such vectors for this particular example.

Regarding the measurement configuration, we assume that the modified Hanoi network is a water transport network: water consumption is metered at all demand nodes, as is likely to be the case if they were DMAs in a sectorised water system. Also, the water level at the tank (x_1) is measured. Note that these settings ensure the system is observable [11]. Moreover, two different scenarios that consider different sets of additional redundant devices are evaluated in this paper: (1) a pressure meter at node 30 (x_{30}), i.e. one degree of redundancy exists, and (2) pressure meters at nodes x_9, x_{18} and x_{30}, and flow meters at pipes Q_{3-4} and Q_{23-24}, i.e. five degrees of redundancy exist. Table 1 gathers the measurements included in each of these settings. In both scenarios all measurements are assumed to be independent, which is reasonable as all are readings from metering devices. In this paper, we assume that flow meters are subjected to standard deviation $\sigma_q = 2\%q$ or $\sigma_Q = 2\%Q$, depending on whether they measure demand values or flows, with q and Q equal to

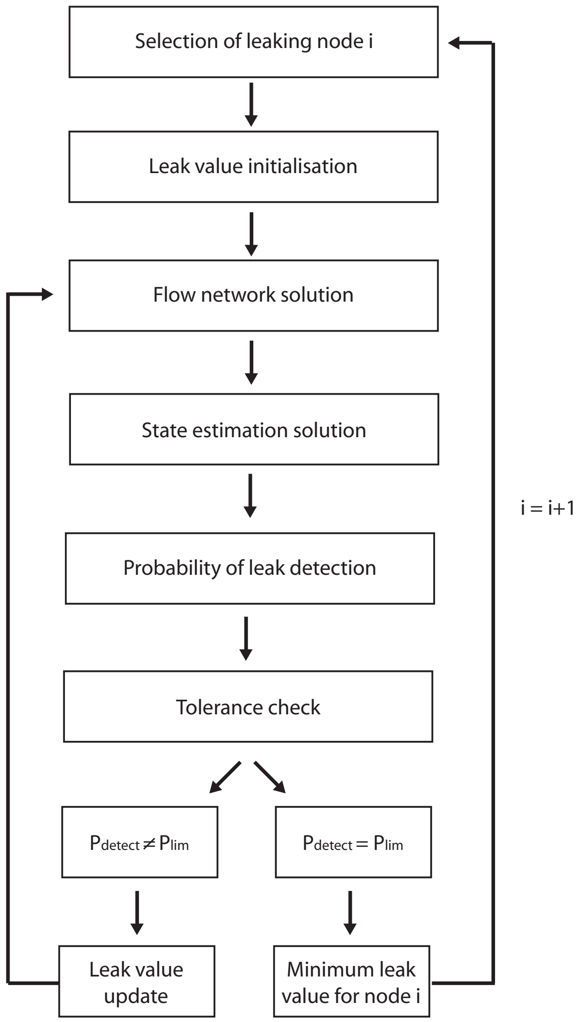

Fig. 3 Flow chart for minimum leak assessment

the value of these variables according to the flow network solution when the leakage is included. Note that the flow network solution must be computed considering extra demand in the leaking node, but the metered demand at the leaking node itself must be the original value (see Appendix S2 for details). This is because leakage takes place in the water transport network, but the flow meter at the DMA remains unaware of its existence. Similarly, pressure meters are associated with $\sigma_x = 0.01$ bar and water level meters are subjected to $\sigma_x = 0.01$ m, which are typical values for current instrumentation. Regarding operating conditions, we assume average demand values, which correspond to those included in Appendix S1.

Once both measurement scenarios have been explained, the aforementioned methodologies are applied. Firstly, the probabilistic leak assessment proposed in this work

Table 1 Measurement scenarios for Hanoi network case study

Measurement scenario	Pressure meters	Flow meters	Demand meters
One degree of redundancy	x_1, x_{30}	-	All nodes
Five degrees of redundancy	x_1, x_9, x_{18}, x_{30}	Q_{3-4}, Q_{23-24}	All nodes

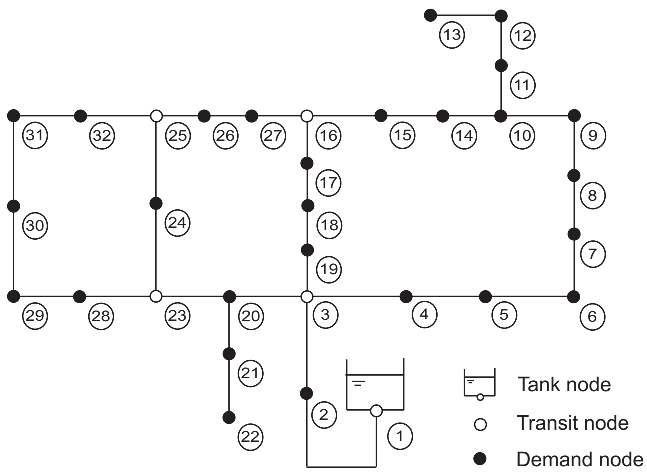

Fig. 4 Case study: Hanoi network

is used to test if a specific leak value could be detected when one or five degrees of redundancy exist in the system. Then, the minimum leak assessment is carried out for the same settings. The confidence level is assumed to be 95% ($\alpha = 0.05$) and the probability threshold is considered to be $P_{lim} = 0.8$ all along.

5.1 Probabilistic leak detectability assessment

In this part of the paper, the probability of detecting a $200\ \mathrm{m}^3/\mathrm{h}$ leak occurring at node 6 is assessed with the two measurement settings previously described. Note that this value corresponds to water loss of approximately 1% of the total system inflow, which is reasonable for a network such as this. The probability of leak detection is computed with both the method presented in this paper and the largest normalised residual test considering a Monte Carlo sampling of 1000 measurement configurations. Note that the Monte Carlo simulation fundamentally consists in applying the largest normalised residual test 1000 times: for each measurement configuration, normalised residuals must be computed according to Eq. (8) and compared with the specified threshold $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$. The probability of detection is computed by counting the number of measurement configurations that lead to residuals greater than the threshold.

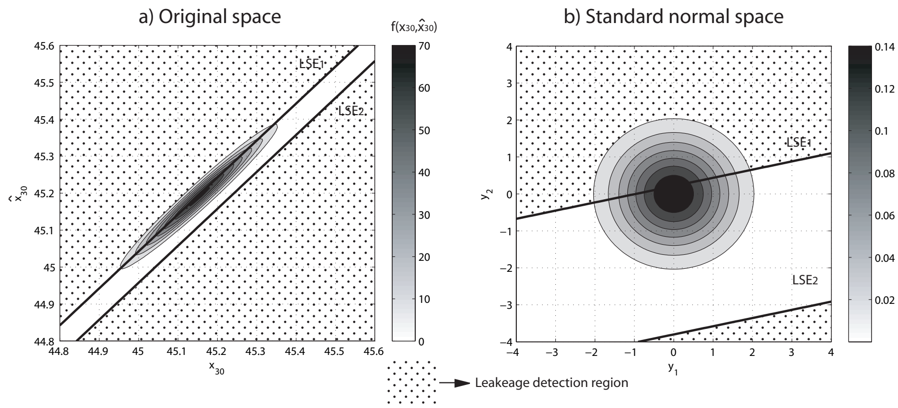

Fig. 5 Measurement-estimate joint probability density function (f) and limit state equations (LSE) for measurement x_{30} when a 200 m^3/h leak exists at node 6 in Hanoi network case study: a) original space, b) standard normal space

5.1.1 One degree of redundancy (x_{30})

As explained before, in order to evaluate the probability of detecting a leak according to the methodology presented in this paper, the joint bivariate covariance matrix must be computed for all the measurements that exist in the system, i.e. head levels at nodes 1 and 30, and all water demand values. The covariance matrix for measurement x_{30} is provided by way of illustration:

$$\boldsymbol{C}_{x_{30},\hat{x}_{30}} = \begin{bmatrix} 0.0100 & 0.0095 \\ 0.0095 & 0.0095 \end{bmatrix}. \tag{22}$$

This matrix has been obtained by selecting the convenient rows and columns from covariance matrices of the metered and estimated variables. Note that element $(1,1)$ corresponds to the measurement variance, which is here assumed to be $(0.1\ \mathrm{m})^2$ for pressure meters. However, element $(2,2)$ shows the variance of the estimated variable, which is lower than the previous one because it refers to the result of the optimisation problem (1)-(6). Element $(1,2)$ is equal to element $(2,1)$, and they show that a correlation $\rho = 0.9763$ exists between measurement x_{30} and its estimated value $\hat{x}_{30}$.

Once the joint covariance matrix is obtained, the measurement-estimate joint probability density function can be plotted as in Figure 5a. Also, limit state equations can be derived and represented in the $x_{30} - \hat{x}_{30}$ space. As this figure shows, most of the joint distribution lies within the space between the two limit state equations, thus the associated probability of leak detection is low. In order to quantify this probability, the joint distribution and the limit state equations are transformed to a standard normal space, where $\beta_{x_{30,1}} = 0.2085$ and $\beta_{x_{30,2}} = 3.7115$ according to Eq. (20). Bearing in mind that limit state equations are at each side of the origin (case c in Eq. (21)), the probability of detecting a leak based on measurement x_{30} can be calculated as $P_{detect_{x_{30}}} = 2 - \Phi(\beta_{x_{30,2}}) - \Phi(\beta_{x_{30,1}}) = 0.4175$.

Nevertheless, there are 28 measurements more that must undergo the same analysis. Table 2 shows the probability of leak detection based on each of the available

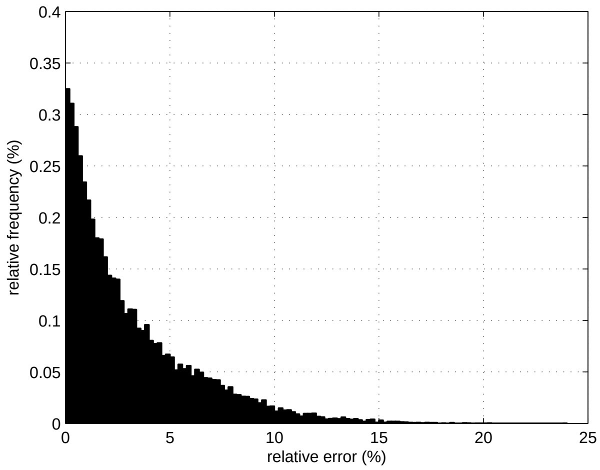

Fig. 6 Histogram of the relative error in the computation of the $\boldsymbol{\Omega}$ matrix in the 1000 Monte Carlo experiment with respect to the mean matrix assumed for the proposed methodology: one degree of redundancy

measurements according to the methodology presented in this paper and the largest normalised residual test considering a Monte Carlo sampling of 1000 measurement configurations. Results prove that the new approach provides a good approximation while considerably reducing the computational time: 5.1 s are required to compute all P_{detect_i} measurements with the new methodology, whereas 2920.0 s are needed for the largest normalised residual test simulation in a MatLab 7.12.0 (R2011a) 64-bits version and a 23.3 GAMS 64-bits version when run in an Intel(R) Core(TM) i7-6700 CPU 3.40 GHz 16 GB RAM desktop computer. Moreover, Figure 6 provides the relative error of the elements in the diagonal of the $\boldsymbol{\Omega}$ matrix in the 1000 simulations with respect to the mean value used in the methodology put forward (gathered in Appendix S2). This figure shows that the relative error remains relatively low, thus validating the assumption of a constant $\boldsymbol{\Omega}$ for the probabilistic assessment presented here.

Finally, the probability of detection in the whole network can be obtained as the maximum value of the probabilities obtained for each measurement: $P_{detect} = \max(P_{detect_i}; \forall i = 1, \ldots, m) = 0.4185$ for the proposed approach and $P_{detect} = \max(P_{detect_i}; \forall i = 1, \ldots, m) = 0.4010$ for the largest normalised residual test Monte Carlo sampling. Note that in this case the probabilities of detection for all measurements are very similar with both methods because of low redundancy, but greater variability will be obtained when more meters are added. In any case, as this probability $P_{detect} \approx 0.4$ is below $P_{lim} = 0.8$, it can be concluded that it is not possible to detect the leak with the available measurement setting.

Table 2 Probability of detection based on all available measurements when a 200 m^3/h leak exists at node 6 in Hanoi network case study: one degree of redundancy

Measurement	Proposed methodology P_{detect_i}	Largest normalised residual test P_{detect_i}
x_{30}	0.4175	0.3990
x_1	0.4175	0.3990
q_2	0.4166	0.3970
q_4	0.4175	0.3990
q_5	0.4170	0.3980
q_6	0.4162	0.3970
q_7	0.4184	0.4010
q_8	0.4179	0.3990
q_9	0.4175	0.3990
q_{10}	0.4175	0.3990
q_{11}	0.4175	0.3990
q_{12}	0.4177	0.3990
q_{13}	0.4173	0.3980
q_{14}	0.4175	0.3990
q_{15}	0.4175	0.3990
q_{17}	0.4178	0.3990
q_{18}	0.4185	0.4010
q_{19}	0.4175	0.3990
q_{20}	0.4166	0.3970
q_{21}	0.4176	0.3990
q_{22}	0.4174	0.3990
q_{24}	0.4172	0.3980
q_{26}	0.4167	0.3970
q_{27}	0.4176	0.3990
q_{28}	0.4174	0.3990
q_{29}	0.4175	0.3990
q_{30}	0.4175	0.3990
q_{31}	0.4175	0.3990
q_{32}	0.4171	0.3980

5.1.2 Five degrees of redundancy (x_9, x_{18}, x_{30}, Q_{3-4}, Q_{23-24})

The same analysis is undertaken considering the five degrees of redundancy measurement setting. Table 3 provides the probability of detection for each measurement according to the methodology presented in this paper and the largest normalised residual test with a Monte Carlo sampling of 1000 measurement configurations. Results prove that the new procedure provides a good approximation while significantly reducing the computational cost: 5.5 s are required to compute P_{detect} with the methodology described herein whereas 2876.2 s are needed in the sampling approach. Note that the approximate probability of detection in both cases is $P_{detect} \approx 0.8$. These

Table 3 Probability of detection based on all available measurements when a 200 m^3/h leak exists at node 6 in Hanoi network case study: five degrees of redundancy

Measurement	Proposed methodology P_{detect_i}	Largest normalised residual test P_{detect_i}
x_9	0.4193	0.4010
x_{18}	0.0503	0.0430
x_{30}	0.1103	0.0940
x_1	0.4565	0.4320
Q_{3-4}	0.0883	0.1020
Q_{23-24}	0.0527	0.0460
q_2	0.4551	0.4320
q_4	0.6796	0.6650
q_5	0.7882	0.7780
q_6	0.8156	0.7950
q_7	0.8170	0.7960
q_8	0.7831	0.7760
q_9	0.7432	0.7280
q_{10}	0.7608	0.7440
q_{11}	0.7608	0.7440
q_{12}	0.7610	0.7460
q_{13}	0.7605	0.7440
q_{14}	0.7713	0.7580
q_{15}	0.6614	0.6640
q_{17}	0.2670	0.2440
q_{18}	0.1541	0.1370
q_{19}	0.2890	0.2510
q_{20}	0.3616	0.3300
q_{21}	0.3624	0.3300
q_{22}	0.3621	0.3300
q_{24}	0.0838	0.0950
q_{26}	0.1479	0.1510
q_{27}	0.1862	0.1860
q_{28}	0.0792	0.0840
q_{29}	0.0500	0.0450
q_{30}	0.0558	0.0430
q_{31}	0.0545	0.0420
q_{32}	0.0531	0.0540

results prove that leakage awareness is possible in the Hanoi case study when more metering devices are included in the system. Also, Figure 7 shows that the $\boldsymbol{\Omega}$ error has decreased with the addition of metering devices, reinforcing the validity of the constant $\boldsymbol{\Omega}$ hypothesis.

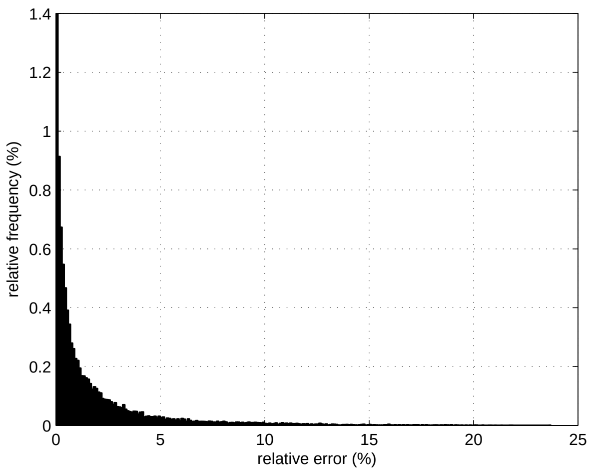

Fig. 7 Histogram of the relative error in the computation of the $\boldsymbol{\Omega}$ matrix in the 1000 Monte Carlo experiment with respect to the mean matrix assumed for the proposed methodology: five degrees of redundancy

5.2 Minimum leak assessment

In this section, the minimum leak that could be detected in different parts of the network is identified for both measurement settings. Note that the process for assessing minimum leaks (as described in Figure 3) fundamentally consists of repeating the probabilistic leak detectability assessment set out for different leakage values until the minimum is identified at every node of the system.

5.2.1 One degree of redundancy (x_{30})

Table 4 shows the minimum leak that could be detected at each node and the associated probability of detection once the algorithm for minimum leak assessment has converged (i.e. $P_{detect} \approx 0.8$) at all nodes. This information is also summarised in Figure 8, which provides the interpolated map of minimum leaks throughout the system. Note that nodes represent DMA themselves, which are connected to each other by means of the water transport network, and the shadowed region represents the area of each DMA. This figure shows that the values obtained are consistent with the measurement distribution throughout the network: leakage of around $80\ \mathrm{m}^3/\mathrm{h}$ could be detected in the surroundings of the redundant measurement x_{30}, but only major losses could be noticed on the right-hand side of the system. Moreover, leakage could not be sensed at all if it was happening at the branch that provides water from the tank. This figure shows that additional meters should be placed on the right-hand side of the network to improve absolute leak detectability in the system according to the selected operating condition.

The previous figure can also be obtained in relative terms by working with the minimum leak value divided by the maximum flow entering each node. Figure 9

Table 4 Minimum detectable leaks in Hanoi network case study: one degree of redundancy

Leaking node	Minimum leak value (m^3/h)	P_{detect}
2	-	0
3	386.7188	0.8006
4	367.1875	0.8043
5	343.7500	0.8048
6	320.3125	0.8033
7	316.4063	0.8071
8	287.1094	0.8034
9	267.5781	0.8041
10	253.9063	0.8042
11	253.9063	0.8042
12	253.9063	0.8041
13	253.9063	0.8042
14	234.3750	0.8055
15	201.1719	0.8039
16	185.5469	0.8008
17	230.4688	0.8067
18	289.0625	0.8040
19	347.6563	0.8049
20	304.6875	0.8048
21	304.6875	0.8048
22	304.6875	0.8048
23	187.5000	0.8074
24	149.4141	0.8038
25	131.8359	0.8040
26	154.2969	0.8009
27	161.1328	0.8040
28	151.3672	0.8032
29	100.5859	0.8051
30	75.1953	0.8017
31	78.1250	0.8043
32	109.3750	0.8042

shows the relative leak detectability map in the Hanoi case study with this one degree of redundancy measurement setting. It shows that on the lower right-hand side of the network there is better behaviour in terms of leakage awareness in this scenario, because even though the minimum leak that could be detected is greater when compared to the left-hand side (see Figure 8), the circulating flow is greater in this area (see Appendix S1). For this reason, around 20% relative leakage can be detected on the left-hand side, but this figure can fall to around 3% on the lower right-hand side. Therefore, additional meters should be added on the left-hand side of the system to improve relative leak detectability in the network based on the flow condition considered in the assessment.

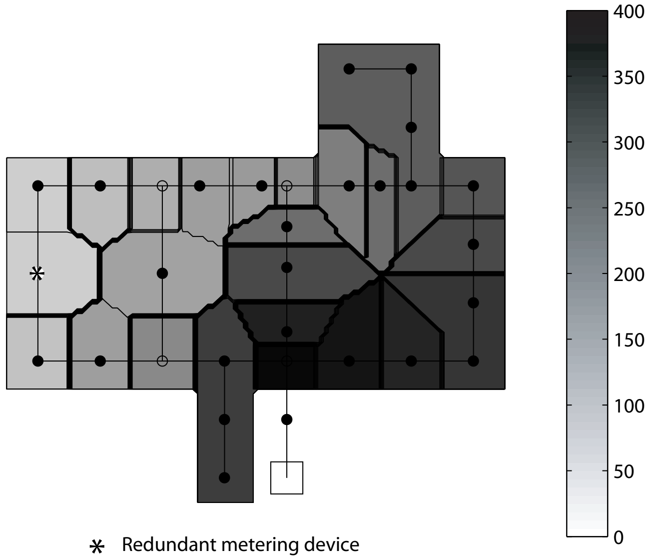

Fig. 8 Leak detectability map (in m^3/h) for Hanoi network case study: one degree of redundancy

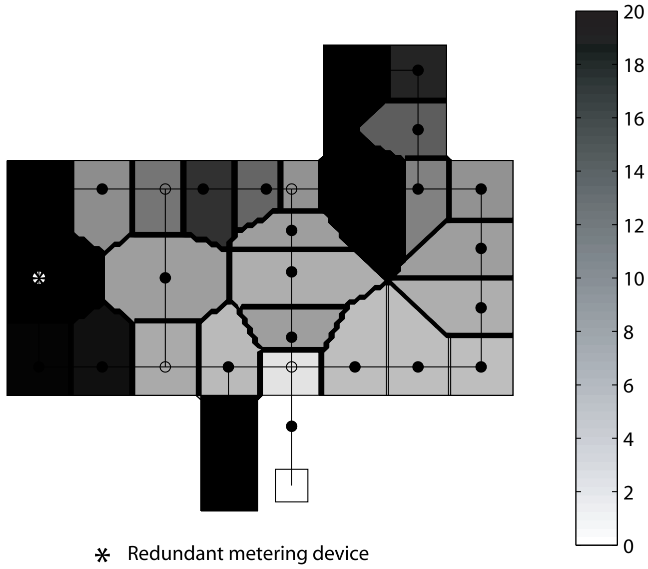

Fig. 9 Relative leak detectability map (in %) for Hanoi network case study: one degree of redundancy

Table 5 Minimum detectable leaks in Hanoi network case study: five degrees of redundancy

Leaking node	Minimum leak value (m^3/h)	P_{detect}
2	-	0
3	279.2969	0.8051
4	250.0000	0.8048
5	230.4688	0.8044
6	197.2656	0.8065
7	187.5000	0.8044
8	142.5781	0.8028
9	115.2344	0.8011
10	127.9297	0.8034
11	127.9297	0.8034
12	127.9297	0.8034
13	127.9297	0.8034
14	144.5313	0.8022
15	187.5000	0.8041
16	187.5000	0.8014
17	160.1563	0.8044
18	121.0938	0.8018
19	201.1719	0.8057
20	253.9063	0.8006
21	253.9063	0.8006
22	253.9063	0.8006
23	154.2969	0.8018
24	125.9766	0.8037
25	116.2109	0.8065
26	146.4844	0.8057
27	156.2500	0.8058
28	123.0469	0.8022
29	74.7070	0.8028
30	51.7578	0.8060
31	54.1992	0.8041
32	87.8906	0.8059

5.2.2 Five degrees of redundancy (x_9, x_{18}, x_{30}, Q_{3-4}, Q_{23-24})

Table 5, Figure 10 and Figure 11 provide the same information when more meters are added. Table 5 and Figure 10 show an improvement in the overall leak detection capability and it is possible to sense leaks of up to 51.76 m^3/h on the left-hand side of the system, which is remarkable considering that the tank provides 17565 m^3/h to the network. This improvement is also noticeable in relative terms (Figure 11). As before, these figures help to see where meters should be placed in order to reduce the absolute and relative leak detectability in the system for the specified operating condition and, hence, are a useful tool for practitioners.

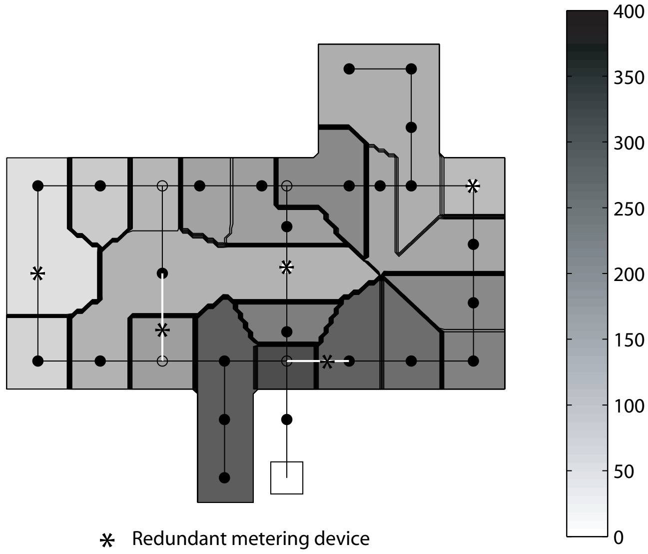

Fig. 10 Leak detectability map (in m^3/h) for Hanoi network case study: five degrees of redundancy

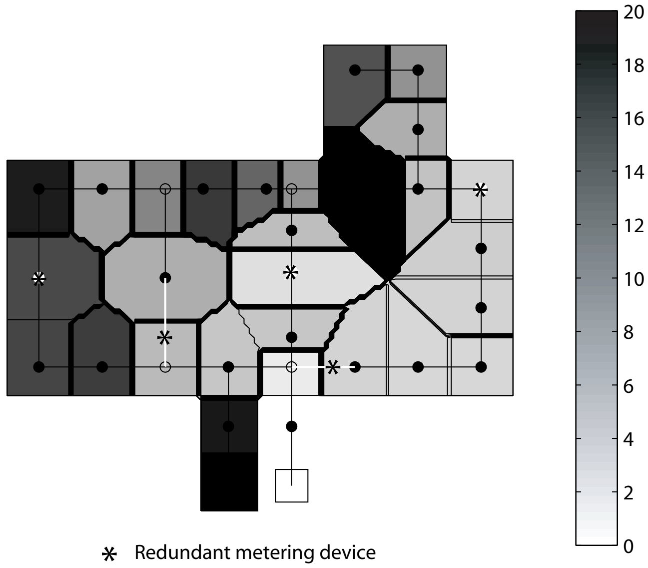

Fig. 11 Relative leak detectability map (in %) for Hanoi network case study: five degrees of redundancy

6 Discussion

The purpose of this section is to discuss the potential and limitations the methodology set out here has to carry out an assessment of a probabilistic leak detectability of a water supply system. As mentioned in the Introduction, the method put forward here has been specifically developed for water transport networks because these systems tend to be better metered than traditional water distribution systems. Note that identifying water transport pipelines may be complicated in some cases, but the International Water Association (IWA) recommends dividing water distribution systems into DMAs that must be connected to each other by means of water transport networks in order to better control water loss. Hence, modern water utilities are currently upgrading their systems to water transport network-DMA schemes, with DMA design itself being an active topic of research [10,33]. In general, at least one flow meter exists at the entrance to each DMA in order to measure the amount of water being provided to each sector, and additional devices are progressively being included to better characterise water flow through the main arteries. Note that instrumentation enhancement is also taking place on a DMA scale thanks to the reduction in terms of cost of both pressure and flow instrumentation [21]. Even if some of these measurements fail (e.g. sensor failure, communication failure), pseudo measurements (i.e. estimations of demand based on historical records) could be used instead [38]. This would mean increasing uncertainty in the problem and hence, it would have an impact on the leak detection potential of the network, but the methodology shown here would still be suitable.

The availability of metering devices plays a crucial role when characterising water loss. If the number of meters were sufficient and not noisy, a simple water balance could be applied. However, their inherent inaccuracy requires more sophisticated methods to take the effect of measurement noise into account. In this particular case, state estimation is used to provide the most likely hydraulic state of the system bearing in mind all the available measurements in the network. Note that state estimation techniques are required to process the on-line information provided nowadays by telemetry systems, but despite their massive use in the power field they have hardly been implemented in the water industry. One of the reasons for this is that the leak detection problem, as well as calibration or topological analysis, have been addressed in isolation from the state estimation conception. This paper represents an effort to tackle one of these traditional water systems problems from the state estimation perspective, emphasizing the appeal of adopting a comprehensive state estimation approach to extract as much information as possible from the available on-line measurements in the network.

Nevertheless, due to the immaturity of real-time state estimation techniques in water systems at present, the methodology explained here is oriented towards assessing if it would be possible to obtain leakage information from state estimation results rather than providing a method already suitable for the on-line detection of leaks. As mentioned in the Introduction, many off-line state estimation approaches have been developed over the past 20-30 years, but such techniques have not been successfully implemented on-line. For this reason, the methodology presented here enables tests to be carried out to find if a specific leak value could be detected at a given location with the available measurement setting, but it does not assess on-line measurements provided by telemetry systems. We believe that this prior analysis is essential to show water utilities and practitioners the genuine possibilities of detecting leakage based on state estimation results.

Moreover, the methodology explained herein can be used as the basis for optimal meter placement. Note that minimum leak value maps quickly show the regions of the network where there is less likelihood of noticing the loss of water, i.e. the regions of the network where more metering devices should be added. However, it must be stressed that the method shown at present provides clues but does not enable the most suitable location for additional metering devices to be identified systematically. This is because only one flow scenario has been considered in the case study for the probabilistic leak detectability assessment. However, water utilities are usually familiar with several normal operating scenarios in their networks and it is advisable to apply this methodology to as many flow configurations as possible in order to further comprehend the leakage awareness capability of the existing measurement setting in different circumstances. For this reason, specific optimal meter placement schemes [24] based on state estimation are a subject for further research. In any case, the minimum leak maps shown here are useful, because with them it can be determined if the available measurement setting is enough for the operating condition under consideration: if the minimum leak values are not within the desired levels, additional meters must be placed, but if they are, this analysis should encourage investors to implement state estimation techniques, which should in turn be adapted to detect and locate water loss in real time.

Note that adaption of state estimation techniques to the detection of leakage on-line is not straight forward. The same strategy of analysing state estimation residuals according to the largest normalised residual test can be used, but several operational aspects must be addressed first. To begin with, the method set out here to identify leaks is analogous to the traditional procedure for bad data identification, but bad measurements and leaks can coexist in real time and both should be addressed to ensure the tool is performing well. Also, the possibility of more than one leak occurring at the same time should be explored, as they may cancel each other out in some cases. For this reason, we believe that consideration of sequent hydraulic states (i.e. based on an extended period simulation) is essential for on-line detection, increasing the confidence in prediction. Similarly, other challenges associated with network modelling should be conveniently solved in order to develop a consistent real-time tool. In this regard, uncertain model parameters (e.g. pipe roughness coefficients), or unknown settings of valves should be conveniently adjusted before using state estimation to detect leaks in real-time. In other words, parameter estimation (i.e. calibration) should periodically be undertaken to ensure the hydraulic model remains valid. Note that a deviation in, for example, pipe roughness coefficients could mask abnormally high residuals. In this respect, [2] and [29] have presented some work on the possible causes for bad data, with the latter having identified five possible types: measurement noise, meter semi-failure, meter total failure, parametric model failure, and topological model failure. Therefore, residual processing tools must also be adapted to identify leaks as an additional cause of bad data. For this reason, the methodology set out here is an initial approach to the leakage awareness problem via state estimation, but further research is required to develop a robust on-line tool.

Once these limitations have been addressed, a suitable platform for on-line detection based on the largest normalised residual test could be used for detecting leaks in water systems. According to the analysis shown in this paper, this test has potential for identifying if there is any leakage, but the reality is that it may also give an idea of where water is being lost (i.e. leak localisation). Note that in Tables 4 and 5 only P_{detect} is shown, but the locations at which the maximum of $P_{detect_i}; \forall i = 1, \ldots, m$ is attained are also important. High detection probabilities are the result of high residuals, which in turn correspond to great differences between measurements and estimated variables. Table 6 ranks the top five measurements associated with the greatest detection probabilities in the minimum leak assessment analysis of the previous case study when one and five degrees of redundancy exist. Results show that with one degree of redundancy, information cannot be extracted about the approximate location of the leak. However, the measurement set associated with greater probabilities of detection varies with the location of the leak when there are five redundant measurements. For example, measurements h_1, q_2, q_{19}, q_{21} and q_{22} are associated with the greatest detection probabilities when the leak is artificially simulated at node 3, whereas measurements q_{32}, q_{31}, q_{30}, q_{29} and q_{26} are selected when the leak is simulated at node 32. Therefore, it can be concluded that when sufficient redundant measurements exist, the largest normalised residual test also has potential to give an idea of where water loss is taking place. This fact, together with a systematic assessment of how the remaining hydraulic variables evolve over time, must be explored in order to set up a consistent methodology for leak localisation, which is beyond the scope of this paper. Furthermore, leak information could be used as an input for reliability assessment [18], which is an ongoing topic of research in the field [32, 19].

7 Conclusions

In this paper, a new methodology for probabilistic leak detectability assessment is set out. This approach consists in analysing state estimation normalised residuals for a particular hydraulic state in the system, which are likely to be high when noticeable leakage exists. The procedure presented herein is conceived as a previous step that enables assessment of the extent to which leakage could be sensed in subsequent state estimation stage with the available noisy measurements. The probability of leak detection is calculated here considering the measurement-estimate joint bivariate distribution rather than undertaking sampling experiments (i.e. Monte Carlo method) with the traditional largest normalised residual test, which is time-consuming. Additionally, a procedure is shown to estimate the minimum leak that could be detected in different parts of the network at a later stage of state estimation.

The potential of this methodology is set out by means of a case study, which corresponds to a water transport network due to the better level of instrumentation of such networks in comparison with conventional water distribution systems. Results show that this alternative approach for computing the probability of leak detection provides sound approximation and is computationally much faster than the sampling procedure. Moreover, minimum leak value maps provide a fair overview of the sys-

Table 6 Ranking of top five measurements associated with the highest probabilities of detection in the minimum leak assessment for Hanoi network case study

	Set of 5 measurements with greatest P_{detect_i}	
Leaking node	One degree of redundancy	Five degrees of redundancy
2	-	-
3	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$h_1, q_2, q_{19}, q_{21}, q_{22}$
4	$q_{18}, q_7, q_8, q_{17}, q_{12}$	q_4, q_5, q_6, q_7, h_1
5	$q_{18}, q_7, q_8, q_{17}, q_{12}$	q_5, q_6, q_4, q_7, q_8
6	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_7, q_6, q_5, q_8, q_{14}$
7	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_7, q_6, q_8, q_{14}, q_{12}$
8	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_8, q_{12}, q_{10}, q_{11}, q_{13}$
9	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_9, q_{12}, q_{10}, q_{11}, q_{13}$
10	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{12}, q_{10}, q_{11}, q_{13}, q_{14}$
11	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{12}, q_{10}, q_{11}, q_{13}, q_{14}$
12	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{12}, q_{10}, q_{11}, q_{13}, q_{14}$
13	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{12}, q_{10}, q_{11}, q_{13}, q_{14}$
14	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{14}, q_{12}, q_{10}, q_{11}, q_{13}$
15	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{15}, q_{14}, q_5, q_6, q_7$
16	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{15}, q_{27}, q_{26}, q_{21}, q_{22}$
17	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{17}, q_{19}, q_{18}, h_1, q_2$
18	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{18}, q_{19}, q_{17}, h_{18}, h_1$
19	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{19}, q_{18}, q_{17}, h_1, q_2$
20	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{21}, q_{22}, q_{20}, h_1, q_2$
21	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{21}, q_{22}, q_{20}, h_1, q_2$
22	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{21}, q_{22}, q_{20}, h_1, q_2$
23	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{28}, q_{29}, q_{21}, q_{22}, q_{20}$
24	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{24}, q_{26}, q_{27}, q_{32}, q_{31}$
25	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{26}, q_{32}, q_{27}, q_{24}, q_{31}$
26	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{27}, q_{26}, q_{24}, q_{32}, q_{15}$
27	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{27}, q_{26}, q_{24}, q_{32}, q_{15}$
28	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{28}, q_{29}, q_{31}, q_{30}, q_{32}$
29	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{29}, q_{31}, q_{30}, q_{32}, h_{30}$
30	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{30}, q_{31}, q_{29}, q_{32}, h_{30}$
31	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{30}, q_{31}, q_{29}, q_{32}, h_{30}$
32	$q_{18}, q_7, q_8, q_{17}, q_{12}$	$q_{32}, q_{31}, q_{30}, q_{29}, q_{26}$

tem response in terms of leak detectability, and they constitute the foundation on which optimal meter placement strategies can be based.

Also, adaption of the largest normalised residual test for use in leakage awareness in real time is discussed. Several issues, such as the simultaneous presence of both erroneous measurements and leaks, the existence of more than one leak at a time, or consideration of parameter or topology uncertainty, must be addressed before using the test for on-line leak detection. Nevertheless, this application shows the potential state estimation has for leak detection, which has been hardly explored to date.

Therefore, it should be used to motivate further research on state estimation related techniques applied to detecting leaks.

References

1. Andersen, J.H., Powell, R.S., Marsh, J.F.: Constrained state estimation with applications in water distribution network monitoring. Int. J. Syst. Sci. **32**(6), 807–816 (2001)
2. Bargiela, A.: On-line monitoring of water distribution networks. Ph.D. thesis, Univ. of Durham, UK (1984)
3. Bargiela, A., Hainsworth, G.D.: Pressure and flow uncertainty in water systems. J. Water Resour. Plann. Manage. **115**(2), 212–229 (1989)
4. Burden, R.L., Faires, J.D.: Numerical analysis, 3rd Edition. PWS Publishers, Boston, USA (1985)
5. Cabrera, E., Almandoz, J., Arregui, F., García-Serra, J.: Auditoría de redes de distribución de agua. Ingeniería del Agua **6**(4), 387–399 (1999)
6. Caro, E., Conejo, A.J., Mínguez, R.: A sensitivity analysis method to compute the residual covariance matrix. Electr. Power Syst. Res. **81**(5), 1071–1078 (2011)
7. Caro, E., Conejo, A.J., Mínguez, R., Zima, M., Andersson, G.: Multiple bad data identification considering measurement dependencies. IEEE Trans. Power Syst. **26**(4), 1953–1961 (2011)
8. Carpentier, P., Cohen, G.: State estimation and leak detection in water distribution networks. Civ. Eng. Syst. **8**(4), 247–257 (1991)
9. Coulbeck, B.: Optimisation and modelling techniques in dynamic control of water distribution systems. Ph.D. thesis, Univ. of Sheffield, UK (1977)
10. Diao, K., Zhou, Y., Rauch, W.: Automated creation of district metered area boundaries in water distribution systems. J. Water Resour. Plann. Manage. **139**(2), 184–190 (2013)
11. Díaz, S., González, J., Mínguez, R.: Observability analysis in water transport networks: Algebraic approach. J. Water Resour. Plann. Manage. **142**(4), 04015,071 (2016). DOI 10.1061/(ASCE)WR.1943-5452.0000621
12. Díaz, S., González, J., Mínguez, R.: Uncertainty evaluation for constrained state estimation in water distribution systems. J. Water Resour. Plann. Manage. **142**(12), 06016,004 (2016). DOI 10.1061/(ASCE)WR.1943-5452.0000718
13. Díaz, S., Mínguez, R., González, J.: Stochastic approach to observability analysis in water networks. Ingeniería del Agua **20**(3), 139–152 (2016). DOI 10.4995/Ia.2016.4625
14. Díaz, S., Mínguez, R., González, J.: Calibration via multi-period state estimation in water distribution systems. Water Resources Management **31**(5), 4801–4819 (2017). DOI 10.1007/s11269-017-1779-2
15. Díaz, S., Mínguez, R., González, J.: Topological observability analysis in water distribution systems. J. Water Resour. Plann. Manage. p. 06017001 (2017). DOI 10.1061/(ASCE)WR.1943-5452.0000762
16. Fujiwara, O., Khang, D.B.: A two-phase decomposition method for optimal design of looped water distribution networks. Water Resour. Res. **26**(4), 539–549 (1990)
17. Giustolisi, O.: Considering actual pipe connections in water distribution network analysis. J. Hydraul. Eng. **136**(11), 889–900 (2010)
18. Goulter, I.: Analytical and simulation models for reliability analysis in water distribution systems. In: E. Cabrera, A.F. Vela (eds.) Improving efficiency and reliability in water distribution systems, pp. 235–266. Kluwer Academics, London, UK (1995)
19. Gouri, R.L., Srinivas, V.V.: A fuzzy approach to reliability based design of storm water drain network. Stoch. Environ. Res. Risk Assess. **31**(5), 1091–1106 (2017)
20. Hasofer, A.M., Lind, N.C.: Exact and invariant second-moment code format. J. Eng. Mech. Div. **100**(EM1), 111–121 (1974)
21. Imschoot, D.P.O., Furnass, W.R., Mounce, S.R., Boxall, J.B.: Flow-pressure sensor placement optimisation for pipe burst localisation in a water distribution network. 14th International Computing and Control for the Water Industry (CCWI) Conference, Amsterdam, The Netherlands (2016)
22. Kang, D., Lansey, K.: Optimal meter placement for water distribution system state estimation. J. Water Resour. Plann. Manage. **136**(3), 337–347 (2010)
23. Kang, D.S., Pasha, M.F.K., Lansey, K.: Approximate methods for uncertainty analysis of water distribution systems. Urban Water J. **6**(3), 233–249 (2009)
24. Kim, S.H., Aral, M.M., Eun, Y., Park, J.J., Park, C.: Impact of sensor measurement error on sensor positioning in water quality monitoring networks. Stoch. Environ. Res. Risk Assess. **31**(3), 743–756 (2017)

25. Melchers, R.E.: Structural reliability analysis and prediction. Wiley, New York, USA (1999)
26. Mínguez, R.: Seguridad, fiabilidad y análisis de sensibilidad en obras de ingeniería civil mediante técnicas de optimización por descomposición. aplicaciones. Ph.D. thesis, Univ. of Cantabria, Spain (2003)
27. Mounce, S.R., Boxall, J.B., Machell, J.: Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. J. Water Resour. Plann. Manage. **136**(3), 309–318 (2010)
28. Nagar, A.K., Powell, R.S.: Observability analysis of water distribution systems under parametric and measurement uncertainty. Build. Partnership pp. 1–10 (2000)
29. Powell, R.S.: On-line monitoring for operational control of water distribution networks. Ph.D. thesis, Univ. of Durham, UK (1992)
30. Puust, R., Kapelan, Z., Savic, D.A., Koppel, T.: A review of methods for leakage management in pipe networks. Urban Water J. **7**(1), 25–45 (2010)
31. Romano, M., Kapelan, Z., Savic, D.A.: Automated detection of pipe bursts and other events in water distribution systems. J. Water Resour. Plann. Manage. **140**(4), 457–467 (2014)
32. Roozbahani, A., Zahraie, B., Tabesh, M.: Integrated risk assessment of urban water supply systems from source to tap. Stoch. Environ. Res. Risk Assess. **27**(4), 923–944 (2013)
33. Savic, D., Ferrari, G.: Design and performance of district metering areas in water distribution systems. Procedia Eng. **89**, 1136–1143 (2014)
34. Savic, D.A., Kapelan, Z.S., Jonkergouw, P.M.R.: Quo vadis water distribution model calibration? Urban Water J. **6**(1), 3–22 (2009)
35. Schweppe, F.C., Handschin, E.J.: Static state estimation in electric power systems. Proc. IEEE **62**(7), 972–982 (1974)
36. Schweppe, F.C., Wildes, J.: Power system static-state estimation, Part I: Exact model. IEEE Trans. Power Appar. Syst. **PAS-89**(1), 120–125 (1970)
37. Vrachimis, S.G., Eliades, D.G., Polycarpou, M.M.: Real-time hydraulic interval state estimation for water transport networks: A case study. 14th International Computing and Control for the Water Industry (CCWI) Conference, Amsterdam, The Netherlands (2016)
38. Yang, T., Shi, P., Yu, Z., Li, Z., Wang, X., Zhou, X.: Probabilistic modeling and uncertainty estimation of urban water consumption under an incompletely informational circumstance. Stoch. Environ. Res. Risk Assess. **30**(2), 725–736 (2016)
39. Yu, G., Powell, R.S.: Optimal design of meter placement in water distribution systems. Int. J. Syst. Sci. **25**(12), 2155–2166 (1994)