Modelling uncertainty for leak localization in Water Networks ⋆

Joaquim Blesa [*,**] **Ramon Pérez** [**]

[*] *Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC, C/. Llorens i Artigas 4-6, 08028 Barcelona, Spain. (e-mail: joaquim.blesa@upc.edu)*
[**] *Research Center for Supervision, Safety and Automatic Control (CS2AC), Universitat Politècnica de Catalunya (UPC), Terrassa Campus, Gaia Building, Rambla Sant Nebridi 22, 08222 Terrassa, Barcelona. (e-mail: ramon.perez@upc.edu)*

Abstract: The performance and success of model-based leak localization methods applied to water distribution networks (WDN) highly depends on the uncertainty of the system considered. This work proposes an original method of modeling the effect of uncertainties in these networks. The proposed method is based on the collection of real data in the water network in the absence of leaks. The discrepancy (residual) between the measured data and the one provided by a simulator of the network in normal operation is used to extrapolate the possible residuals in the different leak scenarios. In addition, indicators for assessing the effect of uncertainty in the performance of leak localization methods based on residual correlation analysis are provided. The error in terms of correlation intervals and leak localzation assessment between the proposed approximation and the real one is studied by means a simplified model of the WDN of Hanoi (Vietnam).

Keywords: water distribution networks, fault detection and izolation, leak localzation, correlation analysis

1. INTRODUCTION

One of the main issues that motivates the use of water distribution models and on-line measurements is the leakage detection and localization. There are methods based on data analysis (van Thienen et al. (2013)) but most of the methods that use on-line measurements rely on models, either transient (Colombo et al. (2009)) or static (Wu and Sage (2006),Goulet et al. (2013),Sanz and Pérez (2015)). The Centre for Supervision, Safety and automatic Control (cs2ac) at the UPC in Terrassa has developed a methodology for leak localization using pressure measurements and hydraulic models. It is based on the fault detection and isolation theory (Gertler (1998)) and it evolved from a first version where binary residuals where generated (Pérez et al. (2011)) to a correlation based method (Quevedo et al. (2011)) . It was successfully applied in real networks (Perez et al. (2014)). The improvement of leak localization includes contributions from other disciplines such as sensor placement demand calibration and the accuracy assessment. The accuracy in the leak localization is determined by the uncertainties present both in measurements and models. These uncertainties have to be modelled (Pérez et al. (2015)) so that the accuracy can be assessed. In (Cugueró-Escofet et al. (2015)) the methodology was adapted to provide the results in the leak localization taking into account the uncertainty. It was demonstrated for a three-node network. The application to a bigger network is not possible without an approximation. Uncertainty in non-leaky scenarios can be computed as the difference between actual pressure measurements and estimations provided by hydraulic water distribution models when no leak is present in the system. When a leak is present in the system, leak detection methods provide an estimation of the range of the leak magnitude by the study of the change of the night flow. Then, the most challenging problem is to determine where the leak is. Model-based methods compare the behaviour of the observed residuals with the ones generated by the model considering all the possible leak locations. The most similar behaviour between the actual residuals and the theoretical ones determines the most probable leak location. As uncertainties are present in the system, they have to be taken into account to perform the leak localization. In this paper, it is proposed that the uncertainty observed in a non-leaky scenario can be extrapolated to leaky scenarios. In particular, the effect of the uncertainty in leaky scenarios can be approximated by the sum of the effect of nominal leak magnitude (provided by the leak detection) in residuals and the uncertainty computed in non-leaky scenarios.

The rest of the paper is organized as follows. Section 2 reviews the leak localization methodology for water distri-

⋆ This work has been partially funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERFD) through the projects DEOCS (ref. MINECO DPI2016-76493), SCAV (ref. MINECO DPI2017-88403-R) and through the grant IJCI-2014-20801. This work has also been partially funded by AGAUR of Generalitat de Catalunya through the Advanced Control Systems (SAC) group grant (2017 SGR 482).

bution networks that is going to be assessed. In Section 3 it is presented the methodology to approximate leaky residual sets considering uncertainty and a correlation error is proposed to measure the effect of residual approximated. Confusion coefficients and an isolability indicator are proposed to assess leak localization methods in Section 3. The proposed methodology is applied to the simplified water distribution network of Hanoi that has been studied in several previous works of the group (Ferrandez-Gamot et al. (2015),Soldevila et al. (2016),Soldevila et al. (2017)) and the error of the approximation is assessed.

2. BACKGROUND

Given the network boundary conditions $\mathbf{h_S} \in \Re^{n_S}$ in the form of heads in n_S nodes and the demands $\mathbf{d} \in \Re^{n_n}$ in the n_n nodes of the network. Pressure in demand nodes $\hat{\mathbf{y}} \in \Re^{n_y}$, where n_y is the number measured node pressures, can be computed as

$$\hat{\mathbf{y}} = g_h(\mathbf{h}_S, \mathbf{d}) \tag{1}$$

where $g_h : \Re^{n_S} \times \Re^{n_d} \to \Re^{n_y}$ is a non-linear hydraulic function. We consider in this work that leaks can appear in demand nodes and the effect of a leak in a node can be modeled as a change of the demand pattern. In a non-faulty scenario (i.e. non-leakage scenario) pressure in demand nodes $\hat{\mathbf{y}}_{nf} \in \Re^{n_y}$ can be computed using (1) considering normal pattern demand

$$\mathbf{d} = \boldsymbol{\alpha} q_{in} \tag{2}$$

where $\boldsymbol{\alpha} \in \Re^{n_n}$ is a vector that contains the weights of demand nodes $\alpha_1, \ldots \alpha_{n_n}$ with $\sum_{i=1}^{n_n} \alpha_i = 1$ and $q_{in} \in \Re$ is the total inflow.

The difference between actual pressure measurements $\mathbf{y}$ and the predicted ones $\hat{\mathbf{y}}_{\mathbf{nf}}$

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}_{\mathbf{nf}} \tag{3}$$

that quantifies the consistency of the measurement with the model prediction is called a *residual*. We will also call it *observed residual* to distinguish it from *predicted residual* as it will bee seen later. If there is no uncertainty in model (1), the absence of leakage implies $\mathbf{r} = 0$. In presence of leak in node i the residual is:

$$\mathbf{r}_{\mathbf{f_i}} = \mathbf{y}_{\mathbf{f_i}} - \hat{\mathbf{y}}_{\mathbf{nf}} \tag{4}$$

In a leakage scenario, only the possibility of one leak of nominal value f_0 in an unknown node of the network is considered. The n_n predictions $\hat{\mathbf{y}}_{\mathbf{f_i}} \in \Re^{n_y}$, where subscript $\mathbf{f}_i$ indicates a faulty scenario corresponding to a leak of nominal value f_0 in node i, can be computed using (1) considering a leak pattern demand

$$\mathbf{d} = \boldsymbol{\alpha}(q_{in} - f_0) + \mathbf{f}_i \tag{5}$$

with $\mathbf{f}_i \in \Re^{n_n}$ a vector whose components are zero except component i which is f_0. The differences

$$\hat{\mathbf{r}}_{\mathbf{f_i}} = \hat{\mathbf{y}}_{\mathbf{f_i}} - \hat{\mathbf{y}}_{\mathbf{nf}} \tag{6}$$

are the predicted residuals for the nominal leak f_0 in node i. If there is no uncertainty in model (1) and the value of the unknown leak to be located is small enough, then the dependency of the observed residual $\mathbf{r}$ can be supposed to be approximately linear in f

$$\mathbf{r}_{\mathbf{f_i}} = \hat{\mathbf{r}}_{\mathbf{f_i}} \cdot \frac{f}{f_0} \qquad i = 1 \ldots n_n \tag{7}$$

Because of linearity of $\mathbf{r}_{\mathbf{f_i}}$ in f, if vectors $\hat{\mathbf{r}}_{\mathbf{f_i}}$ are linearly independent, then each $\hat{\mathbf{r}}_{\mathbf{f_i}}$ characterizes a different leak. Therefore a correlation measure to test linear dependency between $\mathbf{r}$ and $\hat{\mathbf{r}}_{\mathbf{f_i}}$ can be used to select the most consistent leak with $\mathbf{r}$. Thus the selected leak is the one maximizing the correlation measure

$$\rho(\mathbf{r}, \hat{\mathbf{r}}_{\mathbf{f_i}}) = \frac{\mathbf{r}^T \cdot \hat{\mathbf{r}}_{\mathbf{f_i}}}{\parallel \mathbf{r} \parallel \parallel \hat{\mathbf{r}}_{\mathbf{f_i}} \parallel} \tag{8}$$

where $\parallel . \parallel$ denotes the norm associated to the vector dot product. In this work the 2 norm is used. Note that if (7) has additive uncertainty, then the selection of $\hat{\mathbf{r}}_{\mathbf{f_i}}$ by maximizing the correlation measure ρ gives the least squares solution of (7).

In Cugueró-Escofet et al. (2015) uncertainty in pattern demand components $\alpha_i \quad i = 1, ..., n_n$ was considered as

$$\alpha_{imin} \leq \alpha_i \leq \alpha_{imax} \quad \text{for} \quad 1 \leq i \leq n_n \tag{9}$$

subject to

$$\sum_{i=1}^{n_n} \alpha_i = 1 \tag{10}$$

considering nominal demand

$$\mathbf{d^0} = \boldsymbol{\alpha}^0 q_{in} \tag{11}$$

where $\boldsymbol{\alpha}^0$ is a vector that contains nominal weights of demand nodes $\alpha_1^0, \ldots \alpha_{n_n}^0$. The predicted residual calculated in (6) for the nominal weigths $\boldsymbol{\alpha}^0$ is

$$\hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_i}} = \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{f_i}} - \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{nf}} \tag{12}$$

Furthermore observed residual (4) considering uncertainty (9) can be bounded by a set

$$\mathbf{r}_{\mathbf{f_i}} \in \mathcal{R}_{f_i} \qquad i = 1 \ldots n_n \tag{13}$$

where $\mathcal{R}_{f_i} \subset \Re^{n_y}$ and $\hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_i}} \in \mathcal{R}_{f_i}$.

The evaluation of the correlation measures in (8) for all the observed residuals $\mathbf{r}_{\mathbf{f_i}}$ in a set $\mathcal{R}_{f_i}$ and nominal hypothesis $\hat{\mathbf{r}}^0_{\mathbf{f_j}}$ $\forall i, j = 1, \ldots, n_n$ gives a matrix of intervals $[\rho](\mathcal{R}_{f_i}, \hat{\mathbf{r}}^0_{\mathbf{f_j}})$. Rows and columns correspond to the bounding leaky residual sets and to the nominal theoretical leaky residual vectors, respectively.

$$[\rho](\mathcal{R}_{f_i}, \hat{\mathbf{r}}^0_{\mathbf{f_j}}) = \left\{ \frac{\mathbf{r}^T \cdot \hat{\mathbf{r}}^0_{\mathbf{f_j}}}{\parallel \mathbf{r} \parallel \left\| \hat{\mathbf{r}}^0_{\mathbf{f_j}} \right\|} : \mathbf{r} \in \mathcal{R}_{f_i} \right\} \tag{14}$$

A leak localization method based on a falsification process that considers uncertainty bounds in residuals was proposed in Pérez et al. (2016). This method is summarized in Algorithm 1. This algorithm provides the possible leak locations consistent with the considered residual uncertainty given the actual residual $\mathbf{r}$, the n_n nominal leak hypothesis $\hat{\mathbf{r}}^0_{\mathbf{f_i}}$ and correlation boundaries $\underline{\rho}_{i,i}, \overline{\rho}_{i,i}$ between the uncertain set $\mathcal{R}_{f_i}$ and nominal hypothesis $\hat{\mathbf{r}}^0_{\mathbf{f_i}}$ computed as

$$[\rho](\mathcal{R}_{f_i}, \hat{\mathbf{r}}^0_{\mathbf{f_i}}) = [\underline{\rho}_{i,i}, \overline{\rho}_{i,i}] \quad (15)$$

that considering $\hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_i}} \in \mathcal{R}_{f_i}$ implies $\overline{\rho}_{i,i} = 1 \ \forall i = 1, ..., n_n$.

Algorithm 1 Leak localization algorithm

Require: $\mathbf{r}$, $\hat{\mathbf{r}}^0_{\mathbf{f_i}}$, $\underline{\rho}_{i,i}$ $\quad i = 1, ..., n_n$
  leak=ones
    for $i = 1...n_n$
      compute $\rho(\mathbf{r}, \hat{\mathbf{r}}_{\mathbf{f_i}})$
      if $\rho(\mathbf{r}, \hat{\mathbf{r}}_{\mathbf{f_i}}) < \underline{\rho}_{i,i}$
        leak(i)=0
      end
    end
  return leak

As a result of Algorithm 1, Vector **leak** contains 1 for those leak hypothesis assigned with leak.

As was proposed in Cugueró-Escofet et al. (2015) a straightforward way to transmit the uncertainty from demands to the residuals through the pressures is to generate a set of possible demand realizations $\mathbf{d}(l) \quad l = 1 \dots N$ by means of N realizations of $\boldsymbol{\alpha}(l)$ considering (9) and (10) and applying (5) for every different leak $i = 1 \dots n_n$. Once the set of possible demand realizations have been generated, N realizations of pressure residuals $\hat{\mathbf{r}}_{\mathbf{f_i}}(l)$ $l = 1 \dots N$ can be computed following residual equation (6) as

$$\hat{\mathbf{r}}_{\mathbf{f_i}}(l) = \hat{\mathbf{y}}_{\mathbf{f_i}}(l) - \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{nf}} \quad l = 1 \dots N \quad (16)$$

where $\hat{\mathbf{y}}_{\mathbf{f_i}}(l)$ is computed by means (1) considering demand $\mathbf{d}(l)$ with leak in node i (Eq. (5)).

Then, residual set $\mathcal{R}_{f_i}$ can be approximated by the sampled set

$$\hat{\mathcal{R}}_{f_i} = \{\hat{\mathbf{r}}_{\mathbf{f}i}(1), \dots, \hat{\mathbf{r}}_{\mathbf{f}i}(N)\} \quad (17)$$

As the cardinal N of this set increases, the coverage of the set $\mathcal{R}_{f_i}$ improves. Once sampled sets $\hat{\mathcal{R}}_{f_i}$ $i = 1, ..., n_n$ are computed, compact zonotopic approximations of $\mathcal{R}_{f_i}$ Blesa et al. (2012) can be computed as was proposed in Cugueró-Escofet et al. (2015).

Once the sampled sets of possible residuals for each leak are generated, the correlation interval bounds $\hat{\underline{\rho}}_{i,i}$ defined in (15) can be approximated using (14) and $\hat{\mathcal{R}}_{f_i}$ as

$$\begin{aligned} \hat{\underline{\rho}}_{i,i} = \min_{\mathbf{r}} \quad & \hat{\rho}(\mathbf{r}, \hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_i}}) \\ \text{subject to} \quad & \mathbf{r} \in \hat{\mathcal{R}}_{f_i} \end{aligned} \quad (18)$$

If we assume that the actual leak is f_i the correlation of the residual $\mathbf{r}$ with $\hat{\mathbf{r}}^0_{\mathbf{f_i}}$ must be inside this interval.

3. RESIDUAL UNCERTAINTY GENERATION

In this Section a way to compute a sample set $\tilde{\mathcal{R}}_{f_i} \subset \Re^{n_y}$ that approximates the set of residuals $\mathcal{R}_{f_i}$ for every leak $i = 1, ..., n_n$ is presented. This set considers only uncertainty in user demand weights and additive noise for simplicity and understanding, but it could be extended also to uncertainty in boundary conditions and leak magnitude.

Given a set of measured data in a fault free scenario:

$$\mathbf{y_{nf}}(l) \quad l = 1, ..., N_{nf} \quad (19)$$

that can be computed as

$$\mathbf{y_{nf}}(l) = g_h(\mathbf{h}_S, \mathbf{d}(l)) + \mathbf{e}(l) \quad l = 1, ..., N_{nf} \quad (20)$$

where $\mathbf{e}(l)$ is additive noise and $\mathbf{d}(l)$ are actual demand nodes computed as

$$\mathbf{d}(l) = \boldsymbol{\alpha}(l) q_{in} \quad (21)$$

with $\boldsymbol{\alpha}(l)$ actual user demand weights around $\boldsymbol{\alpha}^0$. Then, the following fault free residuals can be computed as:

$$\mathbf{r_{nf}}(l) = \mathbf{y_{nf}}(l) - \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{nf}} \quad (22)$$

The set that contains all computed residuals

$$\tilde{\mathcal{R}}_{nf} = \{\mathbf{r_{nf}}(l)\} \quad l = 1, ..., N_{nf} \quad (23)$$

belong to a set $\mathcal{R}_{nf}$ that bounds the actual residual uncertainty in fault free scenario. Another way to compute an approximation of $\tilde{\mathcal{R}}_{nf}$ would be by means of first order Taylor approximation:

$$\mathbf{y_{nf}}(l) \approx g_h(\mathbf{h}_S, \mathbf{d}^0) + \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0} (\boldsymbol{\alpha}(l) - \boldsymbol{\alpha}^0) + \mathbf{e}(l) \quad (24)$$

Then, considering (22) and $g_h(\mathbf{h}_S, \mathbf{d}^0) = \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{nf}}$

$$\mathbf{r_{nf}}(l) \approx \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0} (\boldsymbol{\alpha}(l) - \boldsymbol{\alpha}^0) + \mathbf{e}(l) \quad (25)$$

and with N_f realizations of $\boldsymbol{\alpha}(l)$ and $\mathbf{e}(l)$ in (25) approximation (23) can be computed. On the other hand, in a faulty scenario

$$\mathbf{d} = \mathbf{d_{f_i}}^0 = \boldsymbol{\alpha}^0 (q_{in} - f_0) + \mathbf{f}_i \quad (26)$$

$$\mathbf{y_{f_i}}(l) \approx g_h(\mathbf{h}_S, \mathbf{d_{f_i}}^0) + \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d_{f_i}}^0} (\boldsymbol{\alpha}(l) - \boldsymbol{\alpha}^0) + \\ + \mathbf{e}(l) \quad (27)$$

considering (4) with $\hat{\mathbf{y}}_{\mathbf{nf}} = \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{nf}}$ and $g_h(\mathbf{h}_S, \mathbf{d_{f_i}}^0) = \hat{\mathbf{y}}^{\mathbf{0}}_{\mathbf{f_i}}$

$$\mathbf{r_{f_i}}(l) \approx \hat{\mathbf{r}}^0_{\mathbf{f_i}} + \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0_{\mathbf{f_i}}} (\boldsymbol{\alpha}(l) - \boldsymbol{\alpha}^0) + \mathbf{e}(l) \quad (28)$$

as

$$\begin{aligned} & \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d_{f_i}}^0} = \frac{\partial g_h}{\partial \mathbf{d}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d_{f_i}}^0} (q_{in} - f_0) \text{ and} \\ & \frac{\partial g_h}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\alpha}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0} = \frac{\partial g_h}{\partial \mathbf{d}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0} q_{in} \end{aligned} \quad (29)$$

considering small leak magnitudes $f_0 << q_{in}$ that implies $q_{in} - f_0 \approx q_{in}$ and $\mathbf{d}^{\mathbf{0}}_{\mathbf{f_i}} \approx \mathbf{d}^0$. Therefore

$$\frac{\partial g_h}{\partial \mathbf{d}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d_{f_i}}^0} (q_{in} - f_0) \approx \frac{\partial g_h}{\partial \mathbf{d}}|_{\mathbf{h}_S, \mathbf{d}=\mathbf{d}^0} q_{in} \quad (30)$$

Then, $\mathbf{r_{f_i}}(l)$ can be approximated $\forall l = 1, ..., N_{nf}$ by

$$\mathbf{r_{f_i}}(l) \approx \hat{\mathbf{r}}^0_{\mathbf{f_i}} + \mathbf{r_{nf}}(l) \quad (31)$$

Finally, for every leak $i = 1, ..., n_n$ a sample set $\tilde{\mathcal{R}}_{f_i}$ that approximates the set of residuals $\mathcal{R}_{f_i}$ can be computed by

$$\tilde{\mathcal{R}}_{f_i} \approx \hat{\mathbf{r}}^0_{\mathbf{f_i}} + \tilde{\mathcal{R}}_{nf} \quad (32)$$

The main advantage of obtaining the approximation of leaky residual sets $\mathcal{R}_{f_i}$ by means of Eq. (32) is that only data in non leak scenario is necessary to determine the uncertainty of the model. The error of this approximation in a complex WDN is rather difficult to evaluate analytically. In this paper a case study for the the analysis of the effect of the approximation error in the leak localization performance will be used.

3.1 Correlation approximation error

In order to measure the effects of the error of approximating residual sets $\mathcal{R}_{f_i}$ by $\tilde{\mathcal{R}}_{f_i}$ for $i = 1, ..., n_n$, an error $e_{i,j}$ that measures the difference between correlations obtained by real sets $[\rho](\mathcal{R}_{f_i}, \hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_j}}) = [\underline{\rho}_{i,j}, \overline{\rho}_{i,j}] \triangleq [\rho_{i,j}]$ and by approximated sets $[\rho](\tilde{\mathcal{R}}_{f_i}, \hat{\mathbf{r}}^{\mathbf{0}}_{\mathbf{f_j}}) = [\underline{\tilde{\rho}}_{i,j}, \overline{\tilde{\rho}}_{i,j}] \triangleq [\tilde{\rho}_{i,j}]$ is defined $\forall i, j = 1, ..., n_n$ as

$$e_{i,j} = \frac{\overline{\rho}_{i,j} - \underline{\rho}_{i,j} + \overline{\tilde{\rho}}_{i,j} - \underline{\tilde{\rho}}_{i,j} - 2lgth([\rho_{i,j}] \cap [\tilde{\rho}_{i,j}])}{\overline{\rho}_{i,j} - \underline{\rho}_{i,j}} \quad (33)$$

where where $lgth([a, b]) = b - a$. This error term satisfies $e_{i,j} \geq 0$. If $e_{i,j} = 0$ the possible errors between approximation set $\tilde{\mathcal{R}}_{f_i}$ and real set $\mathcal{R}_{f_i}$ does not affect to the correlation interval between set of leak i and hypothesis j. From the different n_n^2 error terms $e_{i,j}$ an average error e is defined as

$$e = \frac{1}{(n_n)^2} \sum_{i=1}^{n_n} \sum_{j=1}^{n_n} e_{i,j} \quad (34)$$

4. LEAK LOCALIZATION ASSESSMENT

From leak localization Algorithm 1, it can be deduced that if residual sets $\mathcal{R}_{f_i} \forall i$ are calibrated properly, i.e. condition (13) is guaranteed for all possible leak residuals, the actual leak will never be rejected as hypothesis. However, the main problem of Algorithm 1 is that it can provide more than one leak location candidate i.e. residuals produced by a leak i can also be classified as another leak. Given $[\rho](\mathcal{R}_{f_i}, \hat{\mathbf{r}}^0_{\mathbf{f_j}})$ $\forall i, j = 1, ..., n_n$, for any pair of leaks i, j a confusion coefficient $c_{i,j}$ can be defined as

$$c_{i,j} = \frac{lgth([\rho_{i,j}] \cap [\rho_{j,j}])}{lgth([\rho_{i,j}])} \quad (35)$$

Confusion coefficient satisfies $0 \geq c_{i,j} \leq 1$. If $c_{i,j} = 0$ leak i is never classified as leak j and if $c_{i,j} = 1$ leak i is always classified as leak j (for example $c_{i,i} = 1 \ \forall i$). Finally, if $0 < c_{i,j} < 1$ the value gives an idea about the probability of leak i be classified by algorithm 1 as leak j in one time step. Therefore, condition

$$[\rho_{i,j}] \cap [\rho_{j,j}] = \emptyset \qquad \forall j \neq i \quad (36)$$

must be satisfied in order to guarantee that Algorithm 1 perfectly locates (without confusion) a leak i in one time step ($c_{i,j} = 0 \quad \forall j \neq i$). In order to have a general idea about the degree of isolability an indicator δ can be defined as

$$\delta = \frac{\left(\sum_{i=1}^{n_n} \sum_{j=i+1}^{n_n} (1 - c_{i,j}) + \sum_{j=1}^{n_n} \sum_{i=j+1}^{n_n} (1 - c_{i,j})\right)}{(n_n - 1)^2} \quad (37)$$

Indicator δ satisfies $0 \leq \delta \leq 1$. The two terms of the numerator in (37) appear because elements $c_{i,j}$ and $c_{j,i}$ can be different.

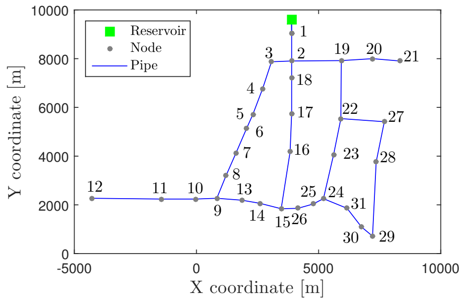

Fig. 1. Hanoi topological network.

4.1 Assessment error

Considering approximation sets $\tilde{\mathcal{R}}_{f_i}$ an approximation of confusion coefficient (35) can be computed as

$$\tilde{c}_{i,j} = \frac{lgth([\tilde{\rho}_{i,j}] \cap [\tilde{\rho}_{j,j}])}{lgth([\tilde{\rho}_{i,j}])} \quad (38)$$

On the other hand, an approximation $\tilde{\delta}$ of isolability index δ can be obtained considering $\tilde{c}_{i,j}$ instead of $c_{i,j}$ in (37).

In order to measure the effect of the error in the leak localization assessment an error of the isolability indicator δ can be defined as

$$e_\delta = \frac{\sum_{i=1}^{n_n} \sum_{j=i+1}^{n_n} |c_{i,j} - \tilde{c}_{i,j}| + \sum_{j=1}^{n_n} \sum_{i=j+1}^{n_n} |c_{i,j} - \tilde{c}_{i,j}|}{(n_n - 1)^2} \quad (39)$$

5. CASE STUDY

In this section the proposed methodology is applied to the simplified model of the Hanoi DMA network. This simplified model, depicted in Fig. 1, consists of one reservoir, 34 pipes and 31 nodes. Two inner pressure sensors placed in nodes 14 and 30 have been considered enough for leak localization as it is detailed in Casillas et al. (2013), i.e. the dimension of the pressure measurement, estimation and residual vectors is $n_y = 2$.

The demand pattern in all demand nodes has been considered known but with an uncertainty of 1% (i.e, $\left(\alpha_{imax} - \alpha_i^0\right)/\alpha_i^0 = \left(\alpha_i^0 - \alpha_{imin}\right)/\alpha_i^0 = 0.01$), with a total water consumption $q_{in} = 5764 l/s$ and a bounded sensor pressure additive noise $|e(l)| \leq 1cm$ has been considered in both pressure sensors. Single-leak scenarios in the 31 nodes of the network have been considered in the leak localization performance.

First, measure residual set $\tilde{\mathcal{R}}_{nf} = \{\mathbf{r}_{nf}(l)\} \quad l = 1, ..., N_{nf}$ is computed by means of (22) with $\mathbf{y}_{nf}(l)$ generated by (20) considering $N_{nf} = 1000$ different realization of user demand weights $\boldsymbol{\alpha}(l)$ and additive errors $\mathbf{e}(l)$ and $\hat{\mathbf{y}}^0_{nf}$ obtained considering nominal demand weights $\boldsymbol{\alpha}^0(l)$. The obtained measure residual set $\tilde{\mathcal{R}}_{nf}$ is depicted in Fig. 2.

Once $\tilde{\mathcal{R}}_{nf}$ has been computed, approximated residuals $\tilde{\mathcal{R}}_{f_i} \approx \tilde{\mathcal{R}}_{nf} + \hat{\mathbf{r}}^0_{\mathbf{f_i}}$ $i = 1, ..., 31$ have been computed for leak

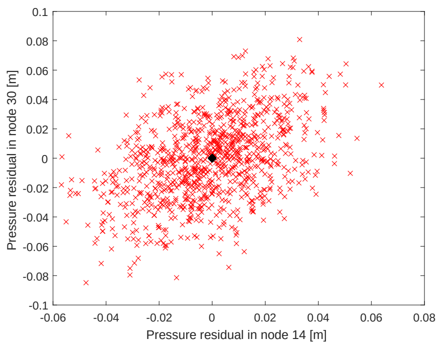

Fig. 2. Approximate non-faulty residual set $\tilde{\mathcal{R}}_{nf} = \{\mathbf{r}_{nf}(l)\} \quad l = 1, ..., N_{nf}$.

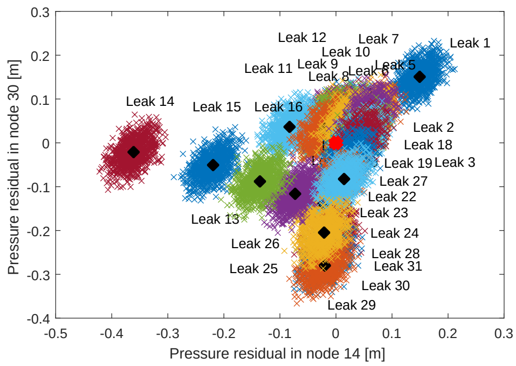

Fig. 3. Residuals considering leak magnitude 10[l/s].

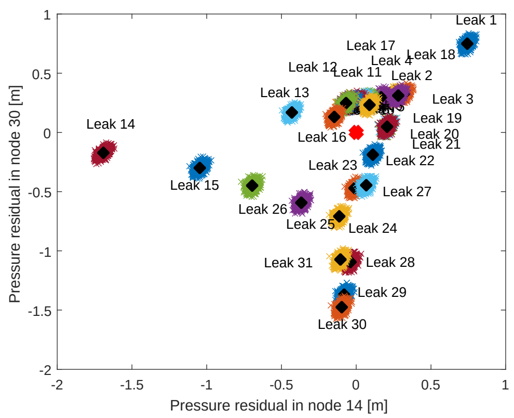

Fig. 4. Residuals considering leak magnitude 50[l/s].

magnitudes f_0 from $10l/s$ to $50l/s$. Approximated residual leak sets $\tilde{\mathcal{R}}_{f_i}$ $i = 1, ..., 31$ are depicted for $f_0 = 10l/s$ (Fig. 3) and for $f_0 = 50l/s$ (Fig. 4).

In order to assess the approximation error, residual sets $\mathcal{R}_{f_i}$ have been computed by a set of residuals $\{\mathbf{r_f}_i(1), \ldots, \mathbf{r_f}_i(N)\}$ generated by means of (4) considering N=1000 realizations of user demands (5) and additive noise $\mathbf{e}(l)$ as proposed in Cugueró-Escofet et al. (2015). Then, the average correlation error e defined in (34) and the isolability indicator error e_δ defined in (39) have been computed for the different leak magnitudes. The evolution of these two errors is depicted in Fig. 5. As it can be seen the average correlation error e increases with the leak magnitude. However the effect of this error in the isolability indicator error e_δ decreases with the leak magnitude. This fact is due to despite the approximation error increases with the leak magnitude, the residual sets are more separated and the effect of errors in correlation is smaller. The evolution of the isolability index δ defined in (37) and its approximated value $\tilde{\delta}$ is depicted in Fig. 6, as it can be seen the index and its approximation increase with the leak magnitude and their difference is almost imperceptible.

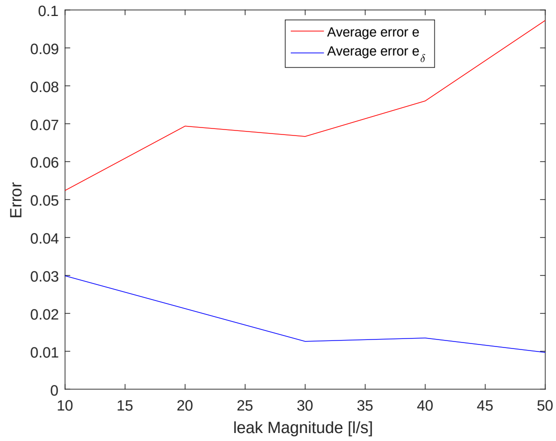

Fig. 5. Error approximations.

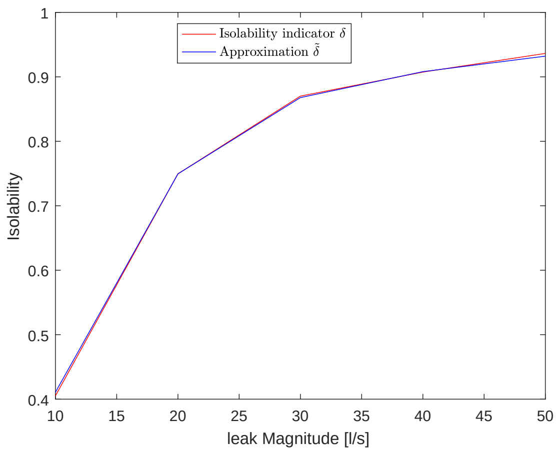

Fig. 6. Isolability Index and approximation

6. CONCLUSIONS

This work proposes one method of modeling the effect of uncertainties in water distribution networks for model-based leak localization purposes. Model-based leak localization are based in the evaluation of residual between actual measurements and values provided by a model. The main problem of these methods is that residuals can be different from zero even in the absence of leaks due to differences between the operation of the WDN and the mathematical model. In this work we consider that the causes of these differences are the unknown distribution of demand and sensor noises. The proposed method is based on the computation of a set of residuals when the WDN is operating in the absence of leaks. Then, by means a linear Taylor approximation the possible sets of residuals considering the different leak scenarios are computed by adding the nominal residual generated with nominal demand distribution to the set of residuals of the non-leak scenario. Therefore, the extrapolation of uncertainty measurements in non-leak scenario to the different leaky scenarios spares a great amount of simulation scenarios.

On the other hand, confusion coefficients and isolability index are defined to assess the effect of uncertainty in the performance of leak localization methods based on residual correlation analysis. Confusion coefficients $c_{i,j}$ give an idea about the probability of classifying leak i as leak j and isolability index δ gives an idea about the probability of providing an exact leak localization in one time step.

The proposed approximation is applied to a simplified model of the WDN of Hanoi (Vietnam). By means of simulation, the results obtained with the proposed approximated sets are compared with the ones obtained with the real sets obtained by means of simulation. Different leak sizes are considered and despite the error in terms of correlation increases with the leak size the effect of this error in the isolability indicator error decreases with the leak magnitude and it is always small enough to consider the approximation a suitable approximation.

Finally, as a future work we plan to apply the proposed methodology to a real WDN considering other causes of uncertainty in boundary conditions and in leak magnitude.

REFERENCES

Blesa, J., Puig, V., and Saludes, J. (2012). Robust identification and fault diagnosis based on uncertain multiple input-multiple output linear parameter varying parity equations and zonotopes. *Journal of Process Control*, 22(10), 1890–1912.

Casillas, M.V., Puig, V., Garza-Castañón, L.E., and Rosich, A. (2013). Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors*, 13(11), 14984–15005.

Colombo, A.F., Lee, P., and Karney, B.W. (2009). A selective literature review of transient-based leak detection methods. *Journal of Hydro-environment Research*, 2(4), 212–227. doi:10.1016/j.jher.2009.02.003.

Cugueró-Escofet, P., Blesa, J., Pérez, R., Cugueró-Escofet, M.À., and Sanz, G. (2015). Assessment of a leak localization algorithm in water networks under demand uncertainty. *IFAC-PapersOnLine*, 48(21), 226 – 231.

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

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*. CRC Press.

Goulet, J.A., Coutu, S., and Smith, I.F. (2013). Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*, 27(2), 261–269.

Pérez, R., Cugueró-Escofet, J., Blesa, J., Cugueró-Escofet, M.A., and Sanz, G. (2016). Uncertainty effect on leak localisation in a dma. In *2016 3rd Conference on Control and Fault-Tolerant Systems (SysTol)*, 319–324.

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

Pérez, R., Sanz, G., Cugueró, M.À., Blesa, J., and Cugueró, J. (2015). Parameter Uncertainty Modelling in Water Distribution Network Models. *Procedia Engineering*, 119, 583–592.

Perez, R., Sanz, G., Quevedo, J., Nejjari, F., Meseguer, J., Cembrano, G., Tur, J.M.M., and Sarrate, R. (2014). Leak Localization in Water Networks. *IEEE CONTROL SYSTEMS MAGAZINE*, (august), 24–36.

Quevedo, J., Cugueró, M., Pérez, R., Nejjari, F., Puig, V., and Mirats, J. (2011). Leakage Location in Water Distribution Networks based on Correlation Measurement of Pressure Sensors. In *8th IWA Symposium on Systems Analysis and Integrated Assessment*, 290–297. IWA, San Sebastián.

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*, 04015020.

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.

van Thienen, P., Pieterse-Quirijns, I., Vreeburg, J., Vangeel, K., and Kapelan, Z. (2013). Applications of discriminative flow pattern analysis using the cfpd method. *Water Science and Technology: Water Supply*, 13(4), 906–913.

Wu, Z. and Sage, P. (2006). Water Loss Detection Via Genetic Algorithm Optimization-Based Model Calibration. In *ASCE 8th International Symposium on Water Distribution System Analysis*. Cincinnati.