Journal of Process Control 55 (2017) 1–9

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Leak localization in water distribution networks using Bayesian classifiers

Adrià Soldevila [a,*], Rosa M. Fernandez-Canti [a], Joaquim Blesa [b], Sebastian Tornil-Sin [a,b], Vicenç Puig [a,b]

[a] Research Center for Supervision, Safety and Automatic Control (CS2AC), Rambla Sant Nebridi, s/n, 08022 Terrassa, Spain
[b] Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain

ARTICLE INFO

Article history:
Received 5 August 2016
Received in revised form 24 March 2017
Accepted 28 March 2017
Available online 7 April 2017

Keywords:
Fault diagnosis
Bayesian classifier
Water distribution networks
Leak localization

ABSTRACT

This paper presents a method for leak localization in water distribution networks (WDNs) based on Bayesian classifiers. Probability density functions for pressure residuals are calibrated off-line for all the possible leak scenarios by using a hydraulic simulator, and considering the leak size uncertainty, demand uncertainty and sensor noise. A Bayesian classifier is applied on-line to the computed residuals to determine the location of leaks in the WDN. A time horizon based reasoning combined with the Bayesian classifier is also proposed to improve the localization accuracy. Two case studies based on the Hanoi and the Nova Icària networks are used to illustrate the performance of the proposed approach. Simulation results are presented for the Hanoi case study, whereas results for a real leak scenario are shown for the Nova Icària case study.

1. Introduction

Water leaks are present to some extent in all water distribution networks (WDNs). They may imply important economic costs because of the amount of water loss, and the location and reparation efforts involved. In many WDNs, losses due to leaks are estimated to account up to 30% of the total amount of extracted water [1]. This is a very significant amount since water is a precious resource in many parts of world that try to satisfy water demands of a growing population.

Several works have been published dealing with leak detection and isolation (localization) methods for WDNs (see [2] and references therein). For example, in [3], a review of transient-based leak detection methods is offered as a summary of current and past work. In [4], a method is proposed to identify leaks using blind spots based on previously leak detection that uses the analysis of acoustic and vibration signals [5], and models of buried pipelines to predict wave velocities [6]. More recently, Mashford et al. [7] have developed a method to locate leaks using support vector machines (SVM) that analyzes data obtained by a set of pressure control sensors of a pipeline network to locate and calculate the size of the leak. The use of *k*-NN and neuro-fuzzy classifiers in leak localization has been recently proposed in [8,9]. Another set of methods is based on the inverse transient analysis [10,11]. The main idea of this methodology is to analyze the pressure data collected during the occurrence of transitory events by means of the minimization of the difference between the observed and the calculated parameters. In [12,13], it is shown that unsteady-state tests can be used for pipe diagnosis and leak detection. The transient-test based methodologies use the equations for transient flow in pressurized pipes in frequency domain and then information about pressure waves is taken into account too.

Model-based leak detection and isolation techniques using stationary models have also been studied, starting with the seminal paper of Pudar and Liggett [14] which formulates the leak detection and localization problem as a least-squares parameter estimation problem. However, the parameter estimation of water network models is not an easy task [15]. The difficulty relies on the nonlinear nature of water network model and the few measurements usually available with respect to the large number of parameters to be estimated that leads to an underdetermined problem. Alternatively, in [16,17], a model-based method that relies on pressure measurements and leak sensitivity analysis is proposed. In this methodology, pressure residuals, i.e. differences between pressure measurements provided by sensors and the associated estimations obtained by using the hydraulic network model, are computed online and compared against associated thresholds that take into account the effects of the modeling uncertainty and the noise.

* Corresponding author.
E-mail address: adria.soldevila@upc.edu (A. Soldevila).

http://dx.doi.org/10.1016/j.jprocont.2017.03.015
0959-1524/

When some of the residuals exceed their thresholds, the residuals are matched against the leak sensitivity matrix in order to discover which of the possible leaks is present. Although this approach has good efficiency under ideal conditions, its performance decreases due to the nodal demand uncertainty and noise in the measurements. This methodology has been improved in [18] where an analysis along a time horizon is taken into account and a comparison of several leak isolation methods is presented. It must be noticed that in cases where the flow measurements are available, leaks can be detected more easily since it is possible to establish simple mass balance relations in the pipes. See for example the work of [19] where a methodology to isolate leaks is proposed using fuzzy analysis of the residuals. This method finds the residuals between the flow measurements and their estimation using a model without leaks. However, although the use of flow measurements is feasible in large water transport networks, this is not the case in water distribution networks where there is a dense mesh of pipes with only flow measurements at the entrance of each district metering area (DMA). In this situation, water companies consider as a feasible approach the possibility of installing only a few pressure sensors inside the DMAs due to budget constraints [20], because they are cheaper and easier to install and maintain.

In this paper, a new method for leak localization in WDNs that uses Bayesian classifiers is proposed. The use of Bayes theory has already been proposed in the context of leak localization but not using Bayesian classifiers as proposed in this paper (see [21,22]). Bayes theory has also been proposed as a tool for diagnosis in [23,24]. Here, probability density functions of pressure residuals are calibrated off-line for all the possible leak scenarios using a hydraulic simulator where leak size uncertainty, demand uncertainty and sensor noise are considered. A Bayesian classifier is applied on-line to the available residuals to determine the location of leaks present in the WDN. A time horizon method combined with the Bayesian classifier is also proposed to improve the accuracy of the leak localization method. Two case studies based on the Hanoi and the Nova Icària networks are used to illustrate the performance of the proposed approach. Simulation results are presented for the Hanoi case study, whereas results for a real leak scenario are shown for the Nova Icària case study.

The remainder of the paper is organized as follows. Section 2 presents an overview of the overall leak localization methodology. Section 3 presents the details of the proposed Bayesian classifier approach. Section 4 shows the application of the method to the two considered WDNs. Finally, Section 5 draws the main conclusions of the work.

2. Methodology overview

The detection and localization of leaks in WDNs is a particular case of the problem of fault detection and isolation (FDI) in dynamic systems. The classical model-based FDI approach considers that this problem can be (on-line) solved by generating and evaluating residuals, i.e. differences between values computed through a system model and values provided by sensors that are indicative of faults.

Following the approach proposed in previous works [16,17], this paper proposes a method for leak localization based on the generation and analysis of pressure residuals. In contrast to these cited works, the analysis of the residuals is not limited to a Boolean or directional analysis and the use of a Bayesian classifier is proposed instead, being this the main contribution of the paper.

It is assumed that a small number of pressure sensors is installed in inner nodes of the network and that a hydraulic model has been built and tuned. On the other hand, it is considered that leaks only appear in the nodes of the network. Although this is not true in practice (leaks can also appear in pipes), the approximation is valid and useful for large and dense networks.

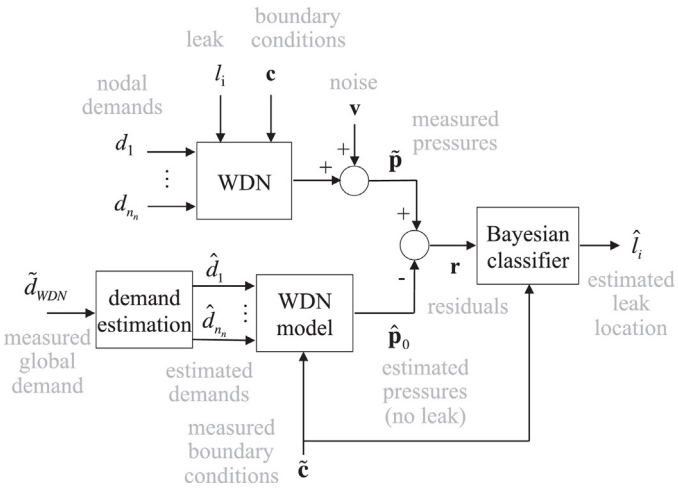

Fig. 1. Leak localization scheme.

Finally, it must be highlighted that the proposed method only addresses the leak localization problem, not the leak detection one, which is assumed to be solved by using any of the available techniques (for instance, leak detection can be based on detecting changes in the night consumption, which is the standard procedure used by most of water utilities [1] to monitor DMAs). Thus, the application of the proposed methodology assumes the existence of a leak detection module that triggers the operation of the leak localization module.

2.1. Basic architecture and operation

The method for on-line leak localization proposed in this paper relies on the scheme depicted in Fig. 1, based on computing pressure residuals and analyzing them by a Bayesian classifier.

Residuals ($\mathbf{r}$) are computed as differences between measurements provided by pressure sensors installed inside the DMA ($\tilde{\mathbf{p}}$) and estimations provided by a hydraulic model simulated under leak-free conditions ($\hat{\mathbf{p}}_0$). The hydraulic model is built using the Epanet hydraulic simulator [25] by considering the DMA structure (pipes, nodes and valves) and network parameters (pipe coefficients). It is assumed that, after the corresponding calibration process using real data, the model represents with accuracy the WDN behavior. However, it must be noticed that the model is fed with estimated water demands (typically obtained by the total measured DMA demand $\tilde{d}_{\mathrm{WDN}}$ and distributed at nodal level according to historical consumption records) in the nodes $(\hat{d}_1, \ldots, \hat{d}_N)$ since in practice nodal demands $(d_1, \ldots, d_N)$ are not measured (except for some particular consumers where automatic metering readers (AMRs) are available). Hence, the residuals are not only sensitive to faults but also to differences between the real demands and their estimated values. Additionally, pressure measurements are subject to the effect of sensor noise ($\mathbf{v}$) and this also affects the residuals. Taking all of these effects into account, the Bayesian classifier must be able to locate the real leak present in the WDN, that can be in any node and with any (unknown) magnitude, while being robust to the demand uncertainty and the measurement noise. Finally, it must be noted that the operation of the network is constrained by some boundary conditions ($\mathbf{c}$; for instance the position of internal valves and reservoir pressures and flows) that are measured ($\tilde{\mathbf{c}}$). These boundary conditions are taken into account in the simulation and can also be used as inputs for the classifier.

2.2. Design

As any other FDI method, the one proposed in this paper operates on-line but it requires a previous off-line design. In standard model-based FDI approaches, the design typically relies on the manipulation of the model (combination and/or projection of equations) with the aim of obtaining enhanced residuals, i.e. residuals that can facilitate isolation under a Boolean (structured residuals) or geometric (directional residuals) framework [26]. Unfortunately, this type of procedure cannot be applied to the localization of leaks in WDNs since the model is a set of non-linear implicit equations that cannot be easily (algebraically) manipulated. Alternatively, what can be done is to generate synthetic data in the residual space by performing extensive simulations of the model under the possible faulty conditions and to analyze the obtained raw residuals. Since the number of installed pressure sensors is small compared to the number of nodes (which are the places where it is considered that the leak can occur) and there are different sources of uncertainty (modeling errors, errors in estimation of consumer demands, measurement noise), the raw residuals do not facilitate a straightforward leak localization. However, their analysis is still possible by using classifiers. In a previous work [8], this was done by training and using a classical k-NN classifier. In this work, a Bayesian classifier is proposed.

The method proposed in this paper considers an off-line design based on the following stages:

- Modeling – A model for the WDN is obtained, calibrated and implemented in Epanet. The model is basically built by taking into account the network structure and by applying flow balance conservation and pressure loss equations, see [16,17] for details. This type of model can be usually provided by the company that manages the water utility.
- Data generation – The model implemented in Epanet is extensively used to generate data in the residual space for each possible leak location. The residuals correspond to the differences between the pressures obtained by considering a leak scenario and the ones obtained in the leak-free scenario.
- Calibration – The data associated to each leak location is used to calibrate a joint probability density function that models the effect of each leak in the residual space. The calibration procedure is detailed in Section 3.2 of the paper.

The data generation stage is critical, since the generated data have to be complete enough to allow the computation of representative probability density functions and the maximum possible degree of isolability. Data is generated by applying the scheme presented in Fig. 2, that is similar to the one presented in Fig. 1 but substituting the real WDN by its Epanet model.

The presented scheme is exploited in order to:

- Generate data for all possible leak locations, i.e. for all the different nodes in the WDN ($\bar{l}_i, \quad i = 1, 2, \ldots, N$).
- For each possible leak location, generate data for different leak magnitudes inside a given range ($\bar{l}_i \in [l_i^-, l_i^+]$).
- Generate sequences of demands and boundary conditions that correspond to realistic typical daily evolution in each node.
- Simulate differences between the real demands and the estimations computed by the demand estimation module ($(\bar{d}_1, \ldots, \bar{d}_N) \neq (\hat{d}_1, \ldots, \hat{d}_N)$).
- Take into account the measurement noise in pressure sensors, by generating synthetic normal distributed noise ($\bar{\nu}$).

Considering all these uncertainties, the residual space becomes overlapped between classes (different leak locations), as it is in practice. To deal with the fact that the residuals obtained for the diagnosis can fall in the residual space of more than one class, the use of the Bayes rule (in form of Bayesian classifier) is proposed. Bayesian reasoning allows to assign different probabilities to different classes and therefore it achieves better results than other approaches that cannot give information about how related is the instance to classify to each class. In addition it allows to fuse in time different but consecutive diagnosis to improve accuracy.

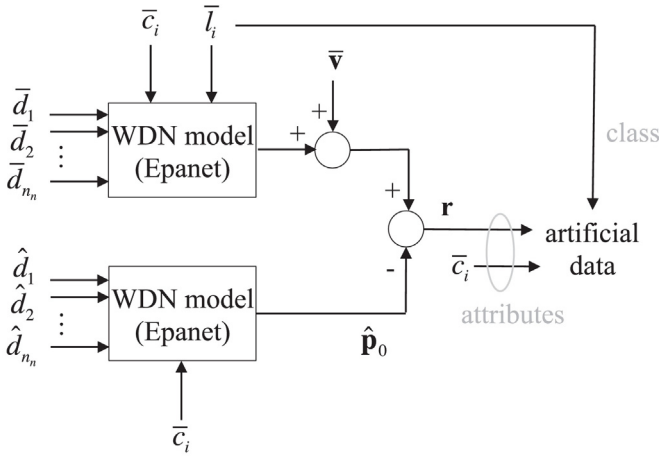

Fig. 2. Data generation scheme.

3. Bayesian leak localization

3.1. Bayesian classification

For a network with N nodes, the potential leak locations are l_i, $i = 1, \ldots, N$. It is assumed that pressure is measured in M internal nodes through sensors that provide the values $\tilde{p}_j$, $j = 1, \ldots, M$. It is also assumed that a model of the system behavior allows the computation, at each time instant k, of M residual signals r_j, $j = 1, \ldots, M$. These residuals are defined as the instantaneous mismatch between the measurements provided by sensors and the behavior predicted by the model. When a leak occurs, all the residuals are activated (diverging from zero) up to some extent. Given the residuals, the objective is to apply a Bayesian leak discriminating procedure in order to identify which leak is the most likely to be behind the observed behavior. Such a diagnosis procedure based on Bayesian reasoning is explained below.

At every time sample k, the probability of each leak can be estimated as a result of the application of the Bayes Theorem

$$P(l_i \mid \mathbf{r}(k)) = \frac{P(\mathbf{r}(k) \mid l_i) P(l_i)}{P(\mathbf{r}(k))}, \quad i = 1, \ldots, N \tag{1}$$

where $P(l_i \mid \mathbf{r}(k))$ is the posterior probability that the leak l_i had caused the observed residual vector $\mathbf{r}(k) = (r_1(k) \cdots r_j(k))^T$, $P(\mathbf{r}(k) \mid l_i)$ is the likelihood of the residual $\mathbf{r}(k)$ assuming that the active leak is l_i, $P(l_i)$ is the prior probability for the leak l_i, and $P(\mathbf{r}(k))$ is a normalizing factor given by the total probability law,

$$P(\mathbf{r}(k)) = \sum_{i=1}^{N} P(\mathbf{r}(k) \mid l_i) P(l_i) \tag{2}$$

Regarding the prior probabilities, unless we have any additional information, an unprejudiced starting point is to consider all the leak locations equally probable, that is, $P(l_i) = (1/N)$, $i = 1, \ldots, N$. To estimate the likelihood value $P(\mathbf{r}(k) \mid l_i)$, we need to perform a previous calibration task in order to obtain the joint probability density function for each leak in the residual space, $P(\mathbf{r} \mid l_i)$, $i = 1, \ldots, N$. The calibration stage is detailed in the next section. Note that, in

contrast to standard naïve Bayesian classifiers, we do not assume independence between the residuals.

The application of (1) produces a set of values $P(l_i \mid \mathbf{r}(k))$, $\sum_{i=1}^{N} P(l_i \mid \mathbf{r}(k)) = 1$, that can be used to decide which leak is acting over the system. Note that, at each time sample k, we have information of which is the probability associated to each leak situation. There can be many competing leaks, each one with a different probability value. The leak with the highest posterior probability can be selected as the most likely leak, or all the leaks with a posterior probability above a prespecified threshold can be selected as leak candidates.

3.1.1. Recursivity

The results can be improved if (1) is recursively applied, that is, if the posterior $P(l_i \mid \mathbf{r}(k))$ is used as the prior probability for the next time sample. This way, as long as new measurement data are available, the probabilities are updated and many of the competing leaks can be discarded.

The only drawback is that if any of the leaks takes the posterior probability value of 1 at any k, then all the remaining leaks take the 0 probability value, therefore preventing them to have a future value different from zero due to the recursive application of (1). This drawback can be easily overcome by forcing all probabilities to present a maximum value of, say, 0.99. When a leak l_i presents the probability $P(l_i \mid \mathbf{r}(k)) > 0.99$, we force it to be $P(l_i \mid \mathbf{r}(k)) = 0.99$ and we can force the remaining leaks to be $P(l_n \mid \mathbf{r}(k)) = ((1 - 0.99)/(N - 1))$, $n = 1, \ldots, N$, $n \neq i$.

3.1.2. Time horizon

Finally, the result can additionally be improved if the diagnosis is performed by applying the recursivity approach described above in a time horizon H. In this case, the posterior probability can be computed on the basis of the H previous time samples, that is, to compute $P(l_i \mid \mathbf{r}(k))$, we recursively can apply the following equation

$$P(l_i \mid \mathbf{r}(k-H+n)) = \frac{P(\mathbf{r}(k-H+n) \mid l_i) P(l_i \mid \mathbf{r}(k-H+n-1))}{P(\mathbf{r}(k-H+n))},$$
$$i = 1, \ldots, N, \quad n = 1, \ldots, H \tag{3}$$

where an unprejudiced starting point may be $P(l_i \mid \mathbf{r}(k-H)) = (1/N)$, $i = 1, \ldots, N$.

3.2. Calibration of the probability density functions

In order to apply the methodology presented in the previous subsection, it is needed to perform a previous off-line calibration task consisting in obtaining the joint probability density function for each leak class in the residual space, $P(\mathbf{r} \mid l_i)$, $i = 1, \ldots, N$. With this aim, we generate a set of residual samples for each leak class and we obtain the probability density function that best fits to them.

3.2.1. Residuals probability density function

The first step is to decide the distribution family. The Law of Large Numbers states that most situations lead to a Gaussian probability density function if the number of samples is high enough. Several tests can be applied to the residual samples in order to assess whether the residual samples are Gaussian distributed or not. For instance, we can apply the well-known Kolmogorov–Smirnov [27] or the Anderson–Darling [28] tests, among others.

Fig. 3 shows an illustrative case where two-dimensional residual data for two different leaks are fitted by means of Gaussian probability density functions. Data for leak 1 is better adjusted because the cross-correlation between residuals r_1 and r_2 is taken into account. On the other hand, data for leak 2 is adjusted by assuming statistic independence between residuals r_1 and r_2 and therefore the fitting is not so accurate. Note also that other probability distribution families different from Gaussian could be used, including multimodal and non-parametric distributions.

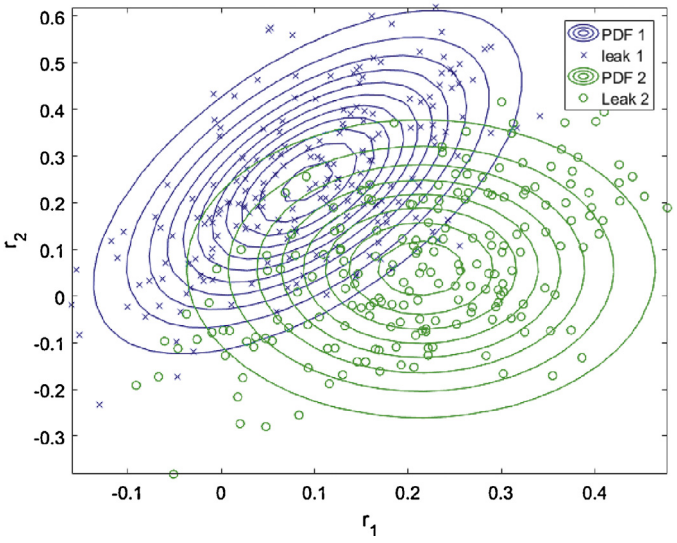

Fig. 3. Calibration for leaks 1 and 2.

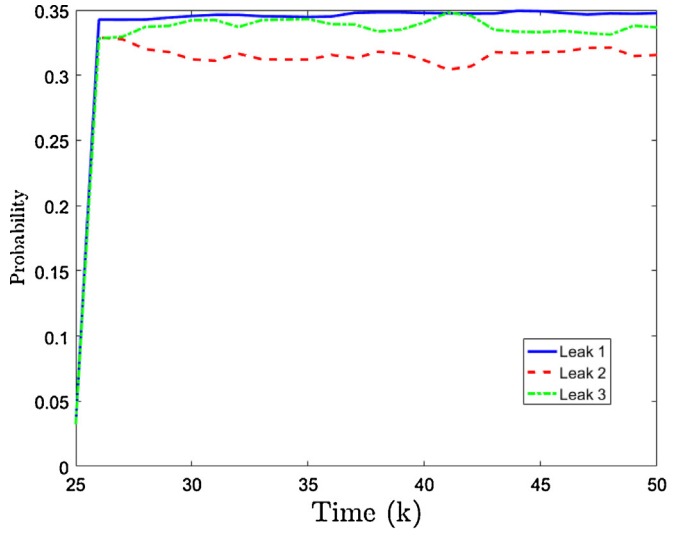

Fig. 4. Probability for the leaks 1, 2 and 3.

3.2.2. Identification of overlapped nodes

In practical situations, leaks in different nodes can provide very similar residual realizations that lead to very close, overlapped or even indistinguishable probability density functions in the residual space. This is mainly due to the sensor location and performance. It must be noticed that a really limited number of installed sensors is frequent and that these sensors are not located in the optimal places.

In these cases, the diagnosis procedure is more difficult and ends up with a result that indicates that the leak is associated to a certain group of nodes whereas the particular node is not completely isolated. See for instance the illustrative Fig. 4. In a network with 31 nodes, a leak in the node 1 has taken place at $k = 25$. The figure shows how the method has triggered not only the probability of node 1, but also the probabilities of nodes 2 and 3, assigning a value around 1/3 to each of them three. This is due to the fact that all three nodes are located in a peripherial branch of the network,

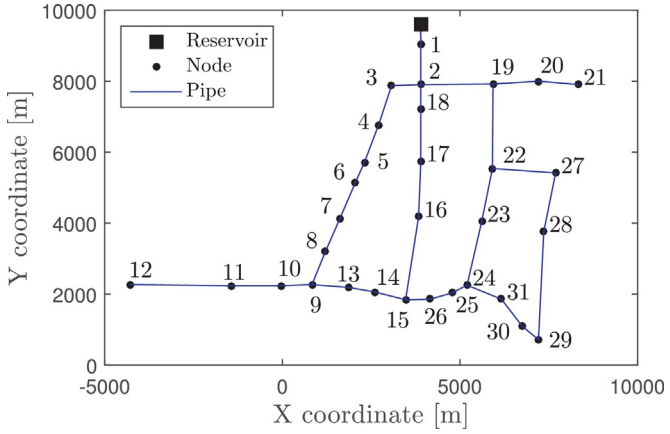

Fig. 5. Hanoi topological network.

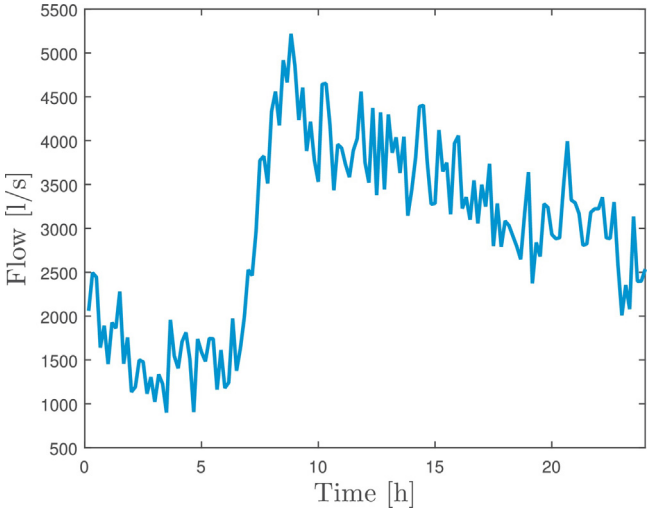

Fig. 6. Flow pattern.

far away the nearest sensor. The method has properly detected the area where the leak is but it cannot clearly distinguish which is the leaking node.

However, given the residual samples, it is possible to identify a priori which nodes will be "overlapped". For instance, we can perform a minimal statistical energy test to the residual samples. This is a non-parametric test which compares multidimensional data from two samples using a measure based on statistical energy [29].

If the test statistic φ is negative, the two samples are closely distributed, being $\varphi=-\infty$ the case where they come from exactly the same distribution. Different values of φ can be associated to different degrees of similarity. Interestingly, the posterior probabilities obtained by the Bayesian method proposed here identify the same groups of nodes, confirming thus the validity of the statistical energy test.

In order to improve the distance between the probability density functions of the overlapped nodes, a filtering of the residual samples can be performed before applying the proposed Bayesian diagnosis procedure. For instance, the mean value of all the hourly measurements can be obtained (see the case studies section).

4. Case studies

In order to assess the performance of the proposed methodology, two case studies based on DMAs of increasing size and complexity, Hanoi and Nova Icària, are presented in this section.

4.1. Hanoi case study

The proposed methodology has been first applied to the simplified model of the Hanoi (Vietnam) network, depicted in Fig. 5. This model consists of one reservoir, 34 pipes and 31 nodes. Measurements of two inner pressure sensors placed in nodes 14 and 30 are available as considered in other works [8].

In order to analyze and illustrate the performance of the proposed methodology, four different studies have been carried out. In each one of the first three studies, the individual effect of a given source of uncertainty is considered. In the last study, all three sources of uncertainty are considered together. The particular conditions for each study are the following:

- Study 1 (Leak magnitude uncertainty): In this study, it is assumed that the leak magnitudes can vary inside a given range and that their precise values are unknown. Hence, for each node, leaks in the range between [25,75] [l/s] (0.84% and 2.51% of the average value of total amount of water demanded, which is 2991.1 [l/s]) are simulated. On the other hand, it is assumed that the nodal demands are perfectly known and that there are not measurement noises.
- Study 2 (Noise in pressure measurements): It is assumed that pressure measurements are corrupted by a uniform zero-mean noise with an amplitude of ±5% of the mean value of all pressure residuals. It is assumed that the leak magnitude is fixed to a nominal value and that the nodal demands are known.
- Study 3 (Demand uncertainty): It is considered that the nodal demands are not perfectly known. In particular, an uncertainty of ±10% of the nominal nodal demand at each node is considered. It is assumed that the leak magnitude is fixed to a nominal value and that there is no noise in the measurements.
- Study 4 (Realistic case): All three uncertainties previously defined are considered to be simultaneously acting on the DMA.

For each study, two complete data sets have been generated for each node (potential leak location), one for estimating the associated probability density function and the other one for testing the leak localization performance. Each set used for testing, associated to a leak in a given node, is called a leak scenario. The variables conforming the data are the input flow $\tilde{d}_{\mathrm{DMA}}$ and the two residuals r_1 and r_2 associated to pressure measurements in nodes 14 and 30, respectively. The sampling time used to simulate the network is 10 min, but hourly average values of variables are used as the result of filtering sensor data in order to improve the leak localization performance. Different daily inner flow patterns have been simulated such as the one depicted in Fig. 6. The distribution of the global pattern demand in all demand nodes has been considered fixed and known although uncertainty is added for the Studies 3 and 4 as it is proposed in [30]. Pressure residuals are obtained by means of a WDN simulator (Epanet), following the scheme presented in Fig. 2. For example, Fig. 7 shows the residuals obtained for the Study 4.

4.1.1. Calibration

In order to determine whether the three classifier inputs (r_1, r_2 and $\tilde{d}_{\mathrm{DMA}}$) can be fit into a Gaussian distribution, a one-dimension Kolmogorov–Smirnov test on a training data set of 480 samples (for each of the 31 leak nodes) has been performed. As a result, the three inputs can be considered Gaussian distributed for a significance level of 3%.

Once the 31 Gaussian probability density functions have been calibrated, a minimal statistical energy test has been applied to these Gaussian probability density functions in order to

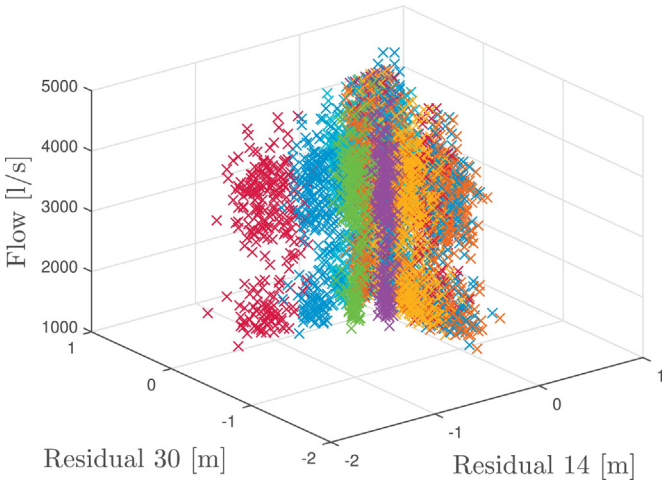

Fig. 7. Residual space for Study 4 (all sources of uncertainty considered).

determine the distance between them and consequently detect possible overlapped nodes. In the considered case study, the statistic φ takes the value $-\infty$ for the combinations of nodes 9-10-11-12 and 20-21-22. This indicates that the nodes 9, 10, 11 and 12 present exactly the same probability distribution and thus they cannot be distinguished in the subsequent diagnosis procedure. And the same conclusion applies for the group 20-21-22. The existence of these two undistinguishable groups is due to the topology of the network and the sensor geographical distribution. Additionally, values of φ different from $-\infty$ could be used to a priori estimate which nodes, even though not belonging to an undistinguishable group, would be more difficult to separate. For instance, and for the available measurement data, nodes 5 and 6 would be the more difficult to isolate since they produce the largest φ value different from $-\infty$ ($\varphi_{5,6} = -0.02205$); hereafter the group 5, 6, 7 and 8 would be difficult to separate between them ($\varphi_{6,7} = -0.02119$, $\varphi_{7,8} = -0.02142$,) and from the 9-10-11-12 group ($\varphi_{8,9} = \varphi_{8,10} = \varphi_{8,11} = \varphi_{8,12} = -0.02163$,. . .). The different degrees of overlapping detected are, again, explained by the network topology and the sensor geographical distribution.

4.1.2. Results

A set of 1000 samples for every leak scenario has been used for testing purposes. The results obtained by the proposed method considering different time horizons, $H = 1, \ldots, 24$ h, where hourly values are obtained as the average values of the last six sensor measurements (obtained at a 10 min sampling rate), have been compared with the ones obtained using the angle method proposed in [18] and with the k-NN method presented in [31]. For the angle method, a different sensitivity matrix for each hour of the day has been considered. For the k-NN method, the value used for k is three, the Euclidean distance is used as metric, and the range of the third attribute (flow) is reduced to a similar scale of the other two attributes by dividing all the values by the maximum flow value divided by three.

The results for each considered study, working with $H = 1$ and $H = 24$, obtained by using each of the three methods are summarized in Table 1. The values in the table indicate the % of well classified leak locations (the highest posterior probability corresponds to the correct hypothesis) defined as "Accuracy". On the other hand, the evolution of the performance with the time horizon H considering all the uncertainties (Study 4) is depicted in Figs. 8 and 9. In the first of these figures, the performance is presented in terms of the already defined accuracy, while in the second figure the performance indicator "Average Topological Distance", which is defined as the average value of the minimum distance in nodes between the node with highest posterior probability and the node with the real leak, is shown.

Table 1
Accuracy results in Hanoi DMA network.

Study	$H = 1$			$H = 24$		
	Bayes	k-NN	Angle	Bayes	k-NN	Angle
Leak uncertainty	99.87	92.57	85.82	100	96.26	93.81
Noise in measurements	98.47	93.86	84.29	99.88	99.39	89.95
Demand uncertainty	68.42	63.41	61.23	92.92	81.24	78.92
All together	63.35	55.78	56.94	91.59	81.47	76.13

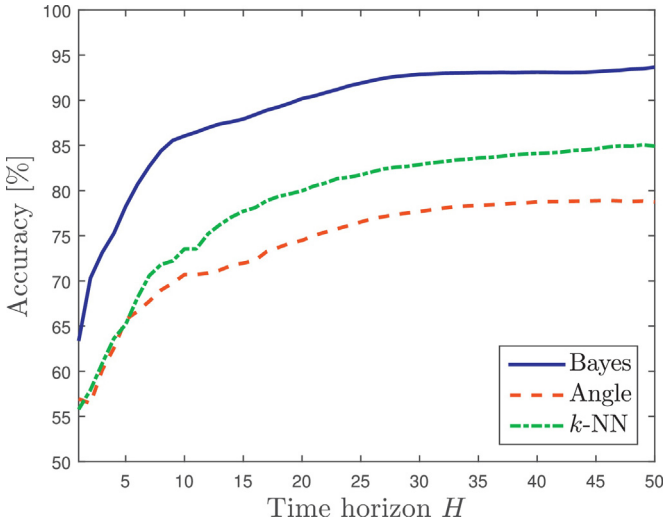

Fig. 8. Accuracy results in a time horizon in Hanoi network.

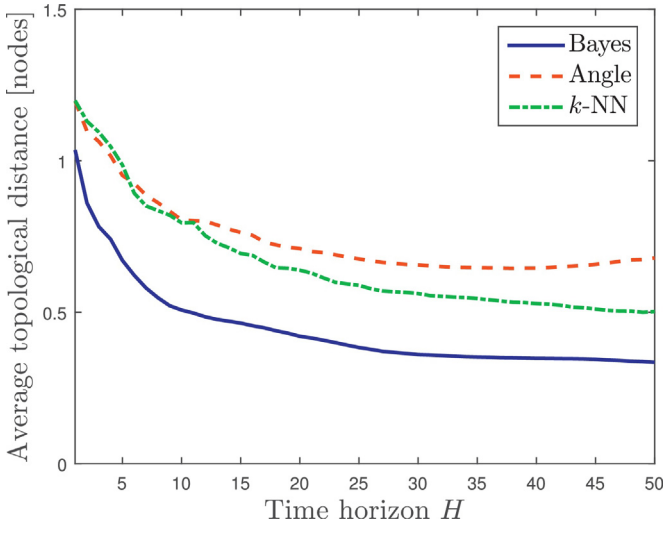

Fig. 9. Average topological distance results in a time horizon in Hanoi network.

As it can be concluded from the presented results, the performances provided by the proposed method are better in all the studied uncertainty scenarios than the ones provided by the angle and k-NN methods. On the other hand, as it was expected, the performance of all methods improves with the increase of available data (increase of time horizon H). Additionally, Figs. 10 and 11 show the evolution with the time horizon H of the leak location (posterior probabilities) provided by the proposed method in two leak scenarios. In the first scenario (leak in node 15) the leak is correctly located in one time instant ($H = 1$). However, in the second leak scenario (leak in node 8) there is a competition between the

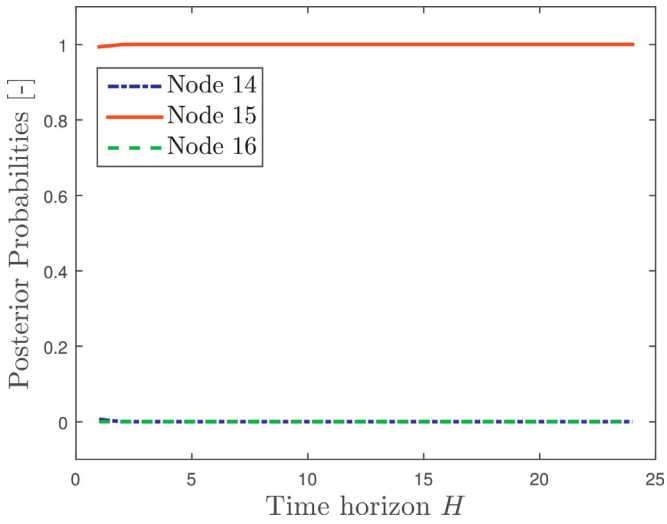

Fig. 10. Posterior probabilities evolution of hypothesis 14, 15 and 16 in node 15 leak scenario.

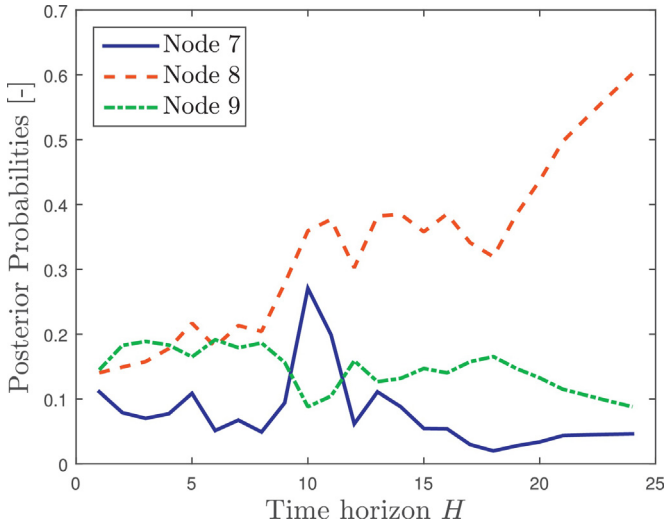

Fig. 11. Posterior probabilities evolution of hypothesis 7, 8 and 9 in node 8 leak scenario.

correct node and some slightly overlapped nodes according to the minimal statistical energy test described in Section 4.1.1. Therefore, some samples are required to provide a correct leak localization in this scenario.

4.2. Nova Icària case study

The Nova Icària network, shown in Fig. 12, is one of the DMAs of the Barcelona WDN. This network consists of 1520 nodes, 1664 pipes, two reservoirs and two pressure reducing valves (PRVs), each one located after the reservoirs with the aim of maintaining a certain pressure control level. Measurements of five pressure sensors (the pressure transducers used are the IMP-S-004-010S model[1] with a resolution of 0.1 m_{H_2O}) installed in nodes 3, 4, 5, 6 and 7 (highlighted in Fig. 12), measurement of the flow of the network entering the DMA and the set points for the PRVs are available every

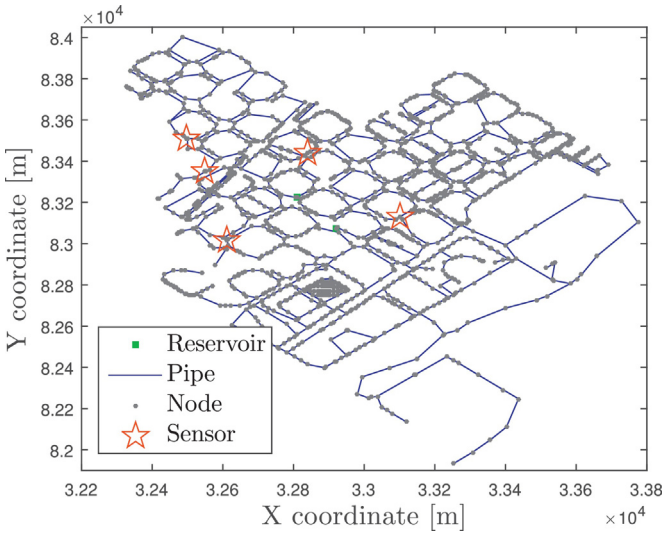

Fig. 12. Nova Icària DMA topological network.

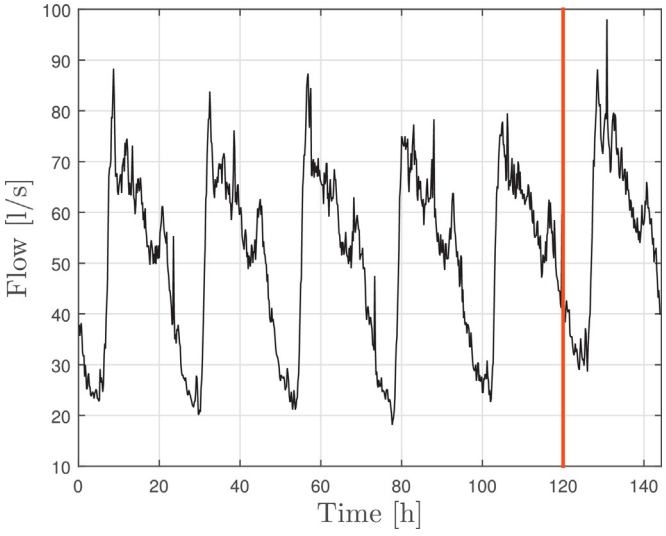

Fig. 13. Nova Icària flow measurements under nominal conditions (before red line) and faulty conditions (after red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

10 minutes using the data logger MultiLogS GSM/SMS[2] device. As in the previous case study, for all the measured variables, the average value of the six samples available each hour is used for leak localization purposes. Single leak scenarios have been considered in the 1520 nodes.

For this case study, real data collected both under normal operation conditions and under the presence of a real leak and provided by the water company have been used. The leak was created by the water company that operates the network by opening a fire hydrant. The experiment took place on December 20, 2012 at 00:30 h and lasted around 30 h with a leak size about 5.6 [l/s], being the total demand of water in the range between 23.5 and 78 [l/s] approximately. For more details see [17]. Moreover, real sensor data for the network in a normal operation scenario of five days before the leak scenario was also provided. The relevant data used to perform the leak localization is shown in different figures: Fig. 13 shows the DMA input flow; Fig. 14 show the pressure references for

[1] http://www.impress-sensors.co.uk/products/sensor-products/pressure-measurement/industrial-pressure-transducers-transmitters/standard-range-pressure-transmitter/imp-industrial-pressure-transmitter.html.

[2] http://hinco.com.au/shop/type/data-loggers/multilog/.

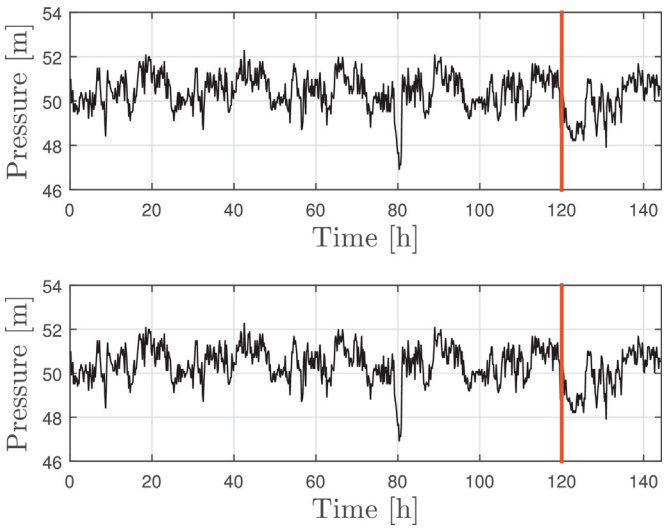

Fig. 14. Nova Icària PRVs set point values under nominal conditions (before red line) and faulty conditions (after red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

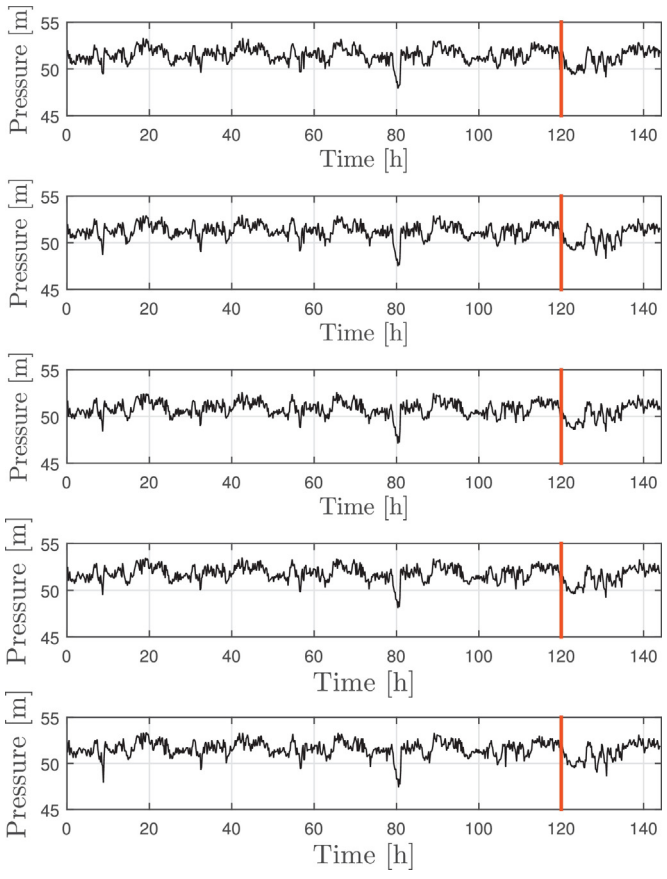

Fig. 15. Nova Icària pressure measurements under nominal conditions (before red line) and faulty conditions (after red line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the two input valves; and, finally, Fig. 15 shows the measurements provided by the five internal pressure sensors. In all these figures, the red line indicates the time instant where the leak is introduced. Finally, an accurate Epanet model of the Network and node demand estimations was provided as well.

The discrepancy between the leak-free real pressure measurements and the pressures provided by the hydraulic model have been used to extract the real uncertainty of the system. This discrepancy and the simulation of a nominal leak of 5.6 [l/s] have been used to calibrate the Gaussian probability density functions of the 1520 nodes.

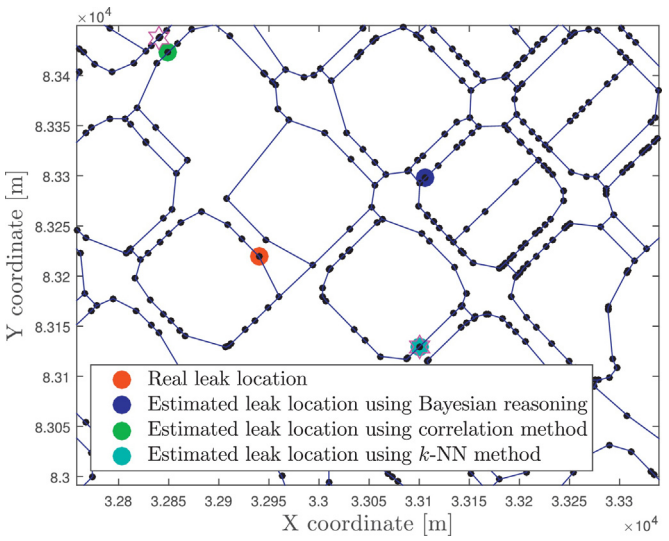

Fig. 16. Nova Icària leak location.

Once the probability density functions have been calibrated, the proposed method has been applied to the data of the real leak scenario. The obtained results are graphically shown in Fig. 16. The real leak was introduced in node 996 (red dot in Fig. 16) and the Bayesian classifier provided in one day time horizon (H = 24) the node 403 (blue dot) as the candidate with highest posterior probability. The topological distance between the two nodes is 10 (265.0 m of pipe) and the linear geometric distance is 183.2 meters. Additionally, for comparison purposes, the correlation method described in [17] and the k-NN method presented in [31] have been also applied to the defined leak scenario. For the correlation method, the selected candidate is node 1036 (green dot in Fig. 16), which is at a topological distance of 17 nodes (433.4 m of pipe) and at a linear geometric distance of 222.0 m from the real leaky node. For the k-NN method, node 3 is selected as node candidate, at a topological distance of 13 nodes (390.8 m of pipe) and a linear geometric distance of 184.0 m from the real leak location. For both cases, the obtained results are worse (according to the three computed distances) than the provided by the proposed method.

5. Conclusion

In this paper, a methodology for leak localization in water distribution networks has been proposed. The methodology relies on the computation of pressure residuals according to a network hydraulic model and their analysis by a Bayesian classifier. In a first off-line stage, the network model is simulated under different uncertainty conditions to obtain residual data for each potential leak location (each network node). Then this data is used to calibrate probability density functions. Additionally, a study of the possible overlapped probability density functions based on the minimal statistic energy test is also proposed. In the on-line stage, a Bayesian classifier provides the time-dependent posterior probability of every possible leak. The effect of the time horizon in the decision has also been studied.

The performance of the proposed method has been tested in the Hanoi network in different scenarios of uncertainty and in the Nova Icària DMA network in a real leak scenario. Moreover, the obtained performance has been compared with the performance provided

by two other state-of-the-art methods and an improvement has been observed.

As future work, a method for sensor placement adapted to the leak location methodology proposed in this paper is being developed.

Acknowledgment

This work has been funded by the Ministerio de Economía, Industria y Competitividad (MEINCOP) of the Spanish Government and FEDER through the projects ECOCIS (ref. DPI2013-48243-C2-1-R) and HARCRICS (ref. DPI2014-58104-R) and through the grant IJCI-2014-20801, by the European Commission through contract EFFINET (ref. FP7-ICT2011-8-31 8556) and by the Catalan Agency for Management of University and Research Grants (AGAUR), the European Social Fund (ESF) and the Secretary of University and Research of the Department of Companies and Knowledge of the Government of Catalonia through the grant FI-DGR 2015 (ref. 2015 FI_B 00591).

References

[1] R. Puust, Z. Kapelan, D.A. Savic, T. Koppel, A review of methods for leakage management in pipe networks, Urban Water J. 7 (1) (2010) 25–45.
[2] Z.Y. Wu, P. Sage, Water loss detection via genetic algorithm optimization-based model calibration, in: Systems Analysis Symposium, ASCE, 2006, pp. 1–11.
[3] P.L. Andrew, F Colombo, B.W. Karney, A selective literature review of transient-based leak detection methods, J. Hydro-environ. Res. (2009) 212–227.
[4] J. Yang, Y. Wen, P. Li, Leak location using blind system identification in water distribution pipeline, J. Sound Vibr. (310) (2008) 134–148.
[5] H. Fuchs, R. Riehle, Ten years of experience with leak detection by acoustic signal analysis, Appl. Acoust. (33) (1991) 1–19.
[6] J. Muggleton, M. Brennan, R. Pinnington, Wavenumber prediction of waves in buried pipes for water leak detection, J. Sound Vibr. (249) (2002) 939–954.
[7] J. Mashford, D. de Silva, D. Marney, S. Burn, An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine, in: Third International Conference on Network and System Security, 2009, pp. 534–539.
[8] L. Ferrandez-Gamot, P. Busson, J. Blesa, S. Tornil-Sin, V. Puig, E. Duviella, A. Soldevila, Leak localization in water distribution networks using pressure residuals and classifiers, IFAC-PapersOnLine 48 (21) (2015) 220–225.
[9] D. Wachla, P. Przystalka, W. Moczulski, A method of leakage location in water distribution networks using artificial neuro-fuzzy system, IFAC-PapersOnLine 48 (21) (2015) 1216–1223.
[10] D. Covas, H. Ramos, Hydraulic transients used for leak detection in water distribution systems, in: 4th International Conference on Water Pipeline Systems, BHR Group, 2001, pp. 227–241.
[11] A. Kepler, D. Covas, L. Reis, Leak detection by inverse transient analysis in an experimental *pvc* pipe system, J. Hydroinform. 13 (2) (2011) 153–166.
[12] M. Ferrante, B. Brunone, Pipe system diagnosis and leak detection by unsteady-state tests. 1. Harmonic analysis, Adv. Water Resour. 26 (1) (2003) 95–105.
[13] M. Ferrante, B. Brunone, Pipe system diagnosis and leak detection by unsteady-state tests. 2. Wavelet analysis, Adv. Water Resour. 26 (1) (2003) 107–116.
[14] R.S. Pudar, J.A. Liggett, Leaks in pipe networks, J. Hydraul. Eng. 118 (7) (1992) 1031–1046.
[15] D.A. Savic, Z. Kapelan, P. Jonkergouw, Quo vadis water distribution model calibration? Urban Water J. 6 (1) (2009) 3–22.
[16] R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, A. Peralta, Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks, Control Eng. Pract. 19 (10) (2011) 1157–1167.
[17] R. Pérez, G. Sanz, V. Puig, J. Quevedo, F. Nejjari, J. Meseguer, G. Cembrano, J. Mirats, R. Sarrate, Leak localization in water networks, IEEE Control Syst. Mag. (2014) 24–36.
[18] M.V. Casillas, L. Garza-Castañón, V. Puig, Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks, in: 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Elsevier, 2012, pp. 570–575.
[19] J. Ragot, D. Maquin, Fault measurement detection in an urban water supply network, J. Process Control 16 (2006) 887.
[20] R. Sarrate, J. Blesa, F. Nejjari, J. Quevedo, Sensor placement for leak detection and location in water distribution network, Water Sci. Technol. Water Supply 14 (5) (2014) 795–803.
[21] Z. Poulakis, D. Valougeorgis, C. Papadimitriou, Leakage detection in water pipe networks using a Bayesian probabilistic framework, Prob. Eng. Mech. 18 (4) (2003) 315–327.
[22] H. Zhang, L. Wang, Leak detection in water distribution systems using Bayesian theory and Fisher's law, Trans. Tianjin Univ. 17 (2011) 181–186.
[23] B. Huang, Bayesian methods for control loop monitoring and diagnosis, J. Process Control 18 (2008) 828–838.
[24] R. Gonzalez, F. Qi, B. Huang, Process Control System Fault Diagnosis: A Bayesian Approach, John Wiley and Sons, Ltd, Chichester, UK, 2016.
[25] L. Rossman, Epanet 2 User's Manual, United States Envionmental Protection Agency, 2000.
[26] J. Gertler, Fault Detection and Diagnosis in Engineering Systems, CRC Press, 1998.
[27] W. Daniel, Applied Nonparametric Statistics, PWS-Kent Boston, 1990.
[28] M. Stephens, EDF statistics for goodness of fit and some comparisons, J. Am. Stat. Assoc. 69 (347) (1974) 730–737.
[29] , in: Statistical energy as a tool for binning-free, multivariate goodness-of-fit tests, two-sample comparison and unfolding, Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip. 537 (3) (2005) 626–636.
[30] P. Cugueró-Escofet, J. Blesa, R. Pérez, M.À. Cugueró-Escofet, G. Sanz, in: Assessment of a leak localization algorithm in water networks under demand uncertainty, IFAC Proc. Vol. (IFAC-PapersOnline) 48 (21) (2015) 226–231.
[31] A. Soldevila, J. Blesa, S. Tornil-Sin, E. Duviella, R. Fernandez-Canti, V. Puig, in: Leak localization in water distribution networks using a mixed model-based/data-driven approach, Control Eng. Pract. 55 (2016) 162–173.