Leak Localization in Water Distribution Networks using Pressure Residuals and Classifiers

Lise Ferrandez-Gamot*, Pierre Busson*, Joaquim Blesa**, Sebastian Tornil-Sin**, Vicenç Puig**, Eric Duviella*, Adrià Soldevila**

*École des Mines de Douai, Bd. Lahure 764, Douai, 59500, France
**Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona
(e-mail :joaquim.blesa@upc.edu)

Abstract: In order to take into account the scarcity of the water resource and the increasing of the population, the management of drinking water networks has to be improved with the use of new tools and actions that allows fighting against wasting water. The monitoring of drinking water networks is based on the use of sensors to locate malfunctions (leaks, quality/contamination events, etc.). Practical implementation has to be carried out by optimizing the placement of the number of sensors and improving the detection and localization of malfunctions. Techniques for the detection and localization of leaks have been proposed in the last years based on the evaluation of residuals obtained by means of the comparison between the measurements obtained by the sensors and the values obtained by simulating the water network in a leak free scenario. In this paper, a data-driven approach based on the use of statistical classifiers working in the residual space is proposed for leak localization. The classifiers are trained using leak data scenarios in all the nodes of the network considering uncertainty in demand distribution, additive noise in sensors and leak magnitude. Finally, the proposed approach is tested using the well-known Hanoi network benchmark.

Keywords: Fault diagnosis, classifiers, water distribution networks, leak localization, pressure sensors

1. INTRODUCTION

Water leaks in water distribution networks (WDN) can cause significant economic losses in fluid transportation leading to increase reparation costs that finally generate an extra cost for the final consumer. In many WDN, losses due to leaks are estimated to account up to 30% of the total amount of extracted water. This is a very important amount in a world struggling to satisfy water demands of a growing population.

Several works have been published on leak detection and isolation (localization) methods for WDN (see e.g. Wu et al. (2011)), Puust et al. (2010) and references therein). Model based leak detection and isolation techniques have also been studied starting with the seminal paper of Pudar and Liggett (1992) which formulates the leak detection and localization problem as a least-squares estimation problem. However, the parameter estimation of water network models is not an easy task (Savic et al. (2009)). The difficulty lies in the non-linear 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 Pérez et al. (2011, 2014), a model based method that relies on pressure measurements and leak sensitivity analysis is proposed. This methodology consists in analyzing the residuals (difference between the measurements and their estimation using the hydraulic network model) on-line regarding a given threshold that takes into account the

modeling uncertainty and the noise. When some of the residuals violate their threshold, the residuals are compared 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 Casillas et al. (2012) where an analysis along a time horizon has been taken into account and a comparison of several leak isolation methods is offered. In case where the flow measurements are available, leaks could be detected more easily since it is possible to establish simple mass balance in the pipes. See for example the work of Ragot et al. (2006) where a methodology to isolate leaks is proposed using fuzzy analysis of the residuals. This method finds the residuals between the measurements with and without leaks. However, although the use of flow measurements is viable 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 solution the possibility to installing some pressure sensors inside the DMAs, because they are cheaper and easy to install and maintain.

In this paper, a new approach for leak localization in WDN is presented. This approach combines the use of models and classifiers. Models are used to generate residuals while

classifiers are used for analysing the residuals taking into account the residual leak sensitivity. Finally, the proposed approach is applied to a case study based on the Hanoi water network and compared against the angle method introduced in Casillas et al. (2012).

This paper is organized as follows: Section 2 presents an overview of the proposed approach. Section 3 describes the proposed methodology. Section 4 presents the results of the application of this methodology in Hanoi case study. Finally, Section 5 draws the main conclusions.

2. OVERVIEW OF THE PROPOSED APPROACH

2.1 Overview

The aim of the proposed methodology is to localize leaks in a water distribution network using pressure measurements and their estimation using the hydraulic network model. This methodology is complementary to the analysis of DMA night consumes that is used for detecting and estimating the leakage level (Puust et al. (2010)).

Model based leak localization method is based on comparing the monitored pressure disturbances caused by leaks at certain inner nodes of the DMA network with the theoretical pressure disturbances caused by all potential leaks obtained using the DMA network mathematical model. Thereby, the residual set, $r \in \Re^{ns}$, is determined by the difference between the measured pressure at inner nodes, $p \in \Re^{ns}$, and the estimated pressure at these nodes obtained using the network model considering a leak-free scenario, $\hat{p}_0 \in \Re^{ns}$

$$r(k) = p(k) - \hat{p}_0(k) \quad (1)$$

The size of the residual vector r, ns, depends on the number of inner pressure sensors of the DMA network. In (Pérez et al. (2011)), an optimal pressure sensor placement for leak localization was presented to achieve the minimum economical costs (number of sensors) keeping a suitable performance of the leak localization method. The number of potential leaks, $f \in \Re^{nn}$, is considered to be equal to the number of network nodes n_n, since from the modeling point of view as proposed in Pudar and Liggett (1992) and Pérez et al. (2014) leaks were assumed to be in these locations.

The fault diagnosis approach proposed in this paper is depicted in Fig. 1. It is based on using pressure residuals (1) between available pressure measurements in the network and estimations of these magnitudes provided by a hydraulic simulator (Epanet) combined with a statistical classifier. This classifier determines the most probable leak that produces the mismatches between the measurements and the estimations.

The hydraulic model simulated by Epanet considers a known structure (pipes, nodes and valves) and network parameters: pipe coefficients and estimated water demands (using historical billing records) in the nodes ($\hat{d}_1, \dots, \hat{d}_{n_n}$) since in general in real practice nodal demands $d_1, \dots, d_{n_n}$ are not measured (except for some particular consumers where

automatic metering readers (AMR) are available). On the other hand, pressure measurements are subject to the effect of sensor noise n. Finally, when a leak f_i appears in the network the leak magnitude $|f_i|$ is not completely known.
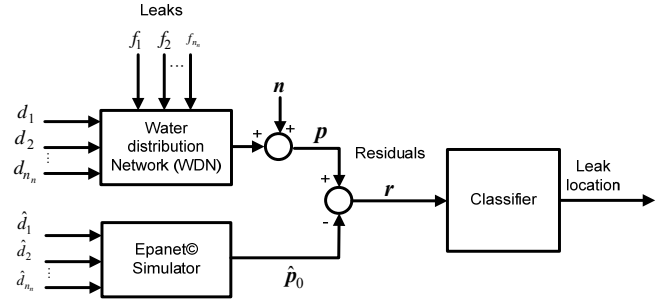

Fig. 1. Leak localization scheme

2.2 Motivation

The proposed approach summarized in Fig. 1 is an alternative to the sensitivity-to-leak analysis (Pérez et al., (2011, 2014)) where instead of using a classifier, the theoretical pressure disturbances caused by all potential leaks are stored in the sensitivity matrix $\Omega \in \Re^{ns \times nn}$ (with as many rows as DMA inner pressure sensors, ns, and as many columns as potential leaks (DMA network nodes, n_n)) is obtained as

$$\Omega = \begin{pmatrix} \dfrac{\partial p_1}{\partial f_1} & \cdots & \dfrac{\partial p_1}{\partial f_{n_n}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial p_{ns}}{\partial f_1} & \cdots & \dfrac{\partial p_{ns}}{\partial f_{n_n}} \end{pmatrix} \quad (2)$$

where each element s_{ij} measures the effect of the leak f_j in the pressure p_i of the node where the inner pressure sensor i is located. The leak isolation is based on matching the residual vector (1) against all the columns of the sensitivity matrix using some metrics (see Casillas et al. (2012) for details).

However, in practice, it is extremely difficult to calculate Ω analytically in a real network because a water network is a large scale problem described by a multivariable non-linear system of equations which may also be non-explicit. Thereby, the sensitivity matrix is generated by simulation of the network model approximating the sensitivity s_{ij} by

$$s_{ij} = \dfrac{\hat{p}_{if_j} - \hat{p}_{i0}}{f_j} \quad (3)$$

where $\hat{p}_{if_j}$ is the predicted pressure in the node where the pressure sensor i is placed when a nominal leak f_j is forced in node j and $\hat{p}_{i0}$ is the predicted pressure associated with the sensor i under a scenario free of leaks (Pérez et al. (2011)). Then, repeating this process for all n_n potential faults the approximation of the sensitivity matrix is obtained.

Another difficulty of the leak sensitivity approach is that the practical evaluation of (3) is highly depending on the nominal leak f_j. If the real leak size (in general unknown) is different from the nominal one, the real sensitivity will be different from the one computed using (3). This will lead to worsen leak localization results. The approach proposed in this paper aims to overcome these difficulties.

3. DESCRIPTION OF THE PROPOSED APPROACH

In this section, the methodology proposed to build a classifier that implements the leak localization and the performance evaluation of the proposed scheme (Fig. 1) are described.

3.1 Classifier training

Given a set of data $\mathbb{X}$

$$\mathbb{X} = \{(r_1, l_1), (r_2, l_2), \cdots, (r_N, l_N)\} \tag{4}$$

where $r_i \in \Re^{ns}$ are the pressure residuals (features) in different leak scenarios and $l_i \in \mathbb{L}$ are the labels (classes) that indicate the node where the leak has been produced i.e. $\mathbb{L} = \{1, 2, \cdots, n_n\}$. Then, the problem of building the classifier in Fig. 1 can be addressed by means of formulating a multilabel supervised classification problem. This problem consists in designing a function

$$g : \Re^{ns} \to \{1, 2, \ldots, n_n\} \tag{5}$$

See (Kotsiantis, 2007), for a complete survey of supervised machine learning algorithms.

In order to get an accurate classifier, the set of data $\mathbb{X}$ has to be enough representative of the real system, i.e. residuals r_i have to take into account all the differences between the real WDN and the Epanet model: different demand distributions, noise realizations and leak magnitudes. As it is not possible to obtain this amount of data for the real system, the residuals can be generated replacing the real system by the Epanet model, in the scheme depicted in Fig. 1, with different realizations of demand distributions $\tilde{d}_1, \ldots, \tilde{d}_{n_n}$, noise realizations $\tilde{n}$ and leak magnitudes $\tilde{f}_1, \ldots, \tilde{f}_{n_n}$.

3.2 Classifier evaluation

Once the classifier has been designed, its performance can be evaluated using a validation data set $\mathbb{X}'$

$$\mathbb{X}' = \{(r_1', l_1'), (r_2', l_2'), \cdots, (r_N', l_N')\} \tag{6}$$

Applying function g defined in (5) to the residuals r_i', the complete confusion matrix Γ defined in Table 1 can be obtained.

Table 1. Confusion matrix of the validation data

	$\hat{f}_1$	$\cdots$	$\hat{f}_i$	$\cdots$	$\hat{f}_{n_n}$
f_1	$\Gamma_{1,1}$	$\cdots$	$\Gamma_{1,i}$	$\cdots$	Γ_{1,n_n}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
f_i	$\Gamma_{i,1}$	$\cdots$	$\Gamma_{i,i}$	$\cdots$	Γ_{i,n_n}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
f_{n_n}	$\Gamma_{n_n,1}$	$\cdots$	$\Gamma_{n_n,i}$	$\cdots$	Γ_{n_n,n_n}

where $\Gamma_{i,i}$ indicates the number of times that a leak scenario in node i, f_i has been correctly diagnosed as $\hat{f}_i$ (true positive) and $\sum_{j=1}^{n_n} \Gamma_{i,j} - \Gamma_{i,i}$ indicates the number of times that a leak scenario in node i has been wrongly classified. From Table 1, the precision of correct diagnosis of every leak f_i can be defined as

$$\text{Prec}(f_i) = \frac{\Gamma_{i,i}}{\sum_{j=1}^{n_n} \Gamma_{i,j}} \tag{7}$$

If the precision index (7) is smaller than a desired threshold σ

$$\sigma = \frac{100 - \alpha}{100} \tag{8}$$

where α is the admissible percentage of wrong leak isolability, then a new class i^* (composed class) that groups class i and the other classes with the biggest $\Gamma_{i,j}$ such that

$$\frac{\sum_{j \in \{i^*\}} \Gamma_{i,j}}{\sum_{j=1}^{n_n} \Gamma_{i,j}} \geq \sigma \tag{9}$$

is created. After checking all the rows of the confusion matrix Γ, if the precision index (7) is bigger or equal to σ $\forall i = 1, \ldots, n_n$ then the classifier defined by Eq. (5) is considered suitable for the leak localization problem. On the contrary, if new classes have been created in order to fit (9) a new set of classes is defined as

$$\mathbb{L}_2 = \{1, 2, \cdots, n_{n2}\} \quad \text{with } n_{n2} < n_n \tag{10}$$

where the classes of $\mathbb{L}_2$ contain the new composed classes obtained in the grouping process described above and the classes of $\mathbb{L}$ that are not contained in the new classes.

Once the new set of classes $\mathbb{L}_2$ has been defined, new sets of data $\mathbb{X}_2$ and $\mathbb{X}_2'$ are generated to train a new classifier

$$g_2 : \Re^{ns} \to \{1, 2, \ldots, n_{n2}\} \tag{11}$$

and, with a new confusion matrix Γ_2 obtained applying (11) to residuals of set $\mathbb{X}_2'$. Then, if the precision index (7) is bigger or equal to σ $\forall i = 1, \ldots, n_{n2}$, the grouping process is stopped. Otherwise, the process of grouping the number of classes,

training a new classifier and evaluating its behavior is repeated until this condition is satisfied.

3.3 Leak localization evaluation

Once a classifier has been obtained

$$g_F : \Re^{ns} \rightarrow \{1, 2, \ldots, n_{nF}\} \qquad (12)$$

that provides a precision index (7) bigger than σ $\forall i = 1, \ldots, n_{nF}$, being n_{nF} the final number of different classes. The composed classes that contain more than one of the original classes in $\mathbb{L}$ (i.e. nodes in the WDN) are studied.

The main drawback of grouping original classes in order to get a given precision in the leak location is that when classifier (12) will classify a leak scenario in a composed class it will not be able to distinguish between the nodes related to this class. Thus, the provided leak localization will have an uncertainty given by the number of these nodes and distance between them.

4. CASE STUDY

This section illustrates the application of the proposed methodology to a benchmark water distribution network: the Hanoi water network. Moreover, the obtained results are compared to the ones obtained by applying the leak-sensitivity method (summarised in Section 2.2) using the angle metric proposed in Casillas et al. (2012).

4.1 The Hanoi network

The Hanoi network (see Fig.2) consists of one reservoir, 34 pipes and 31 junction nodes. Leaks in all nodes are considered. The known variables are the reservoir pressure and the pressures in nodes 14 and 30. The demand in each node is assumed to be uncertain, i.e. the instantaneous values are unknown but inside known limits.
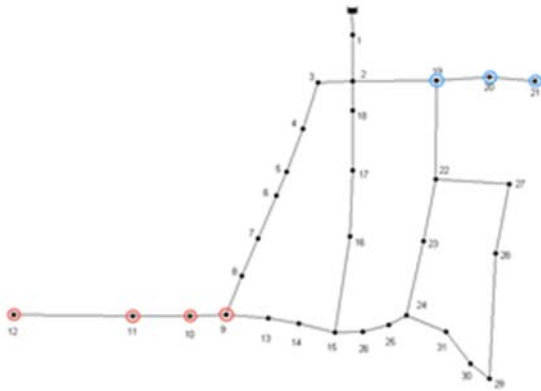

Fig. 2. Hanoi water network.

4.2 Classifier training

For the Hanoi network, the direct use of the raw residuals associated to the measured pressures in nodes 14 and 30 would provide poor isolation results. This can be easily seen in Fig.

3, which shows points in the residual space computed for a leak of 50 l/s (1.7% of the total demand) with 10% uncertainty on demands.
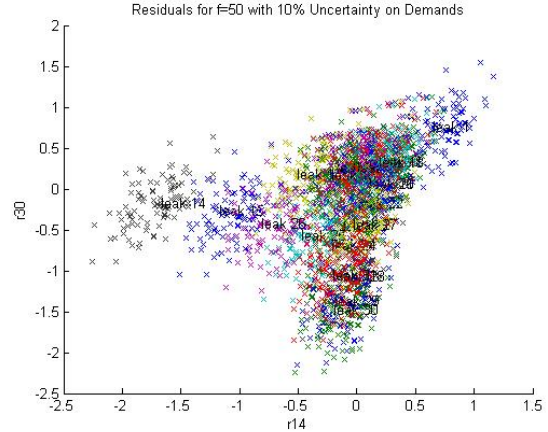

Fig. 3. Raw residuals for a leak size of 50 l/s with a 10% uncertainty in demands.

In order to improve the leak localization results, a FIR filter is applied to the residuals. This procedure aims to geometrically separate the points associated to one leak scenario from the others associated to other leaks. The implemented filter is defined as

$$\bar{r}(k) = \left(h_0 + h_1 q^{-1} + \ldots + h_T q^{-T} \right) r(k) \qquad (13)$$

with the following coefficients:

$$h_0 = \frac{1}{S} + \frac{1}{T+1} \frac{S-1}{S}$$

$$h_i = \frac{1}{T+1} \frac{S-1}{S}, \quad i = 1, \ldots, T \qquad (14)$$

where T is the order of the filter and $S > 1$ is a scaling ratio that provides an extra weight to the current residual $r(k)$.

The performance of the residual filtering depends on the filter parameters T and S. In principle, the bigger the two parameters are the better the leak localization performs. However, it is of interest to determine if there is a point for which an increment in the value of both parameters does not improve substantially the obtained results, whose quality can be associated with the final number of distinguished classes.

Fig. 4 shows the effect of the filter order (or window size) T on the number of classes, considering $\sigma = 0.5$ in Eq. (8), for different leak magnitudes, assuming a scaling ratio S=100. It can be observed that for each considered leak magnitude, the number of classes increases with the filter order at the beginning (small values for the filter order) and then it becomes almost constant around a given maximum value. Moreover, it can be observed that the value of T for which the number of classes stabilizes is quite similar and around T=20 for all the considered leak magnitudes. This value is chosen as the optimal filter order.
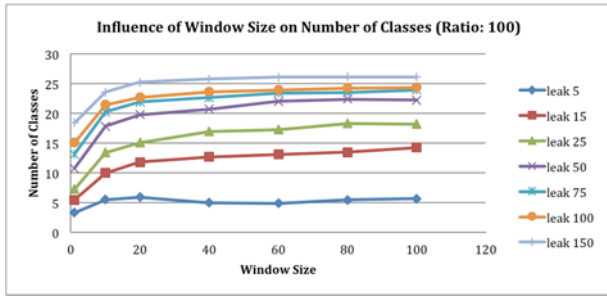
Fig. 4. Influence of the filter order T on the number of obtained classes for different leak magnitudes, with S=100.

Fig. 5 shows the influence of the scaling ratio *S* on the number of classes for different leak magnitudes, assuming the use of the optimal filter order given by *T*=20. It can be seen that for each considered leak magnitude, the number of classes becomes almost constant around a given maximum value when the value for the scaling ratio is higher than a given threshold value. However, this threshold value for *S* is different for different leak magnitudes. Hence, instead of working with a fixed value for *S*, working with a set of different values of *S* associated to different ranges of leak magnitudes is proposed. The set of chosen values is summarized in Table 2.
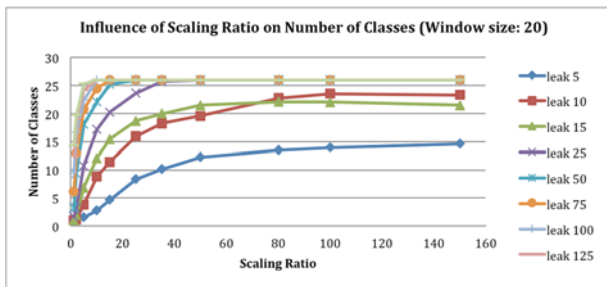

Fig. 5. Influence of the scaling ratio S on the number of obtained classes for different leak magnitudes, with T=20.

Table 2. Selected values for the scaling ratio *S* for different ranges of the leak magnitude.

Leak range (l/s)	Scaling ratio *S*
0-25	100
25-50	50
50-100	20
100-150	10
>150	5

The results of the residual filtering process can be seen in Fig. 6, which shows points in the residual space for a leak of 50 l/s with 10% uncertainty on demands (the same conditions leading to the raw residuals plotted in Fig. 3). Comparing with Fig. 3, it can be seen that much more classes will be easily distinguished, although not as many as possible faults.

Working with leaks in the interval [20,60] l/s (between 0.67% and 2% of the total demand), after applying the proposed data

transformation and reduction of number of classes procedures, the classifier is proven to distinguish up to 28 classes (for 31 possible leak locations), being the most common case one in which 26 classes are identified: 24 classes containing just one leak location, one class associated to three leak locations and one class associated to four. These classes are shown in Fig. 2. Non-circled nodes belong to classes containing only one leak position while nodes circled in a given colour belong to the same class. It must be noticed that leaks associated to a same class correspond to nodes situated in the same area. Summarizing, 24 leaks are expected to be perfectly located while the other seven can only be approximately located.
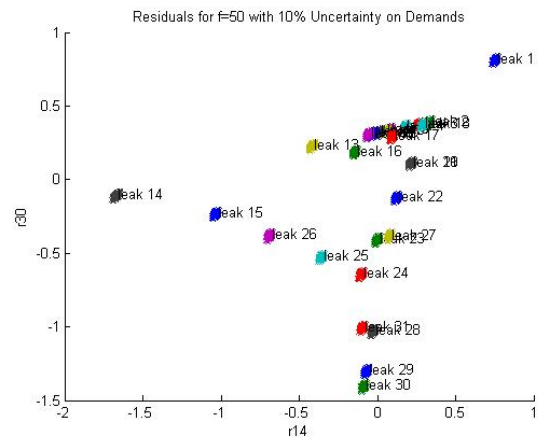

Fig. 6. Transformed residuals for a leak size of 50 l/s with a 10% uncertainty in demands.

4.3 Leak localization results

In order to illustrate the good performance of the proposed methodology, a complete set of results for leak localization in the Hanoi network is summarized in this subsection. Moreover, the obtained results are compared with the results obtained by using the leak-sensitivity analysis with the angle metrics proposed in Casillas et al. (2012).

The comparison of the proposed method with the *angle method* is not straightforward since both methods present different types of results. The *angle method* always indicates (correctly or incorrectly) a precise leak location, whereas the method based on classifiers proposed in this paper may provide as result that the leak can be in several (close between them) possible locations. Due to this, the proposed method is completed with classifiers that are applied in a second round if the class to which the leak is identified to belong is associated to more than one location, trying to determine the exact location.

Table 3 shows average percentages for perfect isolation obtained by using both the proposed method and the angle method under different conditions. The averages are computed based on the results of applying leak magnitudes of 20, 30, 40, 50 and 60 l/s to each node (5*31 single-fault scenarios). Line entitled "Noise 2%" corresponds to values computed with a noise of 2% on measurements and no uncertainties.

"Uncertainty on leak 10%" corresponds to a 10% uncertainty on leak magnitude, none on demands and no noise. "Uncertainty on demands 10%" corresponds to a 10% uncertainty on demands, none on leak magnitude and no noise. Finally, "All" corresponds to 10% uncertainties on leak and demands and 2% noise. Training values are computed with the same nominal leak magnitudes, uncertainties and noise as the values used for testing.

Table 3. Percentage (%) of location of known leaks in the range [20,60] l/s.

Known leak size	Classifier	*Angle method*
Noise 2%	**83,55**	72,87
Leak uncertainty 10%	83,96	**87,74**
Demand uncertainty 10%	**54,48**	25,16
All	**53,02**	25,16

The results presented in Table 3 show that the classifier is likely to be preferred over the angles method, especially when there are demand uncertainties. From a practical point of view, it may be sufficient to localize the leak within a given (small) area. For instance, with a relaxation of two nodes, it may be considered that the location is correct if a leak is attributed to a node that is situated less than two nodes away from the node that really presents the leak. This aspect is considered in the results presented in Table 4.

Table 4. Percentage (%) of location with error distance <2 for leaks in the range [20,60] l/s .

	Classifier	*Angle method*
Noise 2%	**98,21**	95,48
Leak uncertainty 10%	98,04	**98,71**
Demand uncertainty 10%	**82,40**	52,90
All	**82,76**	57,42

The values presented in Table 4 show that when there is no uncertainty on demands the two methods provide quite similar results in all cases. However, the method based on classifiers is significantly more accurate when there is uncertainty on demands. The difference of accuracy in this case seems significant enough to justify the use of the classifier method since the uncertainty on demands is a troublesome problem in practical leak localization applications.

Finally, it has been considered that the magnitude of the nominal leak used to train/tune the methods can be different from the real leak in the system. Results with training values have been computed with a leak magnitude value 10 l/s higher and lower than the tested values. The efficiencies given for leaks of magnitudes 20, 30, 40, 50 and 60 l/s are similar to the obtained considering perfect known leak magnitude, i.e. around 80% using the classifier method and around 50% using the angle method.

5. CONCLUSIONS

In this paper, a new fault localization approach based on the combined use of models, pressure sensors and classifiers has been proposed. The hydraulic model of the network is used to generate residuals by comparing model predictions against the available measurements provided by sensors. Once residuals have been generated they are analysed using supervised classifiers in order to allow the leak localization. The classifiers are calibrated using data in leak scenarios in all the nodes of the network considering uncertainty in demand distribution, additive noise in sensors and uncertainty in leak magnitude. Finally, the proposed approach has been successfully tested using the well-known Hanoi network benchmark.

REFERENCES

Casillas M.V., Garza L.E. and Puig V., Extended-Horizon Analysis of Pressure Sensitivities for Leak Detection in Water Distribution Networks, 8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Safeprocess (2012), Mexico, p.570-575

Kotsiantis, S.B. Supervised machine learning: A review of classification techniques (2007) Informatica (Ljubljana), 31 (3), pp. 249-268.

Pérez R., Puig V., Pascual J., Quevedo J., Landeros E., Peralta A., Methodology for leakage location using pressure sensitivity analysis in water distribution networks, Control Engineering Practice, Volume 19, Issue 10, October (2011), p. 1157-1167.

Pérez, R., Sanz, G., Quevedo, J., Nejjari, F., Meseguer, J., Cembrano, G., Tur, J.M.M., and Sarrate, R.,. Leak Localization in Water Networks. IEEE Control Systems Magazine, Volume 34, Issue 4, August (2014), pp. 24-36.

Pudar S. and Liggett J., Leaks in Pipe Networks, Journal of Hydraulics Engineering, 118 (7), (1992), p. 1031-1046.

Puust, R. Z. Kapelan, D. A. Savic, and T. Koppel, "A review of methods for leakage management in pipe networks," Urban Water Journal, vol. 7, no. 1, pp. 25–45, 2010.

Ragot J. and Maquin D., Fault measurement detection in an urban water supply network. Journal of Process Control, 16 (2006) 887–902.

Savic, D., Kapelan, Z., Jonkergouw, P. Quo vadis water distribution model calibration? Urban Water Journal, Vol. 6, No. 1 (2009). pp. 3-22.

Wu Z. et al. Water Loss Reduction. Bentley Systems 2011.