

Extended-Horizon Analysis of Pressure Sensitivities for Leak Detection in Water Distribution Networks

Myrna Violeta Casillas Ponce*, Luis Eduardo Garza Castañón*,
Vicenç Puig Cayuela**

*Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, 64849
México (Tel: 81-83-582000 ext. 5486; e-mail: mv.casillas.phd.mty@itesm.mx, legarza@itesm.mx).

** Universidad Politécnica de Cataluña, Advanced Control Systems Group,
Barcelona, España, (e-mail: vicenc.puig@upc.edu)}

Abstract: In this paper, a new approach for leak detection in water networks is proposed which considers an extended time horizon analysis of pressure sensitivities. Previous works based on pressure sensitivities analysis were developed by considering time instant evaluation. This fact makes them very sensitive to demand changes and noise in measurements. The proposed approach has been combined with five detection methods: the first is the binarization method, the second, third and fourth are based on the comparison of measured pressure vectors with leak sensitivity matrix using methods of correlation, vector angle and Euclidean distance respectively. And, finally, the fifth is based on the least square optimization method. Another contribution of this paper is the performance comparison between the five detection methods in presence of single leaks in scenarios with noise in measurements and nodal demands. Results showed that in most of the methods effectiveness is high, being the best the vector angle method, with effectiveness higher than 96%. The correlation and optimization method had similar behaviors with effectiveness superior to 90%. Finally, the binarization method is effective only in some scenarios but in presence of noise has a poor performance.

1. INTRODUCTION

Water leaks in networks can cause significant economic losses in the fluid transportation and an increase on reparation costs, giving as a consequence an extra cost translated to the final consumer. In many water distribution systems (WDS), losses due to leaks are estimated to account up to 30 % of the total amount of extracted water. Such burden cannot be tolerated in a world struggling with satisfying water demands of a growing population.

In this research, we present five different methodologies to detect leaks in water distributions system based on an extended horizon analysis of pressure sensitivities. The first method, called *Sensitivity Matrix Binarization*, is based on the transformation of real-valued sensitivities matrix to a binary matrix, according to a threshold. The second is the *Correlation method*, the third and fourth are the *Euclidean distance* and the *Angle between vectors methods*; the last method is the *Least square Optimization method*. The five methodologies were tested in simulation with two different water distribution systems named Hanoi and Quebra.

Several works have been published on leak detection and isolation for WDS. In (Yang et al., 2008), a method to identify leaks is proposed using blind spots based on previously leak detection researches that uses the analysis of acoustic and vibrations signals (Fuchs and Riehle, 1991), models of buried pipelines to predict wave velocities (Muggeleton et al., 2002), among others. In addition, the detection of pipelines leaks can also be possible using the

inverse problem (Pudar and Liggett, 1992) which uses pressure and flow measurements.

In (Mashford et al., 2009), a method to locate leaks is presented using Support Vector Machines (SVM). This research presents a method to analyze data obtained by a set of pressure control sensors of a pipeline network to locate and calculate the size of the leak. In (Covas and Ramos, 2001), a method to detect and locate leaks based on the transitory inverse analysis is presented. The main idea of this methodology is the location of leaks in a network based on the pressure observed data collected during the occurrence of the transitory events and the minimization of the difference of the observed and calculated parameters. In (Ragot and Maquin, 2006), a technique to isolate leaks using fuzzy analysis of the residuals is presented. This method finds the residuals between the nominal measurements (no fault presence) and the faulty measurements. In (Blesa et al., 2010), an LPV model based leak detection approach that consider uncertainty using zonotopes is introduced and tested in a small water network. In (Perez et al., 2010), a method based on pressure measurements and leak sensitivity analysis is presented. In this methodology, first a model free of leaks is obtained off-line and then the residuals are analyzed on-line against a proposed threshold. If any inconsistency is found, an analysis to detect and isolate the leaks begins using an established mapping. Although this approach has good efficiency under ideal conditions, its performance is diminished with changes in demand and noise in measurements. In current paper, the approach proposed by (Perez et al., 2010) is improved by making an extended-

horizon analysis of pressure sensitivities and residuals, and introducing adequate isolation algorithms to locate the leaks.

This paper is organized as follows: In Section 2, the proposed methodology is shown. Section 3 presents the considered case study water networks. Section 4 collects the experiments and results. And, finally, Section 5 presents the conclusions.

2. METHODOLOGY

2.1 Introduction

The main objective of the proposed scheme is to detect, isolate and identify leaks in a hydraulic network, using pressure measurements in the nodes. A leak will be considered as a water flow loss through a defect of an element of the network that is being measured. We consider the existence of a single and continuous leak at a given time. Figure 1 shows the leak detection, isolation and identification process. Data of node pressures are obtained from extensive simulations of normal and leak scenarios. From these data, pressure sensitivities and residuals are obtained for a time window and then analyzed by several leak detection algorithms. In real cases, the proposed methods should be applied taking into account that the number of pressure measurements will be reduced.

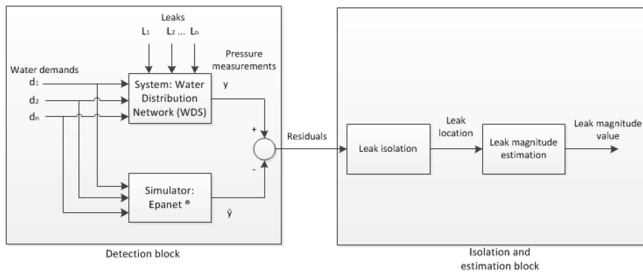


Figure 1 Detection, isolation and estimation diagram

All the leaks are simulated in the nodes of the network. This is performed by adding an extra demand of water at a specific node and by using two patterns of water demands: one to simulate the nominal water demand and the other one to simulate the leak.

A comparison of five different methods to detect and isolate the leaks was performed and applied to a time horizon data set, instead of only at a given time instant. To compare the efficiency of each method, changes in the leak magnitude and noise in the measurements and nodal demands were simulated. A time horizon window of 24 hours was selected for the simulations.

The proposed methods can be divided in *direct* and *indirect methods*. The direct methods can be classified as binary or non-binary. The non-binary direct methods can be divided in three types according to if they are based on residual correlation, Euclidean distance and the angle of the pressure vector with the leak signatures. On the other hand, the indirect method is based on a least-square optimization method. In all of the above methods, to initiate the detection of the leak, a sensitivity matrix that quantifies the effect of

leaks in alls nodes and pressure sensors in the network is needed.

Matlab® and Epanet® are used altogether to simulate the leaks and to obtain and analyze the network data using the algorithms proposed in the paper.

2.2 Sensitivity matrix algorithm

This algorithm is divided in three main stages. The first stage is the construction of the nominal operation scenario of the network. At this stage, two matrices are obtained. One for the pressure in each node of the network and another one for the water demand of each node as follows

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,m} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,m} \end{bmatrix}, P_a = \begin{bmatrix} p_{a,1,1} & \cdots & p_{a,1,m} \\ \vdots & \ddots & \vdots \\ p_{a,n,1} & \cdots & p_{a,n,m} \end{bmatrix} \quad (1)$$

where:

- P is the matrix of pressures in no leak situation
- $p_{i,j}$ represents the pressure of node i at time j in no leak situation
- P_a is the matrix of actual pressures
- $p_{a,i,j}$ is the actual pressure of node i at time instant j
- n is the number of nodes in the network and
- m is the number of samples through the simulation time.

The second part is the construction of scenarios with the presence of a leak. The pressures in nodes when a leak in the network is present are stored in the following matrix

$$P_f^k = \begin{bmatrix} p_{f,1,1}^k & \cdots & p_{f,1,m}^k \\ \vdots & \ddots & \vdots \\ p_{f,n,1}^k & \cdots & p_{f,n,m}^k \end{bmatrix} \quad (2)$$

where

- P_f^k is the matrix of pressures when a leak is present at node k
- $p_{f,i,j}^k$ is the pressure of node i at time instant j when a leak is present at node k
- n, m are the same as defined previously

In the third stage of the algorithm, the residual matrices are obtained as follows

$$\mathcal{R}_k = P - P_f^k = \begin{bmatrix} r_{1,1}^k & \cdots & r_{1,m}^k \\ \vdots & \ddots & \vdots \\ r_{n,1}^k & \cdots & r_{n,m}^k \end{bmatrix} \quad (3)$$

$$\mathcal{R}_a = P - P_a = \begin{bmatrix} r_{a,1,1} & \cdots & r_{a,1,m} \\ \vdots & \ddots & \vdots \\ r_{a,n,1} & \cdots & r_{a,n,m} \end{bmatrix} \quad (4)$$

where

- \mathcal{R}_k is the matrix of residuals when a leak is present at node k
- $r_{i,j}^k = p_{i,j} - p_{f,i,j}^k$ is the residual of node i at time j when a leak is present at node k
- \mathcal{R}_a is the matrix of actual residuals
- $r_{a,i,j} = p_{i,j} - p_{a,i,j}$ is the actual residual of node i at time j

In the leak isolation phase, residuals should be analyzed by using the leak sensitivity matrix defined in the following way

$$S_k = \frac{p-p_f^k}{l} = \begin{bmatrix} s_{1,1}^k & \cdots & s_{1,m}^k \\ \vdots & \ddots & \vdots \\ s_{n,1}^k & \cdots & s_{n,m}^k \end{bmatrix} \quad (5)$$

where

S_k is the matrix of sensitivities when a leak is present at node k

$s_{i,j}^k = \frac{r_{i,j}^k}{l}$ is the sensitivity of node i at time j when a leak is present at node k

l is the assumed leak magnitude (liter/second).

2.3 Binarization sensitivity method

To implement this method we follow these steps:

- Binarize the sensitivity matrices according to an established threshold. See equation (6).
- Extract the actual residuals matrix \mathcal{R}_a and binarize it in a similar way to previous step. See equation (7).
- Compare each of the actual residual vectors against each vector of the sensitivity matrices; when the algorithm find a vector equal to the vector calculated with one of the sensitivity matrices means that the leak occurred at that instant of time
- Because the leaks are analyzed for a time horizon of 24 hours, it is necessary to count the coincidences found on previous step in order to find the maximum number of this coincidences for a given node.
- Once the leaking node is found, the magnitude of the leak is calculated. This is realized by identifying the actual residuals column that corresponds to the node of the leak and dividing each of them by the sensitivity matrix values.

The binarization procedure is represented by:

$$S_{bin_{m,i,j}} = \begin{cases} 1 & \text{if } S_{m,i,j} > 0.95 * \max(S_{m,:j}) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$\mathcal{R}_{bin_{i,j}} = \begin{cases} 1 & \text{if } \mathcal{R}_{i,j} > 0.95 * \max(\mathcal{R}_{:,j}) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where S_m represents the m calculated sensitivity matrices according to the necessary samples of time and \mathcal{R} is the actual residuals matrix with respect to which it is desired to find the leaks.

If $\mathcal{R}_{bin_{:,j}} = S_{bin_{m,:j}}$, then the leak at the time instance of m is assumed to be present in the node with the j subindex. As result of this comparison, the matrix (8) is created in which the binary indicators of the existence or absence of the leak will be saved.

$$leak_{bin_{j,m}} = \begin{cases} 1 & \text{if } \mathcal{R}_{bin_{:,j}} = S_{bin_{m,:j}} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

With the above procedure, an m size vector that contains the leak indexes is obtained, which means that a leak index is detected for each sample time. To calculate the leak in the time horizon, the verification of the index that appears more

times through the time horizon is done and this index is assigned as the leak index during the provided time. This is formulated as follows:

$$vector_{leak} = \sum_{i=1}^m leak_{bin_{j,i}} \quad (9)$$

where $vector_{leak}$ is a vector that contains the number of fault indications for the possible leaks according to the column that is occupying. Thus, if the maximum of this vector is found, the index of the node that contains the leak in the desired time horizon is obtained. This means:

$$leak_{index} = \max(vector_{leak}) \quad (10)$$

2.4 Data correlation method

In this algorithm, it is necessary to obtain the correlation between each of the actual residuals vectors and the sensitivities vectors. Such computation is realized using

$$C_m = \frac{\sum_{i=1}^n (\mathcal{R}_a(m)_i - \overline{\mathcal{R}_a(m)}) * (S_k(m,i) - \overline{S(m,i)})}{\sqrt{\sum_{i=1}^n (\mathcal{R}_a(m)_i - \overline{\mathcal{R}_a(m)})^2} * \sqrt{\sum_{i=1}^n (S_k(m,i) - \overline{S(m,i)})^2}} \quad (11)$$

where C_m is the correlation vector for each sampling time m , $\mathcal{R}_a(m)$ represents the residual, while S_k represents the residual leak sensitivity along the k possible leaks from 1 to n .

Once the correlation vectors are obtained, a correlation matrix is formed. The matrix row that has the higher correlation will correspond to the index of the node where the leak is present at the sampling time indicated by the column in which it is placed

$$leak_node_m = \max(C_m) \quad (12)$$

In order to obtain the leak in the analyzed time horizon, it is necessary to perform the sum of all the obtained correlations at every instant. The above means that: once the correlations of each vector of actual residuals are calculated against the sensitivity matrices, a correlation matrix will be obtained in which each element represents the correlation of a node in the considered time instant with each possible leak. To identify the node where the leak is, the following indicator is computed in the considered time horizon

$$C_h = \sum_{i=1}^m C_m \quad (13)$$

and the leaky node is identified as follows

$$leak_{index} = \max(C_h) \quad (14)$$

Once the node index that contains the leak is obtained, it is possible to obtain the node which it refers making possible the identification of the leak. To calculate the leak magnitude, it is necessary to remind the definition of the leak sensitivity matrix (5). Then, the leak magnitude can be calculated as follows

$$leak_m = \frac{\mathcal{R}_a}{s_i^k} \quad (15)$$

2.5 Angle between vectors method

This algorithm calculates the angle between the vector of the actual residuals against every column (leaky node) of the leak sensitivity matrices. The leaky node is the column (sensitivity vector) that presents smaller angle with the residual vector. The angle is calculated as follows

$$\alpha_{m_i} = \arccos\left(\frac{\mathcal{R}_a(m) \cdot S_k(m, i)}{|\mathcal{R}_a(m)| \cdot |S_k(m, i)|}\right) \quad (16)$$

The above means that there will be a total of m vectors of size i according to the number of possible leaks and time instants to analyze in the established time horizon. $S_k(m, i)$ in equation (16) represents the same than in equation (11). Because the calculus is in a time horizon, it is necessary to calculate the angle mean throughout the time as follows

$$\alpha_h = \frac{\sum_{i=1}^m \alpha_{m_i}}{m} \quad (17)$$

Once the vector of angles of all the possible leaks is obtained, it is necessary to determinate the smaller obtained angle using

$$leak_{index} = \min(\alpha_h) \quad (18)$$

2.6 Euclidean distance method

Alternatively to the angle between vectors method, we can apply the Euclidean distance between them computed as follows

$$\mathcal{D}_{m_i} = \sqrt{\sum_{k=1}^n (\mathcal{R}_a^k(m) - S^k(m, i))^2} \quad (19)$$

where \mathcal{D}_{m_i} represents the distance vector between the current residual and a given column (leak) of the sensitivity matrix found for a given instant time. Then, the distance vector for the time horizon is calculated as

$$\mathcal{D}_h = \sum_{i=1}^m \mathcal{D}_{m_i} \quad (20)$$

As this vector contain each element of the Euclidean distance vector to every possible leak, we conclude that the present leak will be the minimum value of that vector

$$leak_{index} = \min(\mathcal{D}_h) \quad (21)$$

This method works just in the case that the leak has the same magnitude than the one used to compute the sensitivity matrix. If this is not the cases, it does not provide good results. For this reason, the results obtained using this method has not been included in the paper.

2.7 Least square optimization method

It can observed from (5) that it is possible to find the magnitude of a leak present in a node based on

$$leak_m^k = S_k^{-1} * \mathcal{R}_a \quad (22)$$

However, in case that not all nodal pressure measurements are available, the S_k matrix is not invertible. Then, alternatively it is proposed to find the minimum error on the difference from actual residuals against the sensitivity matrix

for the considered leak. This leads to formulate a least square optimization problem with the following objective function

$$\min J = \sum_{i=1}^m (\mathcal{R}_a - S_k * leak_m)^2 \quad (23)$$

where the optimization variable is leak $leak_m$. This method allows knowing the leak node and at the same time obtaining the approximate magnitude of the leak.

3. DESCRIPTION OF WATER NETWORKS FOR EXPERIMENTS

To test the above methodologies, two networks were used: Hanoi (from Rodríguez et al 2006), and Quebra (provided by EPANET).

3.1 Hanoi network

This network is presented in Figure 2. It will allow to analyze the effectiveness of the proposed methods in a network with big flows.

The demand pattern is designed according to (Rodríguez et al 2006). A simulation of 24 hours with a sampling time of 15 minutes is carried out. This is because the demand is measured each 15 minutes. This give a total of 97 samples.

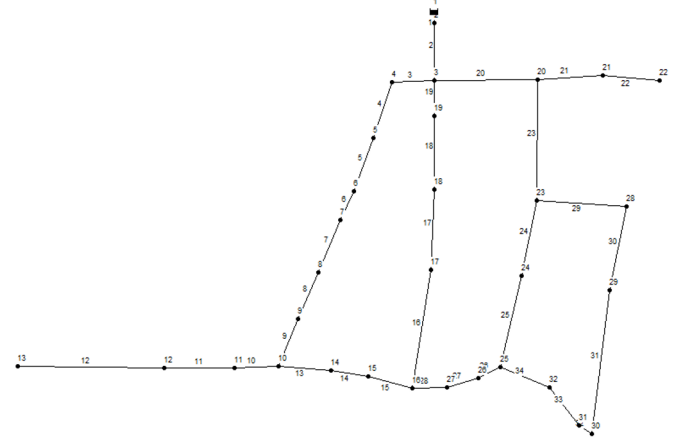


Figure 2 Hanoi network

This network has 31 demand nodes with indexes from 2 to 32. A leak of 50 liters per second magnitude is used to compute the sensitivity matrixes shown in Figure 3.

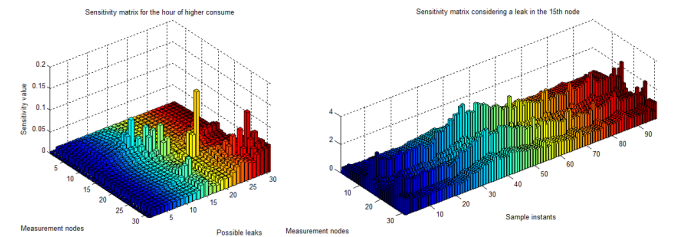


Figure 3 Sensitivity matrixes for Hanoi Network

3.2 Quebra Network

This network is presented in Figure 4. It will allow to analyze the performance of the proposed methods using a network of bigger size than the Hanoi network. Quebra is a network designed according to the method presented in the EPANET webpage.

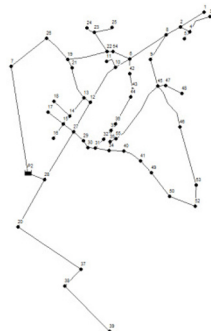


Figure 4 Quebra network

In this network, the demand is measured with a sampling of one hour. The simulation is carried out for 24 hours giving a total of 25 samples. The following parameters used in the simulation of the network are established: The network is composed of 55 nodes, the samples are taken every hour and the leak is placed in the node 34 which corresponds to the 32 index. The sensitivity matrices were calculated with a leak magnitude of 0.01 liters per second. Figure 5 show the values of the sensitivity matrix at the sample instant of maximum consume and where there is the leak.

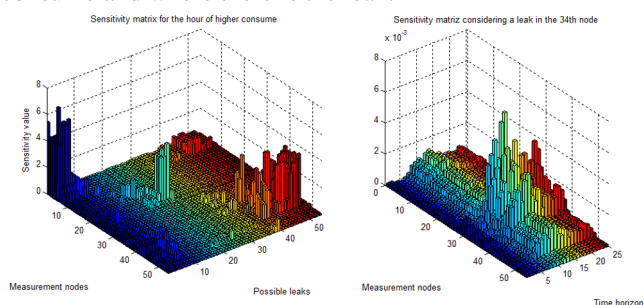


Figure 5 Sensitivity matrixes for Quebra network

4. EXPERIMENTS AND RESULTS

To test the proposed methodologies, the following experiments were developed.

4.1 Applications of the methods to the networks

For every network, each one of the following experiments were carried out:

1. Impact analysis of leak magnitude.
2. Application of random demand noise between $\pm 2\%$ and $\pm 4\%$ of the medium demand along the time horizon.
3. A study of the effect of the measurement noise, applying Gaussian white noise around of $\pm 2\%$ of the nominal pressures.

4. Application of both uncertainties introduced in Step 2 and Step 3

5. Finally, both effects were tested with 200 random leaks location with and without noise whose size depends on the network, i.e., with sizes around 20 to 80 liters per second for the Hanoi Network, and from 0.01 to 1 litter per second for Quebra network.

The results of tests 1, 2, 3 and 4 are shown in Table 1 and 3. It can be observed that each method delivers very good results.

The results of test 5 are shown in Table 2 and 4 where it can be observed that best method is the angle between vectors.

In all the tables, the effectiveness is shown in percentage obtained according to the number of leaks detected satisfactorily divided for the number of tests realized. It is important to mention that the network structure has an important impact in the results. This means that introducing appropriate structural changes in the network, even better performance of the methods could be achieved. After analyzing the result of the tests, it can be concluded that methods based on optimization and vector angle provide excellent results.

EFFECTIVENESS (%) FOR EACH METHOD IN HANOI NETWORK TEST OF EFFECT OF 2% NOISE ON DEMAND AND MEASUREMENTS				
Leak size	Binarization	Correlation	Angle	Optimization
50.00	51.61	96.77	100.00	96.77
10.00	38.71	83.87	90.32	83.87
20.00	38.71	93.55	96.77	87.10
30.00	38.71	90.32	96.77	90.32
40.00	51.61	96.77	100.00	100.00
60.00	58.06	96.77	100.00	93.55
70.00	58.06	96.77	100.00	96.77
80.00	61.29	96.77	100.00	100.00
Average efficiency	48.76	93.84	97.93	93.38

Table 1 Efficiency in tests applied to Hanoi network

EFFECTIVENESS (%) IN RANDOM TESTS FOR HANOI NETWORK				
Test	Binarization method	Correlation method	Angle method	Optimization method
No noise	100.00	100.00	100.00	100.00
Demand noise	86	90	98	94
Measure noise	60	70	98	98
Noise in both	48	60	98	96

Table 2 Efficiency in random tests for Hanoi network

EFFECTIVENESS (%) FOR EACH METHOD TEST OF EFFECT OF 2% NOISE ON DEMAND AND MEASUREMENTS				
Leak size	Binarization	Correlation	Angle	Optimization
0.01	62.96	88.89	92.59	88.89
0.03	46.30	79.63	81.48	85.19
0.02	66.67	94.44	94.44	94.44
0.08	79.63	98.15	98.15	98.15
0.15	83.33	96.30	98.15	98.15
0.20	85.19	96.30	98.15	98.15
Average efficiency	69.20	92.05	93.63	93.69

Table 3 Efficiency in tests applied to Quebra network

EFFECTIVENESS (%) IN RANDOM TESTS FOR QUEBRA NETWORK				
Test	Binarization method	Correlation method	Angle method	Optimization method
No noise	98	100	100	100
Demand noise	98	100.00	100.00	100.00
Measure noise	72	94.5	97.5	98
Noise in both	71.5	94.5	98.5	98.25

Table 4 Efficiency in random tests for Quebra network

5. CONCLUSIONS

This paper has presented five different methods for the isolation of leaks in water networks. The binarization method has the advantage of relating residuals and leaks in a binary form, but gives only satisfactory results when it works in noise-free scenarios.

The correlation method compares the actual residual with the residual sensitivity to each possible leak in order to decide where the leak is located. This method is almost not affected by a change in the leak magnitude and to the presence of noise.

The angle between vectors method is the best of the proposed schemes according to the results obtained from experiments done in this research. The method is not affected by a change in the leak magnitude and the presence of noise.

Finally, the least squares optimization method has the advantage of calculating the magnitude of the leak at the same time as the node where it is located.

REFERENCES

G. Cembrano, J. Quevedo, V. Puig, J.Figueras, R. Pérez. Mejoras en la gestión del agua con técnicas de modelado y control *Journal de Tecnología del Agua* (2006), p. 28-32.

J. Yang, Y. Wen, P. Li. Leak location using blind system identification in water distribution pipeline, *Journal of Sound and Vibration* 310 (2008), p. 134–148.

H.V. Fuchs, R. Riehle. Ten years of experience with leak detection by acoustic signal analysis, *Applied Acoustics*, 33 (1991), p. 1–19.

J.M. Muggleton, M.J. Brennan, R.J. Pinnington. Wavenumber prediction of waves in buried pipes for water leak detection, *Journal of Sound and Vibration* 249 (5) (2002), p. 939–954.

J. Mashford, D. De Silva, D. Marney and S. Burn. An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. *Third International Conference on Network and System Security* (2009), p. 534-539

D. Covas and H. Ramos. Hydraulic Transients used for Leak Detection in Water Distribution Systems, 4th International Conference on Water Pipeline Systems, BHR Group, (2001), p. 227-241.

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

K. Rodríguez, O. Fuentes, M. Jiménez, F. De Luna. Diseño Óptimo de Redes de Distribución de Agua Potable utilizando un Algoritmo Genético Multiobjetivo. *Seminario Iberoamericano sobre Sistemas de Abastecimiento Urbano de Agua*, Junio 5-7, (2006).

J. Blesa, V. Puig, J. Saludes, J. Vento. Leak Detection, Isolation and Estimation in Pressurized Water Pipe Networks Using LPV Models and Zonotopes, *IFAC Symposium in Nonlinear Control Systems, Volume # 1*, (2010), p- 36-41.

Ra. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, A. Peralta. Leakage Isolation using Pressure Sensitivity Analysis in Water Distribution Networks: Application to the Barcelona case study, 12th *IFAC Symposium on Large-Scale Systems: Theory and Applications* (2010), p. 1-6.

J. Gertler, J. Romera, V. Puig and J. Quevedo, Leak Detection and Isolation in Water Distribution Networks using Principal Components Analysis and Structured Residuals. "2010 Conference on Control and Fault Tolerant Systems. Final program and book of abstracts". IEEE Press. Institute of Electrical and Electronics Engineers (2010), p. 191-196.

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