LEAKAGE DETECTION AND LOCALIZATION METHOD BASED ON A HYBRID INVERSE/DIRECT MODELING APPROACH SUITABLE FOR HANDLING MULTIPLE-LEAK SCENARIOS

JORDI MESEGUER (1), JOSEP M. MIRATS-TUR (1) ,GABRIELA CEMBRANO (2) (1), EDUARD BONADA (1), VICENÇ PUIG (2) (3)

(1): CETaqua, Water Technology Center, Carr. Esplugues 75,Cornellà de Llobregat, 08940, Spain. Phone: +34-93-3124829, fax : +34 93 312 48 01 / e-mail: jmeseguer@cetaqua.com

(2): Institut de Robòtica i Informàtica Industrial(CSIC-UPC), Spain.

(3): Dpt, Enginyeria de Sistemes, Automàtica i Informàtica Industrial (UPC), Spain.

When an accurate hydraulic network model is available, direct modeling techniques are very straightforward and reliable for on-line leakage detection and localization applied to large class of water distribution networks. Nonetheless, the assumption of single-leak scenarios is usually considered and may not hold in real applications. This paper presents a leakage detection and localization method suitable for multiple-leak scenarios in a large class of water distribution networks. This method can be seen as an upgrade of a direct-modeling approach in which a global search method based on *Genetic Algorithms* (*GAs*) has been integrated in order to estimate the water loss hotspots and their size. This is an inverse / direct modeling method which seeks to take benefit from both approaches: the exploration capability of *GAs* and the straightforwardness and reliability offered by the availability of an accurate hydraulic model. The application of the resulting method in a district metered area of the Barcelona water distribution network is provided and discussed.

Keywords: Water Networks, Hydraulic Models, Leakage Localization, Genetic Algorithms.

1. INTRODUCTION

Continuous improvements on water loss management are being applied based on the use of new available technologies. Nonetheless, the whole leakage localization process may still require long periods of time (i.e. weeks, months) with an important volume of water wasted before the leak is found[1]. To avoid these inconveniences, leakage detection and localization based on mathematical models may be used [2] which can “compare” the data gathered by installed sensors in the network with the data obtained by a model of this network. The use of flow and pressure sensors together with hydraulic models of the water network for leak detection and localization is a suitable approach for the on-line monitoring of water balance [3][4][5].

[1] presents a straightforward direct modelling methodology for leakage detection and localization in district metered areas (DMAs[1]) of water distribution networks which is inspired by the binary model-based fault diagnosis theory [6] and takes benefit from those available DMA hydraulic models used by water operators. In [7], the method proposed in [1] was extended to work with non-binary fault signatures enhancing the overall performance of the method. Nonetheless, as inheritance coming from the standard model-based diagnosis theory [6], the assumption of single-leak scenarios is made and may not hold in real applications. As a consequence, when dealing with multiple-leak scenarios, inaccurate results might be obtained.

Considering the approach presented by [7], this paper highlights the existing trade-off when using a straightforward direct modeling technique assuming a single leak scenario and also presents an extension of this method suitable for multiple-leak scenarios where a global search method based on *GAs* has been integrated. The resulting method is a hybrid inverse / direct modeling method which tries to take benefit from both approaches: the exploration capability of *GAs* to estimate network water loss hotspots and the size of the leaks and the straightforwardness and reliability offered by the availability of an accurate hydraulic model. Regarding the application, a DMA of the Barcelona water distribution network has been used to illustrate the performance of both methods using simulated single and multiple-leak scenarios and a real multiple-leak scenario.

This paper is organized as follows: *Section 2* presents the case study used to illustrate the performance of the leakage detection and localization methods. *Section 3* recalls the direct modeling method proposed by [7] and describes the proposed extension to handle multiple-leak scenarios highlighting their performances in simulated single and multiple-leak scenarios. Then, in *Section 4*, the performance of both methods is described using a real multiple-leak scenario.

2. CASE STUDY DESCRIPTION

In this paper, *Nova Icaria* DMA of the Barcelona water distribution network) has been used to illustrate the performance of the assessed leakage detection and localization methods considering simulated single and multiple-leak scenarios and a real multiple-leak scenario. *Nova Icaria* DMA has two inlets (Alaba and Llull), *1996* nodes and *3442* pipes. In Figure 1, the water network of Nova Icaria DMA can be seen where the two DMA inlets have been highlighted using red triangle symbols. Regarding the instrumentation, the Nova Icaria DMA is provided by flow and pressure sensors at every inlet and by *6* inner pressure sensors (green star symbols) already deployed in the DMA.

3. MODEL-BASED LEAKAGE LOCALIZATION METHOD

3.1. Mathematical Modeling

The method proposed by [7] works with steady-state models concatenated in an *Extended Period Simulation* (EPS) [8] where the governing laws are determined by the conservation of mass / energy [2] and, a demand model is also considered. Thereby, the demand of node i is determined by the nodal base demand bd_i and the demand pattern $p_{a,i}$ estimated using the billing information. Then, leaks are assumed to be located in the nodes and simulated as an emitter coefficient C_j generating a leakage size depending on the pressure of that node ([9][10]):

---

[1] District Metered Area (DMA) is a defined area of the distribution system that can be isolated by valves and for which the quantities of water entering and leaving can be metered.

$$f_j = C_j p_j{}^{\gamma}, \tag{1}$$

where f_j is the leak size; C_j is the associated emitter coefficient; p_j is the pressure at node j; and γ is an exponent in the range of 0.5 (Hazen-Williams, Darcy-Weisbach, Chezy-Manning formulas). In this method, the DMA EPS hydraulic model is implemented in EPANET [11] updating the boundary conditions of the network at a given time instant k using the inflow and pressure measurements at the DMA inlets.

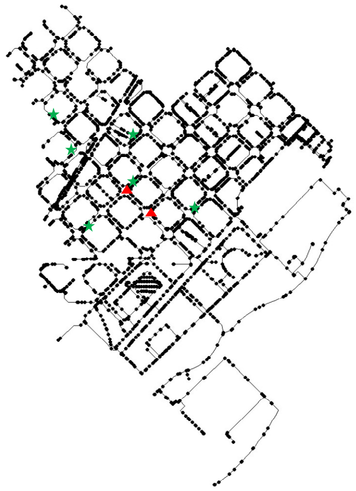

Figure 1 Water network of Nova Icaria DMA (EPANET model) highlighting inner pressure sensors (green stars) and DMA inlets (red triangles)

3.2. Direct modeling method assuming single-leak scenarios

3.2.1. Method description

The method proposed by [7] 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 which are obtained using the DMA hydraulic model. Thereby, the residual set, $\boldsymbol{r} \in \Re^{ns}$, is determined by the difference between the measured pressure at certain network nodes, $\boldsymbol{p} \in \Re^{ns}$, and the predicted pressure at these nodes considering a scenario free of leaks, $\hat{\boldsymbol{p}}_0 \in \Re^{ns}$:

$$\boldsymbol{r} = \boldsymbol{p} - \hat{\boldsymbol{p}}_{\boldsymbol{0}} = \begin{pmatrix} p_1 - \hat{p}_{10} & \cdots & p_{ns} - \hat{p}_{ns0} \end{pmatrix}^t \tag{2}$$

The size of the residual vector $\boldsymbol{r}$, ns, depends on the number of inner pressure sensors of the DMA network. Regarding the number of potential leaks, $\boldsymbol{f} \in \Re^{np}$, it is equal to the number of network nodes, np since from the modeling point of view, leaks are placed in these locations (single-leak scenario assumption). On the other hand, the theoretical pressure disturbances caused by all potential leaks are stored in the theoretical fault signature matrix, $\boldsymbol{FSM} \in \Re^{ns \times np}$ [6], with as many rows as DMA inner pressure sensors, ns, and as many columns as potential leaks (DMA network nodes), np. This matrix can be obtained from a sensitivity-to-leak analysis which evaluates the theoretical effect of all potential leaks f_j in the pressure of all the monitored nodes, p_i [7]:

$$\boldsymbol{FSM} = \begin{pmatrix} \dfrac{\hat{p}_{1f_1} - \hat{p}_{10}}{f} & \cdots & \dfrac{\hat{p}_{1f_{np}} - \hat{p}_{10}}{f} \\ \vdots & \ddots & \vdots \\ \dfrac{\hat{p}_{nsf_1} - \hat{p}_{ns0}}{f} & \cdots & \dfrac{\hat{p}_{nsf_{np}} - \hat{p}_{ns0}}{f} \end{pmatrix} \tag{3}$$

where $\hat{p}_{if_j}$ is the predicted pressure in the node where the pressure sensor i is placed when a nominal leak of size f is forced in node j and $\hat{p}_{i0}$ is the predicted pressure associated with the sensor i under a scenario free of leaks. Both the matrix ***FSM*** and the vector ***r*** depend on the demand and boundary conditions [12] and must be computed at every analysis time step. Regarding the leakage localization process at time instant k, this is based on a correlation process which compares the residual vector ***r**(k)* (Eq. (2)) with the theoretical signatures of all potential leaks (columns of matrix ***FSM**(k)*; Eq. (3)) applying the correlation function[2]. Those potential leaks whose theoretical signatures have the biggest correlation values with the residual vector ***r**(k)* point out the most probable nodes to have the leak.

$$\max_j \left(\rho_{\boldsymbol{r},\boldsymbol{FSM}_j}(k) \right) \quad , \quad j = 1,\ldots,np \tag{4}$$

where $\rho_{\boldsymbol{r},\boldsymbol{FSM}_j}$ is the obtained correlation between the residual, ***r**(k,)* and the j^{th}–column of the theoretical fault signature matrix, FSM_j, associated with a potential leak in node j.

3.2.2. Method performance under multiple-leak scenarios considering the Nova Icaria DMA

In this section, the performance of the direct modeling method (*Section 3.2.1*) is illustrated using a simulated single-leak scenario and an extension of the previous one in which an additional simulated leak is added (multiple-leak scenario). In this analysis, all the simulated leaks have been forced using an emitter coefficient *C=0.92* being *6.5 l/s* the average size of the resulting leak given that the average pressure in this DMA is around *50 m.w.c.* (Eq. (1)).

In Figure 2, the left plot shows the performance of this method in the simulated single-leak scenario highlighting the leak exact location (red cross), the predicted most correlated location (blue spot) and other predicted high-correlated locations (black spots)(> 98% of the highest correlation). In this scenario, this method predicts the exact location of the simulated leak. Nonetheless, a remarkable area of nodes presenting similar correlation values are also highlighted since their corresponding theoretical fault signatures are almost non-distinguishable in regard to the predicted most correlated location. On the other hand, the right plot in Figure 2 shows the performance of this method using the multiple-leak scenario mentioned above. In this case, the most correlated location predicted by the method (blue spot) misses the exact location of the simulated leaks. Nonetheless, the method highlights an area of nodes presenting similar correlation values which does contain the location of one of the two simulated leaks. This result is also valuable for the network operator to find the exact location of the leak (i.e. using acoustic loggers).

3.3. Hybrid inverse/direct modeling method handling multiple-leak scenarios

3.3.1. Method description

Quevedo et al., 2011 assumes single-leak scenarios constraining the number of potential leaks ($\boldsymbol{f} \in \Re^{np}$) to the number of network nodes (*np*) considering one different leak per node at the same time instant; consequently, the ***FSM*** number of columns (Eq. (3)) is also set to *np*. A straightforward procedure to extend this method to deal with multiple-leak scenarios may be increasing the number of accepted potential leakage scenarios considering also the existence of different nodes with leaks at the same time. As a result, the number of columns of ***FSM*** matrix also increases and consequently, so does its computation time. In order to overcome this drawback, an optimization process based on a global search method (i.e. *GAs*) could be integrated in order to select wisely those scenarios that best match the observed fault signature

---

[2] Pearson's correlation coefficient ρ_{x_i,x_j} between two variables $\boldsymbol{x}_i$, $\boldsymbol{x}_j$ may be defined as $\rho_{x_i,x_j} = \mathrm{cov}(\boldsymbol{x}_i,\boldsymbol{x}_j)/\sqrt{\mathrm{cov}(\boldsymbol{x}_i,\boldsymbol{x}_i)\mathrm{cov}(\boldsymbol{x}_j,\boldsymbol{x}_j)}$, where $\mathrm{cov}(\boldsymbol{a},\boldsymbol{b}) = E\left[(\boldsymbol{a}-\overline{a})(\boldsymbol{b}-\overline{b})\right]$ is the covariance function between two variables ***a*** and ***b***, being $\overline{a} = E(\boldsymbol{a})$ and $\overline{b} = E(\boldsymbol{b})$ respectively.

(Eq. (2)). According to [13], the resulting method can be seen as a calibration process (inverse modeling method) where the leak locations and size are considered as model parameters and have to be adjusted. [13] and [8] pointed out the goodness of *Evolutionary Algorithms* (i.e. *GAs*) to carry out the calibration of water network models. In [14] and [15], the procedure presented in [13] was extended to solve the leakage localization problem.

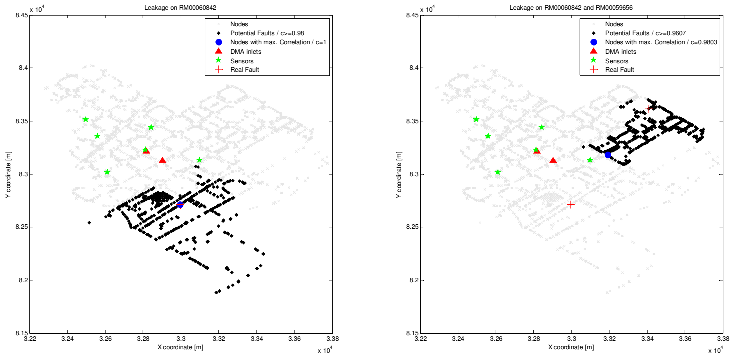

Figure 2 Performance of the direct modeling method using a simulated single-leak scenario (on the left) and a scenario where an extra simulated leak has been added (on the right). The leak exact locations (red cross), the most correlated location predicted by the method (blue spot) and other locations presenting also high correlations (black spots)(> 98% of the highest correlation) have been highlighted.

Considering the approaches presented by [15] and [7], the approach presented in this paper can be considered as an inverse / direct modeling method which tries to take benefit from both approaches: on the one hand, the exploration capability of GAs to estimate water loss hotspots and their size and on the other hand, the straightforwardness offered by the availability of an accurate network hydraulic model to assess the goodness of the potential solutions proposed by the *GAs*. In this way, the result of the proposed method is a estimation of the potential multiple-leakage scenario $\boldsymbol{f_v}^* \in \boldsymbol{F}$ that best matches the observed theoretical signature (Eq. (2)) where the set $\boldsymbol{F}$ is defined as follows:

$$\boldsymbol{F} = \left\{ \boldsymbol{f_v^*} \in (\mathbf{Z}^+ x\Re)^{nl} \,\middle|\, f_{o,v}^* = (Nd_{o,v}, C_{o,v}); o = 1,\cdots,nl; v = 1,\cdots,nF \right\} \tag{5}$$

nl being the number of leakage hotspots that may exist at the same time and nF the maximum number of considered multiple-leakage scenarios. Thereby, a given potential multiple-leakage scenario $\boldsymbol{f_v}^*$, is determined by a set of nl leaks, $f_{o,v}^* \in \boldsymbol{f_v}^*$, which are characterized by the tuple ($Nd_{o,v}$, $C_{o,v}$) where $Nd_{o,v}$ is the node index where the leak is located ($1 \le Nd_{o,v} \le np$) and $C_{o,v}$ is the emitter coefficient value which determines de size of that leak ($0 \le C_{o,v} \le C^{max}$). Then, according to Eq. (3), the *theoretical observed signature* related to $\boldsymbol{f_v}^*$ can be computed as follows:

$$\boldsymbol{FSM}_v^* = \left(\hat{p}_{1 f_v^*} - \hat{p}_{10} \quad \cdots \quad \hat{p}_{ns f_v^*} - \hat{p}_{ns0} \right)^t \tag{6}$$

Thus, the goodness of $\boldsymbol{f_v}^*$ is computed using the correlation function, $\rho_{\boldsymbol{r},\boldsymbol{FSM}_v^*}$ (*Section 3.2.1*) between the *observed fault signature*, $\boldsymbol{r}$ (Eq. (2)) and the associated theoretical fault signature, $\boldsymbol{FSM_v}^*$ (Eq. (6)). In general, the optimization problem solved by the *GA* can be written down as follows:

$$\max_{f_v^* \in \boldsymbol{F}} \left(\rho_{\boldsymbol{r},\boldsymbol{FSM}_v^*}(k) \right), \qquad v = 1,\cdots,nF \tag{7}$$

Regarding the number of considered potential multiple-leakage scenarios, nF, which determines the *GA* search-space, in general, this parameter is determined by $\binom{np}{nl}$ being huge for large-class of networks. Nonetheless, when considering that just a reduce number of *theoretical fault* *signatures* $\boldsymbol{FSM_j}$ (Eq. (3)) are distinguishable given the use of a constrained number of sensors [1], the search-space is also considerable reduced to $\begin{pmatrix} nf \\ nl \end{pmatrix}$ where nf is the number of distinguishable *theoretical fault signatures* $\boldsymbol{FSM_j}$ (Eq. (3)) ($n_f<n_p$ and $n_f \geq n_l$).

In Figure 3, the conceptual scheme of the procedure followed at time instant k by the resulting GA-based inverse / direct modeling method can be seen. In this scheme, those new steps required by the GA-based optimization process are highlighted in green while the others correspond to those steps associated with the direct modeling approach described in *Section 3.2.1*.

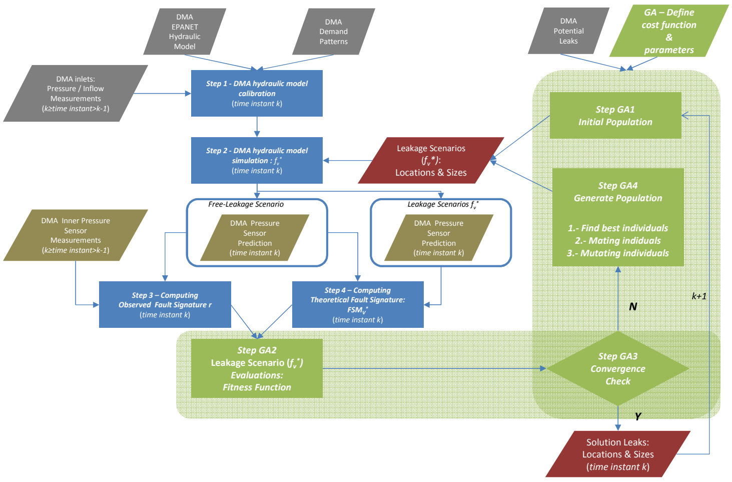

Figure 3 Conceptual scheme of the procedure followed by the leakage location *GA*-based inverse / direct modeling method at time instant *k* highlighting in green those news steps required for the *GA*-based optimization process.

3.3.2. Method performance under multiple-leak scenarios considering the Nova Icaria DMA

In this section, the performance of the method presented in *Section 3.3.1* is illustrated using the simulated multiple-leak scenario described in *Section 3.2.2*. In Figure 4, the obtained results are shown pointing out in blue spots the most correlated locations obtained in hourly runs during a whole day. The locations predicted by the method are close to the exact locations. The discrepancy between the predictions and the exact locations may be explained as follows: *GA* provides only (near) optimal solutions and as seen in Figure 2, there are nodes contained in a certain areas whose theoretical fault signatures are almost non-distinguishable.

4. REAL MULTIPLE-LEAK SCENARIOS

4.1. Leakage scenario description

In this section, the performance of the direct and hybrid modeling methods presented in (*Section 3.2.1* and *Section 3.3.*1) are illustrated using a real multiple-leak scenario affecting the Nova Icaria DMA . During the first half of 2011, an increase of the DMA minimum night flow in *4 l/s* was already noticed by DMA SCADA operators. However, since the beginning of September 2011, the DMA minimum night flow started a progressive worsening reaching an overall increase of 10 *l/s* in October 2011 when both leaks could be found.

4.2. Obtained results using the direct and hybrid modeling methods

In Figure 5, the left plot shows the performance of the direct modeling method using the above mentioned real multiple-leak scenario highlighting the leak exact locations (red cross), the most correlated location predicted by the methods (blue spot) and other locations presenting also high correlations (> 98% of the highest correlation). In this case, the leak location predicted by this method is reasonably close to the locations of the real leaks. Apart from the fact of assuming a single leak, the resulting prediction error is due to the discrepancies between the hydraulic model and the real performance of the DMA network. On the other hand, the right plot (Figure 5) illustrates the performance of the hybrid inverse/direct modeling method which can predict the location of the leaks with a reasonable error. Mainly, this error is due to the mismatch between the DMA hydraulic model and the real water network operation and the fact that the *GA*-based optimization process can only provide nearly optimal solutions.

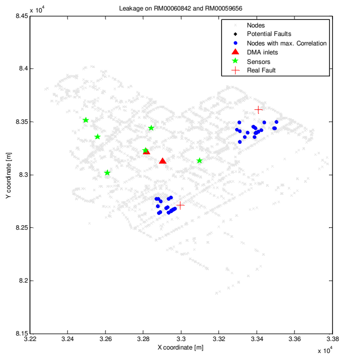

Figure 4 Performance of the hybrid inverse/direct modeling method using a multiple-leak scenario highlighting the leak exact location (red cross) and the most correlated locations predicted by the method (blue spot) obtained at every hourly run over a whole day.

5. CONCLUSIONS

This paper presents a hybrid inverse / direct modeling method suitable for multiple-leak scenarios. The resulting approach is an extension of a straightforward direct modeling method based on the use of available network hydraulic models but constrained to single-leak scenarios. This paper highlights the existing trade-off when using a straightforward direct modeling approach assuming a single-leak scenario regarding the results obtained with the hybrid inverse / direct modeling method capable to handle multiple-leak scenarios.

The Nova Icaria DMA of the Barcelona water distribution network has been used to illustrate the performance of both approaches using simulated single and multiple-leak scenarios and a real multiple-leak scenario. In general, when considering multiple-leak scenarios, the direct modeling method tends to indicate the location of one of the leaks with an acceptable error due to the single-leak assumption and to the modeling inaccuracies. On the other hand, the locations predicted by the hybrid method are satisfactorily close to the exact locations of the leaks and the discrepancy is mainly due to the modeling errors and to the fact that the *GA*-based optimization process can only provide nearly optimal solutions.

REFERENCES

The authors wish to thank the support received by the EFFINET grant FP7-ICT-2012-318556 of the European Commission.

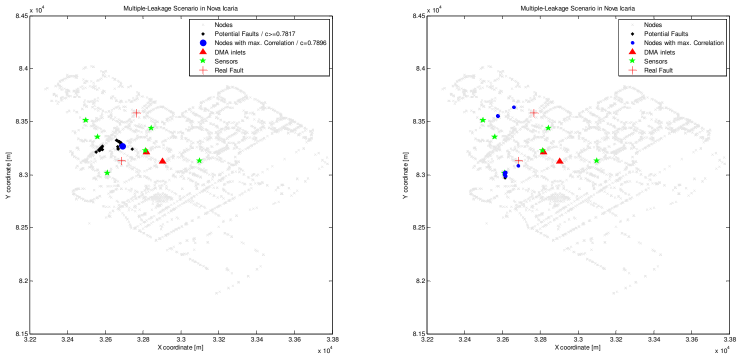

Figure 5 Performance of the direct modeling method (on the left) and the hybrid inverse/direct modeling method (on the right) using real multiple-leak scenario and highlighting the leak exact locations (red cross), the most correlated locations predicted by the methods (blue spot) and other locations presenting also high correlations (> 98% of the highest correlation).

REFERENCES

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

[2] Brdys, M. A., Ulanicki, B., (1994) State and parameter estimation, Operational Control of Water Systems: Structures, Algorithms and Applications. Prentice Hall International (UK) Ltd., pp. 96-131.

[3] Almandoz J., Arregui F., Cabrera E., Cobacho R. (2005). Leakage Assessment through Water Distribution Network Simulation. J. Water Resour. Plng. and Mgmt., Volume 131, Issue 6, pp. 458-466, November/December.

[4] Wu, Z. Y., Sage, P. (2006) Water Loss detection via genetic algorithm optimization-based model calibration. 8th Annual International Symposium on Water Distribution System Analysis, Cincinnati, Ohio, August 27-30.

[5] Bicik, J., Kapelan, Z., Makropoulos, C. and Savic, D. (2013) A Pipe Burst Diagnostics Using Evidence Theory, Journal of Hydroinformatics, pp 596-608, vol. 13(4)

[6] Gertler J.J. (1998) “Fault Detection and Diagnosis in Engineering Systems.” Marcel Dekker. 1998

[7] Quevedo, J., Cugueró, M., Pérez, R., Nejjari, F., Puig, V., Mirats, J.(2011) Leakage location in water distribution networks based on correlation measurement of pressure sensors. IWA Symposium on Systems Analysis and Integrated Assessment, San Sebastián.

[8] Savić, D. A., Kapelan, Z., and Jonkergouw, P. (2009). "Quo vadis water distribution model calibration?" Urban Water Journal, 6(1), 3-22.

[9] R. S. Pudar, and J. A. Ligget, “Leaks in pipe networks”, Journal of Hydraulic Engineering, vol. 118, no. 7, pp. 1031-1046, 1992.

[10] R. Pérez, V. Puig, J. Pascual, A. Peralta, E. Landeros, and Ll. Jordanas, “Pressure sensor distribution for leak detection in Barcelona water distribution network”, Water Science & Technology: Water Supply, vol. 9, no. 6, pp. 715-721, 2009.

[11] L. A. Rossman, EPANET 2 Users Manual. Water Supply and Water Resources Division, National Risk Management Research Laboratory, 2000. Available at http://www.epa.gov/nrmrl/wswrd/dw/epanet.html.

[12] Vento, J., Puig,V.(2009).Leak detection and isolation in pressurized water networks using interval LPV models. In Proceedings of the 7th workshop on advanced control and diagnosis (ACD’09). Zielona-Gora,Poland.

[13] Wu, Z. Y, Walski, T., Mankowski, R., Cook, J. Tryby, M. and Herrin G. (2002) “Calibrating Water Distribution Model Via Genetic Algorithms”, in Proceedings of the AWWA IMTech Conference, April 16-19, Kansas City, MI.

[14] Wu, Z. Y., Sage, P. (2006) Water Loss detection via genetic algorithm optimization-based model calibration. 8th Annual International Symposium on Water Distribution System Analysis, Cincinnati, Ohio, August 27-30.

[15] Wu, Z. Y. (2008) Innovative Optimization Model for Water Distribution Leakage Detection. Watertown, USA: Bentley Systems Inc.