8-1-2014

Practical-Oriented Pressure Sensor Placement For Model-Based Leakage Location In Water Distribution Networks

Eduard Bonada

Jordi Meseguer

Josep M. Mirats

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

PRACTICAL-ORIENTED PRESSURE SENSOR PLACEMENT FOR MODEL-BASED LEAKAGE LOCATION IN WATER DISTRIBUTION NETWORKS

EDUARD BONADA, JORDI MESEGUER, JOSEP M. MIRATS-TUR
CETaqua, Water Technology Centre, Carr. Esplugues 75,Cornellà de Llobregat, 08940, Spain.
Phone: +34 933 422 590 / Fax: +34 933 124 801 / e-mail: ebonada@cetaqua.com

In order to bring remarkable benefits to operation and management of water networks, the analysis of sensed data can be used to locate water leaks using of a model-based methodology. However, the number of sensors installed is usually limited because of budget constraints and hence a strategy for optimizing their number and placement is required. This optimization is tightly coupled to the performance of the real-time model-based leakage diagnosis operation and hence the former should consider the requirements of the latter: (1) high distinguishability among all potential leaks to be detected; and (2) strong robustness in front of model-reality mismatches and other uncertainties. This paper describes a model-based pressure sensor placement optimization technique that focuses on the previous aspects and addresses practicality issues that arise in a real deployment. The technique uses an optimization method based on Genetic Algorithms that, unlike most common approaches in literature, avoids using a binary reasoning process. This increases the information granularity resulting in an improvement of both the leak distinguishability and the method robustness. Moreover, the technique also addresses the practical concerns by deriving an enhanced cost function. Finally, the method is validated in a District Metered Area of the Barcelona water distribution network. Results indicate that a good enough detection accuracy can be achieved with a low number of optimally placed sensors.

1. INTRODUCTION

One of the main concerns of water utilities is to reduce the amount of lost water caused by leaks in the distribution infrastructure [1]. In order to address this issue, water utilities apply monitoring techniques that rely on measurements registered by sensors. A common approach is to analyse the data sensed in District Metered Areas (DMA) and locate the water leaks using a model-based methodology that compares the measured information with the simulated counterparts [2]. Even though the successful results, this type of procedures presents some limitations. First, the number of sensors installed is usually limited because of budget constraints. And second, the sensor devices need to be properly located in order to enhance the performance of the real-time leakage detection module. Therefore, a strategy that optimizes the number and placement of sensors is required. Note that, unlike the leakage diagnosis, the sensor placement optimization is an offline process as it does not need to be executed continuously in a production environment but only once before the real sensor devices are deployed.

Since the objective of the optimization is to enhance the performance of the model-based leakage diagnosis operation, the sensor placement must consider the requirements set to the leakage detection mechanism. In this line, the sensor deployment should provide:

1. ***High distinguishability*** among potential leaks. In general, these techniques are based on the fact that a leak at a certain location causes a characteristic effect across the entire network, which is actually measured by sensors. In consequence, if the effects of different leaks are similar (i.e. different leaks result in same sensor measurements), these cannot be isolated.
2. ***Strong robustness*** in front of model-reality divergences and other uncertainties. In real applications there is always a mismatch between models and reality, sensor measurements may be affected by disturbances, or there may be unknown system inputs.

The model-based leakage location method is based on comparing data gathered by sensors with data simulated using a hydraulic model of the DMA [3]. The use of flow and pressure sensors together with hydraulic models is a suitable approach for the leakage localization problem as described in [4][5]. In particular, as proposed by [6][7], the real data is matched to simulated data of all possible leak scenarios, and the largest similarity identifies the scenario that really occurred. Under this assumption, the method presented in [8] proposes a binary approach to solve the sensor placement optimization problem. In summary, this binary method assumes that a potential sensor either detects a leak or not (hence binary); the optimization algorithm then selects those sensors that minimize the designed cost criteria.

This paper extends the previous binary methodology by introducing changes that (1) focus on the two functional requirements listed above and (2) address practicality issues that arise in a real deployment. First, removing the binary reasoning process allows us to increase the information granularity which results into a higher leak distinguishability. Second, we derive an enhanced optimizer's cost function that aims at enhancing both distinguishability and robustness. Third, we apply a pre-processing of the hydraulic modelled data with the objective of increasing the robustness against modelling errors. And finally, since real sensors have limited resolution, those pressure disturbances lower than this granularity cannot be measured accurately. This aspect shifts the solution closer to reality and, in turn, addresses the robustness requirement.

The method implementation is based on a Genetic Algorithm (GA) and we test it in a District Metered Area (DMA) of the Barcelona distribution network. We present the performance of the optimized locations for different number of sensors in order to understand the behaviour of the placements and decide the minimum number of sensors required.

The remainder of this paper is organized as follows. Section 2 briefly describes the model-based leakage location as well as the binary-based sensor placement optimization. Afterwards, sections 3 and 4 present our proposed method and the analysis results of applying it to the Barcelona DMA. Finally, section 5 conclude the paper recalling the lessons learnt as well as open issues to better understand the problem and improve the technique.

2. BACKGROUND

2.1 Model-based leakage location

The model-based leakage location is based on the standard theory of fault diagnosis [9] and it has been already applied to water distribution network. We base our work on the methods described by [6][7]. The fundamentals behind this methodology are to assume that a leak can be detected by monitoring pressure disturbances at certain inner nodes of the DMA network. These methods simulate all possible leak scenarios and match the real measurements with the simulated equivalents. The scenario that provides closer results to the real values determines the location of the leak.

In fact, the compared values are the pressure variations in the form of residuals. The residual $\boldsymbol{r}$ is determined by the difference between the measured pressure at certain inner network nodes p_i and the estimated pressure at these nodes obtained using the network model considering a scenario free of leaks $\hat{p}_{i,0}$. Note that ns, the length of $\boldsymbol{r}$, is determined by the number of inner points where the measurements are taken (sensors).

$$\boldsymbol{r} = \boldsymbol{p} - \hat{\boldsymbol{p}}_{\boldsymbol{0}} = \begin{pmatrix} p_1 - \hat{p}_{1,0} \\ \vdots \\ p_{ns} - \hat{p}_{ns,0} \end{pmatrix} \qquad \boldsymbol{FSM} = \begin{pmatrix} \hat{\boldsymbol{r}}_{\boldsymbol{1}} & \hat{\boldsymbol{r}}_{\boldsymbol{2}} & \cdots & \hat{\boldsymbol{r}}_{\boldsymbol{nl}} \end{pmatrix} = \begin{pmatrix} \hat{p}_{1,1} - \hat{p}_{1,0} & \cdots & \hat{p}_{1,nl} - \hat{p}_{1,0} \\ \vdots & \ddots & \vdots \\ \hat{p}_{ns,1} - \hat{p}_{ns,0} & \cdots & \hat{p}_{ns,nl} - \hat{p}_{ns,0} \end{pmatrix} \quad (1)$$

The simulated pressure variations caused by all potential leaks are stored in the *Fault Signature Matrix* (*FSM*), with as many rows as sensors, ns, and as many columns as potential leaks, nl. For example, in a network with 3 sensors and 1000 possible leak locations, the matrix has a size of 3×1000. Note that in this case all residuals $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ are simulated and represent the values for a scenario with the leak j. More concretely, $\hat{p}_{i,j}$ is the predicted pressure in sensor i in the scenario with the leak j; similarly, $\hat{p}_{i,0}$ is the predicted pressure associated with the sensor i under a scenario free of leaks.

Regarding the leakage location process, the methods compare the columns of the simulated ***FSM*** (theoretical signature of a given leak) with the residuals of real measurement $\boldsymbol{r}$ (observed signature of the leak) and the column that better matches the measurements indicates the leak detected. Different proposals use different comparisons methodologies: binary comparison [6][10] or Pearson correlation [7][11]. Refer to the survey in [12] for a detailed analysis on different techniques.

2.2 Binary Sensor Placement Optimization

In consonance with the leakage location methodology, the technique described in [6] proposes a binary approach to the sensor placement optimization problem. It initially considers an extended ***FSM*** with as many rows as possible sensors locations: a matrix of size $ns_{all} \times nl$ with ns_{all} sensors and nl leaks. From now on we will refer to this extended version as $\boldsymbol{FSM_{ext}}$.

The objective of the algorithm is to select ns of the ns_{all} rows. Note that the ***FSM*** used by the leakage location described previously has only ns rows because it is the number of sensors that are really taking measures. The optimization is actually about selecting the best possible ns rows. From now, we will refer to this matrix with only the ns sensor rows as $\boldsymbol{FSM_{sens}}$.

In order to choose the rows, each value of the $\boldsymbol{FSM_{ext}}$ matrix is binarized according to a threshold: '1' indicates that the sensor in row i truly detects a leak in column j ('0' indicates no detection). Once the $\boldsymbol{FSM_{ext}}$ is binarized, the optimizer looks for the best combination of rows, hence creating $\mathbf{FSM_{sens}}$, that minimizes the number of identical columns.

3. ENHANCED SENSOR PLACEMENT

This section presents the enhancements that we introduce to the binary sensor placement optimization described above. Figure 1 shows a diagram of the high level execution workflow including (1) the initial phase of hydraulic modelling using EPANET [13], (2) a set of pre-processing actions applied in the ***FSM*** calculation, (3) the optimizer based on a Genetic

Algorithm, and (4) the resulting sensor placement. Note that the grey blocks are those where we have applied our changes in comparison with the binary approach.

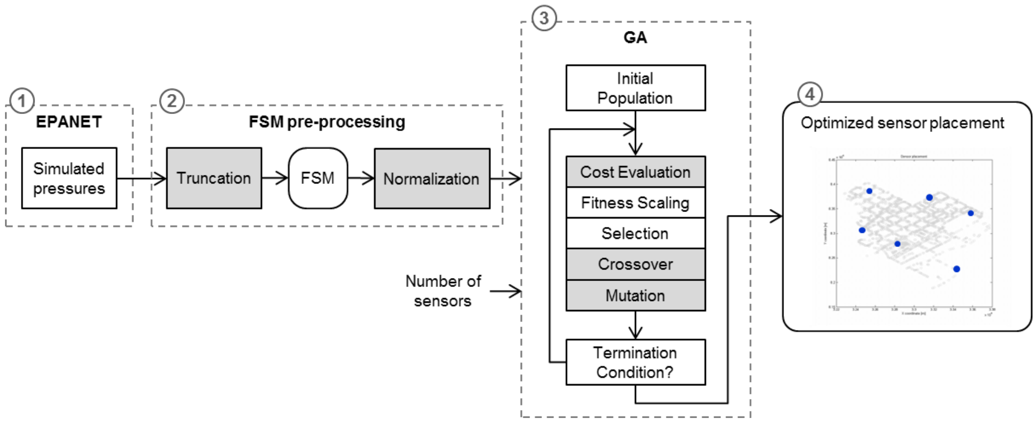


Figure 1 Execution workflow of the sensor placement optimization

3.1 Pressure Measurements Truncation

Our approach is still based on using the $\boldsymbol{FSM_{ext}}$ with the only difference that the simulated pressure values are truncated to the resolution of the real sensor devices (i.e. *0.1m* for the pilot in Barcelona network). Since the hydraulic modelling simulator (EPANET) is based on numerical analysis, the pressure values that it provides have very large precision. This represents a problem as we cannot assume to have such precision in the real networks two pressure values might turn out to be different with the entire precision but equal in the truncated version. With this data pre-processing we not only introduce a more realistic and practical sense, but also increase the robustness of the method because possible small inaccuracies in the hydraulic model are removed by the truncation.

3.2 Leak-wise Normalization

The original binary method normalizes the $\boldsymbol{FSM_{ext}}$ by rows. This action results in considering all the possible sensor locations with the same sensibility in front of a leak. That is, in a leak scenario one node that experiences a high pressure variation is considered equal as another node that just observes a small pressure variation. This is a shortcoming regarding the practicality of the method because one actually wants to take advantage of those sensor locations that are more sensible to leaks. In contrast, we apply a pre-processing consisting of a column-wise normalization with a two-fold objective. First, unlike the binarized method, we keep the information about the different sensibilities so the optimization engine can take advantage of it. And second, we increase the method robustness by reducing the dependence on the leak size.

3.3 Non-binary approach

Working with the binarized matrix implicitly involves losing of information. Note that applying a threshold to the original pressure variations causes these values to be shifted to their extreme representations (‘0’ or ‘1’). In addition, deciding the threshold level at which the 0/1 distinction is done introduces another variable that needs to be analysed as it might be different for different scenarios. Avoiding this binarization step mainly allows us to keep the pressure variations information so it can be used in the optimization phase.

The advantage of the non-binary method is actually the high granularity of the pressure variation values of the ***FSM*** matrices. Recall that the objective of the sensor placement optimization is to select those sensors (or rows) that make the columns as much distinguishable as possible. Therefore, the larger the granularity, the higher the distinguishability.

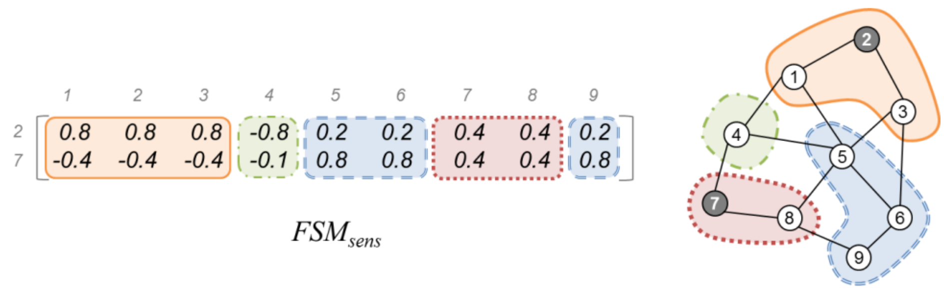


Figure 2 Example of an FSM_{sens} with equal columns (or signatures) and the corresponding groups of nodes created (sensors in nodes *2* and *7*)

3.4 Maximizing Number of Groups

Since the number of sensors is limited, the $\boldsymbol{FSM_{sens}}$ matrix has many more columns than rows. This asymmetry implies that some of the columns end up being identical, which translates into several leaks with the same signature (identical column) and hence not distinguishable. Figure 2 shows an example $\boldsymbol{FSM_{sens}}$ with the not isolable groups of nodes when the sensors are located in the nodes *2* and *7*. Note that two columns with the same signature means that the sensors measure the same pressure variations in the corresponding two leaks (columns of same colour in the matrix).

Since each matrix column represents a different node where a leak might be located, the identical signatures can be easily represented in the network topology as also shown in Figure 2. Note how the coloured columns match the colour circled nodes. In this line, we define a *group* as a set of columns that have the same signature, which correspond to a set of leak locations that cannot be isolated because we cannot distinguish them. For example, if a leak appears in node 5, since it is contained in the group *5-6-9*, the leak location method would tell that the leak has occurred in any of the 3 nodes. Thus, note that big groups are not desired because the distinguishability gets reduced.

In the proposed method, the optimizer selects those sensor rows such that the number of groups with distinguishable signatures are maximized, which in turn minimizes their size. However, using the total number of groups as the indicator to maximize might confuse the optimizer if it is not carefully treated. The problem is that this indicator is an absolute counter of all the groups including the few largest and the many small ones. It is actually the very large amount of small groups that distorts the counter because it introduces an important variability in this absolute number of groups.

We approach this problem in two ways. First, we account for only the largest groups that span a certain percentage of nodes. The reason is that the size of the groups follows the Pareto principle (or 80-20 rule) because approximately 20% of the groups span more than 80% of the nodes (we actually consider a percentage of 90% of nodes because tests indicate better results). Pseudo-code in Figure 3 details the cost function implementation. As detailed in lines *2-3*, the groups are sorted from largest to smallest to afterwards select only those largest whose summated size spans the 90% of nodes. And second, we consider the indicator in a logarithmic scale so large differences in number of groups (because of the small ones) do not heavily affect the indicator. Observe in line *4* how the logarithmic scale is applied to the *numGroups* normalized between its maximum *numNodes* and its minimum *1*.

After studying different indicators (i.e. size of largest group, average geographical extension, etc…) we have concluded that (1) optimizing the number of groups provides the best results and (2) it is also the indicator that better represents other indicators. Refer to [14] for further details on this analysis.

```
cost = costFunction(bitstring sensors, matrix  FSM_ext , integer numNodes)
   1.  Set FSM_sens to rows of FSM_ext according to sensors
   2.  Set sortedAllGroups  to columns of identical signatures (and sort from large to small)
   3.  Set numGroups to first N groups in sortedAllGroups with sum larger than 0.9xnumNodes
   4.  Set cost to log( (numGroups – numNodes)/(1 – numNodes) )
```

Figure 3 Pseudo-code of the cost function that maximizes the number of groups

4. CASE STUDY

This section presents the analysis of applying the described approach in the *Nova Icària* DMA of the Barcelona Water Distribution Network (3381 Nodes, 3457 Pipes, 2 input points). EPANET is the hydraulic modelling software used to simulate the scenarios. The considered leak size is of *6 l/s* and any network node is a potential leak or sensor location (*ns=nl=3381*).

For the optimizer implementation we have used the Matlab GA Toolbox with customized evolutionary functions to improve efficiency and the following parameters: *population size*: 1000, *elite count*: 2; *crossover fraction*: 0.8, *termination condition*: 10 generations with a cost change lower than 1e-6.

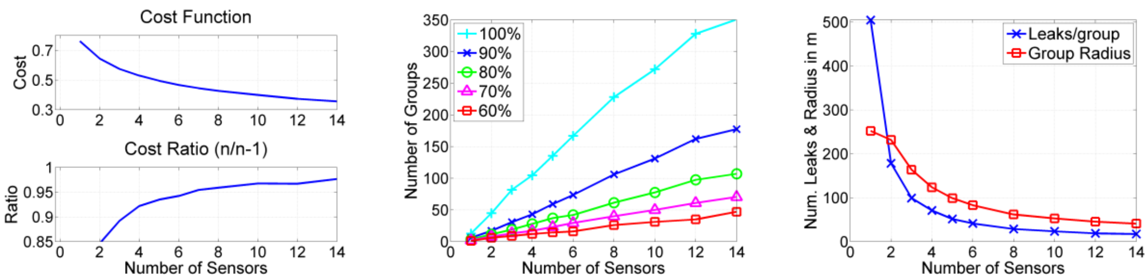
4.1 Performance of optimized placements

Plots in Figure 4 show the performance of the placements proposed by the optimization. All x-axes represent the number of sensors and the y-axes represent a different performance indicator in each plot. The lines correspond to the average value of 10 independent executions.

The top plot in Figure 4(a) contains the cost value of the best sensor placement found by the optimizer at each scenario. We first observe that the cost decreases as we consider more sensors. This makes sense because as more sensors are added, a higher distinguishability can be achieved and more groups are actually created. The question is, however, to determine the point at which adding one more sensor does not increase the performance sufficiently. In order to simplify the decision, the bottom plot in Figure 4(a) depicts the ratio between consecutive cost values ($cost_n / cost_{n-1}$). Note that the ratio reaches values closer to 1 as the curve gets flatter. We arbitrarily adopt that a change in the cost value lower than 5% is enough to decide that an additional sensor will not introduce a significant performance. In the ratio plot, this point can be detected when the line crosses 0.95, which takes between 6 and 7 sensors. Hence, 6 is the proposed number of sensors.

The plot in Figure 4(b) shows the number of groups created. Each line corresponds to a different percentage of nodes in a way that, for example, the 80% line indicates the number of groups that together contain 80% of the nodes. In consequence, the 100% line indicates the total number of groups obtained from each scenario. The important observation to do is that the difference between the 100% and 90% lines denotes that there are lots of small groups (e.g. for 10 sensors, approximately 150 groups contain 90% of the nodes, while another 150 represent the remaining 10%). This confirms that configuring the optimizer to maximize the number of groups with 90% of the nodes provides better results as it is the percentage that better discards the smallest groups but still considering a large amount of nodes.

The plot in Figure 4(c) shows the performance of the sensor placements in terms of average group size considering (1) the number of leaks per group (line with squares), and (2) the group radiuses in meters (line with crosses). Note that both lines are measured by the same left-hand side vertical axis. In both cases we notice the already observed decreasing behaviour with a trade-off point also around 6 sensors.

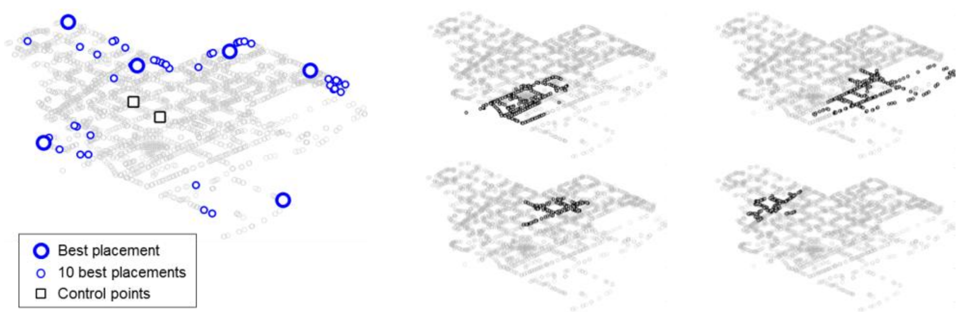


(a) Cost

(b) Number of groups

(c) Average Group Size

Figure 4 Performance of placements proposed by the optimizer for different number of sensors



(a) Sensors placements

(b) Groups created

Figure 5 Geographical location of the placement and extension of the largest groups created

4.2 Proposed sensor placement

The maps depicted in Figure 5 show the DMA under study and different geographical information of the optimized sensor placements. The round blue circles in Figure 5(a) indicate the location of the sensors resulting of 10 executions. The placement that obtained the best cost value (cost=0.464; 77 groups for 90% of nodes; 180 total groups) is depicted with larger circles. From a global perspective, observe that there are some DMA areas where sensors tend to be located. These regions correspond to the edges of the network because these locations are sensible to more leaks (note the water flows from the central control points towards the edges), and hence they allow creating more groups of smaller size.

The maps included in Figure 5(b) show the largest groups, in number of leaks, resulting from the best sensor placement. While the first two large groups appear because of the low sensitivity region, the rest of groups at most measure a few 100s of meters. This represents an important improvement for a water operator that has no sensors installed a leak has to be searched in the entire DMA with a diameter of 1.4km.

5. CONCLUSION

This paper has presented a new method for sensor placement optimization in water distribution networks that works together with a model-based leakage isolation process. The technique is based on the binary approach in [8] and this paper describes the enhancements applied in order to improve leakage distinguishability and method robustness as well as to introduce a more practical perspective.

The technique is validated in a DMA of the Barcelona water distribution network. The analysis indicates that the method provides leakage isolation with enough accuracy to help the water operator. Results show that a few sensors (6 in the studied DMA) are sufficient to achieve leak detection areas of a few hundreds of meters at most. This a significant improvement for the operator compared to searching leaks in an entire DMA with a diameter of 1.4km.

The study has also allowed us to detect open issues and potential improvements. These include: testing the sensor placement in different leak scenarios; designing a more advanced FSM pre-processing in order to reduce the problem dimension; introducing sensor sensitivity data in the cost calculation; or enhancing the method robustness by considering a time horizon in the optimization.

6. ACKNOWLEDGEMENTS

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

REFERENCES

[1] M. Farley and S. Trow, eds. “Losses in water distribution networks: a practitioner's guide to assessment, monitoring and control”, *IWA Publishing*, (2003).

[2] J. Meseguer et al., “Model-driven software tool for on-line leakage localization”, *IWA WaterLoss Conference*, (2014).

[3] M. A. Brdys and B. Ulanicki, “State and parameter estimation, Operational Control of Water Systems”, *Prentice Hall International*, (1994).

[4] J. Almandoz et al., “Leakage Assessment through Water Distribution Network Simulation”, *Journal of Water Resources Planning and Management*, vol. 131-6, (November 2005).

[5] J. Bicik et al., “A Pipe Burst Diagnostics Using Evidence Theory”, *Journal of Hydroinformatics*, vol. 13-4, (2013).

[6] R. Pérez, et al. "Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks", *Control Engineering Practice*, 19.10, 2011.

[7] J. Quevedo, et al. "Leakage location in water distribution networks based on correlation measurement of pressure sensors", *8th IWA Symp. on Systems Analysis and Integrated Assessment*, (2011).

[8] R. Perez et al. "Pressure sensor distribution for leak detection in Barcelona water distribution network", *Water Science & Technology: Water Supply*, 9.6 (2009).

[9] J. Gertler, “Fault detection and diagnosis in engineering systems”, *CRC press*, (1998).

[10] B. Farley et al. "Development and Field Validation of a Burst Localization Methodology." *Journal of Water Resources Planning and Management*, (November 2013).

[11] J. Meseguer et al. “Leakage detection and localization method based on a hybrid inverse/direct modeling approach suitable for handling multiple-leak scenarios”, Submitted for review to *11th International Conference on Hydroinformatics*, (2014)

[12] M. V. Casillas et al, "Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities", *Journal of Hydroinformatics*, (November 2013).

[13] L. A. Rossman, “EPANET 2”, (2000), Available at http://www.epa.gov/nrmrl/wswrd/dw/epanet.html

[14] E. Bonada et al. “On the structure of the objective function for a pressure sensor placement optimization applied to model-based leakage localization in distribution water networks”, Submitted for review to *11th International Conference on Hydroinformatics*, (2014).