International Conference on Hydroinformatics

2014

On The Structure Of The Objective Function For A Pressure Sensor Placement Optimizing Methodology Based On Genetic Algorithms Applied To Model-Based Leakage Localization In Distribution Water Networks

Eduard Bonada

Jordi Meseguer

Josep M. Mirats-Tur

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

ON THE STRUCTURE OF THE OBJECTIVE FUNCTION FOR A PRESSURE SENSOR PLACEMENT OPTIMIZATION APPLIED TO MODEL-BASED LEAKAGE LOCALIZATION IN DISTRIBUTION WATER NETWORKS

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

Real-time monitoring of distribution water networks relies on the deployment of sensors and the availability of their measurements in order to predict the system state and assess its performance. A meaningful application of this methodology is the detection and localization of leaks using model-based approaches. Since the number of sensors is limited because of budget constraints, it is important to place these devices in locations where the effectiveness of the leakage diagnosis is maximized. Finding the best sensor distribution is a global optimization problem defined by an objective function that might depend on different factors. Therefore, deriving the correct structure of such function is a crucial step as a wrong definition would lead towards a confusing optimal solution affecting negatively the monitoring performance. In general, sensor placement optimization methods describe objective functions using factors related to the amount of undistinguishable leaks. More concretely, the methods first compute groups of locations where leaks cannot be differentiated and then maximize this number of groups or minimize their size. In this paper, additional factors are presented to accurately represent the requirements of the leak diagnosis phase. These include other statistical figures related to the size of groups, geographical characteristics like the group's extension area, levels of sensitivity that indicate whether a location is more or less sensible to pressure changes, etc. The objective of this study is to review several factors in order to comprehend their behaviour and justify or discard them for the objective function. The indicators under study are evaluated by means of a cross-correlation analysis applied to the scenario defined by the District Metered Area of the Barcelona water distribution. Results indicate the existence of different independency levels between the indicators that allow us to select those with less redundancy.

1. INTRODUCTION

Water utilities are concerned about reducing the amount of lost water caused by leaks in the distribution infrastructure [1]. This is why they 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.

Finding the best sensor distribution is a global optimization problem defined by an *objective function* that guides the optimizer towards the best solution. Its goal is to search for solutions, evaluate them using the objective function, and select the one that provides the best objective value. Note that depending on the notation used, the objective function is sometimes called *fitness function* (usually if the objective is maximized) or *cost function* (if it is minimized). From now on we will use the term cost function to refer to this concept.

The cost function is represented by a computation of a cost value that is optimized (maximized or minimized). This computation can range from very simple calculations including only one performance variable, to more complex cost evaluations considering several factors that might be weighted and merged into a single cost value. In any case, the factors that compose the cost function must accurately represent the objectives of the optimization. Therefore, in the leakage location frame, deriving the correct structure of the cost function is a crucial step as a wrong definition would lead towards a confusing optimal solution affecting negatively the monitoring performance.

Identifying the indicators to consider in the cost function of the sensor placement optimization is the objective of this paper. One way to achieve such objective is to follow a top-down approach and manually select the indicators that are considered to better represent the performance of a sensor placement. Since this approach is based on the expert's knowledge, it guarantees a cost function design that matches the real network behaviour. However, it has some inconveniences in the sense that it is not able to detect hidden dependencies between indicators or to discover not expected ways of merging them. In order to overcome such drawbacks, this study presents a bottom-up approach based on a computer-assisted performance evaluation. Several indicators are extensively studied with the objective of (1) understanding behaviours and dependencies, and (2) selecting the most representative one(s) for the cost function. Note that this approach is not substituting the former but complementing it. First, the expert's opinion is still needed to select a set of indicators to study. And second, the bottom-up analysis can be used to validate the top-down approach.

The indicators under study are evaluated by means of a cross-correlation analysis applied to the scenario defined by the District Metered Area of the Barcelona water distribution. The correlation values allow us to detect redundancies between pairs of indicators, which in turn helps devising those that are independent and hence eligible for the cost function design. The study shows that there exist several clusters of indicators that point out a certain level of redundancy between them. This fact leads to potential cost function structures that considers only independent indicators (e.g. one from each cluster).

The remaining of this paper is organized as follows. Section 2 presents the context where the study is applied and briefly describes the fundamentals of sensor placement optimization for leakage location. Afterwards, section 3 includes a description of the indicators considered in the study. Accordingly, section 4 contains the correlation analysis of these. Finally, section 5 concludes the paper and gives some insights on how to use the selected indicators in a complex cost function.

$$
\begin{array}{c|ccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
2 & 0.8 & 0.8 & 0.8 & -0.8 & 0.2 & 0.2 & 0.4 & 0.4 & 0.2 \\
7 & -0.4 & -0.4 & -0.4 & -0.1 & 0.8 & 0.8 & 0.4 & 0.4 & 0.8
\end{array}
$$

FSM_{sens}

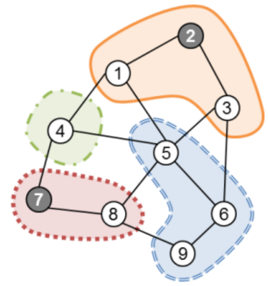


Figure 1 Example of an FSM_{sens} with equal columns (or signatures) and the corresponding groups of nodes created (sensors in nodes *2* and *7*)

2. BACKGROUND

The cost function determines how the optimizer evaluates the potential solutions therefore it must be designed considering the requirements set to the sensor placement. In turn, the optimized placement aims at enhancing the performance of the model-based leakage diagnosis operation. Therefore, and ultimately, the cost function must consider the requirements set to the leakage location mechanism. In this line, the cost function should provide *high distinguishability* among potential leaks, and s*trong robustness* in front of model-reality divergences and other uncertainties.

2.1 Optimal Sensor Placement for Leakage Localization

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 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. 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.

More concretely, the methods compare the columns of the simulated ***FSM*** (theoretical signature of a given leak) with the residuals of real measurement ***r*** (observed signature of the leak) and the column that better matches the measurements indicates the leak detected. Figure 1 shows an example of this matrix where the columns indicate different leaks that can occur, and rows indicate the measurements in the sensors *2* and *7* when each particular leak occurs. For instance, if the measurements in the sensors are *[0.2, 0.8]*, a leak is located in either 5, or *6* or *9*. 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 the graph of Figure 1. 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. In the previous example where 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. Different proposals use different vector comparisons methodologies: binary comparison [6][8] or Pearson correlation [7][9]. Refer to the survey in [10] for a detailed analysis on different techniques.

Under this assumption, the method described in [11] 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. An improvement of this method is presented in [12] and it is based on an additional pre-processing of pressure data in order to increase robustness, and on applying a non-binary approach so as to improve the leak distinguishability. Note that the cost function used in the non-binary approach is a direct result of the study presented in the current text.

2.2 Cost Function Indicators in Literature

Several cost function configurations have been proposed in literature. First, there is a set that relate to the sizes of groups that get created: [11] takes the simplest approach and minimizes the size of the largest group (size in number of leaks/group); [8] considers a preferred number of groups with the corresponding averaged size (in leaks/group) and minimizes how different the created groups are from the evenly distributed equivalents; finally, **¡Error! No se encuentra el origen de la referencia.** focuses on the geographical characteristics and minimizes the average distance between each leak and the group where it is detected.

An alternative approach is to skip the creation of groups and design the cost calculation directly considering the elements in the FSM matrix. In this line, **¡Error! No se encuentra el origen de la referencia.** maximizes a cost function with two indicators: the sum of sensitivities (FSM values) and the difference between each pair of columns. Similarly, **¡Error! No se encuentra el origen de la referencia.** considers both leak and contaminant detection in a multiple-objective optimization that maximizes, again, the sum of sensitivities and the contaminant detection likelihood.

Even though the diversity of cost functions already present in literature, we are not aware of any study that evaluates and compares the different indicators. Hence, with this paper we aim at filling this gap in order to provide a framework that can be used (1) to justify the use of the different factors and (2) by any optimization mechanism.

Table 1 List and brief description of the indicators under study

Topological (number of leaks)	
All groups	
NG	Number of groups created
AvGS	Average group size measured in leaks/group
SdGS	Standard deviation of group size measured in leaks/group
MxGS	Maximum group size
NotDet	Number of not detected leaks (all-zero FSM columns)
Largest groups (largest groups that span 90% of nodes)	
NG90	Number of groups created
AvGS90	Average group size measured in leaks/group
SdGS90	Standard deviation of group size measured in leaks/group

Topographical (geographical coordinates)	
All groups	
$AvGS_{geo}$	Average group radius measured in m
$SdGS_{geo}$	Standard deviation of group radius measured in m
$MxGS_{geo}$	Largest group radius
Largest groups (largest groups that span 90% of nodes)	
$AvGS90_{geo}$	Average group radius measured in m
$SdGS90_{geo}$	Standard deviation of group radius measured in m

Sensibility (FSM values)	
SumSens	Sum of all different FSM columns
DiffSV	Average difference between each pair of rows (sensors)
CorrSV	Average correlation between each pair of rows (sensors)

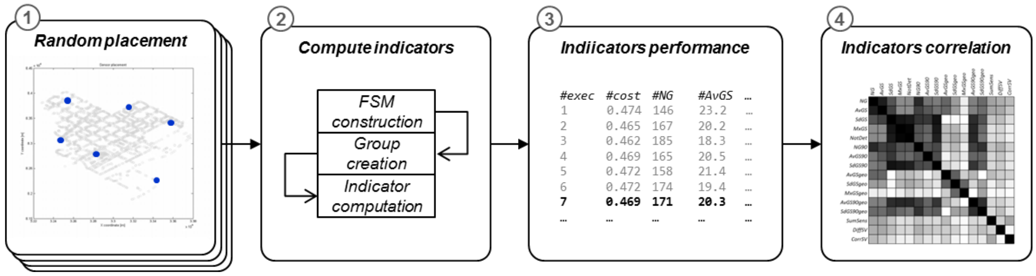


Figure 2 Execution flow steps of the indicators' cross-correlation analysis

3. DESCRIPTION OF INDICATORS

One of the objectives of this paper is to understand the behaviour of the different indicators under study (including those listed in section 2.2 and others proposed in the current section). Table 1 lists and briefly describes the analysed indicators. They are grouped according to their nature:

- The **topological** indicators include the performance measures that relate to the graph-based information. These include aspects such as the number of groups or different statistical figures regarding their size (measured in leaks per group). Note that a leak location corresponds to a network node, hence the graph concept.
- The **topographical** indicators are those that measure the geographical characteristics of the groups: different statistical figures are computed using the geographical coordinates of each node (or leak location). Note that in this case, unlike in the topological equivalent, the group size is measured as the radius in meters.
- The **sensibility** indicators measure the awareness of the sensors in front of different leaks. These are obtained directly calculating using with the values of the FSM (sum of columns, correlation between rows, etc)

Also note the distinction between the indicators computed considering *all* groups and those that only include the *largest* ones. The reason behind this separation is that the groups that get created are very different in size (a few large, many small ones) and we want to understand the effect of the small ones in the performance counters. In this line, we distinguish the indicators that consider all groups and those that account only for largest groups. These largest groups are defined as those that span a certain percentage of nodes: 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.

4. CORRELATION BETWEEN INDICATORS

This section includes the results of applying the indicators analysis to the *Nova Icària* DMA of the Barcelona Water Distribution Network – shown in Figure 3 – with 3381 Nodes, 3457 Pipes and 2 input points. In this paper we only include the results for the case of 6 sensors deployments but additional tests with different sensors and different DMAs indicate similar behaviours.

The performance evaluation methodology is described following. We first generate a random sensor placement (*step 1* in Figure 2) and evaluate it computing all the indicators under study (*step 2*). Note that, as described in section 2, this computation implies the calculation of the FSM matrix and the creation of the groups. The result of this computation is stored in a row

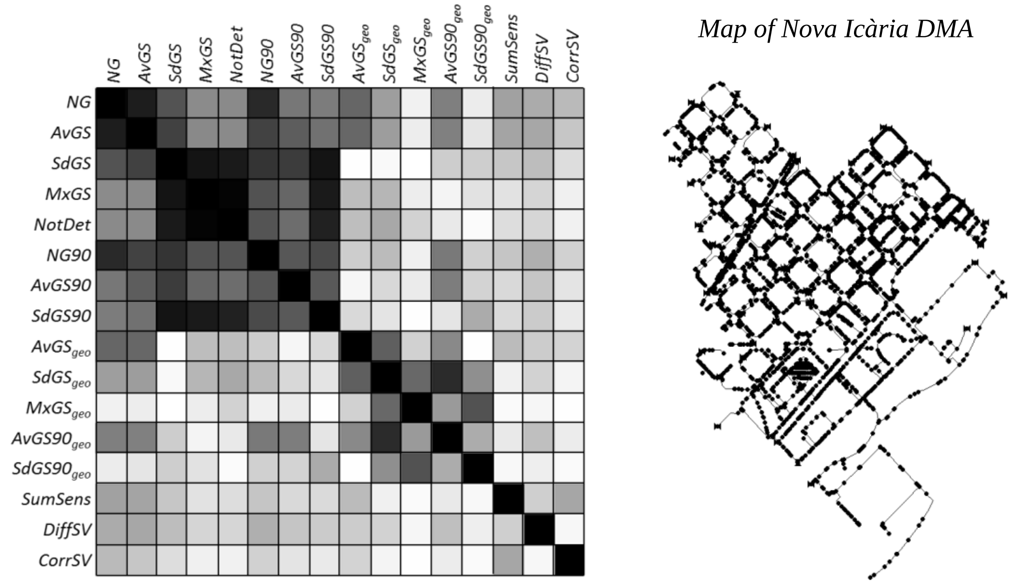
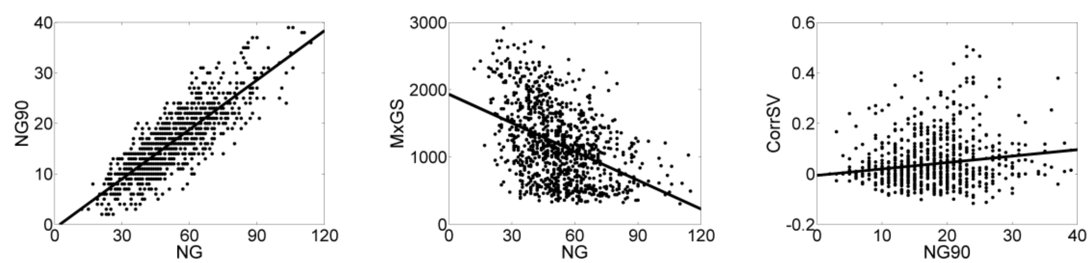


Figure 3 Correlation between indicators in the *Nova Icària* DMA in Barcelona Water Distribution Network (darker indicates either higher correlation or higher anti-correlation)



(a) NG – NG90 correlation (b) NG – MxGS correlation (c) NG90 – CorrSV

Figure 4 Cross-correlation values between different pairs of indicators

of a table with as many columns as indicators (*step 3*). After the generation and evaluation of all random sensor placements, this table grows up to as many rows as placements executed. Finally, the indicators (table columns) are correlated to each other and the result is plotted in a 2-D matrix (*step 4*).

Figure 3 shows the 2-D correlation matrix between each pair of indicators after 10.000.000 executions. Each cell corresponds to the value of the correlation between the indicators in the row and in the column. A lighter colour indicates a correlation value closer to 0. Note, however, that we actually plot the correlations as absolute values so the positive and negative correlations are considered equal (darker cells indicate either positive or negative correlation). The reason is that we are interested in detecting which indicators are independent and which are redundant; hence it is not necessary to distinguish between positive and negative correlations because both indicate that the indicators are dependent.

The first observation to do is the existence of indicator clusters with high correlations. This can be clearly seen in the top-left darker area that contains all the topological indicators (from *NG* to *SdGS90*). What this actually means, for example, is that a sensor placement that results in a large *NG* (number of all groups) also generates a large *NG90* (number of groups spanning 90% of the nodes). On the contrary, a large *NG* results in a small *MxGS* (largest group) – note in this case the two indicators are actually anti-correlated. These two opposite correlations can be observed with more detail in the plots (a) and (b) in Figure 4 where the positive and negative relationships between this pair of indicators are clearly depicted by the least-squares line.

The topographical indicators – from $AvGS_{geo}$ to $SdGS90_{geo}$ in the matrix – behave differently because they are only correlated with themselves (second dark cluster around the diagonal). From these results, a significant conclusion is that the topological and topographical measurements are independent. This has to be considered because the groups that result from a particular sensor placement might be small in terms of number of leak locations, but this does not necessarily imply that these are also small in terms of radius. Nevertheless, tests in other DMAs with different structure (less meshed) indicate that the topographical and topological indicators are correlated to each other. Therefore a further analysis on this discrepancy must be done with the objective to detect which DMA characteristics affect the indicators performance.

Similarly, the sensibility indicators (three most-right matrix columns) are also only correlated to themselves. As an example the plot of Figure 4(c) shows the low correlation between *NG90* and *CorrSV*. This lack of dependency makes the sensibility indicators become good candidates to be considered for inclusion in cost functions with multiple factors.

Finally, the correlation results can also be used to select the most representative indicators in order to consider them for the cost function of the optimizer. We can extract information to make this selection if we sort the indicators by the cumulated correlation values: for a given indicator, the summation of the correlation values with all other indicators (simply put, the sum of the matrix columns). This sorting shows that the *NG* and *NG90* are those that achieve the highest cumulated values; hence the NG-based indicators are a good representative of all the rest. This good performance can also be used as a validation of the manual top-down approach as these indicators are intuitively selected because a higher number of groups implies that there are more different columns in the FSM matrix, and this actually leads to a higher leak distinguishability, which is one of the optimizer's objectives. The optimization method in [12] is actually based on the *NG90* indicator.

5. CONCLUSION

This paper has presented a detailed study on the structure of the cost function to be used in a sensor placement optimization. Several indicators have been reviewed and compared by means of a cross-correlation analysis. Results show that there exists a certain level of redundancy between clusters of indicators, while some others are highly independent.

The existence of correlated indicators indicates that the most representative cost function should contain one indicator from each cluster. However, considering different factors in the cost calculation becomes challenging as the merging technique to apply might not be straightforward. For instance, the simplest option is to compute an arithmetic mean of the indicators. The drawbacks in this case are that: (1) these must be in very different scales and normalization is needed; (2) only indicators with the same units should be summed; and (3), since one might want to weight the indicators to apply different importance levels, the computation of the right weights values arises as an additional problem.

In order to consider all the factors at once, and assuming that the use of the cost function is within an optimizer [12], the merging technique that seems most appropriate is to use a multiple objective optimization. The main disadvantage of this technique is the additional computational cost that it implies considering the complexity of the problem; hence dimension reduction techniques might be applied together with the multiple-objective optimization.

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

[9] 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)

[10] 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).

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

[12] E. Bonada et al. "Practical-oriented pressure sensor placement for model-based leakage location in water distribution networks", Submitted for review to *11th International Conference on Hydroinformatics*, (2014).

[13] R. Pérez et al., "Accuracy assessment of leak localization method depending on available measurements", *12th International Conference on Computing and Control for the Water Industry CCWI 2013*, (2013).

[14] D. Misiunas, et al. "Failure monitoring in water distribution networks.", *Instrumentation, Control and Automation for Water and Wastewater Treatment and Transport Systems IX 53.4*, (2006).

[15] A. Preis et al., "Multi-objective optimization for conjunctive placement of hydraulic and water quality sensors in water distribution systems." *Water Science & Technology: Water Supply 11.2* (2011).