# Optimal Pressure Sensor Placement and Assessment for Leak Location Using a Relaxed Isolation Index: Application to the Barcelona Water Network☆

Miquel À. Cugueró-Escofet, Vicenç Puig, Joseba Quevedo[a]

[a]*Supervision, Safety and Automatic Control Research Center (CS2AC), Polytechnic University of Catalonia (UPC), Terrassa Campus, Gaia Research Bldg., Rambla Sant Nebridi, 22. 08222 Terrassa, Barcelona, Spain (e-mail: {miquel.angel.cuguero,vicenc.puig,joseba.quevedo}@ upc.edu).*

**Abstract**

Water distribution networks are large complex systems affected by leaks, which often entail high costs and may severely jeopardise the overall water distribution performance. Successful leak location is paramount in order to minimize the impact of these leaks when occurring. Sensor placement is a key issue in the leak location process, since the overall performance and success of this process highly depends on the choice of the sensors gathering data from the network. Common problems when isolating leaks in large scale highly-gridded real water distribution networks include leak mislabelling and the obtention of large number of possible leak locations. This is due to similarity of leak effect in the measurements, which may be caused by topological issues and led to incomplete coverage of the whole network. The sensor placement strategy may minimize these undesired effects by setting the sensor placement optimisation problem with the appropriate assumptions (e.g. geographically cluster alike leak behaviors) and by taking into account real aspects of the practical application, such as the acceptable leak location distance. In this paper, a sensor placement methodology considering these aspects and a general sensor distribution assessment method for leak diagnosis in water distribution systems is presented and exemplified with a small illustrative case study. Finally, the proposed method is applied to two real District Metered Areas (DMAs) located within the Barcelona water distribution network.



*Keywords:* Sensor placement, fault detection and isolation, leak location, correlation coefficient, water distribution networks.

## 1. Introduction

An issue of great concern in water drinking networks is the existence of leaks at the distribution stage, highly related with water resource savings and management costs. The classical approach to leak control is passive, i.e. a leak is repaired when it becomes visible. Recently developed acoustic instruments also allow non-visible leak location [1], but the use of such instrumentation in large-scale water networks is expensive and time-consuming. An acceptable approach is to divide the network into District Metered Areas (DMAs), where the *flow* and the *pressure* are measured [2, 3], and use a leak control-system on a permanent basis. Concretely, leaks affecting DMAs increase the flow and decrease the pressure magnitudes at the DMA inputs. Several empirical studies propose mathematical models to characterise the leak flow magnitude with respect to the pressure magnitude at the leak location point [4, 5].

☆This work has been partially funded by the Spanish Ministry of Science and Technology through the Project ECOCIS (Ref. DPI2013-48243-C2-1-R) and Project HARCRICS (Ref. DPI2014-58104-R), and by EFFINET grant FP7-ICT-2012-318556 of the European Commission.

Best practice in the analysis of DMA flows consists in estimating the leak magnitude when the flow is minimum. This typically occurs at night time, when customers' demand is low and hence the leak magnitude over the total DMA flow is at its highest rate [3]. Therefore, an accepted approach by the practitioners is to monitor the minimum DMA night flow in order to detect and repair the leaks when occurring, while also employing techniques to estimate the corresponding leak magnitude [3]. However, the leak detection may not be straightforward, since different kind of phenomena e.g. unpredictable variations in the customers' demand or measurements' noise, long-term trends or seasonal effects, may occur.

Several works in the literature have addressed the leak location problem in DMAs. In [6], a review of transient-based leak detection methods is summarized. In the seminal work [7], a model-based leak detection and location is solved by means of a least-squares estimation problem. However, the latter problem is challenging when considering the non-linear models involved. Alternatively, a method based on pressure measurements and leak sensitivity analysis is proposed in [8], where a set of residuals —generated as the difference between pressure measurements provided by several sensors installed within the DMA and their estimations by the network hydraulic model— is analysed considering a certain threshold, which takes into account practical factors e.g. the model uncertainty and the measurement noise. This approach shows satisfactory results under ideal conditions, but its performance degrades when considering nodal demand uncertainty and measurement noise. This technique is improved in [9], where an extended time horizon analysis is considered and a comparison of the performance using different metrics is presented.

The performance of the leak location approach is highly dependent on the sensor number and placement within the DMA. Hence, the sensor placement strategy is a key issue to consider in the overall leak location process. There is an important trade-off between the number of sensors and the subsequent cost preventing the use of a high number of sensors for leak location purposes. Consequently, this number should be optimised at the sensor placement stage, in order to produce the maximum benefit i.e. maximize the leak location performance at the minimum cost. According to these constraints, the sensors considered are pressure sensors since they are a cheaper alternative to flow meters for the company managing the network. However, the methodology presented here may also be used for alternative sensor placement setups e.g. combining pressure and flow meters as in [10], or chlorine meters for water quality fault diagnosis. Hence, the methodology may be arranged with minor modifications to the placement of other type of sensors.

Regarding sensor placement for fault detection and isolation (FDI) purposes, several works concerning this subject may be found in the literature. Some approaches consider the study of structural matrices in order to locate sensors based on isolability criteria [11]. In [12], an optimal set of sensors for model-based FDI is sought by means of an optimisation method based on binary linear programming. These works are embraced in the general framework of FDI of dynamic systems. However, they are not specially suited to consider the non-explicit non-linear set of equations describing a water distribution network. Alternatively, several works treated the sensor placement problem when applied to water distribution networks, most of them addressing the water contamination monitoring (e.g. [13, 14]), where sensor placement is considered in a large water distribution network in order to detect malicious introduction of contaminants. Regarding leak location, less contributions addressed the problem of sensor placement. This problem is studied in [15], where an strategy based on the leak isolability maximization is considered to optimally place the sensors based on the water network structural model, and in [8], where an optimal sensor placement is formulated as an integer programming problem, similarly as presented here. Also, an entropy-based approach for efficient water loss incident detection is introduced in [16].

Furthermore, leak location in real water networks involves discrimination among a high number of possible leak locations —typically, the DMA nodes—, which often leads to mislabel the actual leak location due to the limited number of sensors available. However, in practical situations there is no need to locate the leak at the exact point where is produced, since final on-the-ground leak location techniques —e.g. ground-penetrating radar, acoustic listening devices [17]— may precisely locate them starting from a close area where the actual leak is occurring. Hence, this calls for a methodology of sensor placement trying to cluster similar leak behaviors geographically, in order to minimize the number of installed sensors and locate the leak within a certain cluster distance precision.

Having this into account, in this paper a new approach for sensor placement focused on leak location in DMAs is proposed, based on the method introduced in [18]. Alternatively to [8], the approach presented here does not binarize the sensitivity matrix, but instead use the complete numerical information of this matrix, leading to better leak location performance as pointed out in [18, 9]. The use of the numerical sensitivity matrix in the sensor placement problem requires the reformulation of the optimisation problem introduced in [8], since even both approaches are formulated as

an integer optimization problem, isolability conditions considered in the former do not apply here. This reformulation leads to a non-linear integer optimization problem of large dimension that cannot be tackled with deterministic solvers, but with heuristic approaches. Here the use of Genetic Algorithms (GAs) is proposed, since they are a well suited approach to handle problems of this nature [19, 20]. The novel aspects of the sensor placement methodology are, first, to reduce the effect of the leak mislabelling at the sensor placement stage, trying to geographically cluster nodes with similar leak signature. Hence, the sensor distribution favouring this clustering is selected, and the rest are discarded. The second novel aspect of the paper is the proposal of an assessment methodology, using new figures of merit in order to provide the goodness of a certain sensor set from the leak location point of view after the sensors are placed. The assessment indices proposed assume that the leak location algorithm will be based on the correlation between leak signatures, but are independent of the methodology used to place the sensors. Hence, the intrinsic leak mislabelling that may occur in real DMAs with a low ratio between sensors and network nodes is taken into account. To the knowledge of the authors, the use of a general assessment in terms of potential number of isolated leaks is not present in the literature. In [10], an assessment based on the isolation distance is presented in a real DMA, but this do not include the goodness of the sensor distribution regarding the number of isolable leaks for the whole network. The methodology presented is first illustrated in a small example network and then evaluated in several DMAs, located within the Barcelona water network.

The paper is organized as follows: the leak location methodology used as the basis for this work is introduced in Section 2. The sensor placement methodology is presented in Section 3, and the isolability assessment used to evaluate the goodness of the sensor set proposed is introduced in Section 4. The application case studies, based on several DMAs, and the results obtained applying the methodology proposed are shown in Section 5. Finally, in Section 6, some concluding remarks and future work are given.

## 2. Leak Location Problem

The leak location problem may be divided in two different levels: the sensor placement stage and the leak location stage, given a set of sensors. The leak location approach is summarised in this section, since it is the basis of the sensor placement algorithm formulation proposed in this work.

The leak location methodology considered here aims to locate leaks within a DMA by means of some pressure measurements gathered from the network and their estimations, obtained by a network hydraulic model. For a given DMA with $N$ demand nodes and $M$ pressure sensors, the leak detection methodology relies on the computation of the residuals $\mathbf{r} = [r_1 \ \dots \ r_M]^T$, where $r_i \in \mathbf{r}$ is the difference between the pressure measurement $p_i$ and its corresponding estimation $\hat{p}_i$ obtained from a leakless simulation using the corresponding network hydraulic model as follows,

$$r_i = p_i - \hat{p}_i, \quad i = 1, \dots, M, \tag{1}$$

having one residual per each available pressure measurement within the DMA. The number and placement of the sensors is a key issue in the performance of the leak location method and is the target of this paper.

On the other hand, the leak location method relies on the study of the residual vector in (1) by means of sensitivity analysis, aiming to determine the effect of each particular leak on every available pressure sensor measurement at a certain time by means of the sensitivity matrix defined as follows,

$$\mathbf{S} = \begin{pmatrix} s_{11} & \cdots & s_{1N} \\ \vdots & \ddots & \vdots \\ s_{M1} & \cdots & s_{MN} \end{pmatrix}, \tag{2}$$

considering $M \leq N$ sensors within the network and $N$ possible faults (assuming leaks only in nodes) with

$$s_{ij} = \frac{\hat{p}_{ij} - \hat{p}_i}{f_j}, \quad i = 1 \dots M, j = 1 \dots N, \tag{3}$$

where $\hat{p}_i$ is the leakless scenario pressure estimation in node $i$ and $\hat{p}_{ij}$ is the pressure estimation in node $i$ due to leak $f_j$ scenario occurring in node $j$.

To obtain the sensitivity matrix $\mathbf{S}$, a leak scenario is generated per each node by numerical simulation, using EPANET hydraulic solver [21]. The sensitivity vector corresponding to each column of the sensitivity matrix $\mathbf{S}$ is then obtained as follows,

$$\mathbf{s}_j = \begin{bmatrix} s_{1j} \\ \vdots \\ s_{Mj} \end{bmatrix}, \quad j = 1, \cdots, N, \tag{4}$$

which is also known as leak signature. Each simulated fault scenario is performed by setting a leak of magnitude $f_j$ in the $j^{th}$ DMA network node. This procedure is repeated for all the $N$ existing network nodes. Then, matching both the residual vector in (1) and the sensitivity vectors in (4), leak location may be performed by checking which node has the highest potential to present a leak. This analysis may be performed by using different metrics [22]. Here, a method presented in [18, 10], based on the correlation between residual and sensitivity vectors, is considered. According to the study in [9], this method presents the best performance for leak location. However, it should be remarked that the sensor placement method presented here could be also applied when considering alternative leak location methods using the sensitivity matrix (2).

The current metric considered here for leak location is based on the correlation function given by the inner product of the regressor vector in (1) and the sensitivity vector in (4), for each particular fault in node $j$

$$\gamma_j = \frac{\mathbf{s}_j{}^T \mathbf{r}}{|\mathbf{s}_j||\mathbf{r}|}. \tag{5}$$

Then, the highest correlation determines the candidate leaky node $k$,

$$\gamma_k = \max(\gamma_1, \cdots, \gamma_N). \tag{6}$$

The objective here is to develop a methodology to place a given number of sensors, $M$, within a DMA in order to obtain a sensor set maximizing leak isolability under realistic conditions. In DMAs with a large number of nodes, the sensitivity to different leaks occurring in different nodes may be very similar. This sensitivity similarity may lead to confusion between different leaks, specially when a low number of sensors is available and uncertainty in the measurements and the model is present, which is generally the actual situation. This situation may be solved e.g. by increasing the number of sensors, in order to increase the dimension of the leak signature, or by placing the sensors with a methodology clustering the nodes presenting sensitivity similarity, as suggested by the methodology presented here. This sensor placement methodology, which relies on the leak location scheme presented in this section, is introduced in the next section.

## 3. Sensor Placement Methodology

### *3.1. Sensor Placement as an Optimisation Problem*

The goal of this methodology is to place the best sensor set in order to locate the leak as precisely as possible within the considered water network. The sensor distribution method is based on the system sensitivity matrix (2). As discussed in the introduction, a former methodology is presented in [8], where the residuals and the sensitivity matrix are binarized by a certain threshold value. In the approach presented here, the complete numerical information is used in order to avoid data loss introduced by the binarization in order to increase leak discriminability [18]. Moreover, the sensor placement method presented here uses a relaxed isolation index to better handle some real-world effects affecting water network systems, such as system non-linearity, sensor measurements resolution and model uncertainty (e.g. in the demands or network element parameters). These real-world effects cause deviation between the modelled and the actual system behavior, which may lead to mislabel the latter, and to confuse different leak scenarios — sensitivity vectors in (4)—. However, if the confusion involves geographically close behaviors, these undesired effects do not severely impact the final leak location result. Hence, the optimal sensor distribution aims to place the sensors to geographically cluster leaks with similar signature (4). The method presented uses the information of the complete sensitivity matrix (2) —i.e. assuming that all the nodes of the network are measured—, which must be computed beforehand by numerical simulation. In order to perform the sensor placement of $M$ sensors, let us define the binary decision vector that represents the selected sensors,

$$\mathbf{x} = \begin{pmatrix} x_1 & \cdots & x_N \end{pmatrix}^T, \tag{7}$$

where $x_i = 1$ if the pressure sensor in node $i$ is installed and 0 otherwise. Defining

$$\mathbf{X}(\mathbf{x}) = diag(x_1, \cdots, x_N), \tag{8}$$

the sensitivity vectors obtained using the set of selected sensors can be represented as follows,

$$\bar{\mathbf{s}}_j(\mathbf{x}) = \mathbf{X}(\mathbf{x})\mathbf{s}_j, \quad j = 1, \cdots, N, \tag{9}$$

where $\mathbf{s}_j$ is the sensitivity matrix obtained when all the $N$ sensors are available, i.e. $M = N$. Hence, the projection between two different leak signatures $i$ and $j$ for a given subset of sensors $\mathbf{x}$ is introduced by their inner product as follows,

$$\gamma_{ij}(\mathbf{x}) = \frac{\bar{\mathbf{s}}_i^T(\mathbf{x})\bar{\mathbf{s}}_j(\mathbf{x})}{|\bar{\mathbf{s}}_i(\mathbf{x})||\bar{\mathbf{s}}_j(\mathbf{x})|} = \frac{\mathbf{s}_i^T\mathbf{X}(\mathbf{x})\mathbf{s}_j}{|\mathbf{X}(\mathbf{x})\mathbf{s}_i||\mathbf{X}(\mathbf{x})\mathbf{s}_j|}, \quad i, j = 1, \cdots, N, \tag{10}$$

where $\bar{\mathbf{s}}_i, \bar{\mathbf{s}}_j$ are vectors corresponding to two different fault signatures (columns) for each class (leak) in the sensitivity matrix (2) and $\gamma_{ij}$ is a measure of similarity between these two classes. From (10), the projection —or cross-correlation— matrix may be stated as follows,

$$\Gamma(\mathbf{x}) = \begin{pmatrix} \gamma_{11}(\mathbf{x}) & \cdots & \gamma_{1N}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \gamma_{N1}(\mathbf{x}) & \cdots & \gamma_{NN}(\mathbf{x}) \end{pmatrix}, \tag{11}$$

containing the projection between all the leak signatures in the sensitivity matrix. For a given sensor configuration, this matrix presents the degree of similarity of each leak with all the leaks considered in the network. The goal is to find a sensor configuration such that the largest values are in the diagonal of this matrix, whilst the values close to these largest values correspond to leaks which are not far geographically. It may be noted that this matrix is symmetric, so $\Gamma = \Gamma^\top$.

In order to evaluate the quality of a sensor allocation setup, $\rho_{ij}(\mathbf{x})$ is applied to (11) as follows,

$$\begin{aligned} \rho_{ij}(\mathbf{x}) = &\left(\gamma_{ij}(\mathbf{x})\left(1 - \frac{d_{ij}}{d_{\max}}\right)\right)^{d_c} \\ &+ \left(\left(1 - \gamma_{ij}(\mathbf{x})\right)\frac{d_{ij}}{d_{\max}}\right)^{d_f}, \quad i, j = 1 \ldots N, \end{aligned} \tag{12}$$

where $\gamma_{ij}$ is the cross-correlation between leak $i$ and leak $j$ signature vectors, $d_{ij}$ is the topological (pipe) distance between leaky nodes $i$ and $j$, $d_{\max}$ is the maximum pipe distance for the whole network, and $d_c$ and $d_f$ are tuning parameters related with the included high-correlated close leaks cluster and the excluded high-correlated distant leaks subset, for a given *i-j* leak pair, respectively. This particular cost function aims to obtain the best sensor configuration in order to locate the leaky node as precisely as possible, by promoting sensor setups that geographically group leaks with similar signature. On the one hand, parameter $d_c$ is related with the isolation zone —or cluster—, i.e. the zone where the leaks may be mislabelled. An acceptable value for this perimeter, which is provided by the company managing the network, is about 200 m in a real DMA. Inside this perimeter, alternative on-the-ground techniques —e.g. ground penetrating radar— are used for finer isolation. Hence, $d_c$ is selected accordingly, taking into account the corresponding DMA $d_{\max}$, so the cost function value decreases when the distance between high correlated leaks is below the selected perimeter. On the other hand, parameter $d_f$ is related with the exclusion zone, i.e. the zone where leaks should not be mislabelled with the leaks in the isolation zone. Hence, $d_f$ is selected such that outside the isolation zone leak mislabelling is penalized, i.e. the cost function value decreases when the distance between

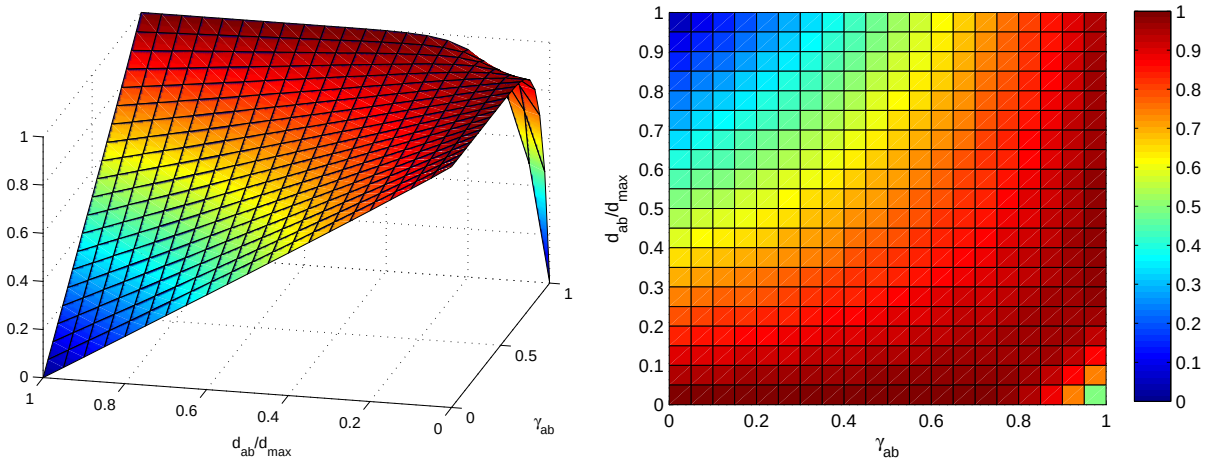
low correlated leaks is above the selected perimeter. The values of $d_c$ and $d_f$ parameters should be selected such that (12) range from zero to one for each *i-j* leak pair. These parameters may also be used to adjust the target of the optimisation. The bigger $d_c$, $d_f$, the narrower the related zone. Generally $d_c$ is chosen bigger than $d_f$, so the slope of the exclusion term is lower, since the exclusion zone embraces all the nodes in the network outside the isolation perimeter. However, if one wants to get focused on the leak isolation zone, $d_f$ may be chosen arbitrarily high in order to penalise arbitrarily distant nodes. Considering (12), the sensor placement may be stated as an optimisation problem as follows,

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad \rho(\mathbf{x}) \\ &\text{subject to} \quad \sum_{i=1}^{N} \mathbf{x}_i = M, \end{aligned} \tag{13}$$

where

$$\rho(\mathbf{x}) = 1 - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij}(\mathbf{x}) \tag{14}$$

is to be optimised over the full $N$ sensors set available, and $M$ is a predefined restriction on the number of sensors to install. The cost function in (14) for a single *i-j* leak pair is depicted in Figure 1 for illustrative purposes. The criterion to select the parameters $d_c$ and $d_f$ may be illustrated with Figure 1b, for a DMA with $d_{\max} = 1000$ m. Parameter $d_c$ is selected so $\gamma_{ij}$ starts decreasing at a normalised distance $d_{ij}/d_{max} = 0.2$, corresponding to a distance of 200 m. Similar criterion is applied for the selection of $d_f$, related with the leak exclusion area. Hence, the use of this cost function aims to achieve a sensor distribution obtaining high-correlation/low-distance (first term in (12)) and low-correlation/high-distance (second term in (12)) leak scenario combinations.




(a) (b)

Figure 1: Cost function for a single $i - j$ pair

The sensor placement optimisation problem (13) is solved using GA, which is a suitable approach for large-scale binary non-linear problems as the one considered here [23]. Further details on the GA parameters utilised to solve this particular problem are given in Section 5.2.

### *3.2. Methodology summary*

Considering the optimisation problem presented in Section 3.1, a summary of the methodology for illustrative purposes is presented in Algorithm 1, which states how this methodology should be implemented. As previously

mentioned in Section 3.1, it may be noted that the sensitivity matrix (2) must be computed beforehand by numerical simulation of the network considered.

**Algorithm 1** Sensor Placement Methodology

**Require:** $\mathbf{S}$, $d_c$, $d_f$, $d_{max}$, $N$, $M$
  Initialise $\mathbf{x}$; {Assign binary random elements to decision vector $\mathbf{x}$ for $M$ sensors and $N$ nodes}
  **while** $\rho(\mathbf{x})$ is not minimum **do**
    Obtain $\mathbf{x}$ for $M$ sensors; {Optimisation algorithm computes a new potential decision vector}
    **for all** $N^2$ i-j node pairs **do**
      Compute $\gamma_{ij}(\mathbf{x})$; {Corresponding projection for i-j pair is obtained}
      Compute $\rho_{ij}(\mathbf{x})$; {Corresponding quality sensor allocation measure for i-j pair is obtained}
    **end for**
    Compute $\rho(\mathbf{x})$; {Corresponding cost value is obtained}
  **end while**
  **return** $\mathbf{x}$ {Return selected decision vector i.e. optimal sensor set}

## 4. Leak Isolability Assessment

In order to assess the leak isolability of a given sensor configuration, a metric based on the confusion matrix is proposed [24]. Each column of this matrix represents the actual leak, whilst each row stands for the degree of similarity with the signature vectors of each considered leak,

$$C = \begin{pmatrix} \kappa_{11} & \cdots & \kappa_{1N} \\ \vdots & \ddots & \vdots \\ \kappa_{N1} & \cdots & \kappa_{NN} \end{pmatrix}, \tag{15}$$

where $\kappa_{ij} \in \{0, 1\}$ for $i, j = 1 \dots N$. The name stems from the fact that this representation allows to check a given sensor configuration, commonly by the number of correct and incorrect predictions achieved by the leak location method compared with the actual leak occurring in the test data. The values of $\kappa_{ij}$ depend on each particular isolation criterion used. Here, a criterion based on the cross-correlation (11) is used to obtain the maximum correlation for each actual fault,

$$\gamma_{ij_{\max}} = \max_{j \in 1 \dots N} \gamma_{ij}, \quad i = 1 \dots N, \tag{16}$$

where $\kappa_{ij}$ is determined as follows,

$$\kappa_{ij} = \begin{cases} 1 & \gamma_{ij} = \gamma_{ij_{\max}} \\ 0 & \text{otherwise} \end{cases}, \quad i, j = 1 \dots N. \tag{17}$$

In order to provide less conservative leak isolation results, while still realistic and well suited to the optimisation criterion stated in (13), the matrix of pipe distances among nodes of the network are considered,

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1N} \\ \vdots & \ddots & \vdots \\ d_{N1} & \cdots & d_{NN} \end{pmatrix}, \tag{18}$$

and the isolation condition (17) may be relaxed by a certain fault isolation cluster distance $d_{\text{cluster}}$ as follows,

$$\kappa_{ij} = \begin{cases} 1 & \max \mathbf{d}_{ij_{\max}} < d_{\text{cluster}} \quad \text{and} \quad d_{ij} < \max \mathbf{d}_{ij_{\max}} \\ 0 & \text{otherwise} \end{cases}, \quad i, j = 1 \dots N, \tag{19}$$

where $\mathbf{d}_{ij_{\max}}$ is the distance between the actual leaky node $i$ and the node (or nodes) with highest correlation $\gamma_{ij_{\max}}$ —i.e. predicted leaky nodes—, and $d_{\text{cluster}}$ is the maximum allowed distance between the actual leaky node $i$ and the predicted leaky nodes, in order to consider that the leak in $i$ is correctly isolated. When several predicted faulty nodes are obtained, the worst case —i.e. $\max \mathbf{d}_{ij_{\max}}$— is considered.

The number of correctly isolated leaks is given by the following isolation index,

$$\zeta = \operatorname{tr}(C), \tag{20}$$

so the correct isolated faults are those which are assigned to its own class and not to any other possible leak occurring in the network. The best isolation index ($\zeta_{best}$) for a given $d_{\text{cluster}}$ is obtained when sensors in all nodes are available i.e. when $M = N$, which states a topological limit,

$$0 \leq \zeta_{opt} \leq \zeta_{best} \leq N, \tag{21}$$

where $\zeta_{opt}$ is the isolation index obtained with the corresponding optimal sensor placement, for a given $d_{\text{cluster}}$. Let us also define a particular $\zeta_{opt}$ and $\zeta_{best}$ considering (17), i.e. $\zeta_{opt_0}$ and $\zeta_{best_0}$, respectively. Then, a more general topological limit which does not depend on the distance between nodes may be given by

$$0 \leq \operatorname{rank} \mathbf{S} \leq \zeta_{opt_0} \leq \zeta_{best_0} \leq N, \tag{22}$$

where $\mathbf{S}$ is the sensitivity matrix obtained when all the $N$ sensors are available, i.e. $M = N$. Previous relation is meaningful since $\zeta_{best_0}$ computation may be infeasible for DMAs with a high number of nodes $N$. Then, $\operatorname{rank} \mathbf{S}$ may provide a useful computationally efficient approximation, specially when this magnitude is close to the DMA number of nodes $N$. It must be noted that, since matrix $\mathbf{S}$ is affected by the pressure sensor resolution, confusion between leaks may be induced (e.g. linear dependency between columns of $\mathbf{S}$) as the DMA size increases.

It may also be noted that the ratio $\phi_{best} = \frac{\zeta_{best}}{\zeta_{best_0}} \geq 1$ suggests the benefit obtained by the geographic relaxation when all the sensors are available (the bigger the better), whilst the ratio $\phi_{opt} = \frac{\zeta_{opt}}{\zeta_{opt_0}} \geq 1$ suggests the geographical relaxation benefit for the sensor subset considered. This benefit may be also obtained from an extra coverage percentage over $\zeta_{\text{best}}$ as follows,

$$\delta = \frac{\zeta_{opt} - \zeta_{opt_0}}{\zeta_{best}} 100, \tag{23}$$

where $\delta$ is the percentage of extra coverage over $\zeta_{best}$, obtained when geographically relaxing the assessment.

A summary of the isolability assessment presented in this section is introduced in Algorithm 2, which states how the sensor placement assessment indices may be evaluated in order to evaluate the quality of the solution obtained.

## 5. Application Examples: Hanoi and Barcelona Drinking Water Networks

### *5.1. Description*

Several DMAs of different level of complexity are used here in order to show the performance of the sensor placement method presented. First, a small DMA is considered to illustrate the method in detail. The Hanoi DMA, an existing benchmark network widely used in the literature (see, e.g. [25]), is considered for this purpose (Figure 2). This DMA has 31 nodes and 34 links, and delivers water to the end consumers by means of a single input point. Also, two different DMAs located in the Barcelona area, with higher nodal density, are used as case studies (Figure 3). On the one hand, the *Canyars* DMA (Figure 4) is located at the pressure level 80 within the Barcelona water supply network. This DMA has 694 nodes and 719 links, and delivers water to the end consumers by means of a single input point. On the other hand, the *Castelldefels Platja* DMA (Figure 5) is located at the pressure level 50 within the Barcelona water supply network. This DMA has 4952 nodes and 5116 links, and covers an area of 606 ha. The DMA has two

**Algorithm 2** Leak Isolability Assessment

**Require:** $D$, $\mathbf{x}$, $\Gamma(\mathbf{x})$, $d_{\text{cluster}}$, $d_{max}$, $N$
  **for all** $N^2$ i-j node pairs **do**
    Obtain $\gamma_{ij_{\max}}$; {Maximum correlation for each leak considering the projection for i-j pairs}
    Obtain $\mathbf{d}_{ij_{\max}}$; {Distance between faulty node $i$ and node/s with highest correlation $\gamma_{ij_{\max}}$}
    **if** $\max \mathbf{d}_{ij_{\max}} < d_{\text{cluster}}$ **and** $d_{ij} < \max \mathbf{d}_{ij_{\max}}$ **then**
      $\kappa_{ij} = 1$; {Leak misslabeling between nodes $i$ and $j$}
    **else**
      $\kappa_{ij} = 0$; {No leak misslabeling between nodes $i$ and $j$}
    **end if**
  **end for**
  Compute $\zeta_{opt}$; {# of correct isolated faults for the optimal sensor set}
  Compute $\zeta_{best}$; {# of correct isolated faults when $M = N$}
  Compute $\zeta_{opt_0}$; {# of correct isolated faults for optimal sensor set without geographic relaxation}
  Compute rank $\mathbf{S}$; {# of correct isolated faults for sensors in all nodes without geographic relaxation}
  Compute $\zeta_{best_0}$; {Obtain the rank of the sensitivity matrix}
  Compute $\phi_{best}$; {Obtain geographic relaxation benefit index when $M = N$}
  Compute $\phi_{opt}$; {Obtain geographic relaxation benefit index for the optimal sensor set}
  Compute $\delta$; {Obtain extra coverage over $\zeta_{\text{best}}$ when geographically relaxing the isolation}
  **return** $\zeta_{opt}$, $\zeta_{best}$, $\zeta_{opt_0}$, rank $\mathbf{S}$, $\zeta_{best_0}$, $\phi_{best}$, $\phi_{opt}$, $\delta$, $\zeta_{\text{best}}$ {Return assessment indices}

inputs —named Ferrocarril and Pi Tort, respectively— delivering water to the end consumers. The current DMA size motivated its reduction to an equivalent hydraulic model of 2828 nodes using skeletonization techniques [26], more suitable for high demanding computation algorithms used here.

As mentioned in Section 3.2, the first step consists in obtaining the sensitivity matrix (2) by numerical simulation of the network under study, i.e. simulating a leak in each node of this network. These scenarios have been generated using EPANET hydraulic simulation software, as introduced in Section 2. In order to simulate these DMAs isolated from the water supply network, the boundary conditions —i.e. pressure and flow measurements from the network— are fixed. Generally, pressure is fixed using a reservoir, and the overall demand is obtained as the sum of the inflow distributed through the DMA using a demand pattern model. The total inflow is distributed using a constant coefficient —named base demand— in each consumption node. Hence, all the consumptions are assumed to share the same profile, whilst the billing information is used to determine the base demand of each particular consumption. A good estimation of the demand model is paramount for the real case application.

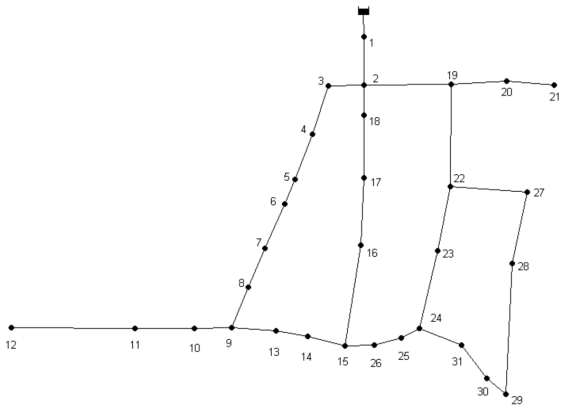



Figure 2: Hanoi DMA

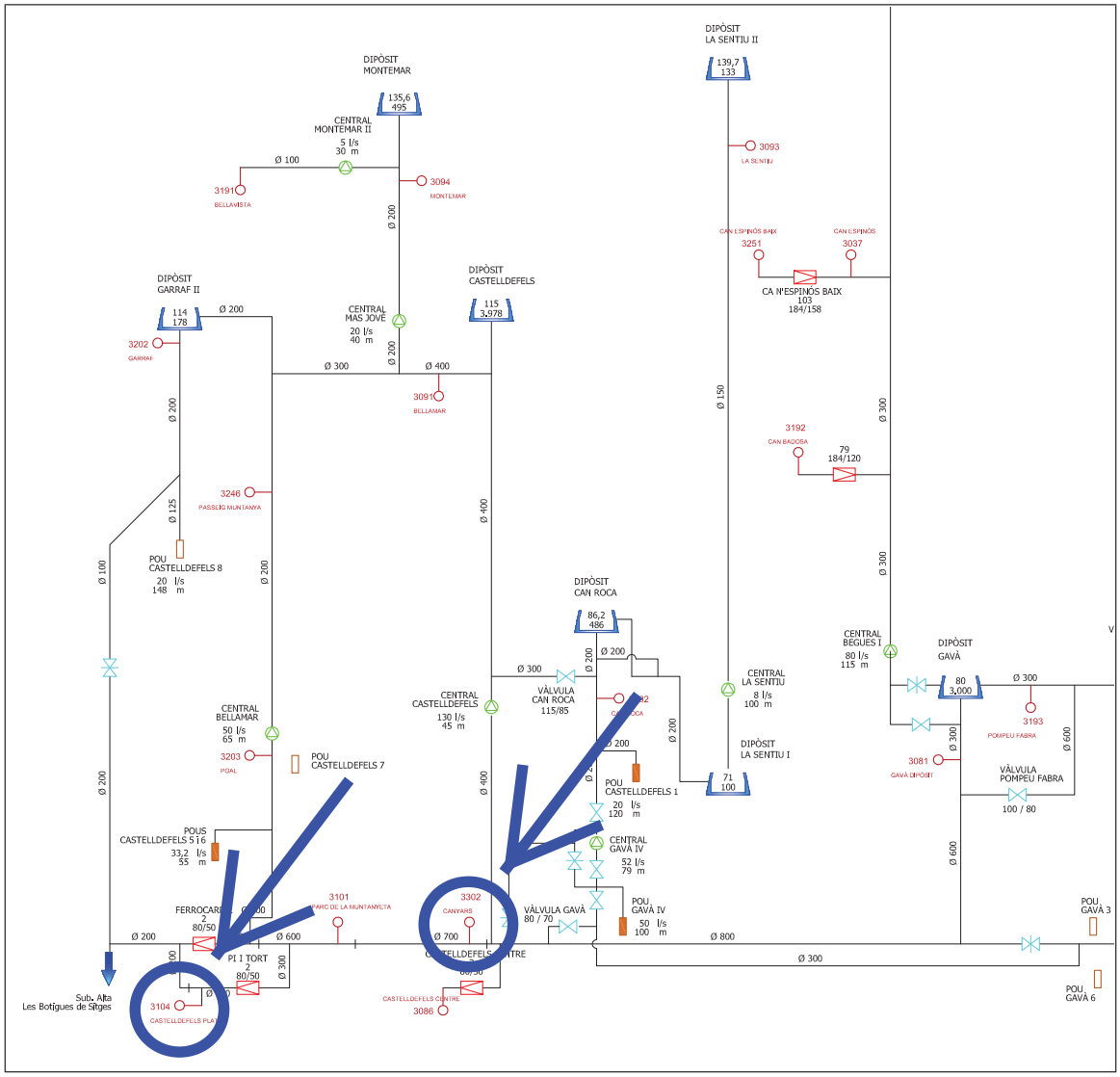



Figure 3: Barcelona Drinking Water Supply Network detail (arrows: Castelldefels Platja and Canyars DMAs, respectively)

### *5.2. Results*

In this section, the results achieved applying the sensor placement methodology described in Section 3 are presented. The sensors considered here are pressure sensors, which may be installed in any node of the network. The maximum isolation distance $d_{\text{cluster}}$, which is a parameter given by the company managing the network, is equal to 200 m for the Hanoi and Castelldefels DMAs, whilst is assumed of 2000 m for the illustrative Hanoi DMA, due to its particularly low nodal density (see Section 4 for a more detailed discussion). For distances below $d_{\text{cluster}}$, alternative more precise methods are more convenient in order to isolate the leak e.g. ground penetrating radar.

To solve the optimization problem (14), GAs available in the global optimization toolbox of MATLAB are used. Regarding GA parameters, an initial population of 100 random sensor sets, including the potential sensors to be used, is employed to seed the GA algorithm. At this stage, already installed DMA sensors may be included to seed the GA. The number of individuals in each generation is set to 100, the maximum number of generations allowed is set to 30,

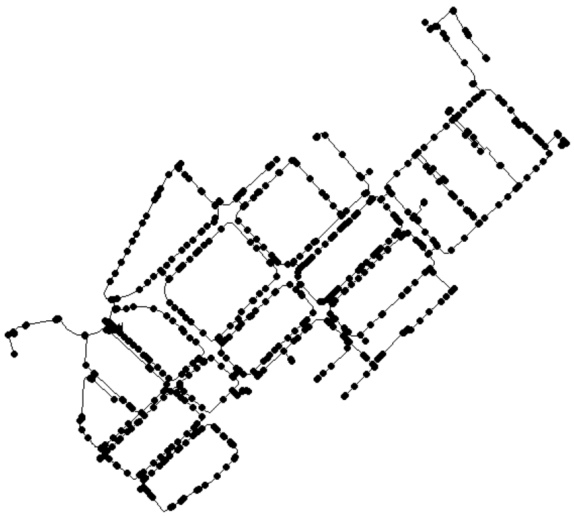


Figure 4: Canyars DMA

Figure 5: Castelldefels Platja DMA

the termination tolerance on the fitness function value is set to $1 \times 10^{-6}$ and the number of generations over which cumulative change in fitness function value is less than the termination tolerance —i.e. stall generations limit— is set to eight. Since the optimum obtained by the GA is not global, consecutive GA optimisations are conducted until fitness function value do not improve between two overall optimisations, aiming to achieve the best possible solution. The selection of these parameters takes into account that the optimisation is dealing with real high dimension DMAs and the problem may be computationally intensive. In order to face such computational issues, the use of local parallel computing is used when multiple labs are available in the host PC, in order to increase computation power. The host PC implemented Intel® Core™ i7 Quad-Core processors and 8 GB of 1600 MHz Dual Channel DDR3 memory, which allowed the use of such technique.

### *5.2.1. Hanoi DMA*

The sensor placement results obtained considering Hanoi DMA (Figure 2) are depicted in Figures 7a to 7d. The sensitivity matrix $\mathbf{S}$ is obtained for a 24 h scenario using an emitter coefficient —i.e. discharge coefficient for emitter placed at junction, representing the flow in liters per second (LPS) occurring at a pressure drop of 1 psi [21]— of 5 LPS/psi$^{0.5}$. The sensitivity $\mathbf{S}$ is concatenated for the 24 hours available, leading to a dimension of $744 \times 31$. The

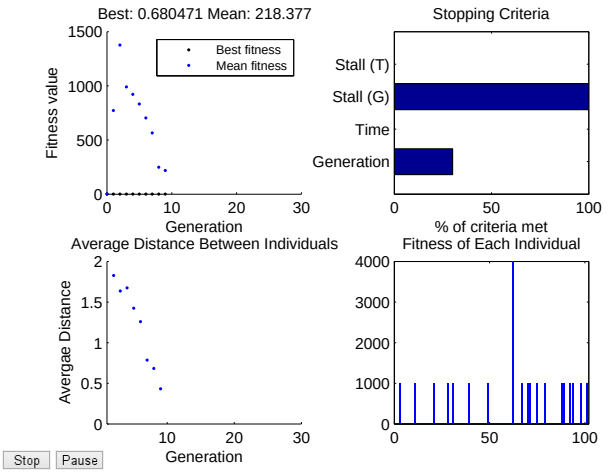
Table 1: Isolation assessment results, Hanoi DMA

| Number of sensors | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| $\zeta_{\text{opt}}$ | 14 | 20 | 22 | 22 |
| % of $N$ | 45 | 64.5 | 70.97 | 70.97 |
| % of $\zeta_{\text{best}}$ | 50 | 71.43 | 78.57 | 78.57 |
| % of $\zeta_{best_0}$ | 45 | 64.52 | 70.96 | 70.96 |
| $\zeta_{opt_0}$ | 30 | 31 | 31 | 31 |
| $\rho$ | 0.6804 | 0.6586 | 0.6445 | 0.6426 |

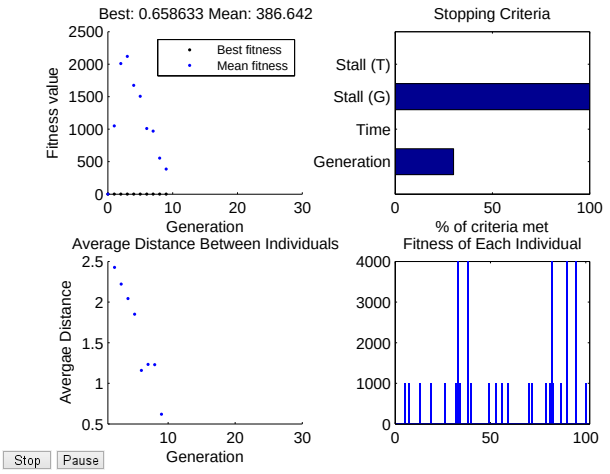
distance used here is the topological distance among nodes, i.e. minimum pipe distance between these elements. This network has a low density of nodes per squared meter, being the minimum distance among the closest nodes of 484 m. Hence, $d_{\text{cluster}}$ should be increased in comparison to a regular network, in order to provide realistic results. For this particular network, the maximum number of isolable faults considering all the sensors available ($\zeta_{best_0}$) is 31, and the maximum number of isolable faults considering all the sensors available and $d_{\text{cluster}}$ = 2000 m ($\zeta_{\text{best}}$) is 28 (90 % of $N$ = 31 nodes forming the network), respectively. According to the specified $d_{\text{cluster}}$ and $d_{\text{max}}$ = 16426 m, the cost function parameters have been chosen of $d_c$ = 5.36 and $d_f$ = 0.57. The evolution of the GA optimisation for each sensor distribution is depicted in Figures 6a to 6d. In the lower row, the latter figures show the evolution of the average distance between individuals among generations in the bottom-left subplot, and the fitness of each individual in the last generation in the bottom-right subplot. In the upper row, the evolution of the best and mean fitness value per generation is depicted in the upper-left subplot, and the GA stopping criteria is depicted in the upper-right subplot. These include the generations limit (set at 30), i.e. the maximum number of generations per optimisation, the time limit (unspecified), i.e. the maximum time in seconds whilst the GA runs before stopping, the stall generations limit (eight), i.e. the number of maximum consecutive generations without improving the average relative change in the best fitness function, over a given function tolerance ($1 \times 10^{-6}$), and the stall time limit (unspecified), i.e. the time interval in seconds after the GA stops if no improvement is obtained in the best fitness value. Time constraints have not been specified since they were not critical parameters in the optimisation. As it is shown in the Figures 6a to 6d, the stopping criteria met in all the optimisations for this particular case is the stall generation limit. Isolation assessment results concerning sensor distribution for different number of sensors (from two to five) are detailed in Table 1. It may be observed how the results obtained between four and five sensors do not improve in terms of $\zeta_{\text{opt}}$, even a better $\rho$ is achieved for five sensors at the optimisation stage. In this case, the benefit of installing extra sensors may obtain reduced isolation clusters, but still bigger than $d_{\text{cluster}}$. Hence, the optimal sensor distribution is obtained for four sensors (Figure 7c) since is the one achieving best $\zeta_{\text{opt}}$ with the minimum number of sensors. For the particular layout of this DMA, which is geographically large ($d_{\text{max}}$ = 16426 m) but has low nodal density ($N$ = 31 nodes, with minimum distance among closest nodes of 484 m), the geographical relaxation is not providing any particular benefit ($\delta$ is negative for all the distributions considered). However, the methodology presented here is useful when leak signature confusion is present, which is not the case in this network ($\zeta_{opt} = N$ for almost all the sensor distributions considered). Hence, a network of this characteristics is useful for illustrative purposes, but it is not a target network for the proposed methodology, more intended to be used in larger DMAs found in real water networks, as the ones presented in the following sections.

### *5.2.2. Canyars DMA*

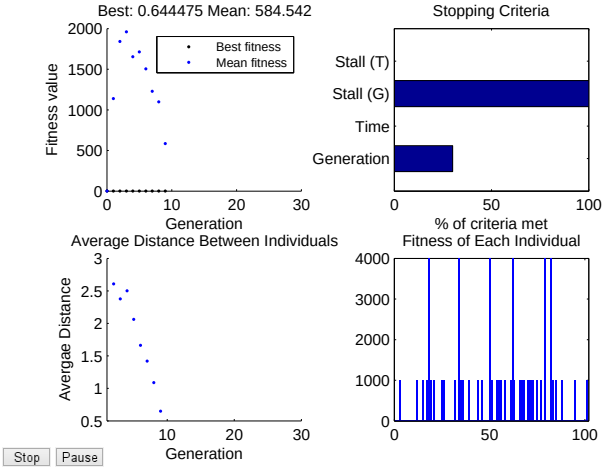
The sensor placement results obtained when considering Canyars DMA (Figure 4) are depicted in Figures 8a to 8c. The sensitivity matrix **S** is obtained for a fixed leak of 6 LPS, in an hourly sampled scenario comprised between 24/02/2014 9h and 25/02/2014 9h. Thus, the sensitivity **S** is concatenated for the 24 hours available, leading to a dimension of 16656 × 694. Also, the information in this matrix considers sensor resolution of 0.1 m in order to take into account current technological constraints of the available pressure sensors in the simulated scenario. The distance used here is the topological distance among nodes, i.e. minimum pipe distance between these elements. For this particular network, the maximum number of isolable faults considering all the sensors available ($\zeta_{best_0}$) is 399, and the maximum number of isolable faults considering all the sensors available and $d_{\text{cluster}}$ ($\zeta_{\text{best}}$) is 398 (57 % of $N$ = 694 nodes forming the network). According to the specified $d_{\text{cluster}}$ = 200 m and $d_{\text{max}}$ = 888 m, the cost function



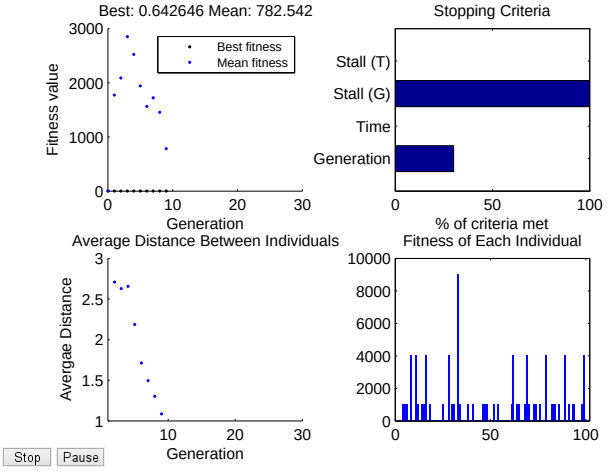

(a) Two sensors




(b) Three sensors




(c) Four sensors



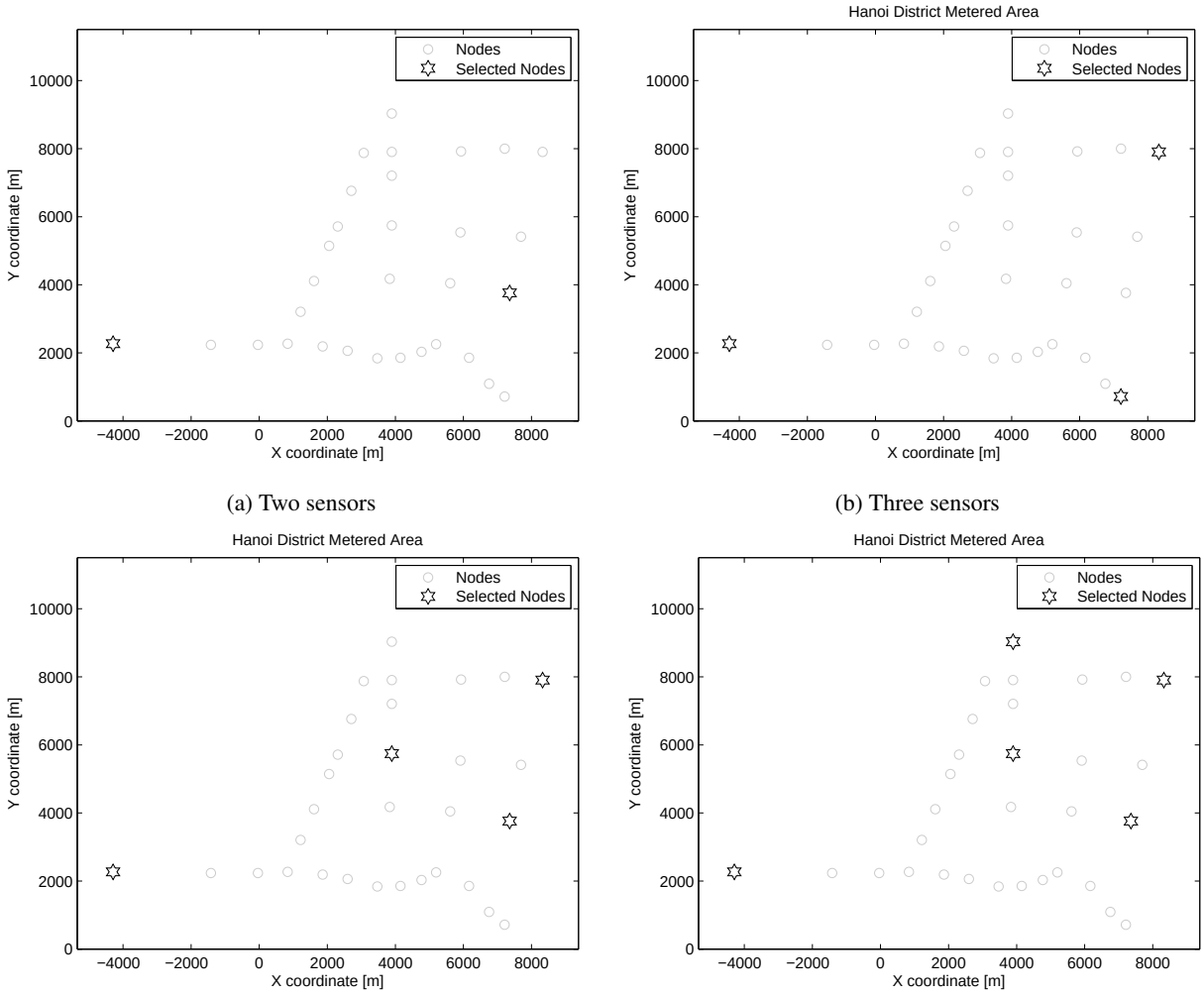

(d) Five sensors

Figure 6: Genetic Algorithms optimisation evolution in Hanoi DMA

parameters have been chosen of $d_c = 8.31$ and $d_f = 1.04$, respectively.

Leak isolation assessment results concerning sensor distribution for different number of sensors (from two to four) are detailed in Table 2. It may be observed how the relaxation by $d_{\text{cluster}}$ does not have much effect when having all the sensors available (i.e. $\phi_{best} \approxeq 1$), but it does for a limited sensor set (see Table 2) e.g. for two sensors, with $\zeta_{\text{opt}} = 267$ and $\zeta_{opt_0} = 116$, $\delta = 38$ % extra coverage over $\zeta_{\text{best}}$ is achieved when geographically relaxing the assessment. It may also be observed how the results obtained between three and four sensors do not improve in terms of $\zeta_{\text{opt}}$, even a better $\rho$ is achieved for four sensors at the optimisation stage. In this case, the benefit of installing extra sensors may obtain reduced isolation clusters, but still bigger than $d_{\text{cluster}}$. Hence, since the coverage of the network is high (97 % of $\zeta_{\text{best}}$), the optimal sensor distribution is obtained for three sensors (Figure 8b), since is the one achieving best $\zeta_{\text{opt}}$ with the minimum number of sensors.

The impact of sensors resolution is also worth to be noted. Although it does not have impact on the maximum number of isolable faults $\zeta_{\text{best}} = 398$ —hence, the maximum achievable coverage is not limited by the sensors resolution but by the topological network setup, when sufficient number of sensors are available—, it does have impact on $\zeta_{\text{opt}}$ for different sensor setups —hence, for limited information gathered from the network, sensor resolution effect is noticeable—. For example, considering five full-resolution sensors setup, almost complete coverage of the network




(a) Two sensors

(b) Three sensors

(c) Four sensors

(d) Five sensors

Figure 7: Sensor placement in Hanoi DMA

is achieved ($\zeta_{\text{opt}} = 395$), against the 388 isolable faults achieved by the five limited-resolution (0.1 m) sensors setup counterpart.

### *5.2.3. Castelldefels Platja DMA*

The sensor placement results obtained considering *Castelldefels Platja* network (Figure 4) are depicted in Figures 9a to 9c. The sensitivity matrix $\mathbf{S}$ is obtained for an emitter coefficient of 0.92 LPS/psi$^{0.5}$, in an hourly sampled scenario comprised between 24/02/2014 9h and 25/02/2014 9h, so $\mathbf{S}$ is concatenated for the 24 hours available and is of dimension 67872 x 2828. Also, the information in this matrix is truncated by sensor resolution (i.e. 0.1 m). The distance used here is the topological distance among nodes, i.e. minimum pipe distance between these elements. For this particular network, the computation of $\zeta_{best_0}$ is not possible due to computational issues related with the network size $N$, as introduced in Section 4. Alternatively, the rank of $\mathbf{S}$ is used, providing a maximum number of isolable faults approximation considering all the sensors available, that is 2828. Since this value is close to $N$, it may be considered a feasible approximation of the maximum number of isolable faults. Also, the maximum number of isolable faults considering all the sensors available and $d_{\text{cluster}}$ ($\zeta_{\text{best}}$) is 2824 (the 99.9 % of $N = 2828$ nodes forming the network). It may be observed how the relaxation by $d_{\text{cluster}}$ does not either have much impact in this DMA when having all the

Table 2: Isolation assessment results, Canyars DMA

| Number of sensors | 2 | 3 | 4 |
|---|---|---|---|
| $\zeta_{\text{opt}}$ | 267 | 388 | 388 |
| % of $N$ | 38 | 56 | 56 |
| % of $\zeta_{\text{best}}$ | 67 | 97 | 97 |
| % of $\zeta_{best_0}$ | 67 | 97 | 97 |
| $\zeta_{opt_0}$ | 116 | 242 | 245 |
| $\rho$ | 0.7375 | 0.7342 | 0.7321 |

Table 3: Isolation results, Castelldefels Platja DMA

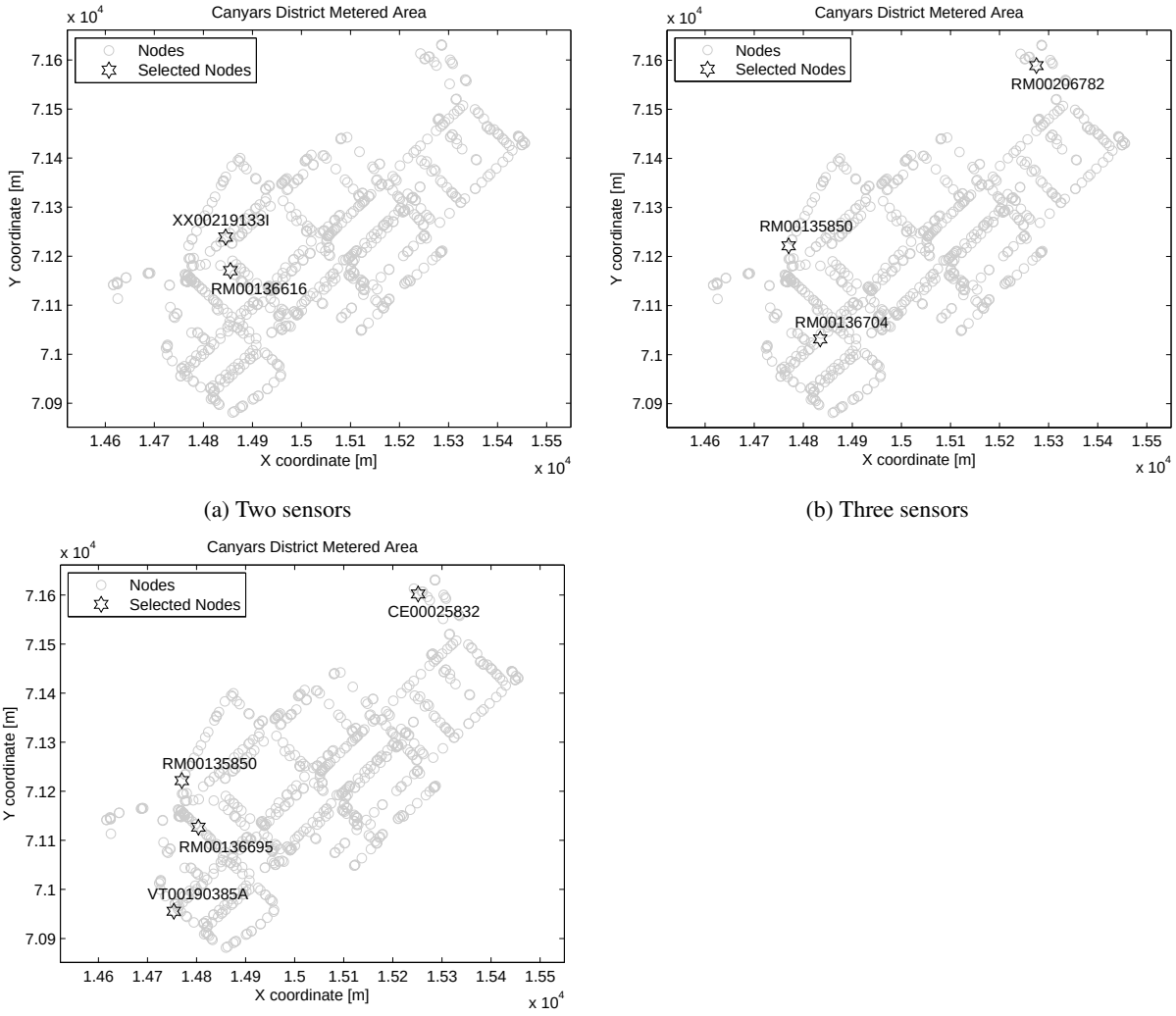
| Number of sensors | 4 | 5 | 6 |
|---|---|---|---|
| $\zeta_{\text{opt}}$ | 2649 | 2665 | 2665 |
| % of $N$ | 93.67 | 94.24 | 94.24 |
| % of $\zeta_{\text{best}}$ | 93.8 | 94.37 | 94.37 |
| % of rank $\mathbf{S}$ | 93.67 | 94.24 | 94.24 |
| $\zeta_{opt_0}$ | 902 | 982 | 1067 |
| $\rho$ | 0.5116 | 0.5086 | 0.5071 |

sensors available (i.e. $\phi_{best} \approxeq 1$), but it does as in Canyars DMA when limited number of sensors are available (see Table 3) e.g. for four sensors, with $\zeta_{\text{opt}} = 2649$ and $\zeta_{opt_0} = 902$, $\delta = 62$ % extra coverage over $\zeta_{\text{best}}$ is achieved when geographically relaxing the assessment. According to the specified $d_{\text{cluster}} = 200$ m and $d_{\max} = 7222$ m, the cost function parameters have been chosen of $d_c = 6.41$ and $d_f = 0.46$, respectively.

Isolation assessment results concerning sensor distribution for different number of sensors (from four to six) are detailed in Table 3. It may be seen how, for five and six sensors, the number of isolable faults for the optimal sensor set ($\zeta_{\text{opt}}$) equals 2665, so according to the criterion presented here, no advantage is obtained from the usage of this extra sensor. Hence, the number of suggested sensors for this network is five (Figure 9b), achieving a theoretical coverage of the 94.24 % of the total possible faults.

## 6. Conclusions

In this paper, a new sensor placement and leak location assessment methodologies have been proposed, in order to improve the performance of leak location in water distribution networks, which may have severe impact on maintenance costs and performance of the water distribution along DMAs. Common problems arising on the leak location in large real water networks can be addressed at the sensor placement stage, e.g. leak discriminability and large location areas, when taking into account real world leak diagnosis trade-offs related with geographic location precision. Hence, a general method of sensor placement is proposed, taking into account these trade-offs by clustering similar leaks geographically, within an acceptable location area from the application point of view. The proposed method achieved promising leak location results —evaluated by an also proposed general assessment method for leak diagnosis in water distribution systems— in a small illustrative DMA in Hanoi and two larger DMAs in the Barcelona urban area. These results motivate the use of the proposed methodology in the actual and similar water networks. Further work involves to consider the number of sensors to install as part of the optimisation problem, as well as to take into account the uncertainty —existing e.g. in the sensor measurements and in the demand model— in the sensor placement algorithm, in order to cope with more realistic assumptions. Also, the extension to multiple leak scenarios may be considered in future steps of this work, by e.g. developing further methods in order to expand the sensitivity matrix accordingly, taking into account that the selection of these new scenarios should be performed carefully, in order to avoid computational issues derived from handling high dimension matrices.




(a) Two sensors

(b) Three sensors

(c) Four sensors

Figure 8: Sensor placement in Canyars DMA

## Acknowledgement

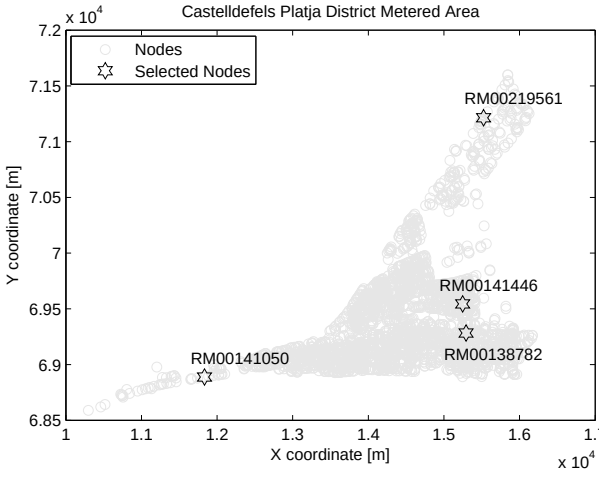
This work has been partially funded by the Spanish Ministry of Science and Technology through the Project ECOCIS (Ref. DPI2013-48243-C2-1-R) and Project HARCRICS (Ref. DPI2014-58104-R), and by EFFINET grant FP7-ICT-2012-318556 of the European Commission. The authors also wish to thank the support received by the Water Technological Center (CETAQUA) of the company managing the Barcelona water network (AGBAR).

## References

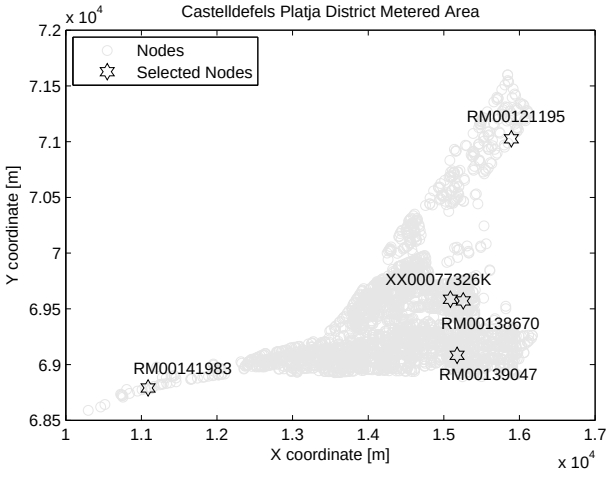
[1] Y. Khulief, A. Khalifa, R. Mansour, M. Habib, Acoustic detection of leaks in water pipelines using measurements inside pipe, Journal of Pipeline Systems Engineering and Practice 3 (2) (2012) 47–54.

[2] M. Lambert, A. Simpson, J. Vitkovsky, X.-J. Wang, P. Lee, A review of leading-edge leak detection techniques for water distribution systems, in: 20th AWA Convention, Perth, Australia, 2003.

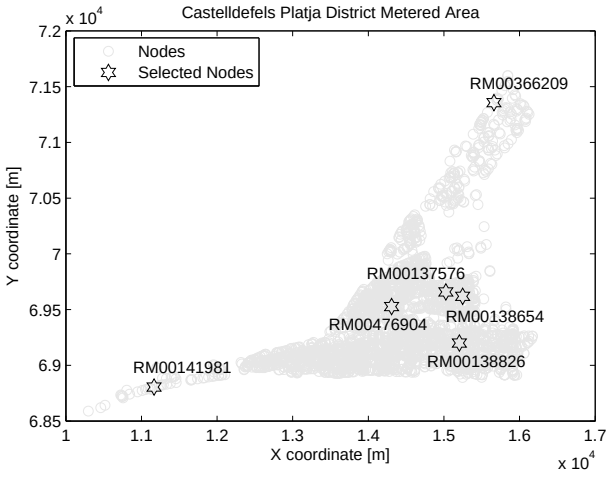
[3] R. Puust, Z. Kapelan, D. A. Savic, T. Koppel, A review of methods for leakage management in pipe networks, Urban Water Journal 7 (1) (2010) 25–45.




(a) Four sensors




(b) Five sensors




(c) Six sensors

Figure 9: Sensor placement in Castelldefels Platja DMA

[4] A. Lambert, What do we know about pressure leakage relationships in distribution systems?, in: IWA Conference System Approach to leakage control and water distribution system management, no. May, Brno, Czech Republic, 2000, pp. 1–9.

[5] J. Thornton, A. Lambert, Progress in practical prediction of pressure: leakage, pressure: burst frequency and pressure: consumption relationships, in: Leakage Conference Proceedings, Halifax, Canada, 2005, pp. 1–10.

[6] A. F. Colombo, P. Lee, B. W. Karney, A selective literature review of transient-based leak detection methods, Journal of Hydro-environment Research (2009) 212–227.

[7] R. S. Pudar, J. A. Liggett, Leaks in pipe networks, Journal of Hydraulic Engineering 118 (7) (1992) 1031–1046.

[8] R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, A. Peralta, Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks, Control Engineering Practice 19 (10) (2011) 1157 – 1167. doi:10.1016/j.conengprac.2011.06.004.

[9] M. V. Casillas, L. Garza-Castañón, V. Puig, Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities, Journal of Hydroinformatics 16 (2014) 649–670.

[10] R. Pérez, M. A. Cugueró, J. Cugueró, G. Sanz, Accuracy assessment of leak localisation method depending on available measurements, Procedia Engineering 70 (2014) 1304–1313. doi:10.1016/j.proeng.2014.02.144.

[11] A. Yassine, S. Ploix, J. M. Flaus, A method for sensor placement taking into account diagnosability criteria, International Journal of Applied Mathematics and Computer Science 18 (4) (2008) 497–512. doi:10.2478/v10006-008-0044-5.

[12] A. Rosich, R. Sarrate, F. Nejjari, Optimal sensor placement for FDI using binary integer linear programming, in: 20th International Workshop on Principles of Diagnosis, DX09, 2009, pp. 235–242.

[13] A. Krause, J. Leskovec, C. Guestrin, J. Vanbriesen, C. Faloutsos, Efficient sensor placement optimization for securing large water distribution networks, Journal of Water Resources Planning and Management 134 (6) (2008) 516–526.

[14] M. Aral, J. Guan, M. Maslia, Optimal design of sensor placement in water distribution networks, Journal of Water Resources Planning and Management 136 (1) (2010) 5–18.

[15] R. Sarrate, F. Nejjari, A. Rosich, Sensor placement for fault diagnosis performance maximization in distribution networks, in: Control & Automation (MED), 2012 20th Mediterranean Conference on, 2012, pp. 110–115.

[16] S. Christodoulou, A. Gagatsis, A. Xanthos, S. Kranioti, A. Agathokleous, M. Fragiadakis, Entropy-based sensor placement optimization for waterloss detection in water distribution networks, Water Resources Management 27 (13) (2013) 4443–4468.

[17] M. Farley, S. Trow, Losses in Water Distribution Networks, IWA publishing, London, 2003.

[18] J. Quevedo, M. A. Cugueró, R. Pérez, F. Nejjari, V. Puig, J. M. Mirats, Leakage location in water distribution networks based on correlation measurement of pressure sensors, in: 8th IWA Symposium on System Analysis and Integrated Assessment (Watermatex), International Water Association (IWA), San Sebastián, 2011, pp. 290–297.

[19] C. R. Reeves (Ed.), Modern Heuristic Techniques for Combinatorial Problems, Mcgraw Hill, UK, 1995.

[20] J. R. Koza, Survey of genetic algorithms and genetic programming, Proceedings of 1995 WESCON Conference (1995) 589–594.

[21] L. A. Rossman, EPANET 2 Users Manual, Environmental Protection Agency (EPA), U.S. (September 2000).

[22] L. Rokach, O. Maimon, Clustering methods, in: Data Mining and Knowledge Discovery Handbook, 2005, pp. 321–352.

[23] K. Gallagher, M. Sambridge, Genetic algorithms: A powerful tool for large-scale nonlinear optimization problems, Computers & Geosciences 20 (7-8) (1994) 1229–1236.

[24] T. Fawcett, An introduction to ROC analysis, Pattern Recognition Letters 27 (8) (2006) 861–874.

[25] M. V. Casillas-Ponce, L. E. Garza-Castañón, V. Puig-Cayuela, Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities, Journal of Hydroinformatics 16 (3) (2014) 649–670. doi:10.2166/hydro.2013.019.

[26] T. Walski, D. Chase, D. Savic, W. Grayman, S. Beckwith, E. Koelle, Advanced Water Distribution Modeling and Management, Haestad Press, 2003.