Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 953 – 962

Procedia Engineering

www.elsevier.com/locate/procedia

Optimal pressure sensor placement in water distribution networks minimizing leak location uncertainty

Fatiha Nejjari[a,*], Ramon Sarrate[a], Joaquim Blesa[b]

[a]*Automatic Control Department, Universitat Politècnica de Catalunya, Rambla de Sant Nebridi, 10, 08222 Terrassa, Spain.*
[b]*Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Parc Tecnològic de Barcelona. C/ Llorens i Artigas 4-6. 08028-Barcelona, Spain.*

Abstract

In this paper an optimal sensor placement strategy based on pressure sensitivity matrix analysis and an exhaustive search strategy that maximizes some diagnosis specifications for a water distribution network is presented. An average worst leak expansion distance as a new leak location performance measure has been proposed. This metric is later used to assess the leak location uncertainty provided by a sensor configuration. The method is combined with a clustering technique in order to reduce the size and the complexity of the sensor placement problem. The strategy is successfully applied to determine the location of a set of pressure sensors in a district metered area in the Barcelona water distribution network.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: sensor placement; leak location; clustering.

1. Introduction

Water leakage in water distribution networks (WDNs), which are already stressed by growing water demand, causes significant loss of water. Leaks in WDNs, can happen sometimes due to damages and defects in pipes, lack of maintenance or increase in pressure. Leaks must be detected and located as soon as possible to minimize their effects. Methods for locating leaks range from ground-penetrating radar to acoustic listening devices [1]. However, techniques based on locating leaks from pressure/flow monitoring devices allow a more effective and less costly search in situ. The need to identify the location of leaks has promoted the development of several techniques based on inverse problems and solving it using pressure or flow measurements. These techniques are based on the sensors installed in the network.

Ideally, a sensor network should be configured to facilitate leak detection and location and maximize diagnosis performance under a given sensor cost limit. Due to budget constraints, it is obvious that only a limited number of sensors can be installed in WDNs. Since improper selections may seriously hamper diagnosis performance, the development of sensor placement strategies has become an important research issue in recent years.

Some contributions have been done regarding leak location. Some of them are based on the fault sensitivity matrix [2,3], which contains the information about how leaks affect the different node pressures. On the other hand, optimal

* Corresponding author. Tel.: +34-93-739-86-19 . ; fax: +34-93-739-86-28 .
E-mail address: fatiha.nejjari@upc.edu

1877-7058
Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.979

pressure sensor placement algorithms that use the sensitivity matrix have been developed to determine which pressure sensors have to be installed among hundreds of possible locations in the WDN to carry out an optimal leak location as in [4,5].

The fault sensitivity matrix can be obtained by convenient manipulation of model equations as long as leak effects are included in them [6]. Alternatively, it can be obtained by sensitivity analysis through simulation [2]. The elements of this matrix depend on the operating point defined by the heads in reservoirs, the inflow, demand distribution, which is not constant, and the leak magnitudes, which are unknown.

In this paper, a methodology is proposed for the sensor placement problem in WDNs, through the fault sensitivity matrix concept. The approach consists in solving a combinatorial optimization problem where the best leak location performance can be achieved by installing a given number of sensors.

This problem was already considered in [7] by maximizing the leak locatability index and a two-step hybrid methodology that combined clustering techniques with an exhaustive search was proposed. The clustering problem has been addressed by researchers in many contexts and disciplines [8]. It is a mature and active research area [9] and many efficient clustering algorithms have been developed in the literature.

In this paper, new leak location performance metrics that take into account the geographical coordinates of every leak node, are defined. The goal of these metrics is to assess the uncertainty degree of a leak location diagnosis in terms of distance among non-discernable leak locations. The methodology proposed in [7] will be recalled and adapted to minimize these distance measures.

This work focuses on the placement of pressure monitoring points as they are more frequently used than flow rate sensors. Collecting pressure data is cheaper and easier, and the pressure transducers give instantaneous readings whereas most flow meters do not react instantaneously to flow changes [10]. Therefore, only pressure sensors will be considered in the sensor placement problem.

The paper is organized as follows: Section 2, introduces the model-based fault diagnosis applied to leak detection and location. The sensor placement problem tackled in this paper is formally presented in Section 3. In Section 4, the sensor placement methodology is applied to a real DMA network in Barcelona. Finally, some conclusions and remarks are given in Section 5.

2. Fault diagnosis principles

2.1. Model-based fault diagnosis

Model-based fault diagnosis techniques are applied to detect and locate leaks in WDNs. In model-based fault diagnosis [11] a set of residuals are designed based on a process model. Fault detection and isolation is achieved through the evaluation of residual expressions under available measurements. A threshold-based test is usually implemented in order to cope with noise and model uncertainty effects. In the absence of faults, all residuals remain below their given thresholds. Otherwise, when a fault is present the model is no longer consistent with the observations (known process variables). Thus, some residuals will exceed their corresponding thresholds, signalling the occurrence of a fault.

Residual fault sensitivities are a key issue for fault diagnosis. Given a set of m target faults $f_j \in \mathbf{F}$ and a set of n residuals $r_i \in \mathbf{R}$, residual fault sensitivities are collected in the *Fault Sensitivity Matrix* (FSM), $\mathbf{\Omega}$

$$\mathbf{\Omega} = \begin{pmatrix} \frac{\partial r_1}{\partial f_1} & \cdots & \frac{\partial r_1}{\partial f_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_n}{\partial f_1} & \cdots & \frac{\partial r_n}{\partial f_m} \end{pmatrix}. \tag{1}$$

A fault can be detected as long as there exists at least a residual sensitive to it. However, isolating faults requires more than one residual being sensitive to them. Fault isolation is achieved by matching the evaluated residual vector pattern to the closest residual fault sensitivity vector pattern (i.e., FSM column vector).

2.2. Leak detection and location

The FSM can be obtained by convenient manipulation of model equations as long as leak effects are included in them [6]. Alternatively, it can be obtained by sensitivity analysis through simulation [2]. The latter approach is used in the present paper and just primary residuals are regarded. Primary residuals are obtained by comparing each actual pressure measurement p_i to the corresponding estimated value in the fault free case $\hat{p}_{i0}$

$$r_i = p_i - \hat{p}_{i0} \tag{2}$$

A model of the WDN is used by a simulation engine to produce the estimated node pressure. An approximate procedure to obtain the FSM involves using as well the simulator to estimate pressure measurements $\hat{p}_{ij}$ for every node i under fault condition f_j

$$\mathbf{\Omega} = \begin{pmatrix} \hat{p}_{11} - \hat{p}_{10} & \cdots & \hat{p}_{1m} - \hat{p}_{10} \\ \vdots & \ddots & \vdots \\ \hat{p}_{n1} - \hat{p}_{n0} & \cdots & \hat{p}_{nm} - \hat{p}_{n0} \end{pmatrix}. \tag{3}$$

Thus, every FSM column corresponds to an estimation of the residual vector in every leak condition. The same nominal leak magnitude is assumed in all simulations. This leak magnitude is not considered in the FSM since it has a scaling factor roll.

Sometimes a binary version of the FSM is used in the leak location procedure [2]. Then, leak location is achieved by looking for the smallest Hamming distance between FSM columns and the binarized actual residual vector. This and several other alternative leak location methods are compared in [3].

In the present paper, a projection based method is considered. Let $\mathbf{r} = [r_1 \cdots r_n]^T$ be the actual residual vector corresponding to all pressure measurement points, and $\boldsymbol{\omega}_{\bullet j}$ be the column of $\mathbf{\Omega}$ corresponding to leak j. Then, leak location is achieved by solving the problem

$$\arg\max_j \frac{\boldsymbol{\omega}_{\bullet j}^T \cdot \mathbf{r}}{||\boldsymbol{\omega}_{\bullet j}||||\mathbf{r}||}, \tag{4}$$

where $||\mathbf{v}||$ stands for the Euclidean norm of vector $\mathbf{v}$. Thus, the biggest normalized projection of the actual residual vector on the fault sensitivity space is sought.

In order to assess the leak detectability and locatability properties, the detectable leak set $\mathbf{F}_D$ and the leak locatability index I were defined in [7].

Given a set of residuals $r_i \in \mathbf{R}$, a set of leaks $f_j \in \mathbf{F}$ and the corresponding leak (fault) sensitivity matrix $\mathbf{\Omega}$, the set of detectable leaks $\mathbf{F}_D$ was defined as

$$\mathbf{F}_D = \{f_j \in \mathbf{F} : \exists r_i \in \mathbf{R} : |\omega_{ij}| \geq \epsilon\}, \tag{5}$$

where ϵ is a threshold to account for noise and model uncertainty. The locatability index I was also defined as

$$I = \sum_{(f_k, f_l) \in \mathbb{F}} 1 - \frac{\boldsymbol{\omega}_{\bullet k}^T \cdot \boldsymbol{\omega}_{\bullet l}}{||\boldsymbol{\omega}_{\bullet k}||||\boldsymbol{\omega}_{\bullet l}||}, \tag{6}$$

where $\mathbb{F} = \{(f_k, f_l) \in \mathbf{F} \times \mathbf{F} : k < l\}$.

Given the leak location criteria defined in Eq. (4), the leak locatability index aggregates the normalized projection degree between the residual leak sensitivity vectors for all leak pairs. Since a minimal normalized projection is desired, the greater the index is, the better it is.

In order to better evaluate the leak locatability performance of a diagnosis system, the uniform projection angle $\bar{\alpha}$ was defined as

$$\bar{\alpha} = \arccos(1 - \frac{I}{\binom{|\mathbf{F}|}{2}}), \tag{7}$$

where, $|\mathbf{A}|$ stands for the cardinality of set $\mathbf{A}$.

$\bar{\alpha}$ provides a reference value for the angle between any pair of leak sensitivity vectors in the FSM, assuming a uniform contribution to the leak locatability index.

In [7] a sensor placement strategy that maximized the locatability index (6) was proposed. The resulting sensor locations led to a maximal uniform projection angle $\bar{\alpha}$. In an ideal case, all pairs of leak sensitivity vectors in the FSM should satisfy this uniform projection angle. This uniform angular separation between leak pairs would allow for a successful leak location method applying (4), even when residuals are affected by modeling errors, sensor noise and other uncertainties.

Nevertheless, in a real case the angle between leak pairs is not uniformly distributed. Some leaks can have similar leak sensitivity vectors, which introduces uncertainty in the leak location results when applying (4). This can become a critical issue for water network utilities, especially when this uncertainty involves distant leak locations, i.e. two distant leaks that have similar leak sensitivity vectors. So, distances between nodes with a similar leak sensitivity vector should be considered in the optimal sensor placement methodologies. In order to take into account these distances, the following properties are defined.

Definition 1 (Leak expansion set)**.** Given a leak $f_j \in \mathbf{F}$ and a projection angle threshold α_{th}, the leak expansion set $\mathbf{F}_j^{\alpha_{th}}$ is defined as

$$\mathbf{F}_j^{\alpha_{th}} = \{f_i \in \mathbf{F} : \frac{\boldsymbol{\omega}_{\bullet j}^T \cdot \boldsymbol{\omega}_{\bullet i}}{||\boldsymbol{\omega}_{\bullet j}||||\boldsymbol{\omega}_{\bullet i}||} > \cos(\alpha_{th})\}. \tag{8}$$

$\mathbf{F}_j^{\alpha_{th}}$ contains the set of leaks whose correlation with leak f_j is bigger than $\cos(\alpha_{th})$. If $f_i \in \mathbf{F}_j^{\alpha_{th}}$, it follows that $f_j \in \mathbf{F}_i^{\alpha_{th}}$.

Definition 2 (Correlated leak pairs ratio)**.** Given the leak expansion sets $\mathbf{F}_j^{\alpha_{th}}$ $j = 1, ..., |\mathbf{F}|$, the correlated leak pairs ratio $\eta^{\alpha_{th}}$ is defined as

$$\eta^{\alpha_{th}} = 100 \frac{\sum_{j=1}^{|\mathbf{F}|} |\mathbf{F}_j^{\alpha_{th}}| - |\mathbf{F}|}{2\binom{|\mathbf{F}|}{2}}. \tag{9}$$

$\eta^{\alpha_{th}}$ provides the percentage of leak pairs from $\mathbf{F}$ whose mutual correlation is bigger than $\cos(\alpha_{th})$.

Definition 3 (Leak node distance matrix)**.** Given the geographical coordinates of every leak node, the leak node distance matrix $\mathbf{D} \in \Re^{|\mathbf{F}|\times|\mathbf{F}|}$ is defined as the matrix whose coefficients $d_{i,j}$ are the geographical distance between nodes i and j.

Matrix $\mathbf{D}$ is a symmetric matrix ($d_{i,j} = d_{j,i}$), with diagonal coefficients equal to zero ($d_{i,i} = 0$). This matrix will be used to compute distances in leak expansion sets.

Definition 4 (Worst leak expansion distance)**.** Given a leak expansion set $\mathbf{F}_j^{\alpha_{th}}$ and the leak node distance matrix $\mathbf{D}$, the worst leak expansion distance $R_j^{\alpha_{th}}$ is defined as

$$R_j^{\alpha_{th}} = \max_{f_i \in \mathbf{F}_j^{\alpha_{th}}} d_{i,j}, \tag{10}$$

$R_j^{\alpha_{th}}$ provides the maximum Euclidian distance between the node of leak f_j and the nodes of leaks whose correlation with leak f_j is bigger than $\cos(\alpha_{th})$. This metric is next used to compute the following overall leak location uncertainty index in terms of leak node distances.

Definition 5 (Average worst leak expansion distance)**.** Given a set of leaks $\mathbf{F}$ and a threshold projection angle α_{th}, leak expansion sets $\mathbf{F}_j^{\alpha_{th}}$ with $j = 1, ..., |\mathbf{F}|$ can be computed applying (8). Then, the average worst leak expansion distance can be computed as

$$\bar{R}^{\alpha_{th}} = \sum_{j=1}^{|\mathbf{F}|} R_j^{\alpha_{th}}, \tag{11}$$

$\bar{R}^{\alpha_{th}}$ provides the average of the worst leak expansion distances considering all the possible leaks in **F**.

In the ideal case, all pairs of leak sensitivity vectors would fulfil the uniform projection angle $\bar{\alpha}$. Then, provided $\alpha_{th} \le \bar{\alpha}$, the leak expansion sets would be

$$\mathbf{F}_j^{\alpha_{th}} = f_j \qquad j = 1, ..., |\mathbf{F}|, \tag{12}$$

which means that the leak expansion sets only contain the nominal leak and therefore $\eta^{\alpha_{th}} = 0$, $R_j^{\alpha_{th}} = 0 \quad \forall j = 1, ..., |\mathbf{F}|$ and $\bar{R}^{\alpha_{th}} = 0$. However, in a real case the leak expansion sets will contain more than nominal leaks, so the average worst leak expansion distances should be minimized by efficient optimal sensor placement methodologies.

3. Sensor placement methodology

3.1. Problem statement

The sensor placement problem for leak location involves solving a combinatorial optimization problem where the best leak location performance that can be achieved installing the cheapest number of sensors is sought. Usually, a budget constraint is imposed by the water utility, which must be taken into account in the sensor placement methodology. In this paper, the approach followed in [5] is adapted to the new performance index defined in Eq. (11).

Let **S** be the candidate pressure sensor set and m_p the maximum number of pressure sensors that can be installed in the network according to the budget constraint. Although fewer pressure sensors could be installed in the network, in this work it is assumed that m_p sensors should be installed. Then, the problem can be roughly stated as the choice of a configuration of m_p pressure sensors in **S** such that the best diagnosis performance is attained. This diagnosis performance depends on the set of sensors installed in the network $S \subseteq \mathbf{S}$ and will be stated in terms of the detectable leak set and the average worst leak expansion distance, i.e., $\mathbf{F}_D(S)$ and $\bar{R}^{\alpha_{th}}(S)$.

To solve the sensor placement problem, some network model information is also required. On the one hand, the leak node distance matrix **D** is assumed to be previously obtained based on the geographical coordinates of every leak node. On the other hand, the leak sensitivity matrix **Ω** corresponding to the complete set of candidate sensors is assumed to be previously computed, following the methodology described in Section 2.2. Hence, the optimal sensor placement for leak diagnosis can be formally stated as follows:

GIVEN a candidate sensor set **S**, a leak node distance matrix **D**, a leak sensitivity matrix **Ω**, a leak set **F**, a desired projection angle threshold α_{th} and a number m_p of pressure sensors to be installed.

FIND the m_p-pressure sensor configuration $S \subseteq \mathbf{S}$ such that:

1. all leaks in **F** are detectable, $\mathbf{F}_D(S) = \mathbf{F}$, and
2. the average worst leak expansion distance is minimized, i.e. $\bar{R}^{\alpha_{th}}(S) \le \bar{R}^{\alpha_{th}}(S^\star)$ for any $S^\star \subseteq \mathbf{S}$ such that $|S^\star| = m_p$.

As this optimization problem cannot be solved by efficient branch and bound search strategies, a suboptimal two-step hybrid methodology that combines clustering techniques with an exhaustive search is followed:

Step 1 Clustering techniques are applied to reduce the initial set of candidate sensors **S** to **S′**, such that next step is tractable.
Step 1 will be described in detail in the next section.

Step 2 An exhaustive search is applied to the reduced candidate sensor set **S′**. This search implies that the diagnosis performance must be evaluated $\binom{|\mathbf{S}'|}{m_p}$ times. The most time demanding test concerns the evaluation of the average worst leak expansion distance which involves computing $\binom{|\mathbf{F}|}{2}$ times the normalized projection of the leak sensitivity vectors. Thus, in all, an exhaustive search is of factorial time complexity, but an optimal solution for the given reduced candidate sensor set is guaranteed.

3.2. Candidate sensor set reduction

In [7], a reduction in the number of candidate sensors has been proposed by grouping the n initial candidate sensors into ℓ groups (clusters) applying the Evidential C-Means (ECM) algorithm [12].

Given a set of objects $\mathbf{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n_e}\}$, clustering consists in partitioning the n_e observations into ℓ sets $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \cdots, \mathbf{C}_\ell\}$ ($\ell \leq n_e$) in such a way that objects in the same group (i.e. cluster) are more similar (in some sense) to each other than those in other groups.

In this work, the criterion used for determining the similitude between elements (sensors) is the sensitivity pattern of their primary residuals to leaks. In particular, according to the procedure described in Section 2.2, this is provided by every row j of the leak sensitivity matrix $\boldsymbol{\Omega}$ defined in Eq. (1). Let $\boldsymbol{x}_j = \frac{\omega_{j\bullet}}{\|\omega_{j\bullet}\|}$, $j = 1, ..., n_e$, where $\omega_{j\bullet}$ be the jth row vector of matrix $\boldsymbol{\Omega}$, $\boldsymbol{x}_j$ the normalized vector of $\omega_{j\bullet}$ and n_e the number of rows of $\boldsymbol{\Omega}$ i.e. $n_e = n = |\mathbf{S}|$. Applying the ECM algorithm defined in [12], a set of ℓ clusters defined by their centroids $\boldsymbol{\mu}_i$ ($i = 1, \ldots, \ell$) are obtained. Additionally, the plausibility matrix $\boldsymbol{\Pi}$ ($n \times \ell$), which provides the membership degree of every element to every cluster, is obtained.

$$\boldsymbol{\Pi} = \begin{pmatrix} pl_1(C_1) & \cdots & pl_1(C_\ell) \\ \vdots & \ddots & \vdots \\ pl_{n_e}(C_1) & \cdots & pl_{n_e}(C_\ell) \end{pmatrix} \tag{13}$$

$pl_i(C_k)$ represents the plausibility (or the possibility) that an object $\mathbf{x}_i$ belongs to cluster C_k. A hard partition can be easily obtained by assigning each object to the cluster with the highest plausibility, i.e

$$\mathbf{g}(i) = \arg\max_j pl_i(C_j) \quad i = 1, \cdots, n_e \tag{14}$$

where $\mathbf{g}$ is the vector that contains the cluster membership of the n_e elements.

Once the set of sensors has been divided into clusters $\mathbf{C}_1, \ldots, \mathbf{C}_N$, N representative sensors will be selected for each cluster, setting up a new candidate sensor set $\mathbf{S}'$ of $N\ell$ elements ($N\ell \leq n$). The number of groups ℓ will be determined by means of a study of the evolution of the validity index provided by the ECM algorithm for different number of groups. Finally, the number N ($N \geq 1$) will be given by

$$N = \left\lceil \frac{n_r}{\ell} \right\rceil \tag{15}$$

where n_r is the desired cardinality of the reduced candidate sensor set and $\lceil\ \rceil$ denotes the nearest integer in the direction of positive infinity.

Let $\mathbf{pl}_i$ be the plausibility vector corresponding to the elements of the cluster set $\mathbf{C}_i$, $\mathbf{Q}_i$ provide a mapping between the sensor index in $\mathbf{C}_i$ and its corresponding row in the sensitivity matrix defined in Eq. (1), and $\mathbf{modw}_i$ be the euclidean norm of this sensitivity matrix row. Algorithm 1 provides the vector $\mathbf{Q}_i^0$ with N representative elements (sensors) of the cluster $\mathbf{C}_i$: $\mathbf{Q}_i^0(1), \ldots, \mathbf{Q}_i^0(N)$. The greater the value of N, the more representative the elements $\mathbf{Q}_i^0$ of the set $\mathbf{C}_i$ are. In addition to the plausibility values, the algorithm takes into account the euclidean norm of the sensitivity matrix sensor rows, so that the sensor candidates that maximize the leak detectability defined in (5) are obtained.

4. Application to a real WDN

4.1. DMA case study

To demonstrate the effectiveness and efficiency of the proposed sensor placement methodology, the whole approach is applied to a DMA located in Barcelona city with 883 nodes, 927 pipes and two inflow inputs modeled as reservoir nodes.

In the network, we consider the existence of 448 potential leaks, corresponding to dummy nodes, that should be detected and located. In order to reduce the problem complexity, a subset of node pressures, corresponding to 311 nodes with demand, is chosen as the candidate sensor set. It is also assumed that there is no sensor already installed in

Algorithm 1 N most representative Cluster elements.

```
Algorithm N-most-representative(pl_i, Q_i, modw_i)
  tempw_i ← modw_i
  pl_i^min ← min(pl_i)
  pl_i^max ← max(pl_i)
  n_i ← length(pl_i)
  for j = 1, ..., N do
    for k = 1, ..., n_i do
      if (pl_i(k) < pl_i^min + (j−1)(pl_i^max − pl_i^min)/N)
        tempw_i(k) ← 0
      end if
    end for
    loc = arg max_k tempw_i(k)
    Q_i^0(j) = Q_i(loc)
    tempw_i(loc) ← 0
  end for
  return Q_i^0
end Algorithm
```

the network before solving the sensor placement problem. The total inflow is distributed using a constant coefficient in each consumption node according to the total demand which is estimated using demand patterns.

A fault sensitivity matrix has been obtained using the EPANET hydraulic simulator [13]. Given a set of boundary conditions (such as water demands) EPANET software has been firstly used to estimate the steady-state pressure at the 311 candidate sensor nodes. Next, 448 leaks have been simulated in the dummy nodes and the steady-state pressure has been estimated again in the 311 candidate sensor nodes. Finally, a fault sensitivity matrix has been obtained as the pressure difference between the leak free case and each leakage scenario, according to the procedure described in Section 2. In this work, the leak sensitivity matrix has been computed for a nominal leak magnitude of 1.5 lps (liters per second).

4.2. Sensor placement analysis

Assume that the water distribution company imposes a budget constraint that allows for the installation of up to 5 pressure sensors. Hence, 5 pressure sensors should be chosen out of 311 such that all leaks are detectable and the average worst leak expansion distance is minimized. Recall from Section 3.1 that an exhaustive search is of factorial time complexity. So, clustering techniques will be applied to set up a reduced set of 25 candidate pressure sensors. With this new setup, time complexity will be reduced $\binom{311}{5}/\binom{25}{5} \approx 440000$ times, which seems reasonably promising.

In order to reduce the number of candidate pressure sensors from 311 to $n_r = 25$, clustering techniques have been applied to the data set (311 normalized rows of the sensitivity matrix $\mathbf{\Omega}$) as described in Section 3.2. First, ECM clustering algorithm [12] has been used to classify the data set in $\ell = 5$ clusters.

Provided the plausibility matrix (13) obtained from the clustering algorithm, a hard partition has been done that assigns each element to its highest plausibility cluster, according to (14). Fig. 3 depicts in different colors the 5 different network node clusters, where the closest nodes to the centroid have been highlighted in every cluster. Finally, Algorithm 1 has been applied to obtain the most N representative sensors of every cluster, with $N = 5$ given by (15). The resulting reduced set $\mathbf{S}'$ with $|\mathbf{S}'| = N \times \ell = 25$ candidate pressure sensor places suggested by the clustering approach is displayed in Fig. 3 as blue circled nodes.

The exhaustive search described in Section 3.1 requires the projection angle threshold to be specified. Table 1 provides a solution of the sensor placement problem for several projection angle thresholds. Case 6 corresponds to setting the projection angle threshold to the uniform projection angle (7) that corresponds to the sensor placement solution found when maximizing the locatability index. Indeed, this uniform projection angle could be conceived as an upper bound on the projection angle threshold.

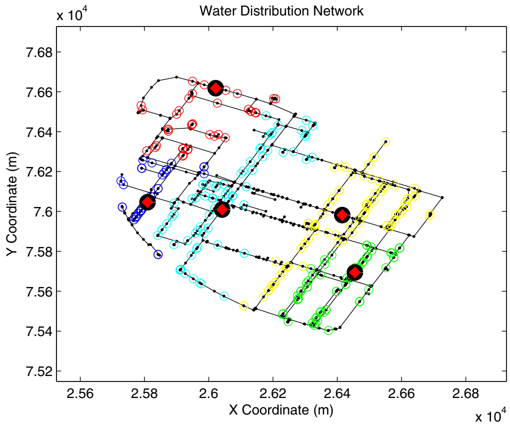

Fig. 1. Clustering results

Table 1. Sensor placement dependency on the projection angle threshold.

Case	α_{th}	S	$\bar{R}^{\alpha_{th}}$ (m)	$\eta^{\alpha_{th}}$ (%)
1	10	$\{3, 8, 67, 89, 207\}$	425.48	15.70
2	20	$\{8, 67, 79, 207, 271\}$	586.27	31.39
3	30	$\{8, 67, 79, 207, 271\}$	677.96	43.27
4	40	$\{8, 67, 207, 249, 271\}$	763.55	52.29
5	50	$\{8, 67, 183, 249, 271\}$	816.80	60.57
6	64.77	$\{8, 10, 159, 207, 271\}$	931.79	69.97

Remark that the greater the threshold, the greater the uncertainty in terms of leak expansion distance and number of correlated leak pairs. Probably the choice of this threshold should take into account implementation requirements of the leak location software module, as well as practical issues concerning the water utility maintenance procedures. On the one hand, the leak location software module will have to deal with sensor measurement noise and network modeling uncertainty. Therefore, the bigger the threshold, the better the performance of the leak location procedure.

On the other hand, the smaller the leak location result uncertainty, the better for the water utility maintenance department. Indeed, upon the occurrence of a leak, the leak location software module will provide a set of leak node candidates to the maintenance department, which then will undergo leak field-testing. Thus, the smaller the leak expansion distance the better, which involves, according to Table 1, specifying a smaller projection angle threshold. Therefore, a tradeoff exists between both criteria.

In order to find a good balanced solution, Figure 2 has been generated. It displays the average worst expansion distance that corresponds to every optimal sensor configuration solution in Table 1, evaluated for all $\alpha_{th} \in \{10, 20, 30, 40, 50, 64.77\}$.

As the optimal sensor configuration that corresponds to Case 4 seems to provide a balanced overall performance for all projection angle threshold values, $S = \{8, 67, 207, 249, 271\}$ has been chosen as a solution to the sensor placement problem. The resulting set of 5 pressure sensors are displayed as red starred nodes in Figure 3. Installing these pressure sensors, all 448 leaks are detectable and the overall average worst leak expansion distance is minimal.

Table 2 provides a comparison between the solution found when minimizing the average worst leak expansion distance and the one found when maximizing the locatability index, according to (6).

As expected the new criteria provides smaller leak location uncertainty in terms of leak expansion distance and number of correlated leak pairs. The expected uniform projection angle decreases, though. But it still seems sufficiently high.

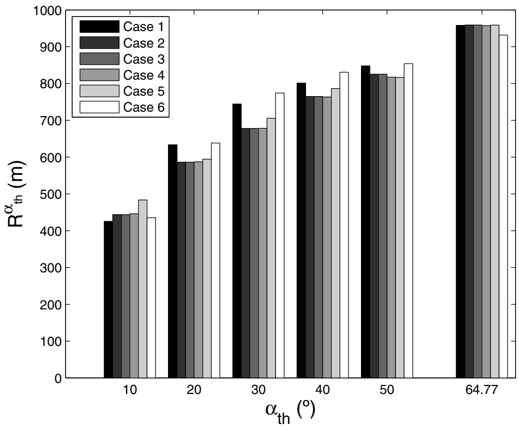

Fig. 2. Comparative average worst expansion distances for all cases.

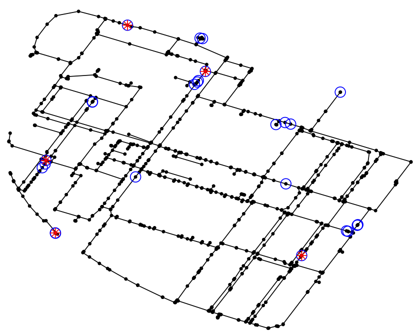

Fig. 3. DMA network sensor placement results

5. Conclusions

This work presents an optimal sensor placement strategy based on pressure sensitivity matrix analysis and an exhaustive search strategy that maximizes some diagnosis specifications for a water distribution network. The methodology has been combined with a clustering technique to reduce the size and the complexity of the sensor placement problem.

An average worst leak expansion distance has been proposed as a new leak location performance measure. This metric is later used to assess the leak location uncertainty provided by a sensor configuration. A comparison to an existent locatability index is also given. The new approach provides a smaller leak location uncertainty in terms of leak expansion distance and number of correlated leak pairs. The new performance measure depends on the choice of the projection angle threshold. Some guidelines are also provided to choose it accounting for software implementation requirements as well as water utility maintenance routines.

Table 2. Sensor placement results comparison.

Criteria	S	I	$\bar{\alpha}$ (°)	$\bar{R}^{10}$ (m)	η^{10} (%)	$\bar{R}^{20}$ (m)	η^{20} (%)	$\bar{R}^{30}$ (m)	η^{30} (%)	$\bar{R}^{40}$ (m)	η^{40} (%)	$\bar{R}^{50}$ (m)	η^{50} (%)	$\bar{R}^{64.77}$ (m)	$\eta^{64.77}$ (%)
min $\bar{R}^{40}$	$\{8, 67, 207, 249, 271\}$	49810	57.09	445.99	15.67	587.37	31.43	678.28	43.27	763.55	52.29	817.40	59.53	957.27	68.83
max I	$\{8, 9, 136, 207, 249\}$	57448	61.75	696.51	35.20	850.93	47.44	905.26	54.35	918.68	59.65	945.61	66.27	961.10	70.34

A secondary contribution of this work is the proposal of a new algorithm to choose the most representative sensor candidates in the clustering procedure.

As a future work, genetic algorithms or other types of optimization methods that provide some guarantee regarding the solution optimality will be investigated. In order to better assess the sensor placement methodology, future work will also target the application of a leak detection and location strategy in a DMA provided with the chosen sensors.

Acknowledgements

This work has been funded by the Spanish Ministry of Economy and Competitiveness through the CICYT projects ECOCIS (ref. DPI2013-48243- C2-1-R) and HARCRICS (ref. DPI2014-58104-R), and by the European Commission through contract EFFINET (ref. FP7-ICT2011-8-318556).

References

[1] T. Farley, Losses in Water Distribution Networks, IWA Publishing UK, UK, 2003.

[2] 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 (2011) 1157–1167.

[3] M. V. Casillas, L. E. Garza, V. Puig, Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks, in: 8^{th} IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, Mexico City, Mexico, 2012, pp. 570–575.

[4] M. V. Casillas, V. Puig, L. E. Garza, A. Rosich, Optimal sensor placement for leak location in water distribution networks using genetic algorithms, Journal of Process Control 13 (2013) 14984–15005.

[5] R. Sarrate, J. Blesa, F. Nejjari, J. Quevedo, Sensor placement for leak detection and location in water distribution networks, Water Science and Technology: Water Supply 14 (2014) 795–803.

[6] J. Blesa, V. Puig, J. Saludes, Robust identification and fault diagnosis based on uncertain multiple input–multiple output linear parameter varying parity equations and zonotopes, Journal of Process Control 22 (2012) 1890–1912.

[7] R. Sarrate, J. Blesa, F. Nejjari, Clustering techniques applied to sensor placement for leak detection and location in water distribution networks, in: 22^{th} Mediterranean Conference on Control and Automation, Palermo, Italy, 2014, pp. 109–114.

[8] A. Jain, M. Murty, P. J., Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323.

[9] R. Xu, D. Wunsch, Survey of clustering algorithms, IEEE Trans. Neur. Net. 16 (2005) 645–678.

[10] W. B. F. de Schaetzen, G. A. Walters, D. A. Savic, Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms, Urban Water 2 (2000) 141–152.

[11] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, Diagnosis and Fault-Tolerant Control, 2nd ed., Springer, 2006.

[12] M. Masson, T. Denoeux, Ecm: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (2008) 1384–1397.

[13] L. A. Rossman, EPANET 2 Users Manual, U.S. Environmental Protection Agency, 2000. URL: `http://www.epa.gov/nrmrl/wswrd/dw/epanet.html`.