Clustering Techniques Applied to Sensor Placement for Leak Detection and Location in Water Distribution Networks

Ramon Sarrate, Joaquim Blesa and Fatiha Nejjari

Abstract— **This work presents an optimization strategy that maximizes the leak locatability performance of water distribution networks (WDN). The goal is to characterize and determine a sensor configuration that guarantees a maximum degree of locatability while the sensor configuration cost satisfies a budgetary constraint. The method is based on pressure sensitivity matrix analysis and an exhaustive search strategy. In order to reduce the size and the complexity of the problem the present work proposes to combine this methodology with clustering techniques. The strategy developed in this work is successfully applied to determine the optimal set of pressure sensors that should be installed in a district metered area (DMA) in the Barcelona WDN.**

I. INTRODUCTION

Leaks and abnormal situation diagnosis is of great importance for distribution network systems. It represents an important factor for quality service, in water distribution networks (WDN). In these systems, it is obvious that only a limited number of sensors can be installed due to budget constraints. Since improper selections may seriously hamper diagnosis performance, the development of sensor placement strategies has become an important research issue in recent years. In particular, leaks in WDNs are an issue of great concern for water utilities. Continuous improvements in water loss management are being applied, and new technologies are developed to achieve higher levels of efficiency [1].

Ideally, a sensor network should be configured to facilitate leak detection and maximize diagnosis performance under a given sensor cost limit.

There are several contributions dedicated to sensor placement in WDNs. Most of the works have addressed the sensor placement problem regarding contamination monitoring. In [2] and [3], the problem of sensor placement in a large WDN is considered in order to detect the malicious introduction of contaminants. On the other hand, less work has been done regarding sensor placement for leak location. In [4] a leak location method based on the pressure measurements and sensitivity analysis of nodes in a network has been proposed. In order to maximize the isolability with a reasonable number of sensors an optimal sensor placement methodology based

on genetic algorithms has also been proposed. The optimization goal consisted in minimizing the size of the larger set of non-isolable leaks.In [5], a strategy based on isolability maximization allows one to optimally locate sensors for leak location based on the structural model of water network. An efficient branch and bound search strategy was developed based on a structural network model. A structural model just considers the relation between variables that exists through equations, so efficient graph-based methods can be then applied to solve the sensor placement problem. However, optimal results can not be guaranteed due to the simplistic nature of structural representation.

In the present paper, a richer model description is proposed for the sensor placement problem through the fault sensitivity matrix concept. However, since no efficient branch and bound search strategy will be possible, a clustering approach to sensor placement is proposed in this paper. In [6], clustering techniques were also applied to leak detection and location, but combined with a structural approach. The clustering problem has been addressed by researchers in many contexts and disciplines [7]. It is a mature and active research area [8] and many efficient clustering algorithms have been developed in the literature.

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

II. FAULT DIAGNOSIS PRINCIPLES

A. Model-based fault diagnosis

Model-based fault diagnosis techniques are applied to detect and locate leaks in WDNs. In model-based fault diagnosis [9] 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. At 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), $\boldsymbol{\Omega}$

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

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).

B. Leak detection and location

The FSM can be obtained by convenient manipulation of model equations as long as fault effects are included in them [10]. Alternatively, it can be obtained by sensitivity analysis through simulation [4]. The latter approach is used in the present paper. 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

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

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 [4]. 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 [11].

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 $\boldsymbol{\Omega}$ 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.

The quality of a leak diagnosis system can be determined through the evaluation of leak detectability and locatability properties.

Definition 1 (Detectable leak set). 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 $\boldsymbol{\Omega}$, the set of detectable leaks $\mathbf{F}_D$ is 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.

Definition 2 (Leak locatability index). 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 $\boldsymbol{\Omega}$, the leak locatability index I is 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 $\mathbb{F} = \{(f_k, f_l) \in \mathbf{F} \times \mathbf{F} : k < l\}$.

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

In order to better evaluate the leak locatability capacity of a diagnosis system the following definition is provided.

Definition 3 (Uniform projection angle). Given a set of leaks $f_j \in \mathbf{F}$ and a leak locatability index I, the uniform projection angle $\bar{\alpha}$, is 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 fault sensitivity vectors in the FSM, assuming a uniform contribution to the leak locatability index. $\bar{\alpha}$ will be later used when comparing different sensor placement solutions.

III. Sensor placement methodology

A. Problem statement

Usually, the sensor placement problem is presented as an optimization problem where the cheaper sensor configuration fulfilling some given diagnosis specifications is sought [12], [13]. Nevertheless, a baseline budget is usually assigned to instrumentation by water distribution companies which constraints the cost of the sought sensor configuration. Thus, in the water distribution domain, companies are not interested in achieving a given diagnosis performance but in the best diagnosis performance that can be reached by installing a specific number of sensors that satisfy the budget constraint.

Let $\mathbf{S}$ be the candidate pressure sensor set and m_p the number of pressure sensors that will be installed in the system. Then, the problem can be roughly stated as the

choice of a configuration of m_p pressure sensors in $\mathbf{S}$ such that the diagnosis performance is maximized. This diagnosis performance depends on the set of sensors installed in the network $S \subseteq \mathbf{S}$ and it will be stated in terms of the detectable leak set and the leak locatability index, i.e., $\mathbf{F}_D(S)$ and $I(S)$.

To solve the sensor placement problem, a network model is also required. The leak sensitivity matrix $\mathbf{\Omega}$ corresponding to the complete set of candidate sensors is assumed to be previously computed following the methodology described in Section II-B. Hence, the sensor placement for leak diagnosis can be formally stated as follows:

GIVEN a candidate sensor set $\mathbf{S}$, a leak sensitivity matrix $\mathbf{\Omega}$, a leak set $\mathbf{F}$, and the 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 $\mathbf{F}$ are detectable, $\mathbf{F}_D(S) = \mathbf{F}$, and

    2) the leak locatability index is maximized, i.e. $I(S) \geq I(S^\star)$ for any $S^\star \subseteq \mathbf{S}$ such that $|S^\star| = m_p$.

This optimization problem can not be solved by efficient branch and bound search strategies. Thus, a suboptimal search algorithm based on clustering techniques will be applied. However, in order to alleviate the suboptimality drawback of clustering techniques a two-step hybrid methodology that combines them with an exhaustive search is proposed:

Step 1 Clustering techniques are applied to reduce the initial set of candidate sensors $\mathbf{S}$ to $\mathbf{S}'$, such that next step is tractable. Step 1 will be described in detail in next section.

Step 2 An exhaustive search is applied to the reduced candidate sensor set $\mathbf{S}'$. 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 leak locatability index for every pair of leaks which involves computing $\binom{|\mathbf{F}|}{2}$ times the normalized projection of the leak sensitivity vectors. Thus, in all, an exhaustive search is of exponential complexity, but an optimal solution is guaranteed.

B. Clustering approach

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 $\mathbf{C} = \{C_1, C_2, \cdots, C_\ell\}$ ($\ell \leq n_e$) in such a way that objects in the same group (called cluster) are more similar (in some sense) to each other than those in other groups (clusters). For example, k-means clustering algorithm [14] minimizes the within-cluster sum of distances by solving the optimization problem

$$\arg\min_{\mathbf{C}} \sum_{i=1}^{\ell} \sum_{\boldsymbol{x}_j \in C_i} d(\boldsymbol{x}_j, \boldsymbol{\mu}_i) \qquad (8)$$

where d is a distance and $\boldsymbol{\mu}_i$ is the centroid of cluster C_i (i.e. it is the mean of observations in C_i according to metric d).

Problem (8) is nonconvex and obtaining the solution is NP-hard, but there are efficient heuristic algorithms that converge quickly to a local optimum. K-means belongs to closed data sets methods, which do not allow overlapping of clusters. This is a drawback when some similar data points originating from different classes cannot be correctly classified into a specific class because of the limitation of the probabilistic framework. This is more precisely due to its inability to make a clear distinction between the full lack of knowledge and the full knowledge of the equiprobable cases. In this sense, some clustering techniques have been developed in the belief functions framework [15]. They provide the centroids of the ℓ clusters and the degree of membership of every element to every cluster $pl_i(C_k)$. $pl_i(C_k)$ represents the plausibility (or the possibility) that object $\boldsymbol{x}_i$ belongs to cluster C_k. One of these algorithms is the Evidential c-means (ECM) [16], which also provides a validity index allowing the determination of the proper number of clusters. A hard partition can be easily obtained by assigning each object to the cluster with highest plausibility i.e

$$\mathbf{g}(i) = \arg\max \{pl_i(C_1), \cdots, pl_i(C_\ell)\} \; i = 1, \cdots, n_e \quad (9)$$

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

In this paper, a reduction in the number of candidate sensors is proposed by grouping the n_e initial sensors into ℓ clusters applying the ECM algorithm. Then N representative sensors will be selected for each cluster, setting up the new candidate sensor set of $N\ell$ elements ($N\ell \leq n_e$). 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

$$N = \left\lceil \frac{n_r}{\ell} \right\rceil \qquad (10)$$

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

In this case, the criterion used for determining the similitude between elements (sensors) is the sensitivity pattern of their primary residuals to faults. In particular, according to the procedure described in Section II, this is provided by every row i of the fault sensitivity matrix $\mathbf{\Omega}$ defined in Eq. (1). So, choosing $\boldsymbol{x}_j = \frac{\boldsymbol{\omega}_{j\bullet}}{\|\boldsymbol{\omega}_{j\bullet}\|}$, $j = 1, ..., n_e$ (where $\boldsymbol{\omega}_{j\bullet}$ is the j row vector of matrix $\mathbf{\Omega}$, $\boldsymbol{x}_j$ the normalized vector of $\boldsymbol{\omega}_{j\bullet}$ and n_e the number of rows of $\mathbf{\Omega}$ i.e. $n_e = n$) and applying the ECM algorithm defined in [16], a set of ℓ clusters defined by their centroids $\boldsymbol{\mu}_i$ ($i = 1, \ldots, \ell$) and the plausibility matrix $\mathbf{\Pi}$ ($n \times \ell$) that contains the membership degree of every element to every cluster are obtained

$$\mathbf{\Pi} = \begin{pmatrix} pl_1(C_1) & \cdots & pl_1(C_\ell) \\ \vdots & \ddots & \vdots \\ pl_{n_e}(C_1) & \cdots & pl_{n_e}(C_\ell) \end{pmatrix} \qquad (11)$$

The groups of sensors with a similar fault sensitivity pattern can be obtained by means of (9). Once the elements x_j (sensors) have been grouped into ℓ clusters, the most N representative sensors of every cluster C_i can be chosen as the ones that have the maximum plausibility.

IV. APPLICATION TO A WDN

A. DMA case study

The sensor placement methodology is applied to a DMA located in Barcelona area (see Fig. 1) with 883 nodes and 927 pipes. The network consists of 311 nodes with demand (RM type), 60 terminal nodes with no demand (EC type), 48 nodes hydrants without demand (HI type) and 448 dummy nodes without demand (XX type). Only dummy nodes can have leaks. Thus, since there are 448 dummy nodes (XX type) in the network, there are 448 potential leaks to be detected and isolated. The network has two inflow inputs modeled as reservoir nodes.

The total inflow is distributed using a constant coefficient in each consumption node according to the total demand which is estimated using demand patterns.
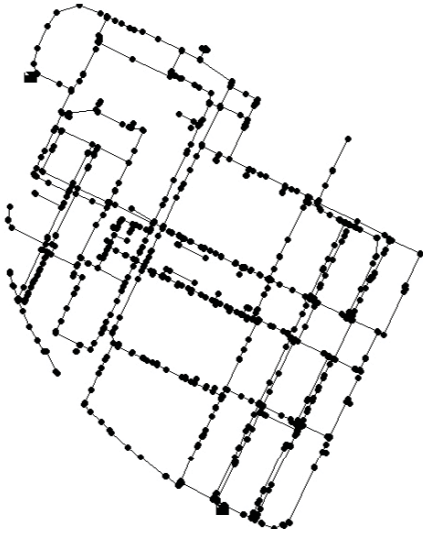

Fig. 1. Case study network map

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 [17]. Therefore, only pressure sensors will be considered in the sensor placement problem. In order to reduce the problem complexity, a subset of pressures is chosen as candidate variables to be measured. This subset consists of pressures at nodes of RM type. There exist 311 pressures that can be measured at these nodes, which represent the candidate sensor set. It is also assumed that there is no sensor already installed in the network before solving the sensor placement problem.

B. DMA network model

A fault sensitivity matrix has been obtained using the EPANET hydraulic simulator. Leaks are simulated in EPANET through the corresponding emitter coefficient, which is designed to model fire hydrants/sprinklers, and it can be adapted to provide the desired leak magnitude in the network, according to the equation:

$$EC = Q/P^{P_{ext}} \qquad (12)$$

where EC is the emitter coefficient, Q is the flow rate, P is the fluid pressure and P_{ext} is the pressure exponent. EPANET permits the value of the Emitter Coefficient to be specified for individual leak sites, but the pressure exponent can be only specified for the entire network. Data of node pressures are obtained from extensive simulations of normal and leak scenarios.

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 RM type nodes. Next, 448 leaks have been simulated in the XX type nodes and the steady-state pressure has been estimated again in the 311 RM type nodes. Finally, a fault sensitivity matrix has been obtained as the pressure difference between the fault free case and each faulty situation, according to the procedure described in Section 2. Although the fault sensitivity matrix depends on the leakage size, the properties are robust against this uncertainty. In this case, the leak sensitivity matrix has been computed for a leak magnitude of 6.3 lps (liters per second).

C. Sensor placement analysis

Assume that the water distribution company has assigned a baseline budget for investment on instrumentation that just makes it possible to install 5 pressure sensors. Hence, 5 pressure sensors should be chosen out of 311 such that all leaks are detectable and the leak locatability index is maximized. Recall from Section III-A that an exhaustive search is of exponential complexity. So, clustering techniques will be applied to set up a reduced set of 25 candidate pressure sensors. With this new setup, 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 Ω) as described in Section III-B. First, ECM clustering algorithm [16] has been used to classify the data set in different numbers of clusters and the evolution of the validity index with the number of clusters has been studied. From this study, it has been concluded that 5 is a proper number of clusters. The algorithm takes 30 seconds in the classification procedure to obtain 5 clusters. In order to verify the validity of the solution provided by the clustering algorithm, it has been run 500 times. Fig. 2 depicts the histogram of the sensors chosen as centroids in the 500 runs. Remark that some sensors are almost chosen in 200 out of

500 runs while other sensors are never chosen. Despite the clusters are different for each run, the leak locatability index (6), computed choosing the closest 5 sensors to the $\ell = 5$ cluster centroids, varies less than 2% (see Fig. 3). Therefore, the result obtained in a single run of the ECM algorithm could be already considered meaningful.
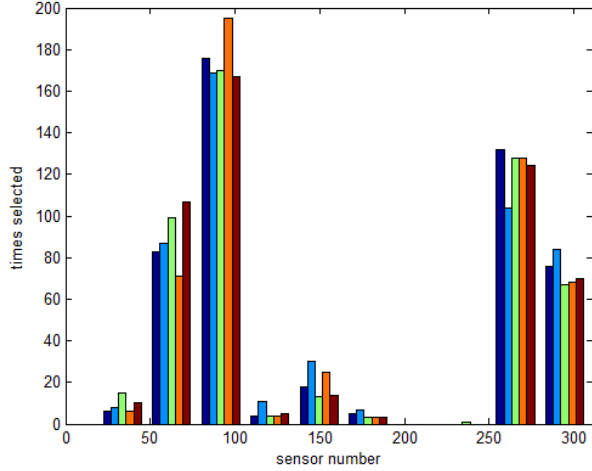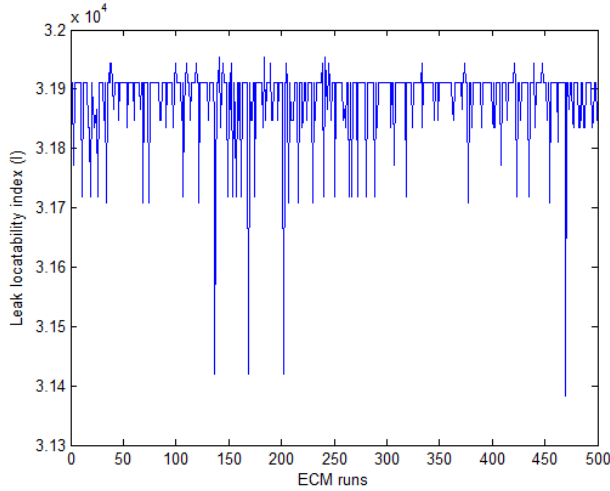

Fig. 2. Sensor centroids histogram

Fig. 3. Evolution of the locatability Index

With the plausibility matrix (11), obtained from the clustering algorithm, a hard partition has been obtained by assigning each element to its highest plausibility cluster applying Eq. (9). Fig. 5 depicts in different colors the 5 different network node clusters, where the closest node to the centroid have been highlighted in every cluster. Finally, the most N representative sensors of every cluster have been chosen as was proposed in Section III-B with $N = 5$ given by Eq. (10).

The reduced set $\mathbf{S}'$ with $|\mathbf{S}'| = N \times \ell = 25$ candidate pressure sensor places suggested by the clustering approach is displayed in Fig. 5 as blue circled nodes. The exhaustive
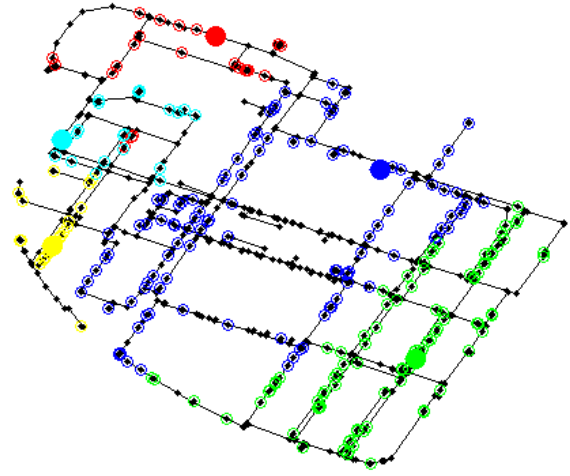

Fig. 4. Clustering results

search is next applied to solve the sensor placement problem, providing the set of 5 pressure sensor places signaled as red starred nodes in the same figure. Installing these pressure sensors, all 448 leaks are detectable and the leak locatability index amounts to 35631.96. According to Eq. (7), its corresponding uniform projection angle is $49.9°$, which seems reasonably good.
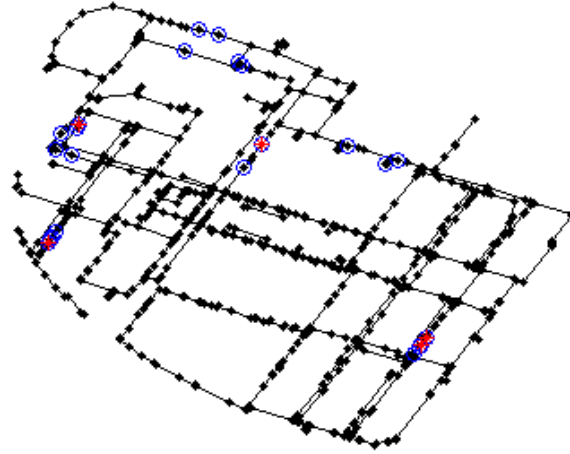

Fig. 5. DMA network sensor placement results

The exhaustive search approach provides an optimal result. However, due to its computational complexity just a reduced candidate pressure sensor set must be provided. Thus, the optimality of this result over the set of 311 original candidate pressure sensors relies on the performance of the clustering algorithm.

The closest node to each class centroid determined by the clustering procedure could be taken as an alternative solution to the sensor placement problem. This solution also guarantees full leak detectability, and the leak locatability index amounts to 31953.49, which corresponds to a uniform

projection angle of 47.1°. Remark that this angle is very close to the one corresponding to the exhaustive search solution. Thus, according to these results, the clustering approach fails to provide the optimal solution, but results are sufficiently satisfactory. The advantage of applying a clustering technique is that complex problems can be addressed in reasonable time. ECM takes 30 seconds to provide a result for the full candidate sensor set, whereas the exhaustive search takes more than 3 hours to get a solution for the reduced candidate sensor set.

V. CONCLUSIONS

The sensor placement problem in WDNs has been addressed in this paper. A distribution network usually describes a mesh topology involving hundreds of interconnected nodes whose behavior follows nonlinear physical laws. Such complexity requires the development of tools applicable to large-scale systems.

This work presents an optimal sensor placement strategy based on pressure sensitivity matrix analysis and an exhaustive search strategy that maximizes some diagnosis specifications. In order to reduce the size and the complexity of the problem, the present work proposes to combine this methodology with clustering techniques.

A first contribution of the paper is the definition of the leak locatability index as a diagnsosis performance measure. This index aggregates the normalized projection degree between the residual fault sensitivity vectors for all fault pairs. The goal is to characterize and determine a sensor set that guarantees a maximum degree of leak locatability while a budgetary constraint is satisfied. As a second contribution, to overcome the complexity of the sensor placement problem, the number of candidate sensors is reduced applying clustering techniques such that it can be tackled through exhaustive evaluation.

The strategy developed is successfully applied to a DMA of the Barcelona WDN. One the one hand, the results show that these combined techniques manage to solve the sensor placement problem in a reasonable time, which otherwise would not be possible. On the other hand, a quick solution of the sensor placement problem, close to the global optimum, can be directly obtained from clustering analysis.

REFERENCES

[1] R. Puust, Z. Kapelan, D. Savic, and T. Koppel, "A review of methods for leakage management in pipe networks," *Urban Water Journal*, vol. 7, pp. 25–45, 2010.

[2] A. Krause, J. L. ans C. Guestrin, J. Vanbriesen, and C. Faloutsos, "Effcient sensor placement optimization for securing large water distribution networks," *J. Water Resourc. Plan. Manag.*, vol. 134, pp. 516–526, 2008.

[3] M. Aral, J. Guan, and M. Maslia, "Optimal design of sensor placement in water distribution networks," *J. Water Resourc. Plan. Manag.*, vol. 136, pp. 5–18, 2010.

[4] R. Pérez, V. Puig, J. Pascual, J. Quevedo, E. Landeros, and A. Peralta, "Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks," *Control Engineering Practice*, vol. 19, no. 10, pp. 1157–1167, 2011.

[5] R. Sarrate, F. Nejjari, and A. Rosich, "Sensor placement for fault diagnosis performance maximization in distribution networks," in 20th *Mediterranean Conference on Control and Automation*, Barcelona, Spain, July 2012, pp. 110–115.

[6] R. Sarrate, J. Blesa, F. Nejjari, and J. Quevedo, "Sensor placement for leak detection and location in water distribution networks," *Water Science and Technology: Water Supply*, 2014, in Press.

[7] A. Jain, M. Murty, and P. J., "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[8] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neur. Net.*, vol. 16, no. 3, pp. 645–678, 2005.

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

[10] J. Blesa, V. Puig, and 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*, vol. 22, no. 10, pp. 1890–1912, 2012.

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

[12] M. Bagajewicz, A. Fuxman, and A. Uribe, "Instrumentation network design and upgrade for process monitoring and fault detection," *AIChE J.*, vol. 50, no. 8, pp. 1870–1880, Aug. 2004.

[13] R. Sarrate, F. Nejjari, and A. Rosich, "Model-based optimal sensor placement approaches to fuel cell stack system fault diagnosis," in 8th *IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Mexico City, Mexico, August 2012, pp. 96–101.

[14] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. on Math. Statist. and Prob., Berkeley*, Univ. of Calif. Press, 2008, pp. 281–297.

[15] Z. G. Liu, Q. Pan, and J. Dezert, "Ecm: An evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 46, no. 3, pp. 834–844, 2013.

[16] M. Masson and T. Denoeux, "Ecm: An evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 41, pp. 1384–1397, 2008.

[17] W. B. F. de Schaetzen, G. A. Walters, and D. A. Savic, "Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms," *Urban Water*, vol. 2, no. 2, pp. 141–152, 2000.