

Robust sensor placement for leak location: analysis and design

Joaquim Blesa, Fatiha Nejjari and Ramon Sarrate

ABSTRACT

In this paper, a nominal sensor placement methodology for leak location in water distribution networks is presented. To reduce the size and the complexity of the optimization problem a clustering technique is combined with the nominal sensor placement methodology. Some of the pressure sensor placement methods for leak detection and location in water distribution networks are based on the pressure sensitivity matrix analysis. This matrix depends on the network demands, which are nondeterministic, and the leak magnitudes, that are unknown. The robustness of the nominal sensor placement methodology is investigated against the fault sensitivity matrix uncertainty. Providing upon the dependency of the leak location procedure on the network operating point, the nominal sensor placement problem is then reformulated as a multi-objective optimization for which Pareto optimal solutions are generated. The robustness study as well as the resulting robust sensor placement methodology are illustrated by means of a small academic network as well as a district metered area in the Barcelona water distribution network.

Key words | clustering, leak location, robustness analysis, sensor placement

Joaquim Blesa (corresponding author)

Fatiha Nejjari

Ramon Sarrate

Automatic Control Department,
Universitat Politècnica de Catalunya (UPC),
TR11, Rambla de Sant Nebridi, 10,
08222 Terrassa,
Spain
E-mail: joaquim.blesa@upc.edu

Joaquim Blesa

Institut de Robòtica i Informàtica Industrial
(CSIC-UPC),
Carrer Llorens Artigas, 4-6,
08028 Barcelona,
Spain

INTRODUCTION

Water loss due to leaks in pipelines is one of the main challenges in efficient water distribution networks (WDNs). Leaks in WDNs can sometimes happen due to damage and defects in pipes, lack of maintenance, or increase in pressure. Leaks can cause significant economic losses and must be detected and located as soon as possible to minimize their effects. Some techniques and methods used to detect and locate leaks 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.

In WDNs, 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 strategy 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 being developed to achieve higher levels of efficiency (Puust *et al.* 2010).

Since Walski's (1983) study, many works that deal with the problem of optimal pressure sensor placement for hydraulic model calibration have been published. In particular, in Kaplan *et al.* (2005), a methodology based on single-objective and multi-objective genetic algorithms was presented. In the case of leak location purposes in WDNs, optimal locations of pressure sensor methodologies based on hydraulic simulations have been proposed (Farley *et al.* 2010). Some of the recent works that deal with the problem of leak location are based on the fault sensitivity matrix (Pérez *et al.* 2011; Casillas *et al.* 2012), which contains the information about how leaks affect the different node pressures. On the other hand, optimal 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 a WDN to carry out an optimal leak location as in Casillas *et al.* (2013) and Sarrate *et al.* (2014b). The fault sensitivity

matrix can be obtained by convenient manipulation of model equations as long as fault (leak) effects are included in them (Blesa et al. 2012). Alternatively, it can be obtained by sensitivity analysis through simulation (Pérez et al. 2011). 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, the robustness of the sensor placement methodology introduced in Sarrate et al. (2014b), where only inner pressure sensors in the district metered area (DMA) were considered, is evaluated. As this methodology was conceived for a nominal case, from now on it will be referred to as the nominal sensor placement methodology. The study is based on the generation of different leak scenarios, taking into account different leak magnitudes on the one hand and several operating points on the other hand. A robustness percentage index, which is based on the leak locatability index, is defined to assess the robustness of the nominal sensor placement methodology.

The robustness study concludes that there is a significant dependency of the nominal sensor placement methodology on the network operating point. This motivates the need for developing a new sensor placement methodology that can cope with uncertain operating conditions. As a result, a robust sensor placement methodology is proposed, which is formulated into a multi-objective optimization strategy. This multi-objective problem provides a set of Pareto optimal solutions, from which a decision-maker chooses the preferred one.

The robustness study and the robust sensor placement methodology are illustrated by means of a simple network with 12 nodes and a DMA in the Barcelona WDN with 883 nodes. In this latter case, a clustering technique is combined with the sensor placement methodology to reduce the size and the complexity of the problem.

SENSOR PLACEMENT FOR LEAK DETECTION AND LOCATION

Leak detection and location in WDNs

Model-based fault diagnosis techniques are applied to detect and locate leaks in WDNs. In model-based fault diagnosis (Blanke et al. 2006) 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, signaling the occurrence of a fault. In model-based fault isolation, the magnitude of the residuals that are inconsistent are compared against the different expected residual fault sensitivities, looking for the most probable fault that leads to model inconsistencies (residuals).

Given a set of m target leaks $f_j \in \mathbf{F}$ (i.e., m possible leak locations) and a set of n residuals $r_i \in \mathbf{R}$ (that compare measurements with model estimations), residual leak sensitivities are collected in the *Fault Sensitivity Matrix* (FSM) denoted by Ω :

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

Primary pressure residuals will be considered in this work. Primary residuals compare each actual pressure measurement vector \mathbf{p} with the corresponding estimated value in the fault free case $\hat{\mathbf{p}}_{nf}$. The FSM can be approximately computed by means of the predicted residual vector $\hat{\mathbf{r}}_{f_j} \in \mathcal{R}^n$ defined as:

$$\hat{\mathbf{r}}_{f_j} = \hat{\mathbf{p}}_{f_j} - \hat{\mathbf{p}}_{nf}, \quad (2)$$

where $\hat{\mathbf{p}}_{f_j}$ is the estimated pressure vector under leak fault f_j . $\hat{\mathbf{p}}_{nf}$ and $\hat{\mathbf{p}}_{f_j}$ are the solutions of the following nonlinear equations:

$$g_{nf}(\hat{\mathbf{p}}_{nf}, \boldsymbol{\theta}) = 0 \quad (3)$$

$$g_{f_j}(\hat{\mathbf{p}}_{f_j}, \boldsymbol{\theta}, f) = 0, \quad (4)$$

where $\boldsymbol{\theta}$ is a vector of dimension n_θ that defines the operating point in the WDN (heads in reservoirs, total inflow, and

demand distribution in nodes), f is the leak magnitude and $g_{nf}: \mathcal{X}^n \times \mathcal{X}^{n_0} \rightarrow \mathcal{X}^{n_c}$ and $g_{fj}: \mathcal{X}^n \times \mathcal{X}^{n_0} \times \mathcal{X} \rightarrow \mathcal{X}^{n_c}$ are non-linear functions derived from n_c hydraulic relations that describe the WDN behavior.

The FSM can be approximated in a nominal operating point θ^0 and for a nominal leak magnitude f^0 by

$$\Omega(\theta^0, f^0) \simeq \frac{1}{f^0} (\hat{\mathbf{r}}_{f_1}(\theta^0, f^0), \dots, \hat{\mathbf{r}}_{f_m}(\theta^0, f^0)), \quad (5)$$

where $\hat{\mathbf{r}}_{f_j}(\theta^0, f^0) \in \mathcal{X}^n$ can be obtained using (2) with $\hat{\mathbf{p}}_{f_j}$ and $\hat{\mathbf{p}}_{nf}$ being the solutions of (3) and (4) for $\theta = \theta^0$ and $f = f^0$, using a hydraulic simulator. In this work, (3) and (4) will be solved using the EPANET hydraulic simulator (Rossman 2000).

Thus, in general, the FSM defined in (1) is not constant but depends on the leak magnitude (f) and the operating point θ , i.e., $\Omega(\theta, f)$.

A leak can be detected as long as there exists at least a residual sensitive to it.

Leak isolation is achieved by matching the evaluated residual vector pattern to the closest residual leak sensitivity vector pattern (i.e., FSM column vector). In the present paper, a projection-based method is considered. Let $\omega_{\bullet j}$ be the column of Ω corresponding to leak j and $\mathbf{r} = [r_1 \dots r_n]^T$ be the actual residual vector corresponding to all n pressure measurement points

$$\mathbf{r} = \mathbf{p} - \hat{\mathbf{p}}_{nf} \quad (6)$$

Then, leak location can be achieved by solving the problem

$$\arg \max_j \frac{\omega_{\bullet j}^T \cdot \mathbf{r}}{\|\omega_{\bullet j}\| \|\mathbf{r}\|}, \quad (7)$$

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 and the most probable leak j is obtained.

The quality of a leak diagnosis system can be determined through the evaluation of leak detectability and locatability properties (Sarrate et al. 2014b).

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 Ω , the set of detectable leaks \mathbf{F}_D is defined as:

$$\mathbf{F}_D = \{f_j \in \mathbf{F}: \exists r_i \in \mathbf{R}: |\omega_{ij}| \geq \epsilon\}, \quad (8)$$

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 Ω , the leak locatability index I is defined as

$$I = \sum_{(f_k, f_l) \in \mathbb{F}} 1 - \frac{\omega_{\bullet k}^T \cdot \omega_{\bullet l}}{\|\omega_{\bullet k}\| \|\omega_{\bullet l}\|}, \quad (9)$$

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

Following the leak location criteria defined in Equation (7), the leak locatability index aggregates the normalized projection degree between the residual fault sensitivity vectors for all combinations of two leaks. Since a minimal normalized projection is desired, the greater the index is, the better it is.

Nominal sensor placement methodology

Usually, the sensor placement problem is presented as an optimization problem where the cheaper sensor configuration fulfilling some given diagnosis specifications is sought (Bagajewicz et al. 2004; Sarrate et al. 2012). Nevertheless, a baseline budget is usually assigned to instrumentation by water distribution companies which constrains the maximum cost of the sought sensor configuration and consequently the achievable diagnosis specifications. Thus, in the water distribution domain, companies rather seek the best diagnosis performance that can be achieved by installing the cheapest number of sensors that satisfy the budget constraint. Henceforth, the sensor placement methodology introduced in Sarrate et al. (2014b) is evoked.

Let \mathbf{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 \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 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 in a given operating point θ and for a given leak magnitude f , as described in the section ‘Leak detection and location in WDNs’. Hence, the nominal sensor placement for leak diagnosis can be formally stated as follows:

GIVEN a candidate sensor set \mathbf{S} , a nominal leak sensitivity matrix $\mathbf{\Omega}(\theta, f)$, 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 $I(S)$ is maximized, i.e.,

$$I(S^*) = \max_{S \subseteq \mathbf{S}, |S|=m_p} I(S)$$

This optimization problem cannot 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 the next step is tractable. Step 1 will be described in the 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 factorial complexity with the number of

candidate sensors $|\mathbf{S}'|$, but an optimal solution for the given reduced candidate sensor set is guaranteed.

Candidate sensor set reduction

In Sarrate et al. (2014a), a reduction in the number of candidate sensors has been proposed by grouping the n initial sensors candidate into ℓ groups (clusters) applying the evidential c-means (ECM) algorithm (Masson & Denoeux 2008).

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

In the present case, 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 the section ‘Leak detection and location in WDNs’, this is provided by every row i of the leak sensitivity matrix $\mathbf{\Omega}$ defined in Equation (1). So, choosing $\mathbf{x}_j = \frac{\omega_{j\bullet}}{\|\omega_{j\bullet}\|}$, $j = 1, \dots, n_e$ (where $\omega_{j\bullet}$ is the j th row vector of matrix $\mathbf{\Omega}$, \mathbf{x}_j the normalized vector of $\omega_{j\bullet}$ and n_e the number of rows of $\mathbf{\Omega}$, i.e. $n_e = n = |\mathbf{S}|$) and applying the ECM algorithm defined in Masson & Denoeux (2008), a set of ℓ clusters defined by their centroids μ_i ($i = 1, \dots, \ell$) and the plausibility matrix $\mathbf{\Pi}$ ($n \times \ell$) that contains the membership degree of every element to every cluster are obtained:

$$\mathbf{\Pi} = \begin{pmatrix} pl_1(C_1) & \dots & pl_1(C_\ell) \\ \vdots & \ddots & \vdots \\ pl_{n_e}(C_1) & \dots & pl_{n_e}(C_\ell) \end{pmatrix} \quad (10)$$

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

$$\mathbf{g}(i) = \arg \max_j pl_i(C_j) \quad i = 1, \dots, n_e \quad (11)$$

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

Once the set of sensors has been divided in clusters $\mathbf{C}_1, \dots, \mathbf{C}_N$, N representative sensors will be selected for

each cluster, setting up the new candidate sensor set 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 numbers of groups. Finally, the number N ($N \geq 1$) will be given by

$$N = \lceil \frac{n_r}{\ell} \rceil \quad (12)$$

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

Given the vectors \mathbf{pl}_i , \mathbf{Q}_i , and \mathbf{modw}_i that contain the plausibility values of the elements of the cluster set \mathbf{C}_i , the row numbers of the sensitivity matrix defined in Equation (1) relate to the elements of this cluster (sensor numbers) and the Euclidean norm of these rows of the sensitivity matrix. Algorithm 1 provides the vector \mathbf{Q}_i^0 with N representative elements (sensors) of the cluster \mathbf{C}_i : $\mathbf{Q}_i^0(1), \dots, \mathbf{Q}_i^0(N)$. The bigger N is the more representative the elements \mathbf{Q}_i^0 of the set \mathbf{C}_i are. In this algorithm, in addition to the plausibility values, the Euclidean norm of the sensor sensitivity matrix is taken into account in order to obtain sensor candidates that maximize the leak detectability defined in Equation (8).

Algorithm 1 N most representative cluster elements

Algorithm N-most-representative (\mathbf{pl}_i , \mathbf{Q}_i , \mathbf{modw}_i)

tempw_{*i*} ← **modw**_{*i*}

$\mathbf{pl}_i^{\min} \leftarrow \min(\mathbf{pl}_i)$

$\mathbf{pl}_i^{\max} \leftarrow \max(\mathbf{pl}_i)$

$n_i \leftarrow \text{length}(\mathbf{pl}_i)$

for $j = 1, \dots, N$ **do**

for $k = 1, \dots, n_i$ **do**

if $\left(\mathbf{pl}_i(k) < \mathbf{pl}_i^{\min} + \frac{(j-1)(\mathbf{pl}_i^{\max} - \mathbf{pl}_i^{\min})}{N} \right)$

tempw_{*i*}(k) ← 0

end if

end for

$loc = \arg \max_k \mathbf{tempw}_i(k)$

$\mathbf{Q}_i^0(j) = \mathbf{Q}_i(loc)$

$\mathbf{tempw}_i(loc) \leftarrow 0$

end for

return \mathbf{Q}_i^0

end Algorithm

ROBUSTNESS ANALYSIS METHODOLOGY

The robustness analysis will concern the leak magnitude uncertainty and the operating point variation. Both analyses will be done separately.

On the one hand, the study concerning leak magnitude uncertainty will involve, for a given nominal operating point θ^0 , evaluating the effect of possible uncertain values of the leak magnitude f^i within a given interval $f^i \in [f_{\min}, f_{\max}]$ on the nominal sensor placement methodology. This analysis considers a finite number sf of scenarios that lead to sf different FSMs $\Omega(\theta^0, f^1), \dots, \Omega(\theta^0, f^{sf})$, where $f^1 = f_{\min}$ and $f^{sf} = f_{\max}$.

On the other hand, the study concerning the operating point variation will involve, for a given nominal leak magnitude f^0 , evaluating the effect of the operating point θ^i variation (total inflow, demand distribution, etc.) on the nominal sensor placement methodology. This analysis considers a finite number $s\theta$ of representative operating point scenarios in the network that lead to $s\theta$ different FSMs $\Omega(\theta^1, f^0), \dots, \Omega(\theta^{s\theta}, f^0)$.

The nominal sensor placement methodology proposed in the section ‘Nominal sensor placement methodology’ will be applied to every scenario. The optimal solution obtained for each scenario is expected to be different. Let S_j^* be the optimal sensor configuration obtained for scenario j and let $I_i(S)$ be the locatability index that corresponds to scenario i when sensor configuration S is installed. Then, the *leak locatability matrix* (LLM) is defined. It has as many rows and columns as scenarios and its elements llm_{ij} correspond to $I_i(S_j^*)$. Based on this matrix, robustness will be evaluated through the *robustness percentage index* ρ defined as:

$$\rho = 100 \max_i \left(\frac{\max_j llm_{ij} - \min_j llm_{ij}}{\max_j llm_{ij}} \right). \quad (13)$$

In order to gain robustness under uncertain operating conditions that have been considered in the leak magnitude and operating point variation studies, the following extended FSM will be used by the clustering procedure proposed in the section ‘Candidate sensor set reduction’ to reduce the number of candidate sensors:

$$\bar{\Omega} = (\bar{\Omega}_1, \bar{\Omega}_2). \quad (14)$$

where

$$\bar{\Omega}_1 = (\Omega(\theta^0, f^1), \dots, \Omega(\theta^0, f^{sf})) \quad (15)$$

$$\bar{\Omega}_2 = (\Omega(\theta^1, f^0), \dots, \Omega(\theta^{s\theta}, f^0)) \quad (16)$$

Fault sensitivity matrices $\Omega(\theta^0, f^i)$ and $\Omega(\theta^j, f^0)$ will be 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:

$$Q = EC \cdot P^{P_{exp}} \quad (17)$$

where EC is the emitter coefficient, Q is the flow rate, P is the available pressure at the considered node, and P_{exp} is the pressure exponent. EPANET permits the value of the emitter coefficient to be specified for individual leak sites, but the pressure exponent can only be specified for the entire network.

Concerning the operating point robustness study, scenarios are generated with EPANET by specifying several values for the network total inflow.

ROBUSTNESS ANALYSIS OF A SMALL WDN BENCHMARK

Case study 1 description

The robustness analysis will be first performed on a small network (see Figure 1). The network has 12 nodes and 17 pipes, with two inflow inputs modeled as reservoir nodes. Thus, 10 potential leaks and 10 candidate pressure sensor locations will be considered at the network nodes (excluding reservoir nodes).

Five scenarios are defined concerning leak magnitude uncertainty ($f^0 = 0.92$, $\mathbf{f} = \{0.5, 0.7, 0.9, 0.93, 0.95\}$) corresponding to leaks of $\{3.5 \text{ lps}, 4.9 \text{ lps}, 6.3 \text{ lps}, 6.51 \text{ lps}, 6.65 \text{ lps}\}$, respectively, and five others related to operating point variation ($\theta^0 = 15.84 \text{ lps}$, $\theta = \{5, 10, 15, 20, 25\} \text{ lps}$). Note that the operating point θ is defined here by the network inflow and leaks are characterized through the emitter coefficient, i.e., $f = EC$.

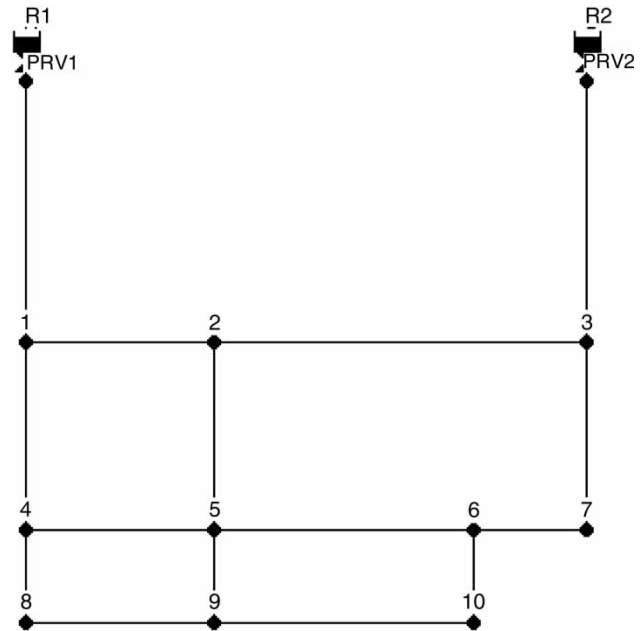


Figure 1 | Case study 1 network map.

Assume that a sensor placement problem with $m_p = 2$ is to be solved. Concerning the fault detectability constraint, $\varepsilon = 0.154 \text{ mm/lps}$ is considered in Equation (8).

Nominal sensor placement analysis and results

Figure 2(a) shows the evolution of row vector components 3 and 8 of $\Omega(\theta^0, f)$ considering $f = f_{min}, \dots, f = f_{max}$ for the 10 possible leaks, i.e. $\omega_{3j}(\theta^0, f)$ and $\omega_{8j}(\theta^0, f)$ for $j = 1, \dots, 10$.

Notice that the normalized vector $\|[\omega_{3j}(\theta^0, f), \omega_{8j}(\theta^0, f)]\|$ for all the considered leaks is almost the same. Thus, a non-significant variation in the locatability index (9) is expected for the different leak scenarios. Figure 2(b) shows the evolution of the same components of $\Omega(\theta, f^0)$ for different operating points $\theta^1, \dots, \theta^{s\theta}$, considering the same leak magnitude f^0 in all the leak scenarios. In this case, the variation that the normalized vector exhibits is remarkable. Thus, some variation in the locatability index (9) is expected for the different operating point scenarios.

In this case the network size is small and the clustering procedure is not needed to solve the sensor placement problem. Thus, just step 2 of the methodology outlined in the section 'Nominal sensor placement methodology' has been directly applied. Table 1 provides the resulting leak

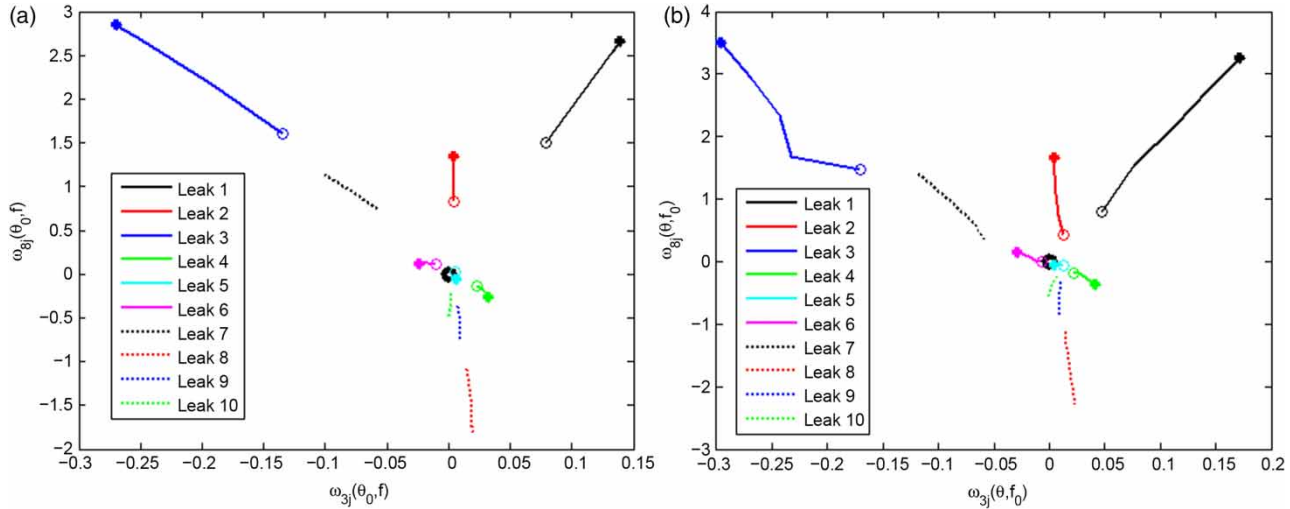


Figure 2 | Evolution of Ω row vector components 3 and 8 in case study 1 for the 10 possible leaks. (a) Leak magnitude dependency and (b) operating point dependency. Please refer to the online version of this paper to see this figure in colour.

locatability matrices. At the top row, the optimal sensor locations for each scenario are provided.

The robustness percentage index for leak magnitude uncertainty is 3.94%, which means that the nominal sensor placement results are very robust against this kind of uncertainty. However, the robustness percentage index for operating point variation is 14.43%, which proves some dependency of the nominal sensor placement results on this kind of variation.

ROBUSTNESS ANALYSIS OF A REAL WDN IN BARCELONA

Case study 2 description

The robustness analysis is performed on a bigger real network. The DMA is located in Barcelona (see Figure 3) and has 883 nodes and 927 pipes. The network consists of 311 nodes with demand (RM type), 60 terminal nodes with no

Table 1 | Leak locatability matrices for case study 1

(a) Leak magnitude uncertainty analysis

	$S_1^* = \{1,4\}$	$S_2^* = \{1,4\}$	$S_3^* = \{3,8\}$	$S_4^* = \{1,4\}$	$S_5^* = \{1,4\}$
Scn1	50.00	50.00	48.03	50.00	50.00
Scn2	50.00	50.00	49.24	50.00	50.00
Scn3	50.00	50.00	50.00	50.00	50.00
Scn4	50.00	50.00	50.00	50.00	50.00
Scn5	50.00	50.00	50.00	50.00	50.00

(b) Operating point variation analysis

	$S_1^* = \{1,9\}$	$S_2^* = \{3,8\}$	$S_3^* = \{3,8\}$	$S_4^* = \{1,4\}$	$S_5^* = \{1,4\}$
Scn1	50.00	49.57	49.57	47.73	47.73
Scn2	49.63	49.97	49.97	48.03	48.03
Scn3	43.50	50.00	50.00	50.00	50.00
Scn4	42.79	50.00	50.00	50.00	50.00
Scn5	43.03	50.00	50.00	50.00	50.00

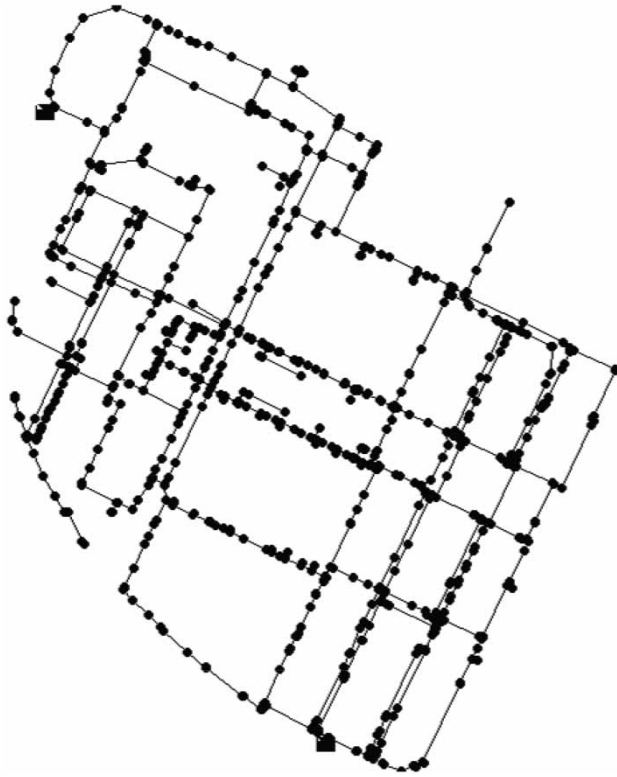


Figure 3 | Case study network map.

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 same scenarios defined for case study 1 are considered for case study 2.

Pressure sensors at RM nodes set up the candidate sensor set. Assume that a sensor placement problem with $m_p = 5$ has to be solved. Concerning the fault detectability constraint, $\epsilon = 0.154$ mm/lps is also considered in Equation (8).

Nominal sensor placement analysis and results

In order to reduce the complexity of the exhaustive algorithm, the number of candidate pressure sensors has been reduced from 311 to $n_r = 25$ using Algorithm 1. Clustering techniques have been applied to the 311 normalized rows of the nominal sensitivity matrix $\Omega(\theta^0, f^0)$ to classify the data set in five different clusters. The same procedure has

been carried out with the extended FSM $\bar{\Omega}$ defined in Equation (14). Figure 4(a) and 4(b) depict in different colors the five different network node clusters obtained with the nominal and extended FSM, respectively, where the closest node to the centroid has been highlighted in every cluster.

Note that there is an appreciable variation between the clustering obtained considering the nominal FSM and the one obtained considering the extended FSM.

Finally, the most N representative sensors of every cluster have been chosen using Algorithm 1 with $N = 5$ given by Equation (12). Figure 5 depicts the obtained reduced candidate sensor set in the network.

The exhaustive search of the section ‘Nominal sensor placement methodology’ is next applied to the reduced candidate sensor set provided by the clustering algorithm based on the extended FSM. Table 2 provides the resulting leak locatability matrices.

Results are similar to case study 1 concerning robustness analysis. Note that the same optimal sensor configuration results regardless of the leak magnitude uncertainty, which implies that the corresponding robustness percentage index is 0%. Thus, the nominal sensor placement results are again very robust against this kind of uncertainty. However, the robustness percentage index for the operating point variation is 34.31%, which also proves some significant dependency of the nominal sensor placement results on this kind of variation.

ROBUST SENSOR PLACEMENT FOR LEAK DETECTION AND LOCATION

Design methodology

Given the significant dependence of the leak detection and location procedure performance on the water network operating point, the goal of this section is to propose a sensor placement methodology that provides a set of pressure sensors that guarantee a robust performance. Thus, the nominal sensor placement methodology stated in the section ‘Nominal sensor placement methodology’ should be reformulated in order to cope with a new robustness

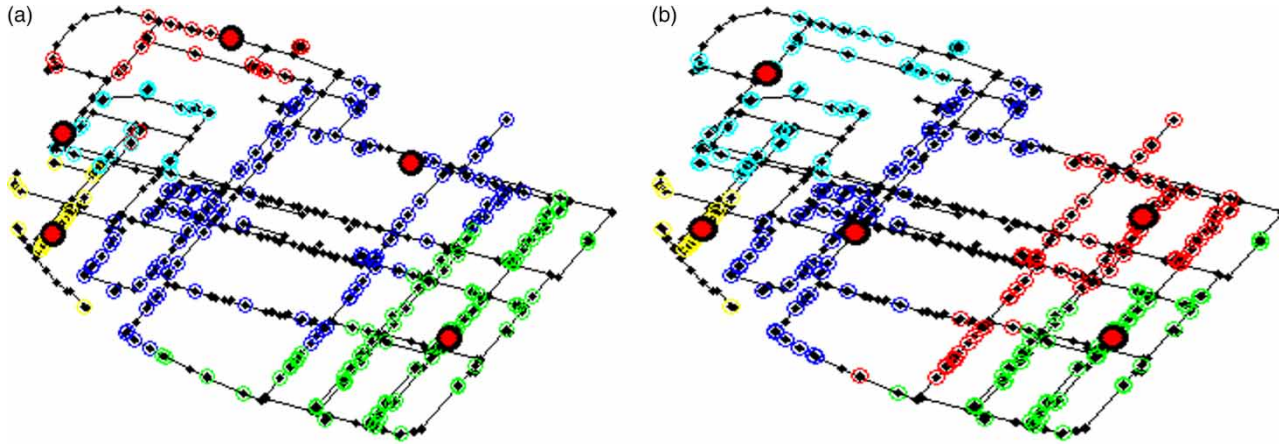


Figure 4 | Case study 2 network map and clustering results. (a) Based on the nominal FSM and (b) Based on the extended FSM. Please refer to the online version of this paper to see this figure in colour.

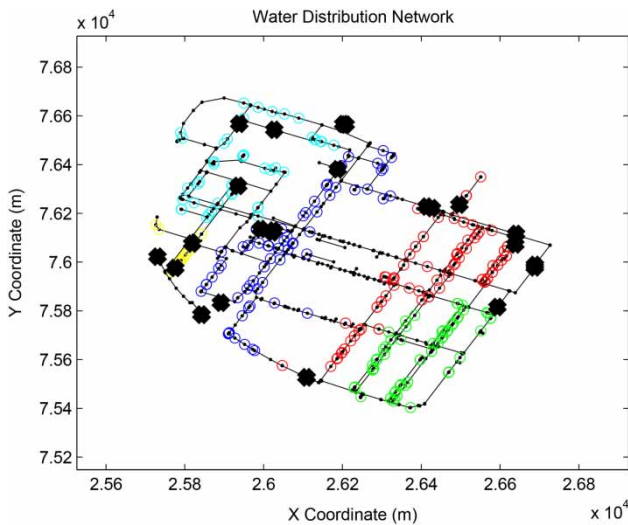


Figure 5 | Reduced candidate sensor set in the network. Please refer to the online version of this paper to see this figure in colour.

specification. The robust sensor placement problem is cast into a multi-objective optimization strategy as follows:

GIVEN a candidate sensor set \mathbf{S} , a set of leak sensitivity matrices $\mathbf{\Omega}(\theta^j, f^0)$ for $j = 1..s\theta$ representative operating point scenarios, 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, $\mathbf{F}_D(S^*) = \mathbf{F}$,
2. the mean leak locatability index $\bar{I}(S) = \text{mean}_{j=1..s\theta} I_j(S)$ is maximized, i.e., $\bar{I}(S^*) = \max_{S \subseteq \mathbf{S}, |S|=m_p} \bar{I}(S)$, and

3. the worst leak locatability index $I_w(S) = \min_{j=1..s\theta} I_j(S)$ is maximized, i.e., $I_w(S^*) = \max_{S \subseteq \mathbf{S}, |S|=m_p} I_w(S)$.

Although two robustness specifications have been stated, other specifications could be included in the sensor placement problem. Since, in a multi-objective optimization set-up, both robustness specifications cannot usually be met at the same time, a Pareto optimality will be sought (Carania & Dell’Olmo 2008). Given a set of objective functions, a solution is said to be Pareto optimal if no other feasible solution that improves an objective function without degrading at least one of the other objectives exists. There is usually a set of Pareto optimal solutions which is often called the Pareto front. The first goal of the robust sensor placement procedure will be to find the Pareto front. From this Pareto set, a decision-maker should examine the optimal solutions and through higher-level reasoning, choose the most suitable solution.

Different techniques exist to generate the Pareto front, such as those based on a linear scalarization (Das & Dennis 1997) or evolutionary algorithms (Deb et al. 2002). In this work, as in the case of the nominal sensor placement methodology, a two-step hybrid methodology that combines clustering techniques with an exhaustive search procedure is proposed. This methodology allows the computation of an approximation of the entire Pareto front for complex problems.

Step 1 Clustering techniques are applied to reduce the initial set of candidate sensors \mathbf{S} to \mathbf{S}' , such that next step is

Table 2 | Leak locatability matrices for case study 2

	$S_1^* = \{2, 8, 9, 83, 249\}$	$S_2^* = \{2, 8, 9, 83, 249\}$	$S_3^* = \{2, 8, 9, 83, 249\}$	$S_4^* = \{2, 8, 9, 83, 249\}$	$S_5^* = \{2, 8, 9, 83, 249\}$
(a) Leak magnitude uncertainty analysis					
<i>Scn1</i>	96,027	96,027	96,027	96,027	96,027
<i>Scn2</i>	100,165	100,165	100,165	100,165	100,165
<i>Scn3</i>	100,328	100,328	100,328	100,328	100,328
<i>Scn4</i>	100,186	100,186	100,186	100,186	100,186
<i>Scn5</i>	100,149	100,149	100,149	100,149	100,149
(b) Operating point variation analysis					
	$S_1^* = \{199, 243, 244, 245, 285\}$	$S_2^* = \{199, 243, 244, 245, 285\}$	$S_3^* = \{2, 8, 9, 83, 249\}$	$S_4^* = \{19, 199, 206, 222, 285\}$	$S_5^* = \{2, 9, 171, 222, 249\}$
<i>Scn1</i>	98,437	98,437	73,060	81,292	74,810
<i>Scn2</i>	99,086	99,086	79,696	80,116	78,817
<i>Scn3</i>	93,134	93,134	100,108	81,544	90,460
<i>Scn4</i>	41,328	41,328	42,355	49,105	46,728
<i>Scn5</i>	33,760	33,760	43,709	46,817	51,396

tractable. The method described in the section ‘Candidate sensor set reduction’ is applied. However, now, in order to cope with the robustness requirement, the extended FSM $\bar{\Omega}_2$ defined in Equation (16) is considered.

Step 2 An exhaustive procedure is applied to the reduced candidate sensor set S' . This procedure implies $\binom{|S'|}{m_p}$ times the computation of the detectable leak set, as well as the mean and worst locatability indices. The most time demanding computation concerns the evaluation of both indices, which involves computing $s\theta\left(\frac{|F|}{2}\right)$ times the normalized projection of the leak sensitivity vectors. Thus, in all, this exhaustive search is of factorial complexity with the number of candidate sensors $|S'|$.

As in the case of evolutionary algorithms, the Pareto optimality of the solutions cannot be guaranteed, but none of the generated solutions dominates the others. For demonstration purposes, one solution, that is relatively good for both objectives, will be chosen from the Pareto front.

Robust sensor placement for case study 1

Step 2 of the robust sensor placement methodology has been directly applied to the small WDN (see section ‘Case study 1 description’). Figure 6 represents the resulting Pareto front on the objective space as black circles, and Table 3 provides the two Pareto optimal solutions found.

Next, assume that a decision-maker chooses from the Pareto front sensor configuration {1,8} as the preferred one. Figure 7 displays the selected pressure sensor locations on the WDN.

Robust sensor placement for case study 2

The robust sensor placement methodology has been applied to the real WDN in Barcelona (see section ‘Case study 2 description’). Step 1 produces the same reduced candidate sensor set displayed in Figure 5. Next, step 2 is applied. Figure 8 displays the resulting Pareto front on the objective space as black circles, and Table 4 provides the eight Pareto optimal solutions found.

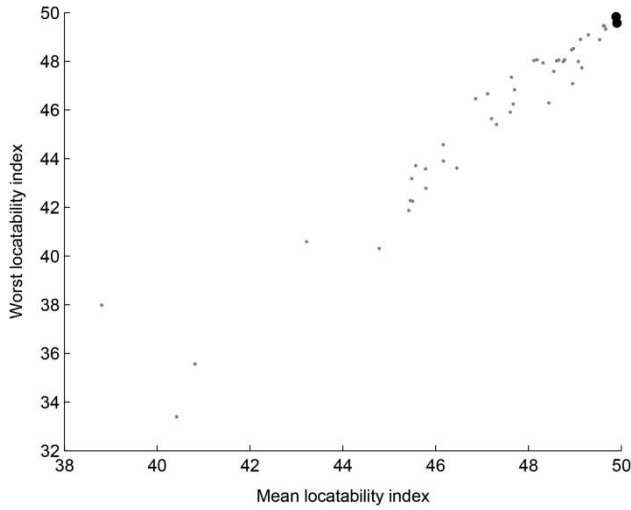


Figure 6 | Feasible solutions (grey dots) and Pareto front (black circles) on the objective space for case study 1.

Table 3 | Pareto optimal solutions for case study 1

	{1,8}	{3,8}
I_w	49.83	49.57
\bar{I}	49.89	49.91

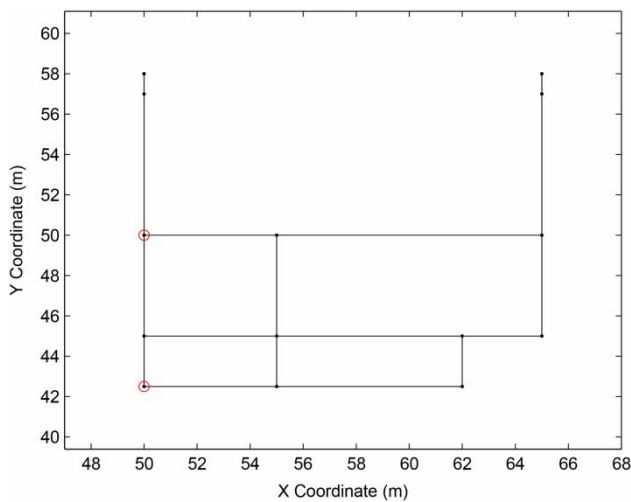


Figure 7 | Robust sensor placement results for case study 1.

Next, assume that a decision-maker chooses from the Pareto front sensor configuration {9, 199, 222, 243, 285} as the preferred one. Figure 9 displays the selected pressure sensor locations on the WDN.

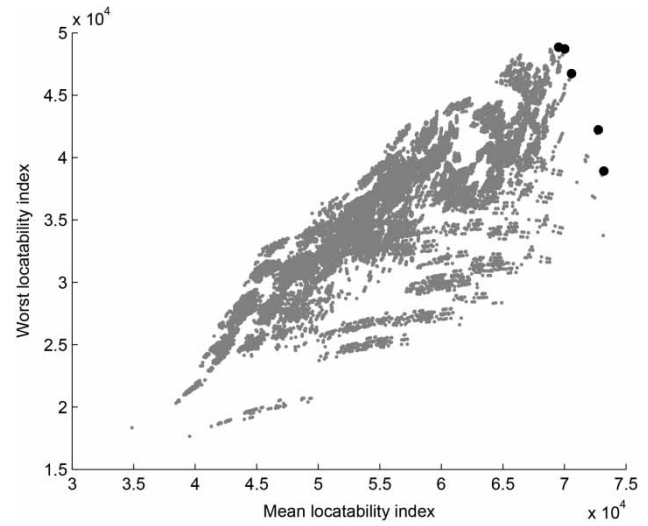


Figure 8 | Feasible solutions (grey dots) and Pareto front (black circles) on the objective space for case study 2.

Table 4 | Pareto optimal solutions for case study 2

	{199,222,243,244,285}	{171,199,222,243,285}	{8,199,222,243,285}	{9,199,222,243,285}
I_w	38,927	42,230	46,738	46,738
\bar{I}	73,194	72,751	70,580	70,580
	{8,171,199,222,243}	{9,171,199,222,243}	{8,171,199,222,285}	{9,171,199,222,285}
I_w	48,713	48,713	48,845	48,845
\bar{I}	70,027	70,027	69,523	69,523

In order to provide evidence of the benefits of the robust sensor placement methodology over the nominal one, the nominal sensor placement methodology has also been applied to the real WDN in Barcelona based on its nominal leak sensitivity matrix. First, a reduced set of 25 candidate pressure sensors has been obtained through clustering techniques. Next, applying an exhaustive search, sensor configuration {8, 199, 206, 245, 285} has been obtained. Figure 10 displays the selected pressure sensor locations on the WDN. The mean locatability index that corresponds to this sensor configuration is 68,491, whereas the worst locatability index is 43,102. According to the leak location performance achieved by the robust sensor placement methodology (see Table 4), this solution is not Pareto optimal. Thus, this confirms that the nominal methodology produces worse results than the robust one.

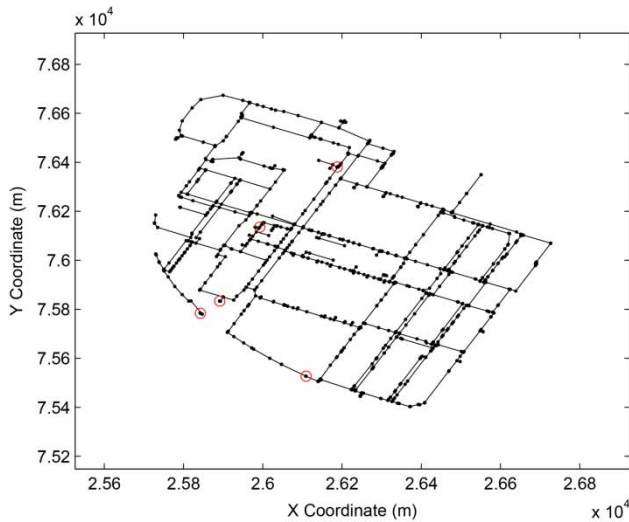


Figure 9 | Robust sensor placement results for case study 2. Please refer to the online version of this paper to see this figure in colour.

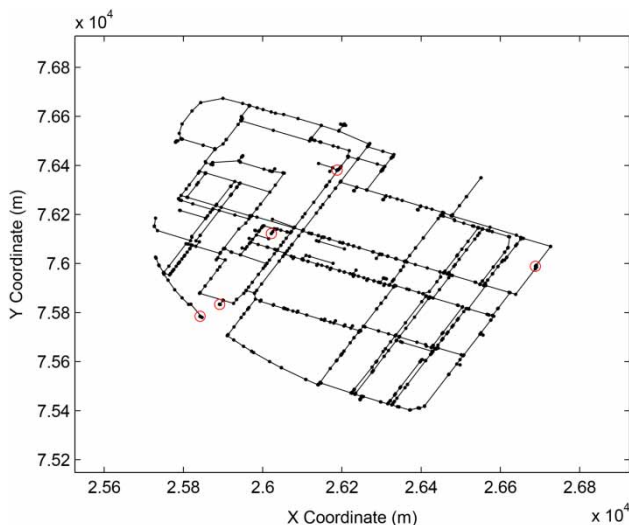


Figure 10 | Nominal sensor placement results for case study 2. Please refer to the online version of this paper to see this figure in colour.

CONCLUSIONS

The robustness analysis of the sensor placement problem based on leak sensitivity matrix analysis in WDNs has been addressed in this paper.

A robustness percentage index has been introduced to evaluate the variation of the leak locatability index achieved

by nominal sensor placement strategies for different leak magnitudes and DMA operating points.

A robustness analysis has been applied to an academic network and to a DMA in the Barcelona WDN. Results show that there is not an important variation of the leak locatability index when different leak scenarios are considered, but the variation can be significant when different operating point scenarios are considered. Therefore, this variation should be considered in optimal sensor placement strategies.

An important contribution of the paper is the development of a robust sensor placement methodology that takes into account the dependency of the leak location procedure on the network operating point. The robust sensor placement problem is formulated as a multi-objective optimization for which Pareto optimal solutions are generated. An exhaustive search is proposed to provide the Pareto front. In order to reduce the size and complexity of the optimization problem, a clustering technique based on an extended sensitivity matrix is applied beforehand to reduce the number of candidate sensors.

The exhaustive search provides the optimal Pareto front for the reduced candidate sensor set, but it is highly inefficient. Future research will focus on genetic algorithms or simulated annealing which will increase efficiency at the expense of global optimality.

The robust sensor placement methodology has focused on decoupled network operating uncertainties. However, the methodology could be easily extended to account for simultaneous uncertainties, provided that a complete set of fault sensitivity matrices could be generated. Of course, this would involve an increase in time complexity.

Future research should improve the formulation of the sensor placement problem presented in this paper. Currently, a solution with exactly m_p pressure sensors has been sought. However, the leak locatability index does not exhibit monotonicity on the cardinality of the sensor configuration. This means that sensor configurations with fewer pressure sensors could eventually improve the leak location performance of the network. Thus, developing algorithms that seek for a maximization of the leak location performance while minimizing the sensor expense would be desirable. Additionally, in the formulation of the multi-objective optimization, other convenient robustness criteria for the decision-maker could be investigated.

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

- Bagajewicz, M., Fuxman, A. & Uribe, A. 2004 Instrumentation network design and upgrade for process monitoring and fault detection. *AIChE J.* **50**, 1870–1880.
- Blanke, M., Kinnaert, M., Lunze, J. & Staroswiecki, M. 2006 *Diagnosis and Fault-Tolerant Control*, 2nd edn. Springer-Verlag, Berlin, Heidelberg, Germany.
- Blesa, J., Puig, V. & Saludes, J. 2012 Robust identification and fault diagnosis based on uncertain multiple input-multiple output linear parameter varying parity equations and zonotopes. *J. Process Control* **22**, 1890–1912.
- Caramia, M. & Dell'Olmo, P. 2008 *Multi-objective Management in Freight Logistics*. Springer, London, UK.
- Casillas, M. V., Garza, L. E. & Puig, V. 2012 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, pp. 570–575.
- Casillas, M. V., Puig, V., Garza, L. E. & Rosich, A. 2013 Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **13**, 14984–15005.
- Das, I. & Dennis, J. 1997 A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct. Multidiscip. Optimiz.* **14**, 63–69.
- Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197.
- Farley, B., Mounce, S. & Boxall, J. 2010 Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water J.* **7**, 345–356.
- Kapelan, Z., Savic, D. & Walters, G. 2005 Optimal sampling design methodologies for water distribution model calibration. *J. Hydraul. Eng.* **131**, 190–200.
- Masson, M. & Denoeux, T. 2008 ECM: an evidential version of the fuzzy c-means algorithm. *Pattern Recogn.* **41**, 1384–1397.
- Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. & Peralta, A. 2011 Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Eng. Pract.* **19**, 1157–1167.
- Puust, R., Kapelan, Z., Savic, D. & Koppel, T. 2010 A review of methods for leakage management in pipe networks. *Urban Water J.* **7**, 25–45.
- Rossman, L. A. 2000 *EPANET 2 Users Manual*. United States Environmental Protection Agency, Washington, DC, USA.
- Sarrate, R., Nejjari, F. & Rosich, A. 2012 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, pp. 96–101.
- Sarrate, R., Blesa, J. & Nejjari, F. 2014a Clustering techniques applied to sensor placement for leak detection and location in water distribution networks. In: *22nd Mediterranean Conference on Control and Automation*, Palermo, Italy, pp. 109–114.
- Sarrate, R., Blesa, J., Nejjari, F. & Quevedo, J. 2014b Sensor placement for leak detection and location in water distribution networks. *Water Sci. Technol. Water Supply* **14** (5), 795–803.
- Walski, T. M. 1983 Technique for calibrating network models. *J. Water Resour. Plann. Manage.* **109**, 360–372.

First received 30 January 2015; accepted in revised form 28 July 2015. Available online 19 September 2015