Sensor placement for leak monitoring

Ramon Sarrate, Fatiha Nejjari and Joaquim Blesa

Summary. The goal of this chapter is to review recent developments on sensor placement for fault diagnosis in drinking water networks. Solving a sensor placement problem for fault diagnosis entails the evaluation of the diagnosis performance of a monitoring system for a certain sensor configuration. In model-based fault diagnosis the monitoring system is built upon a set of consistency indicators, which are designed based on a model of the water network, and evaluated on the measurements provided by the sensors. The diagnosis performance is usually evaluated based on the fault sensitivity matrix (FSM), which states the fault sensitivity of the consistency indicators. Two approaches to sensor placement will be reviewed, which differ on the FSM nature. On the one hand, structural analysis leads to a binary FSM, which can be efficiently handled by the optimisation algorithm applying graph theory tools. On the other hand, the use of an analytical model leads to a more informative non-binary FSM. In both approaches only sub-optimal solutions can be attained. This is due to the fact that in the first approach a structural model is used, which is an abstraction of the analytical model, whereas in the second approach the sensor placement problem must be solved applying clustering techniques. These two approaches will be recalled and compared when applied to a district metered area (DMA) in the Barcelona WDN to decide the best location of pressure sensors for leak monitoring.

1 Introduction

Diagnosing system malfunctioning is of great importance in drinking water networks (DWNs). Leaks, water quality degradation or misbehaviour of pumps, flow meters and pressure sensors can lead to economic losses and consumer complaint. The diagnosis capability for system monitoring highly depends on the set of real-time measurements that are available. Thus, for a drinking water management company, deciding which sensors to install is the key to the success of a monitoring system.

Water loss due to leak in pipelines is one of the main challenges in efficient water distribution networks (WDN). Leaks in WDNs can happen due to damages 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 minimise their effects. Continuous improvements in water loss management are being applied, and new technologies are developed to achieve higher levels of efficiency [1].

Methods for locating leaks range from ground-penetrating radar to acoustic listening devices [2]. 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 the inverse problem and solving it using pressure or flow measurements (see Chapter 7). In the last years, different works that deal with the topic of leak location in WDNs using pressure sensors have been published. Some of these last works tackle with the problem of leak location using the fault sensitivity matrix [3, 4], which contains the information about how leaks affect the different node pressures (see Chapter 7 for a more detailed state of the art in this topic).

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 maximise 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.

The sensor placement problem can be roughly stated as choosing a subset, from a given candidate sensor location set, such that some diagnosis performance is guaranteed or at least maximised. Since installing sensors will involve a cost for the drinking water management company, economic constraints must be additionally taken into account in the choice. Sensor placement entails formulating a combinatorial optimisation problems. In such problems, an exhaustive search of the solution is usually not feasible, since its complexity grows exponentially with the number of candidate sensor locations. A DWN may easily involve several thousands of candidate sensor locations, which poses a severe optimisation challenge. Thus, the sensor placement methodology should be able to cope with such complexity issues.

Some results devoted to sensor placement for diagnosis can be found in [5, 6, 7**?**]. All these works use a structural model-based approach and define different diagnosis specifications to solve the sensor placement problem. A structural model is a coarse model description, based on a bi-partite graph, which can be obtained early in the development process, without major engineering efforts. This kind of model is suitable to handle large scale systems since efficient graph-based tools can be used and does not have numerical problems. Structural analysis is a powerful tool for early determination of fault diagnosis performances [8]. In [7] an algorithm is developed to determine where to install a specific number of pressure sensors in a DMA in order to maximise the capability of detecting and isolating leaks. The number of sensors to install is limited in order to satisfy a budgetary constraint requirement. However, in this case despite using an efficient branch and bound search strategy based on a structural model, the approach applicability is still limited to medium-sized networks. To overcome this drawback the methodology is combined with clustering techniques [9].

On the other hand, optimal pressure sensor placement algorithms based on sensitivity matrix analysis 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 [10], [11] and [**?**]. The fault sensitivity matrix can be obtained by convenient manipulation of model equations as long as leak effects are included in them [12]. Alternatively, it can be obtained by sensitivity analysis through simulation [3]. 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 [13] a robustness analysis of the sensor placement problem in WDNs has been addressed. The study has been achieved by optimal sensor placement strategies for different leak magnitudes and DMA operating points and evaluated through a robustness percentage index.

2 Problem Statement

2.1 Model-based fault diagnosis

Model-based fault diagnosis is a consolidated research area [8]. Most approaches to detect and isolate faults are based on consistency checking. The basic idea behind all these works is the comparison between the observed behaviour of the process and its corresponding model. This is performed by means of consistency relations, which can be roughly described as a function of the form

$$h(\mathbf{y}(t), \mathbf{u}(t)) = 0, \tag{1}$$

where $\mathbf{y}(t)$ and $\mathbf{u}(t)$ are vectors of known variables, denoting respectively process measurements and process control inputs. Function h is obtained from the model and is the basis to generate a residual

$$r(t) = h(\mathbf{y}(t), \mathbf{u}(t)). \tag{2}$$

A residual is a temporal signal indicating how close is behaving the process compared with its expected behaviour predicted by the model. In the absence of faults, a residual equals zero. In fact, a threshold based test is usually implemented in order to cope with noise and model uncertainty effects. Otherwise, when a fault is present the model is no longer consistent with the observations (known process variables) and the residual exceeds the prefixed threshold.

Detecting faults is possible with only one residual sensitive to all faults. However, fault isolation is usually required rather than just detecting the presence of a fault. The fault isolation task is performed by designing a set of residuals based on several consistency relations. Each residual is sensitive to different faults such that the residual fault signature is unique for each fault. Therefore, distinguishing the actual fault from other faults is possible by looking at the residual fault signature. These fault signatures are collected in the Fault Sensitivity Matrix (FSM) denoted by $\boldsymbol{\Omega}$ and introduced in Chapter 7

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{\partial r_1}{\partial f_1} & \cdots & \frac{\partial r_1}{\partial f_{n_d}} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_{n_y}}{\partial f_1} & \cdots & \frac{\partial r_{n_y}}{\partial f_{n_d}} \end{pmatrix}. \tag{3}$$

Where n_y is the number of the available residuals and n_d the number of considered faults. When an element ω_{ij} of $\boldsymbol{\Omega}$ is close to zero then residual r_i is weakly sensitive to fault $f_j \in \boldsymbol{\mathcal{F}}$, being $\boldsymbol{\mathcal{F}}$ the set of leaks that must be monitored, whereas when it diverges from zero then the residual is strongly sensitive to the fault $f_j \in \boldsymbol{\mathcal{F}}$.

In Chapter 7, the FSM $\boldsymbol{\Omega}$ concerning primary residuals (differences between pressure measurements and estimations) was approximated by

$$\boldsymbol{\Omega} \simeq \frac{1}{f^0}(\hat{\mathbf{r}}_{\mathbf{f_1}}, \ldots, \hat{\mathbf{r}}_{\mathbf{f_{n_d}}}), \tag{4}$$

where $\hat{\mathbf{r}}_{\mathbf{f_i}}$ is the predicted residual considering a leak in node i with magnitude f^0.

Sometimes a binary version of the fault sensitivity matrix is used. Then the corresponding binary residuals are usually called structured residuals, whereas in the non-binary matrix they are referred to as directional residuals.

In model-based diagnosis, fault detectability and fault isolability are the main objectives [8]. Assuming structured residuals, a fault is detectable if its occurrence can be monitored, whereas a fault $f_i \in \boldsymbol{\mathcal{F}}$ is isolable from a fault $f_j \in \boldsymbol{\mathcal{F}}$ if the occurrence of f_i can be detected independently of the occurrence of f_j.

2.2 Optimal sensor placement

Sensors measure water physical magnitudes such as pressure, flow, tank level or chlorine concentration. The aim of the sensor placement for monitoring can be roughly stated as the choice of a sensor configuration such that a monitoring performance specification is maximised. Nevertheless, this may lead to a solution involving a large instrumentation cost. A baseline budget is usually assigned to instrumentation by drinking water network companies which should constraint the maximum cost of the sought sensor configuration and consequently will bound the achievable monitoring performance. Thus, drinking water companies rather seek the best monitoring performance that can be achieved by installing the cheapest sensor configuration that satisfies a budget constraint. This chapter focuses on pressure sensor placement for leak monitoring although the methodology could be adapted to other formulations.

Let $\boldsymbol{\mathcal{S}}$ be the set of candidate pressure sensors and m_p the maximum number of pressure sensors that can be installed in the water network according to the budget constraint. Just sensor configurations $\mathcal{S} \subseteq \boldsymbol{\mathcal{S}}$ satisfying $|\mathcal{S}| \leq m_p$ will be considered where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$.

The monitoring specification T will be stated based on two fault diagnosis properties: fault detectability and fault isolability. Single fault assumption will hold (i.e., multiple faults will not be covered).

A drinking water network description M is also required to solve the sensor placement problem. Such description will allow the evaluation of leak monitoring specifications for a given pressure sensor configuration. Hence, the sensor placement for fault diagnosis can be formally stated as follows:

GIVEN a candidate pressure sensor set $\boldsymbol{\mathcal{S}}$, a drinking water network description M, a leak set $\boldsymbol{\mathcal{F}}$, a leak monitoring specification T, and a maximum number of pressure sensors m_p.

FIND a pressure sensor configuration $\mathcal{S} \subseteq \boldsymbol{\mathcal{S}}$ such that:

1. $|\mathcal{S}| \leq m_p$
2. T is maximised, and
3. $|\mathcal{S}|$ is minimal.

3 Proposed Approach

To solve the sensor placement optimisation problem two alternative methodologies based on two different drinking water network description M and involving a different formulation of the leak monitoring specification T will be investigated. The first approach is based on structural analysis and the second one on sensitivity analysis.

The considered optimisation problem is of combinatorial nature and its complexity critically depends on the cardinality of $\mathcal{S}$. In order to reduce the size and the complexity of this optimisation problem the following two-step hybrid methodology is proposed:

Step 1 Clustering techniques are applied to reduce the initial candidate sensor set $\boldsymbol{\mathcal{S}}$, such that the next step is tractable. At this step, a tentative size n_r for the reduced candidate sensor set is proposed to the clustering analysis. The complexity issues concerning Step 2 should be accommodated through this specification.

Step 2 Given the new candidate sensor set, the optimisation problem is solved following either the approach based on structural analysis or the one based on sensitivity analysis.

3.1 Clustering analysis

Given a set of objects $\mathcal{O} = \left\{\mathbf{o}_1, \mathbf{o}_2, \cdots, \mathbf{o}_{n_y}\right\}$ clustering consists in partitioning the n_y objects into ℓ sets $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_\ell\}$ $(\ell \leq n_y)$ 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).

In this case, the criterion used for determining the similitude between elements (sensors) is the sensitivity pattern of their primary residuals to leaks. This information is provided by every row i of the leak sensitivity matrix $\boldsymbol{\Omega}$ defined in (4). In this case, a complete sensitivity matrix will be computed. This matrix considers all possible sensors installed in the system, i.e. $n_y = n_d = |\boldsymbol{\mathcal{S}}|$. As proposed in [11], normalised leak sensitivities are considered, i.e., $\mathbf{o}_j = \frac{\boldsymbol{\omega}_{\mathrm{j}\bullet}}{\|\boldsymbol{\omega}_{\mathrm{j}\bullet}\|}$, $j = 1, ..., n_y$, where $\boldsymbol{\omega}_{\mathrm{j}\bullet}$ is the jth row vector of matrix $\boldsymbol{\Omega}$ and $\|\boldsymbol{\omega}_{\mathrm{j}\bullet}\|$ stands for the Euclidean norm of this vector. Next, applying the ECM (Evidential c-means) algorithm defined in [14], a set of ℓ clusters defined by their centroids $\boldsymbol{\mu}_i$ $(i = 1, \ldots, \ell)$ and the plausibility matrix $\boldsymbol{\Pi}$ $(n_y \times \ell)$ that contains the membership degree of every element to every cluster are obtained.

$$\boldsymbol{\Pi} = \begin{pmatrix} pl_1(\mathcal{C}_1) & \cdots & pl_1(\mathcal{C}_\ell) \\ \vdots & \ddots & \vdots \\ pl_{n_y}(\mathcal{C}_1) & \cdots & pl_{n_y}(\mathcal{C}_\ell) \end{pmatrix}, \tag{5}$$

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

$$\mathbf{g}(i) = \arg\max_j pl_i(\mathcal{C}_j) \quad i = 1, \cdots, n_y, \tag{6}$$

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

Once the set of sensors has been divided into clusters $\mathcal{C}_1, \ldots, \mathcal{C}_\ell$, N representative sensors should be selected of each cluster, setting up the new candidate sensor set of $N \times \ell$ elements $(N \times \ell \leq n_y)$. The number of groups ℓ will be set to the maximum number of installed sensors m_p as long as the validity index provided by the ECM algorithm confirms that this is a suitable number of clusters. Thus, N will be determined by

$$N = \left\lceil \frac{n_r}{m_p} \right\rceil, \tag{7}$$

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.

Let $\mathbf{pl}_i$ be the plausibility values of the elements of the cluster set $\mathcal{C}_i$, $\mathbf{row}_i$ the row numbers of the sensitivity matrix defined in (4) related to the elements of this cluster (sensor numbers) and $\mathbf{modw}_i$ the Euclidean norm of these rows of the sensitivity matrix. Algorithm 1 provides the vector $\mathbf{row}_i^0$ with N representative elements (sensors) of the cluster $\mathcal{C}_i$: $\mathbf{row}_i^0(1), \ldots, \mathbf{row}_i^0(N)$. The higher N is, the more representative the elements $\mathbf{row}_i^0$ of the set $\mathcal{C}_i$ 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 maximise the leak sensitivity. Once Algorithm 1 has been applied to the ℓ clusters, the reduced sensor set is composed by the $N \times \ell$ sensors associated to the obtained variables $\mathbf{row}_i^0$ $i = 1, \ldots, \ell$.

Algorithm 1 $\mathbf{row}_i^0 = \textsf{N-most-representative}(\mathbf{pl}_i, \mathbf{row}_i, \mathbf{modw}_i)$

```
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) then
            tempw_i(k) ← 0
        end if
    end for
    loc = arg max_k tempw_i(k)
    row_i^0(j) = row_i(loc)
    tempw_i(loc) ← 0
end for
return row_i^0
```

3.2 Structural analysis approach

3.2.1 Structural analysis framework

The analysis of the model structure has been widely used in the area of model-based fault diagnosis [8]. Therefore, consistent tools exist in order to perform diagnosability analysis and consequently compute the set of detectable and isolable faults.

The structural model is often defined as a bipartite graph $\mathcal{G} = (\mathcal{M}, \mathcal{X}, \mathcal{A})$, where $\mathcal{M}$ is a set of model equations, $\mathcal{X}$ a set of unknown variables and $\mathcal{A}$ a set of edges, such that $(e_i, x_j) \in \mathcal{A}$ as long as equation $e_i \in \mathcal{M}$ depends on variable $x_j \in \mathcal{X}$. A structural model is a graph representation of the analytical model structure since only the relation between variables and equations is taken into account, neglecting the mathematical expression of this relation.

Structural modelling is suitable for an early stage of the system design, when the precise model parameters are not known yet, but it is possible to determine which variables are related to each equation. Furthermore, the diagnosis analysis based on structural models is performed by means of graph-based methods which have no numerical problems and are more efficient, in general, than analytical methods. However, due to its simple description, it cannot be ensured that the diagnosis performance obtained from structural models will hold for the real system. Thus, only best case results can be computed.

It is well-known that the over-determined part of the model is the only useful part for system monitoring [8]. The Dulmage-Mendelsohn canonical decomposition [15] is a bipartite graph decomposition that defines a partition on the set of model equations $\mathcal{M}$. It turns out that one of these parts is the over-determined part of the model and is represented as $\mathcal{M}^+$.

The system fault diagnosis analysis is performed based on the structural model properties. Specifically, fault detectability and isolability are defined as properties of the over-determined part of the model [16]. First, it is assumed that a single fault $f \in \boldsymbol{\mathcal{F}}$ can only violate one equation (known as *fault equation*), denoted by $e_f \in \mathcal{M}$.

Definition 1. *A fault $f \in \boldsymbol{\mathcal{F}}$ is (structurally) detectable in a model described by the set of equations $\mathcal{M}$ if*

$$e_f \in \mathcal{M}^+ \tag{8}$$

Definition 2. *A fault f_i is (structurally) isolable from f_j in a model described by the set of equations $\mathcal{M}$ if*

$$e_{f_i} \in \left(\mathcal{M} \setminus \{e_{f_j}\}\right)^+ \tag{9}$$

Without loss of generality, it is assumed that a sensor $s_i \in \boldsymbol{\mathcal{S}}$ can only measure one single unknown variable $x_i \in \mathcal{X}$. In the structural framework, such sensor will be represented by one single equation denoted as e_s (known as *sensor equation*). Given a set of sensors $\mathcal{S}$, the set of sensor equations is denoted as $\mathcal{M}_{\mathcal{S}}$. Thus, given a candidate sensor configuration $\mathcal{S}$ and a model $\mathcal{M}$, the updated system model corresponds to $\mathcal{M} \cup \mathcal{M}_{\mathcal{S}}$.

Let $\mathcal{F}_D(\mathcal{S}) \subseteq \boldsymbol{\mathcal{F}}$ denote the detectable fault set when a sensor configuration $\mathcal{S} \subseteq \boldsymbol{\mathcal{S}}$ is installed in the system. Fault isolability can be characterised in a similar way by means of fault pairs. Let $\mathbb{F} : \boldsymbol{\mathcal{F}} \times \boldsymbol{\mathcal{F}}$ be all fault pair permutations from $\boldsymbol{\mathcal{F}}$, then $\mathcal{F}_I(\mathcal{S}) \subseteq \mathbb{F}$ denotes the set of isolable fault pairs when the sensor configuration $\mathcal{S} \subseteq \boldsymbol{\mathcal{S}}$ is chosen for installation (i.e., $(f_i, f_j) \in \mathcal{F}_I(\mathcal{S})$ means that fault f_i is isolable from f_j when the sensor set $\mathcal{S}$ is installed in the system).

From Definition 1, $\mathcal{F}_D(\mathcal{S})$ can be computed as

$$\mathcal{F}_D(\mathcal{S}) = \{f \in \boldsymbol{\mathcal{F}} \mid e_f \in (\mathcal{M}_{\mathcal{S}} \cup \mathcal{M})^+\} \tag{10}$$

and from Definition 2, $\mathcal{F}_I(\mathcal{S})$ can be computed as

$$\mathcal{F}_I(\mathcal{S}) = \{(f_i, f_j) \in \mathbb{F} \mid e_{f_i} \in (\mathcal{M}_{\mathcal{S}} \cup (\mathcal{M} \setminus \{e_{f_j}\}))^+\} \tag{11}$$

It is worth noting that testing different sensor configurations involves different sensor equation sets, $\mathcal{M}_{\mathcal{S}}$, in (10) and (11) while the other sets remain unchanged.

Definition 3. *(Isolability index) Given a sensor configuration $\mathcal{S} \subseteq \boldsymbol{\mathcal{S}}$, the isolability index is defined as the number of isolable fault pairs provided the sensors $s \in \mathcal{S}$ are installed, i.e.,*

$$I(\mathcal{S}) = |\mathcal{F}_I(\mathcal{S})|. \tag{12}$$

3.2.2 Optimal sensor placement algorithm

The optimal sensor placement problem stated on Section 2.2 will be solved under the structural analysis approach. This involves providing a structural model G as a DWN description M and stating the leak monitoring specification T as follows:

1. All leaks are detectable, i.e., $\mathcal{F}_D(\mathcal{S}) = \boldsymbol{\mathcal{F}}$ according to (10).
2. The number of isolable leak pairs is maximised, i.e., the isolability index $I(\mathcal{S})$ is maximised.

Algorithm 2 solves the optimal sensor placement problem, by applying a depth-first branch and bound search strategy. The search involves building a node tree by recursively calling function $\mathsf{searchOp}_C$, beginning at the root node down to the leaf nodes. Each node corresponds to a sensor configuration ($node.\mathcal{S}$) and child nodes are built by removing sensors from their corresponding parent node. Set $node.\mathcal{R}$ specifies those sensors that are allowed to be removed.

Throughout the search, the best solution is updated in $\mathcal{S}^*$ whenever a sensor configuration with a higher fault isolability index than the current best one is found, as long as all faults are detectable and the number of sensors does not exceed m_p. The best solution is also updated whenever a smaller sensor configuration is found that has the same isolability index as the current best one.

The search is initialised as follows: $node.\mathcal{S} = node.\mathcal{R} = \boldsymbol{\mathcal{S}}$ and $\mathcal{S}^* = \emptyset$. During the search, only those branches that can be further expanded to a sensor configuration that improves the current solution are explored. Branch exploration is aborted whenever the fault isolability index can not be improved, a fault is not detectable or the number of sensors can not satisfy the budget constraint.

3.3 Sensitivity analysis approach

3.3.1 Sensitivity analysis framework

Alternatively to the structural method proposed in the previous section, a method that aims at optimising the performance of the leak location method presented in Chapter 7 is proposed in this section. Considering $\mathbf{r} = [r_1 \cdots r_{n_s}]^T$ be the actual residual vector corresponding to all pressure measurement points $n_s = |\mathcal{S}|$, and $\hat{\mathbf{r}}_{f_j}$ be the column of FSM $\boldsymbol{\Omega}$ corresponding to leak j, the leak location was achieved by solving the problem:

$$\underset{j \in \{1, \ldots n_d\}}{\arg\max} \frac{\mathbf{r}^T \cdot \hat{\mathbf{r}}_{f_j}}{\|\mathbf{r}\| \|\hat{\mathbf{r}}_{f_j}\|}. \tag{13}$$

Thus, the biggest normalised projection of the actual residual vector on the fault sensitivity space is sought.

The detectable leak set $\mathcal{F}_D$ was defined in terms of structural analysis in (10). Next, it will be defined in terms of sensitivity analysis as proposed in [11]. Given

Algorithm 2 $\mathcal{S}^* = \mathsf{searchOp}_C(node, \mathcal{S}^*)$

```
childNode.R := node.R
for all s ∈ node.R ordered in decreasing cost do
  childNode.S := node.S \ {s}
  childNode.R := childNode.R \ {s}
  if |childNode.S \ childNode.R| > m_p then
    return S*
  end if
  if I(childNode.S) = I(S*) then
    if |childNode.S \ childNode.R| < |S*| then
      if F_D(childNode.S) = F then
        if |childNode.S| < |S*| then
          S* := childNode.S {update best solution}
        end if
        S* := searchOp_C(childNode, S*)
      end if
    else
      if I(childNode.S) = I(Node.S) then
        return S*
      end if
    end if
  else
    if I(childNode.S) > I(S*) and
    F_D(childNode.S) = F then
      if |childNode.S| ≤ m_p then
        S* := childNode.S {update best solution}
      end if
      S* := searchOp_C(childNode, S*)
    end if
  end if
end for
return S*
```

a set of sensors $\mathcal{S}$, a set of leaks $\boldsymbol{\mathcal{F}}$ and the corresponding FSM $\boldsymbol{\Omega}$, the set of detectable leaks $\mathcal{F}_D(\mathcal{S})$ is defined as:

$$\mathcal{F}_D(\mathcal{S}) = \{f_j \in \boldsymbol{\mathcal{F}} : \exists r_i \in \mathbb{R} : |\omega_{ij}| \geq \epsilon\}, \tag{14}$$

where ϵ is a threshold to account for noise and model uncertainty.

Regarding the leak locatability performance, assuming that the leak location is implemented by means of (13), a uniform projection angle $\bar{\alpha}$, defined as the average between the residual fault sensitivity vectors for all leak pairs, was proposed in [11].

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 (columns of $\boldsymbol{\Omega}$) should satisfy this uniform projection angle. This uniform angular separation between leak pairs would allow for a successful leak location method applying (13), even when residuals are affected by modelling 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 (13). 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 fault sensitivity vectors. So, distances between nodes with a similar fault 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 4. *(Leak expansion set) Given a leak $f_j \in \boldsymbol{\mathcal{F}}$ and a projection angle threshold α_{th}, the leak expansion set $\boldsymbol{\mathcal{F}}_j^{\alpha_{th}}$ is defined as*

$$\boldsymbol{\mathcal{F}}_j^{\alpha_{th}} = \{f_i \in \boldsymbol{\mathcal{F}} : \frac{\hat{\mathbf{r}}_{f_j}^T \cdot \hat{\mathbf{r}}_{f_i}}{\|\hat{\mathbf{r}}_{f_j}\| \|\hat{\mathbf{r}}_{f_i}\|} > \cos(\alpha_{th})\}. \tag{15}$$

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

Definition 5. *(Correlated leak pairs ratio)Given the leak expansion sets $\boldsymbol{\mathcal{F}}_j^{\alpha_{th}}$ with $j = 1, ..., |\boldsymbol{\mathcal{F}}|$, the correlated leak pairs ratio $\eta^{\alpha_{th}}$ is defined as*

$$\eta^{\alpha_{th}} = 100 \frac{\sum_{j=1}^{|\boldsymbol{\mathcal{F}}|} |\boldsymbol{\mathcal{F}}_j^{\alpha_{th}}| - |\boldsymbol{\mathcal{F}}|}{2\binom{|\boldsymbol{\mathcal{F}}|}{2}}. \tag{16}$$

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

Definition 6. *(Leak node distance matrix) Given the geographical coordinates of every leak node, the leak node distance matrix $\mathbf{D} \in \mathbb{R}^{|\boldsymbol{\mathcal{F}}| \times |\boldsymbol{\mathcal{F}}|}$ is defined as the matrix whose coefficients d_{ij} are the geographical distance between nodes i and j.*

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

Definition 7. *(Worst leak expansion distance) Given a leak expansion set $\boldsymbol{\mathcal{F}}_j^{\alpha_{th}}$ and the leak node distance matrix $\mathbf{D}$, the worst leak expansion distance $R_j^{\alpha_{th}}$ is defined as*

$$R_j^{\alpha_{th}} = \max_{f_i \in \boldsymbol{\mathcal{F}}_j^{\alpha_{th}}} d_{ij}. \tag{17}$$

Thus $R_j^{\alpha_{th}}$ 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 8. *(Average worst leak expansion distance) Given a set of leaks* $\boldsymbol{\mathcal{F}}$ *and a threshold projection angle* α_{th}*, leak expansion sets* $\boldsymbol{\mathcal{F}}_j^{\alpha_{th}}$ *with* $j = 1, ..., |\boldsymbol{\mathcal{F}}|$ *can be computed applying* (15)*. Then, the average worst leak expansion distance can be computed as*

$$\bar{R}^{\alpha_{th}} = \frac{1}{|\boldsymbol{\mathcal{F}}|} \sum_{j=1}^{|\boldsymbol{\mathcal{F}}|} R_j^{\alpha_{th}}. \tag{18}$$

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

As discussed in [17], the greater the threshold α_{th} is, the greater the uncertainty is in terms of leak expansion distance and number of correlated leak pairs. 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 modelling 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 specifying a smaller projection angle threshold. Therefore, a tradeoff exists between both criteria.

In order to find a good balanced solution, it is expected to find a sensor placement solution suitable for a range of projection angle thresholds as proposed in [17].

Definition 9. *(Mean average worst leak expansion distance) Given a set* $\mathcal{A}_{th} = \{\alpha_{th}^1, \ldots, \alpha_{th}^{n_\alpha}\}$ *that covers a suitable range of projection angle thresholds. Then, the mean average worst leak expansion distance over this set can be computed as*

$$\bar{R}^{\overline{\alpha_{th}}} = \frac{1}{|\mathcal{A}_{th}|} \sum_{\alpha_{th} \in \mathcal{A}_{th}} \bar{R}^{\alpha_{th}}, \tag{19}$$

where the average worst leak expansion distance $\bar{R}^{\alpha_{th}}$ *is computed applying* (18) *to every projection angle threshold in* $\mathcal{A}_{th}$*.*

3.3.2 Optimal sensor placement algorithm

The optimal sensor placement problem stated in Section 2.2 will be solved under the sensitivity analysis approach. This involves providing a node distance matrix $\mathbf{D}$, the FSM matrix $\boldsymbol{\Omega}$ and a set of projection angle thresholds $\mathcal{A}_{th}$. Then, stating the monitoring specification T as follows:

1. All leaks are detectable, i.e., $\mathcal{F}_D(\mathcal{S}) = \boldsymbol{\mathcal{F}}$ according to (14).
2. The mean average worst leak expansion distance $\bar{R}^{\overline{\alpha_{th}}}$ is minimised, i.e, $\frac{1}{\bar{R}^{\overline{\alpha_{th}}}}$ is maximised.

This optimisation problem cannot be solved by efficient branch and bound search strategies and it is necessary to implement an exhaustive search, i.e., to compute the mean average worst leak expansion distance $\bar{R}^{\overline{\alpha_{th}}}$ by means of (19) for all the possible $\frac{n_d!}{(n_d-n_s)!n_s!}$ sensor configurations with $n_s = 1, ..., m_p$ and choosing the one that provides the best performance (smallest $\bar{R}^{\overline{\alpha_{th}}}$).

4 Simulations and Results

4.1 DMA case study

A leak detection method involves dividing the distribution system into well-defined DMAs. Leak level in a DMA is determined based on the minimum night flow minus the legitimate night demand and estimated unavoidable background leakage. DMAs help identifying areas of the pipe network that suffer from excessive leakage.
The sensor placement methodology is applied to a DMA located in the 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). Additionally, the network has two inflow inputs modelled 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.

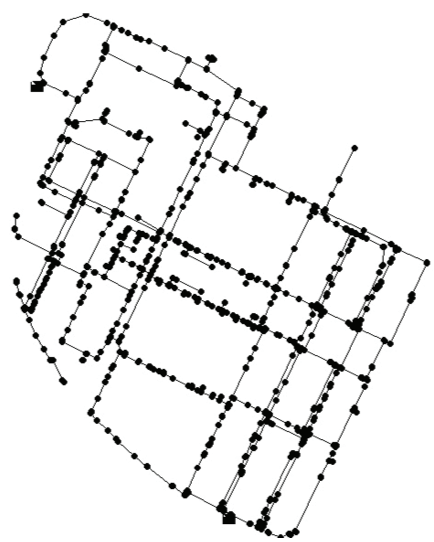

Fig. 1: Case study network map

Leaks might appear anywhere in the water network. However, due to simulation limitations, leaks are represented in the nodes where the flow balances take place. It is assumed that leaks might only occur at dummy nodes, leading to 448 potential leaks to be detected and located. A similar practical reason is applied when defining the possible location of the network monitoring points. Pressure sensors at RM type nodes will be used as network monitoring points, leading to 311 candidate sensors that could be chosen for installation. Additionally, it is also assumed that there is no pressure sensor already installed in the network before solving the sensor placement problem.

The water distribution company has established a maximum budget for investment on instrumentation that makes it possible to install up to 5 pressure sensors. Hence, up to $m_p = 5$ pressure sensors should be chosen out of 311 such that the leak monitoring specifications are maximised. Despite measuring flow rate could also be useful for leak detection, collecting pressure data is cheaper and easier, and pressure transducers give instantaneous readings whereas most flow meters do not react instantaneously to flow changes [18].

4.2 DMA network modelling

The DMA network is represented by a directed graph $\mathcal{G}_N = (\mathcal{V}, \mathcal{J})$ where pipe junctions are nodes, $v \in \mathcal{V}$, and pipes are edges, $j \in \mathcal{J}$. Each node represents, at the same time, a pressure variable and a flow balance equation. Similarly, each edge represents a flow variable and a pipe equation. Therefore, given a node $v \in V$, the following flow balance equation can be derived,

$$\sum_{j \in \mathcal{J}_v} q_j = d_v \tag{20}$$

where $\mathcal{J}_v$ represents the set of edges incident to node v, and d_v is the known flow demand associated to node v. Furthermore, given an edge $j \in \mathcal{J}$, the corresponding pipe equation can be deduced as

$$q_j = \text{sgn}(h_{j,1} - h_{j,2}) \cdot c(|h_{j,1} - h_{j,2}|)^{\gamma} \tag{21}$$

where q_j is the flow of edge j, $h_{j,1}$ and $h_{j,2}$ are the heads ($p_{j,1}$ and $p_{j,2}$ pressures plus elevation offsets) at the nodes adjacent to edge $j = (v_{j,1}, v_{j,2})$, and c and γ are parameters modelling physical properties of the pipe, such as length, inside diameter, minor losses, and roughness.

Therefore, the DMA model comprises 883 flow balance equations and 927 pipe flow equations, that depend on 927 unknown flow variables and 883 unknown pressure variables. The resulting structural model is depicted in Fig. 2 in matrix form where the equation set corresponds to rows and the variable set corresponds to columns. A dot in the (i, j) element indicates that there exists an edge incident to equation e_i and variable x_j. Note that the structural model of the DMA network is a just-determined model where all unknown variables can be computed, i.e. the model can be used for system simulation.

Since leaks are represented in the nodes where a flow balance takes place, flow balance equations corresponding to XX type nodes will be considered as fault equations when applying (10)-(11).

A fault sensitivity matrix has also been obtained using the EPANET hydraulic simulator [19]. 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:

$$q_f = C_e \cdot P^{P_{exp}} \tag{22}$$

where C_e is the emitter coefficient, q_f is the flow rate, P is the available pressure at the considered node and P_{exp} is the pressure exponent. EPANET permits the value

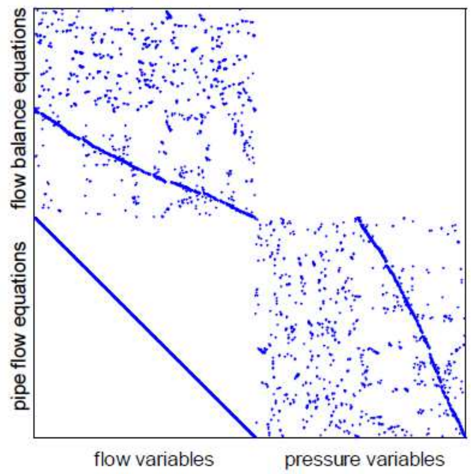

Fig. 2: Structural model of the DMA network

of the emitter coefficient to be specified for individual leak sites, but the pressure exponent can only be specified for the entire network.

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 311×448 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 2.1. Although the fault sensitivity matrix depends on the leakage size, diagnosability properties are robust against this uncertainty. In this work, an average leakage size of 1.5 l/s has been considered in the simulations.

4.3 Clustering analysis

Recall from Section 3 that clustering techniques should be applied beforehand in order to accommodate the time complexity of the optimisation problem. The methodology described in Section 3.1 has been applied to the data set (311 normalised rows of the FSM $\boldsymbol{\Omega}$) in order to set up a reduced set of 25 candidate pressure sensors. As it later will be shown in Section 4.6, $n_r = 25$ is a convenient cardinality for the reduced candidate sensor set. The most time demanding methodology to solve the sensor placement optimisation problem will be the sensitivity analysis approach, requiring a solving time of 27 hours. The structural analysis approach will only require 103 minutes, thus a bigger candidate sensor set could be accepted. However, since a comparison of both approaches is targeted, the size of the candidate sensor set is decided based on the most time demanding one.

Firstly, ECM clustering algorithm has been used to classify the data set into $\ell = 5$ clusters (the same number of clusters as the maximum number of sensors to be installed). Provided the plausibility matrix (5) obtained from the clustering algorithm, a hard partition has been done that assigns each element to its highest plausibility cluster, according to (6). Figure 3 depicts the 5 network node clusters in different colors, where the closest nodes to the centroid have been highlighted in every cluster.

Next, Algorithm 1 has been applied to obtain the N most representative sensors of every cluster, with $N = 5$ given by (7). The resulting reduced set $\boldsymbol{\mathcal{S}}'$ with $|\boldsymbol{\mathcal{S}}'| = N \times \ell = 25$ candidate pressure sensor places suggested by the clustering approach is displayed in Fig. 4 as blue circles.

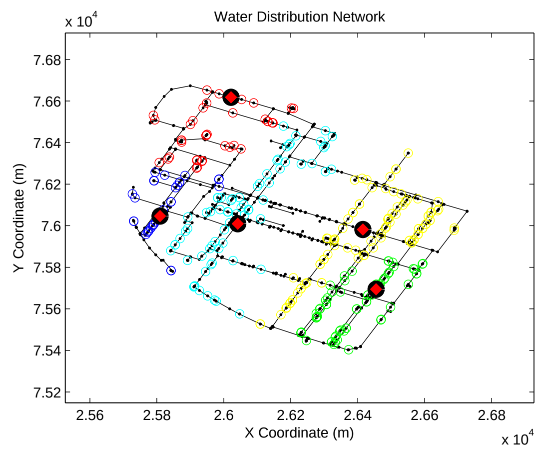

Fig. 3: Clustering results

4.4 Structural analysis approach

The structural analysis approach aims at maximising the isolability index. To fully isolate all 448 possible leaks, the required isolability index should be $\binom{448}{2} = 100128$. However, according to structural analysis, installing all 311 candidate sensors, the isolability index would just be 100099. Achieving a better performance would require installing more sensors than those provided in the candidate sensor set. So, there is a trade-off between the diagnosis performance and the number of installed sensors. Since the maximum number of sensors to install is 5, the maximum achievable isolability index is expected to be less than 100099.

Algorithm 2 is applied to solve the sensor placement problem with $m_p = 5$. The algorithm returns the optimal sensor configuration provided in Fig. 4, highlighted as red starred nodes.

With these 5 sensors all leaks can be detected and the isolability index amounts to 100073, which means that 99.9% of leak pairs are isolable. This is the highest diagnosis performance that can be achieved by a sensor configuration satisfying the

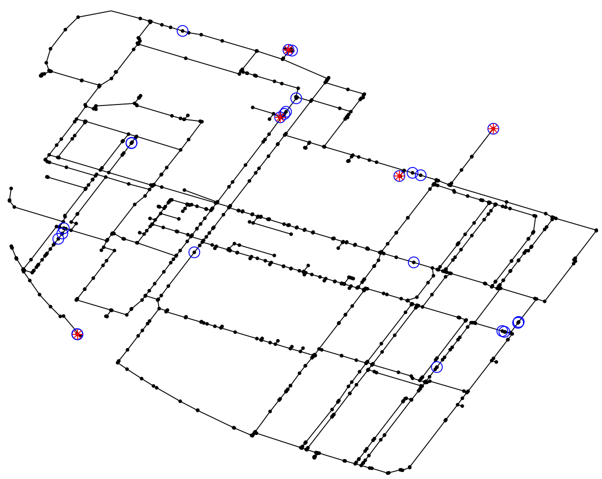

Fig. 4: DMA network sensor placement results under structural analysis approach

stated budgetary constraint. In this case study, no cheaper sensor configuration can achieve better diagnosis performance. Figure 5 provides an evidence to back this assertion up. It displays the highest isolability performance that can be achieved by a sensor configuration of a given size. Remark that, under the structural analysis approach, the isolability performance decreases with the size of the sensor configuration, till null when installing just one pressure sensor into the water network. Note however that with one sensor less, the performance just slightly decreases. Thus a savings in the initial budget could be considered by the water distribution company.

4.5 Sensitivity analysis approach

The sensitivity analysis approach aims at minimising the mean average worst leak expansion distance, which requires a set of projection angle thresholds to be specified. Recall from Section 3.3.2 that, in addition to the FSM matrix $\boldsymbol{\Omega}$, a leak node distance matrix and a set of projection angle thresholds should be specified. Matrix $\mathbf{D}$ has been obtained from geographical data contained in the EPANET model and the following projection angle threshold set has been considered: $\mathcal{A}_{th} = \{10, 20, 30, 40, 50, 60\}$.

As stated in Section 3.3.2, an exhaustive search is applied to solve the sensor placement problem with $m_p = 5$. The optimal sensor configuration is displayed as red starred nodes in Fig. 6.

Installing these 5 pressure sensors, all 448 leaks are detectable and the mean average worst leak expansion distance is 698.23 m. As in the structural analysis

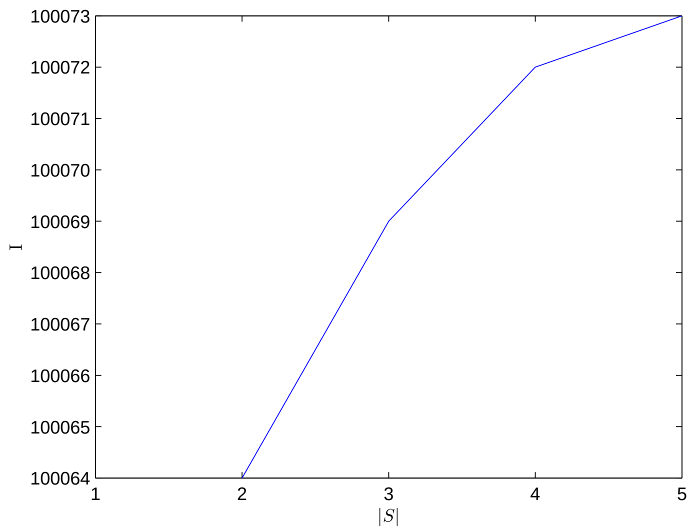

Fig. 5: Branch and bound search results for several sensor configuration sizes.

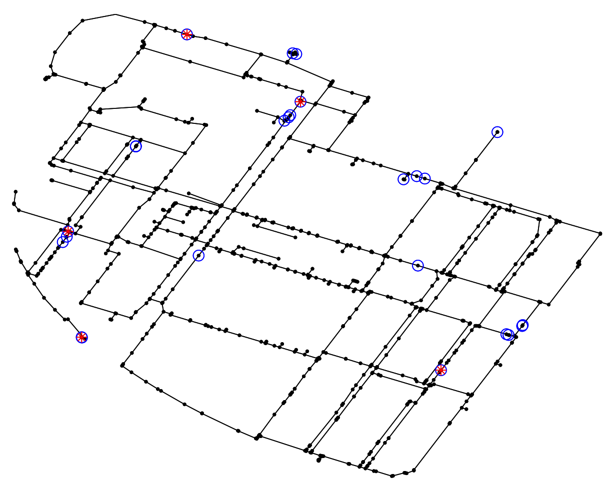

Fig. 6: DMA network sensor placement results under sensitivity analysis approach

approach, the best diagnosis performance is achieved by a 5-sensor configuration. Figure 7 displays the smallest mean average worst leak expansion distance that can be achieved by a sensor configuration of a given size. Remark that, in this case study, the performance index shows a monotonically decreasing trend over values of $|\mathcal{S}|$. Thus the best leak location performance is achieved with the 5 pressure sensors displayed in Fig. 6. Note again that with fewer sensors, the performance just slightly deteriorates. Thus a savings in the initial budget could be considered by the water distribution company.

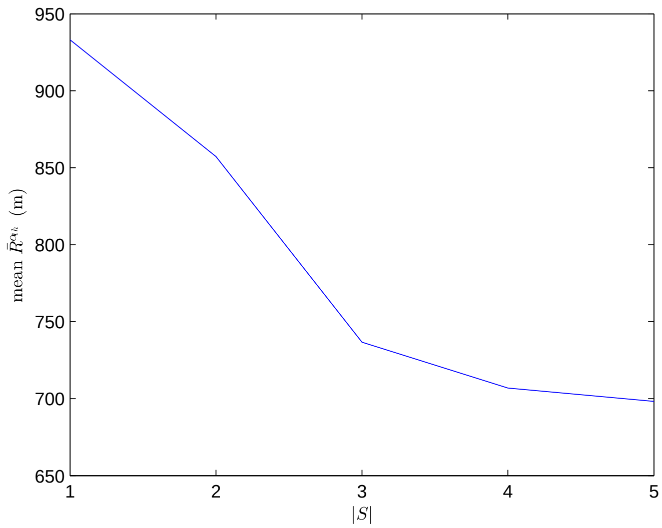

Fig. 7: Exhaustive search results for several sensor configuration sizes.

4.6 Discussion

Regarding the search strategy performance issues, with 25 candidate sensors there are $\binom{25}{5} + \binom{25}{4} + \ldots + \binom{25}{1} = 68405$ possible sensor configurations to check. However, under the structural analysis approach, Algorithm 2 with $m_p = 5$ just needs to inspect 4874 sensor configurations. Algorithm 2 provides the optimal sensor configuration in 103 minutes. Tests have been run in an Intel Core i7-4702MQ @ 2.20GHz HP notebook with 16 GB RAM and 64-bit Windows 10.

Alternatively, as an exhaustive search is applied under the sensitivity analysis approach, the leak monitoring performance index must be evaluated for every sensor configuration and every projection angle threshold. Since there are 68405 sensor configurations to check and 6 projection angle thresholds, the average worst leak expansion distance must be evaluated $68405 \times 6 = 410430$ times. Solving the sensor placement problem under the sensitivity analysis approach takes over 27 hours.

Table 1 provides a summary of the solution found under both approaches.

Table 1: Sensor placement results comparison.

Approach	$\mathcal{S}$	$I(\mathcal{S})$	$\bar{R}^{\overline{\alpha_{th}}}$ (m)	Computation time
Structural analysis	$\{3, 9, 124, 199, 305\}$	100073	793.4	103 s
Sensitivity analysis	$\{8, 67, 207, 249, 271\}$	100023	698.23	27 h

As expected the sensitivity analysis approach provides smaller leak location uncertainty in terms of leak expansion distance at the expense of slightly decreasing the isolability index.

Both approaches provide a different methodology to solve the sensor placement problem stated on Section 2.2. On the one hand, the structural analysis approach is more efficient since the formulation allows for a branch and bound search strategy. However, structural models are a simple description of the network and only best case results can be computed. This methodology is better suited for an early stage of the network design. On the other hand, the sensitivity analysis approach requires a bigger computation time since a highly inefficient exhaustive search is applied. However, the monitoring performance index has a more practical meaning: a leak location uncertainty measure in terms of distance. The search efficiency of the sensitivity analysis approach could be improved by applying other search strategies such as genetic algorithms or simulated annealing at the expense of global optimality.

5 Conclusions

This work presents two optimal sensor placement strategies that maximise some diagnosis specifications for a water distribution network. The goal is to characterise and determine a sensor set that guarantees a maximum degree of diagnosability while a budgetary constraint is satisfied. The first methodology is based on a structural model of the water network and a branch and bound search while the second strategy is based on pressure sensitivity matrix analysis and an exhaustive search strategy. To reduce the size and complexity of the optimisation problem the strategies are combined with clustering techniques. The developed strategies are successfully applied to a DMA in the Barcelona water distribution network to decide the best location of pressure sensors for leak monitoring. However, in both approaches only a sub-optimal solution is attained. Thus, the sensor search could be improved using other types of optimization methods that provide some guarantee regarding the solution optimality.

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] T. Farley, *Losses in Water Distribution Networks.* UK: IWA Publishing UK, 2003.

[3] 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.

[4] M. V. Casillas, L. E. Garza, and 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, August 2012, pp. 570–575.

[5] L. Travé-Massuyès, T. Escobet, and X. Olive, "Diagnosability analysis based on component supported analytical redundancy relations," *IEEE Trans. Syst., Man, Cybern. A*, vol. 36, no. 6, pp. 1146–1160, 2006.

[6] M. Krysander, "Design and analysis of diagnosis systems using structural analysis," Ph.D. dissertation, Linköping Univ., Linköping, Sweden, Jun. 2006.

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

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

[9] 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*, vol. 14, no. 5, pp. 795–803, 2014.

[10] M. V. Casillas, V. Puig, L. E. Garza-Castañón, and A. Rosich, "Optimal sensor placement for leak location in water distribution networks using genetic algorithms," *Sensors*, vol. 13, no. 11, pp. 14 984–15 005, 2013.

[11] R. Sarrate, J. Blesa, and 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, June 2014, pp. 109–114.

[12] 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.

[13] J. Blesa, F. Nejjari, and R. Sarrate, "Robust sensor placement for leak location: analysis and design," *Journal of Hydroinformatics*, vol. 18, no. 1, pp. 136–148, 2016.

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

[15] A. L. Dulmage and N. S. Mendelsohn, "Covering of bi-partite graph," *Canad. J. Math*, vol. 10, pp. 517–534, 1958.

[16] M. Krysander and E. Frisk, "Sensor placement for fault diagnosis," *IEEE Trans. Syst., Man, Cybern. A*, vol. 38, no. 6, pp. 1398–1410, 2008.

[17] F. Nejjari, R. Sarrate, and J. Blesa, "Optimal pressure sensor placement in water distribution networks minimizing leak location uncertainty," *Procedia Engineering*, vol. 119, pp. 953–962, 2015.

[18] 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.

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