

Leak localization method for water distribution networks using a data-driven model and Dempster-Shafer reasoning

A. Soldevila, J. Blesa, T. N. Jensen, S. Tornil-Sin, R. M. Fernandez-Canti, V. Puig

Abstract—This paper presents a new data-driven method for leak localization in water distribution networks. The method uses the information provided by a set of pressure sensors installed in some internal network nodes in addition to flow and pressure measurements from inlet nodes. Pressure measurements are recorded under leak-free network operation and a water distribution network data-driven model of the pressure at each sensed node is adjusted. The pressure estimation from this model is complemented by a Kriging spatial interpolation technique to estimate pressure in the nodes which are not sensed, leading to a pressure reference map. Leak localization is based on the comparison of this reference pressure map with the current pressure map which is obtained by applying Kriging directly to the pressure measurements provided by sensors. The key element in this comparison is the use of Dempster-Shafer theory for reasoning under uncertainty. The successful application of the proposed methodology to two real-data case studies is presented.

I. INTRODUCTION

Water leaks are present to some extent in all Water Distribution Networks (WDNs) and are estimated to account up to 30% of the total amount of extracted water [1]. This is a very significant amount since water is a precious resource, especially in many parts of world where it is necessary to satisfy water demands of a growing population and sometimes in presence of drought periods, which are increased by the climate change. Fault diagnosis and security in water systems are key challenges that will become even more crucial in the years ahead [2].

Several works have been published dealing with the leak detection and isolation (localization) problem in WDNs (see [3] and references therein). For example, in [1], a review of transient-based leak detection methods is offered. In [4], a method is proposed to identify leaks using blind spots based on previous leak detection that uses the analysis of acoustic and

vibration signals [5], and models of buried pipelines to predict wave velocities [6]. More recently, [7] presents a method to localize leaks using Support Vector Machines (SVM) that analyzes data obtained by a set of pressure control sensors installed in a pipeline network to localize and calculate the size of the leak. The use of k -Nearest Neighbors, Bayesian and neuro-fuzzy classifiers in leak localization have been also recently proposed in [8], [9] and [10] respectively.

Some of the recent proposed leak localization methods use pressure sensors inside the WDNs. Pressure sensors are cheaper and easier to install than flow sensors. Therefore, a setup with a few pressure sensors installed in the WDN is an attractive option for utilities. One such setup is presented in [11], where a model-based method that relies on the pressure measurements and a leak sensitivity analysis is proposed. In this methodology, pressure residuals, i.e. differences between pressure measurements provided by sensors and the corresponding estimations obtained by using a hydraulic model, are used. These residuals are computed on-line and compared against associated thresholds that take into account the effects of modeling uncertainty and noise. When some of the residuals exceed their thresholds, the residuals are matched against the leak sensitivity matrix in order to discover which of the possible leaks is present. Although this approach has good efficiency under ideal conditions, its performance decreases due to the nodal demand uncertainty and measurement noises [12], [13]. The methodology has been improved in [14], where a comparison of several geometric approaches for leak localization are compared and an analysis along a time horizon is proposed. Additional improvements are achieved by processing the residuals using classifiers [8], [9].

In this paper, we propose a new method to deal with the leak localization problem in WDNs, or in partitioned zones called District Metered Areas (DMAs) where the flow and pressure at the inlets are measured and some pressure sensors are placed inside. It is here assumed that the leak detection task is already been performed, e.g. using some of the approaches reviewed in [1]. The proposed method is based on three key ideas:

- Use of a data-driven adjusted model to estimate nominal (without leak) expected pressure values according to the current operating conditions in the internal nodes that are equipped with pressure sensors.
- Use of the Kriging spatial interpolation technique to estimate the pressure at the network nodes which are not equipped with sensors based on the hydraulic proximity.
- Use of the Dempster-Shafer theory for reasoning under uncertainty to analyze the differences between the nominal and the currently observed network behaviour.

The proposed method has two clear benefits. First, it is

This work has been funded by the Ministerio de Economía, Industria y Competitividad (MEICOMP) of the Spanish Government through the project DEOCS (ref. DPI2016-76493) and by the Catalan Agency for Management of University and Research Grants (AGAUR), the European Social Fund (ESF) and the Secretary of University and Research of the Department of Companies and Knowledge of the Government of Catalonia through the grant FI-DGR 2015 (ref. 2015 FI B 00591). The second author acknowledges the support from the Serra Hünter program. The third author would like to acknowledge the grant from the Innovation Fund Denmark (IFD) under File No. 6155-00002B

A. Soldevila, J. Blesa, S. Tornil-Sin, R. M. Fernandez-Canti and V. Puig are with the Research Center for Supervision, Safety and Automatic Control (CS2AC), Rambla Sant Nebridi, s/n, 08022 Terrassa, Spain. (e-mail: adria.cbn@gmail.com, joaquim.blesa@upc.edu, sebastian.tornil@upc.edu, rosa.mari.fernandez@upc.edu, vicenc.puig@upc.edu)

J. Blesa, S. Tornil-Sin and V. Puig are with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain.

T. N. Jensen is an independent researcher. Aalborg, Denmark. (e-mail: noergaard_jensen@hotmail.com)

a data-driven method that does not require the computation, the recursive updating and difficult calibration of a hydraulic model. An estimation of the hydraulic resistance between nodes in the network is used instead. Second, it only needs historical data from normal operation behaviour, in contrast to conventional data-driven methods that require historical data for all the considered abnormal behaviours.

The rest of the paper is organized as follows. In Section II, the data-driven model of the network and the spatial interpolation based on Kriging are presented and the leak localization problem is introduced as the evaluation of residuals generated by the data-driven model. Section III proposes an improved localization using a Dempster-Shafer time reasoning. The method is summarized in Section IV. Section V illustrates the application of the leak localization methodology to two real DMAs. Finally, Section VI draws the main conclusions of the work.

Notation: Matrices are denoted using capital and bold letters, while lower and bold letters denote vectors, and italic letters denote scalars. Sets and graphs are represented using calligraphic letters.

II. WDN DATA-DRIVEN MODEL

A water distribution network can be described by a directed graph

$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. Here, $\mathcal{V} = \{v_1, \dots, v_{n_m}\}$ is the set of vertices, which represent connections between the components of the network, i.e. the n_n internal nodes and the n_r inlet nodes, giving a total number of nodes $n_m = n_n + n_r$. The set $\mathcal{E} \in \mathcal{V} \times \mathcal{V}$ with $\mathcal{E} = \{e_1, \dots, e_{n_p}\}$ is the set of edges, which represents the n_p pipes of the network. A path \mathcal{P} in the graph \mathcal{G} is a sequence $\{x_i\}_{i=1}^{\ell}$ such that $x_i \in \mathcal{V}$, $x_i x_{i+1} \in \mathcal{E}$ and $x_i \neq x_j$ for every pair $i, j \in \{1, 2, \dots, \ell\}$. Furthermore, we consider that the edge $e_j \in \mathcal{P}$ if $e_j = x_k x_{k+1}$ with $x_k \in \mathcal{P}$ and $x_{k+1} \in \mathcal{P}$.

Let \mathbf{p} be the vector of absolute pressures at the nodes and $\Delta \mathbf{p}$ be the vector of differential pressures across the pipes, both in meters water column [mwc], then

$$\Delta \mathbf{p} = \mathbf{H}^T \mathbf{p} = f(\mathbf{q}) - \mathbf{H}^T \mathbf{h} \quad (1)$$

where $\mathbf{p} \in \mathbb{R}^{n_m}$, $\mathbf{q} \in \mathbb{R}^{n_p}$ is the vector of volumetric flow in the edges in cubic meters per second [m^3/s], \mathbf{H} with size $n_m \times n_p$ is the incidence matrix¹, $f: \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_p}$ and $f(\mathbf{q}) = (f_1(q_1), \dots, f_{n_p}(q_{n_p}))$. The function $f_i(q_i)$ describes the flow dependent pressure drop due to the hydraulic resistance in the i^{th} edge.

For turbulent flow in the pipes, the Darcy-Weisbach equation is a good approximation of the pressure drop due to the hydraulic resistance of the pipe [16]. The expression in the i^{th} pipe is given by

$$f_i(q_i) = \frac{8C_i L_i}{\pi^2 g D_i^5} |q_i| q_i \quad (2)$$

where C_i is the coefficient of surface resistance (which is in part determined by the flow, but we will here consider it

constant) with adimensional units, L_i is the length of the pipe in [m], g is the gravitational acceleration in [m/s^2] and D_i is the diameter of the pipe in [m].

The term $\mathbf{H}^T \mathbf{h}$ is the pressure drop across the pipes due to the difference in geodesic level (i.e. elevation), in meters [m], between the ends of the pipes, with $\mathbf{h} \in \mathbb{R}^{n_m}$ the vector of geodesic levels at each vertex.

Knowledge of the length and diameter of all the pipes in the network as well as their interconnections (topological structure) and the geodesic level at all nodes is required. Therefore we assume that these values can be obtained from an available database.

A. Interpolation between boundary conditions

Here, we will present the reduced order model which will be used for calculating the expected nominal pressures at the measured internal nodes, based on the known values of flows and pressures at the inlets and the previous measured values in the sensed nodes. The derivation of the model is presented in [17].

Nodes in a network can be classified as inlet nodes (directly connected to tanks or reservoirs) and inner nodes (associated to network internal junctions and consumer nodes). In the following, variables associated to inlet nodes are indicated with a (1) superscript, while variables associated to inner nodes present a (2) superscript. For each node (inlet or inner node), the three variables of interest are the elevation, the pressure, and the head, represented in the paper with the letters "h", "p" and "y", respectively. For each node i , the head is defined as the addition of the elevation and the pressure at the node, i.e: $y_i = h_i + p_i$.

Now, as suggested in [18], consider the case where

$$\mathbf{y}^{(1)} = \mathbf{p}^{(1)} + \mathbf{h}^{(1)} = \kappa \mathbf{1} \quad (3)$$

for some $\kappa \in \mathbb{R}$ which is the total head at the inlets in [mwc] and where $\mathbf{1}$ denote the vector consisting of ones.

The constraint (3) has the physical implication that the total head at all inlets must be equal at all times. At first glance this seems like a hypothetical assumption. However, it appears that at least in some networks this assumption is fulfilled which we have observed through measurements of the inlet pressures (as e.g., the real networks considered as case study in the paper). Moreover, we can refer the reader to the reference [17] in which a discussion is found as to why controllers would have to fulfil this assumption at least in networks with low total consumption. For networks fulfilling (2) and (3), we have the following proposition from [17].

Proposition 1: If the vector $\mathbf{p}^{(1)}$ of pressures at inlet nodes fulfils (3) and $f_i(\cdot)$ fulfils (2), then the pressure at the i^{th} internal node follows the expression

$$p_i^{(2)} = \alpha_i \sigma^2 + \kappa + \gamma_i \quad (4)$$

where α_i and γ_i are parameters dependent on the network topology and the distribution of demands in the network, and $\sigma > 0$ is the total inlet flow (sum of all inlet flows) in [m^3/s]. The total inlet flow σ is typically well-known since inlet flows are measured.

¹The incidence matrix of the graph, that describes the interconnection of vertices and edges, is defined as usual in graph theory, see as e.g. [15].

In general, the parameter α_i in (4) is time-varying. However, if the proportion of the total consumption that any particular end-user at any point in time has is constant then α_i is also constant. The latter is the case if all end-users follow the same profile, for instance if all users are residential.

Since the model (4) of $p_i^{(2)}$ is linear in the parameters α_i and γ_i , standard parameter identification methods [19] can be used to identify these parameters based on measurements of $p_i^{(2)}$, σ and $\mathbf{p}^{(1)}$ over a number of time periods. Subsequently, having identified the parameters, the model (4) can be used to predict the expected pressure $\hat{p}_i^{(2)}$ at a measurement vertex, since the variables σ and $\mathbf{p}^{(1)}$ are measured and $\mathbf{h}^{(1)}$ is known.

B. Pressure estimation in unsensed nodes

In practice, in real WDNs, pressure measurements $p_i^{(2)}$ are only available in a limited number of internal nodes, due to the costs of installing and maintaining pressure sensors. The location of the installed sensors is given by the index set \mathcal{S} of internal nodes equipped with sensors

$$\mathcal{S} = \{s_1, \dots, s_{n_s}\} \quad (5)$$

where n_s is the number of pressure sensors installed.

To obtain the pressure in the nodes without pressure sensors installed, an interpolation technique can be used. In this work, the Kriging approach is used. Kriging is a well-known interpolation method in the area of geostatistics and it can be seen as a multivariate regression approach, see [20] for a recent review. The basic idea of Kriging is to estimate the value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point.

Consider that the WDN is working under some operating conditions \mathbf{c} given by the total inlet flow and user demands as proposed in [21]. Since the user demands are rarely measured, the vector \mathbf{c} contains the total head at the inlets κ , obtained by taking into account PRVs or gravity and the total inlet flow, σ .

The vector \mathbf{q} consisting of flows in all the pipes in the network, can be partitioned as

$$\mathbf{q} = \mathbf{a}\sigma \quad (6)$$

where \mathbf{a} determines the distribution of flow in pipes and is unknown and the total inlet flow σ is typically well-known since inlet flows are measured values.

It was shown in [17] (or in [22] for the single inlet case), that when the distribution of the total inlet flow among the end-users remain constant (i.e. all consumers have the same consumption profile), then the vector \mathbf{a} is constant as well.

Having knowledge of the absolute pressure, say $p_i^{(2)}$, upstream of the k^{th} pipe, then the head $y_j^{(2)}$ downstream the pipe can be obtained from (1) as

$$y_j^{(2)} = p_j^{(2)} + h_j^{(2)} = p_i^{(2)} + h_i^{(2)} - f_k(q_k) \quad (7)$$

Now, we re-write (2) using the definition in (6) and obtain

$$f_k(q_k) = \frac{8C_k L_k}{\pi^2 g D_k^5} |a_k \sigma| a_k \sigma = \frac{8C_k L_k}{\pi^2 g D_k^5} (|a_k| a_k) \sigma^2 \quad (8)$$

where the latter equality is due to the fact that $\sigma > 0$.

If we assume that we have knowledge of the length and diameter of pipes, we can rearrange (8) into known and unknown quantities to obtain

$$f_k(q_k) = \overbrace{C_k (|a_k| a_k)}^{\theta_k} \frac{8L_k}{\pi^2 g D_k^5} \sigma^2 = \theta_k \frac{8L_k}{\pi^2 g D_k^5} \sigma^2 \quad (9)$$

θ_k is now an unknown parameter of the k^{th} pipe and it will be estimated using the Kriging approach. Now (7) can be expressed as

$$\hat{y}_j^{(2)} = p_i^{(2)} + h_i^{(2)} - \hat{\theta}_k \frac{8L_k}{\pi^2 g D_k^5} \sigma^2 \quad (10)$$

where the term $\frac{8L_k}{\pi^2 g D_k^5} \sigma^2$ will be used as the independent variable in the interpolation.

Since multiple pipe paths may connect j^{th} and i^{th} vertices, we suggest to use the shortest “weighted” pipe length (e.g. weighted by pipe diameters) as the interpolation variable in the Kriging interpolation. That is, if we let

$\mathcal{P}_{ij} = \{\mathcal{P}_{ij}(1), \dots, \mathcal{P}_{ij}(n)\}$ denote the set of paths connecting nodes i and j , then we define the shortest weighted pipe length \bar{D}_{ij} as

$$\bar{D}_{ij} = \arg \min_{\mathcal{P}_{ij}(k) \in \mathcal{P}_{ij}} \sum_{e_z \in \mathcal{P}_{ij}(k)} \frac{L_z}{D_z^5} \quad (11)$$

In other words, for every path between the two nodes, we calculate the sum of $\frac{L_z}{D_z^5}$ for all edges e_z in the path. The path which returns the smallest sum is the one used in the interpolation.

Thus, the estimation of an unmeasured head $\hat{y}_i^{(2)}$ in the i^{th} node can be computed by means of a fitted Kriging model [23] as

$$\hat{y}_i^{(2)}(\mathbf{c}, \mathcal{S}) = \mu(\mathbf{c}) + \varepsilon(\chi(\mathbf{c}), \phi(\mathbf{c}), \bar{\mathcal{D}}_{i:}(\mathcal{S})) \quad (12)$$

where \mathbf{c} is the vector that defines the operation conditions (i.e. total head in inlets κ and total inflow σ), \mathcal{S} is the sensor set defined in (5), and $\mu(\mathbf{c})$ provides a value that represents the constant part of the interpolation given a particular operating condition \mathbf{c} . On the other hand, the function $\varepsilon(\chi(\mathbf{c}), \phi(\mathbf{c}), \bar{\mathcal{D}}_{i:}(\mathcal{S}))$ is the spatially correlated part of the variation where $\chi(\mathbf{c})$ is a polynomial function, $\phi(\mathbf{c})$ is the correlation function and $\bar{\mathcal{D}}_{i:}(\mathcal{S})$ is a vector containing the shortest weighted pipe distances, defined in (11), between the node i and the n_s sensors, as $\bar{\mathcal{D}}_{i:}(\mathcal{S}) = (\bar{D}_{i s_1}, \dots, \bar{D}_{i s_{n_s}})$. Both terms $\mu(\mathbf{c})$ and $\varepsilon(\cdot)$ are obtained in the fitting process as well as the functions $\chi(\mathbf{c})$ and $\phi(\mathbf{c})$. The fitting process consists in a least squares error minimization problem considering available pressure measurements and distances between the n_n nodes and the n_s sensors installed. **The reader is referred to the reference [21] for more details about interpolation model** (12).

Finally, the pressure in the unmeasured node i can be obtained from (12) by subtracting the geodesic level as

$$\hat{p}_i^{(2)}(\mathbf{c}, \mathcal{S}) = \hat{y}_i^{(2)}(\mathbf{c}, \mathcal{S}) - h_i^{(2)} \quad (13)$$

C. Basic residual computation and evaluation

The leak localization problem, posed as a Fault Detection and Isolation (FDI) problem in the literature, typically assumes that only one leak can occur at a time. Additionally, it is also usually assumed that leaks can only occur in the nodes of the network (e.g. as considered in [11], [24], or [9]), which makes the number of potential leaks equal to the number of nodes of the considered network.

As stated in the introduction, the proposed leak localization technique is applied after the detection of a leak in the monitored network. Therefore, the used measurements are assumed to be captured under a leaky situation. Consider the presence of a leak l_j with magnitude l and acting on the node j of the network. If leak-free historical data of pressure sensors installed in inner nodes are available for all possible operating conditions, including the current operating conditions \mathbf{c} . Then, residual pressures in internal nodes can be computed as

$$r_i = \tilde{p}_i^{(2)}(\mathbf{c}) - p_i^{(2)}(\mathbf{c}^{l_j}) \quad \forall i \in \mathcal{S} \quad (14)$$

where $p_i^{(2)}(\mathbf{c}^{l_j})$ is the pressure value measured by the inner pressure sensor i under boundary conditions \mathbf{c}^{l_j} (total head and inflow in inlets) under the presence of leak l_j . On the other hand, $\tilde{p}_i^{(2)}(\mathbf{c})$ is its estimation considering boundary conditions \mathbf{c} in a leak-free scenario that can be computed reformulating (4) as

$$\tilde{p}_i^{(2)}(\mathbf{c}) = \hat{\alpha}_i \sigma^2 + \kappa + \hat{\gamma}_i \quad (15)$$

where $\hat{\alpha}_i$ and $\hat{\gamma}_i$ are estimated from historical leak-free data applying classical identification methods based on least squares estimation.

Considering Kriging spatial interpolation to estimate the pressure in the unmeasured nodes (13) both in non-leak $p_i^{(2)}(\mathbf{c}^{l_j})$ and leak $\tilde{p}_i^{(2)}(\mathbf{c})$ conditions $\forall i \in \mathcal{S}$, the residual vector for all internal nodes can be approximated by

$$\hat{\mathbf{r}} = \hat{\mathbf{p}}^{(2)}(\mathbf{c}, \mathcal{S}) - \hat{\mathbf{p}}^{(2)}(\mathbf{c}^{l_j}, \mathcal{S}) \quad (16)$$

where

- $\hat{\mathbf{p}}^{(2)}(\mathbf{c}, \mathcal{S}) = (\hat{p}_1^{(2)}(\mathbf{c}, \mathcal{S}), \dots, \hat{p}_{n_n}^{(2)}(\mathbf{c}, \mathcal{S}))$ is the vector that approximates the pressure map in the WDN under operating conditions \mathbf{c} in a leak-free scenario (reference map).
- $\hat{\mathbf{p}}^{(2)}(\mathbf{c}^{l_j}, \mathcal{S}) = (\hat{p}_1^{(2)}(\mathbf{c}^{l_j}, \mathcal{S}), \dots, \hat{p}_{n_n}^{(2)}(\mathbf{c}^{l_j}, \mathcal{S}))$ is the vector that approximates the pressure map in the WDN under operating conditions \mathbf{c}^{l_j} in a leak scenario with a magnitude l in node j .

Then, leaky node localization can be estimated as the one with the largest pressure residual component (as used in [25] and [26] and analytically justified in [27]), i.e.

$$\hat{j} = \arg \max_{i \in \{1, \dots, n_n\}} \{\hat{r}_i\} \quad (17)$$

where $\hat{r}_i \forall i \in \{1, \dots, n_n\}$ are the components of residual vector (16).

The performance when using the leak localization defined in (17) will depend on the number of sensors, their location in the WDN, and the accuracy of the data-driven model (4) to predict the pressure under the operational conditions \mathbf{c} and leak-free conditions.

Note that according to (10), the consumption (σ) affects the value of the residual, providing larger residuals as the consumption increases for the same leak size. To deal with that issue and to obtain a robust leak localization, the Dempster-Shafer reasoning is used here to integrate the temporal development of the residual (16).

III. DEMPSTER-SHAFFER REASONING

A. Dempster-Shafer evidence theory

The Dempster-Shafer (DS) evidence theory [28], [29] is a general framework for reasoning with uncertainty that allows to fuse information from different sources. In the same way, evidences from the same source but in different time instants can also be considered. In [30], the use of DS reasoning approach is proposed as an alternative to Bayes reasoning when ambiguous hypotheses exist. This is the case of the leak localization approach with a reduced number of sensors, since several leaks presents a similar pressure pattern. On the other hand, the Bayes reasoning considers all of the supported hypotheses must be mutually exclusive. This is the main motivation of using DS reasoning in this paper.

Given a set of individual hypothesis \mathcal{A}_i , the DS theory considers all the possible combinations, i.e. the power set \mathcal{P} . For each subset \mathcal{A} of the power set \mathcal{P} , three quantities are defined: the mass $m(\mathcal{A})$, the belief $bel(\mathcal{A})$, and the plausability $pl(\mathcal{A})$. The mass is associated to the information provided by a given source and it can also be considered as a subjective probability. Hence, a probability mass function (pmf) is a function that satisfies: $m(\emptyset) = 0$; $m(\mathcal{A}) \in [0, 1], \forall \mathcal{A} \in \mathcal{P}$; $\sum_{\mathcal{A} \in \mathcal{P}} m(\mathcal{A}) = 1$. The belief $bel(\mathcal{A})$ is defined as the sum of all the masses of all the subsets of \mathcal{A} , i.e. $bel(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} m(\mathcal{B})$. The plausability $pl(\mathcal{A})$ is the sum of all the masses of the subsets of \mathcal{P} that intersect \mathcal{A} , i.e. $pl(\mathcal{A}) = \sum_{\mathcal{B} \cap \mathcal{A} \neq \emptyset} m(\mathcal{B})$. The belief and the plausability define an interval where the true probability of the considered hypothesis lies, i.e. $bel(\mathcal{A}) \leq m(\mathcal{A}) \leq pl(\mathcal{A})$.

When there are two sources of information the Dempster's rule of combination can be used. Being m_{g1} the mass assignment associated to the first source and m_{g2} the one associated to the other, the joint mass is computed as

$$\begin{aligned} m_{g1,g2}(\mathcal{A}) &= (m_{g1} \oplus m_{g2})(\mathcal{A}) \\ &= \frac{\sum_{\mathcal{B} \cap \mathcal{C} = \mathcal{A} \neq \emptyset} m_{g1}(\mathcal{B}) m_{g2}(\mathcal{C})}{1 - \sum_{\mathcal{B} \cap \mathcal{C} = \emptyset} m_{g1}(\mathcal{B}) m_{g2}(\mathcal{C})} \end{aligned} \quad (18)$$

where \mathcal{B} and \mathcal{C} are subsets of \mathcal{P} and \oplus is the *direct sum* [29]. The constraint $\mathcal{B} \cap \mathcal{C} = \mathcal{A} \neq \emptyset$ limits the interaction of the probability mass functions only to the shared hypotheses and the constraint $\mathcal{B} \cap \mathcal{C} = \emptyset$ contain the rest of combinations of probability mass functions. More precisely, the denominator in (18) is for normalizing the result, i.e. the sum of the resulting

R2-C3

R2-C3

R2-C3

mass probabilities of the resulting set of common hypotheses \mathcal{A} ($m(\mathcal{A})$) is one. This Dempster's rule of combination considering three different sources of information was used in [31] to perform the leak localization task.

The Dempster's rule of combination (18) can be formulated for time reasoning [32] as

$$m_{t_1, t_2}(\mathcal{A}) = m_{t_1}(\mathcal{A}) \oplus m_{t_2}(\mathcal{A}) \quad (19)$$

where t_1 and t_2 are two different time instants for the same set of hypotheses \mathcal{A} . More precisely, for a particular hypothesis \mathcal{A}_i in the set \mathcal{A} we can say

$$m_{t_1, t_2}(\mathcal{A}_i) = \frac{1}{\Lambda} \times \sum_{j=1, \dots, A | \mathcal{A}_i \subseteq \mathcal{A}_j} (m_{t_1}(\mathcal{A}_i) m_{t_2}(\mathcal{A}_j) + m_{t_1}(\mathcal{A}_j) m_{t_2}(\mathcal{A}_i)) \quad (20)$$

where A is the number of hypotheses contained in the set \mathcal{A} and Λ is a normalization factor such that $\sum_{i=1}^A m_{t_1, t_2}(\mathcal{A}_i) = 1$.

B. Probability mass functions for leak localization

The presented Dempster-Shafer framework can be applied to the leak localization problem. Leaks in network nodes can be considered as the individual hypothesis \mathcal{H}_i , and a probability mass function can be associated to normalized pressure residual components of vector (16) as

$$m_t(\mathcal{H}_i) = \frac{\hat{r}_i(t) - \min(\hat{\mathbf{r}}(t))}{\sum_{j=1}^{n_n} (\hat{r}_j(t) - \min(\hat{\mathbf{r}}(t)))} \quad (21)$$

where the size of the hypothesis set \mathcal{H} is the number of internal nodes of the network n_n and

$$\sum_{i=1}^{n_n} m_t(\mathcal{H}_i) = 1 \quad (22)$$

In order to illustrate the proposed leak localization method, we will consider the simplified WDN of Hanoi (Vietnam) introduced in [33] and depicted in Fig. 1. This network has one reservoir, 31 internal nodes ($n_n = 31$) and 34 pipes ($n_p = 34$) and it is assumed that three pressure sensors are placed in internal nodes with indices 2, 8 and 24 ($n_s = 3$). So, for this particular network, the hypothesis set \mathcal{H} will be $n_n = 31$. One for each potential leak location considered i.e. inner node.

However, the use of these probability mass functions presents the particularities that the complete set of hypotheses is composed by singletons (i.e. inside each hypothesis there is no uncertainty among single hypothesis) and fulfills the condition (22). So, the application of the Dempster's rule of combination is analogous to the application of the Bayes rule [34]. Thus, the potential of the Dempster's rule of combination will not be fully exploited. So with that aim, additional hypotheses sets of clustered singleton hypotheses from the set \mathcal{H} are proposed by taking into account spatial and hydraulic information.

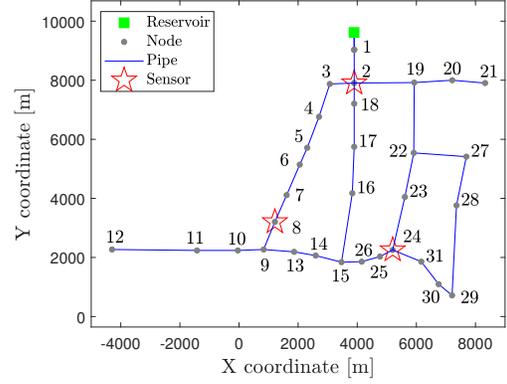


Fig. 1: Simplified Hanoi WDN.

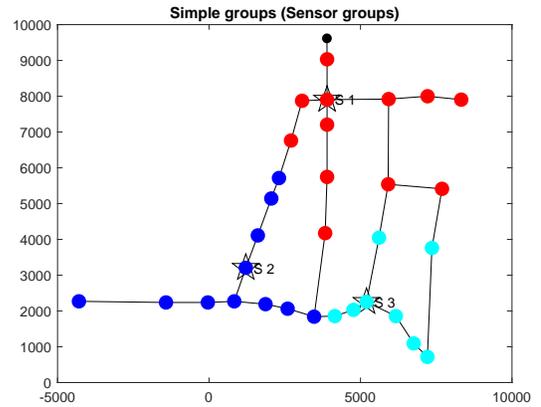


Fig. 2: $\mathcal{K}^{(\mathcal{L})}$ clusters in the simplified WDN of Hanoi. **R2-C5**

C. Clustering

According to section II-B, as nearby nodes present similar hydraulic behaviour [35], the sensor relation to the nodes of the network is analyzed using the hydraulic relation given by the shortest weighted pipe distance \bar{D} . This weighted pipe distance can also be used to infer a set of node clusters that can be considered as additional hypothesis under the Dempster-Shafer framework.

First, n_s clusters are created, each one of them associated to one of the network nodes that are equipped with pressure sensors. Each cluster around a sensed node includes all the network nodes that present a minimum hydraulic distance to this particular sensed node. The obtained set of clusters define a partition of the network similar to a Voronoi diagram. Formally, the cluster associated to a given sensed node l is defined as

$$\mathcal{K}_l^{(\mathcal{L})} = \{v_i^{(2)} \in \mathcal{V}^{(2)} | \arg \min_{j \in \mathcal{S}} \{\bar{D}_{ij}\} = l\} \quad (23)$$

where $l \in \{1, \dots, n_s\}$. The whole set of clusters (one for each sensed node) can be represented as $\mathcal{K}^{(\mathcal{L})}$.

In the simplified WDN of Hanoi, where three pressure sensors are considered, the computed clusters associated to each sensed node are depicted in Fig. 2.

Furthermore, some sensors can be close in hydraulic terms and their measurements can be significantly affected by a leak in the nodes clustered in another cluster from a nearby sensor. To take that into account, a second clustering is performed.

These additional clusters are based on the previous clusters $\mathcal{K}^{(\mathcal{L})}$ and are computed by implementing three steps inside a loop. The loop considers each possible pair $\mathcal{K}_i^{(\mathcal{L})}, \mathcal{K}_j^{(\mathcal{L})} \in \mathcal{K}^{(\mathcal{L})}$ and at each loop iteration the following three steps are implemented.

Firstly, for every cluster $\mathcal{K}_j^{(\mathcal{L})}$ the mean of the shortest weighted pipe distance between the nodes and the sensor is computed as

$$d_j^{(\mathcal{L})} = \frac{1}{|\mathcal{K}_j^{(\mathcal{L})}|} \sum_{v_i^{(2)} \in \mathcal{K}_j^{(\mathcal{L})}} \bar{D}_{ij} \quad (24)$$

where $|\mathcal{K}_j^{(\mathcal{L})}|$ is the number of elements of the cluster $\mathcal{K}_j^{(\mathcal{L})}$.

Secondly, for every couple of clusters $\mathcal{K}_i^{(\mathcal{L})}$ and $\mathcal{K}_j^{(\mathcal{L})}$, with $i \neq j$, a threshold is set as the minimum of the mean of the shortest weighted pipe distance between the couple of sensors as

$$\tau_{ij} = \min\{d_i^{(\mathcal{L})}, d_j^{(\mathcal{L})}\} \quad \forall i \neq j \in \mathcal{S} \quad (25)$$

And thirdly, the elements of a new cluster $\mathcal{K}_k^{(\mathcal{W})}$ can be found from the elements of the couple of clusters $\mathcal{K}_i^{(\mathcal{L})}$ and $\mathcal{K}_j^{(\mathcal{L})}$ as

$$\mathcal{K}_k^{(\mathcal{W})} = \mathcal{K}_{k_i}^{(\mathcal{W})} \cup \mathcal{K}_{k_j}^{(\mathcal{W})} \quad (26)$$

where

$$\mathcal{K}_{k_i}^{(\mathcal{W})} = \{v_l^{(2)} \in \mathcal{K}_i^{(\mathcal{L})} | \bar{D}_{lj} \leq \tau_{ij}\} \quad (27)$$

and

$$\mathcal{K}_{k_j}^{(\mathcal{W})} = \{v_l^{(2)} \in \mathcal{K}_j^{(\mathcal{L})} | \bar{D}_{li} \leq \tau_{ij}\} \quad (28)$$

The potential number of the new resulting clusters is $\frac{n_s!}{2!(n_s-2)!}$ but only some combinations of clusters $\mathcal{K}_i^{(\mathcal{L})}$ and $\mathcal{K}_j^{(\mathcal{L})}$ can lead to a non-empty new clusters $\mathcal{K}_k^{(\mathcal{W})} \neq \emptyset$ in (26). Then, the real number of new clusters $w \leq \frac{n_s!}{2!(n_s-2)!}$ will depend on the sensor configuration.

All the singleton clusters obtained, if any, are not considered since they are already contained in the set \mathcal{H} .

Fig. 3 depicts the elements of cluster $\mathcal{K}_1^{(\mathcal{W})}$ that is one of the three possible additional clusters in the simplified WDN of Hanoi. This cluster is computed by means of Eq. (26) considering clusters $\mathcal{K}_1^{(\mathcal{L})}$ and $\mathcal{K}_2^{(\mathcal{L})}$ associated to sensed nodes 2 and 8, respectively.

The first clusters, $\mathcal{K}^{(\mathcal{L})}$, are converted to a new set of composed hypothesis \mathcal{L} where each cluster is a composed hypothesis from the hypotheses in \mathcal{H} . In the same way, each cluster $\mathcal{K}^{(\mathcal{W})}$ will be a composed hypothesis from the new set of hypotheses \mathcal{W} .

D. Expanded probability mass function for leak localization

Now we have the three sets of hypotheses \mathcal{H} , \mathcal{L} and \mathcal{W} . On the one hand, we have the set of singletons \mathcal{H} where the probability mass functions of each node are computed using (21). On the other hand, we have the sets \mathcal{L} and \mathcal{W}

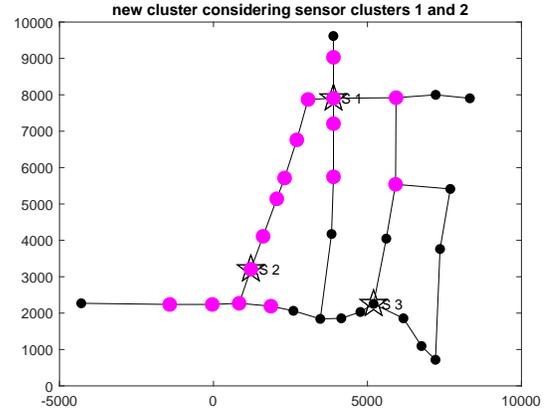


Fig. 3: $\mathcal{K}_1^{(\mathcal{W})}$ cluster in the simplified Hanoi WDN between clusters $\mathcal{K}_1^{(\mathcal{L})}$ and $\mathcal{K}_2^{(\mathcal{L})}$. **R2-C5**

with clustered nodes, but without probability mass functions assigned to them yet.

To assign probabilities into each set of composed hypotheses \mathcal{L} and \mathcal{W} , the mean of the mass probabilities from the set \mathcal{H} of each node inside the cluster is computed as

$$m(\mathcal{L}_i) = \frac{\sum_{v_j^{(2)} \in \mathcal{K}_i^{(\mathcal{L})}} m(\mathcal{H}_j)}{|\mathcal{K}_i^{(\mathcal{L})}|} \quad (29)$$

and the same is done for the composed hypotheses in \mathcal{W} .

Then, the new extended set of mass probabilities $m(\mathcal{H}')$ from the extended set of hypothesis $\mathcal{H}' = \{\mathcal{H}, \mathcal{L}, \mathcal{W}\}$ with size $|\mathcal{H}'| = n_n + n_s + w$ is normalized according to

$$m(\mathcal{H}'_i) = \frac{m(\mathcal{H}'_i)}{\sum_{j=1}^{n_n} m(\mathcal{H}_j) + \sum_{j=1}^{n_s} m(\mathcal{L}_j) + \sum_{j=1}^w m(\mathcal{W}_j)} \quad (30)$$

in order to fulfill the condition $\sum_{i=1}^{|\mathcal{H}'|} m(\mathcal{H}'_i) = 1$.

So, considering the simplified Hanoi WDN, as $n_n = 31$, $n_s = 3$ and the number of additional clusters $w = \frac{3!}{2!1!} = 3$, the number of hypotheses of the extended set is $|\mathcal{H}'| = 37$.

E. Time leak localization reasoning

Given a sequence of pmfs of the hypotheses set \mathcal{H}' at different time instants t_1, t_2, \dots, t_N , the pmfs of the hypotheses set \mathcal{H}' considering all the samples can be computed by the Dempster's rule following (19) to perform a time reasoning and obtain as a result an enhanced diagnosis as

$$m_{t_1, \dots, t_N}(\mathcal{H}') = m_{t_1}(\mathcal{H}') \oplus m_{t_2}(\mathcal{H}') \oplus \dots \oplus m_{t_N}(\mathcal{H}') \quad (31)$$

As an illustrative example, we will describe the computation of probability mass function associated to leak in node 1 of the simplified WDN of Hanoi considering pressure residuals computed from measurements at initial time instant t_1 and time instant t_2 , i.e. $m_{t_1, t_2}(\mathcal{H}'_1)$. As node 1 belongs to the group associated to sensed node 2 ($\mathcal{K}_1^{(\mathcal{L})}$) and to additional cluster $\mathcal{K}_1^{(\mathcal{W})}$ depicted in Fig. 3 that correspond with hypotheses \mathcal{H}'_{32}

and \mathcal{H}'_{35} , the probability mass function of these hypothesis will also be involved in the computation of the pmf of the singleton hypothesis \mathcal{H}'_1 . Applying (31), the probability mass function associated to leak in node 1 can be computed as

$$\begin{aligned} m_{t_1, t_2}(\mathcal{H}'_1) &= \frac{1}{\Lambda} \times (m_{t_1}(\mathcal{H}'_1)m_{t_2}(\mathcal{H}'_1) \\ &+ m_{t_1}(\mathcal{H}'_1)(m_{t_2}(\mathcal{H}'_{32})) + m_{t_2}(\mathcal{H}'_{35})) \\ &+ m_{t_2}(\mathcal{H}'_1)(m_{t_1}(\mathcal{H}'_{32})) + m_{t_1}(\mathcal{H}'_{35})) \end{aligned} \quad (32)$$

To select a given node among all the candidates, the resulting hypothesis with larger probability is chosen as

$$\hat{j} = \arg \max_{j \in \{1, \dots, n_n\}} \{m_{t_1, \dots, t_N}(\mathcal{H}'_j)\} \quad (33)$$

IV. SUMMARY

The application of the proposed leak localization technique can be divided in off-line and on-line stages.

Note that the leak detection task is not considered in this work, neither the problem of data validation, which are problems that have to be addressed first.

The off-line stage can be summarized as follows:

- Use the topological information to calculate the shortest weighted pipe distance between each pair of nodes (11) to generate the \mathcal{D} matrix.
- Perform the clustering process to identify the clusters \mathcal{L} and \mathcal{W} that will be used as composed hypothesis.
- Use historical leak-free data to fit the models (4) that predict the pressure behavior without leak according to the actual network operational conditions \mathbf{c} .

The on-line stage that is triggered when a leak is detected can be summarized as follows:

- Compute the expected pressure behavior without leaks at the sensed nodes using the actual operation conditions and the fitted models (4).
- Apply the Kriging interpolation technique to generate the expected reference (no-leak) map using the predicted values.
- Apply the Kriging interpolation technique using the current pressure measurements to generate the current (leak) pressure map.
- Compute the difference between the two maps (16).
- Compute the probability mass functions of the singleton hypotheses \mathcal{H} by means of (21) and compute the pmfs of composed hypotheses \mathcal{L} and \mathcal{W} by means of (29).
- Build the expanded set \mathcal{H}' and compute the normalized pmfs of the hypotheses of \mathcal{H}' using (30).
- Apply the Dempster-Shafer reasoning (31) to combine diagnosis from different time instants.
- Select a given node among all the candidate nodes using (33).

The whole procedure is summarized in the flowcharts presented in Figures 4 and 5.

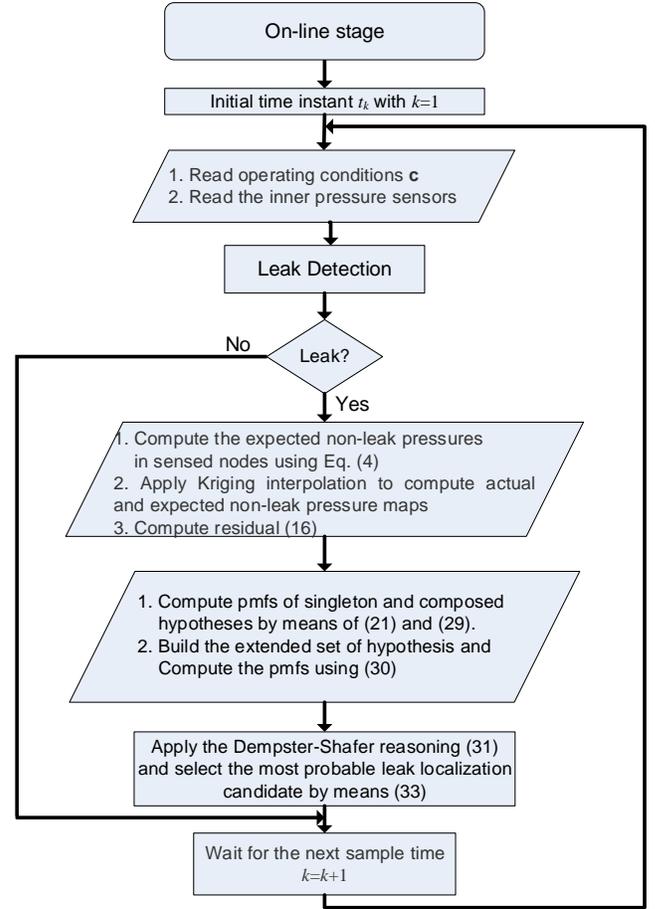


Fig. 4: Proposed Method Flowchart: On-line Stage. R1-C2

V. CASE STUDIES

The proposed leak localization technique is tested in two DMAs from the WDN of the city of Madrid (Spain), where the water company in charge has produced real case studies of water leaks by using fire hydrants. These DMAs are equipped with flow and pressure sensors at the inlet, and pressure sensors inside the network. Also, a flowmeter and a valve at the leak point are installed to measure and control the magnitude of the leak. For every DMA the leak magnitude setpoint was set as a value that the water company considered challenging in leak localization terms. In all the cases, the installed sensors have a sampling rate of two minutes and they have a resolution of 0.1 [mwc] in the case of pressure sensors and a resolution of 0.1 [l/s] in the case of flow sensors. The results of the proposed leak localization method have been compared with the ones provided by other two methods: the proposed method considering the Bayesian reasoning instead of the Dempster-Shafer one (i.e. without considering cluster hypothesis) and a well accepted method proposed in the literature [36] that is based in the use of a hydraulic model to generate leak signatures to be matched with current measurements according to the Angle metric.

R2-C8

R2-C8

R2-C6

R2-C8

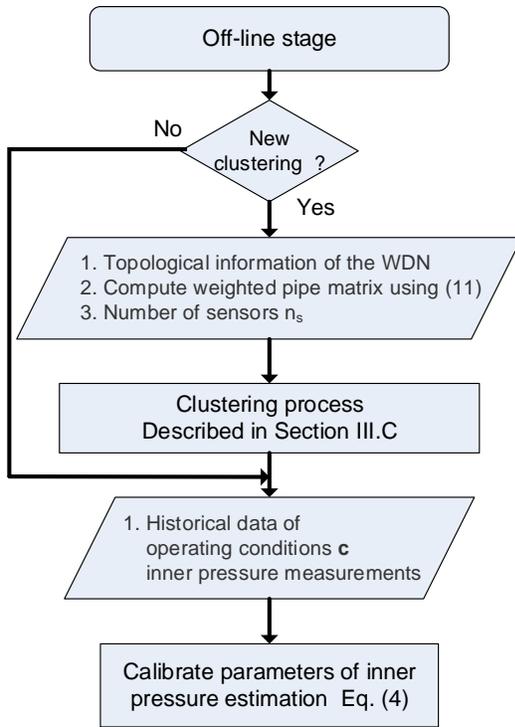


Fig. 5: Proposed Method Flowchart: Off-line Stage. R1-C2

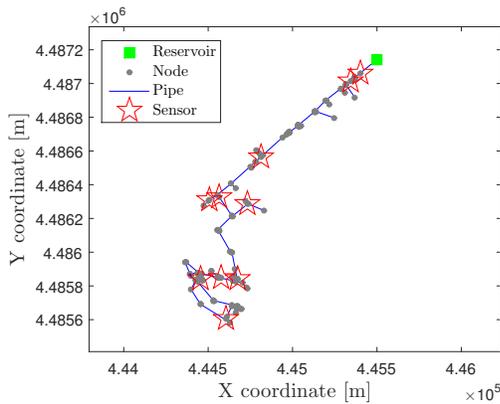


Fig. 6: Madrid DMA1 topological network.

A. Madrid DMA1

Madrid DMA 1 is a small network formed by one reservoir that feeds the network with water by elevation, 169 consumer nodes, and 173 pipes. Ten pressure sensors are placed inside at nodes with indexes 146, 145, **148**, 162, 153, 155, 156, 157, 158 and 168. The topology and the location of the sensors installed are depicted in Fig. 6. Two periods of data, with and without leak, were recorded. Measurements without leak were recorded starting the 19th of December of 2016 at 4:00 pm until the 22nd of December of 2016 at 0:58 am, whereas the leak event was recorded from the 22nd of December of 2016 at 4:00 am until the 23rd of December of 2016 at 8:58 am. The **controlled leak of setpoint magnitude** of 0.8 [l/s] approximately. The leak and the inlet measurements are depicted in Fig. 7, **in addition a detail of the leak flow is**

R2-C1

R2-C8

R2-C6

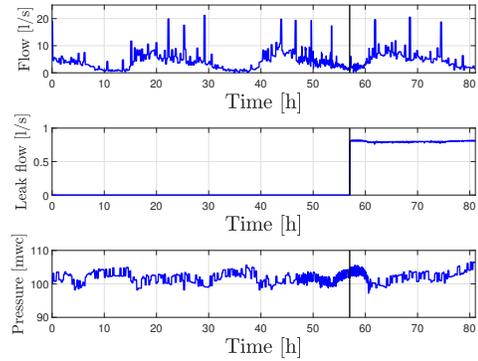


Fig. 7: Madrid DMA1 inlet flow, leak flow and pressure rate measurements without leak (before black line) and with leak (after black line).

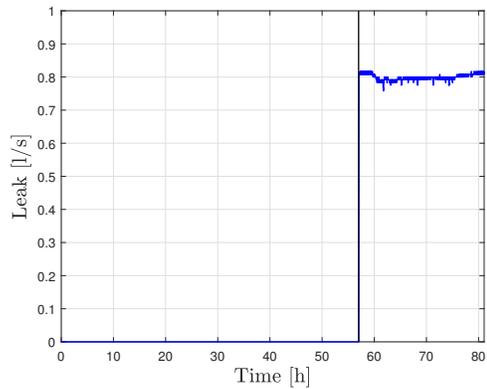


Fig. 8: Madrid DMA1 leak flow detail

provided in Fig. 8.

R2-C8

The measurements inside the network are depicted in Fig. 9 and 10 for both periods (set 1 comprises the nodes from 146 to 153, and set 2 from 155 to 168).

First, the data is downsampled to each hour by computing the average of the value measurements inside. Then, the 57 hours of data without leak are used to fit the models described in (4) for each node equipped with a sensor by means of the least squares fitting technique, which in this case leads to ten pressure models to predict the internal pressure under no leak conditions through the inlet operational conditions.

To calculate the shortest weighted pipe distance contained in the \mathcal{D} matrix, the underlying network graph \mathcal{G} is used but now considering it as undirected. Then, to fill in the matrix, the Dijkstra's Algorithm [37] is used to compute the shortest weighted pipe distance between each pair of nodes.

To estimate the pressure in the nodes inside the network which have no pressure sensors installed the Kriging interpolation technique has been implemented by means of the DACE Matlab toolbox [23]. The total number of clusters $\mathcal{K}^{(W)}$, w , is seven.

The proposed leak localization technique is applied to the Madrid DMA1 real case, where the real leak is placed in the node 165. At the end of the 28 hours that the leak lasted,

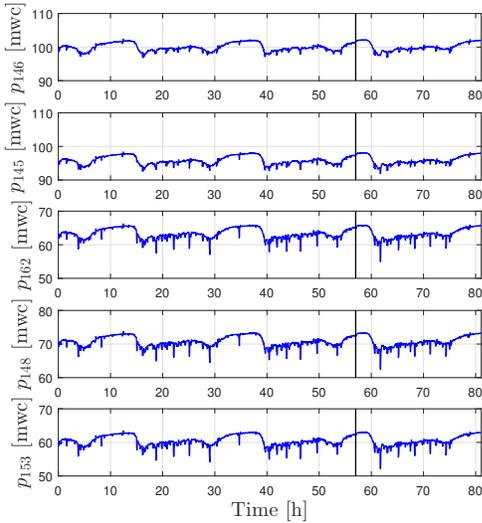


Fig. 9: Madrid DMA1 pressure measurements (set 1) without leak (before black line) and with leak (after black line).

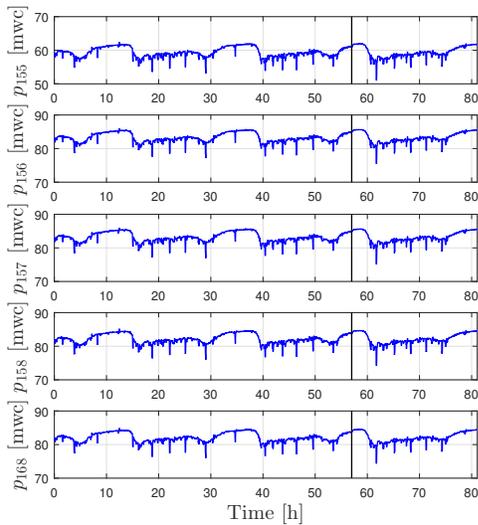


Fig. 10: Madrid DMA1 pressure measurements (set 2) without leak (before black line) and with leak (after black line).

the method provides the node 44 as the candidate to be the leaking node. Compared to the leak location, the candidate node presents a geometric linear distance of 120.8 meters, a pipe distance of 132.8 meters to the real leak location. The result is depicted in Fig. 11, where a zoomed representation of an area of the network surrounding the real leak location and including the proposed candidate node is presented.

The results obtained by the proposed method (labeled as "DS"), the proposed method without considering cluster hypothesis (labeled as "Bayesian") and the Angle method [36] (labeled as "Angle") are presented in Table I. In this table, it can be noted that the proposed method provides better results (measured in linear and pipe distance to the leaky node) than

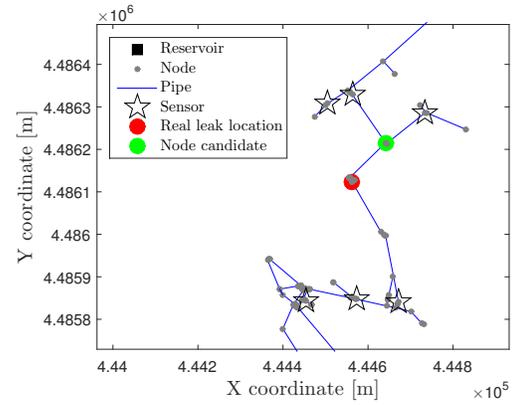


Fig. 11: Leak localization results in Madrid DMA1 real case.

TABLE I: Madrid DMA1 leak localization results comparison.

Method	Linear distance [m]	Pipe distance [m]
DS	120.8	132.8
Bayesian	276.2	474.1
Angle	303.3	363.4

the other two methods.

B. Madrid DMA2

Madrid DMA2 is a medium to large network formed by one reservoir that feeds the network by elevation, 1031 consumer nodes and 1100 pipes. Ten pressure sensors are placed inside at nodes with indexes 886, 877, 864, 971, 880, 898, 917, 933, 943 and 948. The topology and the location of the sensors installed is depicted in Fig. 12.

Measurements have been taken starting the 30th of January of 2017 at 2:00 am until the 31st of January of 2017 at 11:28 pm under no leak conditions and data with a **controlled leak of setpoint magnitude** of 2 [l/s] were taken from the 1st of February of 2017 at 6:00 pm until the 2nd of February of 2017 at 11:58 am. The inlet measurements and the leak flow rate are depicted in Fig. 13, **in addition a detail of the leak flow is provided in Fig.14.**

The pressure measurements inside the network are depicted in Fig. 15 and 16.

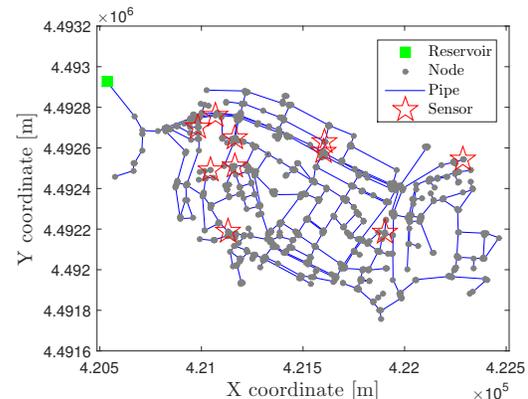


Fig. 12: Madrid DMA2 topological network.

R2-C8

R2-C6

R2-C8

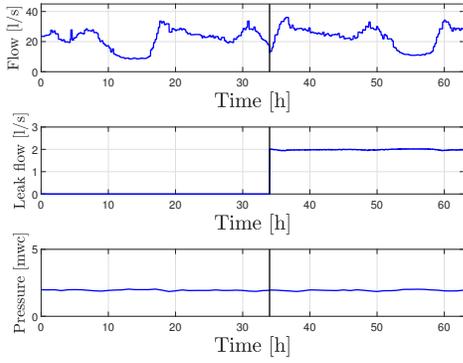


Fig. 13: Madrid DMA2 inlet flow, leak flow and pressure rate measurements without leak (before black line) and with leak (after black line).

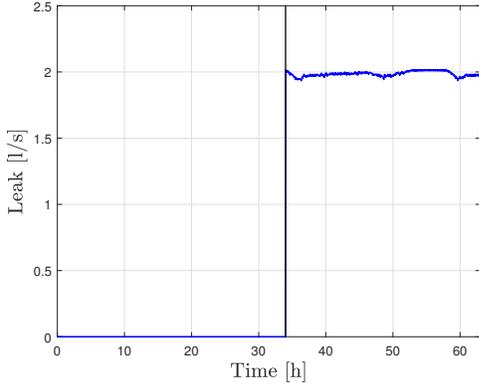


Fig. 14: Madrid DMA2 leak flow detail

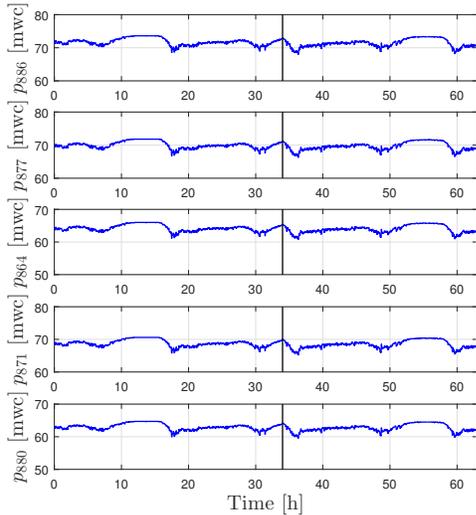


Fig. 15: Madrid DMA2 pressure measurements (set 1) without leak (before black line) and with leak (after black line).

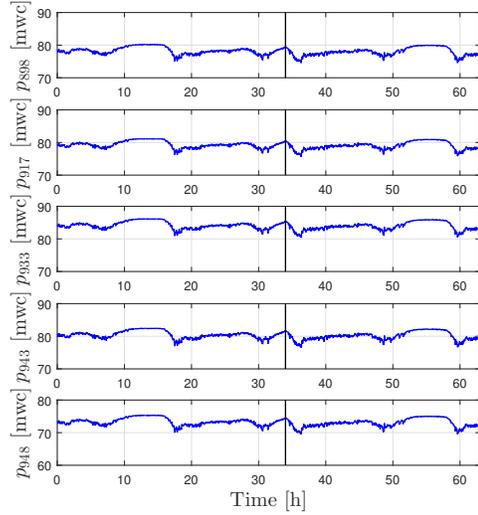


Fig. 16: Madrid DMA2 pressure measurements (set 2) without leak (before black line) and with leak (after black line).

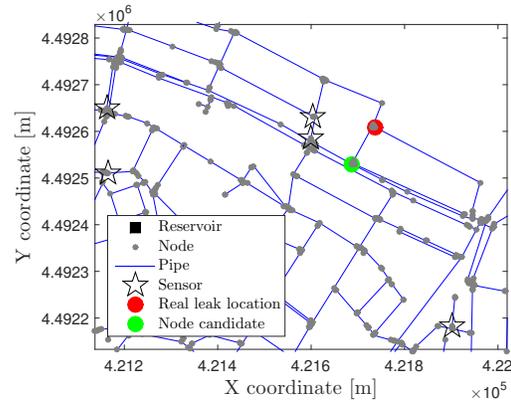


Fig. 17: Leak localization results in Madrid DMA2 real case.

As in the previous case study, the hourly average is computed and the first 54 hours are used to fit the ten pressure models. In this case, 16 clusters $\mathcal{K}^{(W)}$ have been generated. The diagnosis through the 18 hours of leaky data points to the node 5 as candidate whereas the leak is actually in the node 882. In this case, the geometric linear distance is 93.4 meters while the pipe distance is 98.8 meters to the real leak location. The result is depicted in Fig. 17.

The results obtained by the proposed method and the other two methods are presented in Table II. In this table, it can be noted that the proposed method provides better results (measured in linear and pipe distance to the leaky node) than the other two methods.

TABLE II: Madrid DMA2 leak localization results comparison.

Method	Linear distance [m]	Pipe distance [m]
DS	93.4	98.8
Bayesian	506.9	758.8
Angle	594.6	679.2

C. Results Discussion

From the leak localization results obtained in the two real DMAs of Madrid summarized in Tables I and II, we can state that the proposed method outperforms the Angle method presented in [36]. The proposed method is based on the fact that when a leak occurs, there is a loss of pressure across the network, specially in the places where the additional flow produced by the leak passes through the network until arrives to the leak location. This loss of pressure is observed by means the differences (residuals) between estimated pressures (considering non-leak scenario) and actual pressures measured in inner sensors. The pressures of non-measured inner nodes are estimated by the **Kriging** method, therefore there is an additional error in the computation of the residuals of non-measured nodes that increases with the distance of these nodes from sensed nodes. This additional error can produce a deviation in the leak localization to nodes that are closer to sensed nodes. The additional information of clusters by means of the Dempster-Shafer reasoning improves significantly the accuracy in the leak localization. Finally, in absolute terms, the accuracy obtained by the proposed method (around 100[m] of error in the leak localization) is considered reasonable for the water companies because it is possible to send operators to search and exactly pinpoint the leak through acoustic sensors in this bounded area of the DMA, since this technique is too time-consuming to be applied across all the DMA.

VI. CONCLUSIONS

In this paper, a new methodology for a data-driven leak localization problem in water distribution networks has been presented and tested. The proposed approach is based on the hydraulic characteristics that appear when a leak in a particular place of the network occurs. A data-driven model has been used for the prediction of internal pressures without leak based on past measurements. The Kriging interpolation has been proposed to overcome the problem of limited sensor measurements in the network. The application of the Dempster-Shafer reasoning is proposed to enhance the leak localization performance by taking into account the time evolution of the residuals. The method has been successfully tested in two DMAs of real water network providing better leak localization results compared to two other alternative methods. One of them considers the same approach but using a time reasoning based on the Bayes rule while the other is based in the use a hydraulic model to generate leak signatures to be matched with current measurements according the Angle metric. Both methods are outperformed by the proposed approach.

The performance of the proposed methodology has been successfully tested with real measurement data from three DMAs of the Madrid WDN.

As a future work, **the proposed leak localization method will be tested in other DMAs with different topologies and sizes.** In addition, it will be interesting to investigate the application of the proposed data-driven approach to monitor water pollution. Also, the case of multiple leaks occurring at the same time will be considered.

ACKNOWLEDGMENT

This work has been funded by the Ministerio de Economía, Industria y Competitividad (MEICOMP) of the Spanish Government through the project DEOCS (ref. DPI2016-76493) and by the Catalan Agency for Management of University and Research Grants (AGAUR), the European Social Fund (ESF) and the Secretary of University and Research of the Department of Companies and Knowledge of the Government of Catalonia through the grant FI-DGR 2015 (ref. 2015 FI B 00591). J. Blesa acknowledges the support from the Serra Hünter program. The third author would like to acknowledge the grant from the Innovation Fund Denmark (IFD) under File No. 6155-00002B.

REFERENCES

- [1] R. Puust, Z. Kapelan, D. A. Savić, and T. Koppel, "A review of methods for leakage management in pipe networks," *Urban Water Journal*, vol. 7, no. 1, pp. 25–45, 2010.
- [2] D. G. Eliades and M. M. Polycarpou, "A fault diagnosis and security framework for water systems," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 6, pp. 1254–1265, Nov 2010.
- [3] Z. Y. Wu and P. Sage, "Water loss detection via genetic algorithm optimization-based model calibration," in *Systems Analysis Symposium*. ASCE, 2006, pp. 1–11.
- [4] J. Yang, Y. Wen, and P. Li, "Leak location using blind system identification in water distribution pipeline," *Journal of Sound and Vibration*, no. 310, pp. 134–148, 2008.
- [5] H. Fuchs and R. Riehle, "Ten years of experience with leak detection by acoustic signal analysis," *Applied Acoustics*, no. 33, pp. 1–19, 1991.
- [6] J. Muggleton, M. Brennan, and R. Pinnington, "Wavenumber prediction of waves in buried pipes for water leak detection," *Journal of Sound and Vibration*, no. 249, pp. 939–954, 2002.
- [7] J. Mashford, D. de Silva, D. Marney, and S. Burn, "An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine," in *Third International Conference on Network and System Security*, 2009, pp. 534–539.
- [8] A. Soldevila, J. Blesa, S. Tornil-Sin, E. Duviella, R. M. Fernandez-Canti, and V. Puig, "Leak localization in water distribution networks using a mixed model-based/data-driven approach," *Control Engineering Practice*, vol. 55, pp. 162–173, 2016.
- [9] A. Soldevila, R. M. Fernandez-Canti, J. Blesa, S. Tornil-Sin, and V. Puig, "Leak localization in water distribution networks using Bayesian classifiers," *Journal of Process Control*, vol. 55, pp. 1–9, 2017.
- [10] D. Wachla, P. Przystalka, and W. Moczulski, "A method of leakage location in water distribution networks using artificial neuro-fuzzy system," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 1216 – 1223, 2015.
- [11] 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.
- [12] P. Cugueró-Escofet, J. Blesa, R. Pérez, M. A. Cugueró-Escofet, and G. Sanz, "Assessment of a leak localization algorithm in water networks under demand uncertainty," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 226 – 231, 2015.
- [13] J. Blesa and R. Pérez, "Modelling uncertainty for leak localization in water networks," *IFAC-PapersOnLine*, vol. 51, no. 24, pp. 730 – 735, 2018, 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2018.
- [14] M. V. Casillas, L. Garza-Castañón, and V. Puig, "Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks," in *8th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*. Elsevier, 2012, pp. 570–575.
- [15] J. Gross and J. Yellen, *Graph Theory and Its Applications*. New York, USA: Chapman and Hall/CRC, 2005.
- [16] P. K. Swamee and A. K. Sharma, *Design of Water Supply Pipe Networks*, 1st ed. Wiley, 2008.
- [17] T. N. Jensen, C. S. Kallesøe, J. D. Bendtsen, and R. Wisniewski, "Plug-and-play commissionable models for water supply networks with multiple inlets," European Control Conference, 2018.

R2-C1

R1-C1

- [18] T. Jensen, C. Kallesøe, J. Bendtsen, and R. Wisniewski, "Iterative learning pressure control in water distribution networks," in *2018 IEEE Conference on Control Technology and Applications (CCTA), Copenhagen, Denmark, 2018*.
- [19] L. Ljung, *System identification -theory for the user*, 2nd ed. Prentice-Hall, 1999.
- [20] J. P. Kleijnen, "Regression and kriging metamodels with their experimental designs in simulation: A review," *European Journal of Operational Research*, vol. 256, no. 1, pp. 1 – 16, 2017.
- [21] A. Soldevila, J. Blesa, R. M. Fernandez-Canti, S. Tornil-Sin, and V. Puig, "Data-driven approach for leak localization in water distribution networks using pressure sensors and spatial interpolation," *Water*, vol. 11, no. 7, 2019. [Online]. Available: <https://www.mdpi.com/2073-4441/11/7/11500>
- [22] C. S. Kallesøe, T. N. Jensen, and R. Wisniewski, "Adaptive reference control for pressure management in water networks," *EUCA, 14th European Control Conference, 2015*.
- [23] H. B. Nielsen, S. N. Lophaven, and J. Søndergaard, "DACE - a matlab kriging toolbox," Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2002. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?1460>
- [24] M. V. Casillas, L. E. Garza-Castañón, and V. Puig, "Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks: Application to the Barcelona network," *European Control Conference (ECC), 2013, Zurich, Switzerland*, no. 1, 2013.
- [25] T. N. Jensen and C. S. Kallesøe, "Application of a Novel Leakage Detection Framework for Municipal Water Supply on AAU Water Supply Lab," in *3rd Conference on Control and Fault-Tolerant Systems (SysTol), 2016*. IEEE, 2016, pp. 428–433.
- [26] M. Romano, K. Woodward, and Z. Kapelan, "Statistical Process Control Based System for Approximate Location of Pipe Bursts and Leaks in Water Distribution Systems," *Procedia Engineering*, vol. 186, pp. 236–243, 2017.
- [27] C. S. Kallesøe and T. N. Jensen, "On the relation between leakage location and network pressures," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2018, pp. 571–576.
- [28] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.
- [29] —, "Dempster's rule of combination," *International Journal of Approximate Reasoning*, vol. 79, pp. 26–40, 2016.
- [30] R. Gonzalez, F. Qi, and B. Huang, *Process Control System Fault Diagnosis*. Chichester, West Sussex, UK: John Wiley and Sons, Ltd, 2016.
- [31] J. Bicik, Z. Kapelan, C. Makropoulos, and D. A. Savić, "Pipe burst diagnostics using evidence theory," *Journal of Hydroinformatics*, vol. 13, no. 4, pp. 596–608, 2011.
- [32] R. Sadiq and M. J. Rodriguez, "Interpreting drinking water quality in the distribution system using dempster-shafer theory of evidence," *Chemosphere*, vol. 59, no. 2, pp. 177–188, 2005.
- [33] O. Fujiwara and D. B. Khang, "A two-phase decomposition method for optimal design of looped water distribution networks," *Water Resources Research*, vol. 26, no. 4, pp. 539–549, 1990. [Online]. Available: <http://dx.doi.org/10.1029/WR026i004p00539>
- [34] A. Dempster, "A generalization of bayesian inference," in *Classic works of the dempster-shafer theory of belief functions*. Springer, 2008, pp. 73–104.
- [35] J. Blesa, F. Nejari, and R. Sarrate, "Robust sensor placement for leak location: analysis and design," *Journal of Hydroinformatics*, vol. 18, no. 1, pp. 136–148, 2016.
- [36] M. V. Casillas, L. Garza-Castañón, and V. Puig, "Model-based leak detection and location in water distribution networks considering an extended-horizon analysis of pressure sensitivities," *Journal of Hydroinformatics*, vol. 16, no. 3, pp. 649–670, 2013.
- [37] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.