Aalborg Universitet

Leakage localization in water distribution using data-driven models and sensitivity analysis

Jensen, Tom Nørgaard; Puig, Vicenc; Romera, Juli; Kallesøe, Carsten Skovmose; Wisniewski, Rafal; Bendtsen, Jan Dimon

Published in:
IFAC-PapersOnLine

DOI (link to publication from Publisher):
10.1016/j.ifacol.2018.09.657

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 51-24 (2018) 736–741

Leakage localization in water distribution using data-driven models and sensitivity analysis

Tom Nørgaard Jensen * **Vicenç Puig** ** **Juli Romera** **
Carsten Skovmose Kallesøe *,*** **Rafał Wisniewski** *
Jan Dimon Bendtsen *

* *Aalborg University, Fredrik Bajers Vej 7c, DK-9220 Aalborg, Denmark (e-mail: {tnj,raf,dimon}@es.aau.dk).*
** *Department of Automatic Control, Universitat Politècnica de Catalunya (UPC), Barcelona 08028, Spain (e-mail: {vicenc.puig,juli.romera}@upc.edu).*
*** *Grundfos, Poul Due Jensens Vej 7, DK-8850 Bjerringbro, Denmark (e-mail: ckallesoe@grundfos.com).*

Abstract: Water scarcity is becoming an increasing problem worldwide, and an issue compounding the problem is water leakage in the piping networks delivering potable/consumable water to end-users (Sensus, 2012). In this paper, we consider the problem of isolating leakages in water supply networks using reduced network models. Using a reduced order model of the network, the expected behaviour of the network can be estimated and then compared with actual measurements obtained from the network. The result of this comparison is a set of residuals which are used to isolate a leakage to a network node. The localization is based on a sensitivity matrix which captures the residuals' sensitivities to leakages. As the reduced order model is adaptive based on measurements from the network, the reduced order model is plug-and-play commissionable. The calculation of the sensitivity matrix is based on an EPANET model of the network and is performed off-line.

Keywords: Water Supply Systems, Large-Scale Hydraulic Networks, Leakage Localization.

1. INTRODUCTION

Most water utilities use large efforts on leakage detection and leakage localization. A task that becomes increasingly important considering the increasing water scarcity worldwide. The methods commercially available in the sector span from detection based on change in the supply flow typically during the night time, as e.g. proposed in PWiK (2015), over acoustic equipment (Gutermann, 2015), to advanced signal analysis (Scolnicov and Horowitz, 2010). The drawback of many of these methods is the labour needed for either performing the leakage search (acoustic) or for setting up and maintaining the detection systems (signal analysis), or the need for expensive equipment (acoustic).
In this paper, we investigate the performance of a set of residuals, which we proposed in Jensen et al. (2018b) and Jensen and Kallesøe (2016) for the purpose of isolating leakages in a water distribution network. The residuals are generated using an adaptive model whose parameters are obtained from historical data.
While most of the existing literature considers model based methods for detection and localization of leakages in water supply systems, data driven leakage localization is also an option. Mashford et al. (2009) describes a data driven method using support vector machines trained by simulation data. The method in Bicik et al. (2011) can be seen as a hybrid approach where the output of a deterministic hydraulic model is combined with a statistic model of pipe breakage and statistics from consumer contracts. Subsequently, evidence theory is used to produce a spatial distribution of belief and plausibility of failure among the network components.
In model based leakage localization, output prediction together with optimisation in a weighted least squares setting was formulated in Andersen and Powell (2000) where state estimation is performed by solving an optimisation problem. Among the estimated states it is then possible to have the nodal demands in the network including leakages. This method was extended in Kang and Lansey (2009) to include stochastic variables by introduction of probabilistic estimators such as a Kalman filter.
Geometric leakage localization is considered in Casillas et al. (2015) where vectors of residuals are matched against pre-calculated signature vectors to isolate leakages. Leakage localization using fault sensitivity matrices has been explored in Pérez et al. (2011) where the sensitivity of pressure measurements to predefined fault scenarios is employed. The sensitivity matrix is pre-calculated using simulations.
The above mentioned methods require either manual labour or preliminary knowledge in the form of a network model. The approach considered in this paper is plug & play, which limits the installation, commissioning, and operation costs substantially. Our approach utilises pressure

2405-8963 © 2018, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.09.657

measurements at specific nodes together with a reduced network model, whose parameters are identified online, to generate residuals for leakage localization. A detailed derivation of the reduced order model structure is given in Jensen et al. (2018a).

A dual problem to leakage localization is the optimal placement of sensors which is important for any leakage localization method. This topic is also the focus of many works. These include Bonada et al. (2014) where sensor placement is formulated as an optimisation problem which is then solved using genetic algorithms, Bort et al. (2014) where sensitivity matrices constructed from simulations are used for optimal placement and Steffelbauer et al. (2014) where output sensitivity to process noise such as demand fluctuations is used to determine optimal sensor placement. Here, we will assume that the sensor placement problem has been solved prior to the residual generation.

The contribution of the current exposition consists of evaluating the performance of the residuals calculated using a reduced order model presented in a previous paper in a leakage localization scheme on data obtained from a real water distribution network.

The structure of the paper is as follows. Section 2 gives the general network model together with the reduced order model. In Section 3, we then propose a set of residuals and the method used for detection and localization of a leakage. Section 4 gives numerical results obtained using an EPANET model of a real distribution network. Lastly, the paper ends with some concluding remarks in Section 5.

2. PRELIMINARIES

Water distribution networks can be described by a directed and connected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. The elements of the set $\mathcal{V} = \{v_1, \dots, v_n\}$ are denoted vertices and represent pipe connections with possible end-user water consumption. The ith vertex has three values associated to it: p_i, d_i and h_i. The values p_i [mwc] and d_i [m^3/h] are variables describing respectively the absolute pressure and demand at the vertex. The value h_i [m] is the geodesic level at the vertex.

The elements of the set $\mathcal{E} = \{e_1, \dots, e_m\}$ are denoted edges and represent the pipes. The pressure drop due to hydraulic resistance of the edge j is denoted by f_j, and this pressure drop is a function of the flow q_j [m^3/h] through the pipe (edge) j. In the following exposition, we will follow Jensen et al. (2018a) and use the following assumption.

Assumption 1. Each $f_i : \mathbb{R} \to \mathbb{R}$ has the following structure

$$f_i(q_i) = \rho_i |q_i| q_i, \tag{1}$$

where $\rho_i > 0$ is a parameter of the pipe and $|\cdot|$ denotes absolute value.

Associated to the graph $\mathcal{G}$, we have the incidence matrix H, which we recall below

$$H_{i,j} = \begin{cases} -1 \text{ , if the } j^{th} \text{ edge is entering } i^{th} \text{ vertex.} \\ 0 \text{ , if the } j^{th} \text{ edge is not connected to} \\ \qquad \text{the } i^{th} \text{ vertex.} \\ 1 \text{ , if the } j^{th} \text{ edge is leaving } i^{th} \text{ vertex.} \end{cases}$$

The incidence matrix has dimension $n \times m$, where m is the number of edges and n is the number of vertices in the graph. Note that $m \geq (n-1)$ for connected graphs.

The network must fulfil Kirchhoff's vertex law which corresponds to conservation of mass in each vertex, and is described by

$$Hq = d, \tag{2}$$

where $q \in \mathbb{R}^m$ is the vector of flows in edges and $d \in \mathbb{R}^n$ is the vector of nodal demands, with $d_i > 0$ when demand flow is into vertex i. Because of mass conservation in the network, there can only be $n-1$ independent nodal demands which means that $\sum_{i=1}^{n} d_i = 0$.

Let p be the vector of absolute pressures at the vertices and Δp be the vector of differential pressures across the edges, then the "Ohm law" for water networks gives

$$\Delta p = H^T p = f(q) - H^T h, \tag{3}$$

where $p \in \mathbb{R}^n$, $f : \mathbb{R}^m \to \mathbb{R}^m$, $f(q) = (f_1(q_1), \dots, f_m(q_m))$ and f_i strictly increasing. The function f_i describes the flow dependent pressure drop due to the hydraulic resistance. The term $H^T h$ is the pressure drop across the components due to difference in geodesic level between the ends of the components with $h \in \mathbb{R}^n$ the vector of geodesic levels at each vertex expressed in units of potential (pressure).

Next, we partition the n vertices of the underlying network graph into two sets, $\mathcal{V} = \bar{\mathcal{V}} \cup \hat{\mathcal{V}}$. Here, $\hat{\mathcal{V}} = \{\hat{v}_1, \dots, \hat{v}_c\}$ where $c \geq 1$ represents vertices in the graph corresponding to inlet vertices in the distribution network. The set $\bar{\mathcal{V}} = \{\bar{v}_1, \dots, \bar{v}_{n-c}\}$ represents the remaining vertices in the graph.

The m edges of the network graph are likewise partitioned into two sets, $\mathcal{E} = \mathcal{E}_\mathcal{T} \cup \mathcal{E}_\mathcal{C}$. Here, $\mathcal{E}_\mathcal{T} = \{e_{\mathcal{T},1}, \dots, e_{\mathcal{T},n-c}\}$, $\mathcal{E}_\mathcal{C} = \{e_{\mathcal{C},1}, \dots, e_{\mathcal{C},m-n+c}\}$ and the partitioning is chosen such that the sub-matrix, say $\bar{H}_\mathcal{T}$, which maps edges in $\mathcal{E}_\mathcal{T}$ to vertices in $\bar{\mathcal{V}}$ is invertible. Note, that such a partitioning is always possible since any $(n-1)$-by-m sub-matrix of the incidence matrix of a connected graph has full rank $(n-1)$, (Deo, 1974). It is also worth noting that for $c = 1$, the graph $\mathcal{T} = \{\mathcal{V}, E_\mathcal{T}\}$ is a spanning tree of the underlying network graph.

With the chosen partitioning, we can rewrite (2) and (3) as follows

$$\begin{aligned} \bar{d} &= \bar{H}_\mathcal{T} q_\mathcal{T} + \bar{H}_\mathcal{C} q_\mathcal{C} \\ \hat{d} &= \hat{H}_\mathcal{T} q_\mathcal{T} + \hat{H}_\mathcal{C} q_\mathcal{C} \\ f_\mathcal{T}(q_\mathcal{T}) &= \bar{H}_\mathcal{T}^T (\bar{p} + \bar{h}) + \hat{H}_\mathcal{T}^T (\hat{p} + \hat{h}) \\ f_\mathcal{C}(q_\mathcal{C}) &= \bar{H}_\mathcal{C}^T (\bar{p} + \bar{h}) + \hat{H}_\mathcal{C}^T (\hat{p} + \hat{h}), \end{aligned} \tag{4}$$

where $\bar{H}_\mathcal{T}$ ($\hat{H}_\mathcal{T}$) denotes the sub-matrix of H associated with edges $\mathcal{E}_\mathcal{T}$ and vertices $\bar{\mathcal{V}}$ ($\hat{\mathcal{V}}$); $\bar{H}_\mathcal{C}$ ($\hat{H}_\mathcal{C}$) denotes the sub-matrix of H associated with edges $\mathcal{E}_\mathcal{C}$ and vertices $\bar{\mathcal{V}}$ ($\hat{\mathcal{V}}$). To state the reduced order model, we will also need the following definitions of vectors $a_\mathcal{C}$ and ν

$$q_\mathcal{C} = a_\mathcal{C} \sigma \ , \ \bar{d} = -\nu \sigma, \tag{5}$$

where $\sigma = \sum_i \hat{d}_i > 0$ denotes the total inlet to the network.

We associate the following three assumptions to the network partitioning.

Assumption 2. The vectors $\hat{p}$ of inlet pressures and $\hat{d}$ of inlet flows are measured. Furthermore, there exists a vector $y \in \mathbb{R}^o$ where $\{y_1 \dots, y_o\} \subset \{\bar{p}_1, \dots, \bar{p}_{n-c}\}$ of measured pressures at the remaining vertices.

Assumption 3. The total head is the same at all inlets at any time, that is

$$\hat{p}(t) + \hat{h} = \kappa(t)\mathbb{1}, \tag{6}$$

for some $\kappa(t) \in \mathbb{R}$ and where $\mathbb{1}$ denote the vector consisting of ones.

Assumption 4. All non-inlet vertices with non-zero demand have the same consumption profile. That is, the vector ν is constant.

The assumptions listed above lead to the following proposition, proposed in Jensen et al. (2018a).

Proposition 1. Applying Assumptions 1, 3 and 4, the following expression for the pressure at the ith non-inlet vertex applies

$$\bar{p}_i(t) = \alpha_i \sigma^2(t) + \kappa(t) + \gamma_i, \tag{7}$$

with

$$\begin{aligned} \alpha_i &= (\bar{H}_{\mathcal{T}}^{-T})_i f_{\mathcal{T}}(-\bar{H}_{\mathcal{T}}^{-1}\bar{H}_{\mathcal{C}} a_{\mathcal{C}} + \bar{H}_{\mathcal{T}}^{-1}\nu) \\ \gamma_i &= -\bar{h}_i, \end{aligned} \tag{8}$$

where $(\bar{H}_{\mathcal{T}}^{-T})_i$ denotes the ith row of $\bar{H}_{\mathcal{T}}^{-T}$.

Remark 1. Assumption 3 might seem hypothetical at first glance, however, as we will show in Section 4 it is fulfilled at least in some real-life cases. Furthermore, one could design a control structure for the inlet pressures such that Assumption 3 is explicitly fulfilled.

Remark 2. In case Assumption 4 is not fulfilled, α_i will be a time-varying parameter. Typically, the demand profiles in a water distribution network exhibit a periodic behaviour which means that $\nu(t+T) = \nu(t)$ and $\sigma(t+T) = \sigma(t)$ where T denotes the length of the period. Likewise, the parameter $\alpha_i(t)$ will exhibit a periodic behaviour in this case.

We will use the model (7) to predict the expected pressure at a subset of vertices in $\bar{\mathcal{V}}$ where the pressure is measured. To obtain such a prediction, the model parameters need to be known. Since the model (7) of $\bar{p}_i(t)$ is linear in the parameters α_i and γ_i, standard parameter identification methods (Ljung, 1999) can be used to identify these parameters based on measurements of $\bar{p}_i(t)$, $\sigma(t)$ and $\hat{p}(t)$, which are available due to Assumption 2. Subsequently, having identified the parameters, the model (7) can be used to predict the expected pressure $\bar{p}_i(t)$ at a measurement vertex, since the signals $\sigma(t)$ and $\hat{p}(t)$ are known.

3. LEAKAGE DETECTION AND LOCALIZATION METHODOLOGY

With a prediction model in place, we wish to use the model to generate a set of residuals for use in leakage localization in the distribution network. In this work, we will assume the leakages are associated to the vertices in $\bar{\mathcal{V}}$ of the network. In reality, leakages occur in the pipes (edges). However, in practice it is sufficient to assume a simplified situation where leakages occur at the existing vertices of the network.

Assumption 5. A leakage affects the vertices in $\bar{\mathcal{V}}$, and a leakage in the ith vertex is modelled by an additional consumption in the vertex with a magnitude of l_i flow units.

A leakage in the network will therefore change the vector ν which describes the distribution of the demand across the non-inlet vertices in the network. The change in ν will cause a change in the parameter α_i as seen in (8), which in turn affect the model (7). Therefore, the pressure estimated using the model will not correspond to the pressure measured in the measurement vertex and this discrepancy will be captured in the residual.

Let the expected pressure at a measurement vertex be denoted $\bar{p}_i^{(e)}(t)$, then this is given by

$$\bar{p}_i^{(e)}(t) = \alpha_i^{(e)} \sigma^2(t) + \mathbb{1}\kappa(t) + \gamma_i, \tag{9}$$

where $\alpha_i^{(e)}$ is the is the expected value of the parameter α_i obtained from historical measurements. Then, we propose the following set of residuals for leakage localization

$$r_i(t) \equiv \bar{p}_i(t) - \bar{p}_i^{(e)}(t) = (\alpha_i - \alpha_i^{(e)})\sigma^2(t) = \Delta\alpha_i \sigma^2(t), \tag{10}$$

where $\bar{p}_i$ is the measured pressure at the ith measurement vertex and α_i is the actual parameter of the system, see also Jensen et al. (2018b).

3.1 Leakage detection

To detect leakages using the residuals, we will rely on the hypothesis test stated in Jensen et al. (2018b), which is based on the following assumption.

Assumption 6. The residual vector $r(t) = (r_1(t), \ldots, r_o(t))$ is Gaussian distributed with zero mean and co-variance matrix $\Sigma(t)$, and a leakage only affects the mean value of the residual.

The assumption that the leakage only affects the mean value is generally not satisfied as seen in the last equality in (10) since a non-zero $\Delta\alpha_i$ will overlay the residual with the demand curve. However, for small $\Delta\alpha_i$ it is a good approximation.

Assumption 6 motivates the following hypothesis test

$$\mathbb{H}_0 : \; r(t) \in \mathcal{N}(0_o, \Sigma(t)) \; , \; \mathbb{H}_1 : \; r(t) \in \mathcal{N}(\mu_1, \Sigma(t)). \tag{11}$$

The test random variable $T_d(t)$ for detection is given by

$$T_d(t) = r^T(t)\Sigma^{-1}(t)r(t). \tag{12}$$

It is known that $T_d(t)$ has the χ^2-distribution with o-degrees of freedom under the hypothesis $\mathbb{H}_0$, (Anderson, 2003). This means that for a given allowed false alarm-rate, say R_f, the threshold J_d for alarm can be set using the χ^2-table with o-degrees of freedom by choosing $J_d = \chi^2_{R_f}$. Thereby, the probability of generating a false alarm becomes $P\left(T_d(t) > J_d \mid \mathbb{H}_0\right) = R_f$.

3.2 Leakage localization

For the purpose of localization of the leakage to a particular node, we will here rely on the method based on the so-called sensitivity matrix, Casillas et al. (2013) and Pérez et al. (2014). The method is well described in these works and the purpose here is to test the quality of the residuals generated based on the reduced order model (7). The sensitivity matrix S is defined as the change in measured pressures with respect to an occurring leakage, that is

$$S = [s_{ij}] = \begin{bmatrix} \dfrac{\partial y_1}{\partial l_1} & \cdots & \dfrac{\partial y_1}{\partial l_{n-c}} \\ \vdots & & \vdots \\ \dfrac{\partial y_o}{\partial l_1} & \cdots & \dfrac{\partial y_o}{\partial l_{n-c}} \end{bmatrix}. \tag{13}$$

For practical purposes, the entries in the sensitivity matrix S will be approximated using an EPANET model of the network, using the following calculation

$$s_{ij}(k) \approx \frac{y_i^{(l_j)}(k) - y_i^{(0)}(k)}{l_j}, \qquad (14)$$

where $y_i^{(l_j)}(k)$ is the estimated output from EPANET during the kth sample interval with leak l_j; $y_i^{(0)}(k)$ is the estimated output from EPANET during the kth sample interval but with no leak.

In principle, $S(k)$ should be determined for every sample interval. However, due to the following assumption, the number of different sensitivity matrices needed is limited.

Assumption 7. The system is periodic with period $T = \tau T_s$ where T_s is the sample time.

Due to Assumption 7, the number of sensitivity matrices which needs to be calculated is τ since we have $s_{ij}(k + \tau T_s) = s_{ij}(k)$.

To isolate the vertex with the leakage, we use the columns of $S(k)$ as signature vectors and find the signature which has the maximal normalised projection of the residual vector. That is, given the residual vector $r(k)$ and the jth column $s_{*j}(k)$ of $S(k)$, first we calculate the normalised projection

$$\psi_k(j) = \frac{\langle r(k), s_{*j}(k)\rangle}{|r(k)||s_{*j}(k)|}, \qquad (15)$$

where $j \in \mathcal{I} = \{1, 2, \ldots, n - c\}$ and $\langle \cdot, \cdot \rangle$ denotes the inner product.

Having calculated the normalised projections, the set of leak vertex candidates, say $\bar{\mathcal{V}}^l(k) \subset \bar{\mathcal{V}}$, is then given as

$$\bar{\mathcal{V}}^l(k) = \{\bar{v}_i \in \bar{\mathcal{V}} \mid i = \arg\max_{j \in \mathcal{I}} \psi_k(j)\}, \qquad (16)$$

given that the EPANET model is an accurate representation of the real network.

In practice, the set $\bar{\mathcal{V}}^l(k)$ is extended to also include vertices with a normalised projection close to the maximal, that is

$$\bar{\mathcal{V}}^l(k) = \{\bar{v}_i \in \bar{\mathcal{V}} \mid \psi_k(i) \in [\max_{j \in \mathcal{I}} \psi_k(j) - \varepsilon, \max_{j \in \mathcal{I}} \psi_k(j)]\}, \qquad (17)$$

where $1 >> \varepsilon > 0$. This will increase the probability that the leaking node is in the set $\bar{\mathcal{V}}^l(k)$ if the EPANET model is less accurate. The value of ε is a trade-off between the size of the set $\bar{\mathcal{V}}^l(k)$ of potential candidates and the chance that the vertex with the leak is contained in the set. A formal procedure for choosing ε is still an open question and should be based on the uncertainty in the EPANET model.

Given a set of candidates for a number of samples, say $(\bar{\mathcal{V}}^l(1), \bar{\mathcal{V}}^l(2), \ldots, \bar{\mathcal{V}}^l(w))$, one subsequently needs to combine this information to choose a final set of candidate nodes. One possibility is to choose the set $\bar{\mathcal{V}}^l(k)$ for $k \in \{1, 2, \ldots, w\}$ with lowest cardinality.

4. EXPERIMENTAL RESULTS

In this section, we present results obtained using data from an actual water distribution network. The data are collected before and during a controlled test where a leakage is introduced in the network by the opening of a fire hydrant.

The network in question is illustrated in Fig. 1. The network consists of 399 vertices and 429 edges ($n = 399$ and $m = 429$). This particular network has two inlets

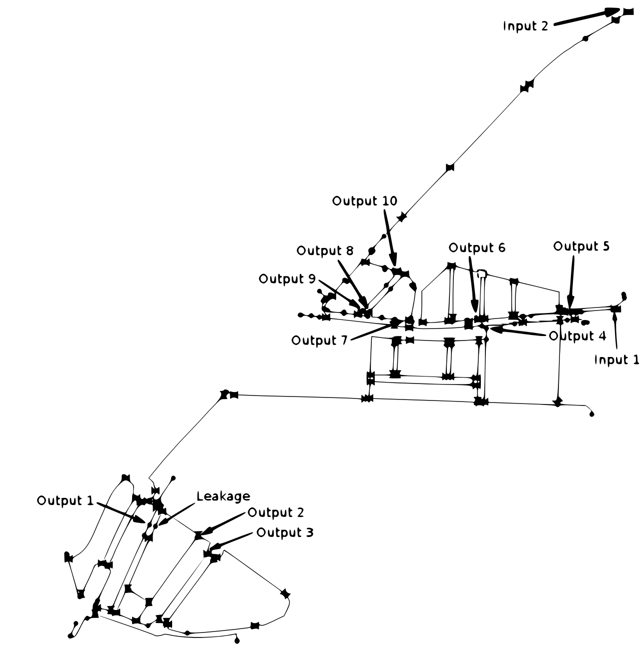

Fig. 1. Illustration of the network with two inlets.

indicated with the arrows marked 'Input #' (so $c = 2$) and $\hat{p} \in \mathbb{R}^2$ ($\hat{d} \in \mathbb{R}^2$) is the vector of pressures (demands) at these two vertices. The vector $\bar{p} \in \mathbb{R}^{397}$ ($\bar{d} \in \mathbb{R}^{397}$) consists of the pressures (demands) at the remaining 397 vertices in the network. Ten of the pressures in the vector $\bar{p}$ are measured, and the measurement points are indicated by the arrows marked 'Output #' (so $y \in \mathbb{R}^{10}$). The vertex emulating the leakage is indicated by the arrow marked 'Leakage'.

Since the network has more than one inlet, we should confirm that Assumption 3 is fulfilled. Otherwise, the reduced order model (7) is not applicable for calculating the expected pressure at the measurement points. Fig. 2 illustrates the total head at the two inlets in the network during the test. As it can be seen in the figure, there

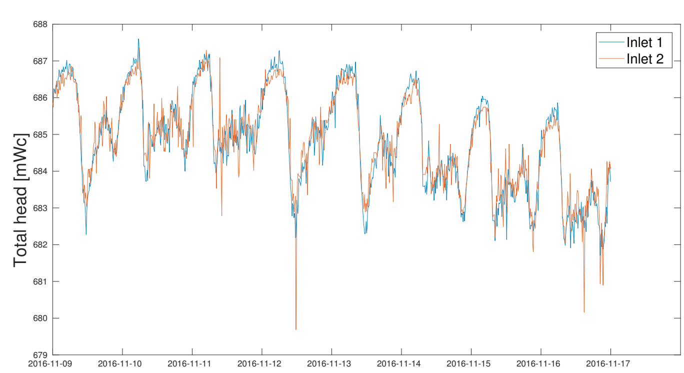

Fig. 2. Time series of the total head at the two inlets during the test.

is some discrepancy between the total heads at the two inlets. However, overall they follow each other quite well. The maximal absolute difference between the two heads at any given time is 2.6 [mWc] and the RMS value of the difference is 0.4 [mWc]. The value of $\kappa(t)$ used for the model (7) in the test is taken as the mean over the two inlet heads at any given sample.

Fig. 3 illustrates the residuals obtained from the data using (9) and (10). In the two top plots, the figure

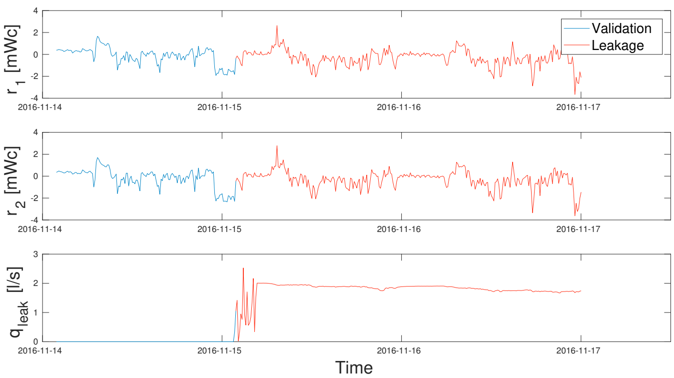

Fig. 3. Plot of the residuals r_1 and r_2 during the test together with the leakage flow q_{leak}.

shows the results for the two residuals obtained from the measurement vertices marked 'Output 1' and 'Output 2' in Fig. 1. The bottom plot in Fig. 3 gives the leakage flow at the vertex marked 'Leakage'. The blue lines in the plot indicates that the result is obtained from a validation data set, where there is no leakage in the network, while the lines in red indicates results obtained during the leak. The maximal size of the leak during the test is 2.5 [l/s], while the mean value of the leak is 1.8 [l/s] (corresponding to respectively ≈57 % and ≈41 % of night-time demand).

As it can be seen in Fig. 3, the difference between the residuals generated during the situation without a leak versus the situation with the leak is visually difficult to establish. However, as illustrated in Fig. 4 there is a statistical difference between the two situations. The top

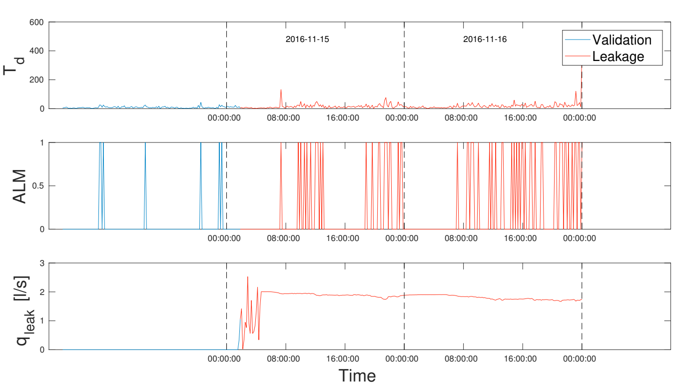

Fig. 4. Result of performing the statistical test on the residuals during the leakage test.

plot in the figure gives the test variable T_d during the test. The middle plot is an alarm signal with the following definition

$$\text{ALM} = \begin{cases} 0 \ , \ T_d \leq J_d \\ 1 \ , \ T_d > J_d \end{cases} . \tag{18}$$

The probability of a false alarm is chosen to be 0.1, which gives a threshold $J_d = 23.21$ from the χ^2-distribution with ten-degrees of freedom. As it can be seen in the figure, a few false alarms are being generated during the validation data set, whereas the frequency of alarms is much higher during the leakage. Furthermore, it appears that it is difficult to detect any difference during the early hours of the day (from 00:00:00 to 08:00:00). We expect this is due to the fact that the total demand is generally low during this period with the consequence that the flow in the system is low. Since we use pressure signals to detect the leakage a low sensitivity towards flow variations during periods with low flow is expected (due to (1)).

In Fig. 5, the localization results obtained using the described sensitivity method on the generated residuals are illustrated. The first six rows gives the result of the local-

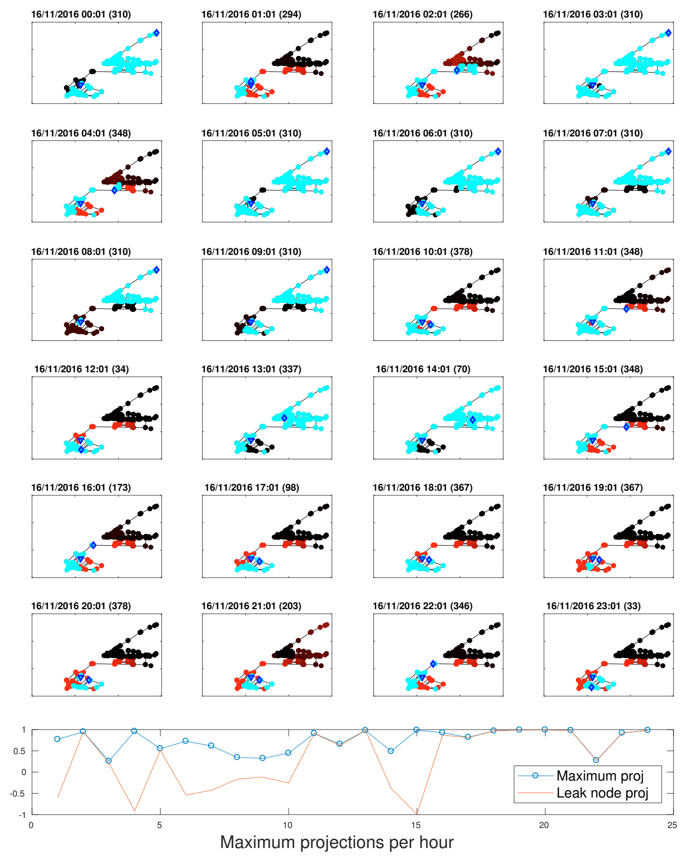

Fig. 5. Leakage localization results obtained from the network.

ization with results obtained from four hours in each row. That is, each subplot in the first six rows corresponds to the results from one given hour. The number in parenthesis above each plot is the index of the selected leak candidate. The results are colour coded in the following way. Vertices which have a small normalised projection of the residual onto its corresponding column in the sensitivity matrix are black. The vertices are increasingly red with larger normalised projections. Vertices in turquoise are in the set $\bar{\mathcal{V}}^l$ of leakage candidates. Here $\bar{\mathcal{V}}^l$ is given by all vertices with an equal or higher projection than the vertex with the leak. This means that the more turquoise vertices in the figure, the worse the performance of the localization is. However, this is of course not possible in practice and a formal method for choosing ε in (17) is needed. The vertex with the leak is marked with a blue triangle and the chosen leak candidate which has maximal projection is marked with a blue diamond. The bottom row in Fig. 5 gives the value of the normalised projection at each hour for the leaking node and for the node with the largest projection.

5. CONCLUSION

In this paper, we have tested the performance of a set of residuals generated using a data driven pressure model in a water supply network. Using the sensitivity based method, the residuals were used in a leakage localization scheme. The results show that it is possible to isolate the leakage to a limited set of candidate vertices which contains the leaking vertex. Furthermore, the most likely candidate vertex appointed using the method is generally close to the actual leaking vertex. The method seems to perform best when the flow in the system is high, which we see as natural since the sensitivity of the pressure drop with respect to the flow is increasing with the flow.

As described in the paper, the residual generation is based on data driven network pressure models while the generation of the sensitivity matrix is based on an EPANET model of the network. Future work will consist of investigating data driven localization methods which are not relying on EPANET models to generate the sensitivity matrix. In addition, parameters such as the age of pipes could be introduced in the leakage localization scheme since it is expected that older pipes are more likely to break than newer ones.

Finally, it is noted that a few isolated false alarms appeared when executing the algorithm on the non-leakage validation data. However, it is noted that these were isolated events, whereas during the leakage, the alarms appear almost constantly. The tuning of the probabilistic threshold, and/or some form of filtering of alarms (considering them as outputs of a stochastic process rather than as independent stochastic variables), should thus also be investigated in the future.

REFERENCES

Andersen, J.H. and Powell, R.S. (2000). Implicit state-estimation technique for water network monitoring. *Urban Water*, 2(2), 123 – 130.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 3rd edition.

Bicik, J., Kapelan, Z., Makropoulos, C., and Savić, D.A. (2011). Pipe burst diagnostics using evidence theory. *Journal of Hydroinformatics*.

Bonada, E., Meseguer, J., and Mirats, J.M. (2014). Practical-oriented pressure sensor placement for model-based leakage location in water distribution networks. *International Conference on Hydroinformatics*.

Bort, C.M.G., Righetti, M., and Bertola, P. (2014). Methodology for leakage isolation using pressure sensitivity and correlation analysis in water distribution systems. *Procedia Engineering*.

Casillas, M.V., Garza-Castañón, L.E., Puig, V., and Vargas-Martinez, A. (2015). Leak signature space: An original representation for robust leak location in water distribution networks. *Water*, 7(3), 1129 – 1148.

Casillas, M.V., Puig, V., Garza-Castañón, L.E., and Rosich, A. (2013). Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *MDPI, Sensors*, 13, 14984 – 15005.

Deo, N. (1974). *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, 1st edition.

Gutermann (2015). Case study: Increasing efficiency with permanent leakage monitoring. White paper. [Online].

Jensen, T.N. and Kallesøe, C.S. (2016). Application of a novel leakage detection framework for municipal water supply on aau water supply lab. *3rd Conference on Control and Fault-Tolerant Systems (SysTol)*.

Jensen, T.N., Kallesøe, C.S., Bendtsen, J.D., and Wisniewski, R. (2018a). Plug-and-play commissionable models for water supply networks with multiple inlets. In *Proc. 16th EUCA European Control Conference*. Limassol, Cyprus.

Jensen, T.N., Kallesøe, C.S., Wisniewski, R., and Bendtsen, J.D. (2018b). Residual generation for isolation of leakages in water supply networks with multiple inlets. In *Proc. 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*. Warsaw, Poland.

Kang, D. and Lansey, K. (2009). Real-time demand estimation and confidence limit analysis for water distribution systems. *Journal of Hydraulic Engineering*, 135(10), 825 – 837.

Ljung, L. (1999). *System identification -theory for the user*. Prentice-Hall, 2nd edition.

Mashford, J., Silva, D.D., Marney, D., and Burn, S. (2009). An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. *Proc. 3rd International Conference on Network and System Security*.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., and Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, 19(10), 1157 – 1167.

Pérez, R., Sanz, G., Puig, V., Quevedo, J., Cugueró-Escofet, M.A., Nejjari, F., Meseguer, J., Cembrano, G., Mirats-Tur, J.M., and Sarrate, R. (2014). Leak localization in water networks – a model-based methodology using pressure sensors applied to a real network in barcelona. *Control Systems Magazine*, 34(4), 24 – 36.

PWiK (2015). Innovativeness' of the environmental innovation. The Water Supply and Sewerage Joint Stock Company of the Częstochowa District (PWiK), White paper. [Online].

Scolnicov, H. and Horowitz, G. (2010). Water network monitoring: A new approach to managing and sustaining water distribution infrastructure. Takadu white paper.

Sensus (2012). Water 20/20: Bringing smart water networks into focus. Report. [Online].

Steffelbauer, D., Neumayer, M., Günther, M., and Fuchs-Hanusch, D. (2014). Sensor placement and leakage localization considering demand uncertainties. *Procedia Engineering*.