CrossMark

# Efficient Sensor Placement for Leak Localization Considering Uncertainties

**David B. Steffelbauer**[1] (ID) · **Daniela Fuchs-Hanusch**[1]

**Abstract** Losses in water distribution systems can be between several percent in well maintained systems up to more than 50 percent in developing countries. Most of the losses originate from leaks. Therefore, a fast detection and localization of leaks is crucial for effectively reducing this losses in water distribution networks. Model-based leak localization has become increasingly popular in recent years. Certainly, the performance of these methods is linked to 1) the measurement locations in the system and 2) uncertainties at these locations. This paper provides a methodology that incorporates uncertainties of different types and sources in the optimal sensor placement problem for leak localization shown by the example of the effect of demand uncertainties on potential pressure measurement points. The problem is solved for different numbers of sensors and different strengths of uncertainties are taken into account. Additionally, to describe the relation between the number of sensors and the leak localization quality, a cost-benefit function is derived based on the different sensor placement results and GoF statistics. It was found that the function follows a power law. Results show that incorporating uncertainties leads to other optimal positions than without uncertainties, but the power law behavior still stays true. Additionally, more sensors are needed than without uncertainties.

---

---

✉ David B. Steffelbauer
david.steffelbauer@tugraz.at

1   Institute of Urban Water Management and Landscape Water Engineering, Graz University
    of Technology, Stremayrgasse 10/I, 8010 Graz, Austria

# 1 Introduction

In well maintained water distribution systems (WDS), only 3 - 7 % of water is lost. Though in developing countries this number can go up to 50 % and more (Colombo et al. 2009). Most of the lost water originates from leaks.

Therefore, water loss reduction through locating and repairing leaks in WDS is becoming increasingly important.

In terms of leak localization water utilities follow two different approaches, the *active* and *passive* leakage control (Puust et al. 2010). In case of the *passive* approach, the utility waits until customers report leaks because they experience supply problems or leakages find their way to the surface and thus become visible. However, not all leaks become visible leading to high unreported leaks and hence to high total water losses.

The better strategy to reduce water losses is *active* leakage control (Farley and Trow 2003). One central task of an *active* leakage control strategy is searching for leaks on a regular basis. Different methods are applied for finding the position of leaks, like e.g. step testing or by temporary placing acoustic loggers. These methods are referred to as leak localization, which has the purpose of limiting the possible location of a leak to several hundred meters. Leak localization has to be followed by pinpointing to get the exact position of the leak e.g. with ground microphones, leak-noise correlators or gas leak detection.

Model-based leak localization methods have been introduced over the past few decades using measurement data from hydraulic sensors along with hydraulic models. In general, two different techniques are available. First, steady state analysis based techniques, second, transient based techniques (Colombo et al. 2009).

This paper is based on steady state leak localization, which was first proposed by Pudar and Liggett (1992). They formulated an inverse problem where the discrepancy between real-world measurements and data obtained from hydraulic simulations was minimized. Since then, leak localization by solving inverse model-based approaches has been extensively investigated (e.g. Wu et al. 2010; Pérez et al. 2011).

In fact, the success of model-based leak localization depends to a great extent on the choice of the measurement positions within a WDS (Kang and Lansey 2010). For that reason optimal sensor placement (OSP) has the potential for significantly improving the localization of leaks.

While much research work has been put into dealing with OSP for contaminant detection and WDS model calibration (see for example Savic et al. 2009), only few studies have focused on the same problem for leak localization.

The first discussion and analysis of OSP for leak localization emerged in 2008 with the work of Farley et al. (2008, 2010, 2013), followed in 2009 by Pérez et al. (2009). Both OSP approaches are based on the leak sensitivity matrix. The approach of Pérez et al. (2009) binarizes this matrix which leads to a loss of information (Quevedo et al. 2011).

To overcome this information loss, Casillas et al. (2013) formulated the OSP as an integer optimization problem based on projections from a non-binarized leak sensitivity matrix solved with a semi-exhaustive search and a genetic algorithm (GA).

Subsequently, Pérez et al. (2014) defined a greedy-search algorithm, which minimizes the maximum distance of leak scenarios to the gravity center of the nodes with projections over 99 % of the maximum value of the projections. This resulted in more robust sensor placements.

Recently, Cagueró-Escofet et al. (2015) extended the approach of Casillas et al. (2013) with a relaxed isolation index taking practical applications like the acceptable isolation distance into account.

Alternatively, Sarrate et al. (2012) invented a strategy maximizing a "leak isolability index" based on structural analysis of a WDS solved by a depth-first search algorithm. Since this strategy is only capable of medium-sized networks, it is combined in Sarrate et al. (2013) with clustering techniques to reduce the size and complexity of the OSP problem. However, due to its simple description, structural analysis cannot ensure that the diagnosis performance obtained from such models hold for real-world systems. Therefore, in Sarrate et al. (2014) the approach is combined with projections calculated again from sensitivity matrices.

Nejjari et al. (2015) invented an algorithm based on pressure sensitivity matrix analysis using an exhaustive search strategy with average worst leak expansion distances as sensor placement performance measure.

Currently, Casillas et al. (2015) placed sensors in an optimal way by minimizing the overlapping signatures in the leak signature space.

A completely different approach was introduced by Christodoulou et al. (2013), who invented an entropy-based OSP algorithm by maximizing the total entropy in the network.

In hydraulic modeling as well as in measuring hydraulic parameters one has to face several sources of uncertainty (Hutton et al. 2014). We therefore assume that these effects should be taken into account in OSP algorithms. Blesa et al. (2014) studied the robustness of the methodology introduced by Sarrate et al. (2014) against sensitivity matrix uncertainties. The main result was that the sensor positions are not sensitive to the size of the leaks, but to the working point of the WDS. Furthermore, Blesa et al. (2016) used multi-objective optimization to place sensors. The objectives were to determine the mean and the worst locatability index.

Both papers investigate the consequences of uncertainties on the OSP algorithms, but neither of them incorporates these uncertainties in the OSP itself. To the best of our knowledge, Steffelbauer et al. (2014) was the first study incorporating uncertainties in the OSP problem. Unfortunately, the problem was only solved for four sensors.

The aim of this paper is to extend the work described in Steffelbauer et al. (2014) providing a methodology that uses cost-benefit functions for predicting the sensor placements quality for leak localization depending on the number of sensors as well as the strength of general and not only demand uncertainties. Therefore, it provides WU with a methodology that helps them to decide how many sensors should be placed in their system for leak localization.

## 2 Methodology

### 2.1 Efficient Sensor Placement - General Approach

Figure 1 depicts an overview of the methods used in this paper and their chronological order.

First, hydraulic simulations are used to calculate the sensitivity matrices $s$ and the residuals $r$ (see Section 2.2). $s$ and $r$ are later used in the OSP algorithm.

Additionally, the hydraulic model is used to calculate the effect of uncertain input parameters on the outputs of the hydraulic model ($\sigma$), called model output uncertainty
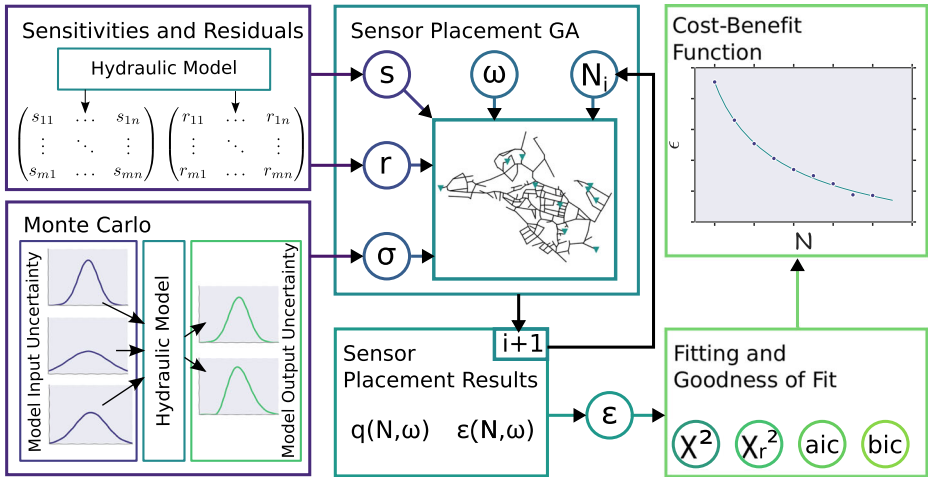
**Fig. 1** Flowchart of the methodology

(MOU) throughout this paper. The MOU calculations are performed through Monte Carlo simulations (MCS) (Section 2.3).

The MOU obtained from MCS are integrated into the sensor placement algorithm of Casillas et al. (2013), by extending this approach to punish measurement points with high uncertainties. The effect of uncertain demands on the pressures of the hydraulic model serves as an example of application throughout this paper.

Together with the number $N$ of sensors and a weighting factor $\omega$ to adjust the strength of the uncertain demands on the sensor placement, $s$, $r$ and $\sigma$ are used with a GA (see Section 2.5) for placing the sensors in an optimal manner (Section 2.4) resulting in a leak localization quality parameter $\epsilon$.

The calculation of $\epsilon$ for different $N$ and $\omega$ allows the derivation of sensor placement cost-benefit functions and different fit-functions are tested to describe the behavior of $\epsilon$ as a function of $N$. Goodness-of-fit (GoF) statistics (Section 2.6) are then used to find the cost-benefit function which fits the simulation data the best.

Sensitivities, residuals, MCS and the OSP are calculated by means of OOPNET, an object-oriented Python network analysis tool for WDS (see Steffelbauer and Fuchs-Hanusch 2015). Leaks are simulated using the power leak law (Ferrante et al. 2014) outflow equation

$$Q = c_e \cdot p^{e_e} \qquad , \qquad (1)$$

where $Q$ is the leak outflow, $c_e$ and $e_e$ the emitter coefficient and exponent and $p$ the pressure at the leak position.

## 2.2 Calculating Sensitivity Matrices and Residual Vectors

The sensitivity matrix $S$ states the key to decide where to install pressure measurements in the system (Pudar and Liggett 1992)

$$S = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \qquad . \qquad (2)$$

$m$ is the number of leak scenarios and $n$ is the number of possible measurement positions. Leak scenarios are the output of hydraulic simulations by introducing a leak of a specific size and on a specific position in the system. The individual elements of $S$ are calculated by subtracting pressures of simulations with no leak $\hat{p}_j$ from the pressures which are calculated at the same position $p_j^{f_i}$ under leak scenario $i$ with a leak of size $f_i$ and normalizing according to the leak size

$$s_{ij} = \frac{\partial p_j}{\partial f_i} = \frac{p_j^{f_i} - \hat{p}_j}{f_i} \quad . \tag{3}$$

In order to estimate the general sensitivity of a potential measurement point with respect to all possible leak scenarios, the mean over the column of $S$ is also calculated

$$\bar{s}_j = \frac{1}{m} \sum_{i=1}^{m} s_{ij} \quad . \tag{4}$$

This is conducted for the graphical representation of the sensitivity calculation results (see for example Fig. 2a).

The leak localization also relies on the computation of residuals, which are expressed as the difference between real-world measurements and the expected behavior of the system described by the hydraulic model. The residuals $r_{ij}$ are thus built in a similar manner to the sensitivity matrices, but in general with a different leak size $f_i$ and without normalization

$$r_{ij} = p_j^{f_i} - \hat{p}_j \quad . \tag{5}$$

This results in a matrix $R$ which represents measurements in the system. Furthermore, the $i$-th column of $R$ is called $i$-th residual vector $\mathbf{r_i}$.

### 2.3 Model Output Uncertainty Calculation with Monte Carlo Simulations

The quality of a hydraulic model relies heavily on the model input parameters. Generally, these input parameters are fraught with uncertainty. Consequently, the uncertain input parameters will lead to MOU.
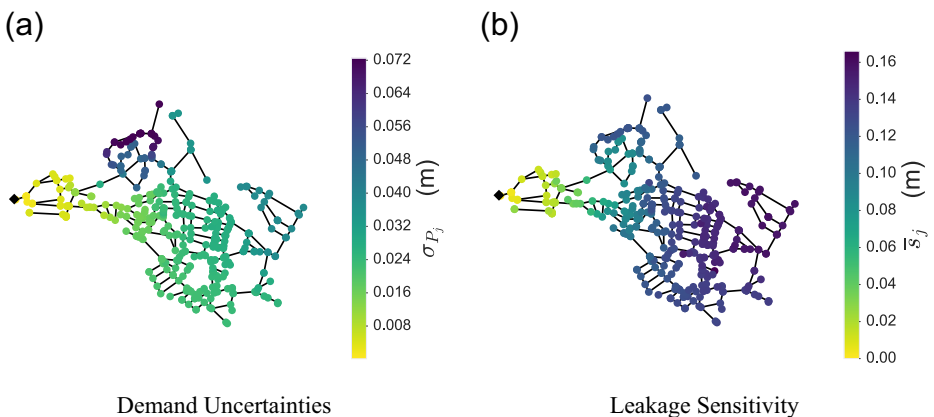


**Fig. 2** (**a**) Effects of demand uncertainties on nodal pressure and (**b**) overall sensitivity of nodal pressures to leaks

Hydraulic solvers like EPANET (Rossman et al. 2000) merely simulate the deterministic forward model without taking uncertainties into account. To compute the MOU, the hydraulic solver must be combined with uncertainty propagation methods such as e.g. first order second moment, MCS or Latin hyper-cube sampling (Kang et al. 2009). In this paper MCS is chosen to calculate the MOU since MCS has the ability to converge to the exact uncertainty estimates if a large number of parameter sets is simulated.

MCS can in general be applied to every combination of model input and model output parameters in a hydraulic model and therefore every kind of input uncertainty can be incorporated in the enhanced sensor placement algorithm in Section 2.4. However, in order to make the method more tangible, it is described on the basis of calculating the effect of uncertain demands on the pressures in the system.

For this example, the parameter set is realized in such a way that each demand in the system is randomly drawn from a normal distribution with a mean $\mu_q$ and a standard deviation $\sigma_q$. Subsequently, EPANET is called using these randomly distributed demands to perform a hydraulic simulation for obtaining the resulting pressures $p_j$ at every node $j$ in the system. This procedure is repeated $mcsteps$-times to get the approximate probability distributions $P(p_j)$ of $p_j$ for every possible measurement position in the system.

Once $P(p_j)$ is calculated, the standard deviation $\sigma_{p_j}$ is taken from this distribution serving as MOU of the pressures in point $j$. The higher $\sigma_{p_j}$, the noisier the signal is at this point, the less ideal this location is for pressure measurements.

## 2.4 Enhanced Sensor Placement Considering Uncertainties

The sensor placement algorithm described in Casillas et al. (2013) is based on calculating projections $\psi_{ij}$ between $\mathbf{r}_i$ and $\mathbf{s}_j$

$$\psi_{ij}(\mathbf{q}) = \frac{\mathbf{r}_i^T Q(\mathbf{q})\mathbf{s}_j}{|\mathbf{r}_i^T Q(\mathbf{q})^T||Q(\mathbf{q})\mathbf{s}_j|} \qquad , \tag{6}$$

where $Q(\mathbf{q})$ is a diagonal matrix constructed from a binary vector $\mathbf{q}$ with the length of the possible sensor positions, where $q_i$ is 1, if a sensor is placed at node $i$ respectively 0 for no sensor. The largest projection value represents the supposed leak spot. If this value lies on the diagonal the correct leak spot is found by the sensor configuration $\mathbf{q}$. Therefore, an error index $\epsilon(\mathbf{q})$ is defined

$$\epsilon(\mathbf{q}) = \begin{cases} 0 \ \dots \ \text{if } \psi_{ii}(\mathbf{q}) = \max(\psi_{i1}(\mathbf{q}), \dots, \psi_{im}(\mathbf{q})) \\ \qquad 1 \ \dots \ \text{otherwise} \end{cases} \qquad . \tag{7}$$

The mean over all leak scenarios takes every possible leak into account

$$\overline{\epsilon}(\mathbf{q}) = \sum_{i=1}^{m} \frac{\epsilon_i(\mathbf{q})}{m} \quad \rightarrow \quad \min_{\mathbf{q}} \overline{\epsilon}(\mathbf{q}) \qquad . \tag{8}$$

This value must be minimized to find the optimal sensor configuration $\mathbf{q}$. The lower bound of $\overline{\epsilon}(\mathbf{q})$ is 0, denoting that every leak in the system is found correctly by the sensor configuration $\mathbf{q}$. If no leak scenario is identified correctly, $\overline{\epsilon}(\mathbf{q})$ is 1.

To incorporate MOU in $\Psi(\mathbf{q})$, the pressure uncertainties $\sigma_{p_i}$ (see Section 2.3) are subtracted from $\mathbf{r}_i$. This punishes points with high MOU, since high MOU leads to a high distortion of the residual vectors $\mathbf{r}_i$ resulting in smaller $\psi_{ij}(\mathbf{q})$. Therefore, points with

sensors at these positions will be less optimal. The extension of Eq. 6 under consideration of demand uncertainties results in

$$\psi_{ij}(\mathbf{q}) = \frac{\left(\mathbf{r}_i^T - \omega\sigma_{p_i}\right) Q(\mathbf{q})\mathbf{s}_j}{\left|\left(\mathbf{r}_i^T - \omega\sigma_{p_i}\right) Q(\mathbf{q})^T\right|\left|Q(\mathbf{q})\mathbf{s}_j\right|} \quad , \tag{9}$$

where $\omega$ represents a weighting factor controlling the strength of the influence of MOU on the OSP.

Furthermore, in a District Metering Area (DMA) flow and pressure of the inlet and outlet of the system are assumed to be known. For this reason fixed pressure sensor positions are incorporated into the sensor placement algorithm. This is achieved by setting $q_j = 1$ if $j$ is the inlet respectively the outlet of the system.

## 2.5 Solving the Optimal Sensor Placement Problem with a Genetic Algorithm

GAs are widely used to obtain optimal solutions to countless problems in water related fields (Nicklow et al. 2010). Meta-heuristics like GA are the only alternative for finding optimal solutions in reasonable time for problems with enormous search space sizes.

For example, in the OSP problem calculating $\overline{\epsilon}(\mathbf{q})$ for every possible combination of sensors can result in an almost countless solution space. If there are $M$ different possible measurement positions for pressure sensors in a network, there are $\mathscr{C}$ possible sensor combinations for placing $N$ sensors

$$\mathscr{C} = \binom{M}{N} \quad . \tag{10}$$

This results in $\mathscr{C}$ independent calculations of $\overline{\epsilon}(\mathbf{q})$, which grows exponentially with $N$.

The GA used in this paper is a simple GA as described in Bäck et al. (2000). The genome of the individuals in the GA consists of $N$ integers, were $N$ is the number of sensors. Population size is set to 100 and the evolution takes place for 100 generations. Single crossover is used with a probability of $p_c = 80\%$ followed by uniform integer mutation with a probability of $p_m = 20\%$. Tournament selection with a tournament size of $k = 3$ is chosen.

## 2.6 Finding and Fitting an Ideal Cost-Benefit Function

Running the GA described in Section 2.5 for OSP defined in Section 2.4 for a fixed $\omega$ and different number of sensors $N$ results in different $\overline{\epsilon}(\mathbf{q})$ values. The values of $\overline{\epsilon}(\mathbf{q})$ serve as a sensor placement quality parameter, since it is connected to the number of leak scenarios which have not been allocated appropriately. The percentage $L_\%$ of correctly found leak scenarios is

$$L_\% = 100 \cdot (1 - \overline{\epsilon}(\mathbf{q})) \quad . \tag{11}$$

As $N$ increases, $\overline{\epsilon}(\mathbf{q})$ has to decrease since more sensors lead to better leak localization. The behavior of $\overline{\epsilon}(\mathbf{q})$ as a function of $N$ can be interpreted as a cost-benefit analysis answering the following questions:

– How many sensors are needed to identify a specific number of leak scenarios correctly?
– How does the OSP improve leak localization if $M$ more sensors are placed?

For reasons of simplicity, the cost is simplified here as the number of sensors whereas the benefit is the quality of the leak localization in terms of leak scenarios found.

Nevertheless, carrying out this analysis is computationally expensive and leads only to single points in dependency of $N$. Finding a cost-benefit function which describes the

behavior of $\bar{\epsilon}(\mathbf{q})$ due to $N$ will lead to computational shortcuts, since the GA does not have to find all possible quality parameters as a function of $N$ and might also lead to better insights into the sensor placement problem itself. Two questions arise; First, how to fit a function optimally to the $\bar{\epsilon}(\mathbf{q})$? Second, how to decide which function best describes the data?

For answering the first question the Levenberg-Marquardt algorithm is applied.

To answer the second question, GoF statistics are generally used. In this paper four different GoF measurements are applied to find the best cost-benefit function. The first one is $\chi^2$ statistics

$$\chi^2 = \sum_{i=1}^{n} (y_i - f(x_i))^2 \quad , \tag{12}$$

where $y_i$ is the datapoint on position $x_i$ and $f(x_i)$ is the value calculated from the fit-function and $n$ is the total number of data points. $\chi^2$ does not solve the problem of over-fitting. Therefore, the degrees of freedom $\nu$, respectively the number of function parameters $(n - \nu)$, must be taken into account. This is done by using the reduced $\chi^2$ statistics ($\chi_\nu^2$)

$$\chi_\nu^2 = \frac{\chi^2}{\nu} \quad . \tag{13}$$

Third GoF measure is Akaike's information criterion ($aic$), which penalizes the number of parameters of the fitting function more than $\chi_\nu^2$

$$aic = n \ln\left(\chi^2/n\right) + 2(n - \nu) \quad . \tag{14}$$

Bayesian information criterion ($bic$) punishes the number of parameters even more than $aic$

$$bic = n \ln\left(\chi^2/n\right) + (n - \nu) \ln(n) \quad . \tag{15}$$

The smaller the value of all four criteria compared to other fitting functions, the better the chosen model describes the data.

## 2.7 Case Study

The hydraulic model of the real-world case study WDS consists of one tank at the inflow point, 392 nodes and 452 pipes with a total length of approximately 37 km. The pipe diameters vary from 70 to 400 mm. The pressure at night minimum flow varies between 29.31 and 43.46 m. Flow and pressure are continuously measured at the inflow and outflow point. The hydraulic model was calibrated in terms of pipe roughness several years ago. The nodal base demands were allocated from the billing information of the particular customers. The area is mainly a residential zone with a few small industries. For the purpose of leak localization and hence for the application of the introduced sensor placement approach, the nodal minimum night consumption is of interest. This was derived for every node from the assigned demand patterns which were also available from the hydraulic model. These patterns were derived during the model calibration process and hence are based on real-world measurements. The uncertainties resulting from this very general demand allocations are taken into account with the method described in Section 2.3.

## 3 Results

First, the effect of demand uncertainties on possible pressure measurement points is investigated according to Section 2.3. MCS with 10000 *mcsteps* are performed with random demands drawn from a normal distribution using the nodal demand at minimum night flow as mean $\mu_q$ and a standard deviation $\sigma_d$ of 10 %. The resulting $\sigma_{p_i}$ of the nodal pressures are depicted in Fig. 2a.

The sensitivity matrix $S$ is calculated as described in Section 2.2. Leakage scenarios were generated with an emitter coefficient of $c_e = 0.5$ in EPANET for every node in the system. Only one leak size $c_e = 0.5$ was chosen to build the sensitivity matrices, since Blesa et al. (2016) has found that the sensor positions are not sensitive to leak sizes. The overall sensitivity at every possible measurement position $\overline{s_j}$ in the system is calculated by Eq. 4 and plotted in Fig. 2b.

In the upper region of the case study area exist points that are sensitive to demand uncertainties Fig. 2a. Therefore, this points should be punished by an OSP algorithm. For solving the OSP problem it is needed to compute the residuals $R$. In this paper $R$ are calculated with the same emitter coefficient $c_e = 0.5$ as $S$.

$\sigma_{p_i}$, $S$ and $R$ are then used for solving the enhanced OSP problem described in Section 2.4. The OSP problem is solved for a different $N$ ranging from 2 to 10 and different $\omega$ (with $\omega = (0.0, 0.125, 0.25, 0.5, 1.0)$).

The range of $N$ was chosen, because for $N = 1$ the ideal sensor position is trivially at the point with the highest overall sensitivity and $N > 10$ was viewed as uneconomical for the size of the investigated WDS.

$\omega$ was chosen between $\omega = 0.0$, which corresponds to the case with no uncertainties, and $\omega = 1.0$, which takes the uncertainties into account with full strength. Simulations with $\omega = 1.0$ have shown that the sensors cluster too much in regions with low demand uncertainties and are not spread anymore over the whole system (e.g. see Fig. 3h). Therefore, $\omega$ was halved so long ($\omega = 1.0 \xrightarrow{1/2} 0.5 \xrightarrow{1/2} 0.25 \xrightarrow{1/2} 0.125$) until sensors appeared again in the upper part of the system with high demand uncertainties.

For each combination of $N$ and $\omega$ the GA described in Section 2.5 is applied 10 times solving the enhanced OSP. Due to the stochastic behavior of the GA, the resulting sensor placement quality parameters $\overline{\epsilon}(\mathbf{q})$ may vary for different optimization runs with same $N$ and $\omega$ parameters. Table 1 thus illustrates the statistics of $\overline{\epsilon}(\mathbf{q})$ found during 10 optimization runs and contains the mean, the minimum and the standard deviation of $\overline{\epsilon}(\mathbf{q})$ of all runs.

In Fig. 3 the optimal sensor positions for selected values of $N$ and $\omega$ are presented[1]. The first four plots (Fig. 3a–d) show the OSP results for $N = 2$ to $N = 5$ with $\omega = 0.0$, whereas the last four plots Fig. 3e and h show results for five sensors under incorporation of uncertainties with varying $\omega$ (between 0.125 and 1.0).

Without uncertainties the algorithm tends to place sensors in regions with high demand uncertainties (upper part in Fig. 2). This can be seen in Fig. 3a–d. The higher $\omega$, the more the sensors move to regions with low demand uncertainties (see Fig. 3d–h). Therefore, the extension of the OSP algorithm by incorporating uncertainties shows the desired behavior by punishing measurement points with high uncertainties. Nevertheless, this decreases the spreading of the sensors over the whole system since sensors tend to cluster in areas with low uncertainties.

---

[1]Due to lack of space the results for all combinations of $N$ and $\omega$ can be found in the supplementary material of this paper.

(a)

$N = 2; \omega = 0.0$

(b)

$N = 3; \omega = 0.0$

(c)

$N = 4; \omega = 0.0$

(d)

$N = 5; \omega = 0.0$

(e)

$N = 5; \omega = 0.125$

(f)

$N = 5; \omega = 0.25$

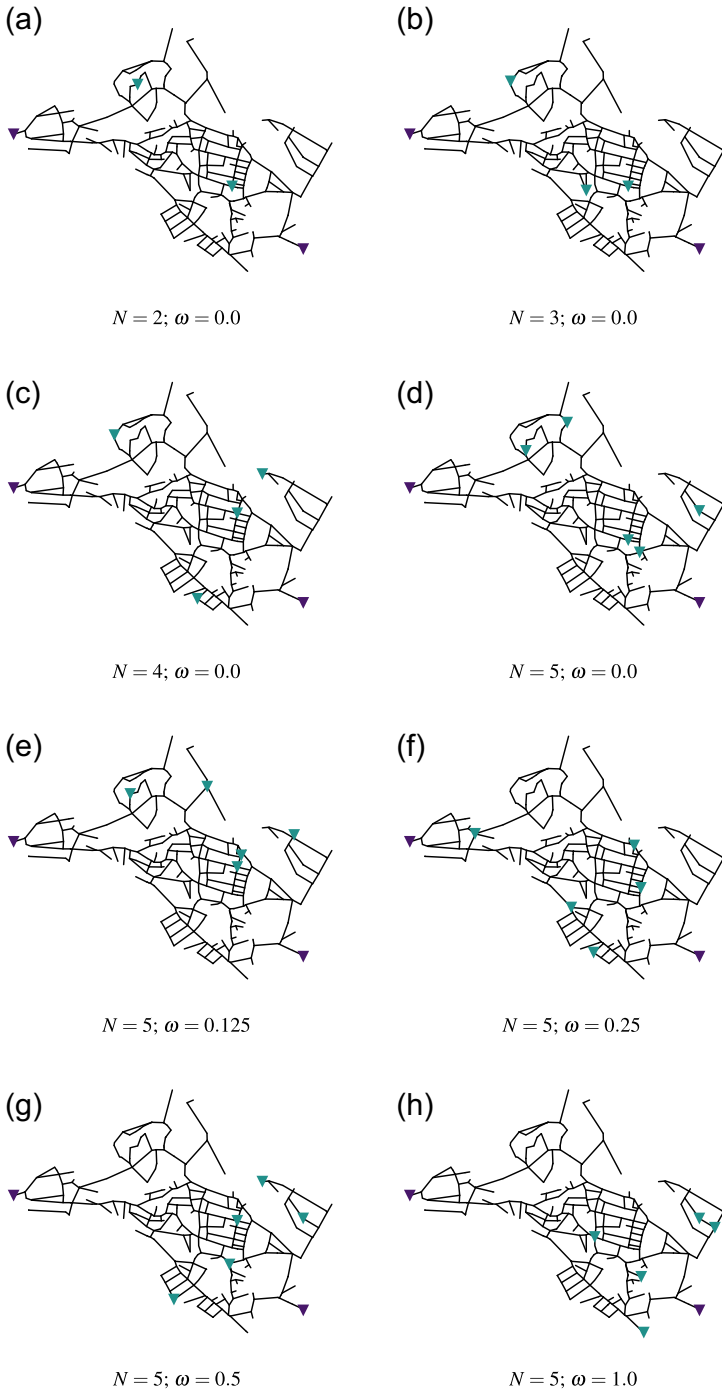(g)

$N = 5; \omega = 0.5$

(h)

$N = 5; \omega = 1.0$

**Fig. 3** Sensor positions (*green triangles*) resulting from the enhanced sensor placement algorithm described in Section 2.4 for different number of sensors $N$ and different weighting factors $\omega$. Already installed sensors at the inflow and outflow point of the DMA are depicted as *purple triangles*

**Table 1** Results for enhanced sensor placement algorithm for different numbers of sensors $N$ and different $\omega$ parameters

| $N$ | $\omega = 0.0$ | | | $\omega = 0.125$ | | | $\omega = 0.25$ | | | $\omega = 0.5$ | | | $\omega = 1.0$ | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | mean | min | std | mean | min | std | mean | min | std | mean | min | std | mean | min | std |
| 2 | 0.327 | 0.316 | 0.007 | 0.345 | 0.334 | 0.007 | 0.399 | 0.378 | 0.009 | 0.467 | 0.459 | 0.005 | 0.559 | 0.546 | 0.008 |
| 3 | 0.265 | 0.247 | 0.010 | 0.284 | 0.270 | 0.009 | 0.312 | 0.301 | 0.009 | 0.372 | 0.355 | 0.010 | 0.444 | 0.423 | 0.014 |
| 4 | 0.227 | 0.207 | 0.011 | 0.233 | 0.222 | 0.008 | 0.271 | 0.260 | 0.009 | 0.316 | 0.304 | 0.008 | 0.385 | 0.367 | 0.016 |
| 5 | 0.203 | 0.189 | 0.011 | 0.211 | 0.196 | 0.008 | 0.237 | 0.219 | 0.011 | 0.275 | 0.263 | 0.011 | 0.338 | 0.314 | 0.012 |
| 6 | 0.185 | 0.171 | 0.006 | 0.191 | 0.173 | 0.011 | 0.212 | 0.202 | 0.008 | 0.249 | 0.230 | 0.010 | 0.312 | 0.291 | 0.012 |
| 7 | 0.175 | 0.158 | 0.008 | 0.172 | 0.156 | 0.013 | 0.188 | 0.176 | 0.009 | 0.228 | 0.202 | 0.014 | 0.291 | 0.268 | 0.013 |
| 8 | 0.162 | 0.145 | 0.008 | 0.160 | 0.151 | 0.009 | 0.173 | 0.161 | 0.009 | 0.207 | 0.191 | 0.011 | 0.261 | 0.230 | 0.019 |
| 9 | 0.144 | 0.130 | 0.007 | 0.150 | 0.140 | 0.005 | 0.161 | 0.133 | 0.011 | 0.189 | 0.179 | 0.007 | 0.249 | 0.222 | 0.016 |
| 10 | 0.143 | 0.135 | 0.006 | 0.140 | 0.128 | 0.010 | 0.156 | 0.140 | 0.012 | 0.181 | 0.156 | 0.015 | 0.241 | 0.207 | 0.025 |

The statistics (minimum (min), mean and standard deviation (std)) of the $\bar{\epsilon}(\mathbf{q})$ values are obtained from 10 GA optimization runs for each $N$-$\omega$ combination

Additionally, in Fig. 3a–d it can be seen that the optimal set for $N - 1$ sensor positions $\mathscr{P}(N - 1)$ is not a subset of the optimal placement $\mathscr{P}(N)$ of $N$ sensors ($\mathscr{P}(N - 1) \not\subseteq \mathscr{P}(N)$). The algorithm leads to different sensor positions for different $N$. The same results are also found by Kapelan et al. (2005) for sensor placement for model calibration.

Finding the mathematical law behind the cost-benefit function describing the OSP is a main objective of this paper since it provides WU with a methodology to answer how many sensors should be placed in a WDS for leak localization. To predict the behavior of $\overline{\epsilon}(\mathbf{q})$ on $N$ for different $\omega$, the resulting mean values were chosen, fitted with different functions and subsequently tested with GoF statistics to figure out which mathematical law the cost-benefit functions follow.

Different functions were taken into account describing the decreasing behavior one can see in Fig. 4 of the results of the GA.

Table 2 shows the result of the fits. In the first column, the different functions which were fitted to the $\overline{\epsilon}(\mathbf{q})$ values as a function of $N$ are shown. $N_p$ is the number of parameters, $\nu$ the number of degrees of freedom. $\chi^2$ is the value resulting from $\chi^2$ statistics calculated by Eq. 12, $\chi_r^2$ is the value from the reduced $\chi^2$-statistic (13), the value for Akaike's information criterion $aic$ is calculated using Eq. 14 and the value for Bayesian information criterion $bic$ by Eq. 15. All values are the resulting means over the simulations with different $\omega$. For example, $\chi^2$ in Table 2 is calculated through

$$\chi^2 = \frac{1}{5}\left(\chi^2_{\omega=0.0} + \chi^2_{\omega=0.125} + \chi^2_{\omega=0.25} + \chi^2_{\omega=0.5} + \chi^2_{\omega=1.0}\right) \quad . \tag{16}$$

According to Table 2, function $f_6(N)$, an extended power law equation, leads to the best $\chi^2$ and $\chi_r^2$ whereas the simple power law equation $f_7(N)$ leads to the best $aic$ and $bic$ values. Table 3 shows the best-fit values of the parameters of functions $f_6(N)$ and $f_7(N)$ together with the estimated standard error for these values. The best-fit values for equation $f_6(N)$ show high standard errors, especially for parameters $a$ and $b$, whereas the best-fit values of
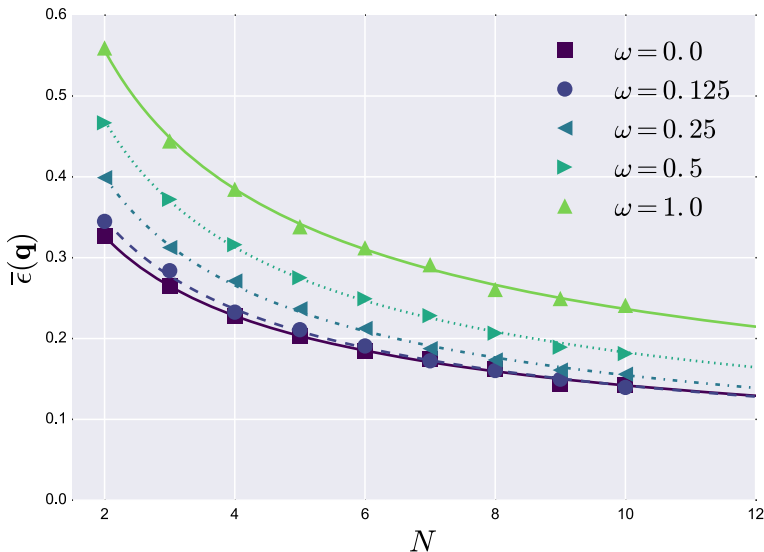


**Fig. 4** Cost-benefit functions of OSP for different $\omega$

**Table 2** Goodness-of-Fit table for different fit-functions $f_i(N)$ containing the number of parameters $N_P$ for every function, the degrees of freedom $\nu$, the resulting values for $\chi^2$ and reduced $\chi_r^2$ statistics, Akaike's ($aic$) and Bayes information criterion ($bic$)

| fit-functions | $N_P$ | $\nu$ | $\chi^2$ | $\chi_r^2$ | $aic$ | $bic$ |
|---|---|---|---|---|---|---|
| $f_1(N) = a\,e^{-b\,N}$ | 2 | 7 | 0.00272 | 0.00039 | −67.68 | −67.28 |
| $f_2(N) = a\,e^{-b\,N} + c$ | 3 | 6 | 0.01346 | 0.00224 | −67.80 | −67.21 |
| $f_3(N) = e^{-a\,N} + b$ | 2 | 7 | 0.00355 | 0.00051 | −64.37 | −63.98 |
| $f_4(N) = a\,N^2 + b\,N + c$ | 3 | 6 | 0.00082 | 0.00014 | −74.96 | −74.37 |
| $f_5(N) = e^{-a\,N} + b\,N + c$ | 3 | 6 | 0.00031 | 0.00005 | −83.12 | −82.53 |
| $f_6(N) = (a + b\,N)^{-c}$ | 3 | 6 | 0.00006 | 0.00001 | −97.65 | −97.05 |
| $f_7(N) = a\,N^{-b}$ | 2 | 7 | 0.00009 | 0.00001 | −97.93 | −97.53 |

$f_7(N)$ show much lower error bounds. Therefore $f_7(N)$ is chosen as a cost-benefit function for the values in Fig. 4.

The power-law behavior of $f_7(N)$ has practical consequences for water utilities. This is shown through an example for $\omega = 0.0$. Placing 3 sensors instead of 2 improves the quality of the sensor placement by

$$\epsilon_{N=2}^{\omega=0}(\mathbf{q}) - \epsilon_{N=3}^{\omega=0}(\mathbf{q}) = 0.327 - 0.265 \approx 0.06 \quad . \tag{17}$$

Values are taken from Table 1. To get again the same improvement for 3 sensors, the water utility has to place double as many sensors as before going from 3 to 5 sensors

$$\epsilon_{N=3}^{\omega=0}(\mathbf{q}) - \epsilon_{N=5}^{\omega=0}(\mathbf{q}) = 0.265 - 0.203 \approx 0.06 \quad . \tag{18}$$

For getting the same improvement for 5 sensors, 4 additional sensors, again double as many as before, have to be placed

$$\epsilon_{N=5}^{\omega=0}(\mathbf{q}) - \epsilon_{N=9}^{\omega=0}(\mathbf{q}) = 0.203 - 0.144 \approx 0.06 \quad . \tag{19}$$

Therefore, for a linear improvement of the quality the number of sensors has to be doubled.

Furthermore, the power-law behavior stays true under incorporation of uncertainties, hence the former example is in general true. The only difference to simulations without uncertainties ($\omega = 0.0$) is that the quality parameter of a sensor placement with a specific $N$ becomes worse if $\omega$ is increased. This makes sense, since leaks are harder to find in real-world networks with high uncertainties compared to theoretical perfect known systems. For example, to achieve the same quality parameter as in simulations with no uncertainties

**Table 3** Resulting fit parameters and error limits for fit-function $f_6(N)$ and $f_7(N)$ from Table 2

| | $f_6(N) = (a + b \cdot N)^{-c}$ | | | $f_7(N) = a \cdot N^{-b}$ | |
|---|---|---|---|---|---|
| $\omega$ | $a$ | $b$ | $c$ | $a$ | $b$ |
| 0.000 | $0.0 \pm 1.0$ | $4.0 \pm 1.0$ | $0.52 \pm 0.05$ | $0.467 \pm 0.006$ | $0.517 \pm 0.009$ |
| 0.125 | $1.0 \pm 0.5$ | $2.2 \pm 0.6$ | $0.62 \pm 0.05$ | $0.512 \pm 0.007$ | $0.557 \pm 0.010$ |
| 0.250 | $0.7 \pm 0.4$ | $1.7 \pm 0.4$ | $0.66 \pm 0.05$ | $0.603 \pm 0.008$ | $0.591 \pm 0.009$ |
| 0.500 | $0.6 \pm 0.2$ | $1.3 \pm 0.2$ | $0.66 \pm 0.03$ | $0.705 \pm 0.007$ | $0.586 \pm 0.007$ |
| 1.000 | $-0.5 \pm 0.5$ | $1.9 \pm 0.3$ | $0.49 \pm 0.03$ | $0.804 \pm 0.009$ | $0.531 \pm 0.007$ |

($\omega = 0.0$) for $N = 2$ sensors, $N = 4$ sensors are needed for $\omega = 0.5$ and $N = 6$ are needed if the uncertainty is fully taken into account ($\omega = 1.0$). This can also be seen in Fig. 4.

## 4 Discussion

Incorporating uncertainties in the OSP problem for leak localization is of great importance. Since points which are sensitive to demand variations can also be points which are sensitive to leaks (upper part of Fig. 2). Hence, leaks can be hidden by demand fluctuations for sensors placed at these points. Consequently, it may be harder to locate leaks with sensors at these points.

Nevertheless, nodes from this region are always chosen by the sensor placement algorithm if the uncertainty effects are not taken into account (see Fig. 3a–d for example). The enhanced algorithm in Section 2.4 supposes a way out of this dilemma. Increasing the $\omega$ value leads to placements avoiding these regions of high uncertainty.

Additionally, it is apparent from Fig. 3d–h that the higher $\omega$ is the more sensors are placed in regions with low uncertainties. Unfortunately, this leads to a clustering of sensors. However, a good sensor placement should also lead to positions that are spread over the whole WDS, since sensors are more sensitive to changes in their close proximity. Therefore, the OSP was solved not only for $\omega = 1.0$, taking the full uncertainty into account, but also for smaller $\omega$ values. Nevertheless, the paper lacks in providing a methodology for choosing this value.

Another important finding is that the optimal set for $N - 1$ sensor positions $\mathscr{P}(N - 1)$ is not a subset of the optimal placement $\mathscr{P}(N)$ of $N$ sensors. Therefore, OSP algorithms that solve the problem by placing one sensor and find the next sensor position through incorporation of the previous one (greedy sensor placement algorithms) may not find global optimal solutions of the problem.

It has to be mentioned that for a high number of sensors the results in Table 1 must be treated with caution, since the search space becomes very large. Placing 10 sensors on 392 measurement locations (network nodes of the case study) results in $\approx 2 \cdot 10^{19}$ possible combinations. Since the GA evaluates approximately 8000 times the fitness functions in one optimization run, this is just a small part of the total solution space.

Another interesting finding is that fitting the functions from Table 2 to the simulated sensor placement cost-benefit values lead to two equations, which both have a power law behavior ($f_6(N)$ and $f_7(N)$). $f_6(N)$ has lower $\chi^2$ and $\chi_r^2$ values while $f_7(N)$ results in better Akaike and Bayes information criteria values. A closer look at the confidence intervals of the fitness coefficients leads to favor equation $f_7(N)$. This curve is therefore chosen as the cost-benefit curve of the OSP for all $\omega$ parameters (see Fig. 4). It is surprising that higher $\omega$ parameters do not change the form of the cost-benefit curve, they merely lead to higher values of the cost-benefit function. One surprising finding is that the form of the function is robust against uncertainties.

The limitations of this study are that it can only be applied to WDS where a hydraulic model already exists since it is necessary to compute the sensitivity matrix. Additionally, the OSP algorithm is only as good as the model. For example, an unknown closed valve can have strong influence on the hydraulics and therefore the OSP result might not be optimal in reality.

Furthermore, for bigger systems the computation time increases for both, calculating the MOU as well as solving the OSP problem, making the method not directly applicable for big real world systems with thousands of nodes. Additionally, the OSP algorithm can only take single leaks into account, not multiple leaks.

## 5 Conclusion and Outlook

This paper set out to determine the effect of MOUs (e.g. the effect of uncertain demands on pressures) on OSP for leak localization in a real-world network. The research has shown that pressure measurement points, which are sensitive to leaks may also be points which are sensitive to uncertain demand and hence less ideal to place sensors on. Incorporating MOU in the OSP algorithm leads to placements avoiding regions of high uncertainty.

The simulations confirmed that the optimal sensor positions for $N - 1$ sensors is not a subset of the optimal locations of $N$ sensors. As a consequence, greedy sensor placement algorithms may fail in finding optimal measurement locations. Therefore, choosing GA to solve the OSP problem in this paper was the right decision.

Furthermore, this paper showed that the effect of the number of sensors on the sensor placement quality has a power law behavior. This even stays true under incorporation of uncertainties.

Although, the work in this paper focuses on the effect of demand uncertainties on the placement of pressure sensors, it should be mentioned that this method is more general in scope. The method is also applicable for other sensors (e.g. flow or quality sensors) and has capability for different sources of uncertainties as well (e.g. roughness, elevation, bulk reactions, ...).

A limitation of this study is the high computation complexity since the search space of the OSP problem can become very large even for small WDS. Clustering algorithms as in Sarrate et al. (2014) can show a way out in future work to reduce the complexity of the problem.

Additionally, further investigations are needed using different networks to provide proof that the power law behavior of optimal sensor positions is true. Finally, what value has to be chosen for the weighting factor $\omega$ in the sensor placement algorithm is still an open question. This will be investigated in future work.

## References

Bäck T, Fogel DB, Michalewicz Z (2000) Evolutionary computation 1: basic algorithms and operators, 1st edn. CRC press

Blesa J, Nejjari F, Sarrate R (2014) Robustness analysis of sensor placement for leak detection and location under uncertain operating conditions. Proc Eng 89:1553–1560. doi:10.1016/j.proeng.2014.11.453

Blesa J, Nejjari F, Sarrate R (2016) Robust sensor placement for leak location: analysis and design. Journal of Hydroinformatics 18(1):136–148. doi:10.2166/hydro.2015.021, http://jh.iwaponline.com/content/18/1/136

Casillas MV, Puig V, LE Garza-Castañón, Rosich A (2013) Optimal sensor placement for leak location in water distribution networks using genetic algorithms. Sensors 13(11):14,984–15,005. doi:10.3390/s131114984

Casillas MV, Garza-Castañón LE, Puig V (2015) Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. Water 7(11):6496–6515. doi:10.3390/w7116496

Christodoulou SE, Gagatsis A, Xanthos S, Kranioti S, Agathokleous A, Fragiadakis M (2013) Entropy-based sensor placement optimization for waterloss detection in water distribution networks. Water Resour Manag 27(13):4443–4468. doi:10.1007/s11269-013-0419-8

Colombo AF, Lee P, Karney BW (2009) A selective literature review of transient-based leak detection methods. J Hydro-environ Res 2(4):212–227. doi:10.1016/j.jher.2009.02.003

Cugueró-Escofet MA, Puig V, Quevedo J, Blesa J (2015) Optimal pressure sensor placement for leak localisation using a relaxed isolation index: application to the barcelona water network. IFAC-PapersOnLine 48(21):1108–1113. doi:10.1016/j.ifacol.2015.09.675

Farley B, Mounce SR, Boxall JB (2008) Optimal locations of pressure meters for burst detection. In: Water distribution systems analysis 2008. American Society of Civil Engineers, pp 1–11

Farley B, Boxall JB, Mounce SR (2010) Field validation of 'optimal' instrumentation methodology for burst/leak detection and location. In: Water distribution systems analysis 2010. American Society of Civil Engineers, pp 1093–1102

Farley B, Mounce SR, Boxall JB (2013) Development and field validation of a burst localization methodology. J Water Resour Plan Manag 139(6):604–613. doi:10.1061/(ASCE)WR.1943-5452.0000290

Farley M, Trow S (2003) Losses in water distribution networks: a practitioner's guide to assessment, monitoring and control. IWA Publ, London

Ferrante M, Meniconi S, Brunone B (2014) Local and global leak laws: the relationship between pressure and leakage for a single leak and for a district with leaks. Water Resour Manag 28(11):3761–3782. doi:10.1007/s11269-014-0708-x

Hutton C, Kapelan Z, Vamvakeridou-Lyroudia L, Savić D (2014) Dealing with uncertainty in water distribution system models: a framework for real-time modeling and data assimilation. J Water Resour Plan Manag 140(2):169–183. doi:10.1061/(ASCE)WR.1943-5452.0000325

Kang D, Lansey K (2010) Optimal meter placement for water distribution system state estimation. J Water Resour Plan Manag 136(3):337–347. doi:10.1061/(ASCE)WR.1943-5452.0000037

Kang D, Pasha MF, Lansey K (2009) Approximate methods for uncertainty analysis of water distribution systems. Urban Water J 6(3):233–249. doi:10.1080/15730620802566844

Kapelan Z, Savic D, Walters G (2005) Optimal sampling design methodologies for water distribution model calibration. J Hydraulic Eng 131(3):190–200. doi:10.1061/(ASCE)0733-9429(2005)131:3(190)

Nejjari F, Sarrate R, Blesa J (2015) Optimal pressure sensor placement in water distribution networks minimizing leak location uncertainty. Proc Eng 119:953–962. doi:10.1016/j.proeng.2015.08.979

Nicklow J, Reed P, Savic D, Dessalegne T, Harrell L, Chan-Hilton A, Karamouz M, Minsker B, Ostfeld A, Singh A, Zechman E (2010) State of the art for genetic algorithms and beyond in water resources planning and management. J Water Resour Plan Manag 136(4):412–432. doi:10.1061/(ASCE)WR.1943-5452.0000053

Pérez R, Puig V, Pascual J, Peralta A, Landeros E, Jordanas L (2009) Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Sci Technol: Water Supply 9(6):715. doi:10.2166/ws.2009.372

Pérez R, Puig V, Pascual J, Quevedo J, Landeros E, Peralta A (2011) Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Eng Pract 19(10):1157–1167. doi:10.1016/j.conengprac.2011.06.004

Pérez R, Cugueró MA, Cugueró J, Sanz G (2014) Accuracy assessment of leak localisation method depending on available measurements. Proc Eng 70:1304–1313. doi:10.1016/j.proeng.2014.02.144

Pudar R, Liggett J (1992) Leaks in pipe networks. J Hydraul Eng 118(7):1031–1046

Puust R, Kapelan Z, Savic DA, Koppel T (2010) A review of methods for leakage management in pipe networks. Urban Water J 7(1):25–45. doi:10.1080/15730621003610878

Quevedo J, Cugueró MÀ, Pérez R, Nejjari F, Puig V, Mirats J (2011) Leakage location in water distribution neworks based on correlation measurement of pressure sensors, 8th IWA Symposium on System Analysis and Integrated Assessment. International Water Association (IWA), San Sebastian, pp 290–297

Rossman LA et al. (2000) EPANET 2: users manual. US Environmental Protection Agency. Office of Research and Development. National Risk Management Research Laboratory

Sarrate R, Nejjari F, Rosich A (2012) Sensor placement for fault diagnosis performance maximization in Distribution Networks. In: 2012 20th Mediterranean conference on control automation (MED), pp 110–115. doi:10.1109/MED.2012.6265623

Sarrate R, Blesa Izquierdo J, Nejjari Akhi-Elarab F, Quevedo Casn JJ (2013) Sensor placement for leak detection and location in water distribution networks. In: Proceedings of 7th IWA specialist conference on efficient use and management of water. International Water Association (IWA), pp 1–10

Sarrate R, Blesa J, Nejjari F (2014) Clustering techniques applied to sensor placement for leak detection and location in water distribution networks. In: 2014 22nd Mediterranean conference of control and automation (MED), pp 109–114. doi:10.1109/MED.2014.6961356

Savic DA, Kapelan ZS, Jonkergouw PM (2009) Quo vadis water distribution model calibration? Urban Water J 6(1):3–22. doi:10.1080/15730620802613380

Steffelbauer D, Fuchs-Hanusch D (2015) OOPNET: an object-oriented EPANET in Python. Proc Eng 119:710–718. doi:10.1016/j.proeng.2015.08.924

Steffelbauer D, Neumayer M, Günther M, Fuchs-Hanusch D (2014) Sensor placement and leakage localization considering demand uncertainties. Proc Eng 89:1160–1167. doi:10.1016/j.proeng.2014.11.242. http://www.sciencedirect.com/science/article/pii/S1877705814023571

Wu Z, Sage P, Turtle D (2010) Pressure-dependent leak detection model and its application to a district water system. J Water Resour Plan Manag 136(1):116–128