Measurement system design for leak detection in hydraulic pressurized networks

Gaudenz Moser[1], Stephanie German Paal[2]* and Ian F.C. Smith[1]

[1]*Applied Computing and Mechanics Laboratory, School of Architecture, Civil and Environmental Engineering, Swiss Federal Institute of Technology, Lausanne, Switzerland*

[2]*Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas, USA*

Corresponding author address: Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU College Station, TX 77843-3136

Email: spaal@civil.tamu.edu

Co-authors Emails: gaudenz.moser@gmail.com, *ian.smith@epfl.ch*

Abstract

This paper compares four sensor placement strategies which differ according to their evaluation criteria. The first involves the minimization of the expected number of candidate models, and the second is based on maximizing joint entropy. The first methodology shows better results in terms of diagnostic performance.

However, the second is promising due to faster execution time. The third strategy is a combination of the first two. Finally, a fourth strategy involves consideration of the cost of the sensor placement at each location in addition to the evaluation criteria of the third strategy. The four strategies are evaluated in terms of performance, computational load and cost. Since there is only mild competition between the three criteria, a hierarchical multicriteria decision making approach is employed to identify the best sensor placement strategy. Two case studies are used for illustration. The results show that the sensor placement strategies are useful for identifying optimized sensor configurations for new configurations as well as for evaluating the performance of existing sensor configurations. Using a hierarchical multicriteria decision-making technique, the fourth sensor placement strategy satisfies all criteria well, making it the best strategy.

Keywords: measurement system design; sensor placement; hydraulic network; joint entropy; sensor placement cost.

Introduction

Water scarcity is a challenge that is faced by approximately one fifth of the world population (Watkins, 2006). In addition, one third of reporting countries lose more than 40% of clean water pumped through distribution systems due to leaks. While the need for leak detection services has been recognized by most global water utilities as part of the solution to this challenging problem, only 40% of utilities have these services (Sensus, 2012). Currently, most utilities react to leakage on an ad-hoc basis, responding to obvious leaks and bursts and repairing infrastructure as required. There is a need for more rational and systematic strategies to manage this infrastructure, and the focus of water supply management is slowly shifting towards rehabilitation of the existing networks and away from new construction (Tscheikner-Gratl et al. 2016). Advanced sensor-based diagnostic methodologies have the potential to provide useful management support. In this direction, cost-effective sensor selection and placement are two of the most essential tasks.

Measurement system design is the task of selecting good configurations of sensors with respect to their type, number and location. Prodon et al. (2010) studied the optimized placement of hydrophones. Vítkovský et al. (2003) presented an approach to determine the optimal measurement site for inverse transient analyses in pipe networks. Farley et al. (2010) developed a pressure sensor method for leak detection. Their methodology determined the sensitivity to leaks of all the possible locations and found the most sensitive locations. Developing a monitoring system for hydraulic and water quality parameters, Preis et al. (2009) combined two types of sensors to decrease cost of installation and maintenance. How the data is interpreted has already been established as an important factor for measurement system design (Laory et al., 2012; Chang et al., 2008). All of these methods involve evaluation of the performance of measurement systems using criteria that do not explicitly include how the data is interpreted for diagnosis. With the objective that measurement system design should take into account interpretation goals, Goulet and Smith (2012) proposed a methodology for sensor placement using model-falsification. While this methodology had the additional advantage to include modeling and measurement uncertainties, it did not include criteria such as cost.

In water supply networks, the cost of placing sensors may vary significantly from one position to another. For example, a flow sensor placed near an accessible part of the network may be significantly cheaper to install than a sensor placed under a main road. Since instrumentation costs for sensors are often reported as very expensive (Dorvash et al., 2014), this should be considered during the measurement system design stage. There are currently, no studies that include the effects of spatial distributions of costs for sensor placement in water supply networks.

The concept of entropy in information theory was first proposed by Shannon and Weaver (1949). Entropy is defined as a measure of information, choice and uncertainty, and several researchers have found that it is a good metric for use in sensor placement, using a greedy algorithm to add sensors where entropy is maximal (disorder is highest) (Cressie, 1992). Furthermore, several studies have been proposed for sensor placement methodologies employing entropy in combination with the model falsification framework (Raphael and Jadhav, 2015; Papadopolou et al., 2014; Robert-Nicoud et al., 2005; Kripakaran and Smith, 2009). However, no one has combined entropy with expected identifiability and these studies also do not include the impact of costs.

For many sequential sensor-placement methodologies, entropy is computed for each potential sensor location individually. Updating calculations once sensor positions have been assigned is usually not performed. This process can lead to selection of sensors with similar information. However, Robert-Nicoud et al. (2005) developed a methodology that updated calculations with each placement. Guestrin et al. (2005) proposed a new criterion based on mutual information to efficiently account for shared information. They developed a sensor placement methodology for Gaussian processes that maximizes the joint entropy. Papadopoulou et al. (2014) also proposed a sensor placement methodology using joint entropy. This technique was developed to support a methodology based on error-domain model falsification for wind behavior prediction around buildings. Further, they extended this work to take into account contributions in performance metrics and multi-criteria decision making (Papadopoulou et al., 2016). These strategies do not take into account interpretation goals such as strategies based on expected identifiability and the cost of the sensor placement. Leskovec et al. (2007) created a sensor placement algorithm which considers a cost-to-benefit ratio but not the interpretation goals of the sensor placement. They introduced a modification of the greedy

algorithm which maximizes the benefit-to-cost ratio of sensor placements to detect contaminants in water distribution networks. The configuration that was identified was sub-optimal, providing lower performance than configurations identified when considering a unit cost for all the nodes. A methodology which accounts for performance and cost in parallel needs to be developed.

When taking into account several criteria (cost, performance, etc.), a mutlicriteria decision making (MCDM) process is necessary. In the Pareto-Edgeworth-Grierson (PEG) MCDM (Grierson, 2008), a trade-off analysis technique detects compromised solutions with mutually satisfied (Pareto-optimal) competing criteria. The preference ranking organization method for enrichment evaluation (PROMETHEE) (Brans, 1982; Brans et al., 1986; Brans and Mareschal, 2005; Behzadian et al., 2010) ranks each Pareto optimal solution by way of a preference index. These techniques are most useful when criteria have similar importance and when the competition between criteria is high. An exception is the work by Adam and Smith (2007) who created a selection strategy which hierarchically reduces the Pareto set until one solution remains that satisfies all criteria well.

There is a need to provide a cost efficient, fast and high-performing strategy to evaluate and respond to damage in water-supply networks. An optimal sensor-configuration strategy which satisfies these criteria in parallel is needed. This paper presents four sensor placement strategies developed to be used with error-domain model falsification for leak detection in water supply networks. All four strategies take into account specific interpretation goals. They differ according to their evaluation criteria (EC) (the optimization objective): EC1 – expected identifiability (a data interpretation metric specific to the model falsification approach used in this work); EC2 – joint entropy; EC3 – a combination of EC1 and EC2; EC4 – a combination of EC3 and non-constant

nodal costs. All four strategies employ a forward greedy algorithm. As the literature has shown, each of these strategies alone has disadvantages; a comparison of the strategies against one another is thus of interest. Comparisons are carried out according to performance, cost and computational load in order to identify the best strategy when accounting for these significant criteria.

The next section outlines the methodologies of these strategies. This is followed by a results section where the four strategies are employed to identify an optimized sensor placement on a real network from the City of Lausanne. Also, the performance on a hypothetical configuration on the same network is evaluated. Finally, the last section discusses limitations and plans for future work.

Methodology

The initial step is to find the best location for a single sensor. Every potential location is tested, and the best one is chosen according to the objective function. In the first strategy, the expected identifiability (EC1) is used to calculate the objective function. In the second, entropy and joint entropy (EC2) are used. In the third strategy these two criteria are used in conjunction (EC3), and in the fourth, the cost, entropy and expected identifiability (EC4) are all used to calculate the function. Once the location for one sensor has been chosen, all remaining potential locations are searched in order to find the second sensor location, and again until an optimized configuration is determined for the desired number of sensors. This algorithm is called "greedy" because previously assigned sensor locations are not changed at each iteration.

The greedy algorithm can also be performed backwards. Instead of beginning with one sensor and adding the best one at each step, the process begins with all the sensors, and at each step, the worst is eliminated. In the case of the water distribution network, a forward algorithm is more applicable. In general, the number of sensors used is low in

comparison with the number of possible locations. If the sensor placement is carried out for 20 sensors in a network of 100 pipes, then the forward process will require 20 steps and the backward 80 steps. Also, a comparison of the two approaches revealed that the forward algorithm provided better results (Papadopoulou et al., 2014).

Sensor placement using expected identifiability (EC1)

In this methodology, the performance of the identification is used to evaluate the sensor configurations at each step of the greedy algorithm. The criterion to evaluate the performance of a sensor configuration is the expected identifiability.

Expected identifiability

Goulet and Smith (2012) first developed the expected identifiability metric for a sensor placement methodology based on error-domain model falsification, a model-based data-interpretation methodology. This metric helps estimate the number of candidate models that is obtainable using measurements, and it is computed by creating simulated measurements. A model instance in the initial model set is randomly selected and the uncertainties (modelling and measurement) are added to predictions using Monte Carlo sampling to produce simulated measurements as defined in Equation (1), and ten thousand simulated measurements were generated:

$$\boldsymbol{y}_s = \boldsymbol{g}(s_i) + (\boldsymbol{u}_{model} + \boldsymbol{u}_{meas}) \qquad\qquad (1)$$

where $\boldsymbol{y}_s$ = simulated measurement; $\boldsymbol{g}(s_i)$ = random model instance; $\mathbf{u}_{model}$ = modelling uncertainty; and $\mathbf{u}_{meas}$ = measurement uncertainty.

A candidate model (CM) is defined as any model instance which can explain measurement values at all measurement locations. A cumulative distribution function (CDF) representing the probabilities associated with obtaining a specific number of CMs

is constructed. This CDF is the expected identifiability. A more in depth explanation can be found in Moser et al. (2015).

EC1 strategy

This strategy employs the forward greedy algorithm, beginning with one sensor. The expected identifiability is computed for each potential sensor location. The location with the best performance (the smallest number of expected candidate models for a probability of 95%) is selected as the first sensor location. The next step is to find the second sensor location. All remaining sensor locations are tested in combination with the first sensor. The methodology searches for the sensor that gives the best results (smallest number of expected candidate models) when it is associated with the first one and the expected identifiability is calculated. This process is repeated at each step of the methodology. The sensor that is added is tested with the sensor configurations obtained in the previous step.

Sensor placement using joint entropy (EC2)

In this strategy the criterion for evaluating the sensor configuration is the joint entropy. The first sensor in the selection process is chosen as the location with maximum entropy in comparison to all other possible locations. Entropy can be used to test the quality of a sensor location by evaluating the disorder in the predictions at that point. A high disorder conveys a high quantity of information.

In order to compute the entropy at a given sensor location, a histogram representing the variation of the predictions is constructed. Then, the entropy for this sensor location is defined as:

$$H = -\sum_{k=1}^{N} p(y_k) \log\big(p(y_k)\big) \qquad (2)$$

where H = entropy; $p(y_k)$ = a probability associated with the kth interval of the histogram; and k = histogram interval. The threshold bounds computed by combining modeling and measurement uncertainties are used to compute the width of the intervals. For all predictions that are in the same interval, there is a range of measurement values that will not lead to falsification of the model instances that generated the predictions.

The first sensor location in the greedy algorithm is chosen as the location with the maximum entropy. The sensor with the highest entropy is the sensor that is able to differentiate between the predictions to the greatest extent. After this, the sensor configurations selected at each sequential step of the greedy algorithm are the configurations with the highest joint entropy value, where the joint entropy is calculated for the first sensor and every sensor after that according to:

$$H(y)_{i,j} = -\sum_{k=1}^{N_K} \sum_{l=1}^{N_L} p\big((y_k)_i, (y_l)_j\big) log_2 \Big(p\big((y_k)_i, (y_l)_j\big)\Big) \qquad (3)$$

$$= H(y)_i + H(y)_j - I(y)_{i,j}$$

where $H(y)_{i,j}$ = the joint entropy between locations i and j; ; k, l = every potential sensor location; $p((y_k)_i,(y_l)_j)$ = the joint probability of location k and l; $H(y)_i$ = entropy at location i; $H(y)_j$ = entropy at location j; and $I(y)_{i,j}$ = the mutual information of i and j.

Joint entropy is the measure of the entropy associated with multiple variables. It measures the information that two or more sensors gather together. For sensor placement, the strategy is to find sensors that have the lowest amounts of mutual information. This methodology is similar to that developed by Papadopoulou et al. (2014); however, it is adapted for the uses described in this paper rather than evaluating performance of sensor configurations in wind studies.

Combined sensor placement strategies

In addition to the strategies which employ a single evaluation criteria, two other sensor placement strategies were considered in this paper: one which combines the first two evaluation criteria, and a second which also includes the cost of the various potential sensor locations. In Figure 1, a flowchart for the sensor placement methodology using combined evaluation criteria is displayed. This methodology is hierarchical. In this flowchart, N represents the total number of potential sensor locations or the number of nodes in the network, n is equivalent to the number of sensors that will be placed, k is the number of solutions the solution space is reduced to at the beginning of each iteration, and i is the iteration in the loop. For each iteration, the solution space is reduced to a predetermined number (in this case, 10), by identifying the 10 locations with highest joint entropy (first evaluation criteria). Then, those ten locations are evaluated using the second evaluation criteria: expected identifiability, in order to find the single best location. This process is repeated for all potential locations or for the number of desired sensors in the network.

Sensor placement using joint entropy and expected identifiability (EC3)

The goal in this strategy is to select sensor configurations using the joint entropy criteria (EC2), and then find the single best location from this smaller subset of locations using the methodology based on expected identifiability (EC1). At each step of the algorithm, the joint entropy criteria is used to reduce the initial number of possible sensor configurations, and then the methodology based on expected identifiability is applied on the smaller subset. For example, for one sensor, the first step is to find the ten sensor locations that are optimal regarding entropy (ten locations with maximum entropy (Eq. 2), and then the selection of the best location is found using the criteria of expected

identifiability. For the following sensor locations, this process is repeated where the ten optimal locations in the subsets are found using joint entropy (Equation (3)).

Sensor placement using joint entropy, cost and expected identifiability (EC4)

An important aspect that engineers consider when selecting optimal sensor configurations is the cost of installing sensors. When every node has equal cost (i.e., unit cost), a commonly used heuristic is the basic greedy algorithm (Leskovec et al., 2007) which has been used in the first three strategies:

$$s_k = \underset{s \in v \backslash A_{k-1}}{\mathrm{argmax}} R(A_{k-1} \cup \{s\}) - R(A_{k-1}) \quad (4)$$

where s_k = the location which maximizes the marginal gain; R = the reward (or benefit) placement, A for all locations, s.

When the costs of the nodes are non-constant Equation (4) must be modified in order to take cost into account. As mentioned in the Introduction, Leskovec et al. (2007) introduced a modification of the algorithm which maximizes the benefit-to-cost ratio in the following manner:

$$s_k = \underset{s \in v \backslash A_{k-1}}{\mathrm{argmax}} \frac{R(A_{k-1} \cup \{s\}) - R(A_{k-1})}{c(s)} \quad (5)$$

where $c(s)$ = the cost function; and s_k = the location which maximizes the marginal gain. The configuration that will be identified using Equation (5) may be sub-optimal, providing lower performance than configurations identified when considering a unit cost for all the nodes.

The fourth strategy includes a cost function according to Equation (5). The smaller subset of locations is found using a modified greedy algorithm which accounts for the

benefit-to-cost ratio for each location. In this way, the performance (entropy) and cost are considered equally. After this step, the sensor placement methodology is the same as that for EC3 (as shown in Figure 1).

Results

In order to assess the four sensor placement strategies, two studies have been carried out on real water networks. The first study investigates performance of the sensor placement strategies on a new network while the second investigates the performance of an existing sensor configuration. In both cases, simulated measurements are generated through a numerical model of the network. As described in Equation (1), simulated measurements are obtained by randomly taking a model instance in the initial model set and adding the combined uncertainties (measurement and modeling) to the predictions. Before the two case studies are discussed, the sources of uncertainty present in these networks and studies are introduced.

Sources of uncertainty in water distribution networks

Although the principal sources of uncertainty in networks that are reported in the literature are pipe roughness and nodal water demand, all the parameters of the model are considered uncertainty sources. These parameters and uncertainty sources are discussed in this section and presented in Table 1. Many have been named previously by Hutton et al. (2004). Pipe characteristics, such as the effective diameter and roughness coefficient are sources of uncertainty. With time, the diameter reduces because of encrusted material due to precipitation of calcium carbonate and oxidation in the case of iron pipes (Hutton et al. 2004). For the same reason, pipe roughness increases with time (Kang et al. 2009).

Water distribution networks are demand-driven systems. The temporal demand at the nodes is often due to random water use; therefore, these two parameters contribute to

the overall uncertainty associated with the system. Pumps, valves and tanks are additional sources of uncertainty. A pump is simulated using a pump curve which describes its performance and is usually given by the pump manufacturer. However, in practice the efficiency often differs from the specifications. Tanks are characterized by minimum and maximum capacity. All these parameters can be considered addition sources of uncertainty (Rossman 2000). The minor loss coefficients used to compute the head loss due to turbulence at bends and fittings can also be considered. In addition to parametric uncertainties, measurement and modelling uncertainties are also taken into account. Measurement uncertainty is due to the inaccuracy of sensors, and modelling uncertainty is due to assumptions, omissions and simplifications associated with the formulation of the model.

Configuration of a new network

For this study, the sensor placement strategies are carried out on a simplified version of the water supply network from the City of Lausanne (Figure 2). The network is reduced by removing all extension pipes and most of the pipes in series (Moser et al., 2015). The resulting network is made up of 123 pipes and 94 nodes. The pump, reservoir and tank locations are marked on the figure. On this network, all four evaluation criteria are assessed in terms of computational time and performance. The locations of 30 sensors are identified using each strategy.

Comparison of EC1 and EC2

In terms of computational time, the joint entropy strategy is advantageous. The joint entropy methodology requires only 13 minutes and eight seconds to find the location of the 30 sensors while the methodology based on the number of expected candidate models requires six hours and 48 minutes. Therefore, the joint entropy strategy (EC2) takes

approximately 2% of the time that the expected identifiability strategy (EC1) takes when carried out on the same computer.

Figure 3 shows the results of the sensor placement in terms of performance. These two curves give the number of candidate models that are expected for a probability of 95% for each of the two strategies. The number of measurements is plotted on the horizontal axis (ranging from one to 30), and the expected number of candidate models is plotted on the vertical axis. The grey curve is for the methodology which uses joint entropy as the optimization parameter, and the black curve is for the methodology based on optimization of the expected number of candidate models. In terms of performance, the results are better for the methodology based on optimization of the expected number of candidate models (black curve). The sensor methodology using the expected identifiability leads to better sensor configurations, outperforming that using joint entropy.

Sensor placement using joint entropy and expected identifiability (EC3)

The third strategy was carried out on the same network (Figure 2) with k = 10 (for the subset reduction step). The results presented in Section 3.1.1 show that each of the first two strategies have advantages. The greedy algorithm associated with expected identifiability leads to higher performing sensor configurations, and the greedy algorithm associated with joint entropy is faster. Therefore, the methodology that combines these two evaluation criteria benefit from each of these advantages.

Figure 4 depicts the results of the combined sensor placement strategy in terms of performance. These two curves give the number of candidate models that are expected for a probability of 95% for the first methodology based on optimization of the expected number of candidate models (in black) and the combined methodology (in grey). The number of measurements is plotted on the horizontal axis (ranging from one to 30), and

the expected number of candidate models is plotted on the vertical axis. The two strategies found the same location for the first sensor. The results in terms of performance are similar except between sensor numbers nine through 13. This area where the two strategies differ most is due to the size of the reduced solution space (k = 10). There is no practical difference between the two curves when considering only performance. The advantage of the combined methodology is the decrease in computational time.

In Figure 5, the two 3D bar plots show the performance of the sensor placement obtained with the combined methodology for leak intensities varying from 20 to 500 l/min and a probability of 95% (left) and 75% (right). As in the previous studies, the number of sensors is varied from one to 30. These figures show that for a given number of sensors, when the leak intensity is higher than 200 l/min, the performance remains constant. In Figure 6, the graphs in Figure 5 have been magnified in order to illustrate more precisely the evolution of the performance for a small leak (from 10 to 200 l/min). These graphs provide useful decision-support for network managers when choosing the number of sensors to place throughout a network.

Sensor placement using joint entropy, cost and expected identifiability (EC4)

The fourth strategy was carried out on the same network (Figure 2) with k = 10 (for the subset reduction step). The cost function (Equation (5)) was defined by assigning a random cost value from 1-100 (minimum value = 1, maximum value = 100) across all the pipes in the reduced network. The cost function was randomly generated in this way a total of ten times and the average of the ten runs was used for the comparisons. Figure 7 illustrates the results of the two combined sensor placement strategies in terms of performance. These two curves give the number of candidate models that are expected for a probability of 95% for the third methodology based on optimization of the expected number of candidate models and joint entropy (in black) and the average of the ten runs

for the fourth methodology (in grey). The standard deviations for the fourth strategy are marked on the plot as well. Due to the nature of the computation, the scatter gradually decreases as the number of sensor locations increases. The number of measurements is plotted on the horizontal axis (ranging from one to 30), and the expected number of candidate models is plotted on the vertical axis. The results in terms of performance are similar except for the first sensor. This shows that the performance is not greatly affected by the additional criteria of cost. These results show that the low-cost sensor configuration performs adequately. The results in terms of computational time are relatively identical to that for the third strategy as the process is the same, only the value used in the argument differs (Equation (5)).

In Figure 8, the simplified version of the water supply network from the City of Lausanne from Figure 2 is shown with a configuration of six sensors determined using the sensor placement methodology which also considers the cost of each location. The optimal sensor locations are represented with boxes. The pump, reservoir and tank locations are all marked on the figure as well as the pipes and nodes (junctions). The performance of the six-sensor configuration shown in Figure 8 is plotted (in grey) in comparison to that for the six-sensor configuration obtained using the third strategy (in black) in Figure 9. The number of candidate models (the number of nodes) is plotted on the horizontal axis, while the cumulative probability is plotted on the vertical axis.

Comparison of sensor placement strategies

A summary of the advantages and disadvantages of the four strategies is shown in Table 2. In the first row of Table 2, the computational time is shown for each strategy for the 30-sensor configuration. Since the Pareto front between computational load and the other two criteria is so steep, the tradeoff position is very clear. There is very little increase in cost or performance for a significant reduction in computational time. For this reason, as

long as the computational load is below a certain limit (e.g., one hour), it is evident that there is no longer a distinct tradeoff. Hence, we can rule out EC1 since EC2 through EC4 have computational loads that are within 2-3% of the time required for EC1 (the first column can be disregarded for the rest of the table).

By ruling out EC1, the problem becomes two-dimensional in terms of cost and diagnostic performance. The diagnostic performance is evaluated as the number of candidate models that are expected using a given sensor configuration. This value is obtained by testing the sensor configuration through error-domain model falsification with a high number of simulated measurements (more details can be found in Section 2.1.1). The best sensor configuration in terms of performance is the configuration with the smallest number of expected candidate models. In the last rows of Table 2, the cost and performance associated with strategies EC2-EC4 are shown for a 30-sensor configuration and a six-sensor configuration. Again, there is very little loss in performance (percentage of expected candidate models) when decreasing the cost (the Pareto front between the cost and performance is very steep). Therefore, we can effectively choose to minimize the cost, choosing strategy EC4 as the best solution. This hierarchical procedure is effective here since the three criteria are not highly competitive. EC4 is thus, a solution that satisfies all criteria, making it the best sensor placement strategy when considering cost, performance and computational load.

Evaluation of existing sensor configurations

The second study investigates the performance of an existing sensor configuration. A configuration of six sensors was chosen for the water supply network of the City of Lausanne to represent an existing configuration. This configuration and the optimal six-sensor configuration determined using EC3 are shown on Figure 10. The optimal sensor locations are represented with boxes, and the existing sensor locations are

shown using four-pointed stars. The pump, reservoir and tank locations are all marked on the figure as well as the pipes and nodes (junctions).In order to test the performance of the initial configuration, the sensor placement strategy was used. The six monitored locations are considered to be potential sensor locations. The sensor placement methodology (EC1) is applied to this set of six locations. This provides a classification of the importance of each sensor in the initial network in addition to a way of comparing the performance of the two configurations.

In Figure 11, the performance of the existing sensor configuration on the City of Lausanne network is shown using the sensor placement algorithm. The expected number of candidate models (the number of identified nodes) is plotted on the horizontal axis, while the cumulative probability is plotted on the vertical axis. The performance of the existing configuration (in grey) is compared with that of the optimal configuration (in black) found using EC3. The plot shows the distinction between the two configurations and can be used to aide decision makers as to whether or not to take action (move sensors). For example, if the desired level of probability 50%, moving the sensors will not add much to the performance. However, for a probability of 80-95%, it appears more beneficial to consider changing the sensor configuration.

Figure 12 displays the results of the sensor configurations in terms of number of expected candidate models (performance) when it is evaluated using the expected identifiability. These two curves give the number of candidate models that are expected for a probability of 95% for the optimal sensor configuration (EC3, in black) and the existing configuration (in grey). The number of measurements is plotted on the horizontal axis (ranging from one to six), and the expected number of candidate models is plotted on the vertical axis.

Evaluating an existing sensor configuration using a greedy sensor placement methodology gives a classification of the importance of each sensor. Such a tool is helpful when deciding which sensors could be moved in order to improve an existing configuration. In this example, the graph shows that sensors one and two could be moved in order to increase diagnostic performance. By moving the first and second sensors in the existing configuration to the optimal locations, the gain in performance is significant (as shown in plot in Figure 12).

Limitations and future work

The greedy algorithm used in the four sensor placement strategies does not necessarily lead to a global optimum. Additionally, the success of sensor placement depends on the model, the model parameters and the estimation of the uncertainties. Thus, if the model is weak, the sensor placement configuration may be sub-optimal. Finally, for the network from the city of Lausanne, the trade-off position between the three criteria (cost, performance and computational load) was very clear. This may not be the case with other networks since the topological configuration of the network influences the importance of the criteria.

The cost function used in this study is a randomly distributed function across each pipe in the network. Additional work should be carried out on similar networks with more realistic cost functions, ranking locations with difficult access or in heavily populated areas higher in terms of cost. In addition, regarding sensor configuration strategies EC3 and EC4, the value of k that is used to reduce the set of potential sensor locations to a smaller subset was chosen to be 10 in this case. Further work will involve varying this value and an analysis will focus on the impact that it has on overall performance.

Summary and conclusions

In order to rapidly evaluate the performance and respond to damage in water-supply networks, a cost efficient, fast and high-performing sensor configuration must first be established. This paper presented four strategies for sensor placement in water supply networks and a comparison of these strategies according to these three criteria. The analysis of the results leads to the following observations and conclusions. All four sensor placement strategies presented in this paper are suitable for sensor placement in water distribution networks. Their performance depends on the evaluation criteria that are employed.

The methodology that is based on the criterion where the expected number of candidate models is minimized at 95% probability, provides sensor configurations that result in the fewest number of candidate leak scenarios. However, computation cost is high. The second methodology uses the criterion where joint entropy is maximized. While this is faster than the first strategy, the configurations result in more candidate leak scenarios than the first strategy. The third strategy involves combining these evaluation criteria, and provides a sensor placement methodology which mixes the advantages of the first two strategies. Nevertheless, there are tradeoffs involving search space reduction, computation load and candidate leak scenario size. The fourth strategy takes into consideration non-constant nodal costs for the pipes in the network. Considering the benefit-to-cost ratio leads to a sensor configuration that is similar to that where the performance is the sole optimization objective.

The sensor placement strategies created and presented in this paper are useful for identifying optimized sensor configurations for new sensor locations as well as for evaluating the performance of existing sensor configurations. Using a hierarchical

MCDM technique, sensor placement strategy EC4 satisfies all three criteria well, making it the best strategy.

Acknowledgments

References

Adam, B. & Smith, I.F.C. (2007). Tensegrity Active Control: Multi-objective Approach. *Journal of Computing in Civil Engineering*, *21*(1), 3-10.

Behzadian, M., Kazemzadeh, R.B., Akbadvi, A. & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *Eur. J. Oper. Res., 2010. 200*(1), 198-215.

Brans, J.-P. (1982). L'ingénierie de la decision:Elaboration d'instruments d'aide à la decision. La méthode PROMETHEE. *Laide a la decision: Nature, instruments et perspectives davenir,* 183-214.

Brans, J.-P., Vincke, P. & Mareschal, B. (1986). How to select and how to rank projects: The PROMETHEE method. *European Journal of Operational Research, 24*(2), 228-238.

Brans, J.-P. & Mareschal B. (2005). PROMETHEE methods. *Multiple criteria decision analysis: state of the art surveys,* 163-186.

Chang, S.-P., Kim, S., Lee, J. & Bae, I. (2008). Health monitoring system of a self-anchored suspension bridge (planning, design and installation/operation). *Journal of Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance, 4*(3), 193-205.

Cressie, N. (2015). Statistics for spatial data. *Terra Nova, 4*(5), 613-617.

Dorvash, S., Pakzad, S., Naito, C. & Hodgson, I. (2014). Application of state-of-the-art in measurement and data analysis techniques for vibration evaluation of a tall building. *Journal of Structure and Infrastructure Engineering: Maintenance, Management, Life-Cycle Design and Performance, 10*(5), 654-669.

Farley, B., Mounce, S. & Boxall, J. (2010). Field testing of an optimal sensor placement methodology for event detection in an urban water distribution network. *Urban Water Journal, 7*(6),345-356.

Goulet, J. & Smith, I.F.C. (2012). Performance-Driven Measurement System Design for Structural Identification. *Journal of Computing in Civil Engineering, 27*(4), 427-436.

Grierson, D.E. (2008). Pareto multi-criteria decision making. *Advanced Engineering Informatics, 22*(3), 371-384.

Guestrin, C., Krause, A. & Singh, A.P. (2005). *Near-optimal sensor placements in gaussian processes*. Presented in the 22nd international conference on Machine learning, Bonn, Germany.

Hutton, C.J., Vamvakeridou-Lyroudia, L.S., Kapelan Z. & Savic, D.A. (2004). Uncertainty Quantification and Reduction in Urban Water Systems (UWS) Modelling: Evaluation. *Environment, 33*(2), 1-15.

Kang, D., Prasha M. & Lansey, K. (2009). Approximate methods for uncertainty analysis of water distribution systems. *Urban Water Journal, 6*(3), 233-249.

Kripakaran, P. & Smith, I.F.C. (2009). Configuring and enhancing measurement systems for damage identification. *Advanced Engineering Informatics, 23*(4), 424-432.

Laory, I., Ali Hadj, N., Trinh, T. & Smith, I.F.C. (2012). Measurement System Configuration for Damage Identification of Continuously Monitored Structures. *Journal of Bridge Engineering, 17,* 857-866.

Leskovec, J., Krause, A. Guestrin, C., Faloutsos, C., VanBriesen, J. & Glance, N. (2007). Cost-effective outbreak detection in networks. *Proceedings of International conference on knowledge discovery and data mining*, 420-429.

Moser, G., Paal, S. G., & Smith, I. F. C. (2015). Performance comparison of reduced models for leak detection in water distribution networks. *Advanced Engineering Informatics, 29*(3), 714-726.

Papadopoulou, M., Raphael, B., Smith, I.F.C. & Sekhar, C. (2016). Evaluating predictive performance of sensor configurations in wind studies around buildings. *Advanced Engineering Informatics, 30*(2), 127-142.

Papadopoulou, M., Raphael, B., Smith, I.F.C. & Sekhar, C. (2014). Hierarchical Sensor Placement Using Joint Entropy and the Effect of Modeling Error. *Entropy, 16*(9), 5078-5101.

Preis, A., Whittle, A. & Ostfeld, A. (2009). *Multi-objective Sensor Network Placement Model for Integrated Monitoring of Hydraulic and Water Quality Parameters.* Paper presented at the meeting of World City Water Forum, Incheon, Korea.

Prodon, A., DeNegre, S. & Liebling, T.M. (2010). Locating leak detecting sensors in a water distribution network by solving prize-collecting Steiner arborescence problems. *Mathematical Programming, 124*(1), 119-141.

Raphael, B. & Jadhav, K.S. (2015). Sensor placement for structural monitoring of transmission line towers. *Frontiers in Built Environment, 1*(24). doi: 10.3389/fbuil.2015.00024

Robert-Nicoud, Y., Raphael, B. & Smith, I.F.C. (2005). Configuration of measurement systems using Shannon's entropy function. *Computers & Structures, 83*(8), 599-612.

Rossman, L.A. (2000). *EPANET 2: users manual*. Cincinati, OH: US Environmental Protection Agency.

Sensus. (2012). *Water 20/20: Bringing smart water networks into focus*. Technical report.

Shannon, C.E. & Weaver, W. *The mathematical theory of communication*. 1949: University of Illinois press.

Tscheikner-Gratl, F., Sitzenfrei, R., Rauch, W. & Kleidorfer, M. (2016). Enhancement of limited water supply network data for deterioration modelling and determination of rehabilitation rate. *Structure and Infrastructure Engineering, 12*(3), 366-380.

Vitkovsky, J.P., Liggett, J.A. & Lambert, M.F. (2003). Optimal measurement site locations for inverse transient analysis in pipe networks. *Journal of Water Resources Planning and Management, 129*, 480.

Watkins, K. (2006). *Human Development Report 2006-Beyond scarcity: Power, poverty and the global water crisis.* UNDP Human Development Reports.

Table 1. Uncertainties and related to and relative importance of secondary parameters (Exp is exponential distribution form, N is Gaussian and U is uniform)

Parameter	Uncertainty distribution	Relative importance for flow predictions [%]	Relative importance for pressure predictions [%]
Nodal demand [l/min]	$\sim Exp(^1/_{3.13})$	99.77	97.81
Node elevation [m]	$\sim N(0,0.015)$	5.45E-04	9.06E-02
Pipe diameter [mm]	$\sim N(0,0.75)$	2.25E-01	5.14E-01
Pipe length [m]	$\sim U(-0.03,0.07)$	3.11E-03	3.34E-02
Pipe roughness	$\sim U(0,0.015)$	5.34E-05	1.49
Tank level [m]	$\sim N0(0,0.32)$	2.80E-04	6.38E-02

Table 2. Comparison of four sensor placement strategies in terms of computational time, cost and expected percentage of CMs for the two sensor configurations. The best result in each row is given in bold.

Strategy	EC1: Expected identifiability	EC2: Joint Entropy	EC3: EC1 + EC2	EC4: EC3 + cost
Computational Time for 30-sensor configuration	06:48:00	**00:13:08**	00:19:16	00:18:32
Cost for 30-sensor configuration	--	1666	1563	**589.5**
Expected percentage of CMs for 30-sensor configuration	--	16.0	14.9	**12.9**
Cost for six-sensor configuration	--	294	316	**53.8**
Expected percentage of CMs for six-sensor configuration	--	45.7	**30.9**	37.0
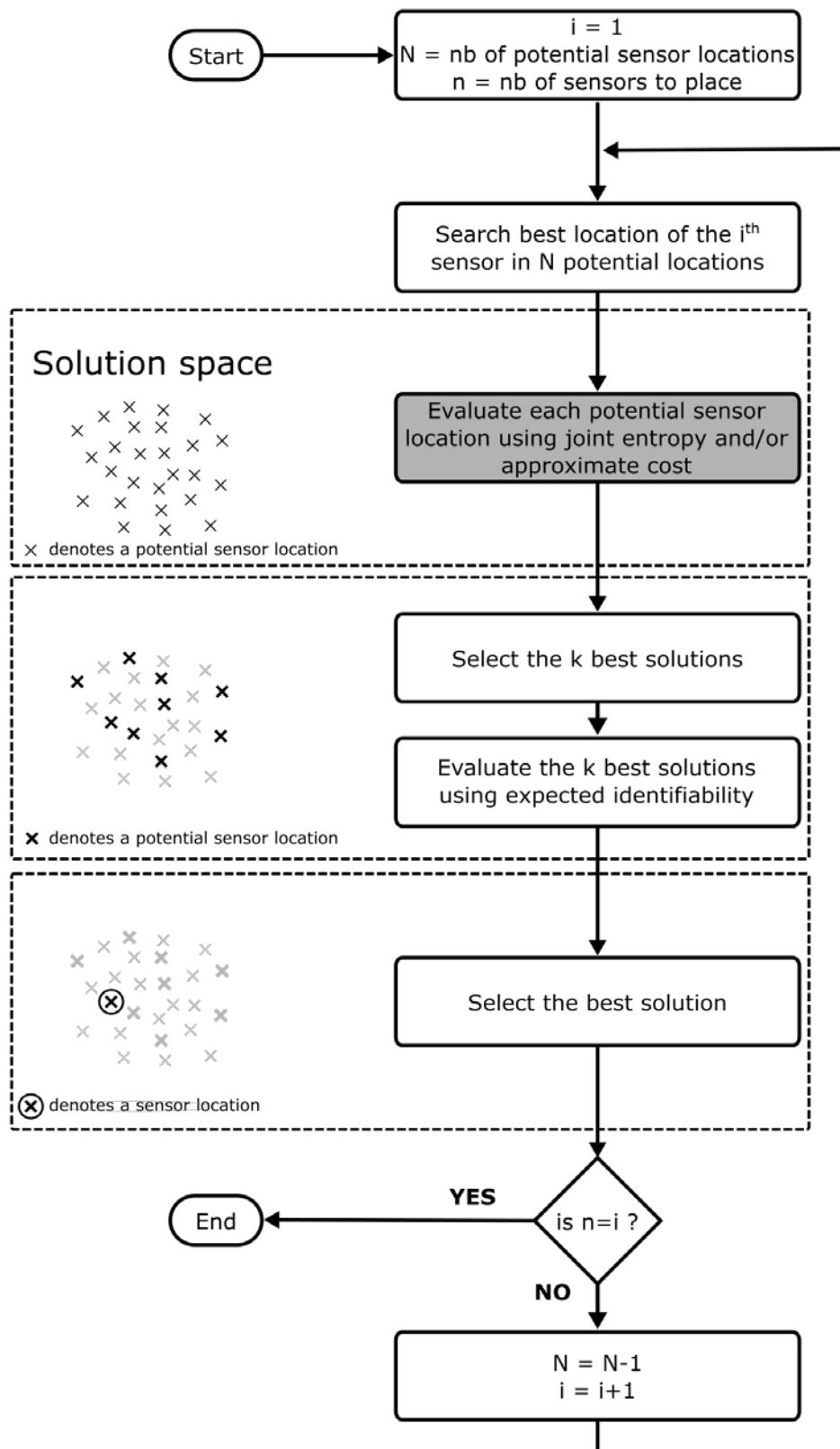
Figure 1. Flowchart for the combined sensor placement strategies (EC3 and EC4).
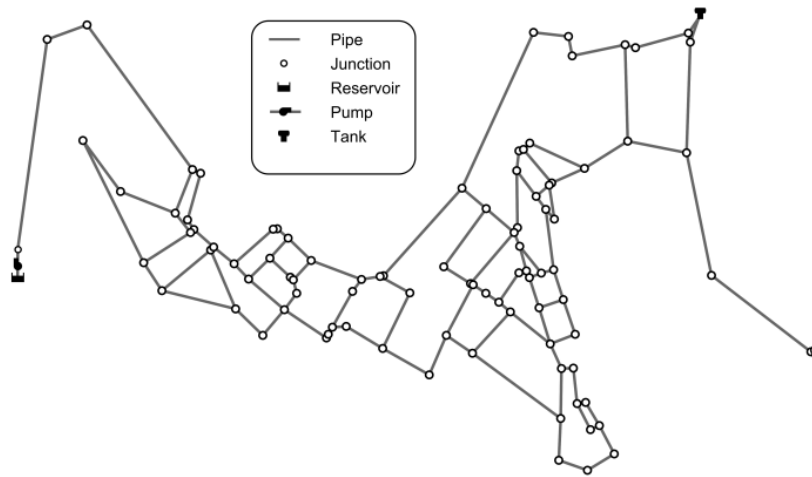
Figure 2. Reduced network used for the sensor placement.
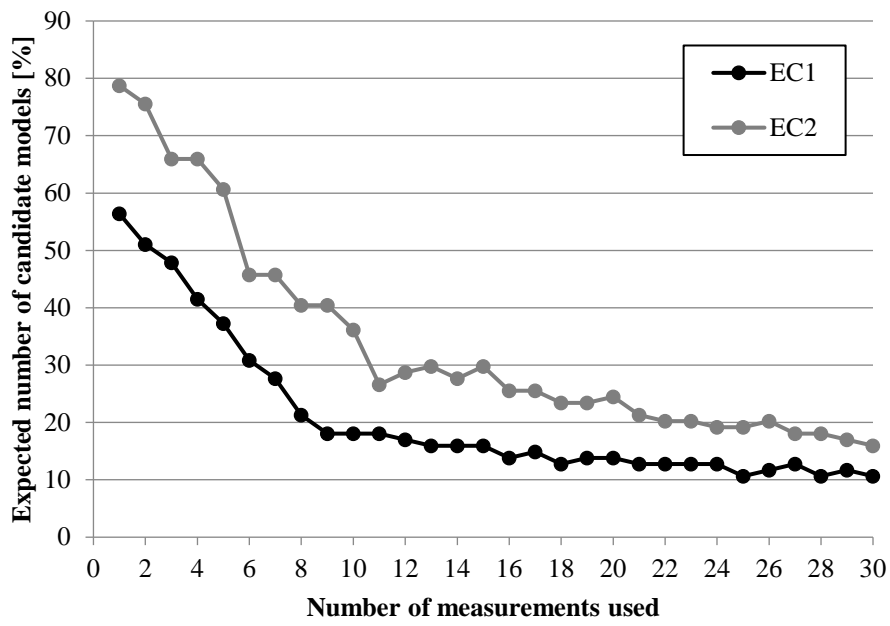
Figure 3. Sensor placement curves for the both sensor placement strategies giving the relation between the expected number of candidate models (with a 95% probability) and the number of measurements used.
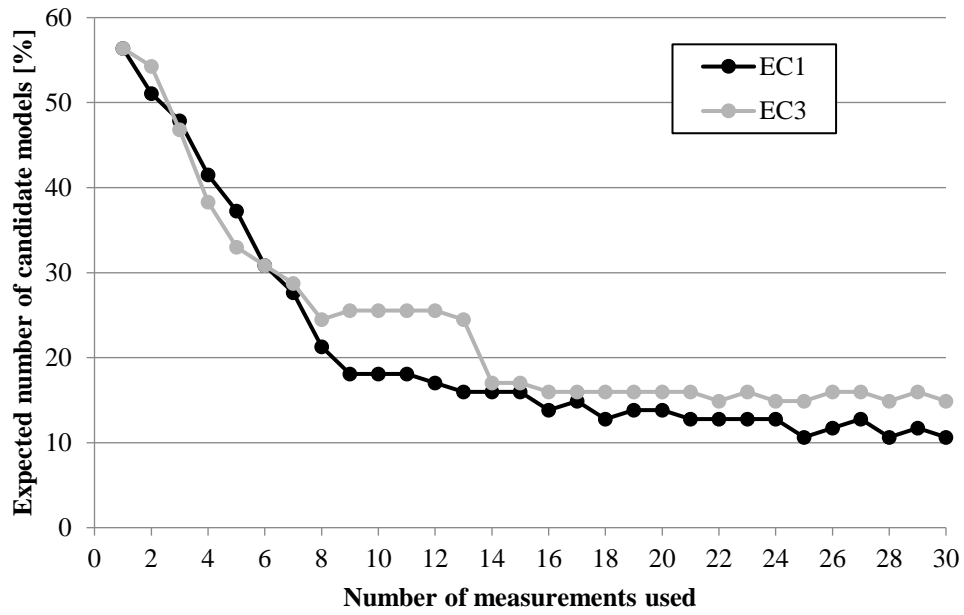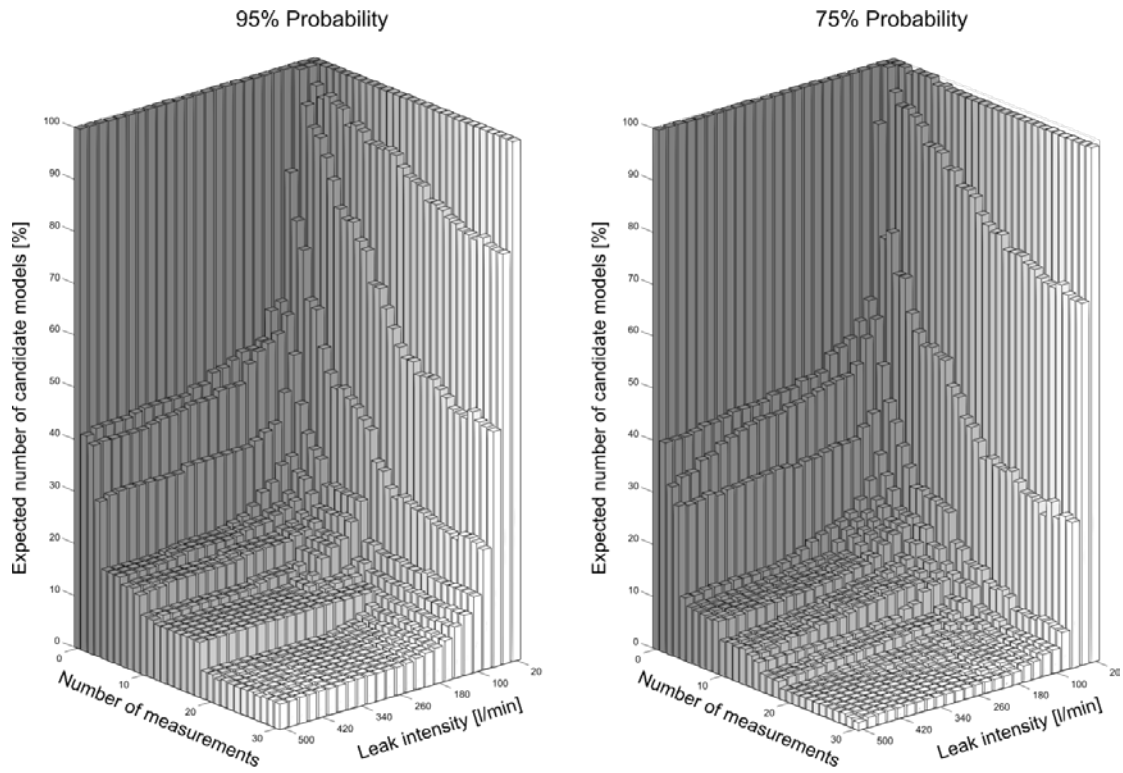
Figure 4. Sensor placement curves for the both sensor placement strategies giving the relation between the expected number of candidate models (with a 95% probability) and the number of measurements used.

Figure 5. Expected number of candidate models in function of the leak intensity (20 to 500 l/min) and the number of measurements (0 to 30) for a 95% probability (left side) and a 75% probability (right side).
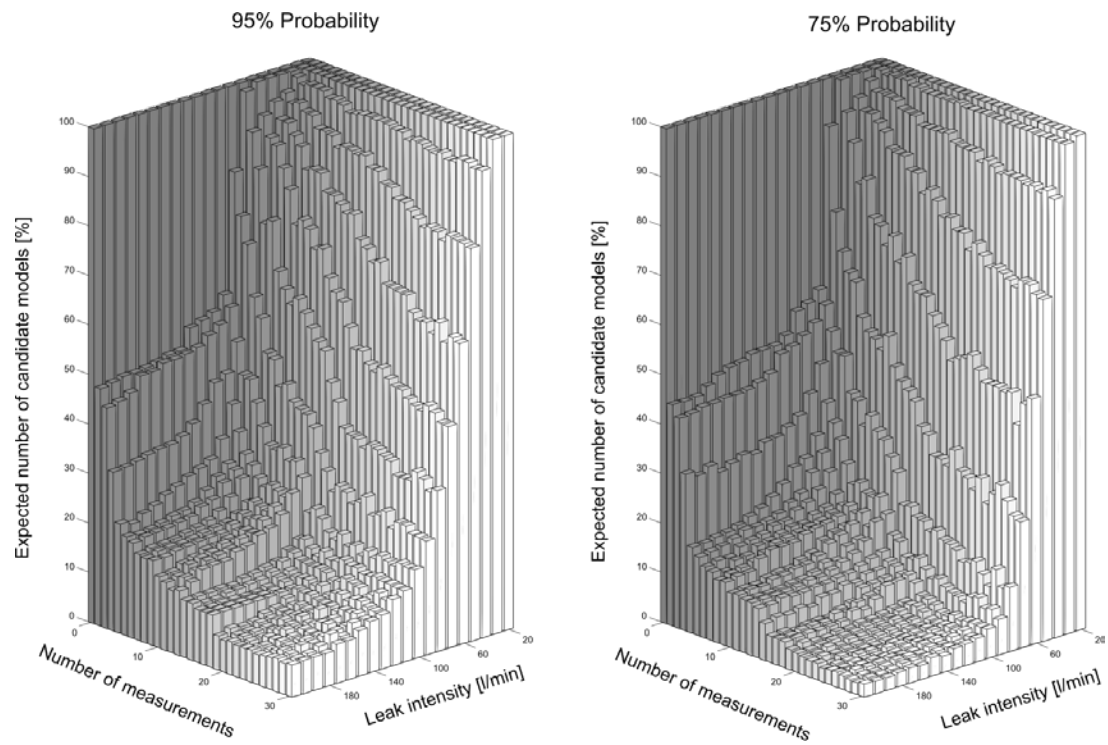
Figure 6. Expected number of candidate models in function of the leak intensity (10 to 200 l/min) and the number of measurements (0 to 30) for a 95% probability (left side) and a 75% probability (right side).
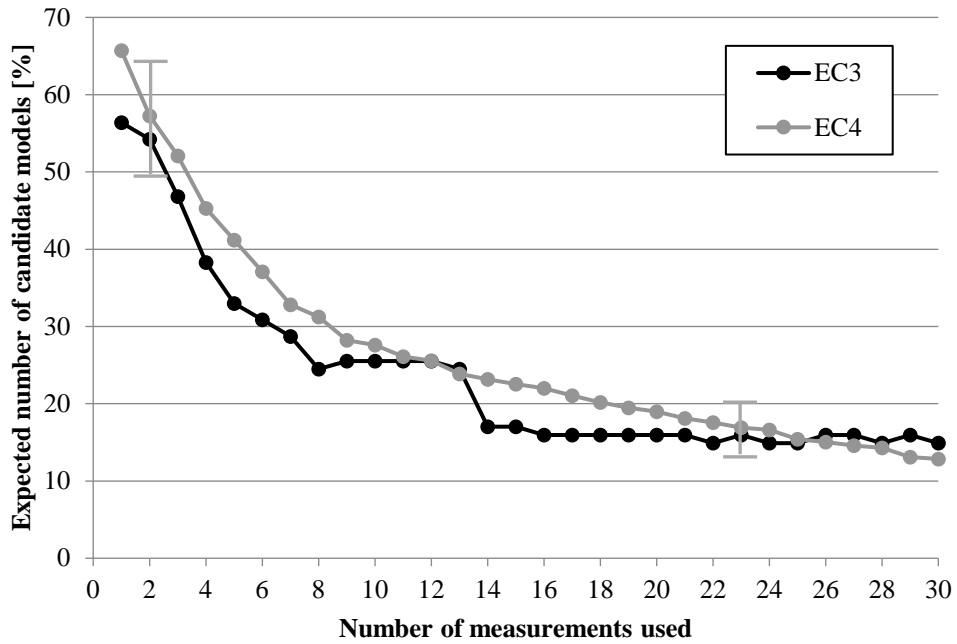
Figure 7. Sensor placement curves for the sensor placement methodology using cost as an additional evaluation criteria giving the relation between the expected number of candidate models (with a 95% probability) and the number of measurements used. For reference, the performance of the cost-based sensor placement methodology is compared with that of the combined (joint entropy and expected identifiability) methodology. The band indicators in grey indicate ±1 standard deviation for ten runs of variable sensor cost distributions.
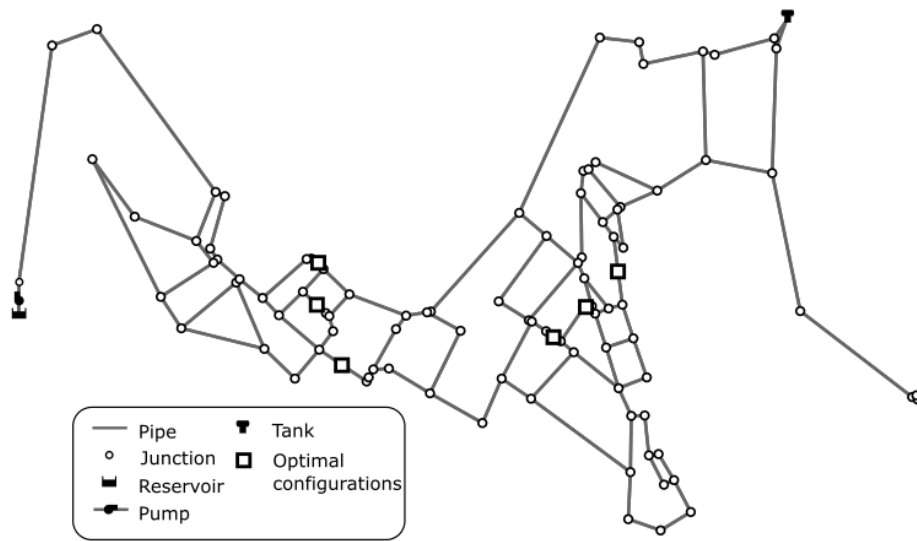
Figure 8. City of Lausanne network showing optimal sensor configuration found using cost as an additional evaluation criteria.
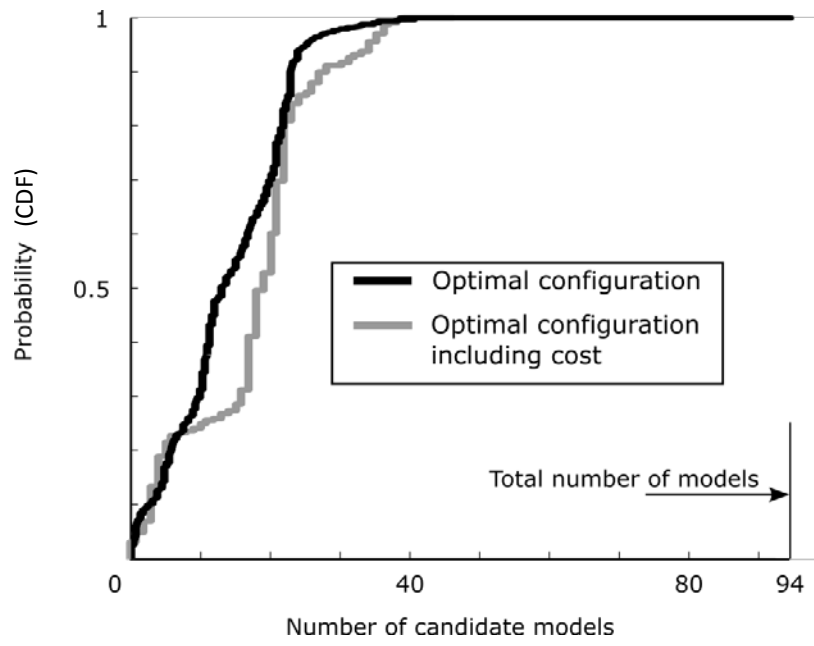
Figure 9. Performance of the proposed sensor configuration shown in Figure 8.
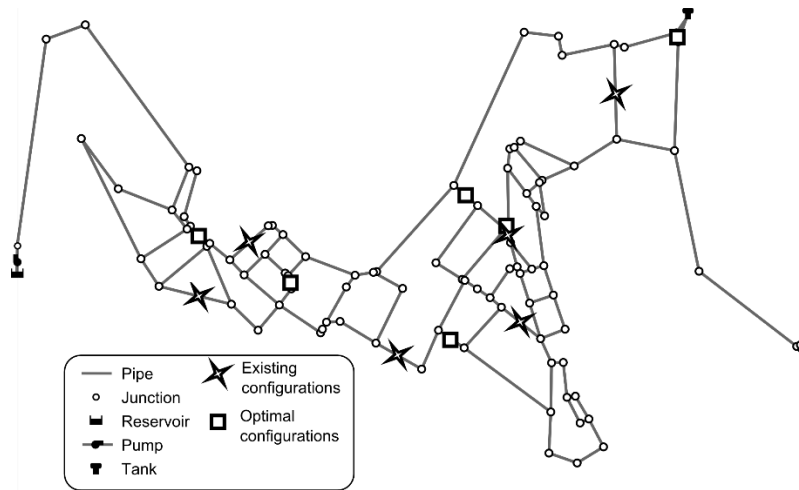
Figure 10. City of Lausanne network showing an optimal sensor configuration (EC3) and an existing sensor configuration.
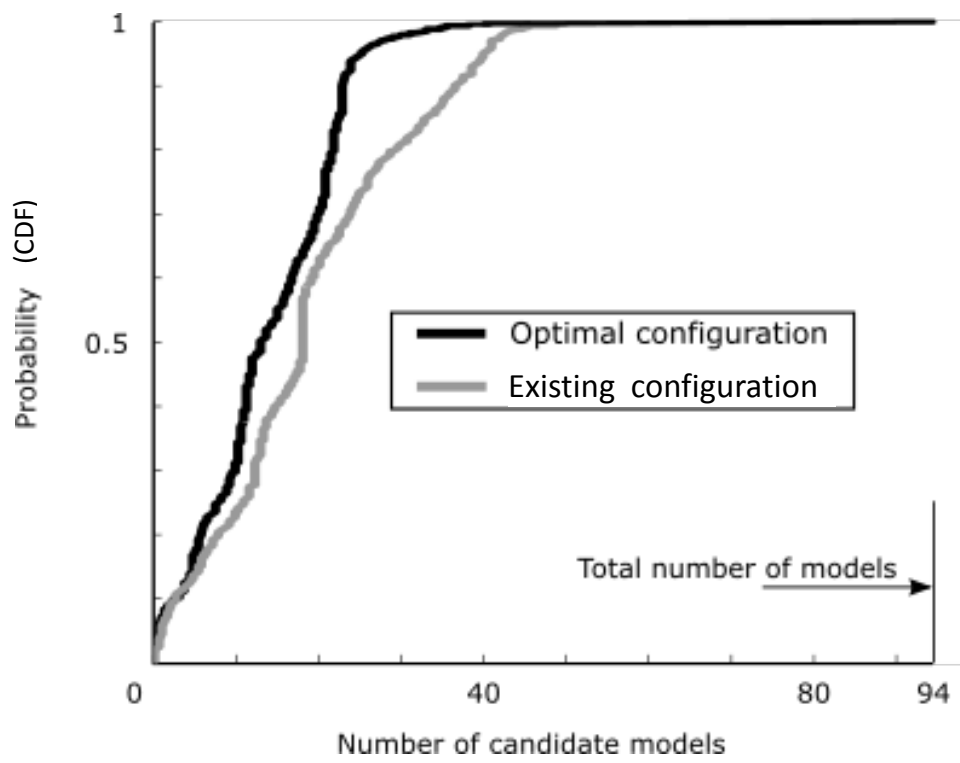
Figure 11. Performance of the existing sensor configuration compared with an optimized six-sensor configuration (EC3) as shown in Figure 10.
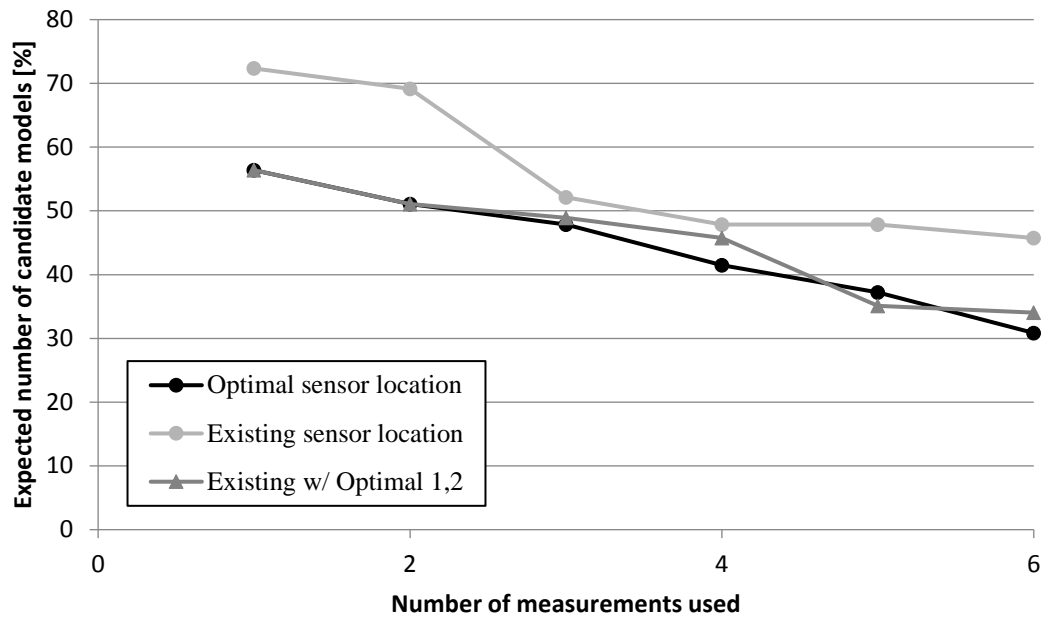
Figure 12. Sensor placement curves giving the relation between the expected number of candidate models (with a 95% probability) and the number of measurements used for the optimal sensor configuration (EC3, in black) and the representative existing sensor configuration (in grey).