Moser, G., Paal, S.G. and Smith, I.F.C. "Leak detection of water supply networks using error-domain model falsification"
Journal of Computing in Civil Engineering, No 2, Vol 32, 2018, pp 04017077-1-18, http://cedb.asce.org Copyright ASCE

Leak detection of water supply networks using error-domain model falsification

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

(1) Applied Computing and Mechanics Laboratory (IMAC)
School of Architecture, Civil and Environmental Engineering (ENAC)
Swiss Federal Institute of Technology (EPFL)
Lausanne, Switzerland
(2) Zachry Department of Civil Engineering, College of Engineering
Texas A&M University
College Station, Texas, USA

Abstract

Pressurized fresh-water fluid-distribution networks are key strategic infrastructure elements. On average, 20% of water is lost by way of leaks around the world. This illustrates the need for more efficient management of pressurized fluid-distribution networks. This paper presents a system identification methodology known as error-domain model falsification adapted for performance assessment of water distribution networks and more specifically, to detect leak regions in these networks. In addition, a methodology to approximate the demand at nodes in water-supply networks is presented and a methodology for estimating uncertainties through experimentation is described. The use of error-domain model falsification for practical use in water distribution networks shows great potential. Finally, two case studies are presented. The first case study is from the water-distribution network of the city of Lausanne. An experimental campaign was carried out on this network to simulate leaks by opening hydrants. The second case study is from a water-distribution network of the commune of Bagnes, and a leak scenario was evaluated. These two case studies illustrate, using full-scale measurements, the potential of error-domain model falsification for the performance assessment of water-distribution networks.

Introduction

Annually, 184 billion USD are spent on clean water supply worldwide. However, collectively, water utilities lose an estimated 9.6 billion USD each year due to water leakage (Sensus 2012). Water supply networks lose an average of 20% of their water supply (Figure 1). The Sensus report also includes an estimate that if leaks were reduced by 5% and pipe bursts by 10%, utilities could save up to 4.6 billion USD.

Currently, most utilities react to leakage on an ad-hoc basis, responding to leaks and bursts and repairing infrastructure only as required by leakage events. There is a need for more rational and systematic strategies for managing infrastructure. Monitoring of water-supply networks could support this need and sensor-based diagnostic methodologies have the potential to provide enhanced management support.

Detecting leaks in water distribution networks is not a new challenge. Several studies over the past century have involved leak detection in fresh-water supply-networks. Hope (1892) studied water losses, and Babbitt et al. (1920) described examples of leak-detection methods such as visual observation and sounding through the soil with a steel rod. Water-hammer techniques and acoustic measurements, considered to be more advanced leak detection techniques, have also been developed. In addition, water loss and related costs have been highlighted for many decades (Niemeyer 1940; Johnson 1947).

Studies involving various leak-detection methodologies have continued into this century. Leak-noise correlation (Grunwell and Ratcliffe 1981; Gao et al. 2006; 2009), pig-mounted acoustic sensing (Mergelas and Henrich 2005) (a “pig” is a device used for cleaning and inspecting pipelines) and ground penetrating radar (Demirci et al. 2012) have all been studied. These techniques are not appropriate for monitoring large networks due to high costs. However, these methods are useful as complements to other methods in order to, for example, locate leaks in network segments that have already been identified to contain leaks.

An additional method, termed water balance, where the network is audited in order to examine the equality between water placed into the distribution system and water taken out was developed by Lambert and Hirner (2000). Morrison created the night flow district metered area (DMA) method (2004). In this method, the network is separated into segments and the water that flows in and out of these segments is metered. Water loss is estimated by taking these measurements when the demand is minimal, usually at night.

In addition, there are several transient-based techniques which use pressure measurements to detect leaks in water supply networks by measuring transient signals. Colombo et al. (2009) reviewed these types of methods and established three sub-categories: inverse-transient analysis (Vitkovsky et al. 2000; 2007), direct transient analysis (Whittle et al. 2010; 2013; Srirangarajan et al. 2010) and frequency-domain techniques. The accuracy of the results is affected by the uncertainties associated with these systems. Thus, many of these techniques are used on single, underground pipelines (Puust et al. 2010) rather than complex water distribution networks. An exception is the study presented by Whittle et al. (2013). However, in this case, slow leak development requires the use of other detection methods.

Another class of techniques are those based on comparisons of measurements with predictions obtained from hydraulic models. This challenge is often framed as an optimization task. The goal is to minimize the differences between predicted values from flow models and network measurements. Such techniques are often based on minimization of least-squares (Pudar and Liggett 1992; Andersen and Powell 2000). Mounce et al. (2009; 2011) used machine learning and fuzzy inference to detect leaks.

Poulakis et al. (2003) proposed a Bayesian system-identification methodology for leakage detection. Rougier (2005), Puust et al. (2006) and Barandouzi et al. (2012) also proposed Bayes-based leak detection methodologies. Romano et al. (2012; 2013; 2014) implemented Bayesian inference in a pipe burst detection framework. Hypotheses made when using traditional residual minimization or Bayesian inference techniques are usually impossible to meet due to the presence of systematic modelling errors and unknown values of induced correlations (Goulet and Smith 2013).

Leak detection is not uniquely carried out in fresh-water-distribution networks. Other pressurized fluid-distribution systems such as oil and gas pipelines may also be subject to leaks. In such cases, the consequences of a leak may be dangerous with a risk of environmental pollution. The Alaska Department of Environmental Conservation (ADEC 1999) presented a review of leak-detection technologies for crude oil-transmission pipelines. For gas pipelines, Murvay and Silea (2012) surveyed leak detection and localization techniques. The techniques presented in these references are comparable to those developed for fresh-water networks. The main difference is the presence of closely-coupled segments in the water network compared with pipelines.

Robert-Nicoud et al. (2005) developed a methodology for sequential sensor placement in water supply networks using entropy. Similarly, Goulet and Smith (2012) developed a model falsification method called error domain model falsification for bridge diagnosis. In addition, they carried out a preliminary study using this method for leak detection on a water supply network (Goulet et al. 2013). This study showed good results for leaks of 100 l/min which is not adequate for full-scale applications; practitioners are interested in detecting much smaller leaks. Moser and Smith (2015) presented a methodology for reduction of water supply networks when paired with error-domain model falsification for leak detection. The reduction process translates the network model into a simpler equivalent model (reducing the number of nodes and connections, and in turn, calculations). Further, Moser et al. (2016a,b) extended this work to electrical networks in order to provide additional case studies for leak detection in water supply networks as the two network types are analogous. In short, previous work has focused on the development of individual technical aspects of the leak detection methodology. What remains is a need to evaluate the performance of these individual aspects as a combined methodology in real-world scenarios. This paper presents the complete scheme with practical case studies that involve real measurements at various leak intensities.

Measurements are not useful unless the data that is generated can be interpreted appropriately. This paper presents a methodology that accommodates systematic uncertainties and is robust in the presence of unrecognized correlations. The objective is to provide a general diagnostic methodology – for water distribution networks, and more generally, for pressurized fluid distribution networks – that is able to locate leak regions. The methodology results in a designated area, or areas, where a leak is known to be present. The size of the region, as well as the number of regions, depends on the number of sensors that are used and the available prior knowledge of the system.

In the following section, the diagnosis methodology is described in more detail. It is then described how this methodology has been modified to support leak detection in water supply networks. The usefulness of the methodology through an application to a part of the city of Lausanne water distribution network is illustrated. In this application, fire hydrants have been opened to simulate leaks. Leak region detection, as well as an estimation of the demand and associated uncertainties is presented. Finally, a discussion of the more general impact of this work is presented.

Methodology

Over nearly twenty years, several researchers have contributed to a methodology for diagnosis called error-domain model falsification (Smith 2016). This methodology is most useful in cases where little information is available to describe the relationships between uncertainties at locations where measurements and predictions are compared. The methodology uses explicit representations of modelling and measurement uncertainty distributions at each location (Goulet and Smith 2012; Pasquier and Smith 2015). Prior knowledge is used to define bounds for the parameter values and to build sets of possible scenarios. A scenario corresponds to a set of parameter values describing the state of the system (e.g. a leak location). Scenarios are generated in order to cover the range of all possible states of the system.

As shown in Equation 1, when comparing predictions $\boldsymbol{g_i}$ from a numerical model with a measurement, $\boldsymbol{y}$, modelling uncertainty ($\boldsymbol{u}_{modelling}$) and the measurement uncertainty ($\boldsymbol{u}_{measurement}$) have to be included (Goulet and Smith 2013).

$$\boldsymbol{g_i} + \boldsymbol{u}_{modelling} = \boldsymbol{Reality} = \boldsymbol{y} + \boldsymbol{u}_{measurement} \quad (1)$$

Equation 1 can be rearranged to obtain Equation 2 in which both uncertainties are combined.

$$\boldsymbol{g_i} - \boldsymbol{y_{sim}}(\boldsymbol{t_j}) = \boldsymbol{u}_{measurement} - \boldsymbol{u}_{modelling} = \boldsymbol{u}_{combined} \quad (2)$$

Using this combined uncertainty ($\boldsymbol{u}_{combined}$), threshold bounds ($\boldsymbol{T_{low}}, \boldsymbol{T_{high}}$) are computed (by taking the 95 % interval of the probability density function). The threshold bounds are used in Equation 3 to determine if the predictions are close enough to the measurements at each measurement location.

$$\boldsymbol{T_{low}} \leq \boldsymbol{g_i} - \boldsymbol{y} \leq \boldsymbol{T_{high}} \quad (3)$$

where $g_i = g(\theta_{i1}, \ldots, \theta_{ik})$

Figure 2 shows the steps of error-domain model falsification. The first step is to define scenarios. The scenarios, $\boldsymbol{s_i}$ are defined by the parameters θ_{ik} that need to be identified. For example, in the case of leak detection, the parameters are leak locations and leak intensity. The scenarios are generated by sampling to cover all the possible behaviors of the system. Some expert input can also be useful to reduce the sampling size. They are then simulated through a numerical model of the system in order to obtain a population of model predictions (termed the

initial model set). These predictions are used for the identification by comparing them with measurements.

The distinction has to be made between physical and simulated measurements. Simulated measurements are necessary in order to estimate the performance of sensor configurations prior to physical measurement. Simulated measurements are obtained by randomly taking a model instance in the initial model set and adding the combined uncertainties (modelling and measurement) to predictions. Measurements are obtained from the sensors on the real/physical network.

For real and simulated cases, measurements are compared with predictions for each model instance. Threshold bounds, see Equation 3, are used to identify model instances that are compatible with the measurements (candidate models). If the difference between the predictions and the measurements (real or simulated) is within the bounds defined by the thresholds, then the model instance is accepted to be part of the candidate model set. Otherwise, the model instance is falsified and removed from the candidate model set.

Expected identifiability is a cumulative distribution function (CDF) that represents the probabilities associated with obtaining a specific number of candidate models. This CDF is built by testing a large number of simulated leaks on the network. For each scenario, the number of candidate scenarios is computed using the error-domain model falsification procedure presented above. This measure indicates the performance of the diagnosis and is used throughout the work presented here.

Uncertainties

The task of estimating correct values for uncertainties is challenging. Sometimes, the information can be determined experimentally (for example, for the sensors). Indications of other uncertainties can be found in the literature. Model simplification uncertainty values are more difficult to estimate. In the case of water distribution networks, estimation of the uncertainty of the demand is challenging, particularly when it is modelled using a statistical distribution and when there is limited data to justify assumptions. The performance of identification using error-domain model falsification is dependent on the estimation of uncertainties. An overestimation of uncertainties will lead to poor identification performance. In extreme cases, either all model instances are accepted because they are all within the range of the thresholds or all model instances are rejected. An incorrect estimation of the uncertainty

(or a wrong hypothesis in the modelling) can lead to an incorrect diagnosis. The methodology created here uses experimental data to obtain an estimation of the threshold values, and, thus the combined uncertainties.

The threshold bounds used in error-domain model falsification to falsify model instances are computed by combining uncertainties that are representative of modelling and measurement errors. Measurement errors are mainly due to sensor resolution since noise and sensor bias can be considered to be negligible. Modelling errors are due to model simplifications and to errors associated with model parameters. Model parameters are often not exactly known. They are based on the network-design plans and estimations. Sometimes, parameters are estimated based on hypotheses made using engineering experience.

Model parameters are categorized into two groups. The first group contains the primary parameters which characterize leak scenarios. For leak detection, these are the leak position and leak intensity, as described above. The other parameters, called secondary parameters, do not explicitly characterize the scenarios. The uncertainties associated with these parameters are included in the combined uncertainty that is used to estimate the threshold bounds.

Secondary parameters include: pipe diameter, pipe roughness, pipe length, nodal demand, node elevation, and the water level in the tank. The propagation of uncertainties of secondary parameters through the model is computed using the Monte-Carlo method. Thousands (10,000) of simulations are carried out by varying the parameters following the uncertainty distributions. An example of a choice of uncertainties is given in Table 1. Three distributions, exponential (Exp), normal (N) and uniform (U) are used. The arguments of these distributions contain defining parameters.

The demand at each node is estimated using an exponential law. The mean of the exponential law (3.13 l/min) represents the minimal demand ratio. Minimal demand ratio is the minimal consumption divided by the number of nodes. By making this hypothesis, the assumption is made that each node has the same “weight”. There is no distinction between large and small consumers; it is the simplest way of modelling the demand without any prior knowledge. For the node elevation and pipe diameter, the uncertainties are represented by a zero-mean normal distributions with standard deviations of 15 cm and 0.75 mm, respectively. These values have been estimated using engineering judgement. Pipe length uncertainty is described by a uniform distribution with minimum and maximum values taken from an ISO norm (ISO 2531 2009).

The values for the pipe roughness and the tank level have been estimated by a review of statistical variation of these data.

Using these uncertainty values, a sensitivity study has been carried out. The relative importance of each parameter is estimated using the surface response technique (Fang et al. 2005; Box and Draper 1959). The technique involves approximating the numerical model by a linear model $\mathbf{Y} \approx \mathbf{M}\boldsymbol{\beta}$ where $\mathbf{Y}$ is the vector of predicted values and $\mathbf{M}$ the model matrix built from the standardized parameters. The vector $\boldsymbol{\beta}$ is the least-square estimator of the parameter vector. Each element of this vector represented the relative importance of the associated parameter.

Results are given in Table 2. This table gives the relative importance of the uncertainties in the computation of the flow predictions and pressure predictions for each parameter. One uncertainty has a much larger impact than the others—the uncertainty on the nodal demand, with more than 99% relative importance. These results show that when the demand is not well known and an exponential law is used to model it, the uncertainty of the demand dominates the uncertainties of the other parameters. In such situations, the uncertainty of the nodal demand can be considered alone to represent the uncertainty of the secondary parameters.

Estimation of threshold values

The overall concept of this methodology is to infer threshold values using measurements that represent known events. Instead of searching unknown leaks, the goal is to start with created leaks. Such experiments can be carried out on a network by opening hydrants and controlling the outcome (flow).

When the leak intensity and its location is known with certainty, remaining sources of uncertainty can be quantified. The difference between predictions of this scenario and the measurements is computed. This difference is then taken to be the combined uncertainty of modelling and measurement errors. By repeating these operations for a range of measurements and various scenarios, knowledge of uncertainties is increased. This process is illustrated in Figure 3.

This methodology has been adapted for two desired quantities in water-supply management: (1) leak region detection; and (2) demand estimation. The two EDMF-based methodologies are described in further detail in the following sections, followed by a discussion of two case studies

in these approaches using sections of real water supply networks in the city of Lausanne and commune of Bagnes.

Leak-region detection

For leak region detection, numerical simulations are performed using the water distribution network simulation software EPANET (Rossman 2000). The goal is to develop a methodology that is capable of identifying and defining areas where a leak is located. The size of leak regions, is dependent on the number of sensors used for identification and the prior knowledge of the system.

Leak scenarios

In order to detect a leak, a set of candidate model instances is built using leak scenarios. For this study, scenarios are constructed following two hypotheses: (1) one leak occurs at a time; and, (2) leaks occur at a node in the network model. The configurations are obtained by varying leak position (the node where the leak occurs) and leak intensity (the flow going out through the node), see Figure 4. The number of scenarios is equal to the number of nodes multiplied by the number of intensities that are considered. It is not necessary to model leaks that occur at intermediate points of pipes since the presence of uncertainties means that only leak regions, which are bounded by nodes, are identified. Intermediate points do not influence the size of the leak region.

The leak scenarios are simulated using EPANET to obtain a population of model instances. The simulations are steady-state only (not transient). For each model instance, predicted values are computed for the physical quantities (flow or pressure) that are measured by sensors on the network. These predictions are then compared with measurements in the error-domain model-falsification process.

Comparison with simulated measurements

This section presents results that have been obtained through a study carried out on part of the fresh-water distribution network in Lausanne, Switzerland. This network contains 263 nodes and 295 pipes. For illustration, in this part of the study, simulated measurements are used (a later section includes two case studies where full-scale measurements are employed). Simulations of measurements and leak scenarios are performed based on the minimum water demand. Analysis of water distribution networks is generally conducted during minimum

demand hours because uncertainties related to the consumption is minimal in this time period. In this case, this minimal demand is 830 l/min. This value was then divided by the number of nodes to obtain the mean consumption for each node. For simulations, the nodal demand is described by an exponential distribution. The exponential distribution is utilized as a logical way to represent this water demand where there is a high probability to have a low consumption and a low probability to have a high consumption.

For this illustration, the number of sensors is chosen to be three, and they were placed using a greedy algorithm (Moser et al. 2016a). Figure 5 shows the results obtained for four leak scenarios. White circles indicate the demand nodes. The links between these nodes are the pipes. Squares are sensor locations on the network. The cross gives the position of the simulated leak. In each of these four examples, the leak intensity used is 100 l/min. The nodes in dark grey indicate the candidate leak scenarios, i.e. those possibilities that have not been falsified.

The four examples in Figure 5 illustrate two situations. First, in Cases (1) and (2), the number of candidate scenarios is important. The size of the regions that are defined by the candidate leak scenarios is too large to be able to satisfactorily identify the leak-region. These results show that the methodology may be useful only to falsify one side of the network. In Case (1) all candidate leak scenarios on the right side have been eliminated, and in Case (2) the entire left side is falsified. These results may be useful in practice if the methodology is combined with local leak detection techniques, such as acoustic methods. Discarding half of the leak locations divides the necessary time for searching with a local technique in half. Nevertheless, more accuracy is desirable. In Cases (3) and (4) the number of candidate leak scenarios is lower than in Cases (1) and (2). In such situations, the region defined by the candidate leak scenarios is small enough to obtain information related to the leak location.

To illustrate the utility of the expected identifiability, a first example is shown in Figure 6. In this figure, the CDF of the expected identifiability is given for two values of the global demand (the demand on the entire network): 830 l/min and 415 l/min. These results are for a leak intensity of 100 l/min and the same sensor configuration shown previously. This graph shows that there is a 95% probability to identify less than 120 candidate leak scenarios (or to falsify more than 145 leak scenarios), for the demand of 415 l/min. This means that in 95% of cases it is possible to reduce the population of candidate leak scenarios by half for a leak intensity of 100 l/min. With a 50% probability, it is possible to reduce the initial model set to less than 69 candidate leak scenarios. In comparison, for a global demand of 830 l/min, a 95% probability

results in 216 candidate models and 136 for a probability of 50%. This graph shows that the identification performance is two times better for the flow rate of 415 l/min because the expected population of candidate models is approximately equal to half when compared to that for the 830 l/min flow rate.

The second example shows the expected identifiability when using pressure sensors rather than flow sensors (Figure 7). The model and the parameters are exactly the same as before, with the same associated uncertainties (the global demand is 830 l/min). The only difference is that the pressure at the nodes are predicted instead of the flow in the pipes. In this graph the performance is given for three sensor configurations. They were obtained with a greedy algorithm that optimized the expected identifiability for the 95% probability (Moser et al. 2016a). The configurations are composed of three, five and 15 sensors to show the evolution of performance when increasing the number of sensors. For comparison, the curves of the three flow-sensor configurations are also given.

These results show that the increase of the performance when increasing the number of sensors is not very pronounced. This is especially true around the 95% probability where the increase is almost null. By comparing these results with the CDF of the three-sensor configuration, the conclusion can be made that in this specific case, measuring flow is more appropriate than measuring pressure. For three flow sensors, the performance is significantly greater than for 15 pressure sensors. The reason is that variations observed in the pressures predictions for leak scenarios are too small in comparison with values for the threshold bounds. Therefore, scenarios cannot be differentiated by pressure predictions. In another situation having lower uncertainty, pressure sensors might be used efficiently. Pressure measurements are more often used with transient models (Whittle et al. 2013).

Demand estimation

In most networks, knowledge regarding the distribution of demand is low. Nevertheless, the global demand is often known. It is computed by measuring the water that goes in and out of the network. Although there are counters for each paying consumer, these counters only record the yearly cumulate consumption; there is no information about how this distribution varies, for example, throughout a day.

The behavior of water distribution networks is governed by the demand of the consumer. For this reason, it is important to increase knowledge regarding demand parameters. This can be

achieved using error-domain model falsification. The methodology developed for this purpose is illustrated using simulated measurements which are retrieved in the manner described in the methodology section.

The methodology is essentially the same as that used for leak detection. The difference is in the identification objectives. The model instances that are used to compute the predictions are different. Instead of leak scenarios, they are built from demand scenarios.

Demand scenarios

In order to estimate the demand, the model instances are built using the nodal demand as the primary parameter. For this reason, the scenarios are referenced as demand scenarios. Each demand scenario represents a specific demand configuration of the system. In the same way as for the leak scenarios, assumptions have to be made in order to limit the sampling size. It is not possible to sample the demand at each node and then to build all the permutations such that all possible behaviors are explicitly considered. The number of scenarios is equal to the number of nodes as based with an exponent of the number of samples for each node. For example, if three sampling values are tested at each node, then, for a total of 265 nodes, the number of scenarios is greater than 18 million (265^3).

In order to achieve consistent sampling, the number of nodes where the demand is estimated needs to be reduced. This is achieved by using a network reduction technique (Moser et al. 2015). The reduction is shown in Figure 8. Only nodes at the three monitored pipes and the nodes connected to the tank and reservoir are included in the reduced network. By performing the sampling only at the nodes connected to the monitored pipes, the task of estimating the demand is reduced to six nodes. This is the upper limit for sampling. For eight sampling values at each node, the number of scenarios is greater than 1 million.

Generally, the global demand is known or can be estimated. That information can be used for the sampling by adding a constraint that fixes the sum of the six nodal demands to be equal to the global demand. If the global demand is not precisely known, then sampling can be repeated for various global-demand values in the same way that it has been done for leak intensity. For this reason, this parameter is termed demand intensity. In order to build scenarios, the demand intensity is divided by the number of nodes in the non-reduced network. Then, the subdivided demands equivalent to the mean nodal demand, are distributed randomly to each of the six nodes in the reduced network. By doing this, the sum will always remain equal to the demand

intensity. In the next section, results using simulated measurements (retrieved in manner presented in the methodology section) to estimate the number of scenarios that need to be built following this procedure in order to have a consistent sampling are presented.

Comparison with simulated measurements

The EPANET model is run with varying demand. The demand of each node is varied following a time pattern that was built randomly. Figure 9 shows the pattern for Node 1. The horizontal axis gives the time in hours, and the vertical axis gives the nodal demand in l/min. The time step for the pattern is five minutes. The black curve represents the demand for Node 1, and the grey curve represents the total demand (i.e., the sum of the six nodal demands). The demand pattern for the other nodes is not shown.

The model of the network is then simulated with these patterns and according to Equation 4.

$$\boldsymbol{y_{sim}(t_j)} = \boldsymbol{h}\left(\theta_1(t_j), \ldots, \theta_k(t_j)\right) + \boldsymbol{u}_{combined} \qquad (4)$$

Simulated measurements are obtained by taking the flow value prediction. In the same way that it has been done for leak detection (and described at the end of the methodology section), the combined uncertainty is added to the predictions to obtain simulated measurements. Simulated measurements are used to test demand estimation through error-domain model falsification. The uncertainties used are the same as in the leak detection. The only difference is that the uncertainty of the demand is removed, because the demand is the parameter which needs to be identified.

Each measurement is compared to the predictions obtained by simulating a population of demand scenarios. Then, after the model falsification process, measurements are associated with a population of candidate demand scenarios. The flowchart in Figure 10 illustrates the process for demand estimation and how to ensure the quality of the sample.

Figure 11 shows results obtained when testing the number of samples necessary for the demand estimation. For each sample size tested, the number of times the correct value is included in the candidate models is computed (the percentage of correct identification). The sample sizes tested are 10,000, 20,000, 40,000 and 80,000. The results show that at 40,000 samples, the increase in the six nodes is over 90% and the increase after that slows. The cost of using more than 40,000 in terms of computation is drastically larger than the amount of increase in performance which would be achieved.

Figure 12 shows the results obtained for the estimation of the demand for the six nodes with 40,000 samples. The solid lines (versus the individual points) represent the 95% confidence intervals of the estimation, and the points are the exact demand values to identify. In accordance with previous results, the real value is inside the confidence interval in most cases.

Case studies

In this section, two case studies involving existing water distribution networks are presented. The first case study is part of the water-distribution network of the city of Lausanne. An experimental campaign was carried out to simulate leaks by opening hydrants. The second case study is from a water distribution network of the commune of Bagnes. With this network, a sensor placement study was carried out. These case studies show the potential of error-domain model falsification for the performance assessment of water distribution networks. The aspects covered by each case study are summarized in Table 3.

Lausanne water distribution network

This case study is based on a section of the water distribution network in Lausanne, Switzerland. The water distribution network in the city of Lausanne is separated into six independent subnetworks. This case study is based on an experimental campaign where leaks are created throughout the network at different locations by opening hydrants.

The model of this network is shown in Figure 13. It consists of 265 demand-nodes and 295 pipes. The pipes are represented by black lines and the nodes, white circles. This network is equipped with three flowmeters. Their positions are given by the black crosses.

The goal of this part of the study is to test the error-domain model-falsification methodology through various leak scenarios. These scenarios have been created by opening nine hydrants at four flow rates in order to simulate leaks at each position. The positions of the hydrants have been chosen such that the main regions of the sub-network are covered. They are given by the numbers in circles in the network representation displayed in Figure 13.

The flowmeter locations have been chosen using a sensor placement methodology based on the greedy algorithm and expected identifiability (Moser et al. 2016a). All three flowmeters are electromagnetic insertion flowmeters: the HydrINS 2 from hydreka (Figure 14). They have been programmed to take measurements every 15 minutes.

The behavior of a water distribution network is often governed by consumer demand. Generally, this data is not readily available. For this study case, the only available information is the global demand of the network. There is no information regarding the distribution of the demand throughout the network. In order to reduce the uncertainty resulting from such a lack of information, measurements are recorded only during the period of the day when consumption is the lowest. This is between 1:00 am and 4:30 am. Figure 15 shows the flow that is measured for one week (9th March to 15th March 2015) during the lowest consumption period.

Demand estimation

The demand of each node in the network is unknown. In order to bring the model behavior closer to the real behavior, error-domain model falsification is used to estimate the demand in the network. This is done by comparing demand scenarios with measurements. It is assumed that there is no leak in the network at the time of this measurement.

The only information available to model demand scenarios is the global demand of the network. It is not possible to use a normal sampling strategy to cover all possible behaviors of the system because the demand of each of the nodes can vary. The result is that the number of scenarios is equal to the number of nodes as a base with an exponent equal to the number of samples for the demand. For three sampling values and 265 nodes there are more than 18 million scenarios.

To overcome this challenge, a two-part solution is proposed. The first part of this solution is to obtain a reduced network keeping only the pipes that are monitored (Moser et al. 2015). Then, the demand is modelled only on the nodes connected to the monitored pipes (six nodes for the City of Lausanne network).

This leads to the second part of the solution—modelling the demand at the nodes randomly in order to build a number of demand scenarios. This is accomplished by using the only information available: the global demand. The condition is fixed such that the sum of the nodal demands at each node equals the global demand. The demand is allocated to each node by following a discrete uniform distribution varying from zero to the global demand. For this case study, 40,000 samples are used. The results in Figure 11 illustrate that this amount of samples is sufficient. This figure shows that above 40,000 samples, the increase of performance slows down significantly.

Figure 16 shows results for the estimation of the demand for the six nodes of the reduced network. In each graph, the horizontal axis represents the time. For this example, the time period is the same used previously to present the measurements. The vertical axis represents the demand for the node, given in l/min. The black curves are the 95% confidence intervals on the demand estimation. This means that, in each plot, 95% of the candidate models are between the two black curves. The dashed black curves display the median of the estimation and the grey curves, the average. The median and the average show that for all nodes except Node 3, the majority of candidate models are concentrated in the lower half of the confidence interval. This means that the estimation of the demand is higher for Node 3.

The results also show that the demand varies at each measurement time. In the following section, the demand used for leak detection is the average of the estimation performed the hour before the experiment, assuming that the demand does not vary significantly throughout the experiment. The graphs show that there are periods where the intensity is stable.

Leak-region detection

As explained previously, in order to test the leak detection methodology, an experimental campaign was carried out. Leaks were simulated on the network by opening hydrants. For each hydrant, four leak intensities were tested: 25 l/min, 50 l/min, 75 l/min and 100 l/min. The procedure was the following. Each night during the campaign, a hydrant was opened. The flow coming out of the hydrant was measured using a flowmeter. Figure 17 shows the setup used for the experiment.

The procedure was to open the hydrant to the first intensity (25 l/min) and then increase the flow each 30 minutes. An illustration of the procedure is given in Figure 18. The blue box that is connected to the hydrant is the flowmeter. Two reasons for waiting 30 minutes are that the perturbation needs time to reach the sensors and a minimum of two measurements are needed for each leak intensity (measurements have been taken every 15 minutes, this frequency was chosen to limit the amount of data and has no effect on the accuracy of the estimation). The positions of the nine hydrants used during the experimental campaign are illustrated on the network in Figure 13. Each hydrant is represented by a circle and a number. The hydrants have been chosen to cover all the regions of the sub-network.

At first, the results obtained when applying error-domain model falsification with the retrieved measurements show that all models were falsified. While initially alarming, this demonstrates

a strength of error-domain model falsification. This provides a warning that there are wrong assumptions associated with the modelling, sampling and uncertainties. Various reasons can explain such behavior. For example the sampling choice might be inappropriate for the primary parameters.

In this case, the main reason for the falsification of all the models is the way in which the demand of the network has been modeled. Since the distribution of the demand at each node is unknown, it has been modeled as an uncertainty using an exponential distribution as described previously. The results show that this model is too far from reality. For this reason, an estimation of the demand has been carried out using error-domain model falsification. The idea is to use the measurements retrieved before the experiment to estimate the demand during the experiment.

More specifically, the hydrant is opened at 2:15 am with a flow of 25 l/min. The period of time used for the demand estimation is between 1:00 am and 2:15 am. At each measurement time step (every 15 minutes), error-domain model falsification is used to obtain the population of demand scenarios. Each time step is computed separately without considering the transient phenomena. For this reason, only steady-state simulations have been used. Building scenarios considering the time parameters would increase the number of scenarios an exorbitant amount. Then, for each node, the average of the demand is computed for the time period and used as an estimation of the demand for the leak detection. Figure 18 illustrates this procedure.

Threshold estimation

A good estimation of the uncertainty is important for error-domain model falsification. The thresholds used in the methodology are built using uncertainty values. In this case study, the estimation of the modelling uncertainties is difficult because parameters that govern the system behavior, such as the demand, are not well-known.

However, in this case study, the target of the identification is fixed. The leak intensity and location are known. For this reason, the problem is inverted in order to estimate the uncertainties. This is achieved by comparing each measurement to the corresponding prediction. Variations between these values give information about the combined uncertainty of the system.

Figure 19 shows the value of measurements and predictions at each step time for two sensors. Measured values are represented by black lines. Predicted values are represented by the short dashed lines for the predictions obtained with demand estimation and long dashed lines for predictions obtained without demand estimation. These graphs show that when the demand is modeled with an exponential law instead of estimating it through model falsification, there is a significant bias for the flow in sensor 1. The results for sensor three shows that the curve of predicted values with demand estimations follows the measured values even when there is a jump in the measured values such as between the two tests presented in Figure 19.

In order to estimate the combined uncertainty, the differences between each measurement and corresponding predictions are computed. Figure 20 shows these results for the three sensors and for the cases when the demand is estimated and when it is not. The results are provided for the nine hydrant tests, listed on the horizontal axes. The vertical axes give the difference between the measured and predicted flows (in l/min). For each hydrant test, this difference is computed for eight measurements, two per leak intensity. The corresponding points are represented by the black squares on the graphs. The variation of the points across one hydrant test is due to modelling and measurement errors. Maximum and minimum values define an interval that can be considered as an estimation of the uncertainty. The evolution of this interval is illustrated on the graphs by the two grey curves.

The results in Figure 20 illustrate that the combined uncertainties are biased. Comparing the results obtained with and without demand estimation shows that estimation of the demand reduces the bias, especially for the first sensor. This demonstrates that estimating the demand will increase the quality of the flow predictions.

As explained previously, the demand is estimated before each hydrant test. The evolution of the bias with the hydrant tests shows that the quality of the demand estimation changes from one test to another. This is due to the fact that in some cases the demand changes faster than the estimated value.

The results in Figure 20 give the estimation of the uncertainties for each hydrant test separately. Each uncertainty estimation is related to a given demand estimation. Instead of considering each hydrant test individually, the difference between measurements and predictions can be considered together. This results in an estimation of uncertainties that include the uncertainty of the demand estimation process. Figure 21 shows these results in the form of histograms. The horizontal axes represent the difference between predictions and measurements. The vertical

axes display the probability. The width of the base of the histogram is a consequence of the variations values, including biases observed in Figure 20.

Estimations of uncertainties obtained in this way could be used directly to build thresholds for error-domain model falsification. However, in this case study knowledge of the demand must be enhanced in order to reduce the width of global combined uncertainties.

Figure 22 shows the results obtained for a simulated leak (simulated in the model) of 100 l/min when using threshold bounds estimated using all the hydrant tests. These results show that when demand estimation is used with threshold bound determination it is possible to identify the leak region.

Bagnes water distribution network

This case study is based on one of the water distribution networks of Bagnes. Bagnes is a commune in Valais State (Switzerland), situated in the mountains. Bagnes is made up of many villages, the most famous being Verbier. The water distribution networks of Bagnes are monitored in order to help with management of the networks. Each network in Bagnes is connected to one or more tanks that supply the water. The water going in and out of each tank is monitored continuously in order to directly gain knowledge of the consumption of each network.

This data can be used to detect leaks by observing abnormal variations in the measurements. However, the information is not sufficient to locate the leak-region, especially for networks connected to only one tank. This case study will initiate with a leak that has been observed and show the gain that can be obtained by installing sensors throughout the network.

Figure 23 shows the model of the water distribution network of Bagnes. It is made up of 900 nodes and 904 pipes. In comparison with the water distribution network of Lausanne shown previously, the number of pipes and nodes seems high. It is not because this network is bigger; in terms of size, it is a smaller network. The reason for the higher number of pipes and nodes is that the representation is more refined for the model of Bagnes. In the case of the Lausanne network, the network is already simplified.

The advantage of working on a smaller network is that the lowest consumption value is smaller. Since Lausanne is a city, even during nights, there are activities that take water from the network. For the Lausanne network the lowest consumption is around 830 l/min. For this

network, the lowest consumption is approximately 100 l/min. This data is displayed on Figure 24 which shows the mean hourly consumption for the month of August. The challenge is not the same in winter because this network is in the village of Verbier, one of the more prominent ski resorts in Switzerland. In winter, the population of Verbier increases to 50,000 inhabitants although it counts only 3,000 official inhabitants. This means that in winter the consumption, as well as the related uncertainty, is higher.

During the summer of 2013, a jump was observed in the flow measurement that records the water coming into the network from the tank. This jump is illustrated in Figure 25. This graph compares the daily measurements taken between the 21st of July and the 24th of September in 2012 and 2013. The jump measured in 2013 is due to a leak. Using this measurement, the leak intensity is estimated to be 200 l/min. These measurements also show that the leak was active for approximately 15 days. That means that it took more than 10 days to detect, locate and repair the leak. That time could be significantly reduced by using a monitoring system for leak detection. This amount of time with such a leak represents a loss of more than 4 million liters of fresh water. This study will show that by using error-domain model falsification, the leak-region can be located.

Leak-region detection

In this section, the leak detection methodology is tested for the leak that occurred in 2013. In order to achieve this, the same leak is simulated. The simulation of the leak is used to build simulated measurements. These measurements are then treated with error-domain model falsification for three sensor configurations: one with two sensors (Figure 26), one with six sensors (Figure 27, top) and one with ten sensors (Figure 27, bottom). The locations of the sensors are given by the black crosses. The nodes in dark represent the candidate leak scenarios. The position of the leak is shown by the four arrows.

The graph in Figure 28 shows the predicted value for the sensor in the middle of the branch where the leak is located (the sensor is represented above the graph). The horizontal axis gives the leak scenarios and the vertical axis, the predicted flow. The predictions are represented by the black points. The dashed horizontal lines are the threshold bounds and the continuous line, the value of the simulated measurement. All the points that are not between the thresholds are falsified. This figure clearly shows that in this case this sensor configuration is sufficient to identify the leak region.

Conclusions and discussion

An error-domain model-falsification methodology that is uniquely adapted for leak-region detection demonstrates potential for practical use.

Error-domain-model falsification is useful for estimating the demand at a small number of nodes. The strategy to estimate combined uncertainty helps aggregate uncertainty sources.

The study of the Lausanne network revealed that demand estimations decrease the uncertainty of the system and lead to better performance for leak detection. The estimation of uncertainties shows that significant bias is present. This supports the use of error-domain model falsification for data interpretation. Demand estimation removes part of the systematic uncertainty. Reducing systematic uncertainty lowers interdependence of measurement locations, and this improves predictions.

The results of the Bagnes case study further show that error-domain model falsification can help locate leak regions within a water supply network. This is demonstrated through the example of a leak that has occurred on the Bagnes network. This example shows that even high intensity leaks take time to be located without such support for the leak detection, and this results in significant water loss.

In addition, this case study shows that a smaller network, situated in a village rather than a city, could have a smaller minimum demand ratio, and thus reduced uncertainty.

In all of the case studies presented in this paper, the networks were dependent on a single reservoir or tank. In reality, networks can be designed in various ways, including, for example, with intermediate reservoirs. The methodology presented is capable of accommodating such networks. The manner in which the network is modelled would need to include these new parameters.

Acknowledgments

This research is part of the Water Resources Innovation Program of EPFL Middle East and is funded by EPFL Middle East. The authors acknowledge the support from Eau Service, the water provider of the city of Lausanne.

References

ADEC, (1999). *Technical review of leak detection technologies - vol. 1 - crude oil transmission pipeline*. Vol. 1. 1999, Alaska: Alaska Department of Environmental Conservation.

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

Babbitt, H., Amsbary, F. and Gwinn, D. (1920). The detection of leaks in underground pipes, *Journal (American Water Works Association), 7*(4): p. 589-595.

Barandouzi, M. A., Mahinthakumar, G., Brill, E. D. and Ranjithan, R. (2012). "Probabilistic Mapping of Water Leakage Characterizations Using a Bayesian Approach," in *Proceedings of World Environmental and Water Resources Congress 2012*, Albuquerque, New Mexico, USA. 3248-3256. ASCE.

Box, G.E. and Draper, N.R. (1959). A basis for the selection of a response surface design, *Journal of the American Statistical Association, 54*(287): p. 622-654.

Colombo, A.F., Lee, P. and Karney, B.W. (2009). A selective literature review of transient-based leak detection methods, *Journal of Hydro-environment Research, 2*(4): p. 212-227.

Demirci, S., Yigit, E., Eskidemir, I. H. and Ozdemir, C. (2012). Ground penetrating radar imaging of water leaks from buried pipes based on backprojection method, *NDT & E International, 47*(0): p. 35-42.

Fang, H., Rais-Rohani, M., Liu, Z. and Horstemeyer, M.F. (2005). A comparative study of metamodeling methods for multiobjective optimization, *Computers and Structures, 83*(25-26): p. 2121-2136.

Gao, Y., Brennan, M.J. and Joseph, P.F. (2006). A comparison of time delay estimators for the detection of leak noise signals in plastic water distribution pipes, *Journal of Sound and Vibration, 292*(3–5): p. 552-570.

Gao, Y., Brennan, M.J. and Joseph, P.F. (2012). On the effects of reflections on time delay estimation for leak detection in buried plastic water pipes, *Journal of Sound and Vibration, 325*(3): p. 649-663.

Goulet, J. and Smith, I.F.C. (2012). Performance-driven measurement system design for structural identification, *Journal of Computing in Civil Engineering, 27*(4): p. 427-436.

Goulet, J.-A. and Smith, I.F.C. (2013). Structural identification with systematic errors and unknown uncertainty dependencies, *Computers & structures, 128*: p. 251-258.

Goulet, J., Coutu, S. and Smith, I.F.C. (2013). Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks, *Advanced Engineering Informatics, 27*(2): p. 261-269.

Grunwell, D. and Ratcliffe, B. (1981). "Location of underground leaks using the leak noise correlator," in *WRC Technical Report*. 1981, WRC.

Hope, W. (1892). The waste of water in public supplies and its prevention, *Minutes of the Proceedings 110*, 260-275.

Johnson, D.W. (1947). Losses in Distribution Systems. *Journal (American Water Works Association), 1947:* p. 157-167.

Lambert, A. and Hirner, W. (2000). "Losses from water supply systems: standard terminology and recommended performance measures," in *the blue pages the IWA information source on drinking water issues*, 2000.

Mergelas, B. and Henrich, G. (2005). Leak locating method for precommissioned transmission pipelines: North American case studies, *Leakage 2005*, p. 12-14.

Morrison, J. (2004). Managing leakage by district metered areas: a practical approach, *Water 21*, 2004(IWA Task Force): p. 44-46.

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

Moser, G., Paal, S.G., Jlelaty, D. and Smith, I.F.C. (2016a). An electrical network for evaluating monitoring strategies intended for hydraulic pressurized networks, *Journal of Advanced Engineering Informatics, 30* (4): pp. 672-686.

Moser, G. Paal, S.G. and Smith, I.F.C. (2016b). Measurement system design for leak detection in hydraulic pressurized networks, *Structure and Infrastructure Engineering*, DOI: 10.1080/15732479.2016.1225312, 30 Aug 2016.

Mounce, S., Boxall, J. and Machell, J. (2009). Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows, *Journal of Water Resources Planning and Management, 136*(3): p. 309-318.

Mounce, S.R., Mounce, R.B. and Boxall, J.B. (2011). Novelty detection for time series data analysis in water distribution systems using support vector machines, *Journal of hydroinformatics, 13*(4): p 672-686.

Murray, P.S. and Silea, I. (2012). A survey on gas leak detection and localization techniques, *Journal of loss prevention in the process industries*, 2012.

Niemeyer, H. (1940). Leak Detection, *Journal (American Water Works Association), 1940*: p. 1354-1358.

Pasquier, R. and Smith, I.F.C. (2015). Robust system identification and model predictions in the presence of systematic uncertainty, *Advanced Engineering Informatics.*

Poulakis, Z., Valougeorgis, D. and Papadimitriou, C. (2003). Leakage detection in water pipe networks using a Bayesian probabilistic framework, *Probabilistic Engineering Mechanics, 18*(4): p. 315-327.

Pudar, R.S. and Liggett, J.A. (1992). Leaks in Pipe Networks, *Journal of Hydraulic Engineering, 118*(7): p. 1031-1046.

Puust, R., Kapelan, Z., Koppel, T. and Savic, D. (2006). "Probabilistic Leak Detection in Pipe Networks Using the SCEM-UA Algorithm", in *Proceedings of Water Distribution Systems Analysis Symposium 2006*, Cincinnati, Ohio, USA. 1-12. ASCE.

Puust, R., Kapelan, Z., Savic, D.A. and Koppel, T. (2010). A review of methods for leakage management in pipe networks, *Urban Water Journal, 7*(1): p. 25-45.

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

Romano, M., Kapelan, Z. and Savić, D.A. (2012). Automated detection of pipe bursts and other events in water distribution systems, *Journal of Water Resources Planning and Management, 140*(4): p. 457-467.

Romano, M., Kapelan, Z. and Savić, D.A. (2013). Geostatistical techniques for approximate location of pipe burst events in water distribution systems, *Journal of hydroinformatics, 15*(3): p. 634-651.

Romano, M., Kapelan, Z. and Savić, D.A. (2014). Evolutionary Algorithm and Expectation Maximization Strategies for Improved Detection of Pipe Bursts and Other Events in Water Distribution Systems, *Journal of Water Resources Planning and Management, 140*(5): p. 572-584.

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

Rougier, J. (2005). Probabilistic leak detection in pipelines using the mass imbalance approach, *Journal of Hydraulic Research, 43*(5): p. 556-566.

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

Smith, I.F.C. (2016). Studies of Sensor Data Interpretation for Asset Management of the Built Environment, *Journal of Frontiers in Built Environment, 2:*8.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B. & Whittle, A. J. (2010). "Water main burst event detection and localization", in *Proceedings of Proceedings of 12th Water Distribution Systems Analysis Conference (WDSA'10)*.

Vítkovský, J. P., Simpson, A. R. & Lambert, M. (2000). Leak detection and calibration using transients and genetic algorithms, *Water Resources Planning and Management*, 258–262.

Vítkovský, J. P., Lambert, M. F., Simpson, A. R. & Liggett, J. A. (2007). Experimental observation and analysis of inverse transients for pipeline leak detection, *Journal of Water Resources Planning and Management*, 133, 519.

Whittle, A. J., Girod, L., Preis, A., Allen, M., Lim, H. B., Iqbal, M., Srirangarajan, S., Fu, C., Wong, K. J. & Goldsmith, D. (2010). WATERWISE@ SG: A testbed for continuous monitoring of the water distribution system in singapore, *Water Distribution System Analysis, WSDA*.

Whittle, A., Allen, M., Preis, A. & Iqbal, M. 2013. "Sensor Networks for Monitoring and Control of Water Distribution Systems", in *Proceedings of The 6th International Conference on Structural Health Monitorinf of Intelligent Infrastructure*, Hong Kong.

726 Table 1. Uncertainties related to secondary parameters (Exp is an exponential distribution, N is Gaussian and U
727 is uniform). The arguments of the distribution contain defining parameters.

Parameter	Uncertainty distribution
Nodal demand [l/min]	$\sim Exp(1/3.13)$
Node elevation [m]	$\sim N(0,0.015)$
Pipe diameter [mm]	$\sim N(0,0.75)$
Pipe length [m]	$\sim U(-0.03,0.07)$
Pipe roughness	$\sim U(0,0.015)$
Tank level [m]	$\sim N0(0,0.32)$

728

729 Table 2. Relative importance of secondary-parameter uncertainties

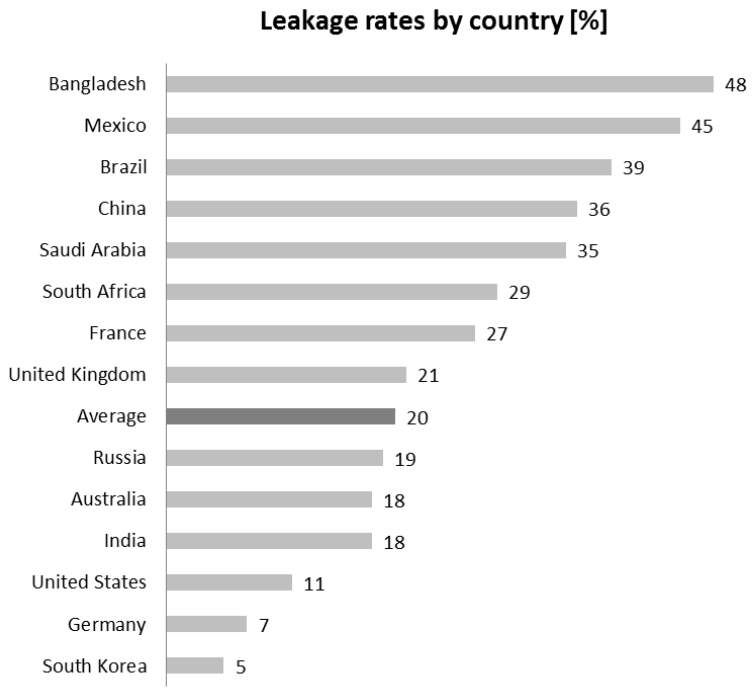
Parameter	Relative importance for flow predictions [%]	Relative importance for pressure predictions [%]
Nodal demand [l/min]	99.77	97.81
Node elevation [m]	5.45E-04	9.06E-02
Pipe diameter [mm]	2.25E-01	5.14E-01
Pipe length [m]	3.11E-03	3.34E-02
Pipe roughness	5.34E-05	1.49
Tank level [m]	2.80E-04	6.38E-02

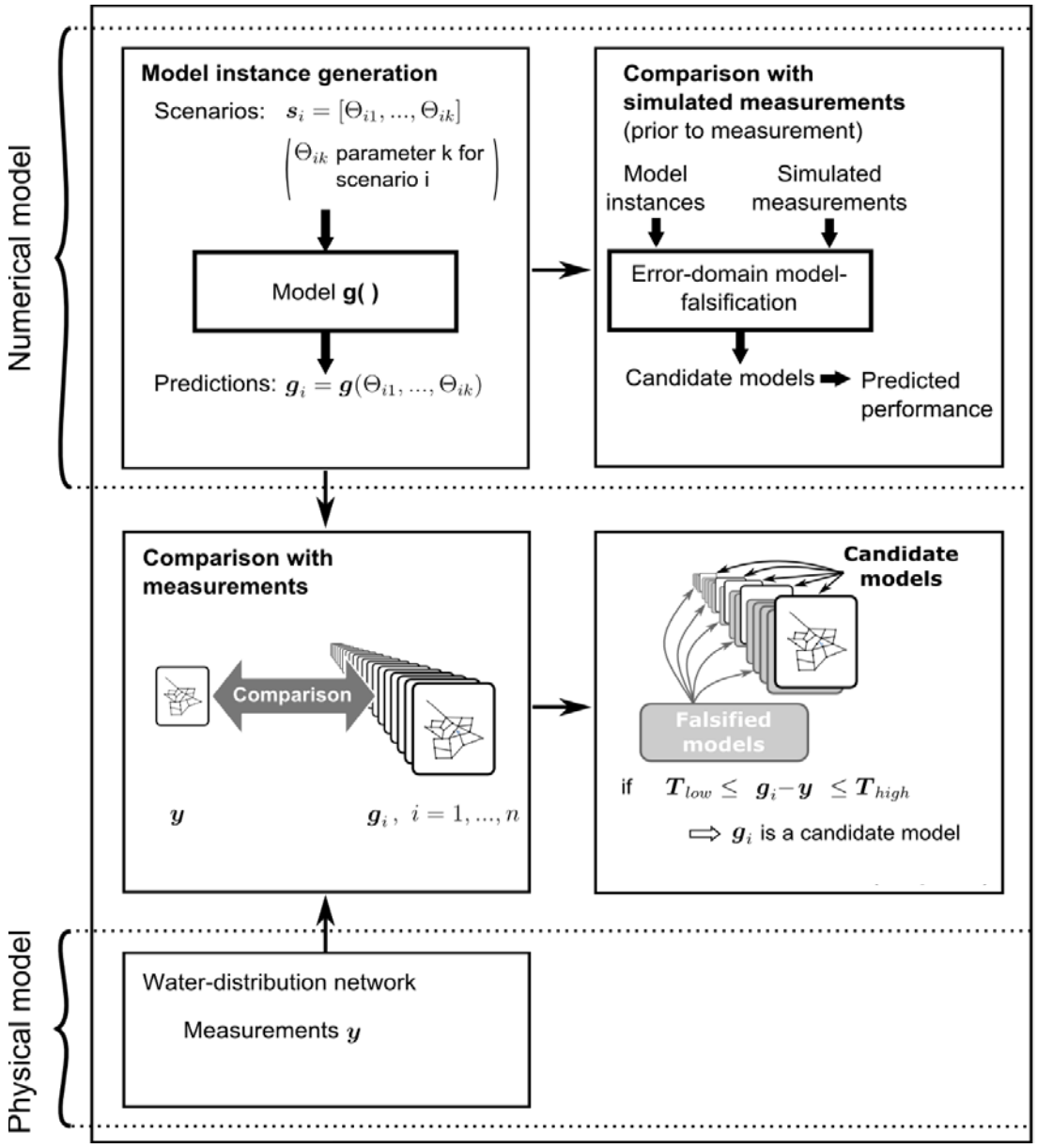
730

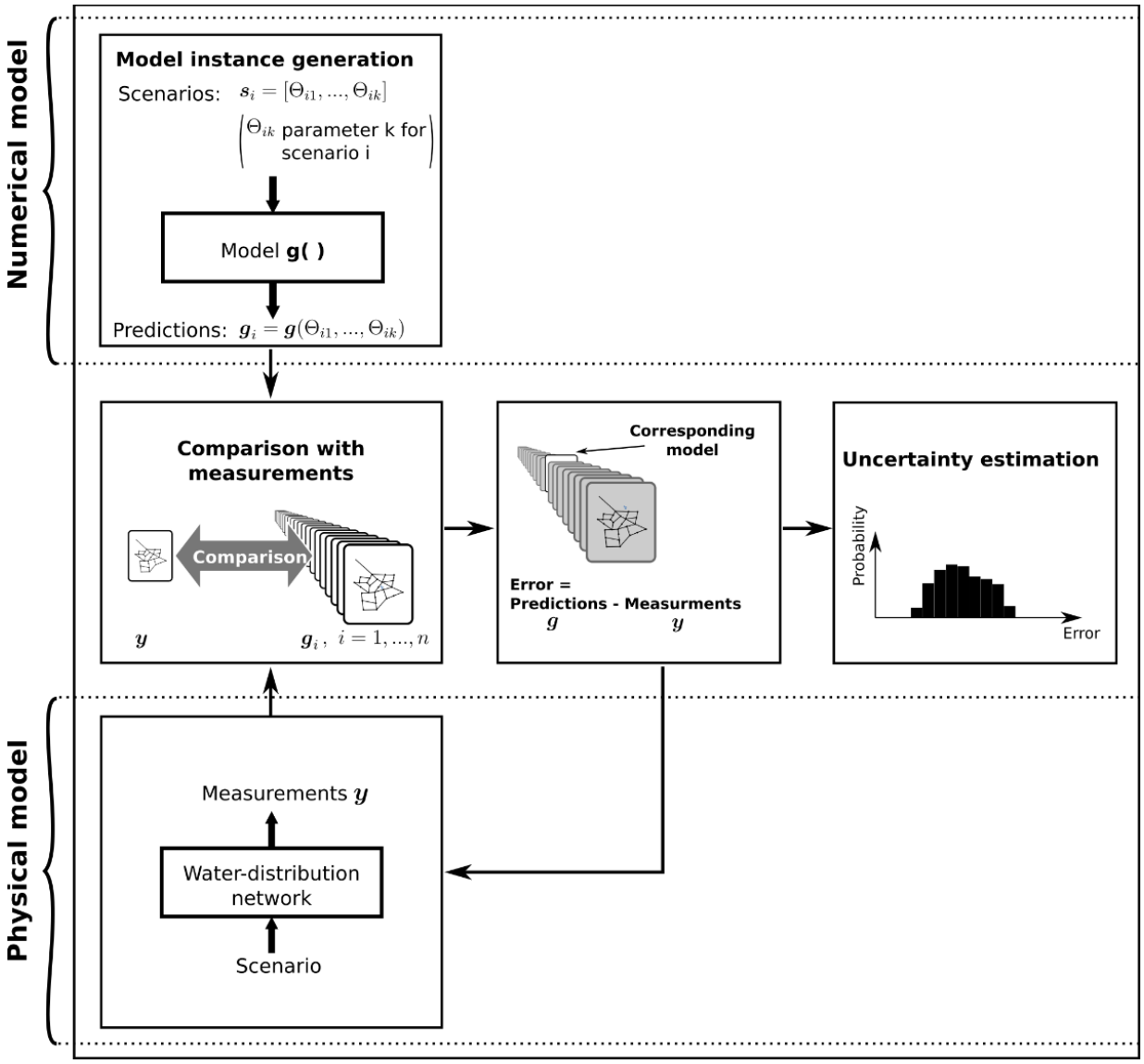
731 Table 3. Overview of aspects covered by each case study

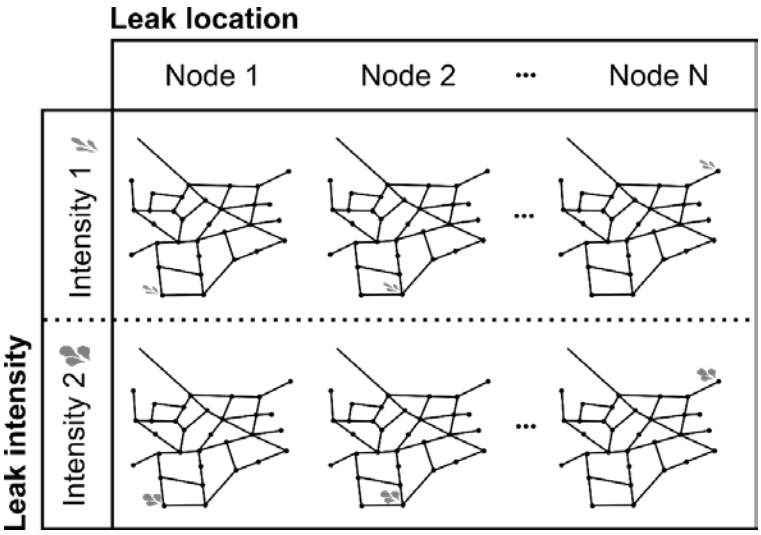
Case study	Aspect illustrated				
	Leak detection	Demand estimation	Uncertainty estimation	Network reduction	Sensor placement
Lausanne network	✓	✓	✓		
Bagnes network	✓			✓	✓

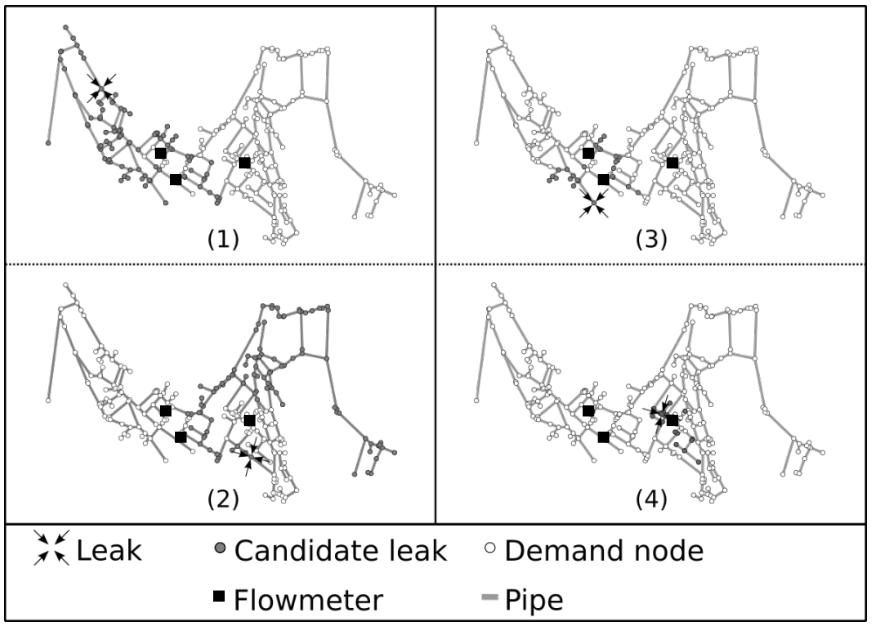
732

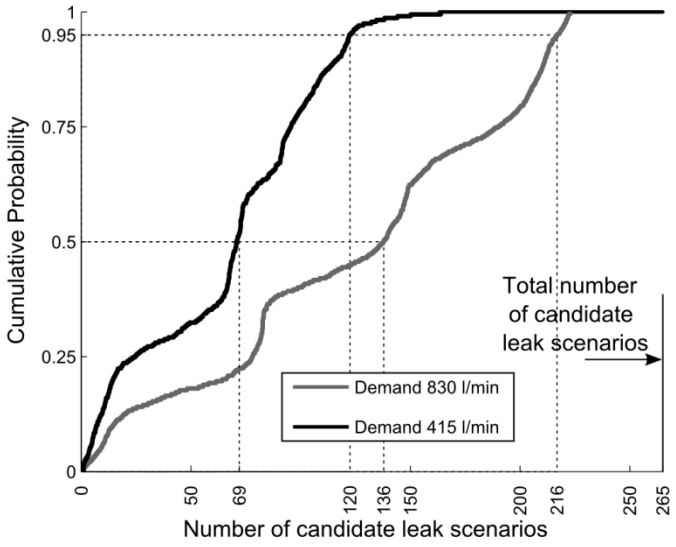

Leakage rates by country [%]
Bangladesh 48
Mexico 45
Brazil 39
China 36
Saudi Arabia 35
South Africa 29
France 27
United Kingdom 21
Average 20
Russia 19
Australia 18
India 18
United States 11
Germany 7
South Korea 5

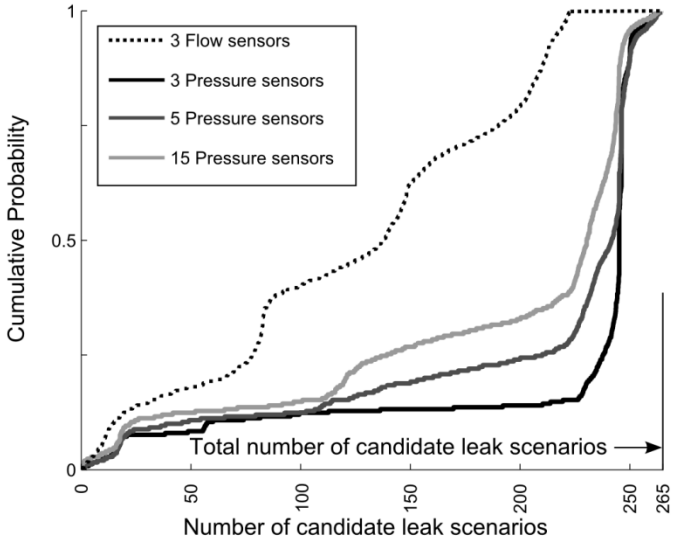

Numerical model
Physical model
Model instance generation
Scenarios: $s_i = [\Theta_{i1}, ..., \Theta_{ik}]$
Θ_{ik} parameter k for scenario i
Model g( )
Predictions: $g_i = g(\Theta_{i1}, ..., \Theta_{ik})$
Comparison with simulated measurements
(prior to measurement)
Model instances
Simulated measurements
Error-domain model-falsification
Candidate models
Predicted performance
Comparison with measurements
Comparison
y
g_i, $i = 1, ..., n$
Candidate models
Falsified models
if $T_{low} \leq g_i - y \leq T_{high}$
g_i is a candidate model
Water-distribution network
Measurements y

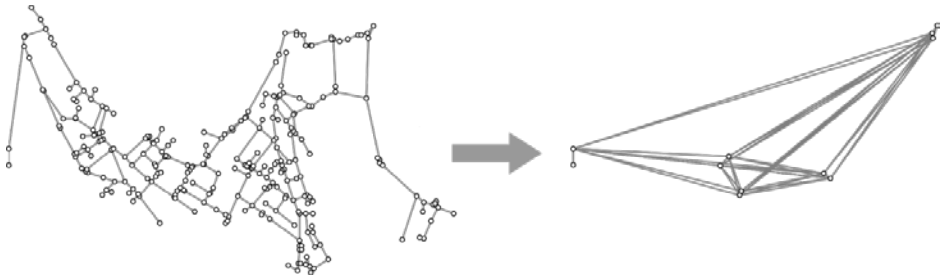

Numerical model
Model instance generation
Scenarios: $s_i = [\Theta_{i1}, ..., \Theta_{ik}]$
Θ_{ik} parameter k for scenario i
Model g( )
Predictions: $g_i = g(\Theta_{i1}, ..., \Theta_{ik})$
Comparison with measurements
Comparison
y
g_i, $i = 1, ..., n$
Corresponding model
Error = Predictions - Measurments
g
y
Uncertainty estimation
Probability
Error
Physical model
Measurements y
Water-distribution network
Scenario

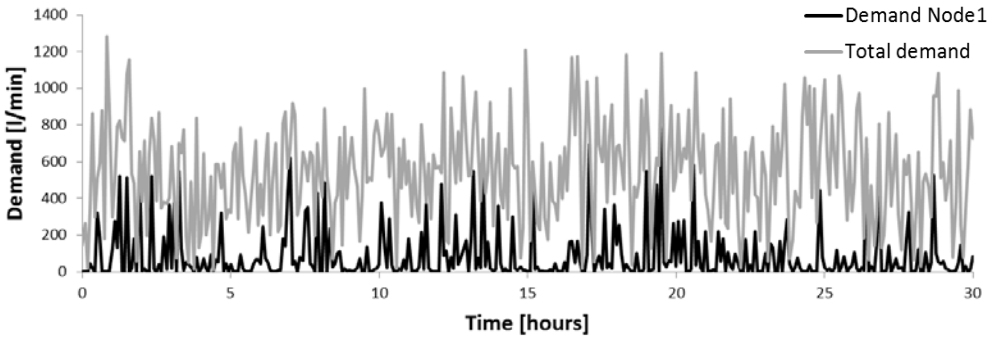

Leak location
Node 1
Node 2
...
Node N
Leak intensity
Intensity 1
Intensity 2

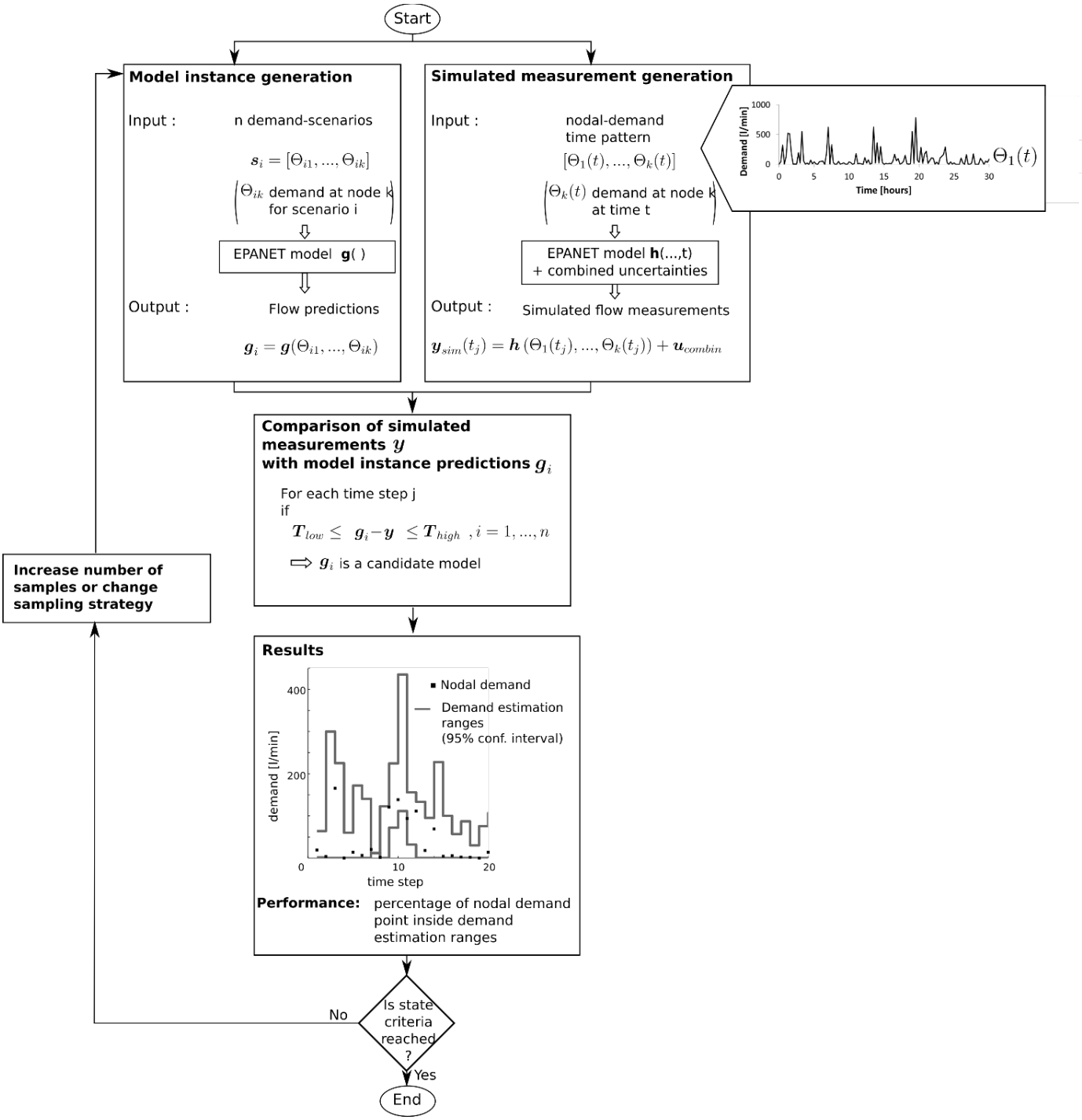

(1)
(3)
(2)
(4)
Leak
Candidate leak
Demand node
Flowmeter
Pipe

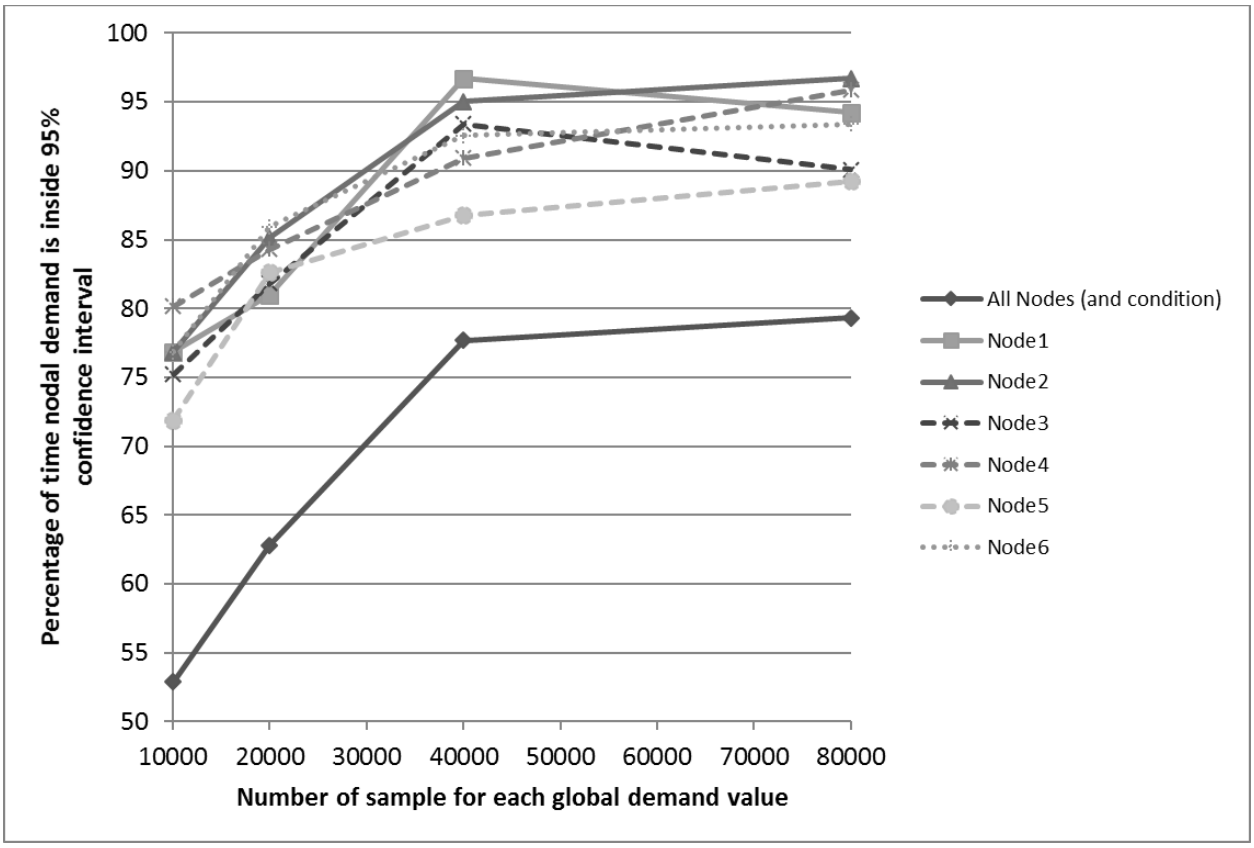

Cumulative Probability
1
0.95
0.75
0.5
0.25
0
0
50
69
120
136
150
200
216
250
265
Number of candidate leak scenarios
Total number of candidate leak scenarios
Demand 830 l/min
Demand 415 l/min

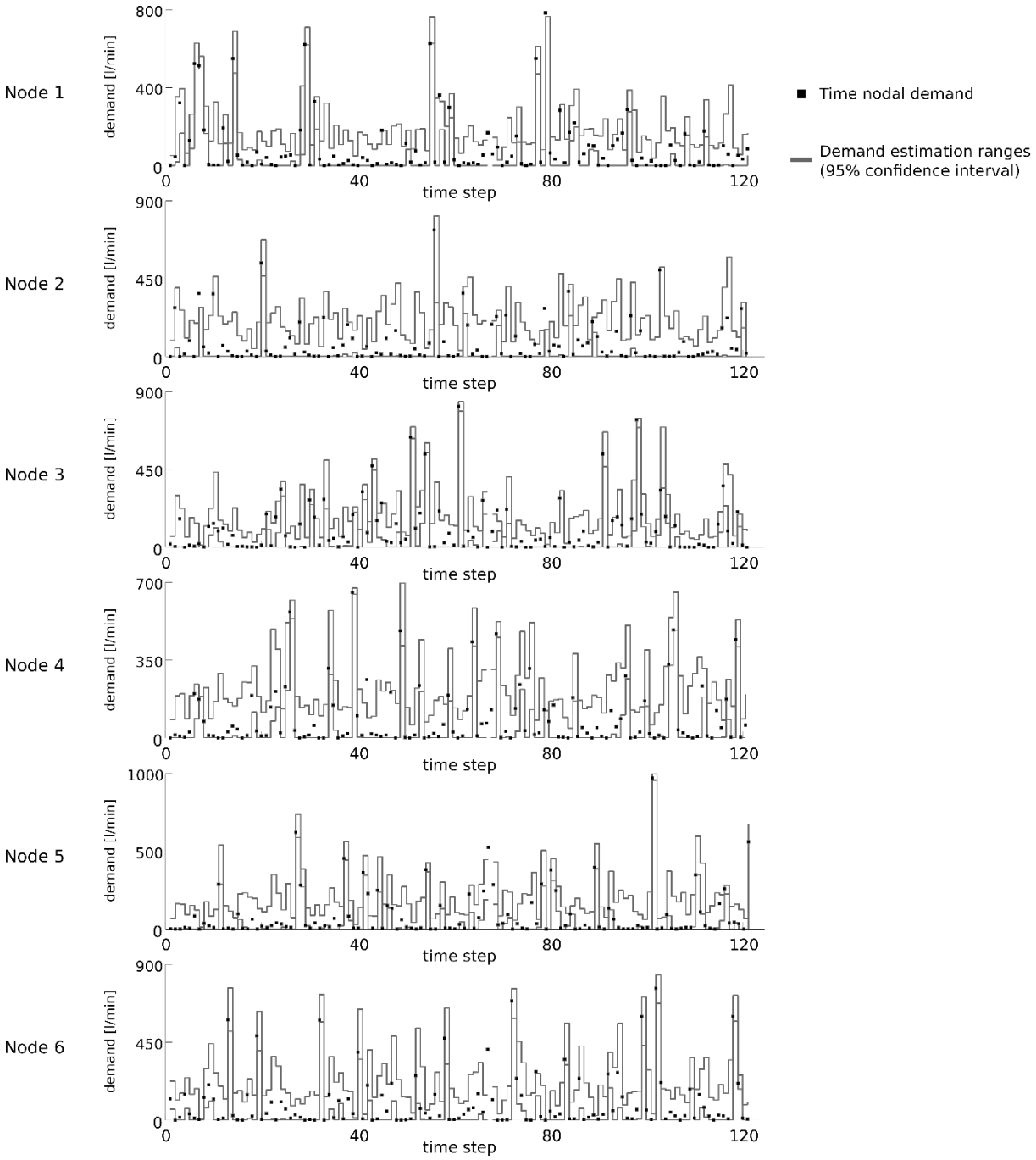

3 Flow sensors
3 Pressure sensors
5 Pressure sensors
15 Pressure sensors
Cumulative Probability
1
0.5
0
Total number of candidate leak scenarios
0
50
100
150
200
250
265
Number of candidate leak scenarios

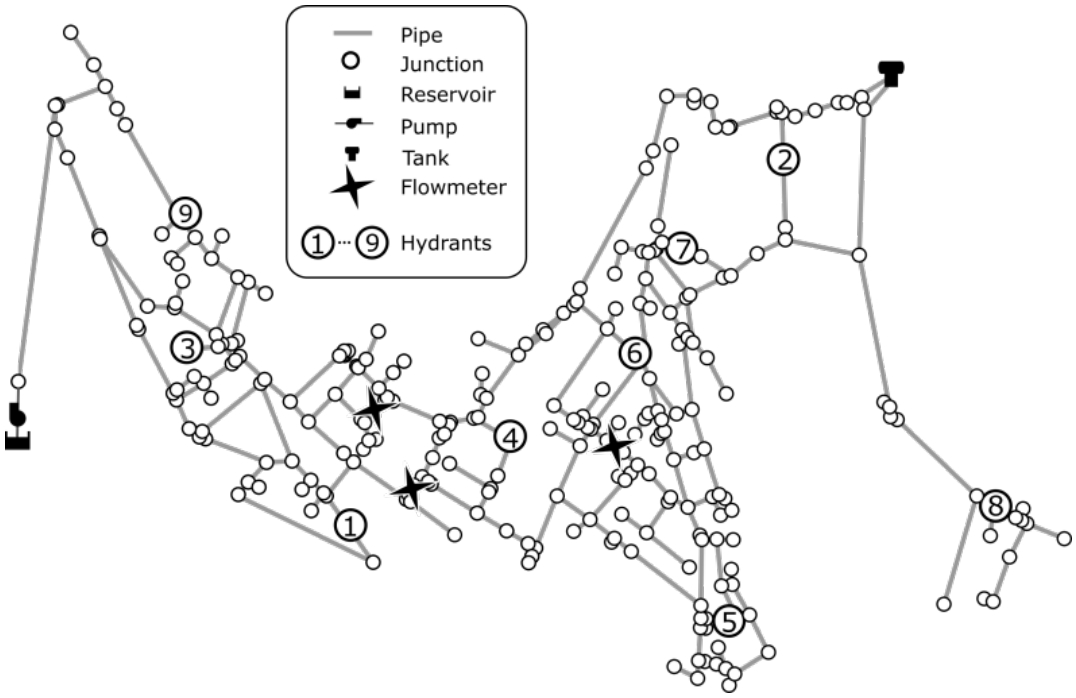



Demand Node1
Total demand
Demand [l/min]
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
30
Time [hours]

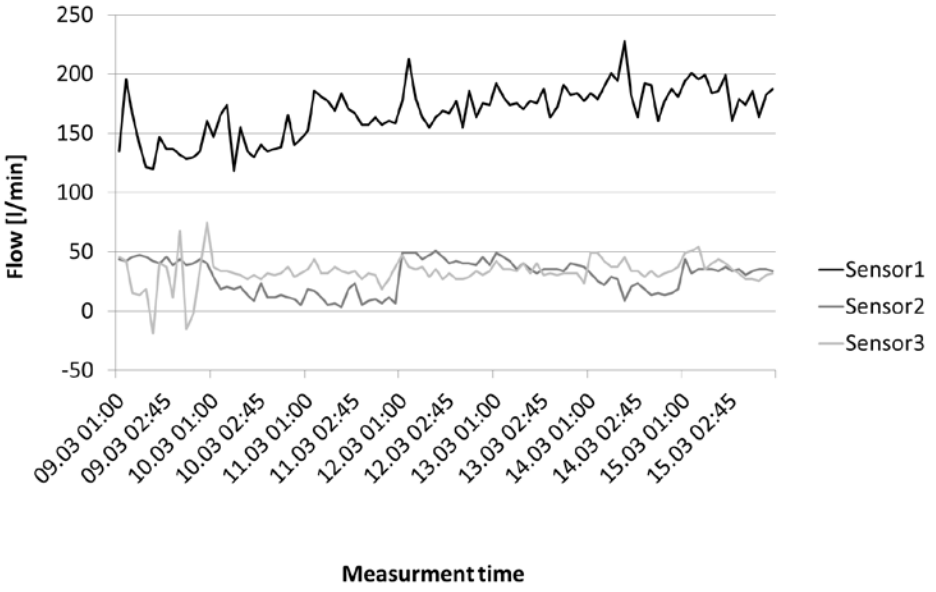

Start
Model instance generation
Input :
n demand-scenarios
$s_i = [\Theta_{i1}, ..., \Theta_{ik}]$
Θ_{ik} demand at node k for scenario i
EPANET model g( )
Output :
Flow predictions
$g_i = g(\Theta_{i1}, ..., \Theta_{ik})$
Simulated measurement generation
Input :
nodal-demand time pattern
$[\Theta_1(t), ..., \Theta_k(t)]$
$\Theta_k(t)$ demand at node k at time t
EPANET model h(...,t) + combined uncertainties
Output :
Simulated flow measurements
$y_{sim}(t_j) = h(\Theta_1(t_j), ..., \Theta_k(t_j)) + u_{combin}$
Demand [l/min]
Time [hours]
$\Theta_1(t)$
Comparison of simulated measurements y with model instance predictions g_i
For each time step j
if
$T_{low} \leq g_i - y \leq T_{high}$, $i = 1, ..., n$
g_i is a candidate model
Increase number of samples or change sampling strategy
Results
Nodal demand
Demand estimation ranges (95% conf. interval)
demand [l/min]
time step
Performance: percentage of nodal demand point inside demand estimation ranges
Is state criteria reached ?
No
Yes
End

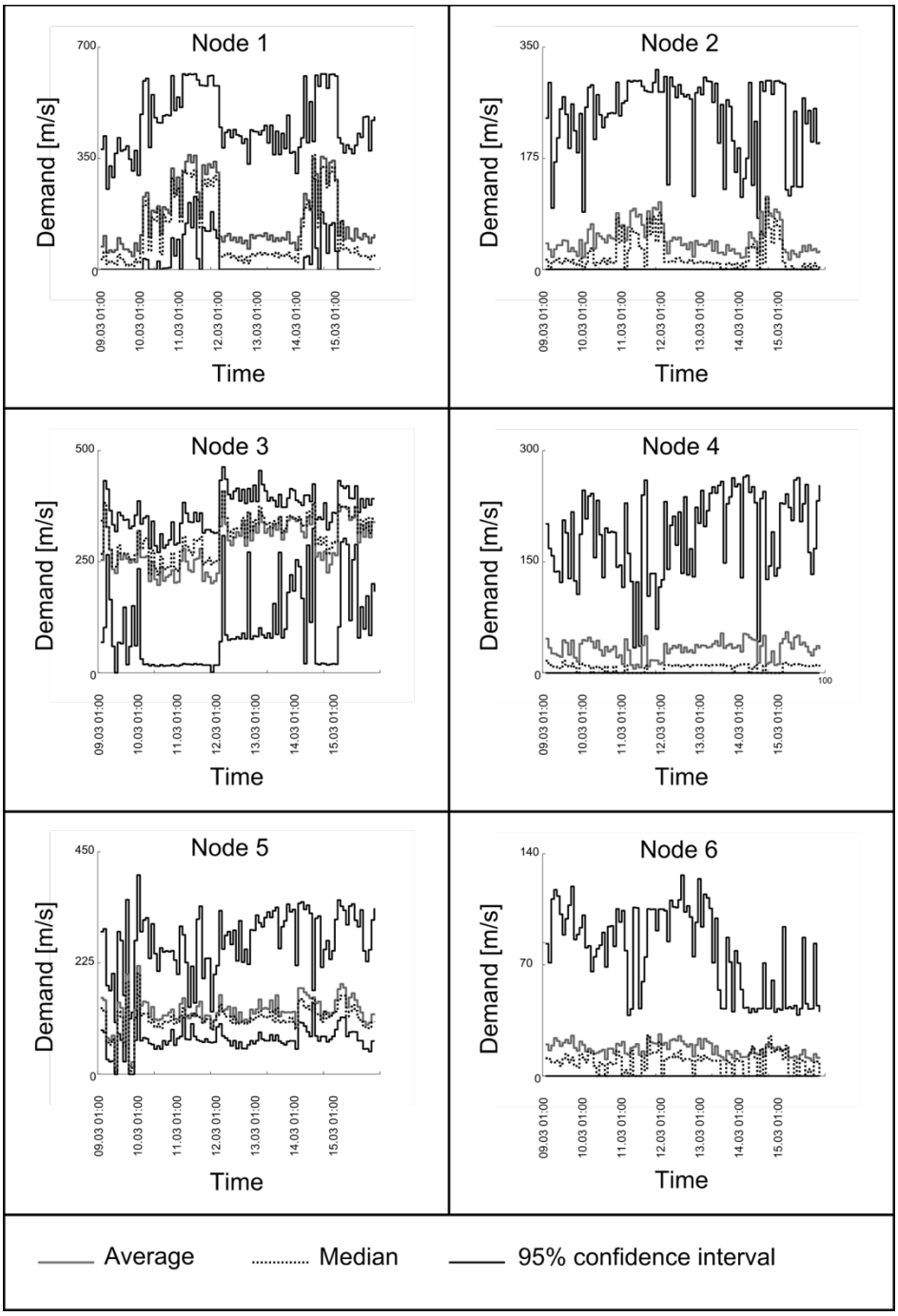

Percentage of time nodal demand is inside 95% confidence interval
100
95
90
85
80
75
70
65
60
55
50
10000
20000
30000
40000
50000
60000
70000
80000
Number of sample for each global demand value
All Nodes (and condition)
Node1
Node2
Node3
Node4
Node5
Node6

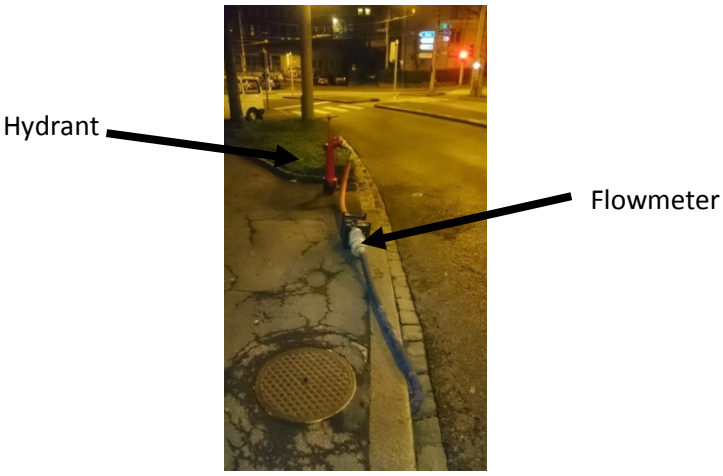

Node 1
Node 2
Node 3
Node 4
Node 5
Node 6
demand [l/min]
time step
0
40
80
120
800
400
900
450
700
350
1000
500
Time nodal demand
Demand estimation ranges
(95% confidence interval)

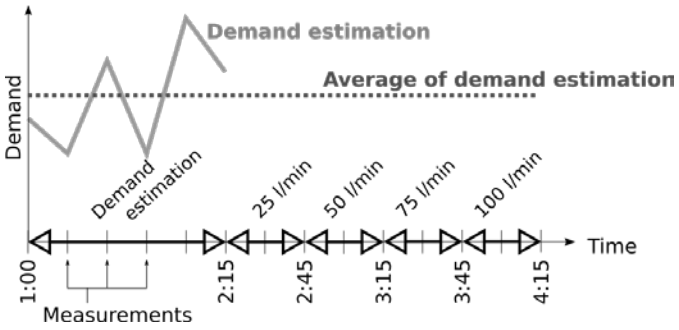

Pipe
Junction
Reservoir
Pump
Tank
Flowmeter
1 ··· 9 Hydrants
1
2
3
4
5
6
7
8
9

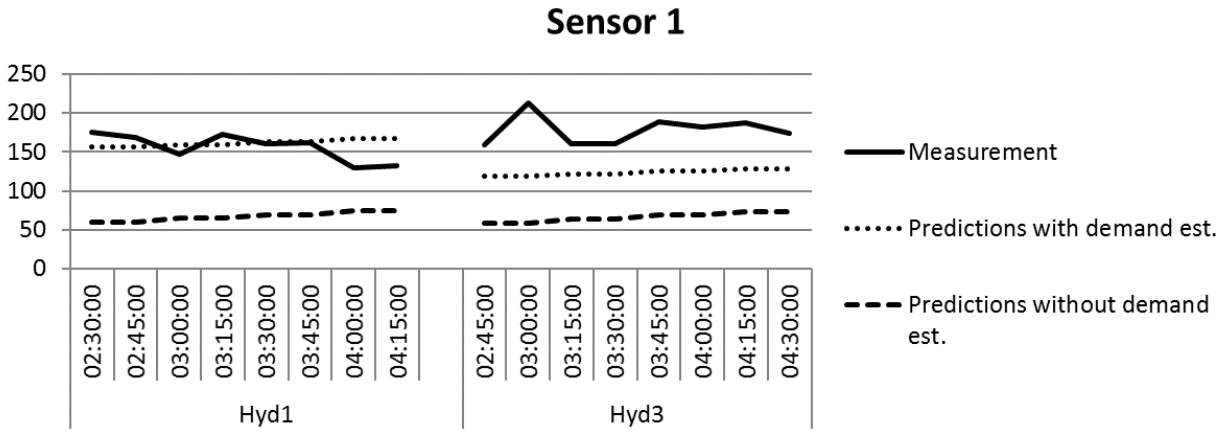

Pipe
Flowmeter

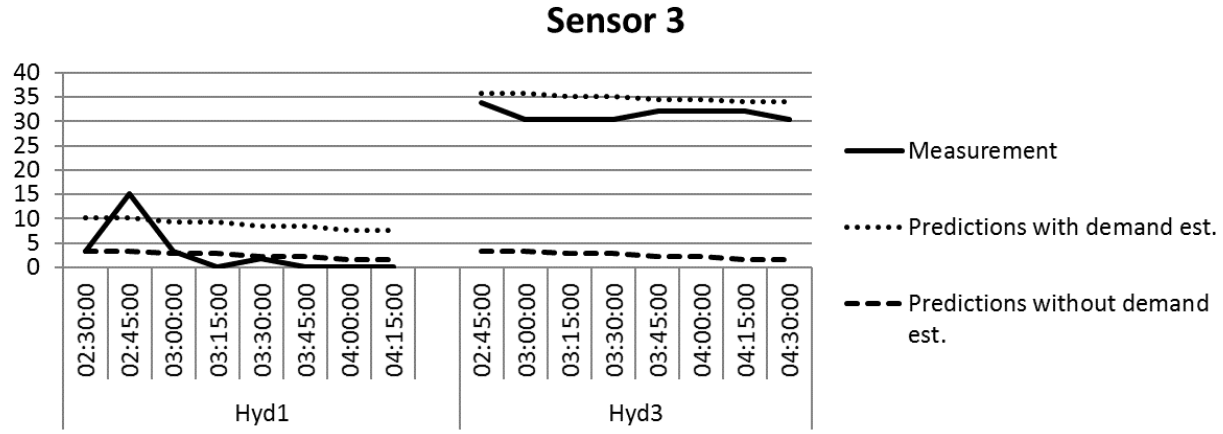

250
200
150
100
50
0
-50
Flow [l/min]
09.03 01:00
09.03 02:45
10.03 01:00
10.03 02:45
11.03 01:00
11.03 02:45
12.03 01:00
12.03 02:45
13.03 01:00
13.03 02:45
14.03 01:00
14.03 02:45
15.03 01:00
15.03 02:45
Measurment time
Sensor1
Sensor2
Sensor3

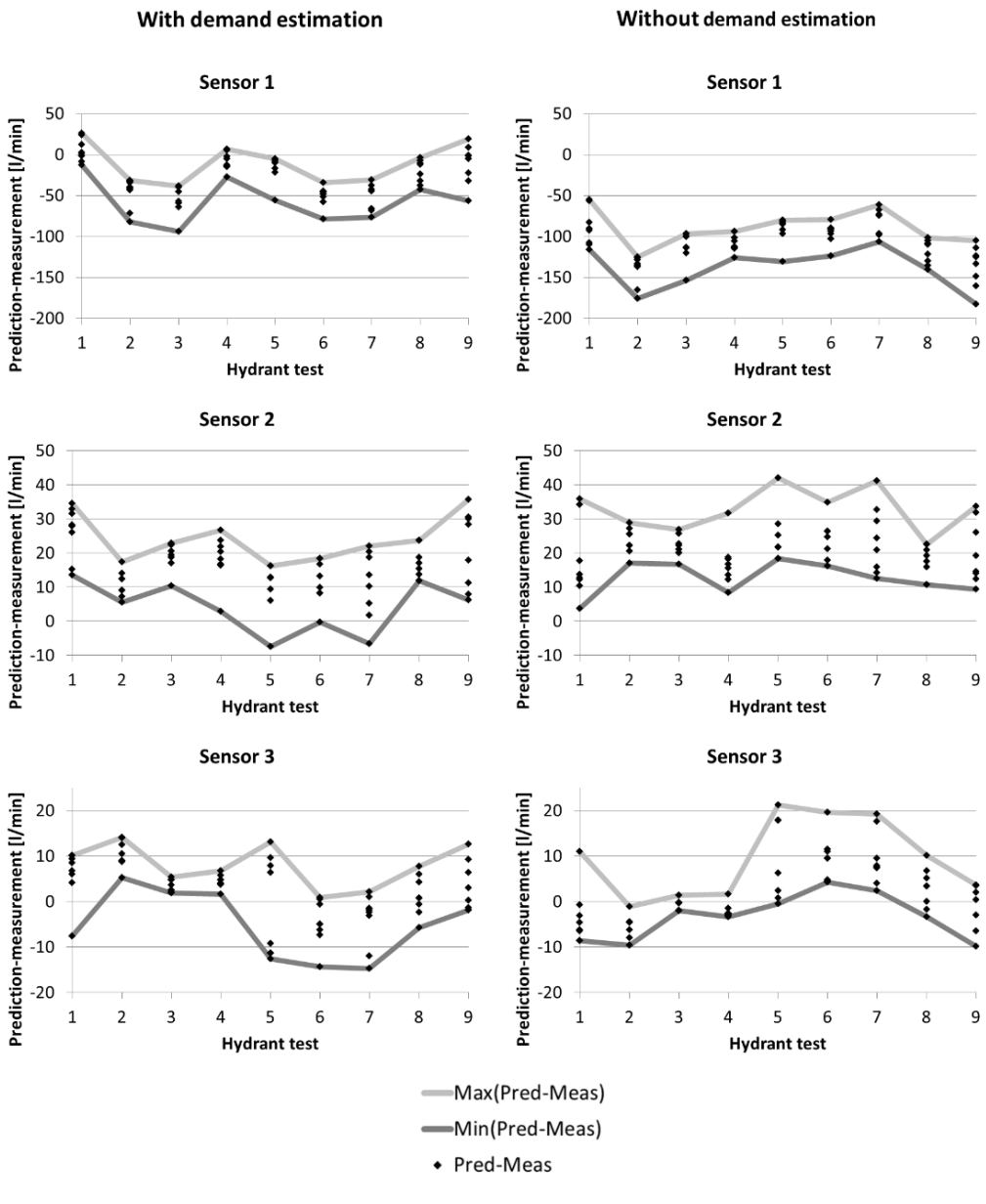

Node 1
Demand [m/s]
700
350
0
09.03 01:00
10.03 01:00
11.03 01:00
12.03 01:00
13.03 01:00
14.03 01:00
15.03 01:00
Time
Node 2
Demand [m/s]
350
175
0
Time
Node 3
Demand [m/s]
500
250
0
Time
Node 4
Demand [m/s]
300
150
0
100
Time
Node 5
Demand [m/s]
450
225
0
Time
Node 6
Demand [m/s]
140
70
0
Time
Average
Median
95% confidence interval

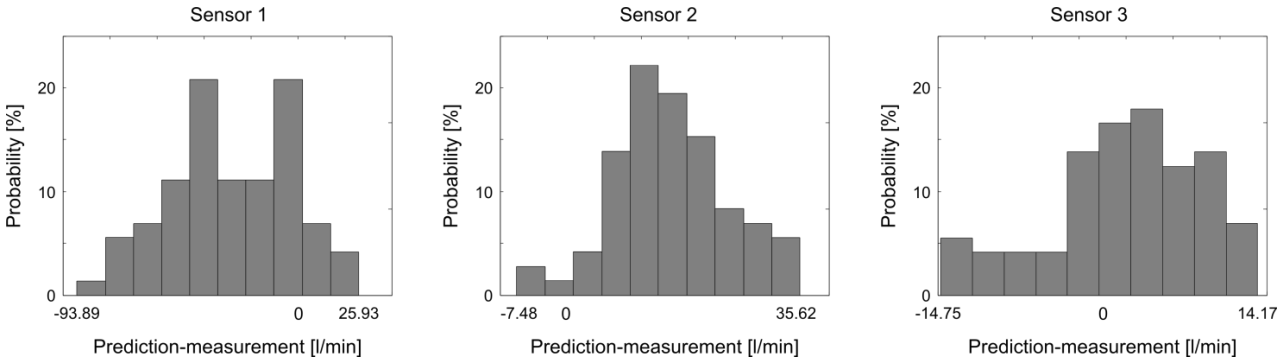

Hydrant
Flowmeter

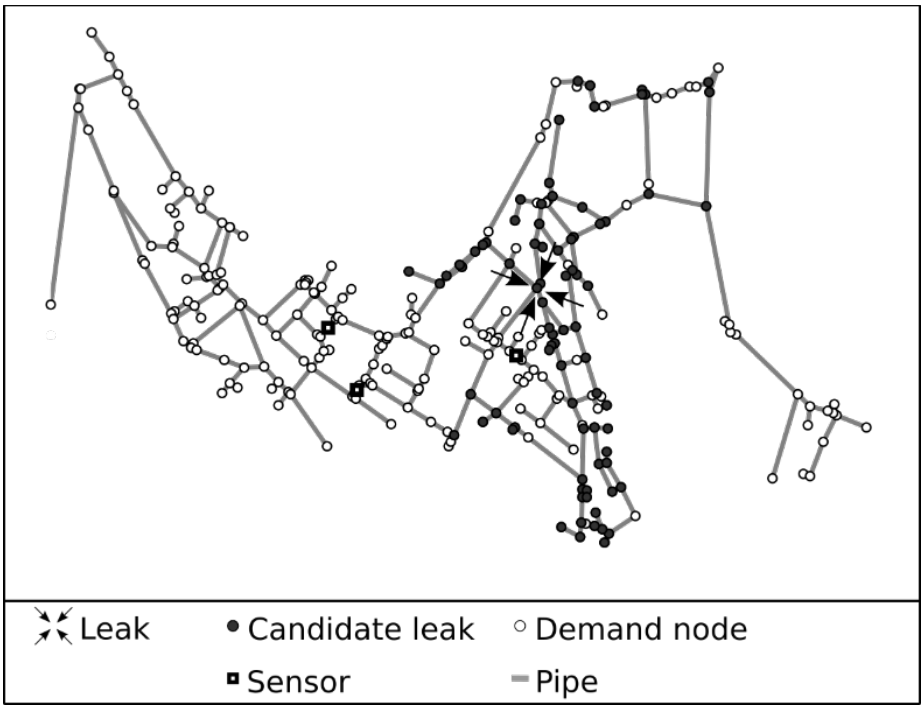

Demand estimation
Average of demand estimation
Demand
Demand estimation
25 l/min
50 l/min
75 l/min
100 l/min
Time
1:00
2:15
2:45
3:15
3:45
4:15
Measurements

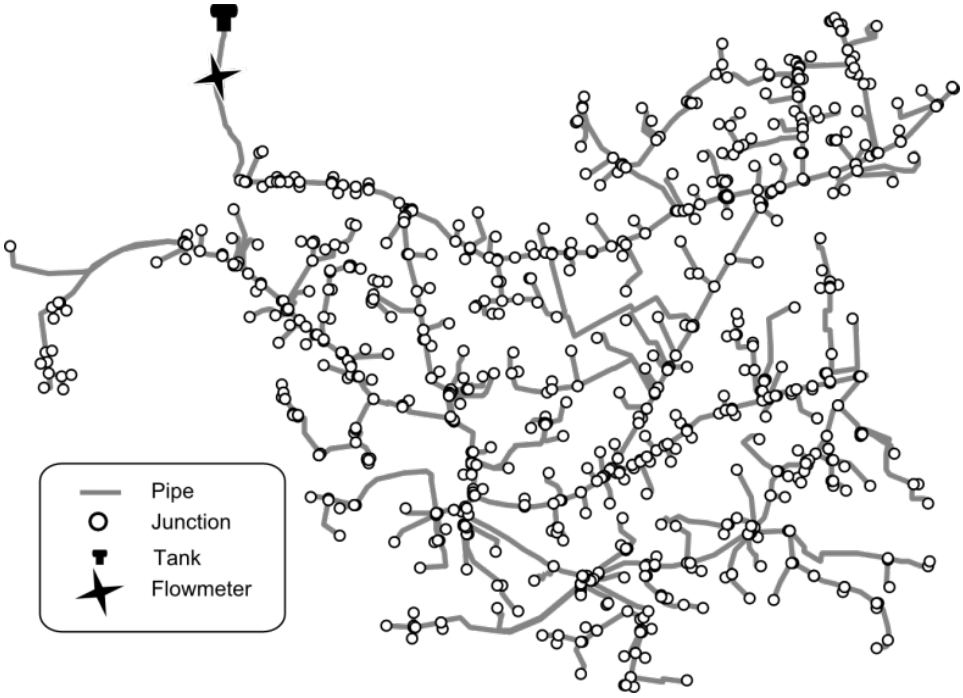

Sensor 1
250
200
150
100
50
0
02:30:00
02:45:00
03:00:00
03:15:00
03:30:00
03:45:00
04:00:00
04:15:00
02:45:00
03:00:00
03:15:00
03:30:00
03:45:00
04:00:00
04:15:00
04:30:00
Hyd1
Hyd3
Measurement
Predictions with demand est.
Predictions without demand est.

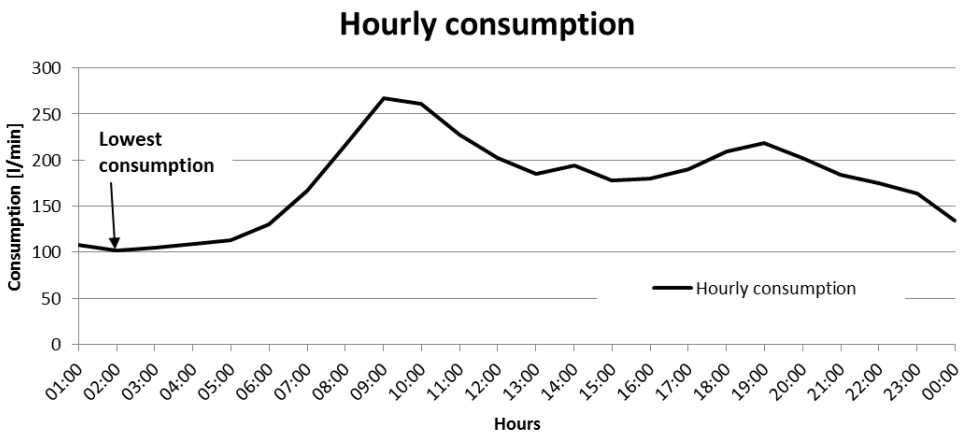

Sensor 3
40
35
30
25
20
15
10
5
0
02:30:00
02:45:00
03:00:00
03:15:00
03:30:00
03:45:00
04:00:00
04:15:00
02:45:00
03:00:00
03:15:00
03:30:00
03:45:00
04:00:00
04:15:00
04:30:00
Hyd1
Hyd3
Measurement
Predictions with demand est.
Predictions without demand est.

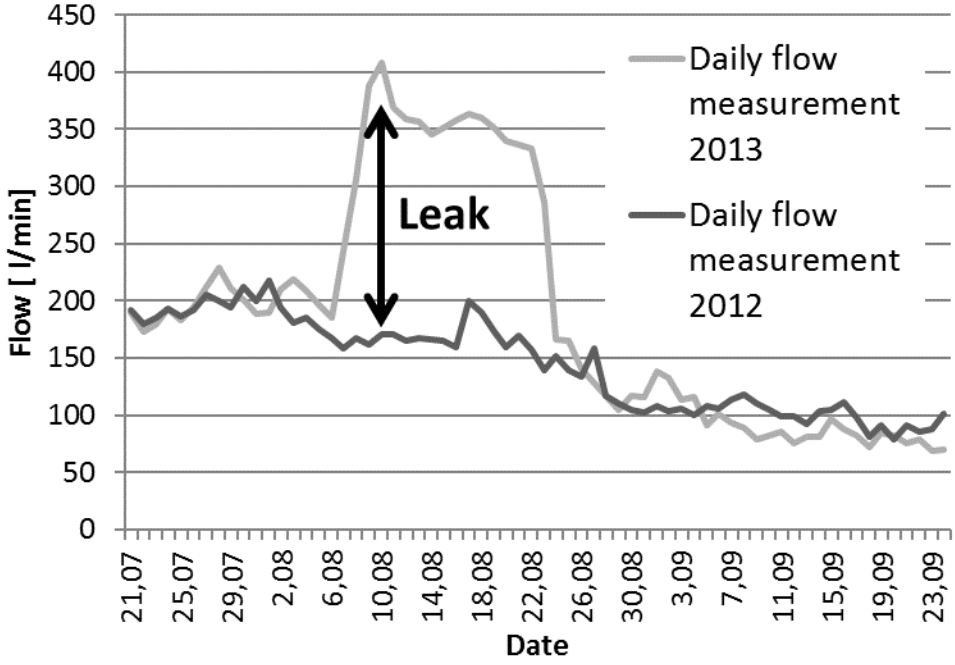

With demand estimation
Without demand estimation
Sensor 1
Sensor 2
Sensor 3
Prediction-measurement [l/min]
Hydrant test
Max(Pred-Meas)
Min(Pred-Meas)
Pred-Meas

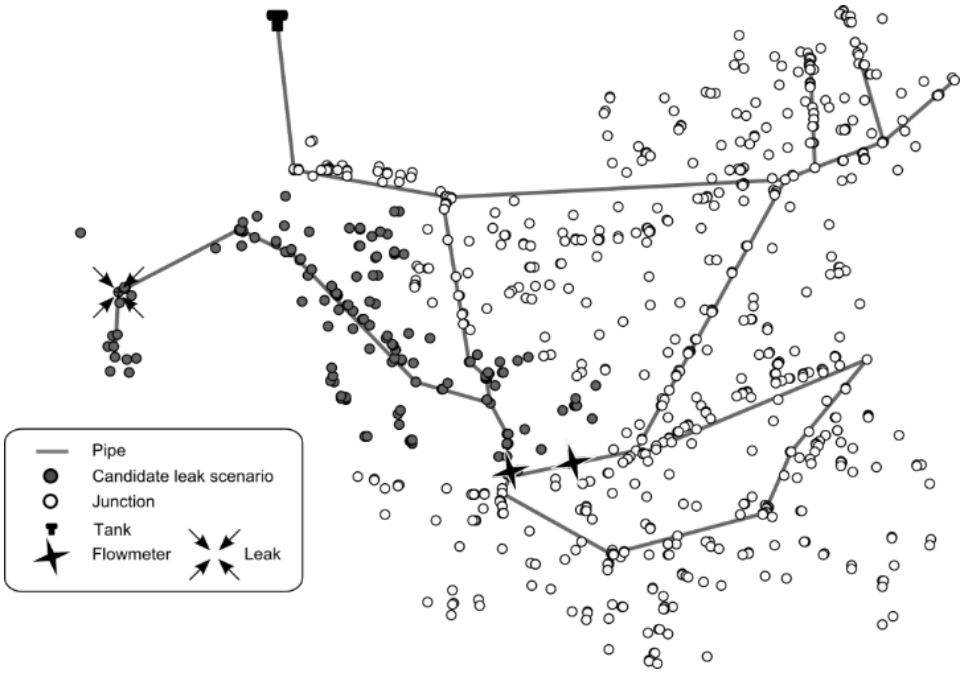

Sensor 1
Probability [%]
20
10
0
-93.89
0
25.93
Prediction-measurement [l/min]
Sensor 2
Probability [%]
20
10
0
-7.48
0
35.62
Prediction-measurement [l/min]
Sensor 3
Probability [%]
20
10
0
-14.75
0
14.17
Prediction-measurement [l/min]

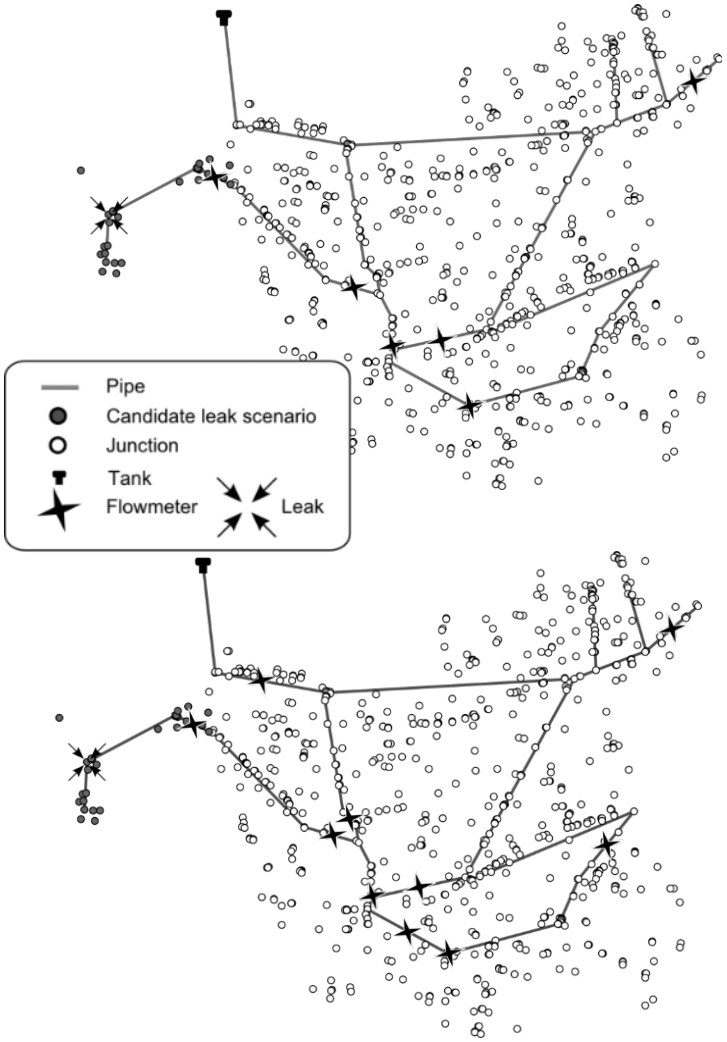

Leak
Candidate leak
Demand node
Sensor
Pipe

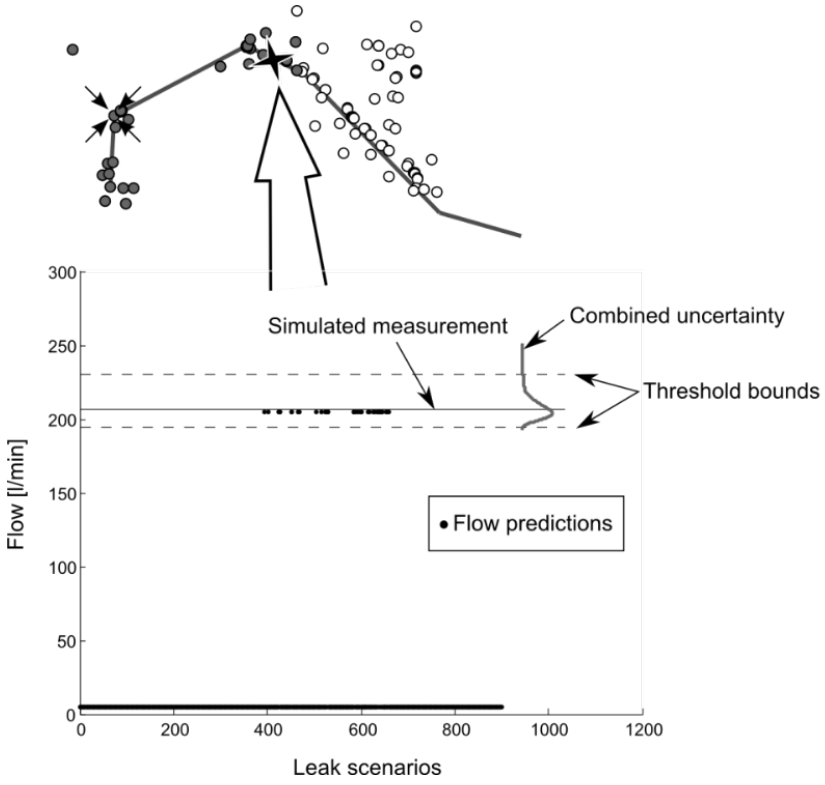

Pipe
Junction
Tank
Flowmeter


Hourly consumption
Consumption [l/min]
300
250
200
150
100
50
0
Lowest consumption
Hourly consumption
01:00
02:00
03:00
04:00
05:00
06:00
07:00
08:00
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
17:00
18:00
19:00
20:00
21:00
22:00
23:00
00:00
Hours


450
400
350
300
250
200
150
100
50
0
Flow [l/min]
Leak
Daily flow measurement 2013
Daily flow measurement 2012
21,07
25,07
29,07
2,08
6,08
10,08
14,08
18,08
22,08
26,08
30,08
3,09
7,09
11,09
15,09
19,09
23,09
Date


Pipe
Candidate leak scenario
Junction
Tank
Flowmeter
Leak


Pipe
Candidate leak scenario
Junction
Tank
Flowmeter
Leak


Simulated measurement
Combined uncertainty
Threshold bounds
Flow predictions
Flow [l/min]
Leak scenarios
300
250
200
150
100
50
0
0
200
400
600
800
1000
1200