Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks

James-A. Goulet[a,*], Sylvain Coutu[b], Ian F. C. Smith[a]

[a]*Applied Computing and Mechanics Laboratory (IMAC)*
[b]*Ecological Engineering Laboratory (ECOL)*
School of Architecture, Civil and Environmental Engineering (ENAC)
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL)
Lausanne, Switzerland

Abstract

Pressurized pipe networks used for fresh-water distribution can take advantage of recent advances in sensing technologies and data-interpretation to evaluate their performance. In this paper, a leak-detection and a sensor placement methodology are proposed based on leak-scenario falsification. The approach includes modeling and measurement uncertainties during the leak detection process. The performance of the methodology proposed is tested on a full-scale water distribution network using simulated data. Findings indicate that when monitoring the flow velocity for 14 pipes over the entire network (295 pipes) leaks are circumscribed within a few potential locations. The case-study shows that a good detectability is expected for leaks of 50 L/min or more. A study of measurement configurations shows that smaller leak levels could also be detected if additional pipes are instrumented.

Keywords: System identification, leak detection, sensor placement, data interpretation, water distribution, uncertainty, error-domain model falsification

1. Introduction

The quantity of fresh water lost due to leaks in water supply networks may reach up to 50% of the input in cases of insufficient maintenance [23, 26]. Leaks involve not only costs; they may also pose a threat to the environment and human health [7, 30]. For these reasons, current research aims to extend the usefulness of computer-aided diagnosis techniques for the detection of leaks in pressurized pipe networks.

Current computer-aided leak-detection techniques can be divided into two categories: *external* and *internal* leak detection systems [1]. Since the 1990's acoustic logging has become the most widely used external detection method [27, 29]. Either vibration sensors or hydrophones are fixed to the pipes to record ambient noise. Leaks are detected when the signal deviates from the normal recordings. Even through these techniques are able to detect small leaks, they usually require a large number of sensors spread over the entire network.

Other external methods such as ground penetrating radar have received an increasing interest in recent years [12, 19]. This non-destructive approach provides cross sectional

*Corresponding author: james.a.goulet@gmail.com

profiles of the soil around pipes in order to detect water leakage. However, its application is time consuming and not suited to large urban areas. Other liquid detection methods use sensing cables buried beside pipes to detect impedance changes when soil gets saturated with fluid [18]. Cables are connected to a central processing system where the data is collected and interpreted. Although accurate, this system has the disadvantage of being invasive.

The second category of leak-detection methods gathers techniques that use continuously monitored data (usually water velocity or pressure) to infer the position of leaks using models. These techniques are referred to as *internal* or inferential methods. One of the first methods was introduced by Ligget and Chen [24] and have since been derived in a number of techniques [8, 9, 13, 14, 22, 28, 33]. These methods are able to take advantage of the interconnectivity of networks to reduce the number of sensors required.

In the field of *model-based data interpretation*, several approaches are calibrating model parameters (e.g. the leak location) by minimizing the discrepancy between predictions and measurements [20, 25]. These approaches are known for their poor predictive capability in case where models contain simplifications compared with the real systems [2, 3, 5]. Approaches such as GLUE [3, 4] and error-domain model falsification [15–17] may be used to identify parameter values such as leak locations using models without having to define completely the error structure associated with model predictions. The error-domain model falsification approach was developed in the field of structural identification. The central idea is to falsify model instances (parameter sets) for which the difference between predictions and measurements are larger than the maximal plausible error. Maximal plausible errors are determined through combining modelling and measurement uncertainties.

This paper builds on the error-domain model falsification methodology to provide a leak detection methodology for pressurised pipe networks. The main objective is to investigate the viability of a detection system at the scale of a city through quantifying the relationship between the number of sensors used and the expected leak-detection performance. Section 2 describes the data-interpretation approach and Section 3 presents a case-study where the leak detection capability is studied using a full-scale water distribution network located in Lausanne, Switzerland. Finally, a discussion of results is provided in Section 4.

2. Methodology

In the context of leak detection, the hypothesis tested is that a leak is occurring at a specific location in the network. Such hypothesis is parametrized in the model of the system as a *leak scenario*. A leak scenario is falsified if the differences between predicted and measured flow velocities in the network are larger than the maximal plausible error, for any measurement location.

Prior to measurement, choices have to be made regarding where to place sensors on the network to most efficiently detect leaks. These decisions are founded on a systematic methodology using simulated measurements. The next subsections describe the system variables (§2.1), provide details on how to falsify leak scenarios (§2.2), and shows how to generate simulated measurements to design optimized measurement systems (§2.3).

2.1. Description of the system variables

When the right values for a vector of model parameters $\boldsymbol{\theta}^* = [\theta_1^*, \ldots, \theta_{\text{max}}^*]$ are known, the predictions returned by a model $\mathbf{g}(\boldsymbol{\theta}^*)$ corresponds to the real quantity $\mathcal{Q}$

plus a modeling error ϵ_{model}. The same happens with observations where a measurement y represents the real quantity $\mathcal{Q}$ plus a measurement error $\epsilon_{measure}$. This relation is expressed in Equation 1.

$$\mathbf{g}(\boldsymbol{\theta}^*) - \epsilon_{model} = \mathcal{Q} = y - \epsilon_{measure} \tag{1}$$

By reorganizing terms of Equation 1, the residual of the difference between a model prediction and a measurement is equal to the difference of model and measurement errors. This relation is expressed in Equation 2.

$$\mathbf{g}(\boldsymbol{\theta}^*) - y = \epsilon_{model} - \epsilon_{measure} \tag{2}$$

The real value for an error ϵ cannot be exactly known. Instead, the probability of error values can be described by a random variable U having a *probability distribution function (pdf)*, $f_U(\epsilon)$. $U_{c,i}$ is a random variable representing the combined uncertainty obtained by computing the difference between modeling and measurement uncertainty sources for a comparison point where predictions and measurements are available $i \in \{1, \ldots, p_{rec}\}$. In the case of pressurized pipe networks, the quantities compared are the fluid velocities in pipes recorded and predicted for p_{rec} locations, where p_{rec} is smaller or equal to the total number of pipes $p_{\max}$. The combined uncertainty represents the expected residual of the difference between predicted and measured values. Techniques available to combine uncertainties are presented in ISO guidelines [21]. Leak scenarios are considered as plausible if the residual outcomes are included in the intervals $[T_{low,i}, T_{high,i}]$. These threshold bounds define the shortest intervals including a target probability $\phi \in]0, 1]$ for the domain $\mathcal{T}$ (see Equation 3).

$$\mathcal{T} = [T_{low,1}, T_{high,1}] \times [T_{low,2}, T_{high,2}] \times \ldots \times [T_{low,p_{rec}}, T_{high,p_{rec}}] \subseteq \mathbb{R}^{p_{rec}} \tag{3}$$

Also, threshold bounds can be conservatively set to be the shortest intervals $[T_{low,i}, T_{high,i}]$ including a target probability $\phi^{1/p_{rec}}$ as presented in Equation 4.

$$\left\{ T_{low,i}, T_{high,i} : \phi^{1/p_{rec}} = \int_{T_{low,i}}^{T_{high,i}} f_{U_{c,i}}(\epsilon_{c,i}) d\epsilon_{c,i} \right\} \forall i \in \{1, \ldots, p_{rec}\} \tag{4}$$

This methodology employs the Šidák correction [32] where the realizations of random variables $U_{c,i}$ have a probability larger or equal to ϕ of simultaneously lying within threshold bounds (see Equation 5). It ensures that the methodology do not wrongly discard a leak scenario with a probability larger or equal to $1 - \phi$. This has been shown to be feasible without requiring the definition of uncertainty dependencies defining the error structure between several comparison points [16].

$$P(\cap_{i=1}^{p_{rec}} T_{low,i} \leq U_{c,i} \leq T_{high,i}) \geq \phi \tag{5}$$

2.2. Leak-scenario falsification

In the following section, $\mathbf{g}_i(\boldsymbol{\theta})$ represents the model of the network where $i \in \{1, \ldots, p_{rec}\}$ corresponds to the pipe number where flow velocity predictions v_i are extracted from the model. $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_{n_{\max}}]$ is a vector containing leak flows for each node location where leak scenarios are simulated and $n_{\max}$ is the number of nodes in the network. An illustrative example of a network is presented in Figure 1. Here, there are $p_{rec} \leq 7$ possible measurement locations and $n_{\max} = 6$ possible leak scenarios, considering that leaks occur at one location at the time.

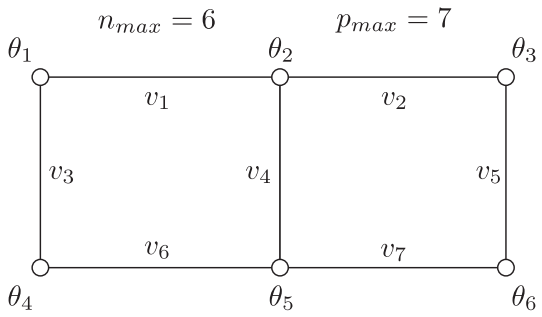


Figure 1: Simple network illustrating the leak flows and and water velocities in pipes. The total number of nodes is $n_{\text{max}} = 6$ and the total number of pipes is $p_{\text{max}} = 7$.

Leak scenarios are generated by simulating a demand of $\theta_j = C$ L/min at a node $j \in \{1, \ldots, n_{max}\}$, when the demand at all other nodes is set to 0, $\forall k \in \{1, \ldots, n_{max}\} \setminus \{j\} : \theta_k = 0$. A leak scenario ($\theta_j \neq 0$) is falsified if the residual of the differences between predicted and measured values is outside threshold bounds for any comparison point $i \in \{1, \ldots, p_{rec}\}$. This corresponds to the situation where Equation 6 is not satisfied.

$$\forall i \in \{1, \ldots, p_{rec}\} : \; T_{low,i} \leq \mathbf{g}_i(\boldsymbol{\theta}) - y_i \leq T_{high,i} \tag{6}$$

Leak scenario falsification is schematically presented in Figure 2, where flow velocity measurements are compared with a database containing p_{rec} flow velocity predictions computed for n_{max} simulated leak scenarios. Based on the model and measurement uncertainties, unlikely leak scenarios are falsified leading to a set containing $CL \in \{1, \ldots, n_{\text{max}}\}$ *candidate leak-scenarios*.

2.3. Measurement system design and simulated measurement

In order to analyze the efficiency of measurement systems, simulated measurements are used to quantify the number of candidate leak-scenarios that are expected to be identified. When a leak is simulated at a node j ($\theta_j \neq 0$), the variation of flow velocity simulated in each pipe due to this leak is stored in a vector $\boldsymbol{v}_{\theta_j}$ of length p_{rec}. p_{rec} is the number of pipes where predictions and measurements are compared.

Based on Equation 1, a vector $\boldsymbol{y}_s$ containing simulated measurements is defined as the difference between the predicted flow velocities v_L for a randomly selected leak scenario and a vector $\boldsymbol{U}_c = [U_{c,1}, \ldots, U_{c,p_{rec}}]^T$ containing realizations of the random variables that correspond to the combined uncertainty for each comparison point. This is presented in Equation 7, where L is a discrete uniform random variable defined between 1 and n_{max}.

$$\boldsymbol{y}_{sim} = \boldsymbol{v}_L - \boldsymbol{U}_c \tag{7}$$

Each instance of simulated measurements $\boldsymbol{y}_{sim}$ is used to emulate the falsification of unlikely leak scenarios according the the methodology presented in §2.2. Each time simulated measurement instances are generated, the number of candidate leak-scenarios CL that remains non-falsified is stored in a vector $\boldsymbol{\Upsilon}_{CL}$. Also, the radius RD including all candidate leak-scenarios is stored in the vector $\boldsymbol{\Upsilon}_{RD}$. After evaluating $\boldsymbol{\Upsilon}_{CL}$ and $\boldsymbol{\Upsilon}_{RD}$ for several instances of simulated measurements (usually several hundreds), the

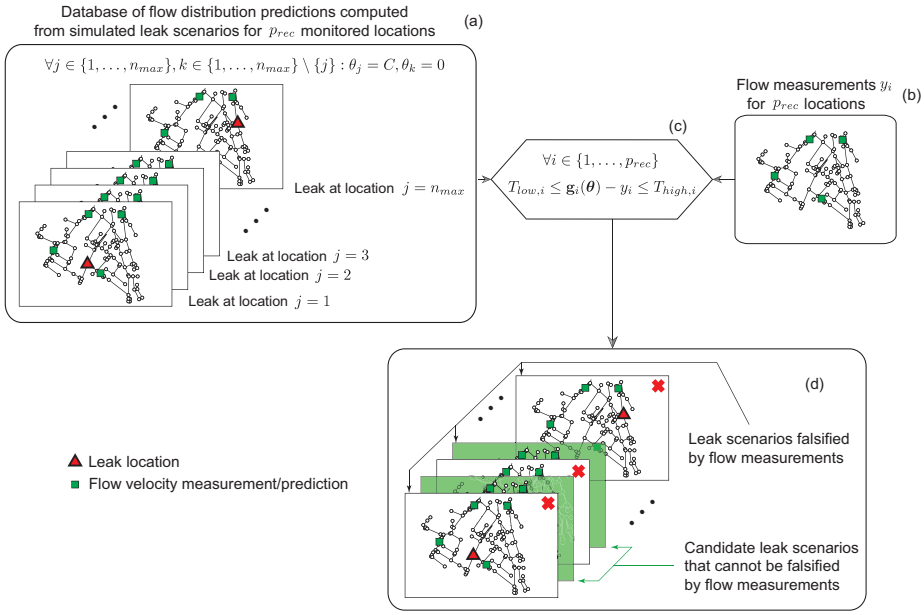


Figure 2: Framework for the identification of leak scenarios compatible with measurements. (a) An initial set of leak scenarios is created by simulating leaks at each location j and predicting flow velocities for p_{max} locations. (b) The network is instrumented with p_{max} flow velocity sensors. (c) Predicted and measured velocities are compared for each location i. Unlikely leak scenarios are falsified leading to a set of candidate leak-scenarios (d).

number of expected candidate leak-scenarios and the expected radius including them are represented by empirical cumulative distribution functions (*cdf*s) $F_{\Upsilon_{CL}}(CL)$ and $F_{\Upsilon_{RD}}(RD)$. These *cdf*s represent, for a given measurement system, the probability of obtaining a maximal number of candidate leak-scenarios and a maximal radius including all leaks. The metric used to quantify the performance of a flow velocity sensor configuration is the inverse *cdf* of Υ_{CL} determined for a probability ϕ_F, $F^{-1}_{\Upsilon_{CL}}(\phi_F)$. In the example of *cdf* presented in Figure 3 there is a probability $\phi_F = 0.95$ or obtaining a candidate leak-scenarios set containing less than 40% of the initial number of scenarios; $F^{-1}_{\Upsilon_{CL}}(0.95) = 40\%$. Examples better and worse expected identifiability are also presented.

For practical applications, it is not possible to test all combinations of sensors. Therefore, optimization algorithms are necessary to reach an acceptable performance while testing a limited number of sensor configurations. For this study, an inverse greedy algorithm [17] is used to find an optimized population of sensor configurations. The optimization method starts by computing the expected number of candidate leak-scenario $F^{-1}_{\Upsilon_{CL}}(\phi_F)$ obtained for the measurement configuration using all $N = p_{rec}$ sensors. After, it iteratively tests each possible combination of $N-1$ sensors to find the set of sensors leading to the lowest number of candidate leak-scenarios. This iterative process is repeated until a single sensor location is left. This algorithm provides a front of optimized sensor configurations in less than $p^2_{\text{max}}/2$ iterations.

The output of the optimization provides multiple solutions having optimized performance and numbers of sensors. Figure 4 illustrates the type of relationship obtained between the number of sensors used and the number of expected candidate leak-scenarios

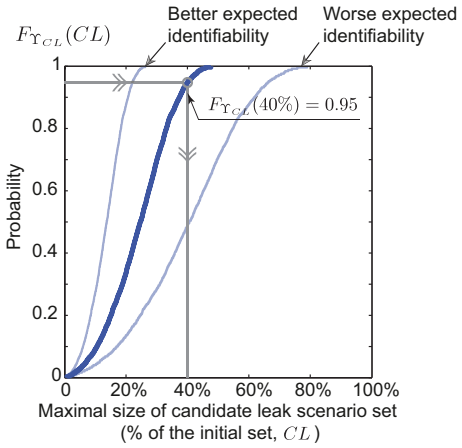


Figure 3: Example of cumulative distribution function $F_{\Upsilon_{CL}}(CL)$ representing the probability of obtaining a maximal candidate leak-scenario set size. In this example, there is a probability of 0.95 of identifying a set of candidate leaks containing less than 40% of the initial set of possible leaks.

$F^{-1}_{\Upsilon_{CL}}(\phi_F)$ computed for a certainty of ϕ_F. Decision-makers may use results presented in this plot to decide which sensor configuration to use based on the performance sought and budget constraints.

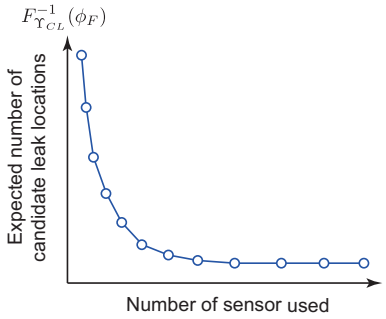


Figure 4: The influence of the number of sensors used on the efficiency of locating a leak within a limited number of locations

Figure 5 presents a flowchart summarizing the steps involved in the evaluation of sensor configuration performance. Note that this flowchart is presented so it is independent of the optimization algorithm used. In the first step, a sensor configuration is chosen. Then the maximal number of candidate leak scenarios $F^{-1}_{\Upsilon_{CL}}(\phi_F)$ is computed using 1000 instances of simulated measurements to emulate the falsification of unlikely leak scenarios. In other similar studies [16, 17] this number of simulated measurements was sufficient to obtain stable empirical cumulative distribution functions $F_{\Upsilon_{CL}}(CL)$ and $F_{\Upsilon_{RD}}(RD)$. When the empirical distribution function is not smooth or diverges significantly from a Gaussian *cdf*, a convergence study may be required to ensure that the number of samples taken is sufficient. If either all or a sufficient number of combinations of sensors have been tested, the algorithm stops and plots the optimal front of measurement configurations. Otherwise, additional sensor configurations are

tested. The criteria used to determine if a sufficient number of combinations of sensors have been tested depends on the optimization algorithm chosen.

3. Case-study

The methodology proposed is applied to the water distribution network of the city of Lausanne, Switzerland. This network is made of several independent sub-networks of different sizes. In this study, we focus on one of the sub-networks, shown in Figure 6. A model of this network is created using the software EPANET [31].

This network contains pumps feeding a tank located at the top right of the network. This reservoir provides water to the network made of 295 pipes and 263 junctions (nodes). Consumption varies between 50 and 250 m^3/h, with an average demand of 150 m^3/h. Leaks in the network are reported at nodes in order to avoid adding nodes in the model for the purpose of simulating leaks. This simplification does not influence the end results because this approach look for possible regions where leaks may occur rather than looking for their exact leak location. In this study, only one leak is considered at a time. Therefore, there are $n_{max} = 263$ possible leak scenarios.

To identify leaks accurately, flow is monitored when water consumption is low. If measurements are taken during high consumption hours, water demands may be too large to differentiate between what is attributed to consumption and what part of the flow is attributed to a leak.

3.1. Monitoring devices

The accuracy of commercially available ultrasonic flow-meter devices is taken to be $\pm 2\%$ of the measured flow. This study considered these devices because of their non-invasive characteristics and their relatively high accuracy. Considering that a flow sensor can be installed on each pipe, there are $P_{\text{max}} = 295$ possible sensor locations. No measurement data is employed; simulated measurements are generated (see §2.3). This approach is necessary to determine good sensor configurations prior to sensor installation.

3.2. Uncertainties

There are several sources of uncertainty associated with the flow velocity model, the measurements and the water network itself. All uncertainties are represented by random variables described as follows.

Secondary-parameter uncertainties are those having a direct influence on the network flow-velocity model. In water distribution networks, common secondary-parameters uncertainties are: node elevations, pipe diameters, minor losses, roughness coefficients and water demand.

The uncertainty in the elevation of nodes is described using a Gaussian distribution with a mean value of 0 mm and standard deviation of 50 mm. Such a level of accuracy on the node position can be obtained using local non-invasive measurements such as ground penetrating radars to accurately locate pipe depth [6, 19]. Uncertainties in pipe diameters are described by a Gaussian distribution with a mean value of 0 mm and a standard deviation of 5 mm. Uncertainties in minor-loss coefficients are described by a Gaussian distribution with a mean value 1.8 and standard deviation 0.1. This value is taken from specifications found in the EPANET documentation [31]. Uncertainty in roughness coefficients is represented by a Gaussian distribution having a mean value of 1 mm and a standard deviation of 0.1 mm. Uncertainty in the water demand at each

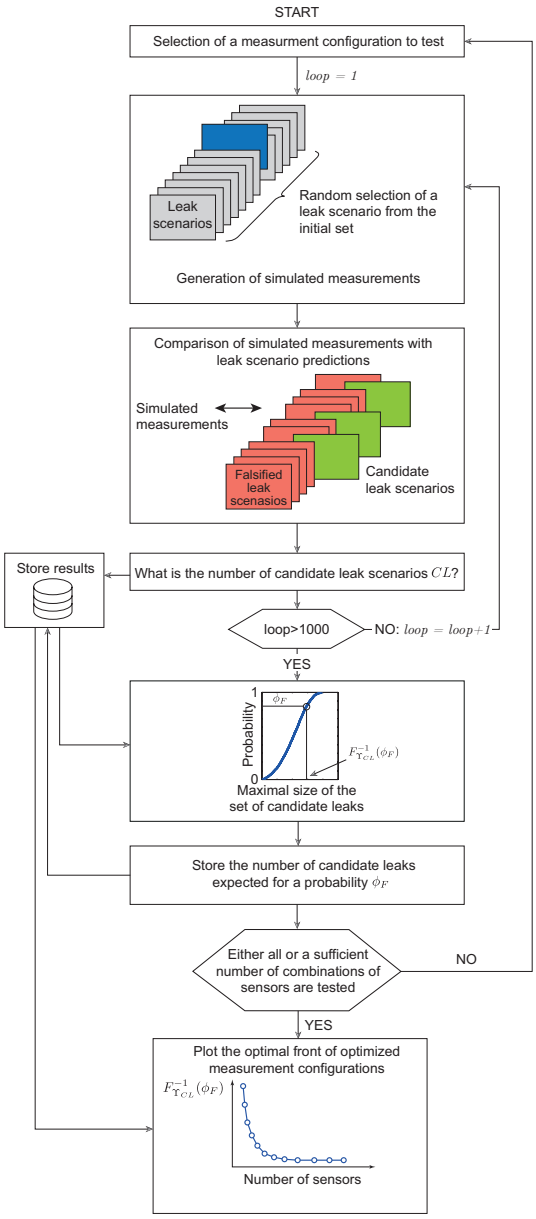


Figure 5: Flowchart summarizing steps involved in the determination of the number of candidate leak-scenarios expected for a given monitoring system.

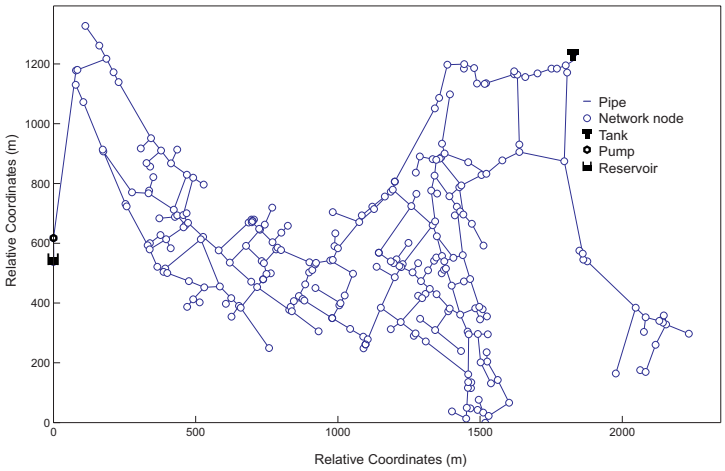


Figure 6: Schematic representation of the water distribution network studied

node is modeled by an exponential distribution with a mean equal to the minimal water demand on the entire network divided by 263 nodes. This distribution is chosen to represent the situation where most consumptions are low and where there are few high consumption locations. The total instantaneous minimal water demand in the network is fixed at 25 m^3/h. This flow demand is based on hourly averaged flow recordings made over a year on the network. The minimum hourly demand is of approximately 50 m^3/h. The values for hourly averaged demand is presented in Figure 7.

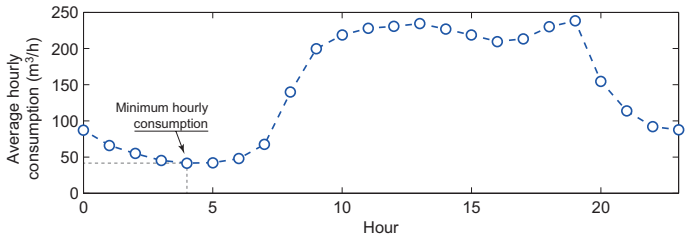


Figure 7: Typical hourly averaged water consumption measured over one day

These uncertainty sources lead to uncertain parameters within the model of the network. Their influence on predicted flow velocities is found by propagating the uncertainties related to parameter values in the model to obtain the uncertainties associated with predicted velocities for each pipe. The sensor resolution uncertainty is taken as a uniform distribution having as lower and higher bound $\pm 2\%$ of the measured value. The uncertainty associated with model simplifications is represented by an extended uniform distribution (EUD [16]) having a lower and higher bound of $\pm 20\%$ of the predicted value and a β factor of 0.3. The EUD represents the uncertainty of bound positions as a fraction $\beta \in [0, 1]$ of the initial interval width. Here the model uncertainty is centered on zero because of the close coupling of the system that makes the systematic bias in model prediction difficult to estimate. Additional uncertainties attributed to other minor sources are also modeled by an extended uniform distribution. The lower and

higher bounds have a value of $\pm 1\%$ of the predicted value and a parameter $\beta = 0.3$. For the purpose of generating simulated measurements, uncertainties are assumed to be independent due to the close coupling of the system. Table 1 presents a summary of uncertainty sources included in the analysis.

Table 1: Summary of uncertainty sources.

Uncertainty source	Distribution type	Distribution parameter	Unit
Node elevation	Gaussian	mean $= 0, \sigma = 50$	mm
Pipe diameter	Gaussian	mean $= 0, \sigma = 5$	mm
Minor loss coefficient	Gaussian	mean $= 1.8, \sigma = 0.1$	-
Pipe roughness	Gaussian	mean $= 1, \sigma = 0.1$	mm
Water consumption	Exponential	mean $= 25$	m^3/h
Sensor resolution	Uniform	$\pm 2\%$	% of the measured flow
Model simplifications	EUD	$\pm 20\%, \beta = 0.3$	% of the mean predicted flow
Additional uncertainties	EUD	$\pm 1\%, \beta = 0.3$	% of the mean predicted flow

3.3. Optimization of measurement-system configurations

Simulated measurements taken from potential leak scenarios are generated to design an efficient measurement-system and to test the applicability of the approach. Iteratively, a random leak scenario is chosen from the 263 possible scenarios. For the purpose of designing the monitoring system, the leak level is taken to be C =100 L/min. The flow velocity predictions associated with each simulated leak scenario are transformed in simulated measurements to emulate the model falsification process. Examples of simulated measurements are presented in Figure 8, where squares represent the location of flow measurements, circles with a cross are the simulated leaks and filled circles are the candidate leak-scenarios.

For this network, it is not feasible to perform an exhaustive search to find the optimal measurement-system because more than 10^{88} combinations of sensors are possible. The inverse greedy algorithm [17] is used to find optimized configurations of sensors. Figure 9 presents the expected number of candidate leak scenarios $F^{-1}_{\Upsilon_{CL}}(\phi_F)$ for each optimized sensor configuration found. Here, the target $\phi_F = 0.95$. In Figure 9, the vertical axis corresponds to the expected number of candidate leak scenarios and the number of sensors used is plotted on the horizontal axis.

In Figure 10, the expected radius for all leak scenarios is presented for the sensor configurations tested. The vertical axis is the expected radius within which all potential leaks are included. For both the number of candidate leak-scenarios and their radius, the expected performance increases rapidly with the number of sensors used until it reaches an asymptote. For engineering purposes, a good tradeoff between the expected performance and the number of sensors used is reached with 14 sensors. The measurement configuration is presented in Figure 11 where the 14 sensors chosen are represented by squares.

Using the optimized sensor configuration found, the expected number of leak scenarios is studied for several levels of leak flow. This expected identifiability of the monitoring system is presented in Figure 11 for leak levels of 100, 75, 50, and 25 L/min. The cumulative distribution function for each leak level is shown in Figure 12.

For high certainty levels ($\phi_F = 0.95$), the expected number of leak scenarios identified remains low for leaks under 75 L/min. For lower certainty levels ($\phi_F = 0.50$) good results are expected up to 50 L/min. The sensor placement optimization procedure is repeated for a leak level of 25 L/min to test if by increasing the number of measurements, the performance may be increased. Results are presented in Figure

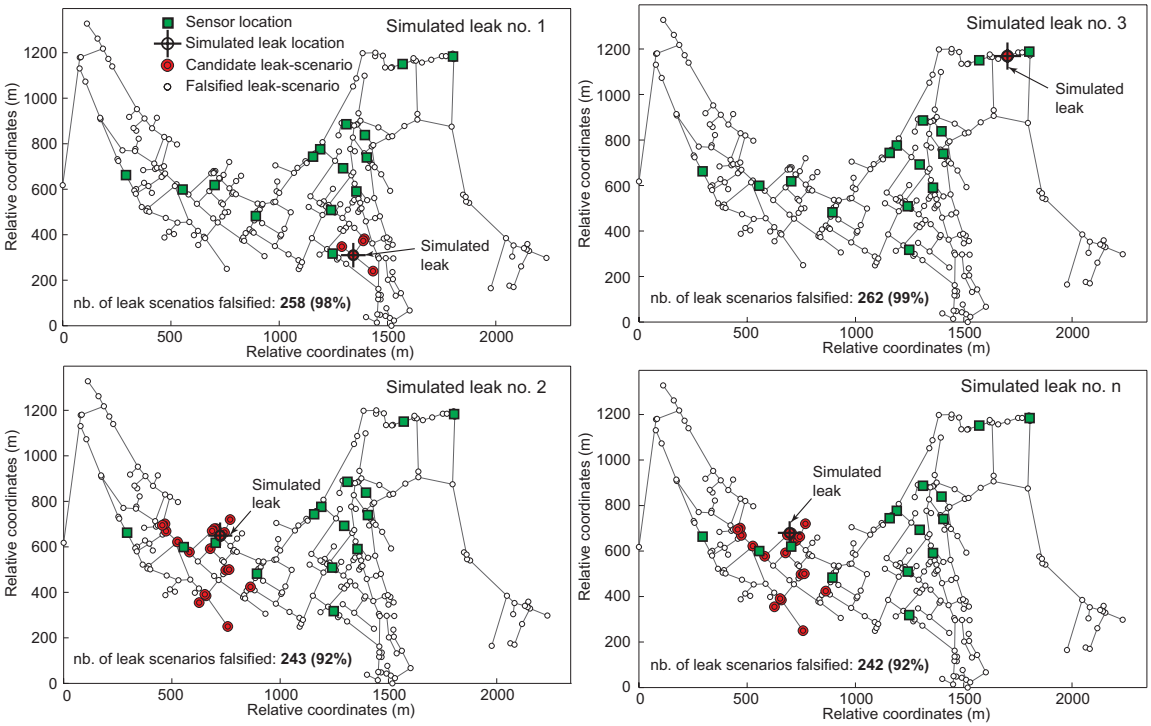


Figure 8: Examples of simulated measurements for Lausanne fresh-water distribution network.

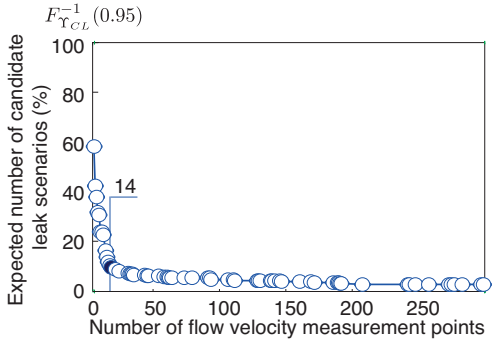


Figure 9: Relation between the expected number of candidate leak-scenarios and the number of flow measurements used.

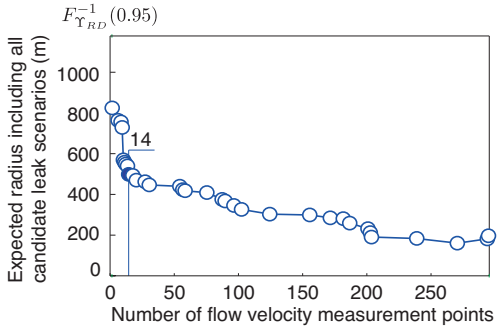


Figure 10: Relation between the radius including all candidate leak-scenarios and the number of flow measurements used.

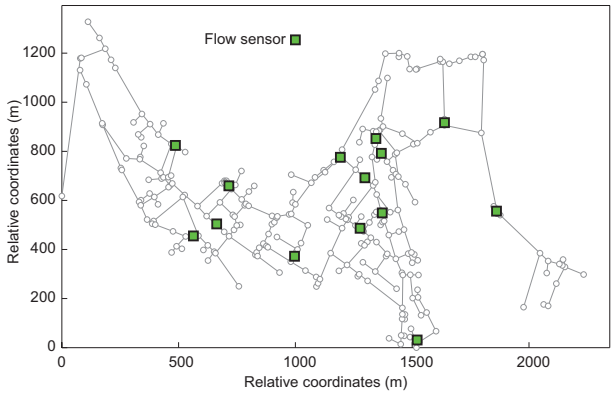


Figure 11: Optimized sensor configuration using 14 flow measurements.

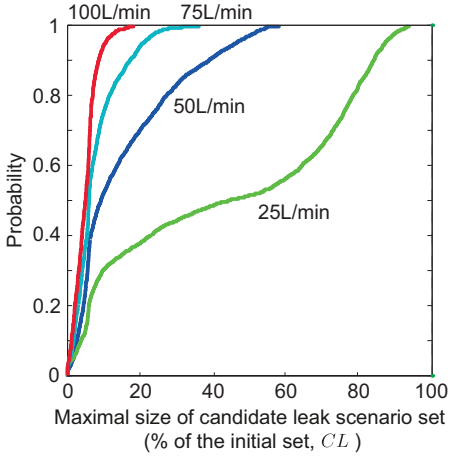


Figure 12: Expected number of candidate leak-scenarios identified for several leak intensities.

13. In this case, when monitoring 100 pipes, the number of expected leak scenarios may be reduced almost by half compared with the situation where only 14 sensors are used. Therefore, efficiently locating leaks in the water distribution network for a leak of 25 L/min is feasible if a sufficient number of pipes can be instrumented.

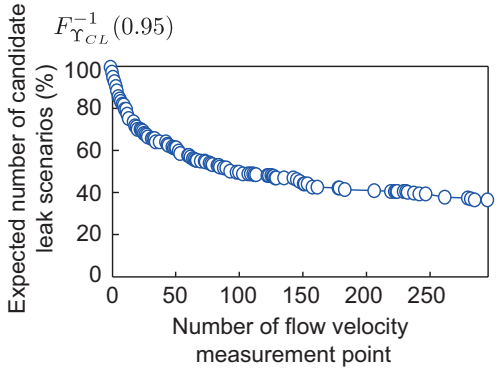


Figure 13: Relation between the expected number of candidate leak-scenarios and the number of flow measurement points used, for a leak level of 25 L/min.

If the identification of lower leak flow levels is required, an option could be to reduce uncertainties associated with the model and measurements. The most important uncertainty sources are water consumption and model simplifications. If these uncertainties are reduced, better performance is expected for lower leak levels. Therefore an important challenge for future work is to quantify more accurately uncertainties associated with water consumption and flow distribution models.

4. Discussion

This work presents a model-falsification methodology for detecting leaks in water distribution network. The falsification approach locates candidate leaks by falsifying improbable leak scenarios. Local investigation techniques such as acoustic detection systems can be used to further refine the actual leak location from candidate scenarios. The approach is expected to be able to identify the location of a leak of 50 L/min using 14 sensors. For lower leak flows, the detection performance of this approach may be improved for instance, by increasing the number of sensors on the network. Also, there is the possibility to couple the information gained using computer-aided data-mining of past breakage history with this leak detection strategy in order to increase its performance [10, 11].

In this paper, the methodology was tested with scenarios where leaks are always located at a single location. Applying the methodology to spatially separated leaks would require more research to maintain the leak detectability to the level obtained in this paper. There is additional complexity associated with searching in a space of possible leak combinations that grows exponentially with the number of simultaneous leaks. Moreover, in addition to detect leaks, this methodology may as well be used for other purposes such as the quantification of demands across networks.

Future work should compare the performance of the measurement system design using the greedy algorithm with other stochastic search approaches. Also, the perfor-

mance of the leak identification approach could be compared in laboratory with other identification methods using reduced-scale benchmark pipe networks.

Conclusion

The leak-detection methodology is able to detect leaks in fluid-distribution pressurized pipe networks and it is able to support the design of measurement systems. In the case of the Lausanne water distribution network, based on simulated data using only 14 flow velocity measurements, leaks of 100 L/min in a radius of 500 m are expected to be detected. The approach is also expected to be able to identify efficiently the location of leaks of 50 L/min. For lower leak flows, the performance of this approach can be improved, for example, by increasing the number of sensors in the network or by reducing modelling and consumption uncertainties.

Acknowledgement

Part of this research is funded by the Swiss National Science Foundation under contract no. 200020-117670/1. The authors acknowledge the support from the City of Lausanne and more particularly from Sébastien Apothéloz and Ibarrola Aitor.

References

[1] ADEC. Technical review of leak detection technologies – vol. 1 – crude oil transmission pipelines. Technical report, Alaskan Department of Environmental Conservation, Alaska, USA, 2000.

[2] ASME. *Guide for verification and validation in computational solid mechanics*. ASME, 2006.

[3] K. J. Beven. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2):18–36, 2006.

[4] K. J. Beven and A. Binley. Future of distributed models: Model calibration and uncertainty prediction. *Hydrological processes*, 6(3):279–298, 1992.

[5] K.J. Beven, P.J. Smith, and J.E. Freer. So just why would a modeller choose to be incoherent? *Journal of hydrology*, 354(1-4):15–32, 2008.

[6] U. Boniger and J. Tronicke. Improving the interpretability of 3D GPR data using target–specific attributes: Application to tomb detection. *Journal of Archaeological Science*, 37:360–367, 2010.

[7] A.F. Colombo. Energy and costs of leaky pipes: Toward comprehensive picture. *Journal of Water Resources Planning and Management*, 128(6):441–450, 2002.

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

[9] D. Covas, I. Stoianov, J. Mano, H. Ramos, N. Graham, and C. Maksimovic. The dynamic effect of pipe-wall viscoelasticity in hydraulic transients. Part II-Model development, calibration and verification. *Journal of Hydraulic Research*, 43(1):56–70, 2005.

[10] D.P. de Oliveira, J.H. Garrett, and L. Soibelman. A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Advanced Engineering Informatics*, 25(2):380–389, 2011.

[11] D.P. de Oliveira, D.B. Neill, J.H. Garrett Jr, L. Soibelman, et al. Detection of patterns in water distribution pipe breakage using spatial scan statistics for point events in a physical network. *Journal of Computing in Civil Engineering*, 25:21–30, 2011.

[12] M. Farley. Finding the difficult leaks. *Water*, Vol. 21:1–80, 2007.

[13] M. Ferrante and B. Brunone. Pipe system diagnosis and leak detection by unsteady-state tests. 1. harmonic analysis. *Advances in water resources*, 26(1):95–105, 2003.

[14] M. Ferrante and B. Brunone. Pipe system diagnosis and leak detection by unsteady-state tests. 2. wavelet analysis. *Advances in water resources*, 26(1):107–116, 2003.

[15] J.A. Goulet. *Probabilistic Model Falsification for Infrastructure Diagnosis*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2012.

[16] J.A. Goulet, C. Michel, and I.F.C. Smith. Hybrid probabilities and error-domain structural identification using ambient vibration monitoring. *Mechanical Systems and Signal Processing*, In press, 2012.

[17] J.A. Goulet and Ian F.C. Smith. Performance-driven measurement-system design for structural identification. *Journal of Computing In Civil Engineering*, (accepted for publication, 2012).

[18] J.M. Henault, G. Moreau, S. Blairon, J. Salin, J.R. Courivaud, F. Taillade, E. Merliot, J.P. Dubois, J. Bertrand, S. Buschaert, et al. Truly distributed optical fiber sensors for structural health monitoring: From the telecommunication optical fiber drawling tower to water leakage detection in dikes and concrete structure strain monitoring. *Advances in Civil Engineering*, (Article ID 930796):13 pages, 2010.

[19] J. Hugenschmidt and A. Kalogeropoulos. The inspection of retaining walls using GPR. *Journal of Applied Geophysics*, 67(4):335–344, 2009.

[20] C.J. Hutton, L.S. Vamvakeridou-Lyroudia, Z. Kapelan, and D.A. Savic. Uncertainty quantification and reduction (UWS) modelling: Evaluation report. Technical Report D3.6.1 PREPARED 2011.005, PREPARED Enabling Change, 2010.

[21] JCGM. *Guide to the expression of uncertainty in measurement supplement 1: Numerical methods for the propagation of distributions*. Number ISO/IEC Guide 98-3:2008/Suppl 1:2008. JCGM Working Group of the Expression of Uncertainty in Measurement, 2008.

[22] S.H. Kim. Extensive development of leak detection algorithm by impulse response method. *Journal of hydraulic engineering*, 131(3):201–209, 2005.

[23] A.O. Lambert. International report: Water losses management and techniques. *Water science and technology: water supply*, pages 1–20, 2002.

[24] J.A. Liggett and L.C. Chen. Inverse transient analysis in pipe networks. *Journal of Hydraulic Engineering*, 120(8):934–955, 1994.

[25] W. Lijuan, Z. Hongwei, and J. Hui. A leak detection method based on EPANET and genetic algorithm in water distribution systems. *Software Engineering and Knowledge Engineering: Theory and Practice*, 114:459–465, 2012.

[26] R. Mamlook and O. Al-Jayyousi. Fuzzy sets analysis for leak detection in infrastructure systems: a proposed methodology. *Clean technologies and environmental policy*, 6(1):26–31, 2003.

[27] R. Pilcher. Leak location and repair guidance notes and..... the never ending war against leakage. *Proceedings Water Loss*, 2, 2007.

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

[29] A. Prodon, S. DeNegre, and T.M. Liebling. Locating leak detecting sensors in a water distribution network by solving prize-collecting steiner arborescence problems. *Mathematical Programming*, 124(1):119–141, 2010.

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

[31] L.A. Rossman. *EPANET 2: users manual*. US Environmental Protection Agency, Cincinati, OH, 2000.

[32] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62:626–633, 1967.

[33] J.P. Viotkovskyo, A.R. Simpson, and M.F. Lambert. Leak detection and calibration using transients and genetic algorithms. *Journal of Water Resources Planning and Management*, 126(4):262–265, 2000.

Notation

N	Number of loops
$\mathcal{Q}$	The true value for a quantity
T	Threshold lower and upper bounds
$\mathcal{T}$	Multidimensional domain where threshold bounds are defined
U	Uncertainty source described by a random variable
$\mathbf{g}(\dots)$	Model of the water network
n_{max}	Number of nodes in the network
p_{max}	Number of pipes in the network
p_{rec}	Number of pipes measured
v	Predicted flow velocity
y	Measured flow velocity
y_{sim}	Simulated measured flow velocity
$\boldsymbol{\Upsilon}_{\mathbf{CL}}\&\boldsymbol{\Upsilon}_{\mathbf{RD}}$	Vectors containing respectively the number of candidate leak scenarios and the radius including all leaks for each simulated measurement instance.
Υ	Random variable describing quantities used during the computation of expected identifiability metrics
ϵ	Error instance
θ	Physical parameter of a model. For leak detection it corresponds to leak flow
ϕ	Target probability content $\in]0, 1]$
ϕ_F	Target certainty used as metric to quantify the performance of measurements $\in]0, 1]$
f_X	Probability density function (*pdf*) of a random variable X
F_X	Cumulative distribution function (*cdf*) of a random variable X
F_X^{-1}	Inverse cumulative distribution function of a random variable X