Advanced Engineering Informatics 27 (2013) 261–269

Contents lists available at SciVerse ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

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), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[b] Ecological Engineering Laboratory (ECOL), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 9 August 2012
Received in revised form 27 October 2012
Accepted 7 January 2013
Available online 8 February 2013

Keywords:
System identification
Leak detection
Sensor placement
Data interpretation
Water distribution

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.

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 1990s 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 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

* Corresponding author.
E-mail address: james.a.goulet@gmail.com (James-A. Goulet).

1474-0346/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aei.2013.01.001

Nomenclature

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
$\mathbf{g}(\cdots)$	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
Υ_{CL} and Υ_{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

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 pressurized 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 (Section 2.1), provide details on how to falsify leak scenarios (Section 2.2), and shows how to generate simulated measurements to design optimized measurement systems (Section 2.3).

2.1. Description of the system variables

When the right values for a vector of model parameters $\boldsymbol{\theta}^* = [\theta_1^*, \ldots, \theta_{\max}^*]$ 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 Eq. (1).

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

By reorganizing terms of Eq. (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 Eq. (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 Eq. (3)).

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

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 Eq. (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 Eq. (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].

$$P\left(\cap_{i=1}^{p_{rec}} T_{low,i} \leqslant U_{c,i} \leqslant T_{high,i}\right) \geqslant \phi \tag{5}$$

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