# **ORE** Open Research Exeter

**TITLE**

Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction

**AUTHORS**

Sophocleous, S; Savić, D; Kapelan, Z

**JOURNAL**

Journal of Water Resources Planning and Management

**DEPOSITED IN ORE**

04 June 2019

This version available at

**WR3759**

# Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction

Sophocles Sophocleous[1]; Dragan Savić[2,3] , M. ASCE and Zoran Kapelan[4,5]

**Abstract**

This research article presents a model-based framework for detecting and localizing leaks in Water Distribution Networks (WDNs). The framework uses optimization and systematic search space reduction. The method employs two stages: (a) the search space reduction stage (SSR), and (b) the leakage detection and localization stage (LDL). During SSR, the number of decision variables is

[1] Eng.D. Student, Centre for Water Systems, College of Engineering, Mathematics, and Physical Sciences, Harrison Building, Univ. of Exeter, North Park Rd., Exeter EX4 4QF, U.K. (corresponding author). E-mail: ss694@exeter.ac.uk.

[2] Chief Executive Officer, KWR Water Cycle Research Institute, Nieuwegein, Groningenhaven 3433 PE, Netherlands. E-mail: dragan.savic@kwrwater.nl.

[3] Professor of Hydroinformatics, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, Harrison Building, Univ. of Exeter, North Park Rd., Exeter EX4 4QF, U.K. E-mail: D.Savic@exeter.ac.uk.

[4] Chair and Professor of Urban Water Infrastructure, Faculty of Civil Engineering and Geosciences, Department of Water Management, Delft Univ. of Technology, Delft 2628 CN, Netherlands. E-mail: Z.Kapelan@tudelft.nl.

[5] Professor of Water Systems Engineering, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, Harrison Building, Univ. of Exeter, North Park Rd., Exeter EX4 4QF, U.K. E-mail: z.kapelan@exeter.ac.uk.

reduced along with the range of possible values, while trying to preserve the optimum solution. Then, at the LDL stage, the size and area of a leak is found. The leak localization method is formulated as an optimization problem, which identifies leakage node locations and their associated emitter coefficients. This is achieved such that the differences between the simulated and field observed values for pressure head and flow are minimized. The optimization problem is solved by using a Genetic Algorithm. A model that has already been calibrated at least according to threshold standards is necessary for this methodology. Two case studies are discussed in this paper including a real WDN example with artificially generated data, which investigates the limits of this method. The second case study is a real water system in the United Kingdom (UK), where the method is implemented to detect a leak event that has actually happened. The results suggest that leaks that cause a hydraulic impact larger than the sensor data error can be detected and localized with this method. The real case outcome shows that the presented method can reduce the search area for finding the leak to within 10% of the WDN (by length). The method can also contribute to more timely detection and localization of leakage hotspots, thus reducing economic and environmental impacts. The optimization model for predicting leakage hotspots can be effective despite the recognized challenges of model calibration and the physical measurement limitations from the pressure and flow field tests.

**Keywords:** Leakage; Optimization; Hydraulic Modelling; Water Distribution Networks

## Introduction

Leaks often remain undiscovered, regularly resulting in large losses of water and revenue for Water Distribution Network (WDN) operators. With time, their impact in the WDN grows and can

provoke catastrophic bursts, which can alter the system's operation causing devastating consequences for the customers and the utility. In many WDNs, losses from leaks are estimated to account for up to 30% of the total volume of extracted water (Puust et al., 2010). This comprises a vital amount in a world struggling to satisfy water demands due to growing population and climate change. Leakage events differ as distinction can be made between bursts, leaks and 'background' leaks. Bursts are normally reported by consumers, as they became visible early following the pipe failure. Yet, the leaks that do not cause water to come up at the surface, can remain undiscovered for a long time and ultimately have similar or larger consequences as large bursts. Finding leaks at an early stage can save water and prevents small leaks turning into bursts, which is important to a water company for economic, environmental and reputational reasons.

Historically, finding leaks has been challenging because even a substantial event can potentially show no manifest signs (Casillas Ponce et al., 2014). A wide range of leak detection and localization techniques exists (Puust et al., 2010), however, there is no universally agreed methodology for finding leaks with the number of techniques currently used by practitioners being limited. The traditional approach to leakage control is a passive one, whereby a leak is repaired only when it becomes visible. This approach is time consuming as it relies heavily on customer contacts and surveillance campaigns by the utility. Nevertheless, automatic leak detection requires pressure and flow field measurements. The smaller the monitored area, the easier it is to detect a leak automatically. Consequently, WDNs are often divided into smaller areas, called 'District Metered Areas' (DMAs), which are easier to monitor and control (Moors et al., 2018). It is normal practice to monitor the DMA for detecting leakages by making water balances between the expected and actual water use. In addition, DMAs are useful to identify any unexpected increase of flow during the night, i.e., when consumption is low and hence the majority of the flow recorded

represents leakage (Farley et al., 2008). Still, leakage detection is not easy, due to unpredictable variations in consumer demands, measurement noise and seasonal effects.

To reduce the search area for finding a leak within a DMA, step-testing is often applied, by redirecting flows through systematic valve openings and closures. This technique can reduce the search area, but requires several steps over multiple nights and is not suitable for all DMAs (Puust et al., 2010). Minimizing the search area is only part of the problem after a leak has been detected, hence, to find its exact location a leak localization technique is necessary. Leak localization techniques divide into internally-based and externally-based (ADEC, 1999). Internally-based methods use field sensors to monitor internal pipeline parameters, such as pressure and flow, and infer the position of leaks using mathematical and non-mathematical models. Though, most frequently leak localization is performed using externally-based methods like acoustic logging, listening rods, noise logging, leak correlators and many other (Li et al., 2015). These methods use data from sensors installed or exploring outside of the pipe to pinpoint a leak. They are very accurate for finding leaks on pipelines, however, they take a long time to find a leak in a large search area (Li et al., 2015). Furthermore, acoustic equipment is less effective and requires an increased effort for new pipe materials (e.g., polyvinyl chloride) because of reduced noise transmission (Gao et al., 2005). Even in small DMAs, leak localization is a time-consuming and labour-intensive process.

To expedite the leak localization process a synergy of internally-based methods, such as mathematical modelling, with externally-based methods is needed. Compared to a leak-free situation, a leak causes more flow in a pipe, which results in larger head loss. Consequently, this leads to different pressures within the WDN, compared to a leak-free situation. In a well-monitored WDN, it creates a unique "signature" on pressure and flow data, which can be used to

automatically find a leak's size and location through a calibrated WDN hydraulic model. This reduces the area of interest and facilitates traditional district audits to find leaks leading to large savings in water volume, time and money.

The literature on model-based leak detection and localization methods for WDNs focuses on how to prioritize areas for leak surveys and facilitate pinpointing of leaks. Such research started with the seminal paper of Pudar and Liggett (1992), which formulates the leak detection and localization problem as a least-squares inverse problem. In an inverse problem for leak detection and localization a number of locations in the system are monitored and some network parameters are assumed known. On the other hand, some parameters are unknown (e.g., leaks, demands, unaccounted nodal outflows), which affect the system's hydraulics. These parameters cannot be determined explicitly by direct measurement, or there is no available data for them. The aim of the inverse problem is to determine the values of the unknown parameters, which in this paper represent the leak size and location(s). Hence, they are determined implicitly by comparing available field observations (typically pressures and flows) with simulated outputs. However, due to the non-linear nature of WDN models and the limited availability of field data relative to the number of parameters to be estimated, this results in an underdetermined problem (Savic et al., 2009). Many of these techniques are based on transient analysis (Liggett and Chen, 1994; Colombo et al., 2009) mainly used on a single, grounded pipeline due to the high effect of the system configuration and uncertainty on results. This results in a costly and time consuming process that is computationally demanding, labour-intensive and requires large amounts of high-frequency data. Non-transient model-based leakage localization techniques have been also developed recently (Wu and Sage, 2006; Wu et al., 2010; Farley et al., 2011; Goulet et al., 2013). These approaches analyse the difference between field observations and modelled outputs from leaky

scenarios and signal an area experiencing leakage by assigning emitter flows as a way to simulate leaks. In some cases researchers identified a leak based on pressure residual analysis (Perez et al., 2011; Casillas Ponce et al., 2012; Farley et al., 2013; Ferrandez-Gamot et al., 2015; Soldevilla et al., 2016), however, a single leak was hypothesized. Others used an optimization technique (Savic and Walters, 1995; Kapelan et al., 2002; Ribeiro et al., 2015). In a real system, it would require the optimization of hundreds of emitter coefficients to include all possible leak locations. Based on the same rationale the leak detection and localization was defined as an inverse problem with the task of determining nodal demands and/or demand patterns (Cheng and He, 2011; Sanz and Perez, 2014; Kun et al., 2015; Do et al., 2016). However, an accurate estimation of consumer demands is necessary, because of their impact on pressure variation, unless the leak size is large relative to the inlet (Moors et al., 2018). To date, model-based methodologies have not reached the maturity required for mainstream adoption by water utilities.

The inverse problem is often ill-posed, characterized by the non-uniqueness of the identified parameters. Even in well-monitored systems multiple combinations of decision variable values could produce equally fit solutions, but inaccurate leak localization. This is a consequence of small impact caused by a leak on the pressure and flow data due to a small velocity change in the pipe experiencing leakage or sub-optimal sensor placement. Currently, there is no optimization algorithm that can efficiently and effectively solve a non-linear inverse problem with thousands of decision variables. A solution to this problem can be to reduce the search space without losing optimum solutions. Wu et al. (2010) used inverse analysis to detect and localize leaks, by reducing the problem dimensionality. This was done by specifying the maximum number of possible leaks within a system, therefore reformulating the inverse problem. They, however, did not narrow down the number of candidate leak locations or the range of flow values prior to conducting the leak

detection. Consequently, the specified number of leaks is still searched among the same number of candidate locations and leak sizes. Furthermore, the reduction process is severely subjective based on engineering judgement. Thus, it is easy for an optimization algorithm to converge prematurely in a large search space, which affects the accuracy of the leak localization. An optimization-based method called step-by-step elimination method was proposed in Nasirian et al. (2013) to calibrate the WDN hydraulic model and detect leaks. The approach was tested on hypothetical and laboratory networks, where nodes that were not reported as leaks during the optimization process were eliminated. Although the search domain was reduced, the process requires iterations to converge to a reduced list of candidate locations before the final optimization analysis. In a real WDN model this would be too time-consuming. Moreover, none of the abovementioned methodologies explicitly minimize the risk of eliminating the global optimum solution from the search space after the reduction, i.e., the best solution to the problem. Therefore, if the specified number of possible leaks is lower than the true WDN state, or the true leak location is eliminated due to premature convergence during the first iterations, the solution would miss the global optimum.

This article presents a new model-based approach for detecting and localizing leaks in DMAs based on systematic search space reduction. Through the use of pressure head and flow data and an optimization technique we aim to increase the reliability in localizing leaks in real WDNs, i.e., reduce the number of false positives. In addition, to speed up the leak search process (both computationally and physically). The method uses a simulation-based technique similar to Perez et al. (2011), to analyse the fault signature matrix and considers the rationale of the subjective simplification approach proposed by Wu et al. (2010) to propose a new objective search reduction approach before implementing leak localization. Based on this, we also try to minimize the number

of computations needed relative to the uncertain number of iterations required by Nasirian et al. (2013). A practical simulation-optimization framework is developed, which highlights the leakage area and makes pinpointing of leaks faster by minimizing discrepancies between observations and model predictions. The proposed methodology is applied to an artificial case and a real WDN to demonstrate the typical steps in undertaking leakage localization.

## Methodology

### Overview

The method detects and localizes leaks in DMAs using inverse analysis where unknown leaks are simulated using emitter coefficients as a means to represent leak flow discharge. The leak location(s) and size is determined using an optimization technique, where decision variables include the candidate leak locations and their corresponding emitter coefficients. The quality of a generated solution is evaluated through the simultaneous comparison of the pressure heads and flows measured by deployed sensors with the simulated values from the hydraulic model. The methodology involves two stages: (a) a Search Space Reduction (SSR) stage for reducing the inverse problem size, and (b) a Leak Detection and Localization (LDL) stage for finding leaks. All simulations were carried out using the EPANET Programmer's Toolkit (Rossman, 2000), while MATLAB was used to implement the developed optimization-based algorithm. During the SSR stage the number of decision variables and the range of their possible values is reduced by implementing three steps. This is to avoid unnecessary evaluation of solutions that cause no impact on model fitness. The three steps include a reduction via: (1) Inverse Problem Simplification, (2) Minimum Detectable Nodal Leakage (MDNL) analysis, and (3) Search Space Optimization (SSO).

Then, at the LDL stage another optimization problem is solved, to indicate the size and location of leaks in the WDN. A schematic framework of the presented approach is given in Figure 1.

**Assumptions**

The developed method relies on the following key assumptions:

- Flow measurements from the inlet and outlet meter(s), as well as heads from pressure measurements at an indicative minimum spatial resolution of at least one sensor per 200 properties, are available (this is the current standard employed by the water utility supporting this work). This is collected at a time interval of 15 minutes and a duration of 24 hours to allow a full Extended Period Simulation (EPS) analysis.

- Any leaks identified by the proposed method are located at network nodes and occurred after the model was calibrated according to the acceptable standards (Ormsbee and Lingireddy, 1997).

- The impact of any existing error in calibration, such as the existence of valves with unknown status or pipe friction coefficients, on the modelled outputs following leak detection and localization is minimal or null.

- The leak flow rate remains unchanged in the time horizon of the data used to detect and localize leaks.

**Search Space Reduction Stage**

***Step 1 – Inverse Problem Simplification***

A node is normally considered a potential leak location if there is no available data associated with its actual demand value. Therefore, all nodes in a WDN model (linked to pipes, valves, pumps and

tanks) could be potential leak locations resulting in a vast search space. In reality the majority of leaks happen on pipes, but the model normally assigns aggregated demands to nearby nodes. In a WDN model not all nodes represent areas of aggregated demand. For example, in EPANET modelling software two nodes are added to a pipe section just upstream and downstream of a valve to indicate its boundaries, without assigning any demand to them. A similar process is used to indicate the physical limits of a pumping station. Although it is true that valves can leak due to a weak stem, the losses are normally insignificant relative to the inlet flow, or to the undetected leaks. Step 1 of the SSR stage reduces the number of potential leak locations, assuming that leaks only happen on pipes, thus, all nodes associated with non-pipe components are excluded.

### Step 2 – Minimum Detectable Nodal Leakage Analysis

Depending on the sensor configuration there is limited observable WDN space, i.e., the length of pipes that can be monitored for leakage. If the leak distance from the sensor increases, the detectable flow decreases due to the lower head loss. Furthermore, all devices are accurate within a specified range indicated by the manufacturers. Thus, if the pressure perturbation from a leak is below the device's accuracy range, the event will remain undetectable regardless of the distance from the sensor. Based on the number, location and accuracy range of sensors there is a minimum detectable flow for each location, which establishes a lower bound when detecting leaks. This is done in Step 2, where the Minimum Detectable Nodal Leakage (MNDL) process starts by imposing the boundary conditions for the leak-free scenario and analysing the resulting pressure ($\widehat{p_{0s}}$) at nodes, $s$, where sensors are present. Then, a leak with large flow, $Q$ (specified by the user), relative to the system's average demand over 24 hours is simulated at every potential leak location,

$l$, and the pressure response, $p_{Q,l,s}$, at each location $s$ is recorded. The number of potential leaks is equal to the remaining nodes after implementing Step 1. Leaks are simulated as emitters following:

$$Q = cp^a \tag{1}$$

where $Q$ is the leakage flow rate (l/s); $c$ is an emitter coefficient representing the leak orifice area; $p$ is pressure (m); and $a$ is a pressure exponent commonly set to 0.5 under ideal conditions, in order to simplify the problem. However, $a$ can vary with the pipe material and leak shape and may impact leak localization. The pressure residuals between a no-leak and leak scenario are determined for all possible leak nodes across all the simulation time steps. These differences between the boundary conditions and leak response pressure at sensor nodes are then averaged over all time steps, using:

$$r_{Q,l}(t) = \begin{bmatrix} p_{Q,l,1}(t) - \widehat{p_{01}}(t) \\ \vdots \\ p_{Q,l,s}(t) - \widehat{p_{0s}}(t) \end{bmatrix} \tag{2}$$

Each time, the simulated leak flow, $Q$, is systematically reduced by adjusting the value of $c$ until the residuals across all sensors do not exceed their accuracy range (e.g., typically ±0.1% measurement error of full scale: 0-10 bar). Ultimately, this establishes the MDNL, $Qmin_l$, for each potential leak node in the WDN model.

$$MDNL = \begin{bmatrix} Qmin_1 \\ \vdots \\ Qmin_l \end{bmatrix} \tag{3}$$

In Step 3 the final search domain for detecting and localizing leaks is established, as explained in the next section.

### Step 3 – Search Space Optimization

*Overview*

The Search Space Optimization (SSO) in Step 3 of SSR is divided into two parts, i.e., Part I and Part II (Figure 2). Here, different optimization problems are solved with the ultimate purpose to minimize the number of candidate leak locations and the range of possible flow values. A Genetic Algorithm (GA), that is a computationally fast and elitist evolutionary algorithm (EA), is used (Deb et al., 2002). An EA is inspired by biological evolution and uses mechanisms such as reproduction, mutation, recombination and selection that can deal with the complexity of such non-linear inverse problems. In both parts of SSO, optimization is formulated with a fixed number of decision variables. For any possible leak two decision variables are defined, one for the discrete location and one for the continuous flow value (Wu et al., 2010). In Part I optimization analysis the decision variables include the candidate node locations from Step 2, optimized as a single emitter coefficient. Using only observed flow measurements, the aim is to detect the total water losses in the WDN based on the simplest case scenario when the entire water loss volume is associated with a single leak. The outcome of Part I defines the maximum emitter flow for each potential leak location, which reduces the range of flow values for the subsequent analyses (Figure 2). Moreover, it allows to falsify any candidate nodes with higher MDNL values than the detected losses from Part I. In Part II a second optimization problem is solved with an updated list of candidates and range of flows, to estimate the maximum number of possible leaks occurring in the WDN, i.e., the most complex case scenario. Both flow and head from pressure observations are considered and the outcome defines the minimum emitter flow for each remaining candidate. Consequently, it leads to further reduction in the number of possible flow values for each potential leak location. Based on the output of Part II optimization analyses, the final search domain is

defined, comprised of the reduced list of potential leak locations and the corresponding reduced range of flow values (including a zero leakage value). The optimization problems are solved similarly in both parts and are subject to two different sets of constraints: (1) the set of implicit type constraints considering system hydraulics; and (2) the set of explicit constraints used as bounds for the algorithm solution search space for each decision variable. The optimization in each part (Figure 2) is formulated as:

Search for: $\quad \vec{X} = (LN_i^n, c_i^n); \quad LN_i^n \in J; \ c_i^n \in K; \ n = 1, ..., NLeak; \ i = 1, ..., NIndex$

(3)

Minimize: $\quad F(\vec{X})$ (4)

Subject to: $\quad [0, \ \underline{c^n} \ \leq \ c_i^n \ \leq \ \overline{c^n}]$ (5) $\qquad P > 0$ (6) $\qquad NLdup^n = 0$

(7)

Where: $LN_i^n$ and $c_i^n$ is the index and the emitter coefficient (equivalent to $c$ in Eq.1), respectively, for node $i$ for the possible leak $n$; $J$ and $K$ are the set of potential leak locations and range of possible values, respectively for any possible leak; $NLeak$ is the number of possible leaks to be identified; $NIndex$ is the number of the candidate nodes for any possible leak; $\underline{c_i^n}$ and $\overline{c^n}$ are the minimum and maximum emitter coefficients, respectively for the node $i$ for the possible leak $n$ along with a possible value of 0 that can be selected to indicate a no-leak case; $P$ is the pressure at any WDN node and $NLdup^n$ is the number of the duplicate nodes that are identified as leakage emitters in the same solution. The constraints given by Equations 6-7 are handled by using a penalty function and are necessary for the GA to search for good and realistic solutions. This leads

to avoidance of solutions that may cause negative pressures in the system or those where a node is identified as a location of multiple leaks if $n > 1$.

*Part I*

In Part I a single leak scenario ($n = 1$) is assumed, i.e., $c_i^1$, to only detect the total water losses in the WDN. Therefore, no information is sought about the precise leak location(s). A series of short (user specified) optimization analyses, $Z$, are undertaken, where at each analysis, $z$, the emitter coefficient value, $c_i^1$, is allowed to vary within a specified range, $d_z$ (user specified), of maximum and minimum flow bounds, while zero is not included as an option. The first analysis, $z = 1$, begins with, $c_i^1$, equal to the maximum difference between the observed and simulated system's demand throughout the 24 hours. This includes both the error caused by leakage as well as the modelling and measurement uncertainty. Each solution's $c_i^1$ values are allowed to vary within, e.g., $d_1 = \pm 50\%$, relative to the $c_i^1$ value that initiates the process. The wider range of values at the start allows to compensate for premature convergence caused by large differences in pressures across the WDN that may result in different optimal $c_i^n$ values for the same leak flow. Furthermore, to mitigate the impact caused by the idealized value of $a$ considered here. When a fit solution is found, the emitter of the next analysis is updated with the optimal solution and the range of adjustment, $d_{z+1}$, is reduced to fine tune the solution for detecting the total water losses. For example, if the optimal solution in analysis, $z$, is fitter than the previous, $z - 1$, the identified emitter is used to specify $c_i^1$ for the next one, $z + 1$. For subsequent analyses the bounds of the $c_i^1$ initiating the optimization process are systematically reduced to eventually, $d_Z = \pm 1\%$ (e.g., $\pm 25\%$, $\pm 10\%$, $\pm 5\%$ and $\pm 1\%$,), relative to the optimum solution of the previous analysis which

allows for both global and local exploration of the search space. The optimization minimizes the

weighted sum of squared flow difference, given by:

Minimize: $\quad F(\vec{X}) = \sum_{t=1}^{T}[\sum_{f=1}^{F} \frac{w_f \left( \frac{Qo_f(t) - Qs_f(t)}{Qpnt} \right)^2}{F}]$

(8)

Where: $Qo_f(t)$ and $Qs_f(t)$ are the observed and simulated flows (l/s), respectively, in link $f$ at

time $t$; $Qpnt$ is the measurement error (l/s), which converts flow differences into a dimensionless

value based on the meter accuracy so that flow error can be compared to pressure error, given by:

$$Qpnt = \bar{Q}_{in} Accuracy_Q$$

(9)

$\bar{Q}_{in}$ is the average global system demand; $Accuracy_Q$ is the flow reading percentage error as

indicated by the meter manufacturer; $F$ is the number of metered pipes, used to mitigate the impact

caused by great differences in the number of available pressure and flow measurements; $w_f$ is a

weighting factor for observed flows, which is used to differentiate the value of information gained

by the measurements during different times of the day, such as the hours of low demand. Following

all optimization analyses the optimal $c_i^1$ corresponds to a leak flow value (l/s), which represents

the global water losses without considering the local modelled pressure variation, which still does

not match the observations. The leak flow value is used in Equation 1 to calculate the final emitter

coefficient based on the average pressure obtained from all sensor measurements and assuming an

average value of $a = 0.5$ across the system. As it is based on a single orifice, it establishes an

upper bound flow, $\overline{c^n}$, for all potential leaks. Any potential leak node with a larger MDNL than

the leak flow, $Q_{\overline{c^n}}$, as a result of $\overline{c^n}$ is removed from the search space. Therefore the potential leak locations are redefined, producing the final list of candidate nodes.

*Part II*

In Part II a series of $n$ leak scenarios are simulated to estimate the maximum number of possible leaks occurring in the WDN and ultimately define the minimum emitter flows for the remaining leak locations. The total number of leak scenarios, $NLeak$, is derived from $Q_{\overline{c^n}}$ and the average MDNL from the remaining candidates, as:

$$NLeak = \frac{Q_{\overline{c^n}}}{MDNL\ avg.}$$

(10)

Optimization analyses are carried out to minimize the weighted sum of squared differences between observed and simulated values for both heads and flows, given by:

Minimize: $\quad F(\vec{X}) = \sum_{t=1}^{T}[\sum_{s=1}^{S} \frac{w_s\left(\frac{Ho_s(t)-Hs_s(t)}{Hpnt}\right)^2}{S} + \sum_{f=1}^{F} \frac{w_f\left(\frac{Qo_f(t)-Qs_f(t)}{Qpnt}\right)^2}{F}]$

(11)

Where: $Ho_s(t)$ and $Hs_s(t)$ are the observed and simulated heads (m), respectively of node $s$ at time $t$; $Hpnt$ is the measurement error (m), which converts head differences into a dimensionless value based on the sensor's reading accuracy, given by:

$$Hpnt = \overline{P_s}Accuracy_p$$

(12)

$\overline{P_s}$ is the average observed pressure at sensor nodes; $Accuracy_p$ is the sensor reading percentage error as indicated by the manufacturer; $S$ is the number of sensors, used similarly as $F$; $w_s$ is a weighting factor for observed heads, used similarly as $w_f$; All flow-related symbols are similar as in Part I formulation. The fittest $n$ scenario establishes the minimum emitter, $\underline{c^n}$, for the candidate leak locations as the total leakage from the identified leaks in the optimal solution should equal the total water losses, $\overline{c^n}$, determined in Part I. The range of flows for any node with larger MDNL than $\underline{c^n}$ is adjusted within its MDNL and $\overline{c^n}$. The process flow chart for SSO is presented in Figure 2.

**Leak Detection and Localization Stage**

At the LDL stage, another inverse problem is formulated, similarly to Step 3 of SSR stage, to detect and localize the emitter flow values in the WDN. The optimization searches for a maximum of $NLeak$ possible leaks based on the fittest, $n$, scenario resulting from Part II of the SSO process. The decision variables include the nodes remaining in the list of leak locations $J$ and the range of their flow values, following Part II. The range of $c_i^n$ values, involves a vector of equally separated discretized values between the upper, $\overline{c^n}$, and lower, $\underline{c^n}$, bounds established in Step 3 of SSR, also including the value of zero for no leakage. Again, the flow and pressure head observations are considered during the optimization analysis to minimize Equation 11. If necessary this stage can be combined with hydraulic model calibration to identify any existing errors carried from previous calibration procedures. It is possible to extend the list of decision variables, by adding calibration parameters such as valves or groups of pipes to optimize the status and friction coefficients, respectively (Wu and Sage, 2006).

## Case Study Applications

### Overview

The leakage detection and localization procedure was implemented in two real WDNs. The first case presents a semi-real case (a real network and simulated observations) undertaken to test the robustness of the methodology with a changing leak size and measurement reading accuracy. The second example presents an application of this methodology on a real leak event based on field data that was captured from deployed flow and pressure devices.

### Semi-Real Case Study

#### *Experimental Design*

A series of desktop experiments was conducted to detect and localize a leak of varying size for a range of sensor reading accuracy values. The aim was to determine the robustness of the methodology and establish the minimum leak flow that could be detected and localized. A number of scenarios based on a single leak of varying size was created where the location of the leak and the sensor devices remained constant. An artificial set of pressure and flow observations was generated for each scenario after an EPS analysis in EPANET considering the system state containing a leak. A number of nodes and the inlet pipe were chosen as locations for monitoring pressure and flow, respectively. This was determined based on the current practice in pressure sensor deployment across the United Kingdom (UK), where 1 logger is installed for every 200 households. The artificial measurements were generated for leak sizes of 50%, 33%, 25%, 20%, 10% and 5% relative to the inlet, here named as flow scenarios. For each flow scenario four datasets were created, resulting in 6x4 scenarios. The first included perfect measurements, whereas the other datasets corresponded to pressure measurements with random noise. The noise was

generated considering a uniform distribution resulting in values ranging between ±0.1%, ±0.25 and ±1% of the reading, here named as noise levels.

## *WDN Hydraulic Model Description*

This example is based on a real city DMA in the UK, which is highly looped and is fed by one inlet. The WDN model (Figure 3) is composed of 158 pipes of total length 9.4km and 202 junction nodes, which in reality would be searched for leakage during an Active Leakage Control Campaign. The average WDN pressure is 25.3m with a maximum difference of 6.9m across parts of the WDN. During a leak-free situation flow from the source node varies between 1.3l/s at MNF and 7.8l/s at morning peak demand, with an average of 3.6l/s without any leakages. The DMA has flow and pressure sensors at the inlet, and seven inner pressure sensors whose placement is also marked in Figure 3 using red circle symbols. Generated field measurements were recorded every 15 minutes for a period of 24 hours totalling to 96 simulated field data from midnight to midnight. Each data set represents a complete snapshot of the system conditions, the observed inflow into the DMA and the eight pressures used for evaluating the quality of leak detection and localization.

## *Leak Localization Problem Set Up*

A leak was introduced at J-33, shown by the blue star symbol in Figure 3. The location at J-33 was chosen to occur on a looped part of the WDN and with loggers placed both upstream and downstream, so that the impact of leak can be observed in most of the pressure data. The method was tested for a leak size of 3.6l/s, 1.8l/s, 1.2l/s, 0.9l/s, 0.65l/s, 0.4l/s and 0.2l/s and four datasets were generated for each flow scenario, based on the noise level. The first set of experiments used perfect pressure and flow data for all flow scenarios. The subsequent datasets involved a systematic

pressure error ranging between ±0.025m, ±0.063m and ±0.250m, respectively, per recorded time step for each device. On the other hand, the typical error range of ±1%, found in flow meter devices, was introduced to the flow measurements of all flow scenarios.

## *Overall Inverse Problem Reduction*

The starting model contained no leakages. Based on the noise level the MDNL differed, leading to a distinct range of possible flow values for each candidate leak location. Hence, this resulted in a different number of candidate nodes for each flow scenario, following Step 3 of SSR. With the end of Part I in SSR the detected water losses for each leak scenario, always equalled the size of the tested leak and any node with higher MDNL value was excluded. Following Part II, in all flow scenarios the fittest optimization analysis corresponded to a WDN with a maximum one leak. Thus, in all cases the LDL part involved two decision variables. The percentage of nodes that remained in the final list of candidates for each flow scenario and for each noise level is presented in Figure 4. The base case used perfect data and involved 42% (84) of the 202 nodes as the potential location for the single leak across all flow scenarios, because of reduction via problem simplification. As noise level increased or the leak size decreased, the number of candidate nodes, which spread across the WDN, was reduced. This is due to the fact that less accurate data provide a lower flow resolution range and opportunities to find leaks if the uncertainty is similar or more than the head loss caused by the event. The downside is that leaks that cause a small hydraulic impact cannot be detected. Interestingly, at a leak size of 5% (0.2l/s) and a noise level of ±1% (±0.25m) there are no candidate nodes for this WDN. The low head loss caused by the small leak, combined with the high noise level results in an undetectable leak situation and the approach fails to perform any leak detection and localization for that flow scenario and noise level. However, in

all other tested cases the optimum solution J-33 always remained in the reduced search space used for LDL.

### *Leak Detectability*

The mean pressure difference between observations and simulated outputs for each leak scenario and noise level, is demonstrated in Figure 5, as a percentage change relative to the leak-free situation. The data labels illustrate the corresponding pressure drop in metres of water head. The mean pressure difference from the leak-free case decreases with a decreasing leak size. Relative to the base case, the difference in pressure changes caused by different flow scenarios decrease with a reduced leak size and an increasing noise level. Interestingly, the pressure change at a noise level of 1% (±0.250m) is similar for a leak size of 5% and 10%. Compared to the mean pressure perturbation, noise has a small impact at larger flows. At a 50% leak scenario there is a pressure drop of 1.68m, while this changes to 1.72m at a noise level of 1%, which comprises around 2.5% of the perturbation. In contrast, a 5% leak only causes 0.06m pressure change, using perfect dataset, whereas at the typical noise level of ±0.1%, the noise accounts for 25% of the variation which rises to 75% at 1% noise level. Under such case it would be impossible to localize a leak due to the large uncertainty. Finally, the small difference between the pressure changes at lower flows can result in a non-unique problem. As the leak size reduces the observable change in the pressure data also decreases, which makes it harder to find a distinct "signature" defined by the leak location J-33.

Table 1 presents the distances of the identified leaks relative to the true leak location at J-33, following the LDL optimization analysis. The leak area was successfully detected and localized up to the size of 10% (0.4l/s) and at a noise level of 0.25% (±0.063m). The latter corresponds to a mean pressure drop of 0.15m, where noise accounts for approximately 40% of the variation (Figure 5). This demonstrates that it is possible to detect and localize a leak if the contribution on the pressure drop is larger than the variation caused by the device's noise. Even with perfect data at a flow of 5% the pressure impact is so small, i.e., less than ±0.3%, which makes difficult to identify a unique pressure signature for J-33 relative to the node detected 42m away. The approach fails to detect any leak at a size of 10% or less and with ±1% noise level as noise impact is larger than the leak for the pressure perturbation. At the current noise level of contemporary logging devices the approach can narrow down the search area for finding a leak of up to 5% within ±2% distance relative to the WDN mains length. This shows that the methodology can find the area of the leak even with consistent noise in the data. From a leakage campaign context in real city WDNs this reduces the leak search distance range from 100% of the DMA to a street level, which can contribute to earlier pinpointing of the leak on the pipe. At a larger noise level the uncertainty was so high that the recorded pressure variation was mainly caused by the reading error and not as a result of the leak. This suggests that more logging would be required for improved leak localization.

**Real District Water System Case Study**

*System Overview*

The methodology was tested on a real DMA (Figure 7) from a UK water utility, under a real leak event, considering that the WDN had historically high levels of leakage recorded. The system is fed by a single source and serves a rural area of approximately 1,000 properties. Flow from the source node normally varies between 2.6l/s at MNF and 8l/s during the morning peak demand. The hydraulic model is composed of 461 pipes of total length 45km and 601 nodes. The DMA has flow and pressure sensors at the inlet, and 10 inner pressure sensors. Typically the utility deploys one pressure logger per 200 properties, however, with the aim to advance in real time modelling of WDNs, the system's logging density was increased to roughly one logger per 100 properties. The DMA contains three Pressure Reducing Valves (PRVs) due to the large range in elevation with the lowest point being at 28m and the highest at 221m, which results in a huge pressure range of 20 and 148 metres of head and an average of 53.8m.

*Leakage Data*

The leak localization process used data from a historical leak, which was already detected and localized by the utility. On November 7, 2016, a leak started at 15:30 hours and lasted for around seven days. The exact location where it was found and repaired by the leak detection engineers is indicated in Figure 7. A systematic flow difference between 2.15 and 2.5 l/s was observed at the MNF throughout the seven days when leak was running, from 2.7 l/s to 5.2 l/s. Moreover, the leak caused a significant head loss around the leak area of more than 30m (Figure 6). Interestingly, a large variation was observed during the morning peak hours, as the leak only caused an increase of approximately 1l/s compared to the day before the leak started. To assess the leak localization

methodology, a calibrated average day model of the system according to the company's standards was used, i.e., a model which simulates pressures at the sensor locations with an accuracy range of ±2m. The observed data for the leak detection and localization part involved the time series for flows into the system and the pressures at 11 locations, on November 8. This was to test whether the leakage localization methodology could report the leak location and flow at an earlier stage. A total of 96 field observation data sets over 24 hour period from midnight of November 8 to midnight of November 9, have been imported into the optimization modelling tool.

### *Leak Search Area Reduction*

Simplification reduced the search space to only 47% of network nodes. Prior to any optimization run for identifying possible leakage hotspots, the system evaluation was conducted by comparing the field observed inflows with the simulated inflows of actual consumptions over 24 hours. This provided a starting point for defining the MDNL flows for each candidate leak location. The MDNL was computed based on a typical reading error of ±0.1% of the utility's devices. Considering the large variation and inconsistency in demand during the morning peak hours and throughout the day, an increased weight was given to the pressure and flow differences during the hours of low demand, i.e., between 00:00 and 06:00. Following Part I of the SSO process a leak of size 2.25l/s was detected, which also corresponds to around 33% of the average inflow readings. The average MDNL, after eliminating the nodes with MDNL values higher than the detected leaks, was approximately 0.14l/s, corresponding to 16 different scenarios for the possible number of leaks. This ranged from a single leak of 2.25l/s up to 16 smaller leaks of 0.14l/s. Part II optimization analysis showed that the most likely leak scenario was the existence of a single large leak in the WDN as this was the fittest scenario. By removing all nodes where the MDNL value

was larger than 2.25l/s the final list of candidates involved only 27% of the network nodes as the potential locations for the leak. LDL was, then, run for finding a single leak, while the setting of the three PRVs was allowed to change by ±10% to emulate the response of a PRV setting to a real leak. The optimum solution indicated a node on the branch where the real leak occurred and 800m downstream (Figure 7). Considering the length of the WDN mains this is an error of only 2% (by length). This suggests that the approach was able to correctly identify the leakage area within a small distance. Furthermore, the search for finding a leak was narrowed down between sensors P6 and P7, which corresponds to a maximum search distance of only 10% of mains length. The leak was found seven days after it happened, thus, the outcome suggests the approach can contribute to a much earlier detection and localization of the leak.

### *Model Calibration*

An EPS analysis was completed with the updated hydraulic model. The simulated inflow after LDL matched observations much better than the original modelled inflow over 24 hours (Figure 8). This is evident especially between 00:00 and 06:00 where a higher weight was assigned. Interestingly, although the leak localization dataset involved measurements on the 8th of November, by comparing the updated inflow data with the inflows between November 9 and November 14, before the leak was found and repaired, an even better match is produced. However, the weekend demand on November 12 and 13 did not match as it follows a different profile, relative to weekdays.

A comparison of the observed and simulated heads at the sensor locations is also given in Figure 9 during the hours of MNF, at 03:00 hours. From the calibration point of view, it is observed that

the error is equal or less than 1% at most sensors, except P6 and P7, where a difference of around 3% exists between the observed and simulated head values. Although this does not represent a fully calibrated model, it serves as a very good baseline for further calibration of the EPS model, following the localization of the leak in the field. Moreover, the approach successfully narrowed down the leak search space and contributed to earlier leak localization. A point to raise considers the fact that PRVs follow a profile variation at their setting throughout the day, a functionality that could not be replicated by the EPANET model used for leak localization. Thus, the profile of the response in the setting of the three PRV's could not be simulated.

## Discussion

## Operational Benefits

The real case application showed that significant savings can be achieved both for the time required to detect and localize a leak, as well as in the lost volume of water. Considering that the leak flow was on average 2.25l/s and ran for a week there were losses of around 1,400m$^3$. If the offline approach was implemented using the data of November 8 and allowed the leak to be found and repaired by the Leakage Technicians on the afternoon of November 9, there would have been volumetric savings in water of more than 70%. Moreover, serious economic impacts can be avoided from costs in fines and insurance payments, and environmental, such as soil erosion. In 2011, Thames Water was fined £8.5m for missing their leakage target (The Guardian, 2018). In addition, a single burst led to insurers to face a £4m payment due to flooding and impacts on transport, homes and businesses (Insurance Times, 2018). Finally, icy weather conditions caused a 4000% increase in Severn Trent's burst alarms and millions of pounds worth of damage (Severn Trent Water Ltd, 2018). On the other hand, a key limitation of this methodology is that it can only

work for new leaks that occur after the model has been calibrated, because background leaks already occurring in the DMA are masked in the calibrated model. Furthermore, the MDNL analysis defines a threshold flow value that is detectable based on the sensor configuration. Therefore, smaller leaks that cause local impact will require additional measurements. Except from that the MDNL analysis provides the water utility an accurate estimate of the level of monitoring a DMA would require for improved leakage control, as it can highlight parts of the WDN where more logging is necessary. Combined with externally-based methods used for pinpointing leaks this methodology provides the opportunity for reducing the time and space required to identify leaks and achieving monetary savings along with higher levels of operational efficiency.

## Computational Benefits

Important benefits were secured through the inverse problem simplification, as unobservable components were removed from the search space. This caused a significant reduction to the number of decision variables and avoidance of unnecessary solution generations during the subsequent optimization analyses. The MDNL flow analysis restricted the range of possible emitter coefficients to only highly sensitive values for the model fitness. To achieve this in the real case study it took 9870 (282 x 35) hydraulic simulations, as the number of the required computations depends on: (a) the number of candidate nodes after simplification and (b) the range of emitter coefficient values after the user specifies the $Q$ (Equation 1). Therefore, the number of solution combinations was minimized. In theory, an over-determined optimization problem including observable parts of the network as decision variables can be solved with a reasonable accuracy.

Further improvements were achieved by using short optimization analyses during the SSR stage. The ability of a GA in considerably improving the objective function value during the initial generations was leveraged for reducing the inverse problem dimensionality. Consequently, following the SSO process the number of combinatorial solutions decreased significantly due to the reduced range of possible values for each decision variable. During Part I, the maximum emitter flow for all candidates was reduced by 22% relative to the maximum difference (2.88l/s) between the observed and simulated system demand, only after five optimization analyses (Population Size: 50, Generations: 30) and a total 7500 solution evaluations. An improvement that none of the previous model-based techniques that use search space reduction achieved. This was based on a single leak scenario, while allowing the possibility for more smaller leaks later on, thus, preserving the optimum solution. The overall effort required to determine the total water losses in Part I relies on $Z\ x\ Population\ Size\ x\ Generations$ hydraulic simulations.

At the end of Part II an approximation of the WDN state was obtained, based on the fittest $n$ scenario (Equation 10), that involved searching for the location of a maximum one leak. This was determined following 1500 solution evaluations for each of the 16 possible leak scenarios (Population Size: 50, Generations: 30), i.e., a total of 24000 solution evaluations in Part II. Eventually, the search domain of the leak detection and localization problem was reduced to $NIndex^{NLeak}x\ (K+1)^{NLeak}$ possible solutions. This was based on an objective approach relative to the subjective approach proposed in Wu et al. (2010) and without perturbing the accuracy. For the real case example discussed, when no reduction in the dimensionality of the inverse problem is undertaken the search space requires $1x10^{601}$ ($K^{NIndex}$) computations to be explored, based on a discretised range of only 10 possible flow values. These are a lot of computations for the algorithm to converge to an optimal solution, which were avoided using the proposed method. The

restriction during the LDL stage to only a number of possible leaks with an additional reduction of the range of possible flow values can contribute to finding leaks earlier, as here, the maximum number of computations for exploring the search space reduced to only 362. The reason lies in the less decision variables and the complexity of the problem, as in practice, cost effective leaks are often associated with a small number of hotspots. Consequently, the initial population of solutions was generated at a better starting point. To verify this, the real case inverse problem (with the known leak) was solved without a reduced search space (i.e., 100% of WDN nodes). The optimization ran for 500 generations and a population size of 200, i.e., 100000 solution evaluations, similarly to the case with a reduced search domain, however, the non-reduced case converged to an incorrect solution. These were carried out on an Intel(R) Core(TM) i7-5600U processor of 2.6GHz speed and 16GB memory. Instead, by integrating the SSR stage only an approximate total number of 42500 hydraulic simulations were needed to solve this leak detection and localization problem. This lasted for about three hours (0.25 seconds per hydraulic simulation), which only allows near-real time applications in a small number of networks. However, this depends on the size of the hydraulic model and the ability of the optimization technique to evaluate solutions. An improvement in computing power can significantly increase these benefits.

## Practicality of the Approach

This method is also applicable to larger WDNs and can detect and localize multiple leaks. However, its accuracy depends on the number and location of available pressure and flow measurements as this can affect the outcome of both the SSR and LDL stages. In addition, to minimize any impact caused by the idealized value of $a$, Zonal Tests can be undertaken (Lambert, Fantozzi and Shepherd, 2018). It is expected that any leak that causes a pressure drop larger than

the uncertainty caused by noise can be detected and approximately localized. The method works best when the number of observations is at least equal to the number of decision variables. Compared to other inverse methods, an advantage of this approach is that it takes into account the uncertainty in observations, which can minimize the number of non-unique solutions and false positives. Furthermore, it reduces the subjective impact coming from the user's choices. Nevertheless, a drawback of the presented method lies in the outcome of Part II during the SSR stage, which may lead to an incorrect estimate of the WDN state. This can occur due to the premature convergence of the algorithm caused by the non-exhaustive optimization analyses. The results suggest that multiple optimization runs for different leakage scenarios prove beneficial for assessing the number of leaks in the network, although in larger systems can be time consuming.

## Model Calibration and Data Challenges

A key requirement for accurate leak localization using this approach and the message from this work is the need to use a well calibrated model of the WDN. However, the traditional methods of calibration, i.e. adjusting pipe roughness coefficients and nodal demands through trial-and-error, have a low chance of meeting the even stricter calibration criterion now required for operational WDN modelling. With the advent of cheaper telemetry and monitoring devices, there are opportunities to further exploit the information captured for estimating decision variables with satisfactory confidence. More data is required due to the increased complexity of large city WDNs and the ill-posedness problem associated with WDN modelling. Although the semi-real case studied a relatively small WDN and used data from the traditional sensor deployment density of 1 per 200 properties, the uniqueness problem was still an issue. This is because the mean pressure change was similar at a smaller leak size. Combined with the effect of noise, various combinations

of solutions led to equally fit solutions and eventually to an incorrect localization. Even using a perfect dataset, at a 5% leak the pressure changes were too small to produce a distinct "signature" that could be used to correctly localize the leak. Significant improvements in the accuracy of the model calibration and leak localization process can be secured with the inclusion of additional flow measurements captured from key flow routes in the WDN. Furthermore, the impact caused by small unknown leaks, or the local effect caused by unknown closed/open valves can be often insufficient to allow detection. An unknown closed/open valve that due to data anomalies may have been left open/closed in the calibrated model, can result in false positive leakage detection. This may be associated with incorrect pipe group roughness values that also exist in same flawed models. Current WDN models are calibrated to simulate observed pressures within ±1m. Thus, hard-to-find leaks and topological anomalies can remain undetected due to small head losses. It is time to move towards a stricter calibration criterion of ±0.1m, equal to the accuracy of field pressure transducers. This will then be able to provide far superior calibration with increased opportunities for more successful detection of previously undetected model anomalies. This will, therefore, allow mathematical optimizers to more clearly distinguish between leak induced head losses and those arising from other faults resulting from theft, unknown status valves, or incorrect pipe sizes and roughness values.

## Conclusions

This article presented an offline model-based methodology that reduces the size of the inverse problem and uses pressure head and flow measurements to detect and localize leaks in WDNs. The approach has been tested on two UK networks using artificial and real leak data, respectively. The desktop case demonstrated the advantages and limitations of this method. The method was then

validated by detecting a real leak event from a UK WDN that run for a week, only using the data from day two. The area for finding the leak on the ground was reduced to only 10% of the WDN (by length), which could have led to savings in water of more than 70%. In both cases presented the optimum solution was preserved in the final search space. A discussion is also provided on the major issues with the traditional calibration of WDN models and the ill-posedness of the calibration problem. The key messages from this research include:

- The methodology proposed is able to effectively detect and localize leaks in a WDN using a calibrated model of the system. This was demonstrated in two case studies where the leaks were successfully identified.

- The method can systematically reduce the number of potential leak locations and the range of possible values, considering the error in observations. This represents an important advantage when compared to similar methods published in the literature as it enables finding leaks in larger, real pipe networks.

- The MDNL analysis establishes a threshold detectable flow value for each candidate leak location, while optimization analyses contribute to accurate detection of the total water losses and an estimation of the maximum number of the leaks in the WDN.

- The method ensures that a good starting point is found, which achieves high computational efficiency leading to early and reliable leak localization.

- The minimum leak size that can be detected and localized is determined by: (a) the pressure change caused by the leaks and (b) the variation caused by the errors/uncertainty in pressure measurement readings. As long former exceeds the latter the leak can be detected and localized.

A future research should assess the robustness of this methodology and its transferability to engineering practice. This can be accomplished through a series of desktop experiments that systematically test the method's accuracy based on different leak size and location, as well as different number and location of sensor devices. It is also of interest to test and validate the methodology on the detection and localization of multiple real leaks. There is also a need to investigate the applicability of this methodology in near real-time. Finally, to quantitatively compare the performance of this method with similar inverse methods that were published previously.

## Acknowledgements

## References

ADEC (1999). "Technical review of leak detection technologies." *Technical report*, Alaska Department of
      Environmental Conservation.

Casillas Ponce, M., Castañón, L. and Cayuela, V. (2012). "Extended-Horizon Analysis of Pressure
      Sensitivities for Leak Detection in Water Distribution Networks." *IFAC Proceedings Volumes*, 45(20), 570-575.

Casillas Ponce, M., Castañón, L. and Cayuela, V. (2014). "Model-Based Leak Detection and

Location in Water Distribution Networks Considering an Extended-Horizon Analysis of Pressure Sensitivities." *Journal of Hydroinformatics*, 16 (3), 649-670.

Cheng, W. and He, Z. (2011). "Calibration of Nodal Demand in Water Distribution Systems." *Journal of*

*Water Resources Planning and Management*, 137 (1), 31-40.

Colombo, A., Lee, P. and Karney, B. (2009). "A Selective Literature Review of Transient-Based Leak

Detection Methods." *Journal of Hydro-Environment Research*, 2 (4), 212-227.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic

algorithm: NSGA-II." *IEEE Transactions on Evolutionary Computation*, 6 (2), 182-197.

Do, N., Simpson, A., Deuerlein, J. and Piller, O. (2016). "Calibration of Water Demand Multipliers in

Water Distribution Systems Using Genetic Algorithms." *Journal of Water Resources Planning and Management*, 142 (11).

Farley, B., Boxall, J. B., and Mounce, S. R. (2008). "Optimal locations of pressure meter for burst detection." *Proceedings on the 10th Water Distribution System Analysis Symposium*, ASCE, Reston, VA, 1–11.

Farley, B., Mounce, S. and Boxall, J. (2011). "Field Validation of Optimal Instrumentation Methodology

for Burst/Leak Detection and Location." *Water Distribution Systems Analysis Conference 2010.*

Farley, B., Mounce, S. and Boxall, J. (2013). "Development and Field Validation of a Burst Localization

Methodology." *Journal of Water Resources Planning and Management*, 139(6), 604-613.

Ferrandez-Gamot, L., Busson, P., Blesa, J., Tornil-Sin, S., Puig, V., Duviella, E. and Soldevila, A. (2015)

"Leak Localization In Water Distribution Networks Using Pressure Residuals And Classifiers." *IFAC-Papers online*, 48 (21), 220-225.

Gao, Y., Brennan M, Joseph, P, Muggleton, J. and Hunaidi, O. (2005). "On the selection of

acoustic/vibration sensors for leak detection in plastic water pipes." *Journal of Sound and*

*Vibration*. 283 (3), 927–941.

Goulet, J., Coutu, S. and Smith, I. (2013). "Model Falsification Diagnosis and Sensor Placement for Leak

Detection in Pressurized Pipe Networks." *Advanced Engineering Informatics*, 27 (2), 261-269.

Insurance Times (2018). "Insurers face £4m payout for south London flood." Available at:

https://www.insurancetimes.co.uk/insurers-face-4m-payout-for-south-london-

flood/1403944.article [Accessed 24 May 2018].

Kapelan, Z. (2002). "Calibration of WDS Hydraulic Models", *PhD Thesis*, Department of

Engineering, University of Exeter.

Kun, D., Tian-Yu, L., Jun-Hui, W. and Jin-Song, G. (2015). "Inversion Model of Water Distribution

Systems for Nodal Demand Calibration." *Journal of Water Resources Planning and*

*Management*, 141(9).

Lambert, A., Fantozzi, M. and Shepherd, M. (2018). "*FAVAD and N1 Update –* *www.leakssuite.com.*"

    [online] Leakssuite.com. Available at: http://www.leakssuite.com/favad-and-n1update [Accessed 6 Aug. 2018].

Li, R., Huang, H., Xin, K. and Tao, T. (2015). "A Review of Methods for Burst/Leakage Detection and

    Location in Water Distribution Systems." *Water Science & Technology: Water Supply*, 15 (3), 429.

Liggett, J.A., and Chen L. (1994). "Inverse transient analysis in pipe networks." *Journal of Hydraulic*

    *Engineering*, ASCE, 120(8), 934-955.

Moors, J., Scholten, L., van der Hoek, J. and den Besten, J. (2018). "Automated leak localization

    performance without detailed demand distribution data." *Urban Water Journal*, 15(2), 116-123.

Nasirian, A., Maghrebi, M. and Yazdani, S. (2013). "Leakage Detection in Water Distribution Network

    Based on a New Heuristic Genetic Algorithm Model." *Journal of Water Resource and Protection*, 5 (3), 294-303.

Ormsbee, L. and Lingireddy, S. (1997). "Calibrating hydraulic network models." Journal of American

    Water Works Association, 89 (2), 42-50.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. and Peralta, A. (2011). "Methodology for leakage

isolation using pressure sensitivity analysis in water distribution networks." *Control Engineering Practice*, 19(10), pp.1157-1167.

Pudar, R.S., and Liggett, J.A. (1992). "Leaks in pipe networks." *Journal of Hydraulic Engineering*, ASCE,

118(7), 1031-1046.

Puust, R., Kapelan, Z., Savic, D. and Koppel, T. (2010). "A review of methods for leakage management in

pipe networks." *Urban Water Journal*, 7 (1), 25-45.

Ribeiro, L., Sousa, J., Marques, A. and Simões, N. (2015). "Locating Leaks with TrustRank Algorithm

Support." *Water*, 7 (12), 1378-1401.

Rossman LA (2000). "EPANET 2, User's Manual, US Environmental Protection Agency." Water Supply

and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH. 45268.

Sanz, G. and Pérez, R. (2014). "Demand Pattern Calibration in Water Distribution Networks." *Procedia*

*Engineering*, 70, 1495-1504.

Savic, D.A., and Walters, G.A. (1995). "Genetic algorithm techniques for calibrating network models."

*Report No. 95/12*, Centre for Systems and Control Engineering, University of Exeter, Devon, U.K.

Savic, D., Kapelan, Z. and Jonkergouw, P. (2009). "Quo vadis water distribution model

calibration?."

*Urban Water Journal*, 6 (1), 3-22.

Severn Trent Water Ltd (2018). "Thank you to all of our customers." Available at:

https://www.stwater.co.uk/news/news-releases/thank-you-to-all-of-our-customers/

[Accessed 24 May 2018].

Soldevila, A., Fernandez-Canti, R., Blesa, J., Tornil-Sin, S. and Puig, V. (2016). "Leak localization

in water

distribution networks using model-based Bayesian reasoning." *2016 European Control

Conference (ECC).*

The Guardian (2018). "Thames Water given maximum £8.5m fine for missing leak target."

Available at:

https://www.theguardian.com/environment/2017/jun/14/thames-water-given-maximum-

fine-for-missing-leak-target [Accessed 24 May 2018].

Wu, Z., Sage, P. and Turtle, D. (2010). "Pressure-Dependent Leak Detection Model and Its

Application to

a District Water System." *Journal of Water Resources Planning and Management*, 136 (1),

116-128.

Wu, Z. Y., Sage, P. (2006). "Water Loss detection via genetic algorithm optimization-based model

calibration." *8th Annual International Symposium on Water Distribution System Analysis*,

Cincinnati, Ohio, August 27-30.

**Table 1.** Comparison of geographical distance of the detected leak for each flow scenario and noise level.

| Leak Size (%) | Distance (m) | | | |
|---|---|---|---|---|
| | 0.00% | 0.10% | 0.25% | 1% |
| 50 | 0 | 0 | 0 | 0 |
| 33 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 | 0 |
| 10 | 0 | 9 | 9 | 420 |
| 5 | 42 | 171 | 171 | N/A |