A model pre-processing approach for improving calibration-based leakage detection using a genetic algorithm

S. Sophocleous[1], D. Savic[2], Z. Kapelan[3], Y. Shen[4], P. Sage[5]
[1,2,3]Centre for Water Systems, University of Exeter, North Park Road, Exeter, EX4 4QJ, UK
[4]Severn Trent Water Ltd, Stratford Road, Longbridge, Warwick, CV34 6QW, UK
[5]WITSConsult Ltd, Milton Rough, Acton Bridge, Northwich, CW8 2RF, UK
[1]*ss694@exeter.ac.uk*

ABSTRACT

The paper presents a systematic approach for narrowing down the search for leaks and unknown closed valves in the water distribution network. The developed approach is applied on a real system and a calibration problem is solved for the ultimate purpose of detecting existing background leakage hotspots. A Genetic Algorithm is used to solve the optimization problem searching for calibration parameter values, while minimizing the differences between observations and model outputs. The optimisation problem is coded in two ways, a scenario-based framework where the maximum number of leaks and closed valves in the network is specified and non scenario-based framework. The leak detection methodology takes advantage of the new pre-processing method to reduce the search space size for the optimisation problems to only significant parameters that contribute to the fitness and hydraulic changes of the model. Artificial calibration data are generated by means of hydraulic modelling employed to mimic planned hydrant discharges during a low demand period. The staged approach demonstrates that the search for location and range of flows for unknown leaks can be reduced to only a small part of the network components. This appears to provide additional benefits towards calibrations problem complexity reduction and reduced time in finding leaks.

Keywords: Hydraulic Models; Leakage Detection; Calibration; Optimisation;

1 INTRODUCTION

Small leaks in water distribution networks (WDNs) often remain undetected, giving way to large amounts of lost water and revenue. Their impact in the WDN grows over time and can result in pipe bursts, having negative consequences for the customers. Detecting leaks within WDNs at an early stage is, thus, of significant importance to a water company. Well calibrated WDN models can be used by performing reliable simulations, "comparing" and analysing the network monitoring data, with the model simulated outputs. Accurate calibration and determination of the WDN state is often associated with the adjustment of model parameters to match simulated pressures within an accuracy range of ±1 metre relative to observations [1]. However, model calibration is based on trial-and-error approaches, due to the lack of major advances from the practitioner's perspective, and a coarse accuracy criterion, compared to the logging accuracy of 0.1%, for supporting operational work at the distribution mains level. A big problem associated with embedded shortfall in the efficacy and calibration of WDN models is the lack of the measurements, i.e. of observed heads and especially, flows from key mains within the WDN. Furthermore, the local demands within the District Metered Areas (DMAs) are often too small to generate significant pipe flows for making accurate measurements with the associated head losses being too small to provide suitable observations for effective calibration. This can, often, lead to the formulation of an ill-posed calibration problem, characterized by non-uniqueness and instability of solutions. Except from that, the widely available and improved information of both the topological representation of the WDN and related boundary conditions has increased the size of the hydraulic model and, as a result, the complexity of network analysis. This has a major impact on system observability, identifiability and, consequently, the ill-posedness of the calibration problem. So far, not much has been done to tackle this problem. Here, we consider a new system for pre-processing the hydraulic model, based on the sensitivitis of model predicted variables and graph theory, with the aim to narrow down the search for leaks in the WDNs and then detect them using an optimization algorithm. The aim is to reduce the inverse modelling problem size and establish a foundation for improved model quality assurance by avoiding unnecessary simulations that impact the model fitness. The approach takes into account suspect valves with unknown status and nodal leakage. The narrowing down approach is applied to a real WDN and the important calibration parameters are highlighted, which formulate the basis of the optimisation problem. A Genetic Algorithm is, then, used to solve the optimization problem searching for calibration parameters values, while minimizing discrepancies between observations and model predictions. The effectiveness of this approach is examined using two different optimisation problem coding approaches. The paper is organized as follows: section 2 provides a literature review on approaches for reducing the inverse modelling problem size, section 3 describes the search space reduction approach and the inverse modelling method, section 4 presents the case study, section 5 compares the optimisation results, followed by conclusions.

2 BACKGROUND

The impact of a leak in a WDN can be modelled by assigning a demand, or emitter to nodes in the system with the aim to match an increase at the inlet flow [2]. Analysing the difference between measurements and modelled outputs from leak scenarios can indicate the probability of a zone to contain leakage. Any parameter associated with an uncertainty in its value is set as candidate for adjustment during the detection process. This can result to an ill-posed problem, due to the larger number of calibration parameters relative to the number of observations. Several authors attempted to reduce the problem size, in order to tackle ill-posedness. Cheng & He [3], applied singular value decomposition to create the sensitivity matrix for the model and then, optimise nodal demands. Goulet et al. [4] proposed a leakage detection and sensor placement methodology based on error domain model falsification. The problem was reduced by falsifying model scenarios for which the difference between predictions and measurements is larger than the maximum plausible error. Nasirian et al. [5] combined a step-by-step elimination method with a GA for calibration and leakage detection, where nodes that provide no contribution in leakage among uncertain parameters of calibration of a WDN were eliminated. Sophocleous et al [6] proposed a graph theory-based partitioning methodology, where the leak detection inverse modelling problem after the WDN is partitioned into different trees, or clusters based on its structural and connectivity properties (i.e., topology and hydraulics). Sophocleous et al. [7] used preliminary topological analysis and sensitivity-based methods to simplify the calibration problem for leak detection purposes.

3 METHODOLOGY

3.1 Artificial Generation of Field Test Data

An artificial set of noise-free pressure and flow observations was generated after an EPANET2 [8] hydraulic simulation analysis considering the "true" system state, i.e., the network with leaks and unknown closed valves. A number of nodes and pipes were chosen for monitoring pressure and flow, respectively, with the observations being used in the calibration process. The dataset was generated to emulate a Night Fire Flow Field Test (NFFFT) situation, where hydrants are flushed during periods of minimum demand to cause a controlled hydraulic stress to the system. Water discoloration risks were taken into consideration with regards to maximum hydrant velocities [9].

3.2 Minimum Detectable Leakage

The Minimum Detectable Leakage (MDL) was determined, based on the configuration of the pressure and flow monitoring devices. A simulation-based framework was applied. The background leakage figure for the WDN was considered as a single leak in a model with no leakages and was simulated at every node. Each time the emitter value was reduced until a point is reached where the head loss from the simulated leak flow across all sensors does not exceed its accuracy range (e.g. ±0.1m). This establishes the MDL for each node in the WDN.

3.3 Narrowing Down Approach to Reduce the Search Space

As a starting point any node with an associated uncertainty in its emitter value is considered as a candidate leak for adjustment during the detection process. This excludes any pressure monitored or flushed locations during the NFFFT as they are assumed to have been checked for leakage. A systematic preliminary analysis of the WDN model was performed to have as few calibration parameters as possible to avoid unnecessary simulation of solutions that do not cause any impact on model fitness. The first step concerns valves in the WDN. Each valve in the model has two nodes associated with it. Assuming that leaks do not happen at valves, but only at demand nodes, all nodes associated with valves were removed as potential leak locations (Figure 1). In the second step all nodes on pipes that cannot supply the MDL flow are removed, as including them in the search space would lead to simulations that result in pressure variations similar, or less than the pressure reading accuracy range. A sensitivity analysis is, then, carried out on the remaining parameters at step three, taking into account engineering constraints. This provides insight to the observability of the different parts of the WDN according to the location of available measurements, while falsifying potential leaks locations that do not meet the constraints. The final step involves a graph theory-based approach. Depth First Search is applied on the remaining nodes to eliminate locations that are close together within a threshold distance, based on the acceptable search range during leakage campaigns, e.g. ±100m. With regards to the candidate unknown closed valve locations, a similar sensitivity analysis was performed to assess the effect of any change in topology and reduce the search space. Furthermore, any branched component, where no pressure measurements at terminal nodes were available were classified as unobservable from the available measurements and were also excluded, as calibration cannot be actually performed. From the remaining valves, only those on loops were included in the search space. This is because, in reality, unknown fully closed valves on any branch of the WDN

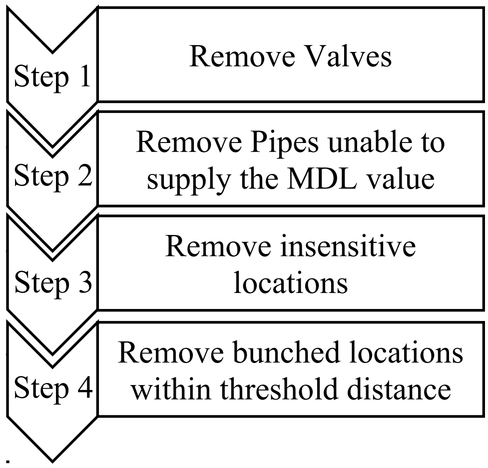

Figure 1. The narrowing down approach for leak

would be sensed by the customer. The remaining valves and nodes were considered as calibration parameters for the optimization problems.

3.4 Optimisation Problem Formulation

A MATLAB optimization code for inverse modelling was linked to the EPANET2 tool-kit. A non-dominated sorting genetic algorithm II (NSGA II) [10], was used where valve status and the emitter coefficients of the candidate valves and nodes, respectively, were considered as decision variables. The calibration was defined as a nonlinear optimization problem with the single objective to minimize the weighted sum of absolute differences between the observed and simulated values for nodal heads and pipe flows. The calibration problem was subject to two sets of constraints: (1) the set of implicit type constraints considering mass and energy balance equations; and (2) the set of explicit constraints used as bounds for the algorithm solution search space for each decision variable. The optimization problem genes were coded in two ways and runs were performed for both cases. The first case listed all potential leak and valve locations as decision variables for calibrating the emitter flows and statuses, respectively. The second involved defining a maximum threshold for the number of leaks and closed valves in the network, i.e. a scenario-based method. This is because in reality leaks and closed valves occur at a small number of locations and not everywhere. The optimization problems were formulated as follows:

Search for: *Case 1:* $\vec{X} = (s_{k,t}, K_i^n)$ *Case 2:* $\vec{X} = (s_{k,t}, LN_i^n, K_i^n)$

$$k = 1,..,NK; i = 1,..,NI; n = 1,..,NGroup$$

(1)

Minimize:

$$F(\vec{X}) = \sum_{t=1}^{T}(\sum_{nh}^{NH} w_{nh}(\frac{Ho_{nh}(t) - Hs_{nh}(t)}{NH}) + w_{nf}(\frac{Qo_{nf}(t) - Qs_{nf}(t)}{NF}))$$

(2)

Subject to: $s_{k,t} \in \{0,1\}$ *(3)* $0 \leq K_i^n \leq K_{max}^n$ *(4)*

Where $\vec{X}$ represents a set of model calibration parameters, s_k, is the status of link k at time step t from a number of candidate links NK, belonging to a vector with values 0 and 1, K_i^n in case 1 is the emitter coefficient for leakage node location i in demand group n for a number of demand groups $NGroup$, or in case 2 for the specified leak n for a number of specified leaks $NGroup$, with 0 and K_i^{max} being the minimum and maximum values the emitter coefficient for group n can take. LN_i^n in case 2 is the leakage node index for the node location i for the specified leak n, NI is the number of the candidate leakage nodes locations in node group n, $F(\vec{X})$ is the objective function to be minimized, corresponding to weighted (W_{nh}, W_{nf}) goodness-of-fit between the field observed and simulated values for NH nodal heads (H_o−H_s) and NF pipe flows (Q_o−Q_s), respectively.

3.5 Inverse Modelling Leak Detection Approach

Two calibration problems were solved to predict system state and status variables as accurately as possible using a well calibrated model. Each problem was associated with the coding framework used in the optimisation process. The results were compared to assess how leak detection and topo-logical calibration (i.e. detection of valve status) performs in terms of accuracy and computational time. Calibration parameters involved the emitters for the candidate leakage nodes for each demand group and the candidate valves for the detection of their initial status. Here, a single demand group was used, as the nodal demand mainly involves domestic consumption, along with a group of emit-ter coefficients determined based on the minimum detectable leakage and maximum leakage based on the background leakage level of the DMA. For the first optimisation analysis all remaining candidates following the narrowing down approach were considered as decision variables, whereas in the second optimisation analysis the scenario-based approach used a maximum threshold of five unknown leak locations and five unknown closed valves, leading to a total 15 calibration parameters. Five optimization run were carried out for each calibration problem. Following optimization it is expected that the simulated model predictions for pressure and flow match the field test data as closely as possible, while all leaks within the observable part of the network have been accurately detected and located. The fittest result was considered as the solution closest to the representation of the true system state. The hydraulic model that was considered for leakage detection assumed all valves are open, except from known closed boundary valves, and that no leaks exist in the network. The following GA parameters were used for the optimisation runs: population size of 200, 500 generations, binary tournament selection operator, random mutation with the probability of 0.1 and single-point crossover with the probability of 0.90.

4 CASE STUDY

4.1 The Water Network

The WDN model layout is shown in Figure 2. It involves a real-life DMA composed of 202 junction nodes, 158 pipes, 59 valves and has one inlet, which is subject to pressure reduction. The total mains' length is 9.4 km. Flow from the source node varies between 2.34 l/s at Minimum Night Flow (MNF) and 8.72 l/s at morning peak demand. Three leakage hotspots were introduced at J21, J35 and J109 leading to a global leakage of around 1 l/s during MNF, or 20 % relative to the average inlet flow. Moreover, two valves were closed. Three hydrants (e.g. HYD1-3), included in the EPANET model as nodal demands, were flushed between 01:00 – 04:30 at flows up to a maximum 6 l/s. Generated field measurements were obtained from six locations (e.g. P1-6) as well as upstream and downstream of the PRV (P7, P8), recording pressures every 15 minutes, while inlet flow was also obtained. A total of 96 data sets over 24 hrs have been used for the calibration process.

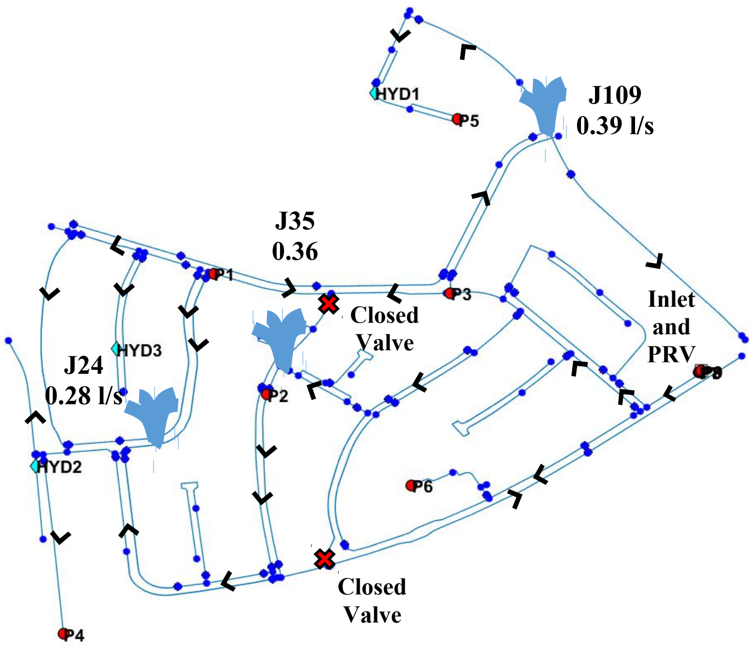

Figure 2. The True system state illustrating the leakage hotspot locations and flows, the closed valves, sensor locations the flushed

5 RESULTS

For this WDN, 30 % of the average inlet flow, i.e. around 1.5 l/s was used as a starting point for the determining the MDL. The narrowing down approach resulted in total 92 % reduction in the population of potential leak locations and 71 % reduction in unknown closed valve locations. Following steps 1 and 2, only 40 % and 31% of nodes, respectively, remained in the search space, after the removal of valves and mainly branched pipes that have small capacities to supply MDL. Sensitivity analysis lead to only 21% of the WDN being a potential leak location, while an introduction of a ±100m distance threshold at Step 4 lead to the final reduced list of candidates. The best solutions from optimisation runs between the non-scenario and scenario-based approaches have been selected for a comparison. These are presented in Figure 3 illustrating the detected leakage hotspots and closed valves. The results were obtained after applying all steps of the narrowing down approach and significantly reducing the problem. Interestingly, following Step 4 the true leak location J35 was removed from the search space due to the distance threshold. The non scenario-based optimisation lead to a best solution with an objective error of *F=1.19*. The two unknown closed valves were correctly identified with no false positives. However, leak detection was not correct although the close match of the model outputs with the observations. Two leaks were identified at a total flow equal to the true water losses. One was located 25m away from the leak J35, while the second was detected more than 300 m away from the closest leak J24. This led to an incorrect detection of emitter values. The best solution was reached after 1081 seconds. On the other hand, when a scenario-based method was used a fitter solution of *F=0.82* was achieved only after 224 seconds. The J109 was correctly detected in terms of location and emitter flow, while the node closest to J35 was also reported as a leak. For the second leak the detected flow did not match the true emitter flows, but the total losses from both leaks, again equalled the true water losses.

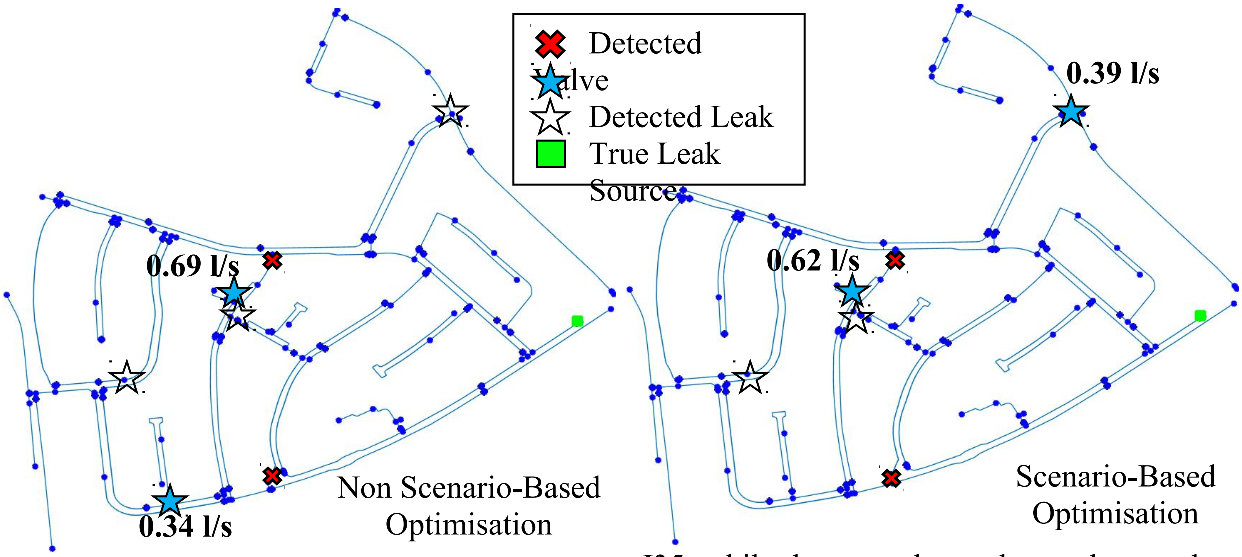

Figure 3. Comparison of best optimisation runs between the two approaches.

After the observation that the correct amount of water losses could be detected an additional search space reduction stage was introduced, where the upper bound for the range of emitter flows was adjusted to equal the maximum detected leakage flow from the best optimization runs, i.e. a total water loss of approximately 0.7l/s. This reduced the range of flows by more than 50%. Moreover, the status of the two frequently detected closed valves was fixed and the optimization was reran. In both cases the best result lead to *F=0.66***,** the lowest achieved even though a true leak location was not in the search space. Three leaks were detected at J24

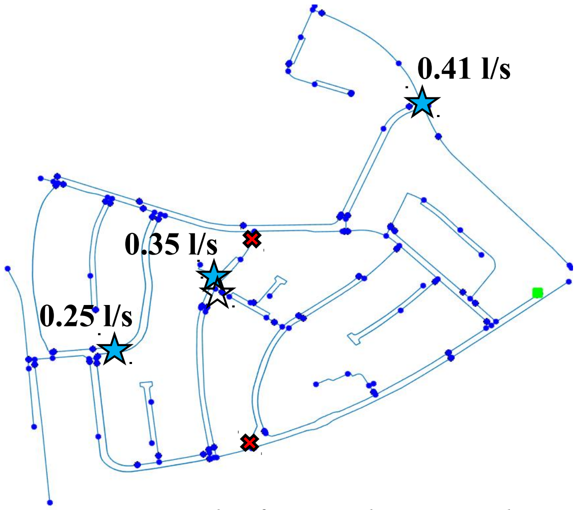

Figure 3. Optimisation result after reduction in the range of flows and fixed closed valves.

Figure 3. Result after a reduction in the range of flows and fixing valve status.

and J109 detected with flows very close to the true leaks (Figure 3). The third leak was detected at the node closest to J35, as the true leak node was removed from the search space after applying step 4.

6 DISCUSSION

6.1 Pros and Cons of Narrowing Down the Search Space

In real systems the inverse calibration problem is often under-determined, due to a larger number of calibration parameters relative to the available measurements, which must be grouped to produce an even/over- determined problem. This often leads to non-uniqueness of the identified parameter values. A non-uniqueness problem occurs when multiple parameter vectors correspond to similar objective function values, i.e. to near-optimum values of roughly equal magnitude, leading to a non-unique solution. This was also observed here, where two leaks, detected at equal total flow compared to the true situation, produced a similarly fit objective function value relative to the best result of $F=0.66$ achieved after the reduction in the range of flows. However, all leaks have a unique pressure "signature" effect on the WDN, which should be identified to correctly localize them. Thus, a well monitored WDN is necessary to accurately detect leakage hotspots and unknown closed valves that cause small and local head losses. In addition, increased flow monitoring through waste meters, or use of prior information can significantly improve the uniqueness of the problem and consequently the detection accuracy. Through the narrowing down approach, which considers topological, sensitivity and graph-based analyses, important benefits were secured, as unobservable components were removed from the search space causing a significant reduction to the number of calibration parameters and avoidance of unnecessary solution generations. The scenario-based approach was able to correctly identify the two larger leaks in the WDN within a small distance and less computations relative to the non-scenario based approach. The reason lies in the less decision variables and complexity of the problem, which was established after specifying the number of maximum leaks and unknown closed valves in the WDN, which resulted in a better starting point for the initial population of solutions for the GA. However, as the above case study demonstrates, the accuracy of result can be significantly affected if true leak locations are removed from the problem during search space reduction. If few observations are available this can lead to false positive detection of leakage hotspots and incorrect identification of emitter flows. On the other hand, in both cases although the true leak J35 was removed the total water losses were identified. A subsequent reduction of the search space associated with the range of flows led to much more accurate leakage localization for both cases. Both approaches have shown that if the solution search space is sufficiently reduced, the optimum, or near-optimum solution can be found. The scenario-based approach, however, seems to provide additional benefits in terms of less computations. This is mainly observed if the search space is increased, due to the fact that the number of decision variables would remain fixed, whereas in a non scenario-based approach the number of calibration parameters will increase. In theory, an over-determined optimization problem including observable parts of the WDN as calibration parameters should be able to be solved with a reasonable accuracy with both approaches. In practice, the synergy between the narrowing down approach and inverse modelling leak detection can significantly reduce the time to find leaks and minimize the chance for supply interruptions.

6.2 Improved Detection Through Better Quality Datasets

Optimisation techniques can contribute to earlier and automated leakage detection if accompanied by sufficient and good quality field data, which is necessary for the accurate determination of calibration parameters. Thus, the impact caused by small unknown leaks, or the local effect caused by unknown closed/open valves would often require large datasets to allow detection due to the measurement noise levels compared to model accuracy. In reality, this comes into conflict with the financial, resource and time constraints faced by water companies. Furthermore, the current calibration accuracy threshold for simulated pressure outputs reduces the chance of detecting hard-to-find leaks and topological anomalies. Introducing known interventions during field tests, such as hydrant flushing during NFFFTs can contribute to better quality datasets. By taking into account discoloration risk, flushing hydrants at key locations during MNF, i.e., when leakage is at its highest value, causes a controlled hydraulic stress on the WDN, while emphasizing the hydraulic impact arising from existing topological and leakage-related anomalies. This can create a more unique pressure signature for the existing situation. The selected locations aim to hydraulically stress as many pipes as possible, increasing the probability to detect the faults. Choosing the locations for the flushed hydrants, as well as the sensor placement based on smart analyses, such as the sensitivity, can provide enhanced opportunities for more successful detection of those previously undetected model anomalies, through the abovementioned inverse modelling methods.

7 CONCLUSIONS

A search space reduction approach was presented that reduces the complexity of the calibration problem and contributes to earlier leak detection through inverse modelling associated with an optimisation method. The approach has been formulated and applied to a real network, where the search for unknown leaks and closed valves was significantly reduced to around 10% and 30% of the network components respectively, appearing to provide additional benefits towards calibrations problem complexity reduction. The method identifies significant parameters that contribute to the fitness and hydraulic changes of the model, while providing a systematic method to eliminate calibration parameters. Discussions were also made on the ill-posedness of the calibration problem and how improved accuracy can be achieved even if true leak locations are removed from the search. The artificial case study has been successfully used to test for the early detection of unknown valve statuses and leakage hotspots. The results suggest that a scenario-based optimisation can bring benefits in both computational time and accuracy, especially in a more complex problem with more decision variables. In practice, the promising approach can be lead to a useful tool for network operations for reducing the time to find existing leaks.

8 ACKNOWLEDGEMENTS

This work is part of the first author's STREAM Engineering Doctorate project sponsored by the UK Engineering and Physical Science Research Council, Severn Trent Water and WITS Consult Ltd.

References

[1] D. Savic, Z. Kapelan and P. Jonkergouw, "Quo vadis water distribution model calibration?", *Urb. W. J.*, vol. 6, 1, pp. 3-22, 2009.
[2] Z. Wu and P. Sage, "Water Loss detection via genetic algorithm optimization-based model calibration", 8th WDSA, 2006.
[3] W. Cheng and Z. He, "Calibration of Nodal Demand in Water Distribution Systems", J. W. Res. Pl. Mgmt., vol. 137, pp. 31–40, 2011.
[4] J. Goulet, S. Coutu and I. Smith, "Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks", Adv. Eng. Inform., vol. 27, pp. 261–269, 2013.
[5] A. Nasirian, M. Maghrebi and S. Yazdani, "Leakage Detection in Water Distribution Network Based on a New Heuristic Genetic Algorithm Model", *J. W. Res. Pr.*, vol. 5, 3, pp. 294-303, 2013.
[6] S. Sophocleous, D. Savic, Z. Kapelan, Y. Shen, P. Sage, "A graph-based analytical technique for the improvement of water network model calibration", Pr. Eng., vol. 154, pp. 27-35, 2016.

[7] S. Sophocleous, D. Savic, Z. Kapelan, O. Giustolisi, "A two-stage calibration for detection of leakage hotspots in a real water distribution network", Pr. Eng, vol. 186, pp. 168-176, 2017.
[8] L. Rossman, "EPANET2 users' manual", Risk Reduction Eng. Lab. Office of Res. and Dev., U.S. EPA, 2000.
[9] J. Boxall, and A. Saul, "Modelling discolouration in potable water distribution systems", *J. Env. Eng.*, vol. 131, pp. 5, 2005.
[10] K. Deb, A Pratap, S. Agarwal, T. Meyarivan, "A fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", IEEE*, Trans. Evol. Comp.*, vol. 6, 2, pp. 182-197, 2002.