

Anomaly Localization by Applying Data-Driven Analysis and Parallel Optimization of Hydraulic Model Calibration [†]

Ashley Hui Zhang ¹, Fred Cao ¹, Alvin Wei Ze Chew ¹ , Zheng Yi Wu ^{2,*}, Rony Kalfarisi ¹, Xue Meng ¹, Jocelyn Pok ¹, Juen Ming Wong ¹, Kah Cheong Lai ³, Lennis Seow ³ and Jia Jie Wong ³

¹ Bentley Systems Singapore Pte Ltd., 3 Harbourfront Pl, Singapore 099254, Singapore; ashley.zhang@bentley.com (A.H.Z.); fred.cao@bentley.com (F.C.); alvin.chew@bentley.com (A.W.Z.C.); rony.kalfarisi@bentley.com (R.K.); meng.xue@bentley.com (X.M.); jocelyn.pok@bentley.com (J.P.); juen.wong@bentley.com (J.M.W.)

² Bentley Systems Inc., 76 Watertown Rd, Suite 2D, Thomaston, CT 06787, USA

³ Water Supply (Network) Department, PUB, Singapore 228231, Singapore; lai_kah_cheong@pub.gov.sg (K.C.L.); lennis_seow@pub.gov.sg (L.S.); wong_jia_jie@pub.gov.sg (J.J.W.)

* Correspondence: zheng.wu@bentley.com; Tel.: +1-2035651740

[†] Presented at the 3rd International Joint Conference on Water Distribution Systems Analysis & Computing and Control for the Water Industry (WDSA/CCWI 2024), Ferrara, Italy, 1–4 July 2024.

Abstract: This paper presents an integrated approach using both data-driven and hydraulic model-based methods to localize anomaly events in near real time (NRT). Upon detecting an NRT anomaly event, the pressure drops at sensor locations are calculated, followed by estimating the pressure drops at junction nodes via an inverse-distance weighted interpolation method. Clustering is then performed based on pressure drops at junction nodes and network topology to segregate and reduce the search areas. Afterwards, a genetic algorithm optimization is performed with hydraulic model simulations to further pinpoint the anomaly hotspots. The integrated method has been tested on real leakage events with field data, where the localized leak hotspots are within 300 m of the ground-truth leaks.

Keywords: water distribution networks; leak localization; nonrevenue water; near real-time operation



Citation: Zhang, A.H.; Cao, F.; Chew, A.W.Z.; Wu, Z.Y.; Kalfarisi, R.; Meng, X.; Pok, J.; Wong, J.M.; Lai, K.C.; Seow, L.; et al. Anomaly Localization by Applying Data-Driven Analysis and Parallel Optimization of Hydraulic Model Calibration. *Eng. Proc.* **2024**, *69*, 6. <https://doi.org/10.3390/engproc2024069006>

Academic Editors: Stefano Alvisi, Marco Franchini, Valentina Marsili and Filippo Mazzoni

Published: 29 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In urban cities, water utility companies supply potable water to users via water distribution networks (WDNs). It is inevitable to have water losses in WDNs such as underground pipe leakages and overflows from storage tanks, which contributes to the nonrevenue water (NRW). Detecting and localizing unreported hidden leaks at an early stage can generally reduce NRW and prevent unreported leaks from developing into service-disruptive water main break events. Recently, the rapid development and reduced cost of the internet of things (IoT) have made it possible for the real-time or near-real-time detection and localization of leaks in WDNs. The centralization of different data sources such as IoT sensors and SCADA systems also makes various types of data (flow, pressure, acoustic, operation conditions, et al.) available for leak analysis in near real-time (NRT). Consequently, there are growing numbers of studies that focus on data-driven approaches to model the complex problem.

Despite decades of efforts in methodology improvements for water loss reduction, we identified that the main challenges as well as trends in leak localization research and software implementation include (1) localizing simultaneous multiple leaks by taking the advantages of both spatial correlation and hydraulic model-based approaches; (2) conducting leak localization based on anomaly event detection results; (3) efficient and effective leak localization in near real time; and (4) validating the effectiveness of leak localization approaches in real systems with real data.

In this study, an integrated approach is developed to combine the benefits and meanwhile to overcome the challenges of both data-driven and model-based approaches for anomaly (leak or sudden demand increase) localization. We target to localize anomaly events (including simultaneous multiple events) in real WDNs with real data accurately and efficiently. The entire localization framework is fully automated and incorporated in the software named Anomaly Leak Finder (ALF). In collaboration with PUB, Singapore National Water Agency, near-real-time anomaly detection and localization are conducted using ALF. This paper focuses on anomaly localization only, and the entire framework from data preprocessing to outlier detection, system leak event detection, and model recalibration is detailed in previous publications [1,2].

2. Methodology

An integrated framework has been developed for anomaly localization, mainly consisting of data-driven and model-based methods, as shown in Figure 1. This framework is implemented in the ALF software with the purpose of not only accurately determining the leak locations but also to facilitating the daily operations of WDNs for leak searches.

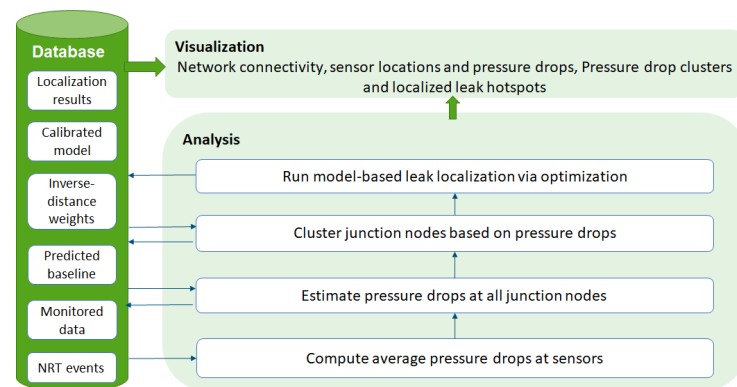


Figure 1. Integrated anomaly localization framework.

Once there is an NRT anomaly event detected by the upstream functions of the software, data-driven localization is triggered. The pressure drops of sensors are calculated based on the NRT monitored pressures and the baseline pressures, and the pressure drops at junction nodes are estimated using inverse-distance interpolation. In the case study with field data, the baseline pressure of a sensor is predicted using the deep belief network with the extended Kalman filter (EKF-DBN) in our software [3], which is able to adapt to the WDN operational changes.

In the above data-driven approach, junctions with the highest estimated pressure drops represent the areas of possible anomaly events or hidden leaks. To facilitate the field search for leaks, it is important to narrow down the search areas and meanwhile to avoid missing possible multiple concurrent leaks. As such, a network topology-based search method is formulated to cluster the pressure drops into individual search areas. The search begins with the junction node having the maximum pressure drop among the available junctions, and the connected chains are traced down until the boundary is reached, when the pressure drop at the next junction starts to increase with a preset tolerance. The visited junctions in all the chains are grouped into a cluster. The available junctions are updated by excluding the clustered junctions, and the search proceeds until the maximum pressure drop in the available junctions falls below a preset threshold.

Then, the hydraulic model-based localization searches for a few leak hotspots and their individual emitter coefficients via genetic algorithm optimization [4]. The leak hotspots are emulated as emitters at nodes where leaks are modelled as pressure-dependent demands in addition to real consumptions. The emitter locations and coefficients at possible leak nodes are the decision variables to be optimized such that the cost function is minimized. To speed up the optimization, the GA optimization is accelerated by integrating Darwin Optimization

Framework (DOF) [5], in which the GA engine is decoupled from the fitness evaluation, so that the manager–worker MPI parallelization scheme is adopted for accelerating the overall optimization process.

3. Applications

The case studies are based on a real WDN in Singapore, as shown in Figure 2a. The WDN consists of 11,410 pipes (total 313.8 km), 9299 junctions, 1875 valves, and 2 service reservoirs. The pipe diameters range from 50 mm to 2500 mm. There are 35 pressure sensors and 1 flow meter at the inlet of the WDN.

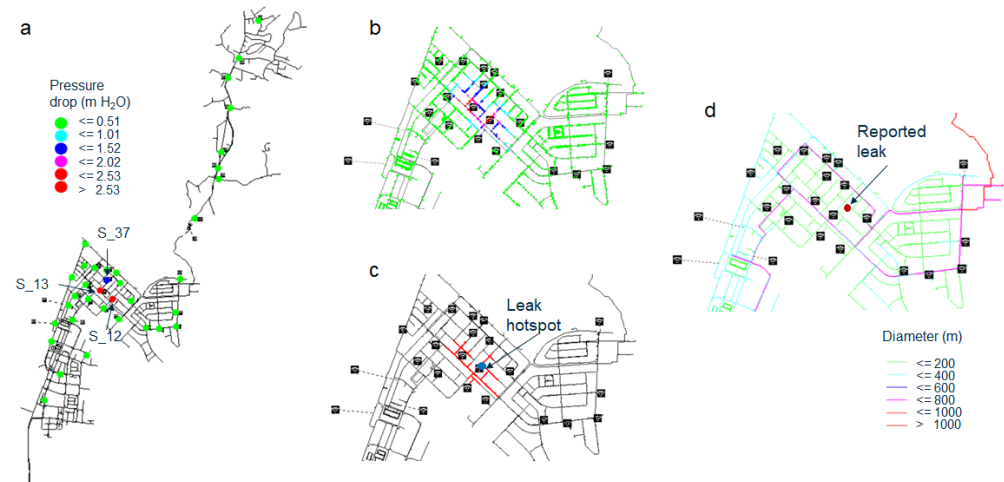


Figure 2. (a) Pressure drops of sensors across the WDN on 16 August 2022; (b) Estimated pressure drops of junction nodes; (c) Leak search areas (red-coloured pipes) after clustering and leak hotspot (blue dot) via Darwin localization; (d) Reported ground-truth leak (red dot) by WDN operators on 19 August 2022.

In this test case with field data, the flow data and pressure data are collected every 5 min and resampled every 15 min. The raw data are preprocessed to deal with data errors such as missing or duplicated time steps. The NRT duration is emulated with the historical data during August–September 2022. There are two events identified by the NRT anomaly event detection in the software that started on 16 August 2022 and 13 September 2022. Table 1 summarizes the localization results. For both detected leaks, the search area after clustering contains the ground-truth leak, and the localized hotspot is within 300 m to the ground-truth leak, which would significantly reduce the leak search time.

Table 1. Localization results of the detected leak events in field tests.

Starting Date of Detected Leak	Date of Reported Leak	Clustered Search Areas Covering the Ground-Truth Leak?	Distance of Localized Hotspot to Ground Truth (m)
16 August 2022	19 August 2022	Yes	250 m
13 September 2022	14 September 2022	Yes	220 m

The localization results for the 16 August leak event are depicted in Figure 2. Figure 2a shows the pressure drops of the three event sensors: S_13 shows the highest pressure drop at 2.53 m H₂O, followed by S_12 at 2.25 m H₂O and S_37 at 1.21 m H₂O. The pressure drops at junction nodes are estimated, and there is an area of high pressure drops near the three event sensors, as shown in Figure 2b. After clustering, there appears a cluster forming the search area of the anomaly event in Figure 2c, consisting of 8.6 km of pipes, which are used as input to further pinpoint the leak hotspot via hydraulic model-based optimization. The leak hotspot in Figure 2c is located near S_12, although the highest pressure drop is at S_13, suggesting good accuracy of the hydraulic model-based localization. The detection and

localization results are verified by a reported leak event on 19 August 2022 near S_12, as shown in Figure 2d, which is only 250 m to the localized hotspot. The leak hotspot serves as a good reference to the starting point of the field search.

Author Contributions: Conceptualization, Z.Y.W., A.H.Z. and A.W.Z.C.; methodology, Z.Y.W., A.H.Z., F.C. and A.W.Z.C.; software, J.M.W., F.C., R.K., X.M. and J.P.; validation, A.H.Z. and A.W.Z.C.; formal analysis, A.H.Z.; investigation, A.H.Z. and A.W.Z.C.; data curation, A.H.Z. and A.W.Z.C.; writing—original draft preparation, A.H.Z.; writing—review and editing, Z.Y.W., A.H.Z., K.C.L., L.S., J.J.W. and A.W.Z.C.; project administration, Z.Y.W., K.C.L., L.S. and J.J.W.; funding acquisition, Z.Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Singapore National Research Foundation under its Competitive Research Program (CRP) (Water) and administered by PUB (PUB-1804-0087), Singapore's National Water Agency.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some data, models, or code generated or used during this study are proprietary or confidential in nature. All the field data including pressures, reservoir outflows, pump configurations and tank levels, customer billing information, historical leakage records, and hydraulic models are confidential and cannot be provided without third-party agreement.

Conflicts of Interest: Author Zheng Yi Wu was employed by the company Bentley Systems Inc. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wu, Z.Y.; Chew, A.; Meng, X.; Cai, J.; Pok, J.; Kalfarisi, R.; Lai, K.C.; Hew, S.F.; Wong, J.J. Data-driven and model-based framework for smart water grid anomaly detection and localization. *AQUA Water Infrastruct. Ecosyst. Soc.* **2022**, *71*, 31–41. [[CrossRef](#)]
2. Chew, A.W.Z.; Wu, Z.Y.; Walski, T.; Meng, X.; Cai, J.; Pok, J.; Kalfarisi, R. Daily model calibration with water loss estimation and localization using continuous monitoring data in water distribution networks. *J. Water Resour. Plan. Manag.* **2022**, *148*, 04022019. [[CrossRef](#)]
3. Li, Q.; Wu, Z.Y.; Rahman, A. Evolutionary deep learning with extended Kalman filter for effective prediction modeling and efficient data assimilation. *J. Comput. Civ. Eng.* **2019**, *33*, 04019014. [[CrossRef](#)]
4. Wu, Z.Y.; Sage, P.; Turtle, D. Pressure-dependent leak detection model and its application to a district water system. *J. Water Resour. Plan. Manag.* **2010**, *136*, 116–128. [[CrossRef](#)]
5. Wu, Z.Y.; Wang, Q.; Butala, S.; Mi, T.; Song, Y. *Darwin Optimization User Manual*; Bentley Systems, Incorporated: Watertown, CT, USA, 2012.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.