Theme: Digital transformation of urban water systems
IAHR Thematic Priority Area: [TPA-4] Digital Transformation
https://doi.org/10.3850/iahr-hic2483430201-522

Integrating Data-Driven and Hydraulic Modelling with Acoustic Sensor Information for Improved Leak Location in Water Distribution Networks

Ignasius Axel Hutomo[1*], Ioana Popescu[2], Leonardo Alfonso[3]

[1,2,3] IHE Delft Institute for Water Education, Delft, Netherlands, Delft, 2611 AX, Netherlands

* *Corresponding author: axelhutomo@gmail.com*

Abstract. Despite ongoing research and practical efforts, water losses in water distribution networks remain alarmingly high, impacting water quantity and quality. Traditionally, water utilities have employed separate approaches using acoustic sensors and hydraulic models to address leaks. However, the integration of these methods and their potential for mutual improvement have not been thoroughly studied. This research proposes a novel approach to improve leak location accuracy by integrating acoustic sensor data and hydraulic modelling within a machine learning framework, using data from an actual use case. Results show that by combining selected acoustic statistical data in time and frequency domains, and various hydraulic features from physical modelling as inputs, a 94% accuracy in the leak location of leaks above 1 L/s can be achieved. This represents a substantial improvement relative to the accuracy achieved by the acoustic methods alone (84%) or by the hydraulic modelling data alone (64%).

Keywords: Acoustic Sensor, Artificial Neural Network, Hydraulic Model, Leakage, Water Distribution Networks.

1 Introduction

Several researchers have processed and analysed acoustic file data to classify leakage incidents in the water network using machine learning approach [1], [2], [3]. However, as reviewed by Fan, further development should focus on study cases that rely on the actual event and not only in an experimental setting [4]. This is substantiated by Bykerk and Valls, who comprehensively review studies analysing acoustic information and include a test in a real District Metered Area (DMA) in Sydney without any hydraulic modelling analysis [5].

Only Saqib et al., (2017) proposes an integration between acoustic information and hydraulic modelling. They approach this integration by introducing an initial screening of leaks based on pressure data generated with EPANET, followed by a simulation of leak vibrations using the acoustic sound propagation mathematical model which developed by Brennan et al. [6]. The drawbacks of the study is that a theoretical network is modelled, and that no real acoustic data is analysed.

This is the gap this paper aims to address. In particular, the extent to which the combination of these two data sources can improve the overall performance of the leak location method is therefore an open question. We propose a novel method that integrates hydraulic modelling-generated data with real acoustic sensor information to improve leak location in a real DMA of a WDN in a city in Eastern Europe. The framework to make this integration is a data-driven model, specifically an Artificial Neural Networks (ANN).

2 Methods

This study adopts a systematic and sequential methodology, encompassing three distinct steps: data preparation, ANN model building, and experimental setup, described in detail below.

2.1 Data Preparation

This step is intended to prepare the acoustic and hydraulic data for training and applying the ANN model. On the acoustic side, four variables are extracted from the noise logger file, namely Mean-Time Domain, Peak-Time Domain, Peak-Frequency Domain, and Kurtosis-Frequency Domain. Conversely, on the hydraulic side, a leakage scenario is simulated at a location where the noise logger is deployed, with leak rates ranging from 0.125 to 5 L/s. In this leakage modeling process, several variables are taken into consideration, including pressure and velocity.

2.2 ANN Model Building for Predicting Leak Occurrence

The second phase of the methodology consist of the construction of the ANN for the purpose of predicting leak occurrence based on binary classification (leak/no leak). Selected acoustic and hydraulic properties are the set of independent variables from which the target (output's dependent variable) is the leak / no leak condition.

The dataset preparation process is automated using Python scripting language in conjunction with various packages, including NumPy, Pandas, SciPy, TensorFlow, Keras, and SciKit Learn. From the data ratio perspective, 80% of the data is used as a training set, while the rest is used as a testing set. The input layers are formed by sets of acoustic and hydraulic features; two hidden layers are used, each consisting of eight nodes, to calculate the relation weight among features. Finally, the output layer consists of one layer with a single node. The hidden layers are activated using the rectified linear unit (ReLU) activation function, and the output is activated using a sigmoid function. Since the leakage determination is a binary classification problem, the Binary Cross Entropy method was used as the loss function. Adam's optimiser [7] was employed to optimise the forward and backpropagation to reduce error.

2.3 Experimental Setup

The final phase of this research involves training and evaluating the performance and applicability of the ANN in accurately classifying leak and non-leak incidents within WDNs, utilizing various feature inputs as presented in Table 1. In each experiment, different sets of acoustic and hydraulic attributes (and their combinations) are used. Note that the hydraulic attributes include the incremental variation of leak rate.

Table 1. Experiments using Different Input Data Features for ANN Leakage Classification

Experiment ID	Input Attribute			Output Attribute
	Acoustic Data Input	Hydraulic Data Input	Combined Data Input	
A - 1	Mean – TD Peak – TD Kurtosis – FD	Pressure at leak	Mean – TD Peak – TD Kurtosis – FD Pressure at leak	Leak / no leak
A – 2	Mean – TD Peak – TD Kurtosis – FD	Delta pressure at leak	Mean – TD Peak – TD Kurtosis – FD Delta pressure at leak	Leak / no leak
A – 3	Mean – TD Peak – TD Kurtosis – FD	Pressure at leak Delta pressure at leak	Mean – TD Peak – TD Kurtosis – FD Pressure at leak Delta pressure at leak	Leak / no leak
A – 4	Mean – TD Peak – TD Kurtosis – FD	Pressure at leak Delta pressure at leak Upstream velocity Downstream velocity	Mean – TD Peak – TD Kurtosis – FD Pressure at leak Delta pressure at leak Upstream velocity Downstream velocity	Leak / no leak

B – 1	Mean – TD Peak – FD Kurtosis – FD	Pressure at leak	Mean – TD Peak – FD Kurtosis – FD Pressure at leak	Leak / no leak
B – 2	Mean – TD Peak – FD Kurtosis – FD	Delta pressure at leak	Mean – TD Peak – FD Kurtosis – FD Delta pressure at leak	Leak / no leak
B – 3	Mean – TD Peak – FD Kurtosis – FD	Pressure at leak Delta pressure at leak	Mean – TD Peak – FD Kurtosis – FD Pressure at leak Delta pressure at leak	Leak / no leak
B - 4	Mean – TD Peak – FD Kurtosis – FD	Pressure at leak Delta pressure at leak Upstream velocity Downstream velocity	Mean – TD Peak – FD Kurtosis – FD Pressure at leak Delta pressure at leak Upstream velocity Downstream velocity	Leak / no leak

3 Results

3.1 Features Selection

On the acoustic data input, four variables have been explored: Mean–TD, Peak–TD, Peak–FD, and Kurtosis–FD. Previous research has shown that these variables strongly relate to a leak signal in WDN [3], [4], [8]. While on the hydraulic data input, pressure, delta pressure, and velocity have been selected. Table 2 provides a detailed overview of the multicollinearity analysis of input features in experiments using the Variance Influence Factors (VIF). It is observed that employing these features individually leads to a substantial reduction in the value of VIF, signifying a decreased level of dependency.

Table 2. Multicollinearity Analysis of Input Features in Experiments using VIF

Input Variables Feature	Experiment ID							
	A-1	**A-2**	**A-3**	**A-4**	**B-1**	**B-2**	**B-3**	**B-4**
Mean -TD	1.03	1.02	1.04	1.04	1.03	1.00	1.04	1.04
Peak -TD	1.67	1.50	1.85	2.10	-	-	-	-
Peak – FD	-	-	-	-	1.23	1.12	1.25	1.37
Kurtosis – FD	1.15	1.14	1.15	1.26	1.08	1.11	1.11	1.24
Pressure wo Leak	-	-	-	-	-	-	-	-
Pressure w leak	1.51	-	1.61	2.3	1.19	-	1.40	2.21
Delta Pressure	-	1.36	1.45	2.4	-	1.13	1.33	2.41
Velocity Upstream	-	-	-	3.08	-	-	-	3.13
Velocity Downstream	-	-	-	2.02	-	-	-	1.91

3.2 ANN Performance for Different Input Features

ANN models were trained and tested using the inputs in Table 1 as independent variables and the output leak / no leak as the dependent variable. The confusion matrix summary as seen in Figure 1 shows that the use of different feature sets and their impact on the model's performance underscored the importance of feature selection and representation in achieving accurate classification results. Based on the experiment result, it can be conluded that a combination of acoustic features including Mean-Time Domain (Mean-TD), Peak-Time Domain (Peak-TD), and Kurtosis-Frequency Domain (Kurtosis-FD) is considered alongside hydraulic features such as pressure at leak and delta pressure at leak gives better performance compared to others input combinations.

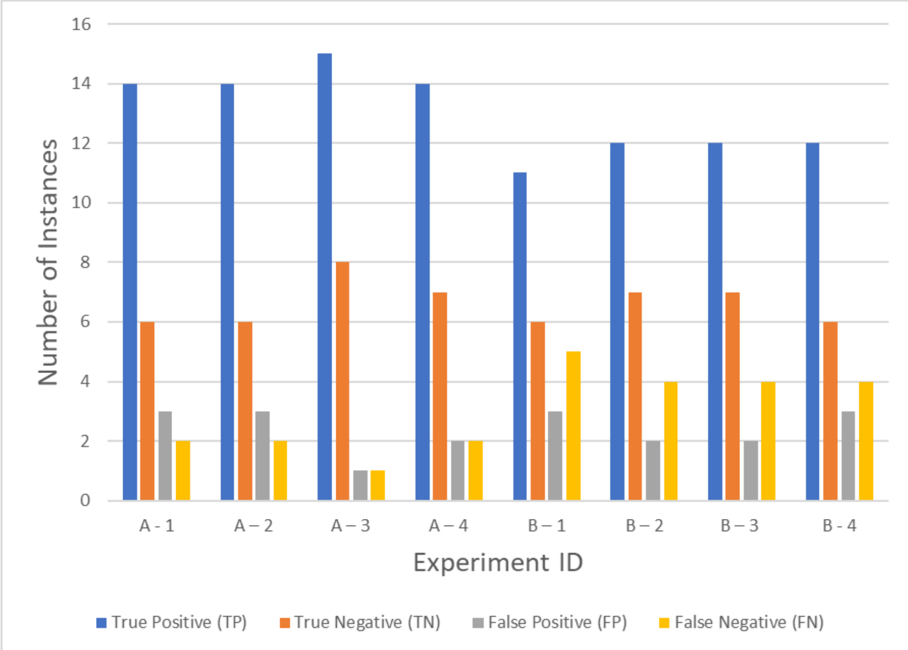

Figure 1. Summary of the confusion matrices from all experiments with combined input

4 Conclusion

In conclusion, our analysis of the confusion matrix from multiple experiments revealed variations in the model's performance across different conditions. The consistent TP and TN values across similar experiments indicated stability in the model's classification ability, while discrepancies in FP and FN values highlighted areas where the model could benefit from further improvement. The use of different feature sets and their impact on the model's performance underscored the importance of feature selection and representation in achieving accurate classification results.

References

[1] S. El-Zahab, A. Asaad, E. Mohammed Abdelkader, and T. Zayed, "Development of a clustering-based model for enhancing acoustic leak detection," *Can. J. Civ. Eng.*, vol. 46, no. 4, pp. 278–286, Apr. 2019, doi: 10.1139/cjce-2018-0229.

[2] Y. Jin, W. Yumei, and L. Ping, "Approximate entropy-based leak detection using artificial neural network in water distribution pipelines," in *2010 11th International Conference on Control Automation Robotics & Vision*, Dec. 2010, pp. 1029–1034. doi: 10.1109/ICARCV.2010.5707291.

[3] S. Li, Y. Song, and G. Zhou, "Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition," *Measurement*, vol. 115, pp. 39–44, Feb. 2018, doi: 10.1016/j.measurement.2017.10.021.

[4] H. Fan, S. Tariq, and T. Zayed, "Acoustic leak detection approaches for water pipelines," *Autom. Constr.*, vol. 138, p. 104226, Jun. 2022, doi: 10.1016/j.autcon.2022.104226.

[5] L. Bykerk and J. Valls Miro, "Detection of Water Leaks in Suburban Distribution Mains with Lift and Shift Vibro-Acoustic Sensors," *Vibration*, vol. 5, no. 2, Art. no. 2, Jun. 2022, doi: 10.3390/vibration5020021.

[6] M. Brennan, F. Lima, F. Almeida, P. Joseph, and A. Paschoalini, "A virtual pipe rig for testing acoustic leak detection correlators: Proof of concept," *Appl. Acoust.*, vol. 102, pp. 137–145, Jan. 2016, doi: 10.1016/j.apacoust.2015.09.015.

[7] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." arXiv, Jan. 29, 2017. Accessed: Jul. 13, 2023. [Online]. Available: http://arxiv.org/abs/1412.6980

[8] I. A. Tijani, S. Abdelmageed, A. Fares, K. H. Fan, Z. Y. Hu, and T. Zayed, "Improving the leak detection efficiency in water distribution networks using noise loggers," *Sci. Total Environ.*, vol. 821, p. 153530, May 2022, doi: 10.1016/j.scitotenv.2022.153530.