


Proceeding Paper

Attributing Minimum Night Flow to Individual Pipes in Real-World Water Distribution Networks Using Machine Learning [†]

Matthew Hayslep ^{1,*} , Edward Keedwell ¹, Raziye Farmani ² and Joshua Pocock ³

¹ Department of Computer Science, University of Exeter, Exeter EX4 4QF, UK; e.c.keedwell@exeter.ac.uk

² Department of Engineering, University of Exeter, Exeter EX4 4QF, UK; r.farmani@exeter.ac.uk

³ South West Water, Exeter EX2 7HR, UK; jpocock@southwestwater.co.uk

* Correspondence: mh989@exeter.ac.uk

[†] 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 article introduces an explainable machine learning model for estimating the amount of flow that each pipe in a district metered area (DMA) contributes to the minimum night flow (MNF). This approach is validated using the MNF of DMAs and pipe failures, showing good results for both tasks. The predictions from this model could be used to guide leak management or intervention strategies. In total, 800 DMAs ranging from rural to urban networks and representing nearly 12 million meters of pipe from a UK water company are used to train, validate, test, and evaluate the methodology.

Keywords: machine learning; minimum night flow; water distribution network; leakage; pipe failure



Citation: Hayslep, M.; Keedwell, E.; Farmani, R.; Pocock, J. Attributing Minimum Night Flow to Individual Pipes in Real-World Water Distribution Networks Using Machine Learning. *Eng. Proc.* **2024**, *69*, 112. <https://doi.org/10.3390/engproc2024069112>

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

Published: 10 September 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

Minimum night flow (MNF) is an important metric commonly used to estimate and understand leakage [1] within district meter areas (DMAs) and is the most common leakage assessment methodology used in the UK [2]. Leakage is a pervasive problem with economic and environmental consequences [3], making it important for water companies, regulators, and governments. One of the main drawbacks to MNF is the resolution of data; even with the complete coverage of smart meters, it is difficult to attribute the remaining water balance to particular pipes. An alternative way to understand leakage is to study known cases of historic leaks and bursts [3], i.e., pipe failures. This relies on the records of utility companies regarding engineering work conducted on the infrastructure. One of the downsides of using historic pipe failures to understand leakage is that this will not include unreported or background leakage (unlike MNF). MNF approximates leakage under the assumption that legitimate water usage is lowest at night, and therefore that most of the flow during this period is leakage.

Previous studies have used machine learning methods to estimate the MNF of entire DMAs based on various factors such as total customers, total pipe length, etc. [4,5]. This paper differs from these previous works by predicting the contribution of individual pipes to MNF. Furthermore, this approach uses data from 800 real-world DMAs that are readily available to water companies, making it applicable to a wide range of real-world scenarios. Finally, by attributing MNF to particular pipes in an explainable manner, this methodology provides more information to decision makers and practitioners, which could contribute to improved leakage assessments, leak localization practices, and sustainable water supply management.

2. Materials and Methods

This paper presents a linear regression model that predicts the amount of MNF a pipe is responsible for, henceforth referred to as pipe-MNF. The data used in this study cover the average MNF for these DMAs in September 2023, pipe failure, and pipe asset data. The following asset data were used for each pipe: diameter, age, material (grouped into metal, plastic, and other), number of domestic connections, number of commercial connections, number of hospital connections, and number of agricultural connections. Separate diameter and age features were created for each material (i.e., metal age, plastic age, metal diameter, etc.) to ensure that all features were numeric. This also allowed for the linear regression algorithm to have different coefficients for these various aspects; for example, the model could assign more importance to the age of metal pipes than that of plastic pipes. Pipe failures were also recorded and were loosely defined as any repair or replacement action undertaken by the water utility company that had some leakage component. These actions were then associated with the closest pipe.

The dataset was split by DMAs, with 70% of DMAs in the training set and the remainder in the test set. The same DMA split was used to filter the pipe asset data into training and testing sets. The pipe asset data and MNF data were used to train a linear regression model that predicted the pipe MNF. However, because MNF is observed over an entire DMA, there was no direct way to evaluate the accuracy of the model's pipe MNF predictions. Therefore, two different validation methods were used: (1) taking the sum of the pipe MNF predictions over a whole DMA and comparing it to the observed MNF, and (2) using the pipe MNF prediction as a prediction of the likelihood that a pipe failure had occurred. The loose definition of pipe failure was used because the model attempted to predict the pipe's flow contribution to MNF, not burst or leak likelihood. The application of these two validation methods to the linear regression model showed promising results.

3. Results

Once trained, the resulting linear regression model made pipe MNF predictions such as those shown in Figure 1. This figure shows that the model highlighted a small number of high-risk pipes that contributed significantly more to the MNF of the DMA than most of the other pipes in this DMA. These predictions could then be used to direct or inform further action within the DMA such as leak localization. If a larger area of a DMA was highlighted, it suggested that leak management strategies such as pressure-reducing valves would be useful. Table 1 shows the regression metrics for this model using validation method (1). Figure 2a shows the predictions versus observed values for validation method (1).

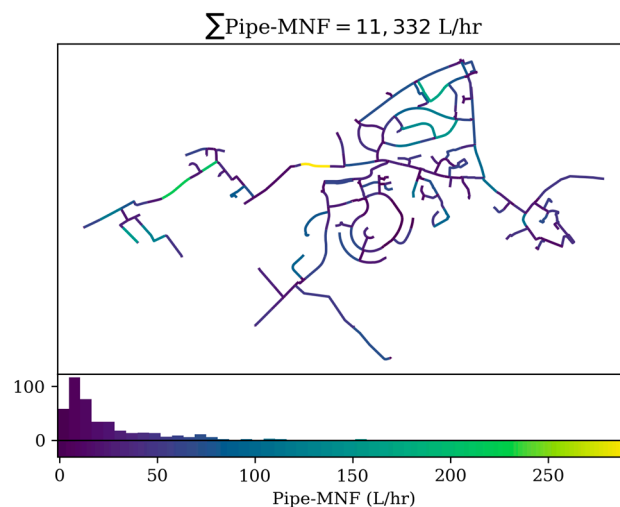


Figure 1. Predictions for an example DMA. The sum of pipe MNF predictions (i.e., DMA-MNF) is at the top. A histogram of prediction values is shown at the bottom.

Table 1. Performance metrics for the linear regression model using validation method (1).

Metric	Training Set	Test Set
R ²	0.674	0.611
MAPE	0.336	0.332
RMSE	4314	4409

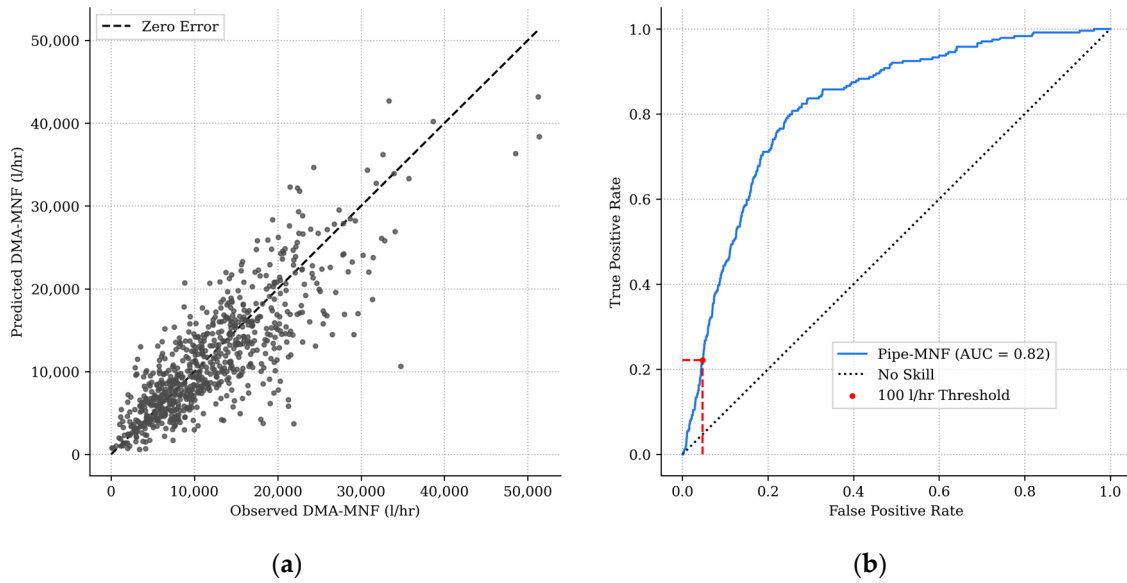


Figure 2. Performance plots for the pipe MNF prediction model: (a) observed vs. predicted DMA–MNF values using validation method (1); (b) ROC curve for pipe failure classification, i.e., validation method (2).

Figure 2b shows the ROC curve for the linear regression model’s pipe MNF predictions using validation method (2). The pipe MNF predictions were not modified for this task but were interpreted as a measure of how likely a pipe was to have failed, i.e., a higher pipe MNF meant a higher likelihood of pipe failure. This was based on the assumption that MNF approximates, or is proportional to, leakage, and that pipe failure is a direct measure of where leakage has been found. Figure 2b indicates that the pipe MNF predictions have predictive power for pipe failure and show good results in comparison with those of other studies [6], which further validates the pipe MNF predictions that the linear regression model made. Although validation method (2) and Figure 2b only give an indication of accuracy with regard to reported leaks and bursts, in combination with the regression results from Table 1 and Figure 2a, they suggest that the pipe MNF predictions are reasonably accurate. Table 2 shows the coefficients of the linear regression model, which correspond to the L/h increase in MNF for each feature. This table clearly shows the large and expected impact of different types of consumers on MNF. In addition, according to this model, plastic pipes, per mm in diameter, have a lower impact on MNF than metal pipes do, but have a higher impact with increasing years of age.

Table 2. Shows the de-scaled coefficients of the linear regression model. The diameters are expressed in mms, and age is expressed in years (rounded to the nearest month).

Feature	Coefficient	Feature	Coefficient
Domestic Count	3.426	Hospital Count	452.181
Commercial Count	24.342	Agricultural Count	166.036
Metal Diameter	0.080	Metal Age	0.049
Plastic Diameter	0.001	Plastic Age	0.803
Other Diameter	0.094	Other Age	0.124

4. Conclusions

This paper presents a model that predicts the contribution of individual pipes to MNF and that achieved good results using two validation methods, showing that the predictions are accurate. By using a linear regression model, the impact of each feature on the final prediction was determined and could be explored and explained. Predictions from this model could be used to direct or inform leak management strategies for water utilities, and by understanding leakage through pipe-level MNF estimates, water utility companies will be able to make better and more informed decisions on how and where to tackle leakage.

Author Contributions: Conceptualization, M.H., E.K., R.F. and J.P.; methodology, M.H.; software, M.H.; validation, M.H., E.K. and R.F.; formal analysis, M.H.; investigation, M.H.; resources, M.H.; data curation, M.H.; writing—original draft preparation, M.H.; writing—review and editing, E.K. and R.F.; visualization, M.H.; supervision, E.K., R.F. and J.P.; project administration, M.H., E.K. and R.F.; funding acquisition, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by South West Water Ltd. through a PhD Studentship with University of Exeter for Matthew Hayslep.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because they are commercially sensitive. Requests to access the datasets should be directed to Joshua Pocock.

Acknowledgments: The authors would like to acknowledge and thank members of South West Water for their help, enthusiasm, and constructive feedback. Furthermore, the authors would like to acknowledge the University of Exeter, Centre for Water Systems for creating a positive and welcoming research environment.

Conflicts of Interest: The authors declare no conflicts of interest. The funder collected the data used in the study as part of their routine operations, and not specifically for the study. The funders had no role in the design of the study; selection, analyses, or interpretation of data; the writing of the manuscript; or the decision to publish the results.

References

1. Farley, M.; Trow, S. *Losses in Water Distribution Networks: A Practitioners' Guide to Assessment, Monitoring and Control*, 1st ed.; IWA Publishing: London, UK, 2005; pp. 1–273. [[CrossRef](#)]
2. Farrow, J.; Jesson, D.; Mulheron, M.; Nensi, T.; Smith, P. *Achieving Zero Leakage by 2050: Basic Mechanisms of Bursts and Leakage*, 1st ed.; UK Water Industry Research Limited: London, UK, 2017; pp. 1–147.
3. Puust, R.; Kapelan, Z.; Savic, D.; Koppel, T. A review of methods for leakage management in pipe networks. *Urban Water J.* **2010**, *7*, 25–45. [[CrossRef](#)]
4. Alkaseh, J.M.A.; Adlan, M.N.; Abustan, I.; Aziz, H.A.; Hanif, A.B.M. Applying Minimum Night Flow to Estimate Water Loss Using Statistical Modeling: A Case Study in Kinta Valley, Malaysia. *Water Resour. Manag.* **2013**, *27*, 1439–1455. [[CrossRef](#)]
5. Hayslep, M.; Keedwell, E.; Farmani, R. Multi-Objective Multi-Gene Genetic Programming for the Prediction of Leakage in Water Distribution Networks. In Proceedings of the Genetic and Evolutionary Computation Conference, Lisbon, Portugal, 15–19 July 2023. [[CrossRef](#)]
6. Bakker, M.; Vreeburg, J.H.G.; Van De Roer, M.; Rietveld, L.C. Heuristic burst detection method using flow and pressure measurements. *J. Hydroinform.* **2014**, *16*, 1194–1209. [[CrossRef](#)]

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.