A Novel Approach for Classifying Leakage in Water Networks

Tom Jefferis*[1], Eamon J. Wilkinson[1], and Jack D. Saunders[1]

[1] Trident Intelligent Solutions Ltd.
* Corresponding author: tom@trident-intelligence.com

Abstract. Water loss is arguably one of the biggest challenges currently facing water distributors globally, with water lost through leakage a key area of concern. Leakage not only impacts customers, but has significant environmental and financial implications. Whilst much has been written on the topic of how to calculate or predict leakage, water distributors face a challenge in their approach to response, with no consensus on the categorisation of leakage. That is, whilst there are established approaches to stating the amount of water lost through leakage, determining which areas to respond to as a priority is non-trivial, especially when limited resources are a concern. In this work, an approach to classifying leakage using a simple statistical approach is presented, banding leakage from least to most severe. Domain experts in the UK water industry have verified this approach on real-world data.

Keywords: Leakage Classification, Water Network Management, Data Analysis.

1 Introduction

Efficient leakage management in water networks is key for improving water retention within the environment, leading to decreased energy use in its treatment and distribution [1]. However, the persistent challenge of excessive leakage poses significant hurdles in the water industry. This is because, not only does excessive leakage strain vital resources, but it also jeopardises environmental sustainability and incurs substantial economic losses. OFWAT reports that leakage causes the loss of roughly 20% of all water flowing through pipes [2], and 3 billion litres of water are lost solely in the UK on a daily basis due to leakage [1]. There are various approaches to the physical detection of leakage, each with their own advantages and drawbacks. For a recent review of the state-of-the-art sensing technology, see [3].

Whilst there is comprehensive literature for detection and prediction of leakage there are few examples of classifying or prioritising leakage. This oversight is significant as classifying leakage by severity allows the efficient allocation of resources to repairs that would have the greatest impact of reducing leakage. This paper demonstrates an effective method for classifying leakage by severity that has been verified on real-world data by domain experts in the UK water industry.

2 Background

Various approaches to calculating, measuring, and predicting leakage have been proposed in the literature, such as recent work that utilised a multi-objective multi-gene genetic programming system, described in [4]. This study focuses on predicting long-term unreported and background leakage in water networks, and involves utilising Minimum Night Flow (MNF) measurements in District Metered Areas (DMAs) to estimate leakage. Other recent work, described in [5] looked to address the drawbacks faced in current approaches to leakage calculations by incorporating valve manoeuvres, which impact network calmness with the potential introduction of transients [6].

It has been shown that the physical characteristics of a pipe - such as length, diameter, age, volume, and material - can influence leakage levels [7,8]. Yet, there are also external factors that contribute to leakage within a DMA, such as freezing weather or soil expansion [2].

3 Methods

The first step of the proposed leakage detection method is to calculate the DMA leakage from flow data. Firstly, the DMA flow data is processed to attain only the data from 3 - 4am, referred to as the nightline flow. Calculating the base nightly flow is done in multiple different ways depending on the data and area used. Generally, a continuous low (multiple days at the same or very similar flow) or taking the lowest point over a long period for measurement is a good indicator of the base nightly flow. This is standard practice within the water industry in the UK, although there are several variations on this in the literature [9-11]. We took the 40th percentile of our 3 - 4am flow as the nightly base rate as the data is in 15 minute intervals. The nightly base rate is then subtracted from the nightly flow data, which will give a leakage value per day, as shown in figure 1.

The next step is calculating the technically achievable leakage level (TAC). This is calculated by firstly Z-scoring the leakage values across all available data for a single DMA. The formula for Z scoring is seen in the equation below.

$$z = (x - \mu) / \sigma$$

Where x is the value being evaluated, μ is the mean, and σ is the standard deviation.

After calculating the Z-scored values for each data point, the 10th percentile is used to serve as the level of leakage that is maintainable, the TAC. Any value above this is considered for leakage alert classification and notification. This TAC is then used in the following leakage classification stages as shown in figure 2. Note that only 28 days of data is shown in the plot but the TAC is calculated on all available values in the DMA leakage array.

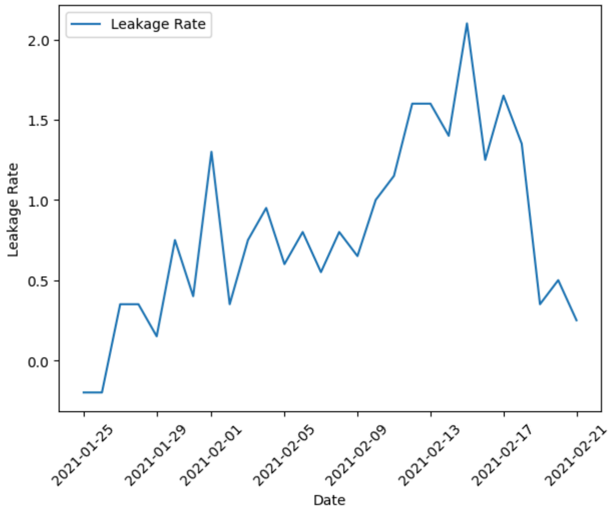

Figure 1: Daily leakage rate for 28 days. Y axis shows the leakage rate and X axis shows the date in days.

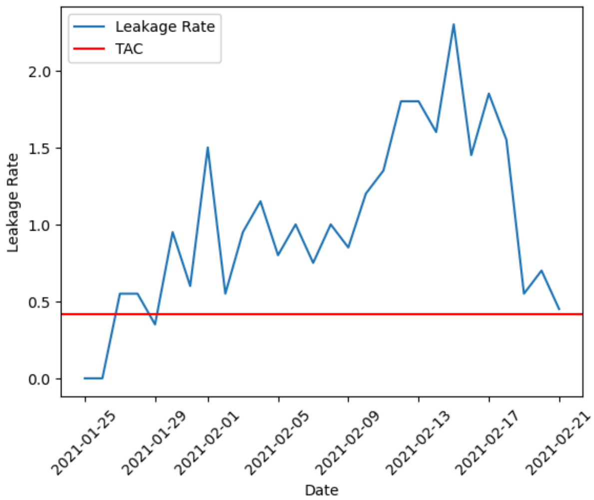

Figure 2: Daily leakage shown with the technically achievable level (TAC). The Y axis shows the leakage rate and the X axis shows the date in days. The TAC is shown as the red horizontal line on the plot.

Once the daily leakage and the TAC have been calculated, the leakage levels can be classified. The leakage classification is divided into three levels of increasing severity.

Level 1 is a notification that the DMA might be be exhibiting signs of a minor fault, level 2 is an alert that the DMA is now leaking at a substantially higher level than usual and should be addressed, level 3 is an alert to take immediate action on the DMA as a catastrophic burst (or similar event) has likely occurred. Succinctly, in practice, these levels were informally referred to as ‘possible’, ‘probable’, and ‘definite’.

The level 1 alert is calculated by measuring the number of increasing data points, the number of which being dependent on the intervals of the data. If the leakage data is in daily intervals the level 1 alert requires 7 days of increasing (or if it has increased and maintained that level) leakage above the TAC to be raised. If the data is weekly data, 3 weeks of increasing leakage above the TAC will trigger a level 1 alert. Note that the threshold for an alert to be activated can be configured automatically or a manual override can be set.

The level 2 alert is dependent on a level 1 alert, only being raised if a level 1 alert is active. The level 2 alert is raised when the leakage is greater than the mean of all previously observed data points plus the standard deviation. The level 2 alert can be stopped by the leakage returning to or below the mean of all previously observed data minus the standard deviation.

Level 3 alerts are the highest level of alert, and are independent of the other two levels. A level 3 alert can be triggered in two ways; firstly by having a leakage point with a Z-score of greater or equal to two, meaning that the leakage is two standard deviations above the mean of all observed data. Alternatively, the alert can be triggered by having a leakage greater than 100 times the TAC plus standard deviation. A level 3 alert can be stopped by the leakage returning below a Z-score of 2.

There are built in overrides to allow level 1 and 2 alerts to trigger if the data is fluctuating a lot as seen in the figure above. Specifically, if the data would trigger a level 2 alert but a level 1 alert is not raised, the level 1 counter is set to trigger if the next datapoint is the same or more than the current leakage level thus, immediately triggering a level 2 alert if the leakage level is maintained.

Note that the parameters of this method were developed in conjunction with domain experts in the UK water industry.

4 Discussion

Figure 3 shows the resulting leakage bands when this approach is applied to specific leakage data. It is evident that it has correctly identified the section of the data that has a fault. Interestingly, a rising trend is observed of the data that is not classified as any leakage level. This is due to the fluctuations in leakage. That is, one day it is high but the next it is falling, thus not triggering any level 1 alerts until it has surpassed the consecutive data points threshold. The built in override did not trigger early on in the data as the following point after the early spike is much lower leakage, thus resetting the level 1 counter. This non-trigger is by design to stop the system alerting the user to potential false alarms.

The level 1 alert is raised after there are 3 consecutive days of increasing leakage which quickly becomes a level 2 alert. This demonstrates the ability of the system to identify trends in the data and its ability to ignore highly fluctuating data that could be caused by a faulty reading or other non-leakage related events not known by the system.

The level 3 alert is also raised once the leakage has passed its threshold. Note that the level 2 alert is also active for this period, demonstrating the independent high level alert. The level 3 alert is only active for 3 data points as the leakage seems to drastically fall after spiking, this could be because a repair was rapidly carried out.

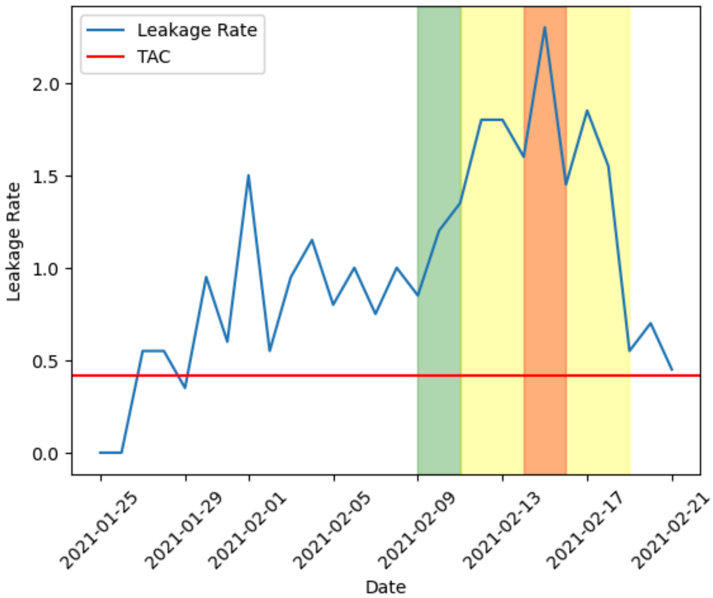

Figure 3: Daily leakage shown with the technically achievable level (TAC) and leakage bands. The Y axis shows the leakage rate and the X axis shows the date in days. The TAC is shown as the red horizontal line on the plot. The leakage bands are: Level 1 - Green, Level 2 - Yellow, Level 3 - Red.

A potential drawback of this approach is that it is parametric. That is, the standard deviations above which leakage must fall to trigger an alert, and the number of data points for which the leakage must rise, have been specified rather than using a more intelligent approach, such as a heuristic based method. However, this allows the user of the approach to specify system sensitivity, tweaking these values to their needs. Future work could look to explore the integration of an approach to automatically set these values on a DMA by DMA basis. This approach and the default parameters were developed and verified on 5,000 DMAs in a UK water network.

5 Conclusion

Overall, the methods employed to classify leakage work well on the example data and have shown promise in analysing real-world data examples. While the method has limitations in handling highly fluctuating data, these can be mitigated with tailored adjustments to alert level triggers. This customisation demonstrates the system's adaptability and its potential for effective use across various datasets with varying intervals. Another limitation is that this method only uses the leakage rate in the DMA with no other factors attributed like weather and soil conditions or pipe age and material. This could however be an avenue for future work where some of those features could be integrated into the system to get a measure based not only on historic flow data but also environmental and network conditions.

References

1. Ofwat: Service Delivery Report 2020-21, Service and delivery report 2020-21 - Ofwat, last accessed: 2024/03/26
2. Ofwat: Leakage in the water industry, https://www.ofwat.gov.uk/leakage-in-the-water-industry/, last accessed: 2024/03/26.
3. Islam, M.R., Azam, S., Shanmugam, B. and Mathur, D., 2022. A review on current technologies and future direction of water leakage detection in water distribution network. IEEE Access, 10, pp.107177-107201.
4. Hayslep, M., Keedwell, E. and Farmani, R., 2023, July. Multi-Objective Multi-Gene Genetic Programming for the Prediction of Leakage in Water Distribution Networks. In Proceedings of the Genetic and Evolutionary Computation Conference, pp. 1357-1364.
5. Fuertes-Miquel, V.S., Arrieta-Pastrana, A. and Coronado-Hernández, O.E., 2024. Analyzing Water Leakages in Parallel Pipe Systems with Rapid Regulating Valve Manoeuvres. Water, 16(7), p.926.
6. Starczewska, D., Collins, R. and Boxall, J., 2014. Transient behavior in complex distribution network: A case study. Procedia Engineering, 70, pp.1582-1591.
7. Hussein, M., Yoneda, K., Othman, N., et al., 2017. Effects of number of connections and pipe length to the water losses in Melaka. Jurnal Teknologi, 79(3), pp.45-59.
8. Hayslep, M., Keedwell, E., Farmani, R., et al., 2023. Understanding district metered area level leakage using explainable machine learning. In IOP Conference Series: Earth and Environmental Science, 1136 (1), pp. 1-10.
9. Farah, E. and Shahrour, I., 2017. Leakage detection using smart water system: Combination of water balance and automated minimum night flow. Water Resources Management, 31, pp.4821-4833.
10. AL-Washali, T., Sharma, S., AL-Nozaily, et al., 2018. Modelling the leakage rate and reduction using minimum night flow analysis in an intermittent supply system. Water, 11(1), pp.48-63.
11. Alkasseh, J.M., Adlan, M.N., Abustan, I., et al., 2013. Applying minimum night flow to estimate water loss using statistical modeling: a case study in Kinta Valley, Malaysia. Water resources management, 27, pp.1439-1455.