DATA-DRIVEN SOLUTION TO IMPROVING SCHOOLS' WATER EFFICIENCY: CHALLENGES AND INSIGHTS FROM A STATEWIDE PROGRAM

Yohan Kim[1], Ahlam El Halabi[2], Alexandra Butler[1], Pierre Mukheibir[1]
[1] Institute for Sustainable Futures (ISF), University of Technology Sydney (UTS), Sydney, NSW, Australia
[2] Yarra Valley Water (YVW), Mitcham, VIC, Australia

KEYWORDS

leakage identification, water efficiency, data analytics

ABSTRACT

Water leakage detection can be a significant challenge for schools. Sometimes persisting undetected for months, leaks are often only identified through the arrival of an abnormal and significantly high quarterly bill. The assessment of real-time water consumption allows for the detection and handling of leaks as they occur. This paper presents how a state-sponsored program has systematically reduced leakages at schools across Victoria with the analysis of water consumption data, through the Schools Water Efficiency Program (SWEP). Users are alerted of potential leaks and the leakage estimates are summed to calculate water and monetary savings generated through the program. This paper presents a framework for the development of a retrospective leak detection algorithm that can reliably estimate the leakage duration and volume from the quarterly flow records, as well as the savings incurred from SWEP's intervention.

BACKGROUND

Approximately 10% of total water usage in Australia is counted as losses [1]. While this figure is one of the lowest in the world, growing concerns around the impact of climate change on water availability, as well as the impact of population growth on leakage severity, makes leakage prevention one of the key issues for Australia [2].

Continued efforts have been made to improve consumption efficiency and minimize loss in Victoria. Of particular focus is the non-residential sector, which accounts for 25% of total demand in Melbourne [1]. In 2007, the Victorian government announced a roll out of mandatory water-savings program across all state schools expecting to save up to 750,000 million litres of water in each school through a combined reduction in water leakage and appliance upgrades[2].

Victorian water utilities continued to focus on promoting a range of water saving activities and behaviours targeted specifically to schools to encourage a reduction in potable water use. These are similar in nature to residential water saving initiatives as many schools have the same appliances one may find in their home; for example: installing aerators on taps, implementing water sensitive design in gardens, irrigating user dripper systems and installing rainwater tanks[3]. By 2012, per capita usage had dropped by 49% compared to 2000 in Victoria.

While programs such as these have been impactful for water consumption, they do not address the issue of lost water. The National Performance Report Data set reveals that real losses caused by leakage in the Victorian water system were reported to be 778 litres per service connection per day and 32600 litres per km of water main per day during the 2020-2021 financial year. Detecting, fixing and avoiding these leaks is critical to reducing unnecessary water consumption. In August 2016, the Water for Victoria plan was launched and has since made significant progress in meeting water supply challenges. This paper presents how the Schools Water Efficiency Program (SWEP), part of the action 5.3 in the plan, has played a fundamental role in meeting strategic water efficiency goals.

Schools Water Efficiency Program

The Schools Water Efficiency Program in Victoria provides a means for schools to be made aware of leaks as they occur, as well as to understand daily water consumptions to promote water savings. The program is driven by data loggers installed on existing water meters, which captures consumption data in 15-minute intervals for daily analysis. This data is used to alert schools of leaks where there have been night flows observed, as well as for the subsequent calculation of water and financial savings incurred for a quarterly report.

[1] https://www.water.vic.gov.au/liveable/using-water-in-business-and-industry

[2] https://www.abc.net.au/news/2007-03-25/schools-to-comply-with-compulsory-water-program/2225386

[3] https://www.melbournewater.com.au/water-data-and-education/learning-resources/browse-resources-year-level/using-and-saving-water

The loggers continuously record water use and provide access to an interactive web portal; this information can be used as part of the school curriculum. Participating schools receive access to a tailored curriculum program which uses the school's water data in mathematics and science, as well as providing students with robust water saving messages to use at school and home. SWEP has registered over 1302 schools in Victoria since it started in 2012. The program has helped these schools save over 9.6 billion litres of water which would have cost them more than $30.6 million in water and wastewater charges.

ANALYTICAL REQUIREMENTS

SWEP utilizes insights gained from the logger data to perform its key functions. These insights are used to directly prevent leaks as they occur, as well as to quantify the value of SWEP intervention for each school. Given this, SWEP needs information on 1) leaks detected as they occur, 2) the estimated volume of the leak, and 3) the estimated volume of avoided leaks.

Detection of leak event

In the program, leaks are detected from meter usage patterns and schools are notified to take action. The most common indicator for triggering the detection of leakages and wasted water is Minimum Night Flow (MNF) [3]. If a sustained period of non-zero flows is observed overnight then the school is alerted of a potential leakage the following morning. Schools are encouraged to scan their facilities for any visible leaks to identify the leakage source. Once the fix is performed, each school can view the SWEP online portal to ensure that water flows reflects the fix.

Estimated volume of leak event

In addition to the detection of leak events as they occur, the total volume of each leak needs to be estimated. While an MNF approach has low implementation costs, it cannot estimate the volume or duration of leakage. To reliably estimate leakage volume, three key components are required: 1) start of leak event, 2) end of leak event, and 3) leak rate between start and end of leak.

There are multiple challenges to deriving accurate leak events for schools that are applicable to other non-residential sites. First is the presence of irrigation, which usually occurs during the night (between 6pm and 10am)[4]. An irrigation schedule varies for each school, hindering the MNF approach to identify start of leaks. This limitation requires a method that can reliably detect irrigation as normal use and not mislabel scheduled usage as overnight leaks.

Another challenge is the lack of labelled data i.e. a secondary dataset accompanying the meter logger data that flags when the leaks have actually occurred. Such data would allow for testing the accuracy of any leak detection algorithm. Such a dataset is not feasible to achieve and would be very costly to acquire with any reliability across the 1000+ schools involved in the SWEP program. Given this, the authors have developed a *semi-supervised* decision tree algorithm as shown in Figure 9, where a small sample of results are manually checked to ensure the correct classification of leaks. The algorithm is described in detail in the *methodology: leakage detection algorithm* section of this paper.

Estimated volume of avoided leaks

As part of SWEP's data-driven approach, any savings in wasted water (in both litres and dollar) due to SWEP's intervention must be quantified. The volume of avoided leaks is estimated to be:

$$V_{avoided} = V_{undetected} - V_{unavoided} \qquad \text{(Eq. 1)}$$

Where $V_{unavoided}$ refers to the volume of the actual leak that could not be avoided due to the time it takes to call in a plumber and wait for a response. $V_{undetected}$ refers to the volume of the leak had it gone unnoticed. Without SWEP, it is assumed that leaks would go unnoticed until the next water bill is received showing abnormal levels of usage. This is estimated to be:

$$V_{undetected} = R_{mean} * D_{mean} \qquad \text{(Eq. 2)}$$

Where R_{mean} is the mean leak rate of the actual leak event, and D_{mean} is the average number of days until the next bill arrives for each given school. The algorithm is described in detail in the *methodology: avoided leaks algorithm* section of this paper.

METHODOLOGY

We developed an algorithm using the R language that iterates through individual school data to detect and quantify leaks. Historical leak data from various schools was fed into the algorithm to check for accuracy of leak detection. Improvements were made to conventional methods to reduce false-positives and improve the detection rate for the context of school use. We also established a bespoke algorithm to estimate the savings from avoided leaks based on the calculated duration of the leak if it had been left unaddressed.

[4] https://www.water.vic.gov.au/liveable/using-water-wisely/advice-and-rules/permanent-water-saving-rules

Leakage detection algorithm

An overview of the decision-tree based leak detection algorithm, explained in this section, is located in figure 8. For each day, the data is split into nighttime and daytime periods, where nighttime is defined as from 7pm to 5am, and daytime from 5am to 7pm. Separate analysis is performed for each period. Having a separate decision tree for nighttime allows for capturing of intuition that during nighttime, there should be no normal usage other than irrigation. In some cases, rainwater tank and/or pool top-ups can also exhibit similar usage patterns.

For each period, data is further grouped into segments using changepoint analysis [4]. Changepoint analysis estimates points in time at which the statistical properties of a sequence of observations change, thus creating segments with the data based on observed water use. This also means that the datapoints within each segment are relatively stable and predictable. By considering changes between segments, rather than changes between each datapoint, the computational speed and analytical simplicity is improved. Figure 1 shows an example of how the segmentation would occur. In this example, the nighttime period (red arrow) is grouped into just one segment. This reflects the fact that water consumption is relatively stable throughout the night. The daytime period (blue arrow) is grouped into four segments.

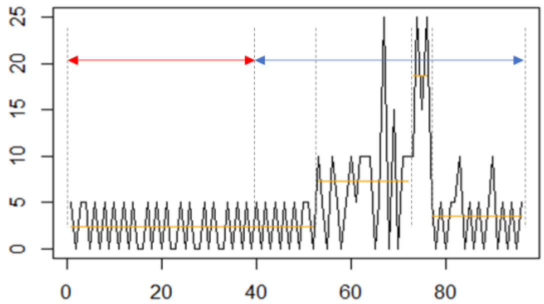

Figure 1. Example of one day's worth of data (black) split into night/day periods (red/blue) and further into segments (yellow).

Nighttime analysis: if nighttime is grouped into one segment, as it is in Figure 1, then it can be understood that there is stable flow throughout the night indicating no periods of irrigation. If there are no leaks, the mean flow rate of this segment should be zero. If the mean flow rate of this segment is above a certain threshold (i.e. 30% of the pulse rate[5] for a given meter), then we can assume a leak is present. This segment is therefore labelled as a leak.

In the case where a nighttime period is split into more than one segment and there has been a change in volume between these segments it can be assumed that irrigation has occurred. To further check if any leaks have occurred within this period, the algorithm additionally checks for presence of consecutive zeros. The number of consecutive zeros depends on the pulse rate; for meters with smaller pulse rates (i.e., a higher resolution of data capture), the threshold for consecutive zeros can be set higher. If there are any segments with consecutive zeros, the algorithm assumes that this segment and any before it does not have a leak. We iterate from the last to the first segment detecting for any leaks, stopping once the first consecutive zeroes are found (and label the remaining as non-leaks).

Figure 2 shows an example of how this logic would apply. The rightmost segment is first checked and labelled as a leak, since it is non-zero. The next segment contains a zero, so the search process stops. It should be noted that this approach may also capture irrigation manually left on until morning. Such false positives have been accounted for and are filtered out at a later point in the process.

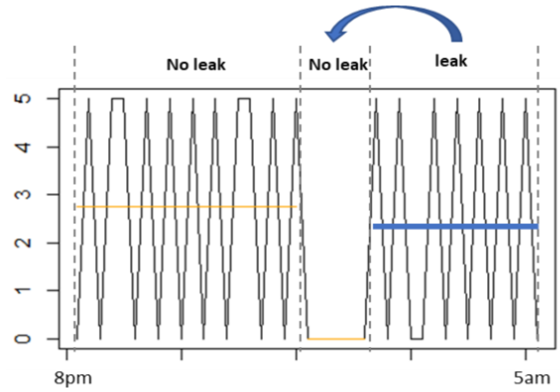

Figure 2. Example of last nighttime segment flagged as a leak, indicating a potential start of a leak event.

Figure 3 shows another example where segments 1 to 4 are all flagged as leaks. The actual rate of this leak is estimated to be the lowest mean rate of all segments labelled as leaks overnight. This removes over-estimation from irrigation that may simultaneously occur.

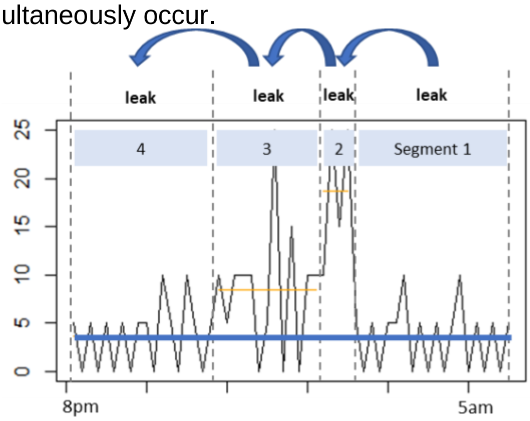

Figure 3. Example of all nighttime segments flagged for leak, with mean leak rate of 4L/15min.

Daytime analysis: Once the nighttime analysis is complete, the following daytime period is analysed. If there is no leak from the night before, the daytime

[5] Each pulse that is output by a meter represents a pre-determined volume of water passing through that meter.

period is ignored and analysis skips forward to the next nighttime period.

If a leak was detected from the night before, the algorithm checks to see if the leak is ongoing. The check is similar to the nighttime process, except that the algorithm checks segments in a forward order, not backwards. Figure 4 shows an example of daytime analysis. Segment 1 is first checked and labelled as a leak, since it contains no consecutive zeros and there was leak overnight. Segments 2 to 4 are also labelled as leaks as they do not contain consecutive non-zeros. The actual rate of the leak is calculated to be the lowest mean rate of all segments labelled as leaks here. This removes over-estimation from daytime activity that occurs alongside the leak.

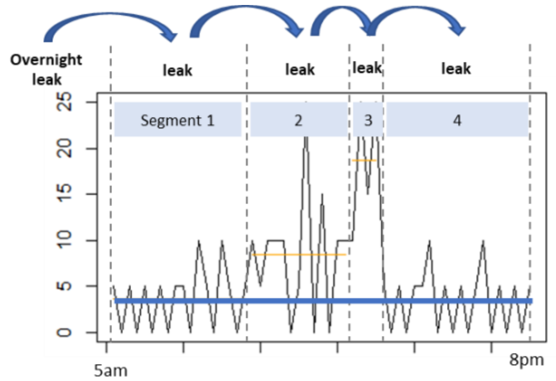

Figure 4. Example of daytime analysis, with leak flagged from the night before.

Additional improvements to the algorithm: Once all segments are labelled, leaks that are close to one another are joined and labelled as a continuous leakage period. This prevents double counting of benefits from SWEP. In the example below, the algorithm has found neighboring leaks on the 28th and 30th. The following conditions are checked:

- Is the break segment less than a threshold? (default: 48 hours)
- Are the adjacent leaks both longer than a threshold? (default: 24 hours)

If both conditions are met, the break segment is re-labelled as a leak and the two leak events are joined.

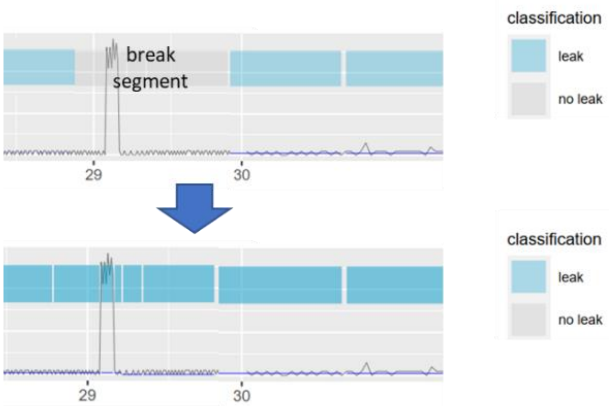

Figure 5. Joining of two leak events

An additional check has been developed for *wasted water* (from taps left running for example). Any leak that is shorter than a threshold (default: 48 hours but 78 hours if the leak started on a Friday) is re-classified as wasted water.

In the example in Figure 6, a leak event spanning from Friday night to Monday morning has been detected. As this period is less than 78 hours, it is re-classified as wasted water. Wasted water events are not used to calculate avoided leak benefits.

Figure 6. Example of a leak being re-labelled as wasted water.

Avoided leaks algorithm

In addition to the identification and quantification of leaks, SWEP requires an estimation of the leak volume that was avoided as a result of SWEP's intervention. Equations 1 and 2 have been previously used to quantify the avoided leaks but the key to getting a reliable estimate on avoided leaks hinges on acquiring a reliable estimate on D_{mean}, i.e. the mean number of days until the leak is detected from quarterly bills. Two approaches were considered to estimate D_{mean}: a simple linear approach and a probability-adjusted approach. Given how quarterly bills arrive on average every 92 days, the linear approach assumes the average time from leak start to bill arrival is $\frac{92}{2}$ = 46 days. D_{mean} is thus estimated to be:

$$D_{mean} = 46 - x \qquad \text{(Eq. 3)}$$

Where x is the duration of the unavoided leak (in days). While this is an unbiased approach, it does not fully capture the probability-weighed benefits that occur for leaks that are 46 days or longer.

Figure 8 shows an example of how this can happen, in this instance for a leak that is 22 days long. In the best case scenario, the leak would occur right at the start of a billing period, which would result in avoided leak worth 70 days long. In a worst case scenario, the end of leak would coincide exactly at the end of a billing period, leading to no avoided leak. It is also assumed that if a quarterly bill occurs within the leak period, it will not incur any avoided leaks. This reflects on the uncertainty whether the leak was stopped due to SWEP or quarterly bills. As such, the duration of possible avoided leak (D_t) range from 0 to 70 days.

Given this, the expected duration of avoided leaks estimate is as follows:

$$D_{mean} = p_x * D_{arrive} \qquad \text{(Eq. 4)}$$

where p_x is the probability that the bill has not arrived during the actual leak period, and D_{arrive} is the expected number of days until the bill arrives after the leak has ended. Equation 4 can be expanded as following:

$$D_{mean} = \frac{92-x}{92} * \frac{92-x}{2} \qquad \text{(Eq. 5)}$$

Where x is the duration of the unavoided leak in days. Figure 7 shows that the probability-based approach can capture benefits for leaks that are greater than 46 days. For Equation 5, 92 days is replaced with $\frac{365}{3}$ = 121.6 days for schools billed 3 times a year.

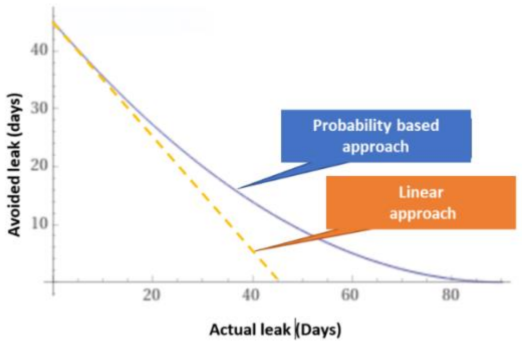

Figure 7. Estimation of avoided leak duration as a function of actual leak duration.

RESULTS/OUTCOMES

Since its inception in 2012, SWEP has introduced significant savings for schools through improving water efficiency by detecting and alerting schools to leaks. The accurate reporting of quantified and avoided leaks is crucial in feeding information back to the user and encouraging water savings, as well as justifying the water efficiency program.

This paper demonstrates the viability of developing an algorithm that can detect and quantify leaks as they occur, as well as estimate savings from avoided leaks.

From an analytical perspective, clustering time series data into dynamic segments to improve computational speed, and setting a baseline leak rate to isolate leakage from normal use were found to be key in improving leak dection accuracy. In additon, specifying bespoke analysis by time of day and day of week to capture behavioural differences, and joining neighbouring leak periods to reduce double counting of leak events further enhanced the analytical reliability of the algorithm. Finally, to avoid wasted water (from taps left running for example) and irrigation events being counted as short leak events, specific rules were set to guide the model.

CONCLUSION

For a water efficiency program to be effective in a school context, it must deploy a feedback mechanism that allows users to understand and observe the impact of water saving measures. Given this, it is crucial for the program to ensure the accuracy and reliability of quantifying leaks and savings. Modelling that can differentiate leaks from similar behaving events, such as wasted water and irrigation, can reduce false positives and prevent the overreporting of leaks.

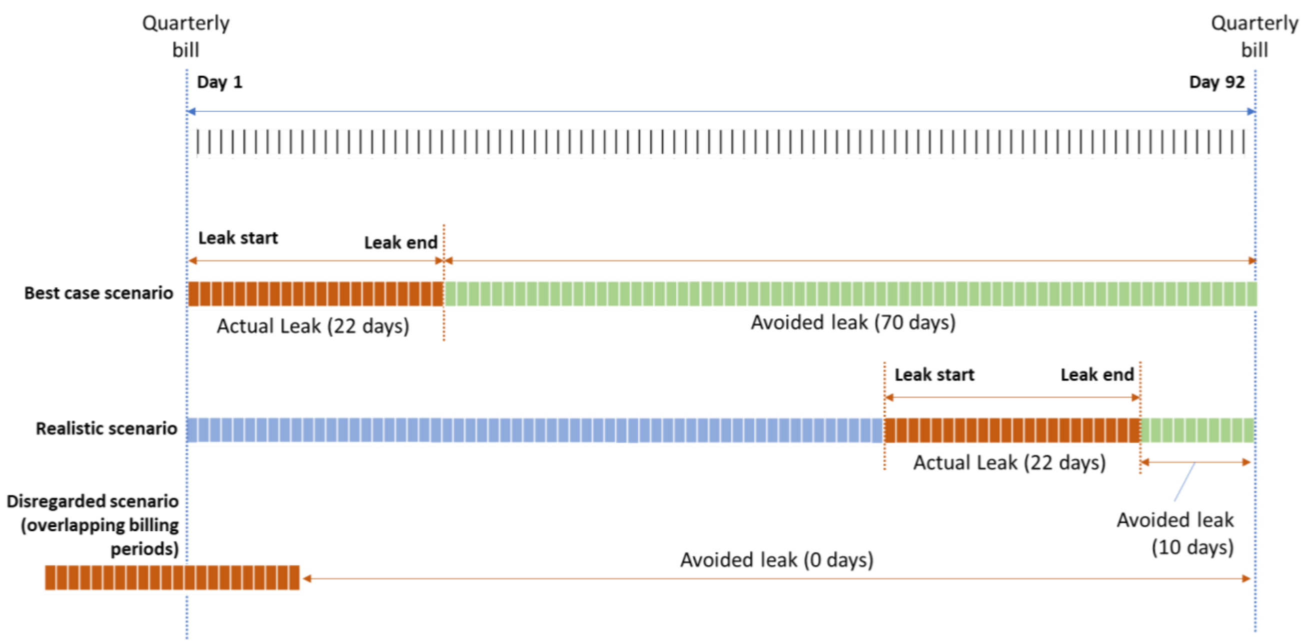

Figure 8. Scenarios showing how the timing of actual leak (red) can affect the length of avoided leak period (green).

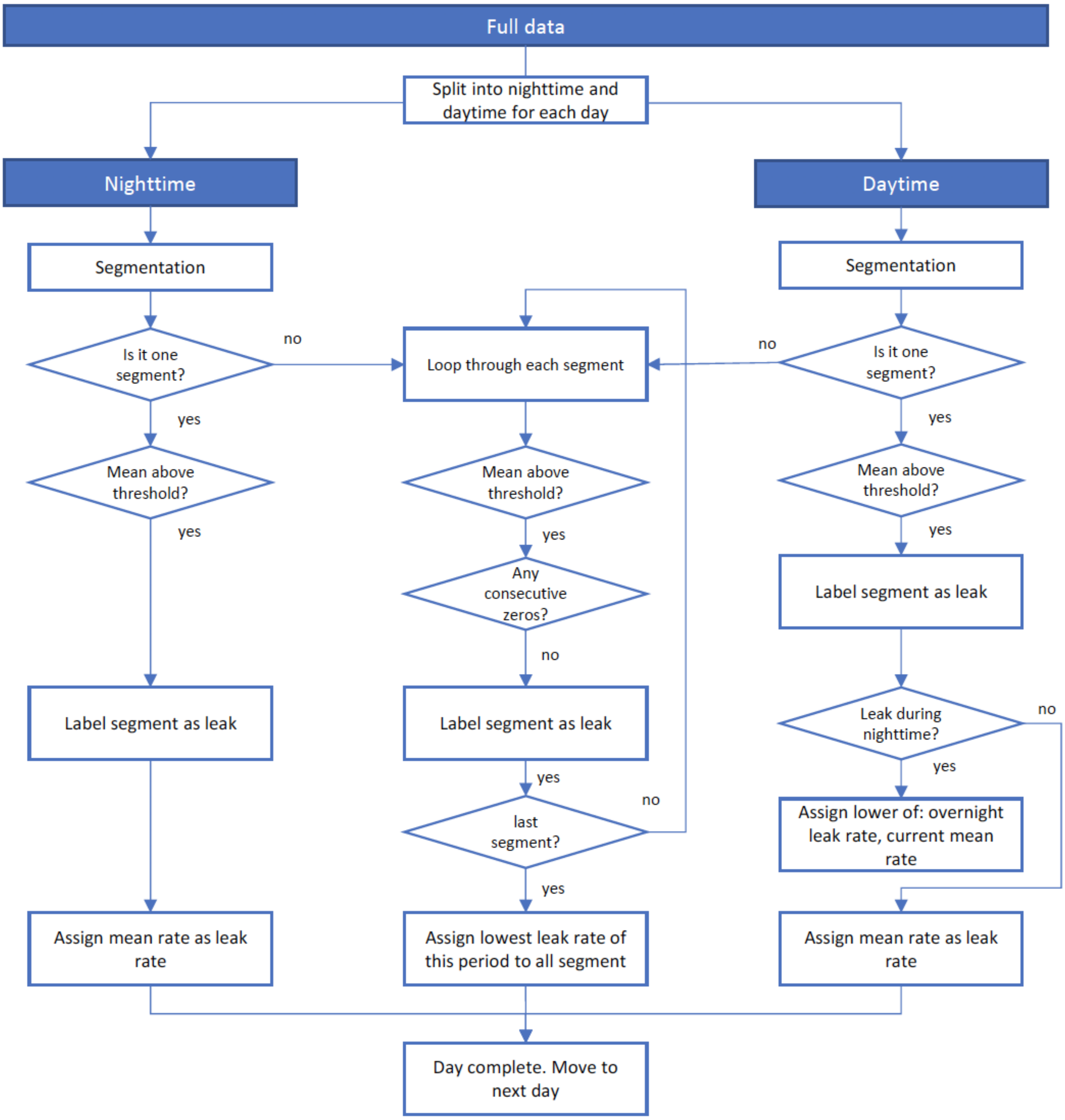

Figure 9. Overview of the decision tree based leak detection algorithm

References

1. Meteorology, B.o. National performance report 2017–18: urban water utilities, part A. **2019**.
2. Ananda, J. Determinants of real water losses in the Australian drinking water sector. *Urban Water Journal* **2019**, *16*, 575-583.
3. Brandt, M.J.; Johnson, K.M.; Elphinston, A.J.; Ratnayaka, D.D. *Twort's water supply*; Butterworth-Heinemann: 2016.
4. Taylor, W.A. Change-point analysis: a powerful new tool for detecting changes. **2000**.