

Online leakage detection system based on EWMA-enhanced Tukey method for water distribution systems

Xi Wan ^{a,*}, Raziye Farmani ^a and Edward Keedwell^b

^a Centre for Water Systems, College of Engineering, Mathematics and Physical Science, University of Exeter, Harrison Building, North Park Rd., Exeter, Devon EX4 4QF, UK

^b College of Engineering, Mathematics and Physical Science, University of Exeter, Exeter EX4 4QF, UK

*Corresponding author. E-mail: xw355@exeter.ac.uk

 XW, 0000-0002-3988-8970; RF, 0000-0001-8148-0488

ABSTRACT

Real-time leakage detection based on pressure and flow data has become increasingly essential for water distribution systems (WDSs). Recent data-driven leakage detection approaches have largely focused on burst detection characterised as sudden outflow or sudden pressure drops but did not mention the ability to detect gradual leakage events that do not have sudden change and could cause more water loss. This study proposes an online leakage detection system based on the exponential weighted moving average (EWMA)-enhanced Tukey method to help monitor gradual leakage events of WDSs. The proposed online system comprises three main parts: data pre-processing, the online detection sub-system, and the parameter updating sub-system. The proposed online system is based on lightweight and powerful statistical tools without complex model construction. The effectiveness of the proposed system is demonstrated on leakage datasets under various real-world scenarios, including gradual leakages and bursts. The results showed that the proposed EWMA-enhanced Tukey method could detect gradual leakage events quickly while generating low false alarms. The proposed method is computationally effective and able to deal with non-stationary behaviours automatically.

Key words: data-driven methods, leakage detection, online detection, statistical methods, water distribution systems

HIGHLIGHTS

- This paper proposed a leakage detection method that focused on gradual leakage events.
- The proposed leakage detection method could adapt to the time-varying characteristics of flow monitoring data, such as the demand variation caused by weather-related issues.
- This paper proposed a method that is robust to the data noises and could successfully detect leakage events without generating many false alarms.

GRAPHICAL ABSTRACT

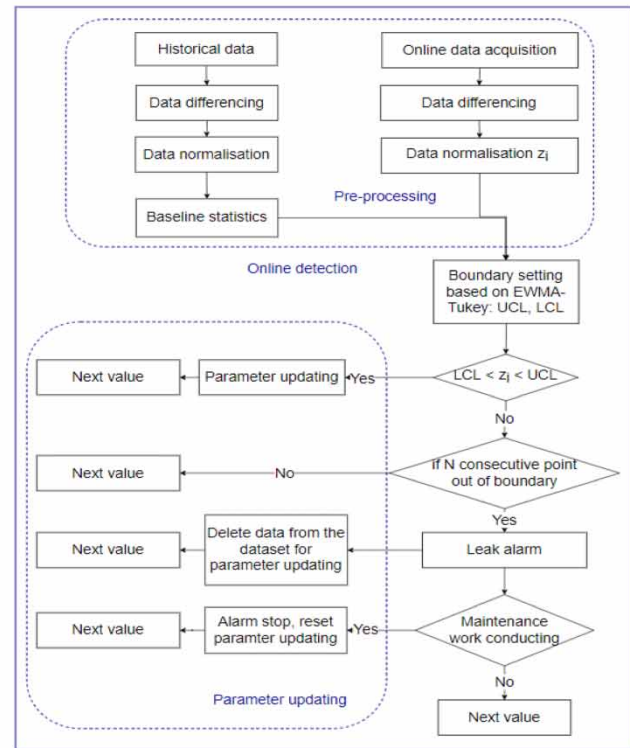
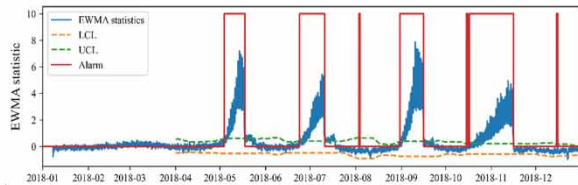
Online Leakage Detection System Based on EWMA-Enhanced Tukey Method for Water Distribution Systems

Aim:

- Online leakage detection framework
- Gradual leakage event that gradually developed over time

Results:

- Detect gradual leakage events successfully
- Low number of false alarms
- Update model parameter automatically and analyse data in a real-time manner



INTRODUCTION

In modern society, water is distributed within well-established water distribution networks (WDN). From the first pipes in Crete around 3,500 years ago to today's complex pipeline system, the water distribution system (WDS) has become one of the inevitable parts of human life. Although the freshwater available for direct human consumption only accounts for less than 1% of the Earth's water resources (Romano 2012), the demand for water resources continues to grow due to urbanisation, population, economic growth, etc. (Gupta & Kulat 2018). Furthermore, high financial and environmental costs are involved in infrastructure construction, power and chemical investment for water treatment, energy costs for pumping water, etc. (Lambert 2002). Therefore, it is becoming more and more important to reduce water loss in all aspects. The International Water Association (IWA) (Farley 2003) has defined water loss as the sum of real and apparent losses plus unbilled authorised consumption, and among them, leakage in pipelines was identified as the primary contributor of water loss. Furthermore, UK Water Industry Research (UKWIR) has identified leakage as one of their strategic priorities and has raised the question, 'How will we achieve zero leakage in a sustainable way by 2050?' Therefore, it is vital to develop methods to support this strategy and to reduce leakage as much as possible in WDSs.

In recent years, leakage detection for WDSs based on data-driven methods has received increasing attention. With the development of hydraulic sensor technology and data acquisition system, it has become possible to monitor a WDS in real-time using pressure and flow monitoring devices that have been permanently installed in the pipeline system. Abundant data sources have become available to represent the complex condition of the real-world system. However, these data are usually too numerous and complicated for humans to handle (Wu & Liu 2017). Thus, data-driven methods are needed to automatically extract valuable information and detect patterns from the big data. Moreover, unlike model-based methods (Mohammed *et al.* 2021) that require a well-calibrated hydraulic model, data-driven strategies do not require specific in-depth knowledge about the WDS and are more suitable for online control.

Various data-driven methods have been studied and developed to detect burst events in WDSs and reduce the time to awareness (Wan *et al.* 2022). With the availability of the real-time flow or pressure monitoring data collected from the Supervisory Control and Data Acquisition (SCADA) system, data-driven methods could be used to mine the historical data and find

a model to represent the condition of a WDS. Based on the representation of the distribution system's normal behaviour, new leakage events could be determined if the system's behaviour is substantially different from its normal state. Therefore, some prediction models could be used to learn historical flow or pressure data pattern and provide a prediction for reference. Mounce *et al.* (2002, 2007) introduced a burst detection system based on an artificial neural network (ANN). In addition, Kalman filter (KF) (Ye & Fenner 2011), nonlinear KF (Jung & Lansey 2015), support vector regression (SVR) (Mounce *et al.* 2011), long short-term memory (LSTM) model (Wang *et al.* 2020), and other prediction models (Bakker *et al.* 2014; Ye & Fenner 2014; Karray *et al.* 2016) have been explored for burst detection for WDSs.

Statistical process control (SPC) charts, with a set of control limits, provide intuitive and cost-effective tools to monitor and display the unusual behaviour of a process. Claudio *et al.* (2015) applied an exponential weighted moving average (EWMA) model to detect a leakage event in a DMA equipped with automated meter reading (AMR), and the awareness time is approximately 1 week after its occurrence. Borges *et al.* (2017) applied the Western Electric Company (WEC) rules for pipe burst detection, but the detection probability (DP) can be as low as 40%. Jung *et al.* (2015) compared the performance of three univariate and three multivariate SPC methods for burst detection in WDSs with consistent system operation. The result showed at least one false alarm per day, which is not promising. In their later research, Ahn & Jung (2019) proposed a hybrid approach that combined the results generated from WEC rules and the cumulative sum (CUSUM) method, but the detection time (DT) for burst events is at least 5 h. Therefore, the accuracy and applicability of the conventional SPC method are still not promising.

It should be noted that in the previous studies, leakage detection, especially burst detection, is usually defined as outlier detection, expressed as a single data point that significantly deviates from its predicted value. However, a leakage event should be described as a continuous collection of data deviating from the data pattern caused by consumers' behaviour under normal conditions. Moreover, current data-driven methods that have been proposed in the literature are mainly focused on burst detection but do not consider gradual leakage events. Unlike burst events characterised by the sudden change in a really short time period, gradual leakage events develop progressively from small seeps and weeps to noticeable leak events. As shown in Figure 1, the blue line represents the flow monitoring data, and the red area represents the duration that the distribution system contains a leakage event. It could be observed that gradual leakage could be more challenging to detect due to the small amplitude and the slow increase trend in the beginning stage. Furthermore, the seasonal trend in the flow or pressure data caused by weather change has made gradual leakage detection even more challenging. The effectiveness of a leakage detection method depends on whether the leakage can be accurately detected in a timely manner. However, rarely have studies evaluated the effectiveness of the leakage detection method for gradual leakages. Eliades & Polycarpou (2012) proposed a Fourier analysis-based fault diagnosis framework to detect leakages in WDSs and mentioned the ability of their method for gradual leakage events but did not provide a comprehensive study.

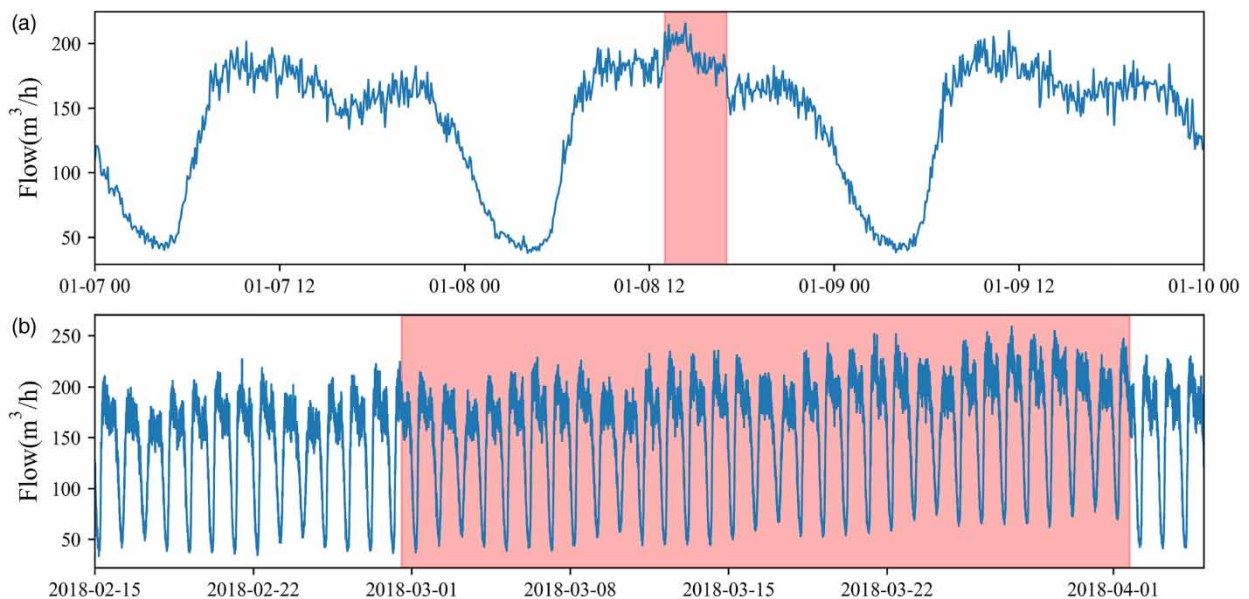


Figure 1 | Flow data contain different types of leakages: (a) burst; (b) gradual leakage. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2022.079>.

Traditional burst detection methods that aim to capture the sudden change in the data may not be suitable for gradual leakage detection. The task of detecting sudden flow increases is now transformed to detecting the trend shift. The most common method used in literature for leakage detection is the Shewhart method, which only considers the current measurement and does not retain any memory of the historical data (Kadri *et al.* 2016). Thus, the Shewhart method is not very effective when detecting gradual leakage events that cause small or moderate process mean shifts. EWMA chart is proposed by Roberts (2000) to mitigate the shortcomings of the Shewhart chart by incorporating the information from past measurements. EWMA exponentially weighted the average of all prior data so that it could utilise all the available information based on their importance to sense small changes in the process mean. Therefore, EWMA is an advanced method to monitor small or moderate shifts in process mean (Nam *et al.* 2019).

In addition, Wu & Liu (2017) provided a review of the data-driven methods for burst detection, and they pointed out that any burst detection or leakage detection method developed must be able to work in an online environment. The data come in sequence, and a decision is also required to be made in sequence. Moreover, the information of new data should be learned by the model in real time so that the model can adapt to the time-varying characteristic of data. For example, in the northern hemisphere, the consumers' demand in summer is generally higher than in winter. Thus, the control limits set in summer should be automatically different from winter. In addition, the information that comes from leakage events should be considered separately.

To address issues mentioned above, this paper proposes an online detection framework to adapt to the time-varying condition and provides an early warning system to detect gradual leak events. The proposed online system comprises three main parts: data pre-processing, the online detection sub-system, and the parameter updating sub-system. In the step of data pre-processing, data differencing and data de-seasonalisation are used to transform the raw monitoring data for the online detection phase. Then, a robust statistic-based approach called the EWMA-enhanced Tukey test is proposed to assess the sequential data and detect gradual leakage events in an online manner. The proposed approach is based on lightweight and powerful statistical tools without complex model construction. A rolling time window is used to adapt to the time-varying characteristics of the monitoring data. The parameter updating sub-system provides an information interaction to inform the online detection sub-system to absorb information and update parameters according to the system condition. The main contributions of this paper are:

1. Based on the EWMA method, this paper proposed a leakage detection method that focused on gradual leakage events;
2. Based on the use of data differencing, the proposed leakage detection method could adapt to the time-varying characteristics of flow monitoring data, such as the demand increase and decrease caused by weather-related issues;
3. Based on the use of robust statistics and the Tukey test, this paper proposed a leakage detection method that is robust to the data noises and could successfully detect leakage events without generating many false alarms;
4. Proposed an online detection framework that could analyse data in sequential order and automatically update model parameters based on the information received selectively.

Leakage datasets under various real-world scenarios are used in this research to evaluate the applicability and effectiveness of the proposed online leakage detection method. Various real-life scenarios are considered, such as daily patterns, weekly patterns, seasonal behaviour, etc. One year of real-time monitoring flow data has been used to detect gradual leakages and bursts within the WDS. Leakage events with different increasing rates and amplitude have been generated from the benchmark model to demonstrate the early warning capability and the limit of the proposed detection system.

METHODOLOGY

A diagrammatic representation of the proposed online leakage detection system is shown in Figure 2. The proposed method contains three main parts: data pre-processing, online detection sub-system, and parameter updating sub-system.

Data pre-processing

Before the detection process, the measured data need to be pre-processed so that the data can be adapted to the requirements for the chosen data mining algorithm. It should be noted that conventional SPC methods, such as Shewhart, CUSUM and EWMA charts (Jung *et al.* 2015; Ahn & Jung 2019) are based on a fundamental assumption that the data are statistically independent or stationary. Thus, before diving into the proposed methodology, an exploratory analysis is conducted to show the existence of seasonality and trend in the flow monitoring dataset. Pipe flow and pressure at demand nodes in WDS have

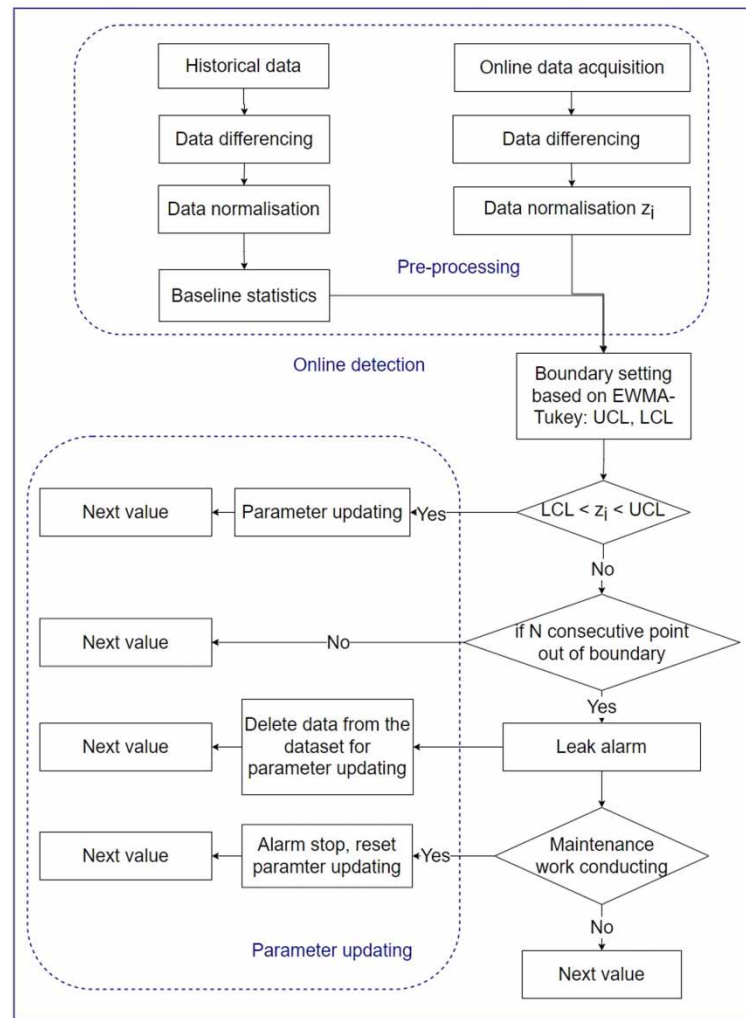


Figure 2 | Framework of the proposed online leakage detection method.

varying statistical characteristics (such as mean and standard deviation) due to regular human behaviour. Figure 3(a) shows a violin plot for the data value at each time of every day across a year of flow monitoring data. Figure 3(b) shows a violin plot for the flow value at each day of the week. It could be observed that the data in the different times of day and different days of a week follow different distributions, and the accuracy of the SPC method could be highly affected if it is applied to raw monitoring data directly. Thus, the diurnal and weekly patterns in water consumption should be considered so that dynamic thresholds can be set according to real-time behaviour. Figure 3(c) shows the average flow value at each week of the year. It could be observed that the flow data show an increasing trend over time, and this trend will cause additional requirement for a leakage detection method, namely a leakage detection method should be able to adapt to this trend without generating false alarms.

In conclusion, the trend and periodicity of the flow monitoring data pose a great challenge, and data pre-processing is inevitable for the performance of the SPC-based leakage detection methods. Therefore, the identification and removal of periodic trends and seasonal effects are conducted at this stage to prepare the data for SPC testing.

Firstly, data differencing is used to eliminate the trend in the data. It is known that water consumption varies as seasonal changes of the year. The flow monitoring data will show a growing trend in the summer and a decreasing trend in the winter. The seasonal change in the data can significantly affect the detection performance, especially for statistical analysis-based detection methods, because of the non-stationary behaviour. Data differencing is known to help eliminate (or reduce) the trend (Hyndman & Athanasopoulos 2018). According to Figure 3, diurnal pattern and weekly pattern exist in flow

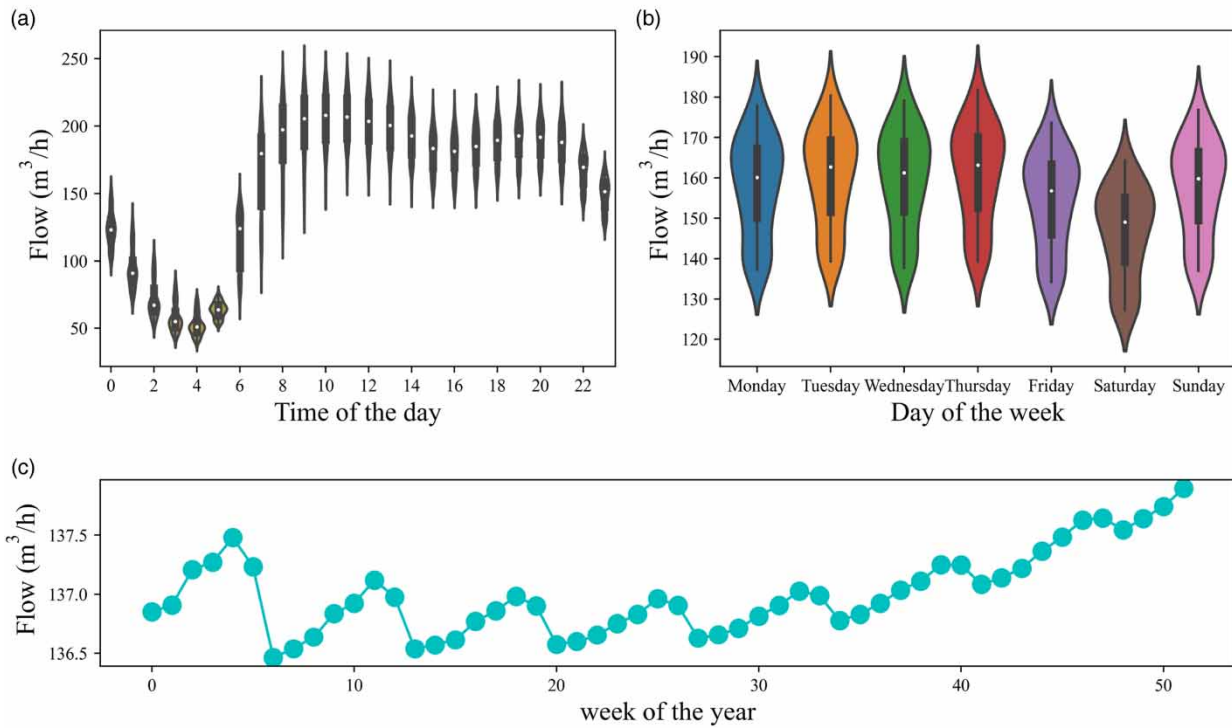


Figure 3 | Violin plot of (a) time of the day and (b) day of the week seasonal average for one of the time series; (c) the average flow of each week of the year.

monitoring data. In order to consider both seasonality, weekly differencing is adopted to preserve the trend variation caused by gradual leakage events while removing the influence of the gross trend in the monitoring data caused by the changing weather condition. The differenced flow series can be written as:

$$x_i = f_i - f_{i-m} \quad (1)$$

where x_i is the differenced value at time i represents the difference between current value and the last week's value for the same time slot; f_i is the measured flow value at time i ; m is the lag number of one week, for example, if the flow data were recorded every 5 min, $m = 7 \times 288 = 2016$.

Therefore, a normalisation process is performed after data differencing to reduce the influence of the regular pattern in the monitoring data. The normalisation process transforms the raw data into normalised scores based on the designated periodicity, such as one day or one week. One day of periodicity is commonly adopted in statistical-based leakage detection methods (e.g. Jung *et al.* 2015; Ahn & Jung 2019; Nam *et al.* 2019; Wu *et al.* 2020). This is based on the assumption that the data at the same time every day follow the same distribution. The normalised score z_i of the measured value x_i is calculated as:

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (2)$$

where μ_i and σ_i are the mean and the standard deviation of the data at the i th time period every day.

The traditional normalisation process uses mean and standard deviation, and these statistics could be easily influenced and skewed by the outlier values. Thus, instead of using mean value and standard deviation, robust statistics are adopted in this paper to reduce the influence of noises and outliers during the calculation. Moreover, one week of cyclicity is employed based on a more realistic assumption. Therefore, at first, the differenced data at the same time every week are assumed from the same data distribution and are grouped together. Then, the ordered grouped data are broken into four quarters, and the boundaries of each quarter are defined as the 1st quartile, the 2nd quartile, and the 3rd quartile. The 2nd quartile

is used to replace the mean value, and the interquartile range is used to replace the standard deviation. The interquartile range is defined as the distance between the 1st quartile and the 3rd quartile (Wang *et al.* 2011). Therefore, the measured data at time t on the k th day of the week $x_{t,k}$ are normalised by:

$$z_{t,k} = \frac{x_{t,k} - Q_{2,k}}{Q_{3,t,k} - Q_{1,t,k}} \quad (3)$$

where $z_{t,k}$ is the robust normalised score at time t in the k th day of week; $Q_{2,k}$ is the 2nd quartile or median of the data at time t in the k th day of every week; $Q_{1,t,k}$ and $Q_{3,t,k}$ are the 1st and 3rd quartile of the data at time t in the k th day of every week.

Online detection based on EWMA-enhanced Tukey method

An EWMA chart assigns different weights to the historical data, with heavier weights for more recent measurements and tailing off exponentially further in history (Dunn 2019). EWMA estimate of the normalised score z_i obtained from pre-processing $ewma_i$ could be calculated as:

$$ewma_i = \lambda z_i + (1 - \lambda)ewma_{i-1} \quad (4)$$

where λ is a weighting factor and $0 < \lambda < 1$, $ewma_0$ is the process mean of historical data.

Traditionally, the upper control limit (UCL) and lower control limit (LCL) for the EWMA are calculated in a similar way as the Shewhart limits. However, the Shewhart method is based on the assumption that the analysed data follow Gaussian distribution. It uses statistics such as mean and standard deviation that could be easily affected by one or more outliers. In order to solve this problem, the Tukey method uses quartiles that are less affected by outliers. For real-time monitoring, a rolling fixed-size data window will be considered for threshold calculating, and the threshold will be updated for each newly collected data. The Tukey UCL and LCL for the normalised score at time i is defined as:

$$UCL_i = Q_{3,i} + k(Q_{3,i} - Q_{1,i}) \quad (5)$$

$$LCL_i = Q_{1,i} - k(Q_{3,i} - Q_{1,i}) \quad (6)$$

where k is an appropriately chosen scaler for threshold setting; the $Q_{1,i}$ and $Q_{3,i}$ are the 1st and 3rd quartile of data in the i th time window $\{ewma_{i-l}, ewma_{i-l+1}, \dots, ewma_i\}$; l is the length of time window. In this step, threshold scaler k and window length l are two parameters that need to be determined by the user. Threshold scaler k determines the normal range of data value, and a rule of thumb is that k could be 2, 2.5, 3, or more (Wang *et al.* 2020). Window length l assumes that data within this time window could be considered as stationary or nearly stationary, and a rule of thumb for l could be 10 days, 20 days, 30 days, or more. User could also define these parameters based on their dataset and their engineering experience.

Parameter updating

The online detection system allows for real-time information selection by parameter updating. Three parameters should be set for this method: threshold scalar, outlier tolerance, and online window length. These parameters could be selected based on empirical experience or intuition, and the robustness of the parameter will be discussed in the next section. The information provided by online detection will be used for parameter updating. For every newly collected data, online detection can flag it as if it belongs to a leakage event or not. It should be noted that a single anomaly has a high probability that caused by consumers' behaviour or noises. Thus, it may not be the best representation of the occurrence of leakage events. In this study, a parameter called outlier tolerance N is used to provide a buffer for leakage detection, which means that the alarm will be triggered when N consecutive outliers are being detected.

If the system remains healthy, the new information will be used to update the model, and old information outside the time window will be abandoned. A rolling time window with length l is used to ensure the data used for statistics calculating is stable and unbiased, and the rolling time window strategy can provide the dynamic threshold that can adapt to the time-varying behaviour of the flow monitoring data. If a leakage event is detected, the detection system will flag it and not use it for parameter updating until there is no more alarm or the algorithm being notified that the leakage has been repaired.

Performance evaluation

Three criteria are used in this study to evaluate the performance of the proposed leakage detection system, which are DP, the number of false alarms (NF), and the DT.

DP is defined as:

$$DP = \frac{\text{number of detected leakages}}{\text{number of total leakages}} \times 100\% \quad (7)$$

In order to demonstrate the capability of the proposed online detection system in dealing with the varying property of flow data caused by weather factors, the number of false alarms is calculated on a yearly basis. A good leakage detection method should maximise the DP while minimising the number of false detections.

The DT is also an important criterion because it reflects how quickly the method can respond to an abnormal event, especially for gradual leakage events. The DT is defined as the elapsed time from the start of the leakage event to the time when the event is first detected.

RESULTS AND DISCUSSION

Description of the study area

The L-Town WDS (Figure 4) is a small hypothetical town with around 10,000 people (Vrachimis *et al.* 2020). The distribution network consists of 905 pipes, 782 junctions, 3 pressure reduction valves (PRV), 2 reservoirs, and 1 water tank. The distribution network can be divided into three areas based on the boundary conditions. A pump and a water tank have been installed in the higher part of town (Area C), and a PRV has been installed in the lower part of town (Area B). Two reservoirs are used to supply water to the network, and one water tank is located at the entrance of Area C to provide enough pressure to consumers of that area.

The monitoring system of the L-Town network contains 1 tank water level sensor, 3 flow sensors, 33 pressure sensors, and 82 automated metered readings (AMR) in Area C. It provides a dataset of pressure and flow monitoring time series and contains multiple leakage scenarios under varying conditions. It should be noted that the smart meters in Area C can provide valuable information on the consumers' behaviour. However, because most water companies do not readily have these measurements, the information provided by smart meters will not be used in this study. Therefore, this paper focuses on

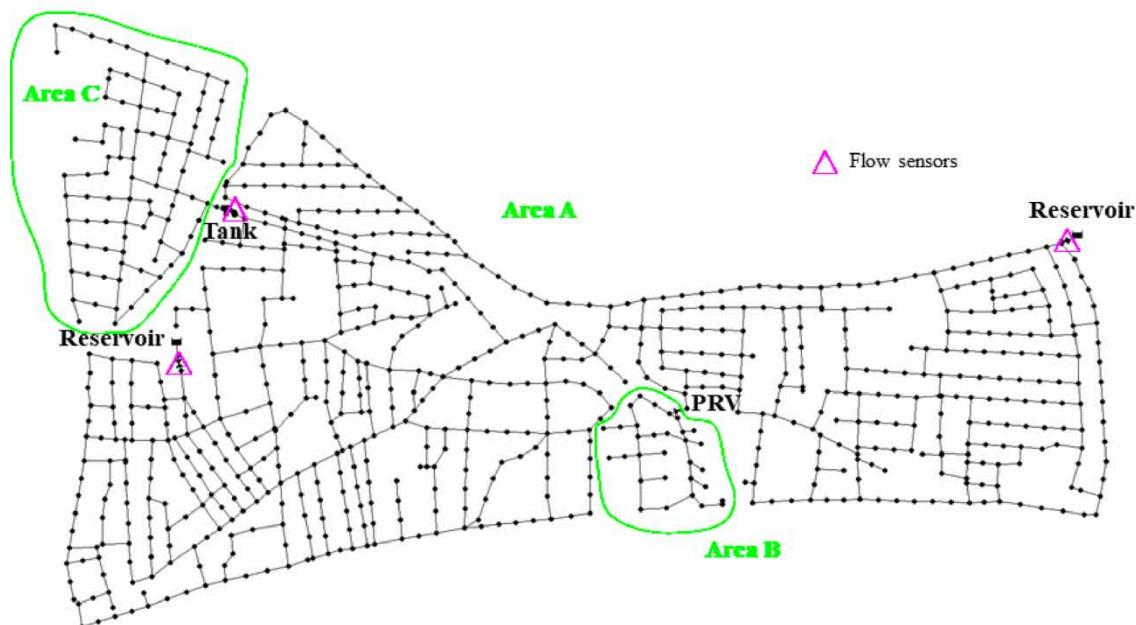


Figure 4 | Layout of the L-Town network and the location of flow sensors.

detecting leakage based on flow data only. Flow sensors are located downstream of water sources, as shown in Figure 4. In addition, flow data are mainly independent of system changing operation controls (such as pump operations) and are more related to the consumers' behaviour (Jung & Lansey 2015). Based on the inlet and outlet flow monitoring data, leakage could be detected by estimating subarea demands. The total number of monitoring stations is dependent on the number of inlets and outlets of the DMA. In this paper, three flow sensors are used to detect leakages.

Dataset generation

The hydraulic model of L-Town contains 1 year of demand pattern at each node with a sampling rate of 5 min. There are three consumer types in L-Town: residential, commercial, and industrial. Different demand patterns have been well modelled in the network, such as daily, weekly, and seasonal behaviour. EPANET is used in this paper to model and simulate the hydraulic condition of the L-Town distribution system. There are two types of hydraulic analysis: demand-driven analysis (DDA) and pressure-driven analysis (PDA). The DDA simulation assumes that node demands are known and satisfied. However, this assumption may not be valid when the system is under low-pressure conditions, such as pipe leaks or firefighting. PDA generates a more realistic representation of the pressure-leakage relationship, and many studies have reported the superiority of PDA over DDA (Baek *et al.* 2010). In pressure-dependent demand simulation, the node demand is dependent on the pressure and calculated as (Wagner *et al.* 1988):

$$q_i = \begin{cases} 0 & p_i \leq P_i^{\min} \\ q_i^{\text{req}} \left(\frac{p_i - P_i^{\min}}{P_i^{\text{req}} - P_i^{\min}} \right)^{1/e} & P_i^{\min} \leq p_i \leq P_i^{\text{req}} \\ q_i^{\text{req}} & p_i \geq P_i^{\text{req}} \end{cases} \quad (8)$$

where q_i is the actual supplied demand at node i , q_i^{req} is the required demand at node i , p_i is the pressure at node i , $P_i^{\min}$ and P_i^{req} are the minimum pressure and required pressure at node i , respectively, $1/e$ is the pressure exponent, usually set equal to 0.5 (Klise *et al.* 2017).

The leakage scenario is created using the Water Network Tool for Resilience (WNTR) (Klise *et al.* 2017) which is an open-source Python package. Leakages are modelled by splitting the pipe into two sections and adding a junction, and additional outflow is added to the leak junction. For leak scenario i , the flow rate is assumed to follow the orifice outflow formula (Crowl & Louvar 2001) as:

$$q_{\text{leak},i} = C_i A_i \sqrt{2gh_i} \quad (9)$$

where $q_{\text{leak},i}$ is the leak flow rate; C_i is the discharge coefficient with default value of 0.75 (assuming turbulent flow) (Lambert 2000); h_i is the head; g is the acceleration of gravity; and A_i is the area of the hole. For burst events, the leak area A_i remains constant during the leak period. For gradual leakage events, the leak area A_i will gradually increased overtime until it reaches its maximum value $A_{i,\text{max}}$. The leak area $A_{i,\text{max}}$ is defined as a circular orifice with a diameter D_i . The relationship of the size of the leak hole change with time is assumed to follow a quadratic function so that the growth rate of leakage is increased as the leak area increases (Vrachimis *et al.* 2020).

$$A_i(t) = \frac{A_{i,\text{max}}}{T_i^2} \times t^2 \quad (10)$$

$$A_{i,\text{max}} = \frac{1}{4} \pi D_i^2 \quad (11)$$

where T_i is the growing duration that the i th gradual leakage needed to reach its maximum value.

A number of leakage events with different leak magnitudes and different growing durations that happened at different times in a year have been simulated by WNTR. Dataset 0 contains six leakage events, including four gradual leakage events, and two burst events. Figure 5(a) shows the flow monitoring dataset that contained those six leakage events, and Figure 5(b) shows the leakage profile in the system. Detailed information about the leakages in Dataset 0 is listed in Table 1. The start and end times are when the leakage is started and fixed in the distribution system. The peak time is defined as the time

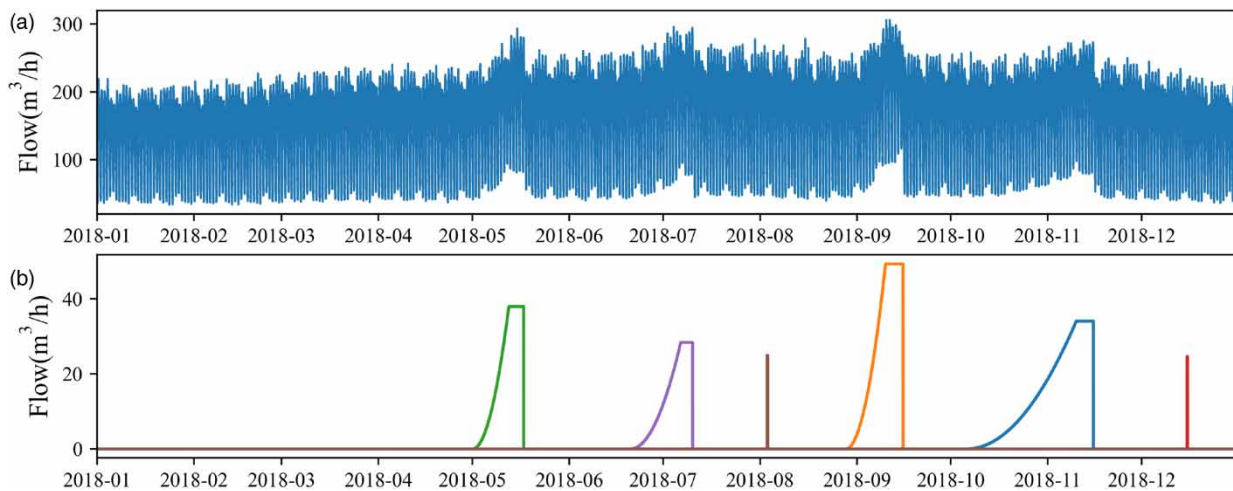


Figure 5 | Monitoring data of area A of L-Town with six leakages: (a) flow data; (b) leakage profile.

when the leak diameter reaches its maximum level. From Table 1, it could be observed that even though the gradual leakage events have higher leak volume than burst events, the daily increase of leak volume is much slower than the burst events. For example, leak number 1 is a slowly developed leakage event that reaches its peak value over about 11 days. This leak increased by 23.55% of the average water demand, which means a daily increase of just about 2.09%. As shown in Table 1, the daily increase of leak volume of gradual leakage events is less than 3% of the average demand, and the daily increase of burst events is more than 20% of the average demand. The slow-developing pattern makes the gradual leakage events more difficult to detect.

The longer the duration between the start time and the peak time, the slower the growth process of the leak, which means that it will be more difficult to detect the leak. In order to demonstrate the limit of the proposed method, 40 other datasets have been generated. Each dataset contains two gradual leakage events with the same leakage size and the same leakage duration, one occurred in the summer, and one occurred in the winter. Leakage with different leak diameter {13mm, 14mm, 15mm, ..., 22mm} is generated with different leak growing duration {10days, 15days, 20days, 25days} for each dataset. After several leakage scenarios were inserted in the distribution network, monitoring data could be generated with a sampling interval of 5 min.

Application results

Parameter estimation

There are three parameters that need to be determined for the proposed leakage detection method, which are threshold scalar k , outlier tolerance N , online window length l . The best value for parameters are identified based on a sensitivity analysis that can be seen in Figure 6. Candidates of threshold scalar k are selected as $k = 2, 2.25, 2.5, 2.75, 3$. Candidates of outlier tolerance N are selected as $N = 1, 2, 3, 4, 5, 6$. Candidates of online window length l are selected as $l = 10$ days, 20 days, 30 days, 40 days, and 50 days. Therefore, 150 different combinations of the aforementioned parameters are used to conduct the sensitivity analysis of the proposed EWMA-enhanced Tukey online leakage detection method.

Figure 6 shows the DP and NF (per year) of the proposed method with different parameter combinations. Threshold scalar k determines the range of threshold, and a larger k indicates a wider range of data will be considered as normal. As shown in Figure 6, with the same parameter setting for N and l , a larger k indicates a higher probability of undetected leakage events but could reduce the number of false alarms. Outlier tolerance N represents the size of the buffer for raising a leakage alarm. A larger N indicates greater tolerance for outliers, reduces the probability of false alarms but may delay the DT. Online window length l determines the length of historical data that will be considered at each time step for the statistics calculation so that the threshold could adapt to the time-varying characteristic of flow data caused by environmental factors (such as weather change). If the window length is too short, there are not enough data samples for accurate statistics estimation, but if the window length is too long, the assumption about stationery could be easily disobeyed. Furthermore, the three

Table 1 | Leak information considered in Dataset 0

Leak	Type	Leak diameter (mm)	Peak leak volume (m ³ /h)	Peak leak volume of average demand	Start time	Peak time	End time	Leak growth volume (m ³ /h /day)	Leak growth volume of average demand
1	Gradual	20.11	37.97	23.55%	1 May 09:20	12 May 16:05	17 May 09:20	≈3.366	2.09%
2	Gradual	17.39	28.40	17.61%	20 Jun. 15:45	6 Jul. 15:45	10 Jul. 10:25	≈1.775	1.10%
3	Burst	20.00	24.98	15.49%	3 Aug. 07:00	3 Aug. 07:00	3 Aug. 11:00	≈24.96	15.48%
4	Gradual	22.92	49.32	30.59%	28 Aug. 10:35	10 Sep. 02:45	15 Sep. 17:30	≈3.884	2.41%
5	Gradual	19.04	34.06	21.12%	06 Oct. 02:35	10 Nov. 02:35	15 Nov. 13:35	≈0.973	0.60%
6	Burst	20.00	24.61	15.26%	15 Dec. 13:00	15 Dec. 13:00	15 Dec. 17:00	≈24.61	15.26%

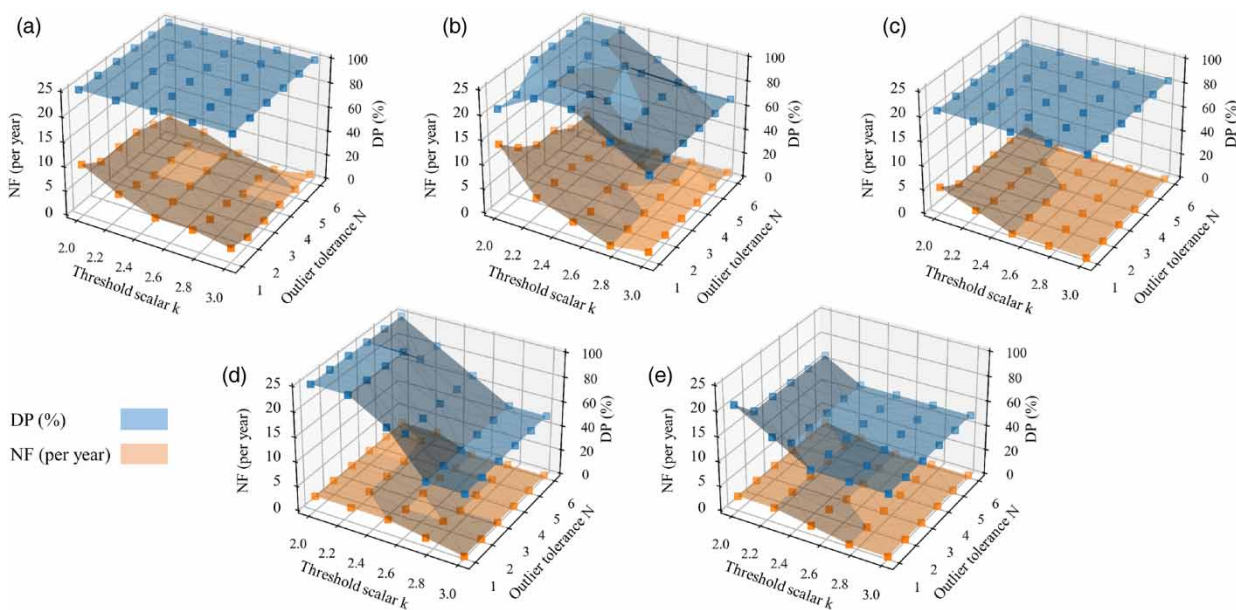


Figure 6 | EWMA-enhanced Tukey method’s detection performance for different threshold scalar k , outlier tolerance N , and online window length l : (a) $l = 10$ days; (b) $l = 20$ days; (c) $l = 30$ days; (d) $l = 40$ days; (e) $l = 50$ days.

parameters have influences on each other. Therefore, these three parameters need to be tuned to achieve a relatively high DP, low number of false alarms per year, and quick DT.

From Figure 6, it could be observed that the proposed method is robust to threshold scalar settings and outlier tolerance settings with a designated rolling window length. For example, when the window length equals 10 days and 20 days, the DP is maintained at 100% no matter the value of k and N (within a certain range). In most parameter settings, the number of false alarms is less than five per year. Based on the sensitivity analysis, there are two parameter sets that could achieve 100% accuracy and 0 false alarm at the same time: (1) $k = 2.5$, $N = 4$, $l = 20$ and (2) $k = 2.75$, $N = 6$, $l = 10$. Table 2 shows the final detection results for Dataset 0 based on the two potential parameter settings. It could be observed that parameter set (2) has outperformed parameter set (1) in DT for all gradual leakage events, but is slightly lower for burst detection. Since this paper aims to develop an early warning system for gradual leakage, DT for gradual leakage events is one of the primary considerations for the proposed method. Therefore, the EWMA-enhanced Tukey method’s parameters k , N , l were set to 2.5, 4, 20, respectively.

Comparison between EWMA-based method and Shewhart-based method

Conventional Shewhart-based method or three-sigma rule only uses the last data sample to make decisions and does not have any memory of previous data (Kadri et al. 2016). In contrast, EWMA-based monitoring charts take into account the historical information by using a weighted average of past observations, which makes it more suitable for detecting gradual anomalies.

Table 2 | Detection results for dataset 0 based on two different parameter settings of the proposed method

Leak	Type	DT (Parameter 1)	DT (Parameter 2)
1	Gradual	3 d 3 h 40 min	5 d 55 min
2	Gradual	3 d 14 h	3 d 16 h 25 min
3	Burst	3 h 50 min	1 h 55 min
4	Gradual	2 d 18 h 40 min	3 d 18 h 10 min
5	Gradual	9 d 3 h	9 d 3 h 10 min
6	Burst	2 h 45 min	2 h

In order to compare the proposed EWMA-enhanced Tukey method to the commonly-used three-sigma rule (Dunn 2019) in detecting gradual leakage events, the Shewhart method has been extended with the proposed online framework using the same pre-processing stage and the parameter updating sub-system.

Figure 7(a) shows the detection results of the online detection process by the proposed method. The blue line represents the EWMA statistics for the transformed flow data obtained from the pre-processing stage. The green and orange dashed lines represent the UCL and the LCL calculated online for the EWMA statistics, respectively. The red line represents the final alarm generated by the proposed algorithm. The alarm is generated as a binary result, which means that there is no leakage when the alarm is 0. Otherwise, the alarm will be raised when the algorithm detects leakage. Compared Figure 7(a) and 7(b), it could be observed that Shewhart method fails to detect several leakage events due to its inefficiency in dealing with data noise and data uncertainties. Furthermore, the Shewhart method is based only on the current data, and all past information is ignored, which makes it unsatisfactory when used for detecting small changes (Harrou *et al.* 2020). However, the gradual leakage event does not have the ‘spike’ that distinguishes the event from the normal condition. Because of the nature of slow-changing pattern of gradual leakage events, the probability of the Shewhart method detecting leakage events is low. In contrast, the proposed EWMA-enhanced Tukey method considers the historical information by assigning weights to past observations.

Table 3 shows the detection results of the proposed method and the online version of the Shewhart method. As shown in Table 3, the Shewhart method detects gradual leakage events at a much lower rate than the proposed EWMA-enhanced Tukey method. The proposed method successfully detects all the leakage events within a short duration. All burst events are successfully detected, and almost all gradual leakage events are detected within 4 days, except leak No. 5. The DT of leak No. 5 is around 9 days. This is because the daily growth rate of leak No. 5 is about 0.5% of the average water demand, which is a very small increase. Table 3 also shows the amount of leakage flowrate when the leakage is first being detected. The proposed method detects all gradual leakage events before they reach 4 m³/h. However, the Shewhart method raises alarms when flow rates of leakages reach more than 15 m³/h. The results show the superiority of the EWMA-based method for gradual leakage events detection.

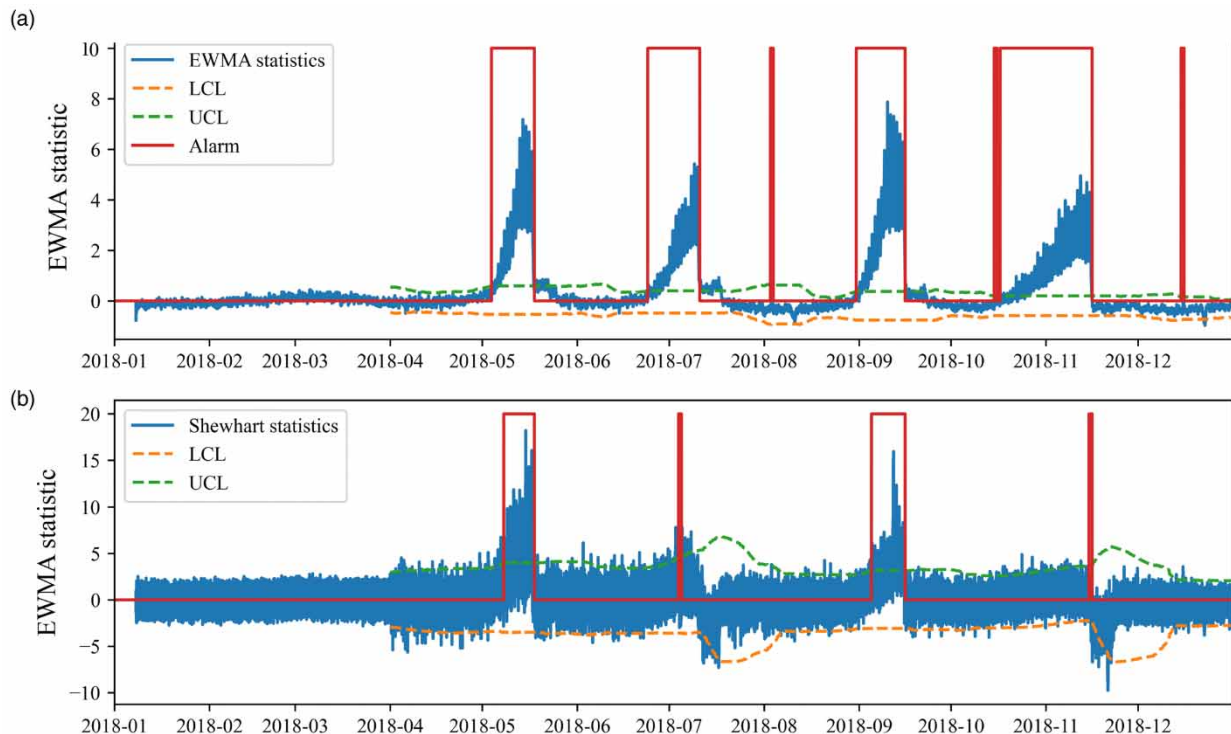


Figure 7 | Detection process based on (a) the proposed EWMA-enhanced Tukey method; (b) the Shewhart-based method. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/hydro.2022.079>.

Table 3 | Detection results for dataset 0 based on the proposed method and the Shewhart method

Leak	Type	Peak leak volume (m ³ /h)	DT (EWMA-enhanced Tukey)	LF (EWMA-enhanced Tukey)	DT (Shewhart)	LF (Shewhart)
1	Gradual	37.97	3.16 days	3.35	7.03 days	15.16
2	Gradual	28.40	3.61 days	1.43	13.57 days	20.42
3	Burst	24.98	3.83 h	24.98	–	–
4	Gradual	49.32	2.78 days	2.37	7.61 days	17.83
5	Gradual	34.06	9.16 days	2.32	–	–
6	Burst	24.61	2.75 h	24.61	–	–

– means the detection method fails to detect the leakage event.

LF means the amount of leakage flowrate when first being detected (m³/h).

Comparison between EWMA-enhanced Tukey method with or without data differencing

Data differencing is one of the important parts of the proposed leakage detection method. Seasonal differencing can effectively reduce the influence of global trends in the data caused by weather factors. Water usage increases from spring to summer and decreases from summer to winter gradually. One difficulty for gradual leakage detection is that the global trend could be easily confused with the increasing trend caused by gradual leakage. In order to satisfy the assumption of stationarity (or as close as possible), data differencing is used in this paper to eliminate the global trend in the flow data while preserving the increasing trend caused by gradual leakages. To evaluate the usefulness of this procedure, methods without data differencing and methods with data differencing were compared.

The performance of the differencing-based and the non-differencing-based EWMA-enhanced Tukey method is demonstrated in Figure 8. Figure 8(a) shows the detection process based on the proposed EWMA-enhanced Tukey method with data differencing, and Figure 8(b) shows the detection process based on the EWMA-enhanced Tukey method without data

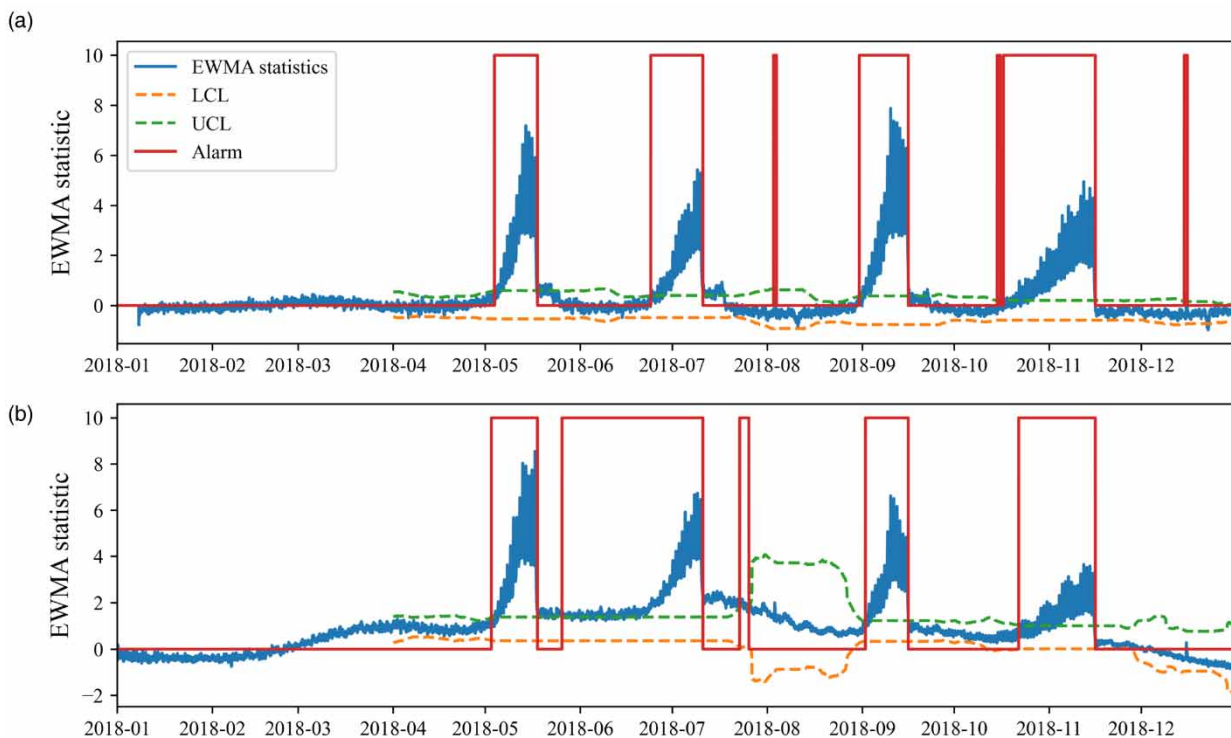


Figure 8 | Detection process based on (a) the proposed EWMA-enhanced Tukey method; (b) non-differencing-based EWMA-enhanced Tukey method.

differencing. In Figure 8(b), the false alarm lasts for months after detecting the first gradual leakage. Gradual leakage could be extended for weeks or months, and the global trend is coupled with the trend caused by leakage events. Thus, right after the leakage events are finished, it is not easy to estimate the threshold accurately with the existence of a global trend. Figure 8(a) shows that data differencing has eliminated most of the global trends compared with non-differencing statistics. Therefore, it could be concluded that data differencing has ensured the accurate detection of the proposed method.

Comparison between EWMA-enhanced Tukey method with or without robust statistics

Another important part of the proposed methodology is the adoption of robust statistics. The use of robust statistics in the data transformation and the use of the Tukey method in the threshold setting stage enhanced the robustness of the proposed online leakage detection method. The traditional methods (Jung *et al.* 2015; Ahn & Jung 2019; Nam *et al.* 2019) use non-robust statistics such as mean and standard deviation that are not robust to noises or outliers, leading to unsatisfying detection performance. Therefore, the proposed method uses robust statistics, such as the median, which is robust against noise.

Figure 9(b) shows the detection results based on non-robust statistics. Compared with Figure 9(a), it could be observed that without using robust statistics, the results generated by the proposed method contained more uncertainties and have obviously generated more false alarms. The results show that robust statistics have greatly reduced false alarms from 7 per year to 0 per year during the detection process. Compared Figures 5(a)–9(a), the proposed EWMA-enhanced Tukey method has successfully transformed the raw data and highlighted the abnormal behaviour caused by leakage events. Furthermore, the parameter updating sub-system ensured the proposed method could process data in real-time and automatically include the information of new data.

Detection ability

In order to show the robustness of the proposed method, all six leakage events in Table 1 have been shuffled randomly in the dataset. The order of the six leakage events and the start time of each leakage event have been shuffled randomly three times. Table 4 shows the detection results of three shuffled datasets. According to Table 4, the proposed algorithm maintained 100% DP while the Shewhart-based method missed several leakage events on each dataset. The proposed algorithm did not raise

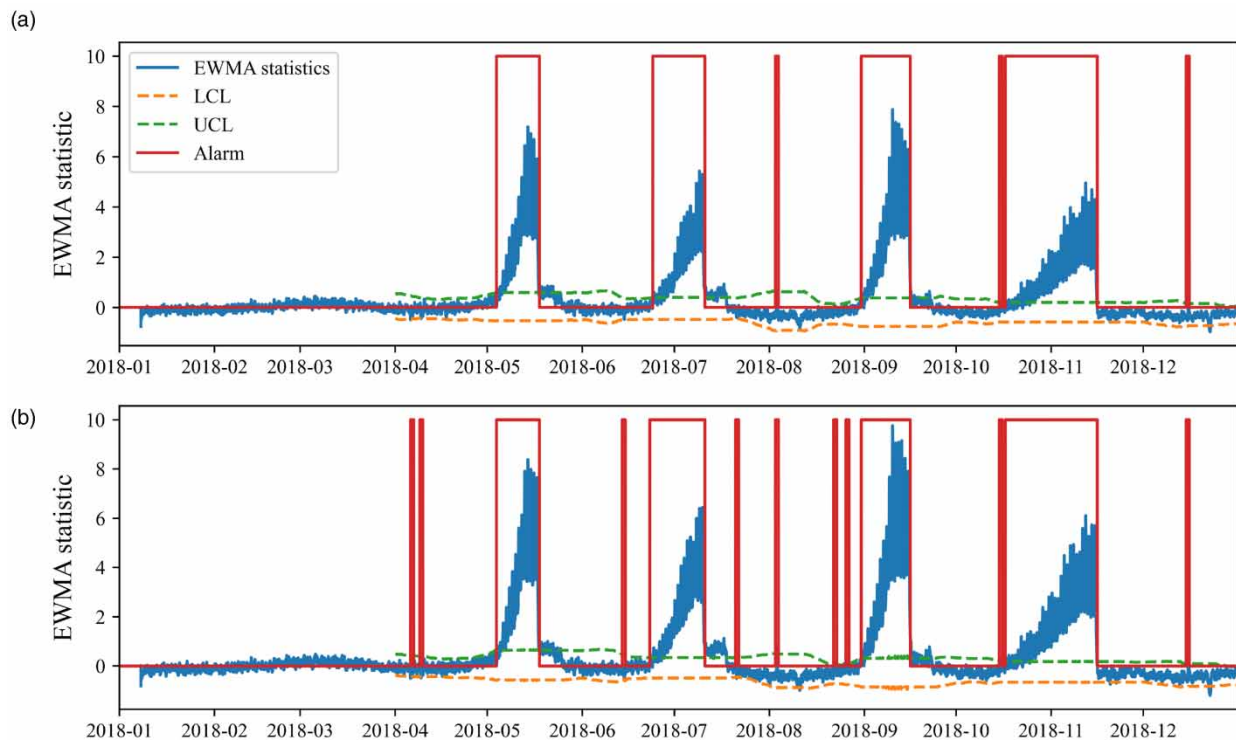


Figure 9 | Detection process based on (a) the proposed EWMA-enhanced Tukey method; (b) non-robust statistics-based EWMA detection method.

Table 4 | Detection results of three shuffled datasets based on the EWMA-enhanced Tukey test and the Shewhart-based method

Shuffled dataset	NF (per year)		Leak	Type	Leak growth volume of average demand	DT	
	EWMA-enhanced Tukey	Shewhart				EWMA-enhanced Tukey	Shewhart
1	0	0	1	Burst	15.26%	1 h	1.83 h
			2	Gradual	2.41%	4.73 days	8.80 days
			3	Burst	15.48%	2.75 h	–
			4	Gradual	0.60%	20.12 days	–
			5	Gradual	2.09%	5.18 days	10.28 days
			6	Gradual	1.10%	8.56 days	–
2	1	0	1	Gradual	2.41%	3.73 days	7.69 days
			2	Burst	15.26%	–	–
			3	Gradual	2.09%	5.01 days	8.05 days
			4	Gradual	0.60%	9.05 days	–
			5	Burst	15.48%	0.92 h	2.25 h
			6	Gradual	1.10%	8.60 days	–
3	0	0	1	Burst	15.48%	1.5 h	–
			2	Gradual	2.09%	4.10 days	9.01 days
			3	Gradual	0.60%	25.13 days	–
			4	Burst	15.26%	1.58 h	–
			5	Gradual	1.10%	8.48 days	–
			6	Gradual	2.41%	5.85 days	8.69 days

– means the detection method fails to detect the leakage event.

any false alarms except one false alarm in shuffled dataset 2, while the Shewhart-based method did not raise any false alarms. Furthermore, for the detected leakage events, the proposed method detected all leakage events quicker than the Shewhart-based method all the time.

From Tables 3 and 4, it can be observed that for a small leakage event, the method needs a longer time to detect its occurrence. For example, the No. 5 leakage event has the most prolonged grow duration and relatively small magnitude, making it the most difficult to detect. When the amplitude of a leakage event becomes too small, it could be easily confused with normal consumption behaviour or variation caused by weather and resulting in undetectable leakages. To demonstrate the limit of the proposed methods, another 40 datasets that contain leakage events with different diameter {13mm, 14mm, 15mm, ..., 22mm} and different grow duration {10 days, 15 days, 20 days, 25 days} have been used. The grow duration is defined as the duration between the leak start time and leak peak time. Table 5 shows the detection results

Table 5 | Average DT for dataset 1–40 based on EWMA-enhanced Tukey test and the Shewhart-based method

Leak diameter (mm)	Peak leak magnitude (m ³ /h)	DT based on the proposed method				DT based on the Shewhart-based method			
		d = 10	d = 15	d = 20	d = 25	d = 10	d = 15	d = 20	d = 25
13	10.63	4.82	–	–	–	–	–	–	–
14	12.32	4.78	6.77	–	–	–	–	–	–
15	14.14	4.75	6.76	–	–	–	–	–	–
16	16.08	4.67	6.76	19.73	–	13.75	–	–	–
17	18.14	4.62	6.76	19.73	–	13.73	–	–	–
18	20.33	2.96	5.74	12.86	–	13.73	14.77	–	–
19	22.64	2.96	4.82	6.76	–	12.78	14.77	–	–
20	25.06	2.96	4.78	6.76	26.69	12.78	19.73	–	–
21	27.61	2.95	4.78	6.76	26.69	6.75	19.73	–	–
22	30.27	2.88	4.77	6.76	12.86	6.72	19.73	–	–

– means the detection method fails to detect the leakage event.

d means the time duration it takes for the leak to grow to its maximum value.

for those 40 datasets. Each dataset contains two leakage events with the same growth duration and amplitude, but one happened in summer, and one happened in winter.

Table 5 shows that the DT could be affected by both leakage magnitude and leakage growth rate. As the leakage magnitude becomes smaller, the DT could be delayed and eventually it will remain undetectable in the system. This is because the leakage volume is small, and it will be regarded as normal water consumption by the algorithm. Furthermore, as the leakage growth duration becomes longer, the DT could also be delayed and eventually fail to raise the alarm. This is because the growing trend caused by leakage is too small, and eventually, it becomes indistinguishable from the growing trend caused by weather factors. In most cases, the detection algorithm could raise the alarm at least 13 days before the leakage reaches its maximum level.

In comparison, the detection results based on the traditional Shewhart method are presented in Table 5. It could be observed that the traditional method failed to detect most of the gradual leakage events, especially when the growing trend of the leakage is relatively slow. In addition, the DT based on the Shewhart-based method is much slower than the proposed method, with detection times at least 10 days slower in each case.

CONCLUSIONS

Gradual leakage events are more challenging to be detected than burst events due to their long-term influence and unnoticeable behaviour. This study proposed a novel EWMA-enhanced Tukey method to detect gradual leakage events in a real-time manner for water distribution systems. This method used simple, computationally lightweight but powerful and robust statistics to detect leakage events in the WDS. First, the raw monitoring data were pre-processed using data differencing and data transformation techniques. Then, the transformed data were processed based on the EWMA-enhanced Tukey method. Based on the results of the online leakage detection, parameter updating was proposed to update the model automatically and ensure the proposed method could process streaming data in an online manner.

The proposed method has been successfully applied for a case study with a year of monitoring data and proved effective in real-time monitoring. The 1-year monitoring datasets contain four gradual leakage events and two burst events, and the detection results showed that the proposed method has successfully detected all leakage events. All leakage events were detected with a short DT and did not generate any false alarms in a year, which showed a promising future for this method.

The comparison analysis showed that the improvement of the method had a significant impact on the detection performance. Data differencing has proven effective when dealing with seasonal behaviour in the consumers' demand, and it ensured the proposed algorithm could be successfully applied in real-time. Robust statistics have proved to be helpful in eliminating the influence of data uncertainties and reducing the number of false alarms.

There is no leakage detection method that could detect all leaks of any magnitude. Therefore, in this paper, the detection ability was tested for the proposed method. In this study, leakage events that continuously happened for weeks have been successfully detected with leak diameter larger than 20 mm or growth duration less than around 20 days.

It should be noted that even though the proposed methodology has presented promising behaviour in dealing with the real-time gradual leakage event detection, the real-life monitoring data may contain more challenging uncertainties, such as missing data, system changing operation, and human behaviour changing caused by pandemics such as Covid-19. This will be studied in our future research. In addition, the proposed detection algorithm could be failed if the leakage event is too small or slowly happens in the system, and a method with higher accuracy could be further developed in future research.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Ahn, J. & Jung, D. 2019 Hybrid statistical process control method for water distribution pipe burst detection. *J. Water Resour. Plann. Manage.* **145**, 06019008. doi:10.1061/(ASCE)WR.1943-5452.0001104.

- Baek, C. W., Jun, H. D. & Kim, J. H. 2010 Development of a PDA model for water distribution systems using harmony search algorithm. *KSCE J. Civ. Eng.* **14**, 613–625. doi:10.1007/s12205-010-0613-7.
- Bakker, M., Vreeburg, J. H. G., Van De Roer, M. & Rietveld, L. C. 2014 Heuristic burst detection method using flow and pressure measurements. *J. Hydroinf.* **16**, 1194–1209. doi:10.2166/hydro.2014.120.
- Borges, A. P. A., Jung, D. & Kim, J. H. 2017 Smart WDS management: pipe burst detection using real-time monitoring data. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, pp. 1–4.
- Claudio, K., Couallier, V., Leclerc, C., Le Gat, Y., Litrico, X. & Saracco, J. 2015 Detecting leaks through AMR data analysis. *Water Sci. Technol. Water Supply* **15**, 1368–1372. doi:10.2166/ws.2015.071.
- Crowl, D. A. & Louvar, J. F. 2001 *Chemical process safety: fundamentals with applications*. Pearson Education, USA.
- Dunn, K. 2019 *Process improvement using data*. Available from: <http://learnche.org/Pid>.
- Eliades, D. G. & Polycarpou, M. M. 2012 Leakage fault detection in district metered areas of water distribution systems. *J. Hydroinf.* **14**, 992–1005. doi:10.2166/hydro.2012.109.
- Farley, M. 2003 Non revenue water – international best practice for assessment, monitoring and control. In *12th Annual CWWA Water, Wastewater & Solid Waste Conference*. pp. 1–18.
- Gupta, A. & Kulat, K. D. 2018 A selective literature review on leak management techniques for water distribution system. *Water Resour. Manage.* **32**, 3247–3269. doi:10.1007/s11269-018-1985-6.
- Harrou, F., Sun, Y. S., Hering, A., Madakyaru, M. & Dairi, A. 2020 *Statistical process monitoring using advanced data-driven and deep learning approaches*. Elsevier.
- Hyndman, R. J. & Athanasopoulos, G. 2018 *Forecasting: principles and practice*. OTexts. Available from <http://OTexts.com/fpp2/>.
- Jung, D. & Lansey, K. 2015 Water distribution system burst detection using a nonlinear Kalman filter. *J. Water Resour. Plann. Manage.* **141**, 04014070. doi:10.1061/(asce)wr.1943-5452.0000464.
- Jung, D., Kang, D., Liu, J. & Lansey, K. 2015 Improving the rapidity of responses to pipe burst in water distribution systems: a comparison of statistical process control methods. *J. Hydroinf.* **17**, 307–328. doi:10.2166/hydro.2014.101.
- Kadri, F., Harrou, F., Chaabane, S., Sun, Y. & Tahon, C. 2016 Seasonal ARMA-based SPC charts for anomaly detection: application to emergency department systems. *Neurocomputing* **173**, 2102–2114. doi:10.1016/j.neucom.2015.10.009.
- Karray, F., Garcia-Ortiz, A., Jmal, M. W., Obeid, A. M. & Abid, M. 2016 EARNPIPE: a testbed for smart water pipeline monitoring using wireless sensor network. *Procedia Comput. Sci.* **96**, 285–294. The Author(s). doi:10.1016/j.procs.2016.08.141.
- Klise, K. A., Hart, D., Moriarty, D. M., Bynum, M. L., Murray, R., Burkhardt, J. & Haxton, T. 2017 *Water Network Tool for Resilience (WNTR) User Manual*.
- Lambert, A. 2000 What do we know about pressure: leakage relationships in distribution systems? In *IWA Conference Systems Approach to Leakage Control and Water Distribution System*, pp. 1–8.
- Lambert, A. O. 2002 International report: water losses management and techniques. *Water Sci. Technol. Water Supply* **2**, 1–20. doi:10.2166/ws.2002.0115.
- Mohammed, E. G., Zeleke, E. B. & Abebe, S. L. 2021 Water leakage detection and localization using hydraulic modeling and classification. *J. Hydroinf.* **23**, 782–794. doi:10.2166/hydro.2021.164.
- Mounce, S. R., Day, A. J., Wood, A. S., Khan, A., Widdop, P. D. & Machell, J. 2002 A neural network approach to burst detection. *Water Sci. Technol.* **45**, 237–246. doi:10.2166/wst.2002.0595.
- Mounce, S. R., Boxall, J. B. & Machell, J. 2007 An artificial neural network/Fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation. In: *Proc. Comb. Int. Conf. Comput. Control Water Ind. CCWI2007 Sustain. Urban Water Manag. SUWM2007*, January, pp. 313–320.
- Mounce, S. R., Mounce, R. B. & Boxall, J. B. 2011 Novelty detection for time series data analysis in water distribution systems using support vector machines. *J. Hydroinf.* **13**, 672–686. doi:10.2166/hydro.2010.144.
- Nam, K. J., Ifaei, P., Heo, S., Rhee, G., Lee, S. & Yoo, C. K. 2019 An efficient burst detection and isolation monitoring system for water distribution networks using multivariate statistical techniques. *Sustainability* **11**. doi:10.3390/su11102970.
- Roberts, S. 2000 Control chart tests based on geometric moving average. *Technometrics* **42** (1), 97–101.
- Romano, M. 2012 *Near real-time detection and approximate location of pipe bursts and other events in water distribution systems*. University of Exeter, Exeter.
- Vrachimis, S. G., Eliades, D. G., Taormina, R., Ostfeld, A., Kapelan, Z., Liu, S., Kyriakou, M., Pavlou, P., Qiu, M. & Polycarpou, M. M. 2020 BattLeDIM: battle of the leakage detection and isolation methods. In *2nd International CCWI/WDSA Joint Conference*. pp. 1–6.
- Wagner, J. M., Shamir, U. & Marks, D. H. 1988 Water distribution reliability: simulation methods. *J. Water Resour. Plann. Manage.* **114** (3), 276–294.
- Wan, X., Kuhanestani, P. K., Farmani, R. & Keedwell, E. 2022 Literature review of data analytics for leak detection in water distribution networks: a focus on pressure and flow smart sensors. *J. Water Resour. Plann. Manage.* **148** (10), 013122002.
- Wang, C., Viswanathan, K., Choudur, L., Talwar, V., Satterfield, W. & Schwan, K. 2011 Statistical techniques for online anomaly detection in data centers. In: *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*. pp. 385–392.

- Wang, X., Guo, G., Liu, S., Wu, Y., Xu, X. & Smith, K. 2020 Burst detection in district metering areas using deep learning method. *J. Water Resour. Plann. Manage.* **146**, 1–12. doi:10.1061/(ASCE)WR.1943-5452.0001223.
- Wu, Y. & Liu, S. 2017 A review of data-Driven approaches for burst detection in water distribution systems. *Urban Water J.* **14**, 972–983. Taylor & Francis. doi:10.1080/1573062X.2017.1279191.
- Wu, Y., Liu, S. & Asce, A. M. 2020 Burst detection by analyzing shape similarity of time series subsequences in district metering areas. **146**, 1–12. doi:10.1061/(ASCE)WR.1943-5452.0001141.
- Ye, G. & Fenner, R. A. 2011 Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *J. Pipeline Syst. Eng. Pract.* **2**, 14–22. doi:10.1061/(ASCE)PS.1949-1204.0000070.
- Ye, G. & Fenner, R. A. 2014 Weighted least squares with expectation-Maximization algorithm for burst detection in U.K. water distribution systems. *J. Water Resour. Plann. Manage.* **140**, 417–424. doi:10.1061/(asce)wr.1943-5452.0000344.

First received 7 June 2022; accepted in revised form 21 December 2022. Available online 30 December 2022