Gradual Leak Detection in Water Distribution Networks Based on Multistep Forecasting Strategy

Xi Wan[1], Raziyeh Farmani[2], Edward Keedwell[3]

[1]Ph.D. Student, Centre for Water Systems, College of Engineering, Mathematics and Physical Science, University of Exeter, Harrison Building, North Park Rd., Exeter, Devon, EX4 4QF, United Kingdom (corresponding author). Email: xw355@exeter.ac.uk

[2]Centre for Water Systems, College of Engineering, Mathematics and Physical Science, University of Exeter, Harrison Building, North Park Rd., Exeter, Devon, EX4 4QF, United Kingdom. Email: R.Farmani@exeter.ac.uk

[3]College of Engineering, Mathematics and Physical Science, University of Exeter, Exeter EX4 4QF, United Kingdom. Email: E.C.Keedwell@exeter.ac.uk

ABSTRACT

With the availability of real-time monitoring data, leakage detection for water distribution networks (WDNs) based on data-driven methods has received increasing attention in recent years. Accurate forecasts based on historical data could provide valuable information about the condition of the WDN, and abnormal events could be detected if the observed behaviour is substantially different from the typical behaviour. Therefore, an accurate forecast model is essential for prediction-based leakage detection methods. While most data-driven methods focus on burst detection, it is also important to develop an early warning system for gradual leakage events as they will cause more water loss due to a longer time to awareness. Therefore, a real-time early leakage detection technique based on a multistep forecasting strategy is proposed in this study. A multistep flow forecasting model is introduced to capture the diurnal, weekly and seasonal patterns in the historical data. The generated multistep forecasting is further compared with the observed measurements, and residuals are calculated based on cosine distance. Based on the analysis of the residual vector, the gradual leakage event could be

detected in a timely manner. The proposed method is applied to the L-town datasets containing one year of real-life flow monitoring data. The results prove the superiority of the proposed multistep prediction model-based method over the traditional one-step prediction model for gradual leakage detection. In addition, the results show that the proposed methodology can detect small gradual leakage events within just a few days while generating no false alarms. The method is further applied to a real-life network and showed consistent results.

INTRODUCTION

Pipeline leakage is likely to have several negative influences on water distribution networks (WDNs): i) enormous economic loss (Wan et al. 2022a); ii) wasted energy consumption from pumping (Puust et al. 2010) ; iii) water quality contamination by introducing infection during hydraulic transients (Colombo et al. 2009). Therefore the problem of detecting and resolving leaks is of great concern to water companies and authorities. The international water association (IWA) has set four leakage management strategies to reduce real losses from WDNs (Farley 2003), namely: 1) active leakage control (ALC); 2) pressure management; 3) speed and quality of repairs; 4) Targeted renewal of infrastructure. Active leakage control plays a vital role in proactive leakage management, as it involves regular surveys and continuous monitoring of the distribution system partitioned into district metered areas (DMAs). One of the benefits of DMAs is that flows and pressures can be monitored in DMAs, and unreported bursts and leaks could be detected. The quicker the water utility can analyse the flow and pressure data of a DMA, the faster bursts or leaks can be located, and the volume of water lost can be reduced (Charalambous et al. 2014).

Leakage detection for WDN has become a growing field of interest that plays a vital role in water asset management. Various methods have been explored and proposed for leakage detection in WDNs. Zaman et al. (2020) provide a comprehensive review of leakage detection strategies for pressurised pipeline systems. These methods can be categorised into hardware-

based methods and software-based methods. Since most hardware-based methods require technicians to interpret the results and have higher installation costs than software-based methods, the hardware-based methods are mainly used for regular inspections rather than continuous monitoring (Romano et al. 2012). However, software-based methods can overcome the aforementioned inability to realise the long-term real-time monitoring of a large-scale water distribution network (Wan et al., 2022a).

In recent years, data-driven methods have received considerable attention due to the development of wireless sensor networks (WSN) (Moridi et al. 2018). The application of WSN for water distribution networks allows continuous measuring and collecting of hydraulic measurements, and the representation of the conditions in the distribution system can be continuously updated. The update frequency is depended on the adopted method. For the hydraulic model-based leak detection method, the update frequency is usually depending on the organization's need and they will adjust the model as necessary, since hydraulic model calibration is one of the most difficult and most important parts of modelling (Zimoch and Bartkiewicz 2018). In the case of the availability of an extensive amount of monitoring data from the real-life network, data-driven methods have provided an alternative solution for leak detection in a cost-effective way without the need for a well-calibrated hydraulic model (Fu et al. 2022). Based on the analysis of pressure, flow, consumer demand or acoustic data collected from supervisory control and data acquisition (SCADA) systems, data-driven methods can extract useful information from the ample amount of monitoring data that are too numerous and complicated for humans to handle.

A popular approach to detect leaks based on pressure or flow monitoring data is based on prediction or regression models (Wan et al., 2022a). By mining the historical data collected under a no-leak situation, the prediction models can learn the historical data pattern or data variation. The well-trained model will be used to predict future pressure or flow values, and

the prediction value will be regarded as a reference and compared with the observed measurements. If the prediction value is substantially different from observed measurements, there is likely an abnormal situation in the system, and inspections may need to be conducted. Ye and Fenner (2011) estimated the normal flow and pressure data based on the Kalman filter, and detected burst events based on the difference between the estimated flow and the measured flow. Mounce and Machell (2006) proposed a leakage detection system based on artificial neural networks and Mounce et al. (2011) used support vector regression for novelty detection on flow and pressure data for a WDN. Wang et al. (2020) applied the long short-term memory (LSTM) model to make more accurate flow data predictions and achieve higher burst detection probability.

All previous studies are based on the one-step prediction strategy, which means that they only focus on making the inference for the system condition of the next time slot. If a burst event happens in the WDN, a significant difference between the measurement value and the predicted value will be observed, and the residual of the prediction model will exceed certain threshold values. The one-step prediction strategy is effective for burst detection. However, this strategy is insufficient for gradual leakage detection, where the events gradually develop from small seeps or weeps to noticeable leak events, unlike burst events that are characterised by sudden changes in a short period. As shown in Fig. 1, the gradual leakage events could last for weeks or even months, and the slowly changing pattern makes it more challenging before detection. The traditional one-step forecasting strategy might not work for gradual leakage detection because it cannot adapt to the slowly changing pattern of gradual leakage and will not generate a large residual, which is essential for leakage detection.

There are limited studies that consider gradual leakage events during leak detection, and aforementioned studies are all focused on burst detection. BattleDIM (Vrachimis et al. 2020) has provided a good assessment of leakage detection methods for gradual leakage events, burst

events, background leakages and simultaneous leakages. Among the methods proposed for BattleDIM, Most of them detected leakages based on the results of hydraulic model calibration. Steffelbauer et al. (2022) proposed a so-called dual approach that use additional virtual valves and reservoirs to translate pressure head drops to leakage outflows. Ma et al. (2022) and Huang et al. (2022) detected leakages based on the analysis of the difference between measurements and calibrated values. The accuracy of the calibrated hydraulic model plays an important role in these model-based methods. However, well-calibrated hydraulic models are usually not widespread in water companies (Wan et al., 2022a). Other candidates have proposed data-driven leakage detection methods. Marzola et al. (2022) detected leakages by visually analysing the monitoring data. Li et al. (2022) identified burst and gradual leakage events based on Seasonal and Trend decomposition using Loess (STL decomposition) and the k-means clustering method. The flow and pressure data were firstly decomposed into three components by STL, and then the gradual leakage events were identified by clustering the trend components based on k-means clustering. However, the STL decomposition is an offline method and batch algorithm, which works on the entire dataset and do not have online mechanism and cannot work on sequential data (Mishra et al. 2022), which means that post event information has been used to identify the leak. Wan et al. (2022b) proposed a method for gradual leakage detection based on statistical analysis. The flow data was transformed into normalised scores and then analysed by exponential weighted moving average (EWMA) to detect gradual leakages.

To address the aforementioned issue, this paper proposed a real-time gradual leakage detection method for WDSs based on a multistep forecasting strategy. A window of data has been predicted at once to include more information so that the trend variation caused by gradual leakage events could be captured. Multistep forecasting is a challenging extension of traditional one-step forecasting. When the prediction horizon is extended from a single point to a time window, various problems will be introduced to the forecasting, such as error accumulation,

126 reduced accuracy, and increased uncertainty (Ben Taieb et al. 2012). Artificial neural networks
127 (ANN), which provide direct support for multistep forecasting, are adopted in this paper for
128 predicting multiple steps forward.

129 In the proposed methodology, the observed flow data is predicted in the long-term using the
130 ANN model first. Next, the comparison between the observed vector and the predicted vector
131 is calculated based on cosine distance. Finally, the EWMA is used to smooth residual vectors
132 to reduce the influence of noises and uncertainties. A threshold is set for the smoothed residual
133 vectors and raises alarms during leakage events. The proposed method is applied to a case study
134 called L-Town which contained one year of real-life flow monitoring data in an online manner.
135 The results demonstrate the applicability of the proposed method for gradual leakage detection
136 and proves its effectiveness by detecting small gradual leakage events within just a few days
137 while generating no false alarms. The proposed method doesn't require information such as
138 smart meter readings and hydraulic models that are usually not widespread in water companies.

139 ## METHODOLOGY

140 The traditional one-step prediction models accurately detect burst events characterised by
141 abrupt changes in monitoring datasets (decreasing in pressure and increasing in flow) but miss
142 the scenario that some leakage events gradually develop over time. This paper proposes a novel
143 methodology for detecting gradual leakage events using an ANN-based multistep forecasting
144 approach to consider a window of data simultaneously. A residual value is generated at each
145 timestamp to represent the distance between the predicted time window and the observed time
146 window. Based on the analyses of the generated residual vector, gradual leakage events could
147 be detected. Fig. 2 shows the flowchart of the proposed gradual leakage detection method.
148 There are two major steps in the proposed algorithm, namely the prediction stage and the
149 classification stage, and the steps of the proposed problem-solution algorithm are listed as
150 follows:

Step 1. Collect real-time monitoring data from sites;

Step 2. Specify the topology of the ANN model based on grid search, and train the multistep forecasting model with months of leak-free historical data;

Step 3. Compute the difference between the output vector and the observed measurements based on cosine distance;

Step 4. Compute the EWMA score of the residual vector;

Step 5. Compute the threshold value in leakage-free historical data;

Step 6. Detect if the system contains any leakage events based on the comparison of the threshold value and the EWMA score.

Multistep Forecasting Strategy

A multistep ahead time series forecasting task means predicting the next H value $[x_{N+1}, x_{N+2}, \dots, x_{N+H}]$ based on a historical time series $[x_1, x_2, \dots, x_N]$ composed of N observations (Ben Taieb et al. 2012). This is a challenging task because of the potential for error accumulation. There are several ways of carrying out the multistep prediction, and three well-known approaches (Andrawis et al. 2011) are: the recursive approach, the direct approach, and the multi-input multi-output (MIMO) approach. The mechanism of the three approaches for multistep prediction is shown in Fig. 3. The grey line in Fig. 3 represents the historical data used for predicting future data, and the blue lines and dots represent the predicted data.

The recursive approach makes predictions based on one-step-ahead prediction. The predicted values will be used as input variables for the next prediction. To forecast h steps ahead, it takes the preceding output as input to forecast the next value and continues in this manner. Suppose I is the length of the input, and Q is the length of the output. The input sequence is represented as $\mathbf{X} = \{x_1, x_2, \dots, x_I\}$, the predicted output sequence is represented as $\widehat{\mathbf{X}} = \{\hat{x}_{I+1}, \hat{x}_{I+2}, \dots, \hat{x}_{I+O}\}$, and $f(x)$ represents the forecasting model. Eq. (1) shows the process of the recursive strategy.

$$\hat{x}_{I+h} = \begin{cases} f(x_1, x_2, \dots, x_I) & h = 1 \\ f(x_h, x_{h+1}, \dots, x_I, \hat{x}_{I+1}, \dots, \hat{x}_{I+h-1}) & 1 < h \leq I \\ f(\hat{x}_h, \hat{x}_{h+1}, \dots, \hat{x}_{I+h-1}) & I < h \leq Q \end{cases} \quad (1)$$

Due to its iterative nature, the output error will accumulate as the prediction moves forward (Li et al. 2019).

The direct strategy uses the same features to predict different target variables. Multiple models are trained with different targets, one for each forecast step. Eq. (2) shows the process of the direct strategy.

$$\hat{x}_{I+h} = f_h(x_1, x_2, \dots, x_I) \quad (2)$$

The direct strategy is also based on a one-step forecasting model, but uses different models to forecast each horizon independently, where each model is trained separately (Sahoo et al. 2020). The number of models is equivalent to the length of the output vector, which means that each added predictive time step is an extra computational demand and maintenance burden, especially when the number of output variables is not trivial. Moreover, this approach considers each prediction horizon independently. Therefore, the dependencies and correlations between the predictions are ignored and cannot be considered during prediction.

In contrast, the MIMO approach learns the mapping function between multiple inputs and multiple outputs directly. Eq. (3) shows the process of the direct strategy.

$$\{\hat{x}_{I+1}, \hat{x}_{I+2}, \dots, \hat{x}_{I+O}\} = f(x_1, x_2, \dots, x_I) \quad (3)$$

The MIMO model's performance depends on the model's capacity. MIMO doesn't suffer from error accumulation. In addition, MIMO outputs all variables at once, which makes it more computationally efficient than direct strategy. The comparative studies (Ben Taieb et al., 2012; Li et al., 2019) proved that the MIMO approach gives the best performance among the three approaches mentioned before. Therefore, the MIMO approach is adopted in this study.

Artificial Neural Network

Classical time series forecasting methods, such as autoregressive moving average (ARMA) (Kadri et al. 2016), autoregressive integrated moving average (ARIMA) (Yaacob et al. 2010), generalized autoregressive conditional heteroscedasticity (GARCH) (Maharaj et al. 2019) have several limitations that limit their applications. For example, classical methods such as ARIMA assume a linear relationship and do not consider more complex joint distributions (Jason 2018). Most importantly, they are focused on one-step prediction and don't support multiple outputs directly, which makes them not suitable for MIMO long-term forecasting. However, ANN could provide direct support for multivariate forecasting by simply increasing the number of neurons in the output layer. Moreover, based on the comparison studies conducted by many researchers (De Nadai and Van Someren 2015; Siami-Namini et al. 2019), ANN models showed better performance than classical statistical methods. Moreover, the inference time of machine learning methods is much faster than some commonly used statistical methods such as ARIMA (Braei and Wagner 2020). Inference time measures the time a model needs to generate outputs, which is important for real-time monitoring. Therefore, ANN has been widely applied in financial marketing (Tealab et al. 2017), energy management (Ahmad et al. 2014), etc. A typical and powerful neural network, multilayer perceptron (MLP), is used in this study to approximate the mapping function to link input variables with output variables.

An ANN can be regarded as a network of "neurons", which are layered and connected. As shown in Fig. 4(a), the simplest network contains no hidden layers and can process only linearly separable functions. The nodes, namely neurons, are linked through connections. The weight of each connection is adjusted iteratively using a "learning algorithm" that minimises a "cost function" such as mean squared error (MSE). Once we add hidden layers between the input layer and the output layer, the neural network becomes nonlinear, as shown in Fig. 4(b). The numbers of hidden layers, connections, and neurons of an ANN depend on the data's complexity (Grekousis 2019). The more complex the data, the more likely the ANN will need

additional hidden layers and neurons. Due to its flexibility, the ANN can be configured to support a defined number of inputs and outputs in the mapping function, as shown in Fig. 4(c). The mathematical relationship between inputs and outputs can be expressed as:

$$\boldsymbol{Y}_j = f\left(\theta_j + \sum_{i=1}^{n} w_{ji}\boldsymbol{X}_i\right) \quad (4)$$

where $\boldsymbol{X}_i$ is the *i*th input variable; $\boldsymbol{Y}_j$ is the *j*th output variable; θ_j is the bias in the hidden layer; n is the number of neurons in the hidden layer; w_{ji} is the connection weight; f is the transfer function.

The MLP used in this paper has been designed to make multistep predictions for flow monitoring data. The topology of the MLP was defined based on the grid search strategy. The algorithm searched in the grid of parameters and across all points to get the optimal combination. The optimal combination of hyperparameters was selected based on the performance metrics using cross-validation (Yasin et al. 2016).

Residual Comparison

The traditional residual vector generated by a one-step prediction model can be expressed as:

$$r_t = x_t - \hat{x}_t \quad (5)$$

where x_t is the observed value at time t; $\hat{x}_t$ is the predicted value at time t; r_t is the residual at time t.

It describes the difference between two single values. The detection of leakage events using the residual calculated in Eq. (5) is based on the assumption that the difference between the predicted value and observed value during the leakage scenario will be substantially larger than those in the normal situation. Therefore, it is useful when one needs to identify the sudden, instantaneous behaviour in the system, which refers to burst events. However, it could be challenging for gradual leakage detection, because there is no memory contained in the residual.

Furthermore, gradual leakage events don't generate sudden outflow in WDNs, which makes it more difficult to be detected based on Eq. (5).

Therefore, in this paper, a memory-contained residual is generated based on the multistep prediction model. The residual represents the difference between the predicted time window and the observed time window. There are three commonly-used approaches to describe the distance of two time series: Manhattan distance, Euclidean distance and cosine distance.

$$d_{manhattan}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) = \left\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\right\|_1 = \sum_{i=1}^{l} |x_i - \hat{x}_i| \tag{6}$$

$$d_{euclidean}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) = \left\|\boldsymbol{X} - \widehat{\boldsymbol{X}}\right\|_2 = \sqrt{\sum_{i=1}^{l} (x_i - \hat{x}_i)^2} \tag{7}$$

$$d_{cosine}(\boldsymbol{X}, \widehat{\boldsymbol{X}}) = 1 - \frac{\boldsymbol{X} \cdot \widehat{\boldsymbol{X}}}{\|\boldsymbol{X}\| \|\widehat{\boldsymbol{X}}\|} = 1 - \frac{\sum_{i=1}^{l} x_i \hat{x}_i}{\sqrt{\sum_{i=1}^{l} x_i^2} \sqrt{\sum_{i=1}^{l} \hat{x}_i^2}} \tag{8}$$

where $\boldsymbol{X}$ and $\widehat{\boldsymbol{X}}$ represent the observed data vector and the predicted data vector, respectively; l represents the length of the output vector.

Manhattan distance represents the sum of the absolute difference in a time window. Euclidean distance measures the distance between two vectors. Cosine distance is a metric to evaluate the angular distance between two vectors and is the approach used here. The superiority of cosine distance will be proved in the results section.

EWMA-Based Event Detection

At each time step, a residual r_t calculated based on Eq. (8) is obtained and appended to a one-dimensional vector of residuals:

$$\mathbf{R} = [r_{t-l}, r_{t-l+1}, \dots, r_t] \tag{9}$$

where l is the number of historical data that are used to set up the threshold and evaluate current residuals.

In order to further reduce the probability of false alarms, the vector of residuals is then smoothed to dampen spikes in residuals that frequently occur due to noises in the data

(Hundman et al. 2018). Exponentially-weighted moving average (EWMA) is used to calculate the smoothed residuals. EWMA exponentially weighted the average of prior data as:

$$r_i^s = \lambda r_i + (1-\lambda) r_{i-1}^s \tag{10}$$

where r_i^s is the smoothed value of r_i, and $r_0^s = r_0$; λ is a weighting factor and $0 < \lambda < 1$. Therefore, a smoothed residual vector $\boldsymbol{R}^s = [r_{t-h}^s, r_{t-h+1}^s, \dots, r_t^s]$ is obtained.

To determine whether there is a leakage event happened in the distribution system, a threshold is set to classify the normal situation and the abnormal situation. The threshold ϵ is determined based on the analysis of historical residual vectors, and calculated as:

$$\epsilon = \mu(r^s) + z\sigma(r^s) \tag{11}$$

where $\mu(\cdot)$ is the mean value of a time series; $\sigma(\cdot)$ is the standard deviation of a time series. Values for z depend on the context, the rule of thumb of z is 2, 2.56, 3, 3.5 etc. (Wang et al. 2020). The most common rule used for z is 3, which is equivalent to a false alarm rate of 0.27% (Harrou et al. 2020), and this value is used in this paper to set the threshold in an unsupervised way. If users could have information about existing leakage event, they could tune the threshold based on the historical event.

Performance Evaluation

Different metrics are used to evaluate the accuracy and effectiveness of the two stages (the prediction stage and the classification stage) of the proposed gradual leakage event detection method.

For the prediction stage, mean absolute percentage error (MAPE), root mean squared error (RMSE), and the coefficient of determination (R^2) are used to measure the prediction accuracy of the prediction model. MAPE calculates the mean of the ratio of absolute differences between the actual values and the predictions. RMSE represents a quadratic score of the average magnitude of the absolute differences. R^2 explains the level of variability of the predicted value can be caused by the actual value. The calculation of MAPE, RMSE and R^2 are represented as

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_t)^2} \quad (13)$$

$$R^2 = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (y)^2]}} \quad (14)$$

For the classification stage, true positive rate (TPR), false positive rate (FPR), detection accuracy (DA), and detection time (DT) are three metrics used to evaluate the accuracy of event detection. The definition of TPR and FPR are:

$$\text{TPR} = \frac{TP}{TP+FN} \quad (15)$$

$$\text{FPR} = \frac{FP}{FP+TN} \quad (16)$$

where TP is true positive, which means the number of data points during a leakage event that have been flagged as abnormal correctly; TN is true negative, which means the number of data points during a leakage event that did not trigger the alarm; FP is false positive, which means the number of data points during normal situations that triggered the alarm wrongly; FN is false negative, which means the number of data points during normal situations that did not raise the alarm correctly. TPR reflects how much of the data during the leakage event has been identified correctly, and FPR reflects the rate of false alarms during the detection process.

DA represents how many leakage events among all leakage events have been detected, calculated as:

$$\text{DA} = \frac{the\ number\ of\ detected\ leakage\ events}{the\ number\ of\ totle\ leakage\ events} \quad (17)$$

When an alarm is raised during a leakage event, the leakage event could be regarded as correctly identified. In order to further evaluate the accuracy of the detection methodology, the DT is used to evaluate the effectiveness of a leakage detection method, especially for gradual leakage detection. As gradual leakage events will always become apparent, provided enough time has passed for the rate of leakage to build up. The detection time is defined as the elapsed

time from the start of a leakage event to the time when the event is first detected. Detection time reflects how quickly the leakage detection algorithm responds, and the volume of water saved.

SYNTHETIC CASE STUDY

Description of Study Area

The case study L-Town water distribution network was provided by the organizers of BattLeDIM (Vrachimis et al. 2020). The graphical representation of L-Town is represented in Fig. 4. This hydraulic model is based on a real WDN located in Cyprus with around 10,000 consumers. This L-Town WDN consists of 905 pipes with a total length of 42.6 km, and 782 junctions. Two reservoirs are responsible for providing clean water with a pressure head of at least 20m to all water consumers. As shown in Fig. 4, L-Town can be divided into three distinct hydraulic areas: (1) Area A, which is the main area and receives water from two reservoirs; (2) Area B, located downstream of Area A, a pressure reduction valve (PRV) has been installed to reduce the water pressure in Area B; (3) Area C, located in an area with higher elevation, and receiving water from a water tank located between Area A and Area C.

There are three types of consumers in L-Town, which are residential, commercial and industrial. All consumers' behaviour patterns are well-modelled in the hydraulic model with a sampling rate of 5 minutes, allowing researchers to have a more reliable and realistic tool for generating monitoring data and testing the ability of leakage detection methods. In this paper, only flow data at the exits of water resources are recorded. As represented by the red triangle in Fig. 5, three flow sensors are assumed to be equipped in the WDN. Two flow sensors are located at the exits of two reservoirs, and one flow sensor is located at the exit of the water tank.

Dataset Generation

EPANET simulation

There are two types of hydraulic analysis methods: demand-driven analysis (DDA) and pressure-driven analysis (PDA). DDA assumes that nodal demands are known and satisfied regardless of the available pressure at nodes. In comparison, PDA prioritises the pressure requirements at nodes than demand requirements. It is known that PDA could provide more realistic representation of the pressure-leakage relationship (Baek et al. 2010). Thus, this paper adopts PDA as the approach to generating synthetic flow data.

The original BattleDIM dataset contains lots of simultaneous leakages. For data-driven leakage detection methods, the situation of simultaneous leakages is still a difficulty. Most of the time, the existence of multiple leakages is still based on expert annotation (Daniel et al. 2022; Marzola et al. 2022). Therefore, the original dataset has been changed to suit the single leakage assumption of the proposed method. The "real" hydraulic model published by Vrachimis et al. (2022) is used in this paper to generate a monitoring dataset to maintain the uncertainties caused by time-varying consumer demand patterns.

Simulation of gradual leakages

Water network tool for resilience (WNTR) (Klise et al. 2017) is used in this paper to simulate gradual leakage events in the L-Town. 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 (Zhou et al. 2019) as:

$$q_{leak,i} = C_i A_i \sqrt{2gh_i} \tag{18}$$

where $q_{leak,i}$ is the leak flow rate; C_i is the discharge coefficient with default value of 0.75 (assuming turbulent flow) (Lambert 2001); h_i is the head; g is the acceleration of gravity; and A_i is the area of the hole.

The leakage area A_i is assumed to increase linearly as the time increases:

$$A_i(t) = \frac{A_{i,max}}{T_i} t \tag{19}$$

where T_i is the growing duration that the ith gradual leakage needed to reach its maximum value $A_{i,max}$.

Leakage events with different growth rates are simulated, and $A_{i,max}/T_i$ determines the growth rate of a leakage event. To model different growth rates, each gradual leakage event has the same growing duration T, which is 30 days, but with different maximum leak flow rates controlled by the maximum leak area in the pipeline. The maximum leak flow ranges from around 5 m^3/h to around 60 m^3/h. It should be noted that the maximum leak flow is used to control the daily leak increasing rate. The larger the daily leak increasing rate, the easier the leak could be detected. Therefore, it is important to evaluate when the algorithm raises the alarm and how early could the leakage be detected.

As shown in Table 1, six different leaking levels are modelled in the dataset, based on the percentage of the increase in the daily leakage rate to the average daily water flow rate. Six leak levels corresponding to six different daily leak increasing rates, ranged from 0.1-1% of daily average demand. From Table 1, it can be seen that even though the leakage flow rate of level 6 has reached over 47 m^3/h after 30 days, the daily increase of leakage flow rate is only 1.57 m^3/h, which only accounts for 1% of the average daily water demand. This minor flow rate increase could be easily regarded as a normal increase in water consumption. As the leakage rates become smaller, detecting those leakages will be more challenging.

As shown in Fig. 6, a total number of 600 leakage events are modelled with different leak levels. Each leaking level contained 100 sets of monitoring data with different leak magnitudes and locations, and the locations of leakages were randomly selected in the network. Each dataset contained a half-year of monitoring data, from 1 January to 1 July, with a sampling rate of 5 minutes. The leakage happened from 15 April to 15 May, and gradually increased from 0 to its designated maximum leakage outflow. Fig. 7 shows the plot of the flow monitoring datasets when the system contained no leakage, a level 1 leakage, and a level 6 leakage. The

grey area shows the time window that contained a leakage event. It could be seen that the gradual leakage event does not show sudden spikes in the data. The dataset with level 1 leakage almost shows no difference from the dataset with no leakage.

Uncertainties of sensor accuracy

It should be noted that the "real" hydraulic model here is only used to generate monitoring dataset. Since the proposed method is a data-driven method, which means that the detection is purely based on data mining, the uncertainties of pipe roughness or pipe diameters do not have major influence the detection results. The uncertainties of pump scheduling, demand uncertainties have been considered during model construction. However, the uncertainties related to the sensor accuracy should be considered. Sensor uncertainties refer to the errors or deviations that exist in the measurements taken by a sensor. It is important to consider sensor uncertainties when using the simulated sensor data. The accuracy of a sensor is usually specified by the manufacturer, and modern flow meters typically publish error of $\pm 2\%$. Therefore, after collecting the monitoring data generated from hydraulic model, a random number which follows a uniform distribution of $\pm 2\%$ is added to each datapoint.

Multistep Prediction VS Single-Step Prediction

In this section, the reason why traditional single-step prediction cannot be used to detect gradual leakage events will be demonstrated. Two ANN models are developed in this section, and the hyperparameters of both models are chosen based on the results of the grid-search method.

Based on the result of the grid search, the MIMO ANN model consists of: (1) input of 7 days of monitoring data, which is 7×288 data points as the sampling interval is 5 minutes, (2) two hidden layers of size 350, 300, (3) output of one day of data, which corresponds to 288 data points. The activation function of the first and second layers and between hidden layers was rectified linear unit (ReLU), and the activation function of the hidden layer to the output

layer was linear. The learning rate was set as 0.01, and the batch size was 256. The parameters of the one-step-prediction ANN model are mostly the same as the MIMO ANN model, except that the one-step-prediction ANN model only contained a single hidden layer with two neurons. It should be noted that one can always use a model with stable performance instead of restricting it to this specific topology. The first three months of flow monitoring data are used to train ANN models, and the last three months of data are used as test data.

Fig. 8 shows the residual vectors generated by the one-step forecasting model and the multistep forecasting model on datasets that contain different levels of leakage. It can be observed that the residual generated by the one-step forecasting model can barely reveal the existence of any leakage events. This is because of the slowly-developing nature of the gradual leakage events. Furthermore, the one-step forecasting model evaluates a single data point at each time, with no memory contained in the detection process. Thus, the one-step forecasting model will regard the slowly-growing trend of leakage as a normal consumer demand increasing and adapt to this trend gradually.

In contrast, multistep forecasting generates large residuals during leakage duration that are essential for detection. The multistep model incorporates the information within a day of a time window, and the residual clearly shows the existence of most of the leakage events. As the growth rate of leakage becomes faster, it becomes clearer to detection algorithm. The comparison clearly shows the superiority of the multistep forecasting strategy for gradual leakage event detection.

Prediction Performance Evaluation

There are three different forecasting approaches mentioned in the methodology, which are recursive, direct, and MIMO. The prediction horizon designated in this paper for leakage detection is a day of flow demand, which means the model will predict 288 data points at once. For the direct approach, it is very computationally demanding to develop 288 ANN models at

444 once, not to mention the considerable amount of time and resources to tune the parameters for
445 all 288 ANN models. Since it is almost impossible to apply this approach in real-life, this
446 approach will not be assessed in this paper.

447 The prediction results of water flow data based on the recursive approach and MIMO
448 approach with different prediction horizons (6 hours, 12 hours, 24 hours and 48 hours) are
449 reported in Table 2. For the recursive approach, the one-step ANN model is used recursively
450 to predict forward until the end of the prediction horizon. For the MIMO approach, four
451 different ANN models were trained to generate results. As shown in Table 2, the MIMO
452 approach is superior to the recursive approach among all prediction horizons.

453 In order to take a closer view of the prediction accuracy of each output, Fig. 9 shows the
454 performance of the MIMO-ANN-24h and the recursive-ANN-24h at each prediction step. It
455 could be observed that the MIMO approach achieves the most stable performance in the whole
456 prediction horizon. In comparison, as the forecast moves forward, the forecast performance of
457 recursive-ANN deteriorates. This is because error accumulation due to noise is unavoidable in
458 the recursive approach.

459 **Influence of Prediction Horizon**

460 One of the key parameters set for the proposed algorithm is the prediction horizon. The
461 length of the prediction window determines how much information should be included to reveal
462 the deviation of the trend component in the data. The prediction horizon used in this paper is
463 one day of flow data. In this section, the performance of the proposed method with different
464 prediction horizons is tested. Table 3 presents the proposed ANN-based method's detection
465 time with different prediction horizons. It could be observed that when the time window is
466 small, such as 6 hours and 12 hours, the information contained in the time window is not
467 enough to reveal the abnormal behaviour caused by small leakages in level 1 and level 2. When
468 the window becomes larger, leakage with a small size could be detected. However, it should

be noted that as the prediction horizon becomes larger, the difficulty of developing such a MIMO model will also increase, making it more difficult to maintain the model. Therefore, a balance should be considered, and in this paper, one day of information is enough for decision-making. As shown in the table, daily forecasting is enough to detect all the gradual leakage events.

Residual Analysis

After the prediction values are obtained from the prediction model, a comparison should be conducted between the prediction values and the true measurements. This comparison statistic is called residual. This section compares the performance of gradual leakage detection based on the three different residual analysis approaches (Manhattan distance, Euclidean distance, and cosine distance).

Fig. 10 shows the plots of the residuals generated based on those three approaches. The residuals generated by Manhattan distance and Euclidean distance show a growing trend. This is because there are general trends in the monitoring data caused by the regular consumer demand increasing from winter to summer. This general trend needs further analysis so that it can be eliminated and prepared for the threshold setting, and unnecessary computational demand should be invested in the threshold setting. In contrast, the residual generated by cosine distance shows a more stable behaviour, and the threshold could be set more easily.

Results of Gradual Leak Detection

The threshold should be set for the residual statistics. If the threshold is set with a large value, the probability of generating false alarms will be reduced, but it will also reduce the ability to detect a possible leakage event, and vice versa. Therefore, a ROC curve plotted by connecting TPR and FPR points is used to evaluate the performance of the proposed methodology.

Fig. 11 shows the ROC curves on datasets with different levels of leakage events. As shown in Fig. 11, for the leakage in level one, the best TPR is 0.556 while FPR is 0.008. For the leakage in level 6, the best TPR reaches 0.848 while FPR is 0.006. The data points used to plot each ROC curve is obtained based on the results calculated based on various threshold value that is calculated based on Eq. (11). A higher threshold value will assign more tolerance of data noise and results in lower FPR but will decrease the TPR, and vice versa. The red star sign represents the results based on the three-sigma rule, which is very close to the optimum TPR-FPR point. The results also prove the eligibility of the unsupervised threshold-setting strategy.

From Fig. 11, it could be observed that for the leakage in level six, the best TPR reach over 80%, and FPR is 0%, which means no false alarms exist. As the growth rate of the leakage increases, the best TPR is closer to 100%, and the optimum TPR-FPR point is closer to the ideal point. Higher TPR means more parts of the leakage events could be detected. As the growth rate of the leakage becomes smaller, it will be more difficult for a detection method to distinguish a leakage event from normal consumer consumption. It should be noted that the beginning stage of leakage could be extremely difficult to be detected because the growth rate and amplitude of the water loss are too small, and this part of undetected data points will reduce the TPR value. Therefore, an additional metric, namely detection time, should be used to evaluate how early it could be to detect a gradual leakage.

Table 4 shows the results of the detection time regarding each level of leakage. Compared with the statistical method proposed by Wan et al. (2022b), the ANN-based method could successfully detect leakage events with lower growth rates, which shows the superiority of the proposed method. In addition, the ANN-based method shows a higher detection probability than the statistical method. For the small leakages in levels 1-4, the detection time of the ANN-based method is around 1-2 weeks, but the statistical method failed to detect these leakages with such small growth rates. For the relatively large leakages in level 5-6, the statistical

518 method could raise alarm quicker than the proposed method. This is might because in the early
519 stage of a leakage, the ANN model is still regarding the trend in the data as normal growing
520 trend and did not generate large residuals. Therefore, it is recommended to combine the
521 statistical method and machine learning method in the future to have better performance. In
522 general, the proposed method shows more robustness than the statistical method.

523 The amount of water loss by the time the leakage has been detected is also presented in
524 Table 4. It could be observed that the proposed algorithm could save much more water
525 resources than the statistical method, which is the first goal of a leakage detection algorithm.
526 Even though the maximum leak flow rate reaches 60 m^3/h, all the leakage events have been
527 detected when the leak flow rate is less than 7 m^3/h, which falls into the category of
528 background leakage determined in (Vrachimis et al. 2022). The results proved the early
529 detection capability of the proposed method.

530 **Results of Monitoring Data with Different Sampling Rates**

531 In reality, the data might be collected with different sampling intervals, such as 15 minutes,
532 30 minutes or more. Flow data are usually reported as average values (Wan et al., 2022a), and
533 it could be smoother with a high sampling interval, which might introduce a benefit for leakage
534 detection, but, in doing so, might lose informative components. Mounce et al. (2012) conducted
535 an analysis of the influence of sampling rate on event detection in WDN. The results showed
536 that a sampling interval of 5 minutes does not greatly improve the detection results for burst
537 detection. Similarly, this section investigates the influence of the sampling interval on the
538 gradual leakage detection based on the proposed methodology.

539 Table 5 shows the detection results of datasets with different sampling intervals (i.e. 15
540 minutes, 30 minutes, and 60 minutes) based on the proposed method. The results showed that
541 the proposed method benefits from higher sampling frequencies. The best detection results are

obtained for the dataset with 5 minutes interval, and the performance are deteriorating as fewer data points are obtained from sites.

REAL CASE STUDY

In order to further demonstrate the application of the proposed method to a real-life water distribution system, monitoring data collected from a DMA in a real-life water distribution network in the UK is presented in this section. The DMA has a total of 1939 properties, which contain 1889 household properties and 50 non-household properties. The average night pressure is 37.74 m. The pressure data were collected from the SCADA system over a period of two months, from October 3 to December 04, 2021. The pressure data were measured every 15 minutes. The water company has provided information about some events recorded during this period. Fig. 12 shows the profile of the pressure recording during the whole period. Despite some burst events, it could be observed that the pressure gradually decreased starting around November 22 or 23, and becomes obvious in the data profile from manual observation on November 27. The event is reported on November 30.

In order to reduce the influence of burst events, based on the customer report, the events on November 11 and November 17 have been deleted from the dataset. The first month is used as the training dataset and the second month which contains a gradual leakage event is used as the testing dataset. Two ANN models have been developed for this dataset. One is the traditional single-step prediction model and another is the proposed MIMO model. Two models have used the same ANN hyperparameters as mentioned in the synthetic dataset except for the number of neurons in the hidden layers has been reduced to 300 and 200 since a higher sampling interval is used, i.e. 15 minutes in the real dataset.

Fig. 13 shows the residual vectors generated by the single-step model and the MIMO model that are smoothed by EWMA. The blue and yellow vertical lines are the time when the alarm was raised by the single-step model and the MIMO model separately. The results show that the

567 single-step model detects the event when it reaches a relatively large size at 16:45 on November
568 27, and the proposed MIMO model captured the growing trend in the dataset and detects the
569 leakage event at 15:00 on November 26, before the leakage becomes a burst event. Moreover,
570 the residual generated by the single-step model shows more volatility and uncertainties, thus
571 requiring more careful threshold tuning to reduce the number of false alarms.

572 ## DISCUSSION

573 Most of the previous data-driven leakage detection methods were focused on burst events
574 and did not evaluate the performance of their methods on gradual leakage detection. Among
575 the limited research that considers the existence of gradual leakage events, most methods
576 developed for the BattleDIM problem are tailored for the problem only since they used various
577 information such as AMR, pressure values, well-calibrated hydraulic model, and flow values.
578 Therefore, this paper proposed a method that is more practical for a water company that doesn't
579 have much information and is able to detect gradual leakage events in an online manner based
580 on the flow data only.

581 There are three main assumptions behind this methodology. The first assumption is that the
582 prediction model is trained with flow monitoring data under fault-free conditions. This is
583 because detecting leakage events relies on the difference between the prediction and the
584 observed measurements. If there are leakage events exist in the historical data, those historical
585 events should be eliminated based on historical customer reports or data preprocessing
586 techniques such as statistical process control (Romano et al. 2014) so that the model could learn
587 the normal behaviour more accurately. Same to other prediction-classification leakage
588 detection methods (Romano et al. 2014; Wang et al. 2020; Ye and Fenner 2011), the main
589 limitation of the proposed method is that it cannot be directly applied to the situation when the
590 distribution system experiencing local changes (e.g., isolation of pipe segments, expansion of
591 network) as changing operation condition will introduce different pattern in the flow data that

592 hasn’t been learnt by the ANN model. Under such conditions, the ANN model should be
593 retrained or updated to learn the new pattern of the monitoring data.
594 The second assumption is that the model needs to be reset after the leakage events detection.
595 The reset time depends on the length of the input vectors. For example, the input vector is a
596 week of data, and thus the reset time is a week. This is because just after the leakage events,
597 the input vector cannot represent the normal situation of the distribution system. Thus, the
598 prediction model will continue to generate abnormal output until there are enough input data.
599 The third assumption is that the system operation condition doesn't change during detection.
600 If the operation condition changed after the model developed from historical data, a separate
601 model needs to be developed for the changed situation. For example, if there is a pressure
602 reduction valve installed in the system, the system behaviour will follow another pattern, and
603 the old model cannot be applied after the installation.

604 ## CONCLUSION

605 In this work, a method is proposed to detect gradual leakage events. The proposed method
606 contains two main stages: prediction and detection. In the prediction stage, a MIMO-ANN is
607 trained on the historical dataset to model the system behaviour under healthy conditions, and
608 then the residual vector is obtained by comparing the output vector with the observed
609 measurements based on cosine distance. In the classification stage, the EWMA is used to
610 smooth the residuals and eliminate spikes caused by noises. Then, leakage could be detected
611 by comparing the residual values against a user-designated threshold. The proposed method
612 has been tested on the L-Town distribution network.
613 The reason why a multi-step forecasting approach is needed for gradual leakage detection
614 has been illustrated. This is because the residual vector obtained from the one-step prediction
615 model doesn’t contain any memory of the historical data. However, by extending the prediction
616 horizon, the multi-step model presented in this paper successfully detected all the leakage

events. The superiority of the MIMO approach against the recursive prediction strategy has been proved. The MIMO reaches the most stable performance and the highest accuracy. Furthermore, the machine learning technique supports multiple outputs in its nature, and provides a more convenient technique to design a MIMO model.

Compared with Manhattan distance and Euclidean distance, the cosine distance used in this paper can generate more stationary residuals, and is thus more suitable for the detection stage. Compared with the statistical method proposed by Wan et al. (2022b), the detection results showed that the proposed method could detect gradual leakage events with a lower level of growth rate with higher detection probability and, thus save more water loss caused by leakage.

The results showed that the proposed method has the capability of detecting gradual leakage events with a small growth rate. All the leakage events could be detected before it reaches 7 m3/h, which is categorized as background leakage in the BattleDIM. For leakage with a large increasing rate, the statistical method shows quicker detection time. However, the proposed method has higher detection accuracy and is more robust than the statistical method. It is recommended to combine the statistical method and machine learning method in the future.

The proposed method has been demonstrated on a pressure data from a real-life DMA. The results showed that the proposed method showed superior performance than traditional one-step-prediction model.. In practical application, it can be applied for real-time monitoring of pressure or flow data. The training dataset should be pre-processed based on statistical process control rules or historical customer report to satisfy the assumption of the proposed method. For the real-time streaming data, the machine learning model could be updated when new data comes so that the model could be adapted to the non-stationary characteristics of the online data.

DATA AVAILABILITY STATEMENTS

The hydraulic model used in this study is available at https://battledim.ucy.ac.cy/. The following data and the model used in this study can be made available by the corresponding author on request: data of synthetic experiments, and codes for the proposed method in Python language.

ACKNOWLEDGEMENT

The first author is funded by the China Scholarship Council (No. 202006370080), and the work is supported by a Royal Academy of Engineering Industrial Fellowship to resource Raziyeh Farmani's involvement (IF\192057).

REFERENCES

Ahmad, A. S., M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin, H. Abdullah, and R. Saidur. 2014. "A review on applications of ANN and SVM for building electrical energy consumption forecasting." *Renew. Sustain. Energy Rev.*, 33: 102--109.

Andrawis, R. R., A. F. Atiya, and H. El-Shishiny. 2011. "Forecast combinations of computational intelligence and linear models for the NN5 time series forecasting competition." *Int. J. Forecast.*, 27 (3): 672–688. Elsevier B.V. https://doi.org/10.1016/j.ijforecast.2010.09.005.

Baek, C. W., H. D. Jun, and J. H. Kim. 2010. "Development of a PDA model for water distribution systems using harmony search algorithm." *KSCE J. Civ. Eng.*, 14 (4): 613–625. https://doi.org/10.1007/s12205-010-0613-7.

Braei, M., and S. Wagner. 2020. "Anomaly detection in univariate time-series: A survey on the state-of-the-art." *arXiv*.

Charalambous, B., D. Foufeas, and N. Petroulias. 2014. "Leak Detection and Water Loss Management." *Water Util. J.*, (8): 25–30.

Colombo, A. F., P. Lee, and B. W. Karney. 2009. "A Selective Literature Review of Transient-Based Leak Detection Methods." *J. Hydro-Environment Res.*, 2 (4): 212–227. Elsevier B.V. https://doi.org/10.1016/j.jher.2009.02.003.

Daniel, I., J. Pesantez, S. Letzgus, M. A. Khaksar Fasaee, F. Alghamdi, E. Berglund, G. Mahinthakumar, and A. Cominola. 2022. *A Sequential Pressure-Based Algorithm for Data-Driven Leakage Identification and Model-Based Localization in Water Distribution Networks*. *J. Water Resour. Plan. Manag.*

Farley, M. 2003. "Non Revenue Water - International Best Practice for Assessment , Monitoring and Control." *12th Annu. CWWA Water, Wastewater Solid Waste Conf.*, 1–18.

675 Fu, G., Y. Jin, S. Sun, Z. Yuan, and D. Butler. 2022. “The role of deep learning in urban water
676 management: a critical review.” *Water Res.*, 223: 118973.
677 Grekousis, G. 2019. “Artificial Neural Networks and Deep Learning in Urban Geography: A
678 Systematic Review and Meta-Analysis.” *Comput. Environ. Urban Syst.*, 74: 244–256.
679 https://doi.org/10.1016/j.compenvurbsys.2018.10.008.
680 Harrou, F., Y. Sun, A. S. Hering, M. Madakyaru, and A. Dairi. 2020. *Statistical Process*
681 *Monitoring using Advanced Data-Driven and Deep Learning Approaches*.
682 Huang, L., K. Du, M. Guan, W. Huang, Z. Song, and Q. Wang. 2022. “Combined Usage of
683 Hydraulic Model Calibration Residuals and Improved Vector Angle Method for Burst
684 Detection and Localization in Water Distribution Systems.” *J. Water Resour. Plan.*
685 *Manag.*, 148 (7): 1–14. https://doi.org/10.1061/(asce)wr.1943-5452.0001575.
686 Hundman, K., V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom. 2018. “Detecting
687 spacecraft anomalies using lstms and nonparametric dynamic thresholding.” *Proc. 24th*
688 *ACM SIGKDD Int. Conf. Knowl. Discov. \& data Min.*, 387--395.
689 Jason, B. 2018. “Deep Learning For Time Series Forecasting.” *Ml*, 1 (1): 1–50.
690 https://doi.org/10.1093/brain/awf210.
691 Kadri, F., F. Harrou, S. Chaabane, Y. Sun, and C. Tahon. 2016. “Seasonal ARMA-based SPC
692 Charts for Anomaly Detection: Application to Emergency Department Systems.”
693 *Neurocomputing*, 173: 2102–2114. https://doi.org/10.1016/j.neucom.2015.10.009.
694 Klise, K. A., D. Hart, D. M. Moriarty, M. L. Bynum, R. Murray, J. Burkhardt, and T. Haxton.
695 2017. *Water Network Tool for Resilience (WNTR) User Manual*.
696 Lambert, A. 2001. “What Do We Know About Pressure-Leakage Relationships in Distribution
697 Systems.” *IWA Conf. n Syst. approach to leakage Control water Distrib. Syst. Manag.*
698 Li, Y., H. Shi, F. Han, Z. Duan, and H. Liu. 2019. “Smart wind speed forecasting approach
699 using various boosting algorithms, big multi-step forecasting strategy.” *Renew. Energy*,

135: 540–553. Elsevier Ltd. https://doi.org/10.1016/j.renene.2018.12.035.

Li, Z., J. Wang, H. Yan, S. Li, T. Tao, and K. Xin. 2022. “Fast Detection and Localization of Multiple Leaks in Water Distribution Network Jointly Driven by Simulation and Machine Learning.” *J. Water Resour. Plan. Manag.*, 148 (9): 05022005.

Ma, X., Y. Li, W. Zhang, X. Li, Z. Shi, J. Yu, J. Wang, and J. Liu. 2022. “A Real-Time Method to Detect the Leakage Location in Urban Water Distribution Networks.” *J. Water Resour. Plan. Manag.*, 148 (12). https://doi.org/10.1061/(asce)wr.1943-5452.0001628.

Maharaj, E. A., P. D’Urso, and J. Caiado. 2019. *Time Series Clustering and Classification*. *Time Ser. Clust. Classif.*

Marzola, I., F. Mazzoni, S. Alvisi, and M. Franchini. 2022. “Leakage Detection and Localization in a Water Distribution Network through Comparison of Observed and Simulated Pressure Data.” *J. Water Resour. Plan. Manag.*, 148 (1): 1–11. https://doi.org/10.1061/(asce)wr.1943-5452.0001503.

Mishra, A., R. Sriharsha, and S. Zhong. 2022. “OnlineSTL: Scaling Time Series Decomposition by 100x.” *Proc. VLDB Endow.*, 15 (7): 1417–1425. https://doi.org/10.14778/3523210.3523219.

Moridi, M. A., M. Sharifzadeh, Y. Kawamura, and H. D. Jang. 2018. “Development of wireless sensor networks for underground communication and monitoring systems (the cases of underground mine environments.” *Tunn. Undergr. Sp. Technol.*, 73: 127–138.

Mounce, S. R., and J. Machell. 2006. “Burst Detection Using Hydraulic Data from Water Distribution Systems with Artificial Neural Networks.” *Urban Water J.*, 3 (1): 21–31. https://doi.org/10.1080/15730620600578538.

Mounce, S. R., R. B. Mounce, and J. B. Boxall. 2011. “Novelty Detection for Time Series Data Analysis in Water Distribution Systems Using Support Vector Machines.” *J. Hydroinformatics*, 13 (4): 672–686. https://doi.org/10.2166/hydro.2010.144.

De Nadai, M., and M. Van Someren. 2015. “Short-term anomaly detection in gas consumption through ARIMA and Artificial Neural Network forecast.” *2015 IEEE Work. Environ. Energy, Struct. Monit. Syst. EESMS 2015 - Proc.*, 250–255. IEEE. https://doi.org/10.1109/EESMS.2015.7175886.

Puust, R., Z. Kapelan, D. A. Savic, and T. Koppel. 2010. “A review of methods for leakage management in pipe networks.” *Urban Water J.*, 7 (1): 25–45. https://doi.org/10.1080/15730621003610878.

Romano, M., Z. Kapelan, and D. A. Savić. 2012. “Real-time Leak Detection in Water Distribution Systems.” *Water Distrib. Syst. Anal. 2010 - Proc. 12th Int. Conf. WDSA 2010*, 1074–1082. https://doi.org/10.1061/41203(425)97.

Romano, M., Z. Kapelan, and D. A. Savić. 2014. “Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems.” *J. Water Resour. Plan. Manag.*, 140 (4): 457–467. https://doi.org/10.1061/(asce)wr.1943-5452.0000339.

Sahoo, D., N. Sood, U. Rani, G. Abraham, V. Dutt, and A. D. DIleep. 2020. “Comparative Analysis of Multi-Step Time-Series Forecasting for Network Load Dataset.” *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, (July). https://doi.org/10.1109/ICCCNT49239.2020.9225449.

Siami-Namini, S., N. Tavakoli, and A. Siami Namin. 2019. “A Comparison of ARIMA and LSTM in Forecasting Time Series.” *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, 1394–1401. IEEE. https://doi.org/10.1109/ICMLA.2018.00227.

Steffelbauer, D. B., J. Deuerlein, D. Gilbert, E. Abraham, and O. Piller. 2022. “Pressure-Leak Duality for Leak Detection and Localization in Water Distribution Systems.” *J. Water Resour. Plan. Manag.*, 148 (3): 1–13. https://doi.org/10.1061/(asce)wr.1943-5452.0001515.

Ben Taieb, S., G. Bontempi, A. F. Atiya, and A. Sorjamaa. 2012. “A review and comparison

750 of strategies for multi-step ahead time series forecasting based on the NN5 forecasting
751 competition." *Expert Syst. Appl.*, 39 (8): 7067–7083. Elsevier Ltd.
752 https://doi.org/10.1016/j.eswa.2012.01.039.
753 Tealab, A., H. Hefny, and A. Badr. 2017. "Forecasting of nonlinear time series using ANN."
754 *Futur. Comput. Informatics J.*, 2 (1): 39--47.
755 Vrachimis, S. G., D. G. Eliades, R. Taormina, Z. Kapelan, A. Ostfeld, S. Liu, M. Kyriakou, P.
756 Pavlou, M. Qiu, and M. M. Polycarpou. 2022. "Battle of the Leakage Detection and
757 Isolation Methods." *J. Water Resour. Plan. Manag.*, 148 (12): 1–20.
758 https://doi.org/10.1061/(asce)wr.1943-5452.0001601.
759 Vrachimis, S. G., D. G. Eliades, R. Taormina, A. Ostfeld, Z. Kapelan, S. Liu, M. Kyriakou, P.
760 Pavlou, M. Qiu, and M. M. Polycarpou. 2020. "BattLeDIM : Battle of the Leakage
761 Detection and Isolation Methods." *2nd Int. CCWI/WDSA Jt. Conf.*, 1–6.
762 Wan, X., P. K. Kuhanestani, R. Farmani, and E. Keedwell. 2022a. "Literature Review of Data
763 Analytics for Leak Detection in Water Distribution Networks: A Focus on Pressure and
764 Flow Smart Sensors." *J. Water Resour. Plan. Manag.*, 148 (10).
765 https://doi.org/10.1061/(ASCE)WR.1943-5452.0001597.
766 Wan, X., R. Farmani, and E. Keedwell. 2022b. "Online Leakage Detection System Based on
767 EWMA-Enhanced Tukey Method for Water Distribution Systems. " *J. Hydroinformatics.*
768 https://doi.org/10.2166/hydro.2022.079.
769 Wang, X., G. Guo, S. Liu, Y. Wu, X. Xu, and K. Smith. 2020. "Burst Detection in District
770 Metering Areas Using Deep Learning Method." *J. Water Resour. Plan. Manag.*, 146 (6):
771 1–12. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001223.
772 Yaacob, A. H., I. K. T. Tan, S. F. Chien, and H. K. Tan. 2010. "ARIMA based network anomaly
773 detection." *2nd Int. Conf. Commun. Softw. Networks, ICCSN 2010*, (1): 205–209. IEEE.
774 https://doi.org/10.1109/ICCSN.2010.55.

775 Yasin, H., R. E. Caraka, Tarno, and A. Hoyyi. 2016. “Prediction of Crude Oil Prices Using
776 Support Vector Regression (SVR) with Grid Search—Cross Validation Algorithm.” *Glob.*
777 *J. Pure Appl. Math.*, 12 (4): 3009--3020.
778 Ye, G., and R. A. Fenner. 2011. “Kalman Filtering of Hydraulic Measurements for Burst
779 Detection in Water Distribution Systems.” *J. Pipeline Syst. Eng. Pract.*, 2 (1): 14–22.
780 https://doi.org/10.1061/(ASCE)PS.1949-1204.0000070.
781 Zaman, D., M. K. Tiwari, A. K. Gupta, and D. Sen. 2020. “A review of leakage detection
782 strategies for pressurised pipeline in steady-state.” *Eng. Fail. Anal.*, 109 (October 2019):
783 104264. Elsevier. https://doi.org/10.1016/j.engfailanal.2019.104264.
784 Zhou, X., Z. Tang, W. Xu, F. Meng, X. Chu, K. Xin, and G. Fu. 2019. “Deep Learning
785 Identifies Accurate Burst Locations in Water Distribution Networks.” *Water Res.*, 166:
786 115058. Elsevier Ltd. https://doi.org/10.1016/j.watres.2019.115058.
787

Table 1. Leak information with different levels

Leaking level	Leak duration	Maximum leak flow (m^3/h)	Average daily leak increase (m^3/h/day)	Percentage of daily leakage to average daily water demand
1	1 month	< 9.4	0.33	< 0.2%
2	1 month	9.4 - 18.8	0.33-0.63	0.2%< 0.4%
3	1 month	18.8 - 28.2	0.63-0.94	0.4%< 0.6%
4	1 month	28.2 – 37.6	0.94-1.25	0.6%< 0.8%
5	1 month	37.6 - 47	1.25-1.57	0.8%< 1%
6	1 month	> 47	>1.57	> 1%

Table 2. Performance of models on testing data

Model	MAPE				RMSE				R^2			
	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h	6 h	12 h	24 h	48 h
Recursive-ANN	0.13	0.15	0.16	0.37	7.13	8.12	8.96	28.66	0.85	0.76	0.81	0.05
MIMO-ANN	0.05	0.05	0.06	0.05	0.70	0.57	0.80	0.46	0.97	0.97	0.96	0.97

Table 3. Detection time of the proposed method with different prediction horizons

Leakage level	Detection time (days)			
	ANN – 6 h	ANN – 12h	ANN – 24 h	ANN – 48 h
1	-	-	13.96	10.69
2	-	-	12.11	8.76
3	27.14	24.77	6.53	6.58
4	21.67	19.56	5.60	5.90
5	18.88	13.46	5.17	5.21
6	16.03	11.88	4.67	5.16

Table 4. Detection results of the proposed method and Wan et al. (2022b)

Leakage level	Detection probability		Average detection time (days)		Leakage flow rate when detected (m^3/h)		Water loss when detected (m^3)	
	The proposed method	The statistical method	The proposed method	The statistical method	The proposed method	The statistical method	The proposed method	The statistical method
1	100%	0%	13.96	--	3.13	--	529.83	--
2	100%	0%	12.11	--	5.10	--	776.92	--
3	100%	0%	6.53	--	3.92	--	375.76	--
4	100%	39.2%	5.60	5.24	5.10	4.96	348.53	311.88
5	100%	100%	5.17	5.00	5.84	7.03	371.60	421.80
6	100%	100%	4.67	2.21	6.15	3.84	367.50	101.84

Table 5. Detection results of datasets with different sampling intervals (SI) based on the proposed method

Leakage level	Detection probability				Average detection time (days)			
	SI=5min	SI=15min	SI=30min	SI=60min	SI=5min	SI=15min	SI=30min	SI=60min
1	100%	100%	0%	0%	14.59	23.21	--	--
2	100%	100%	100%	100%	12.71	12.93	12.93	26.14
3	100%	100%	100%	100%	7.98	12.58	12.62	18.69
4	100%	100%	100%	100%	5.72	11.92	12.19	12.93
5	100%	100%	100%	100%	5.30	8.45	11.84	12.39
6	100%	100%	100%	100%	4.83	5.76	9.55	11.79

Figure1

Click here to access/download;Figure;Fig. 1..tif

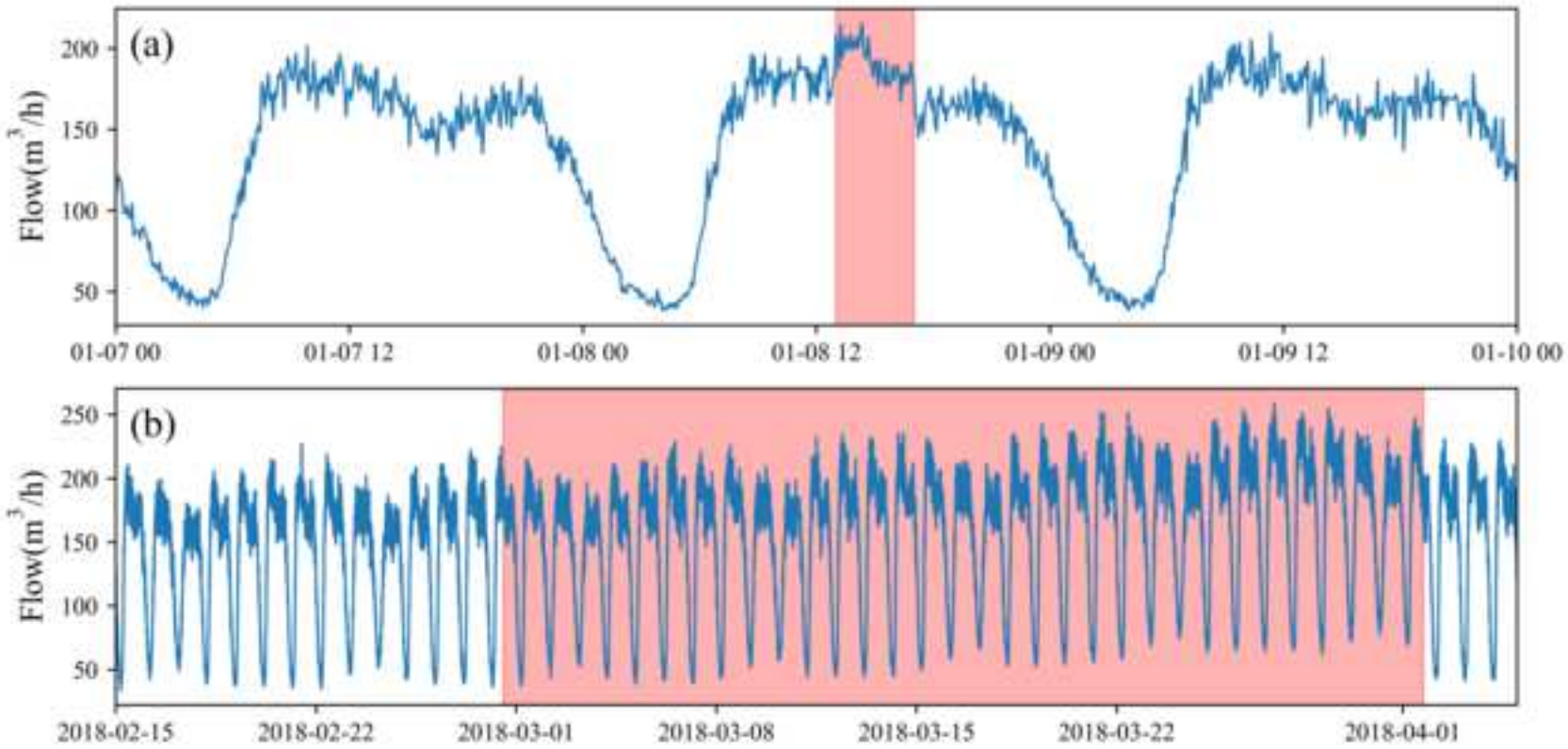


Figure2

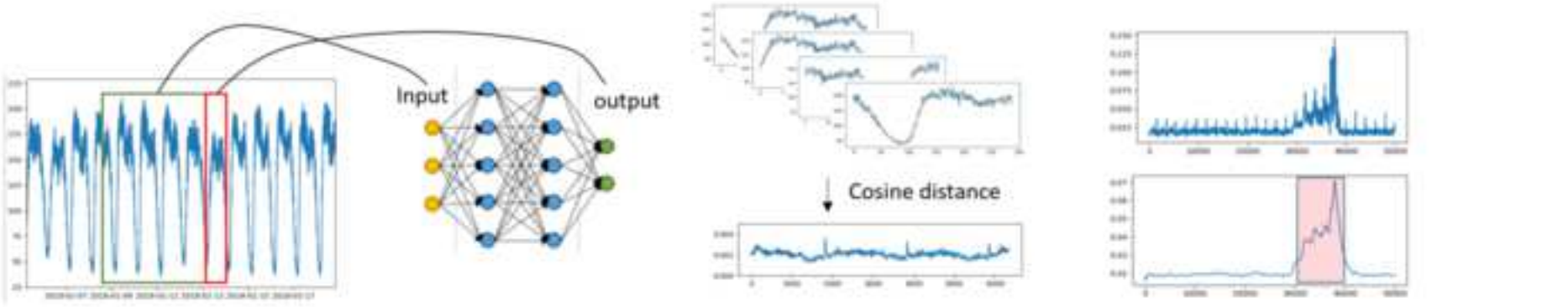
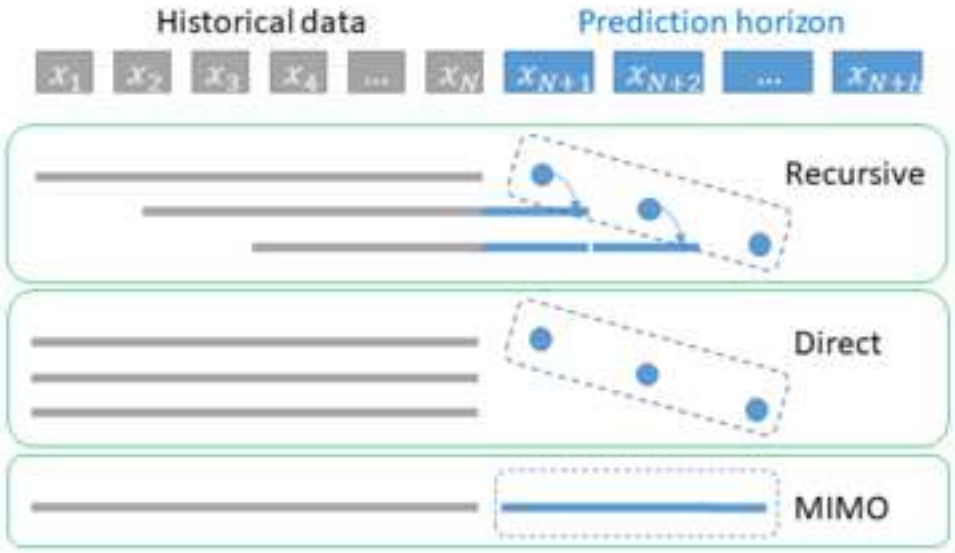


Figure3

Click here to access/download;Figure;Fig. 3..tiff



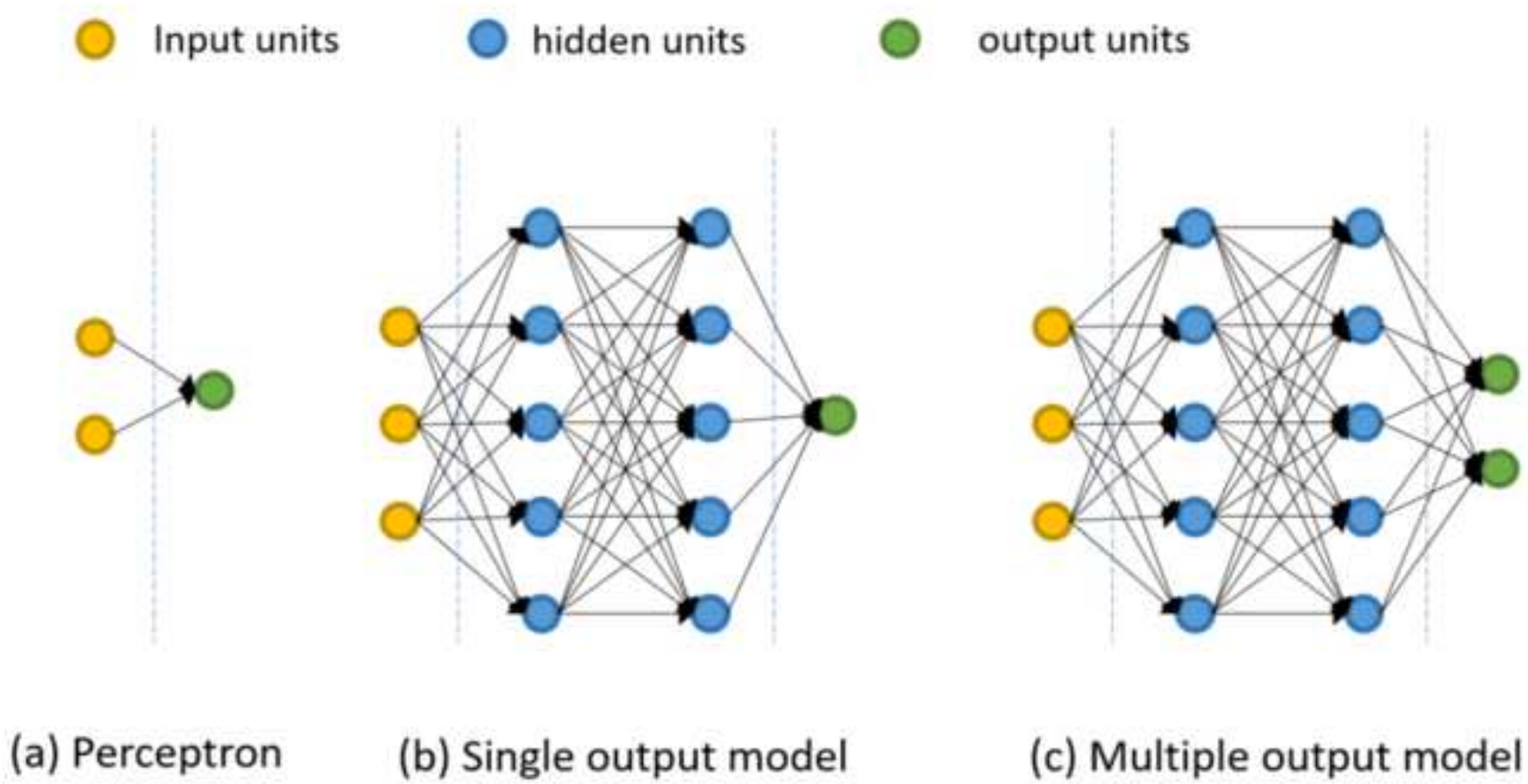


Figure5

Click here to access/download;Figure;Fig. 5..tiff

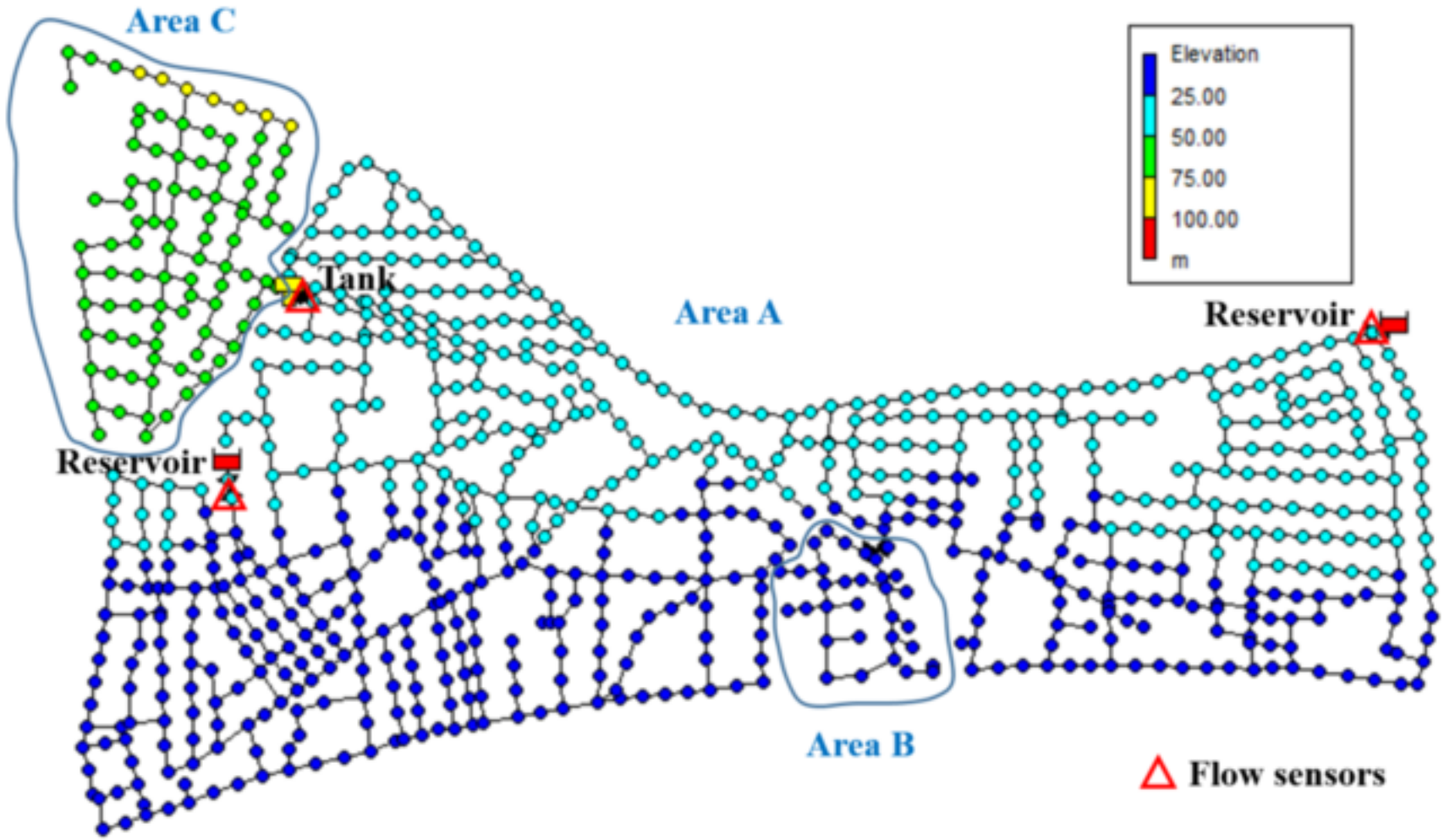


Figure6

Click here to access/download;Figure;Fig. 6..tif

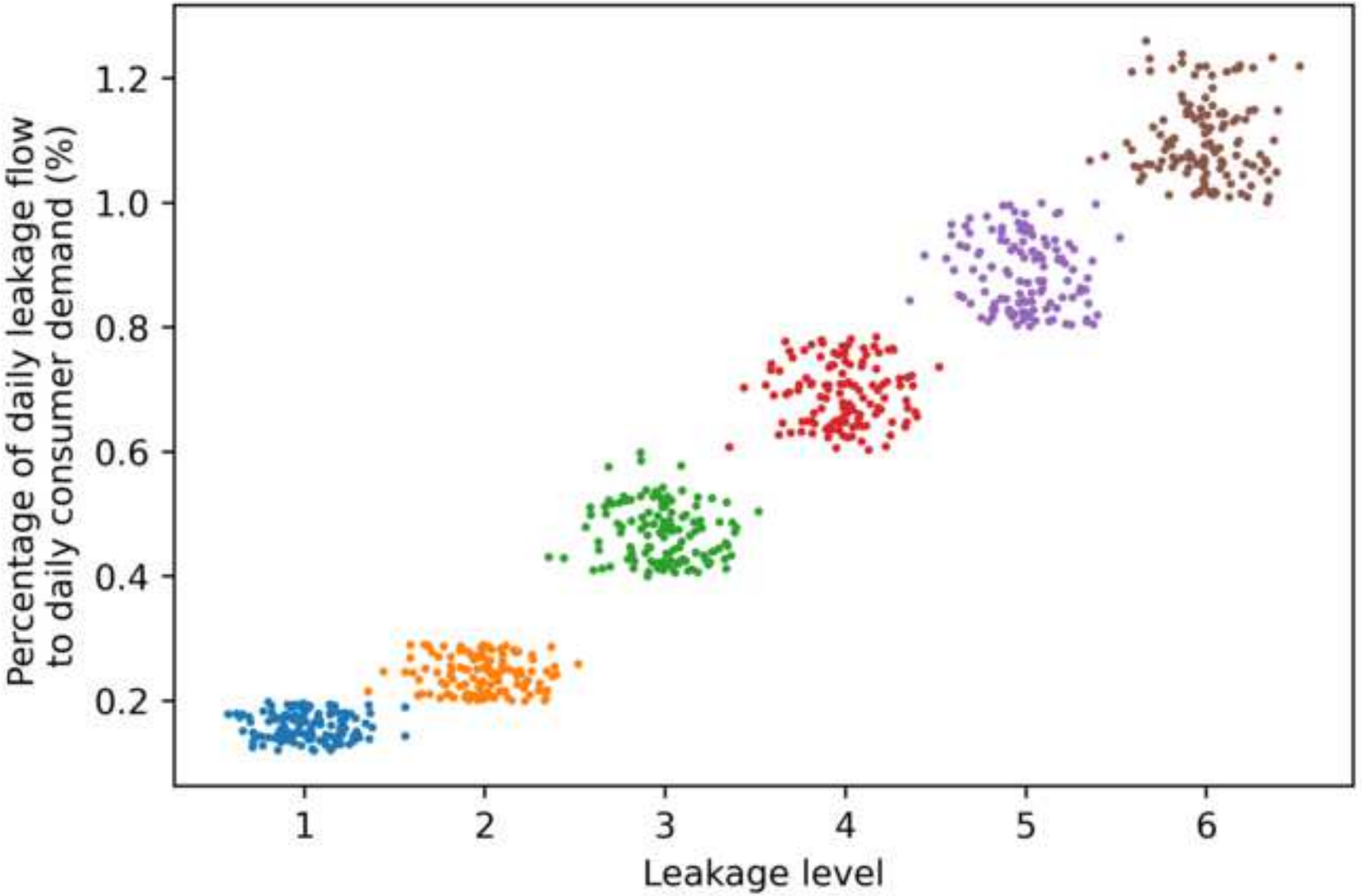


Figure7

Click here to access/download;Figure;Fig. 7..tif

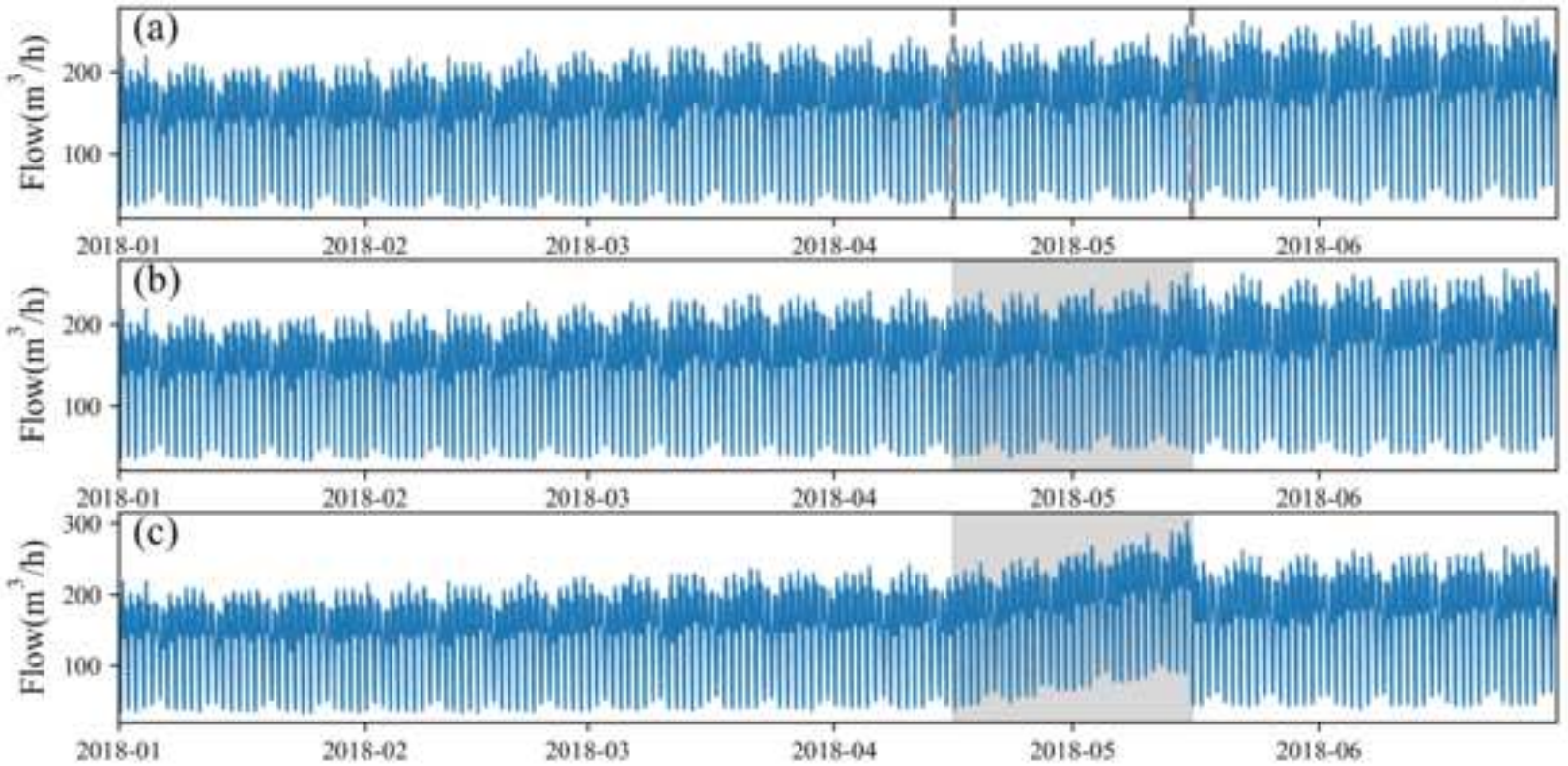


Figure8

Click here to access/download;Figure;Fig. 8..tiff

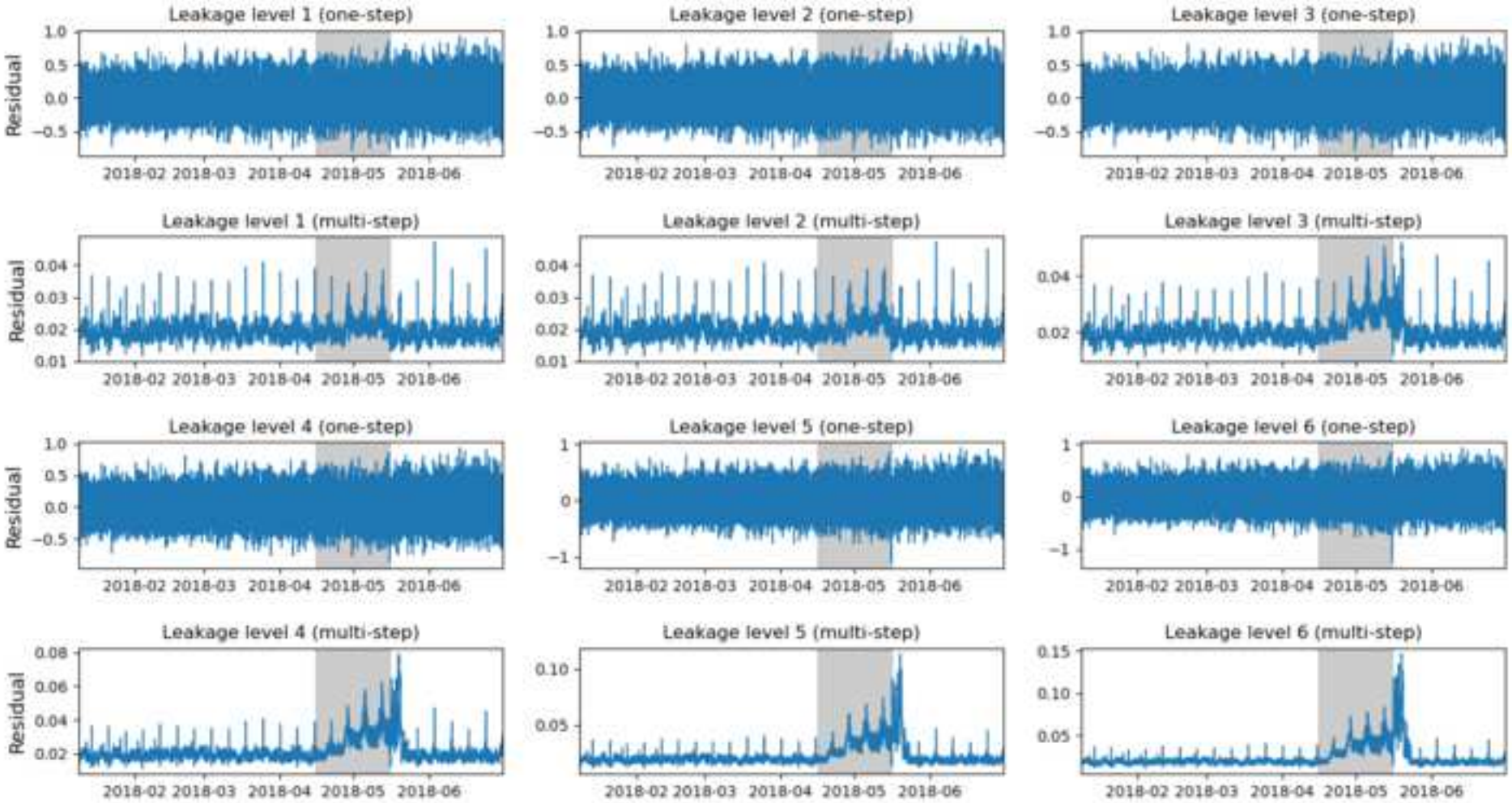


Figure9

Click here to access/download;Figure;Fig. 9..tiff

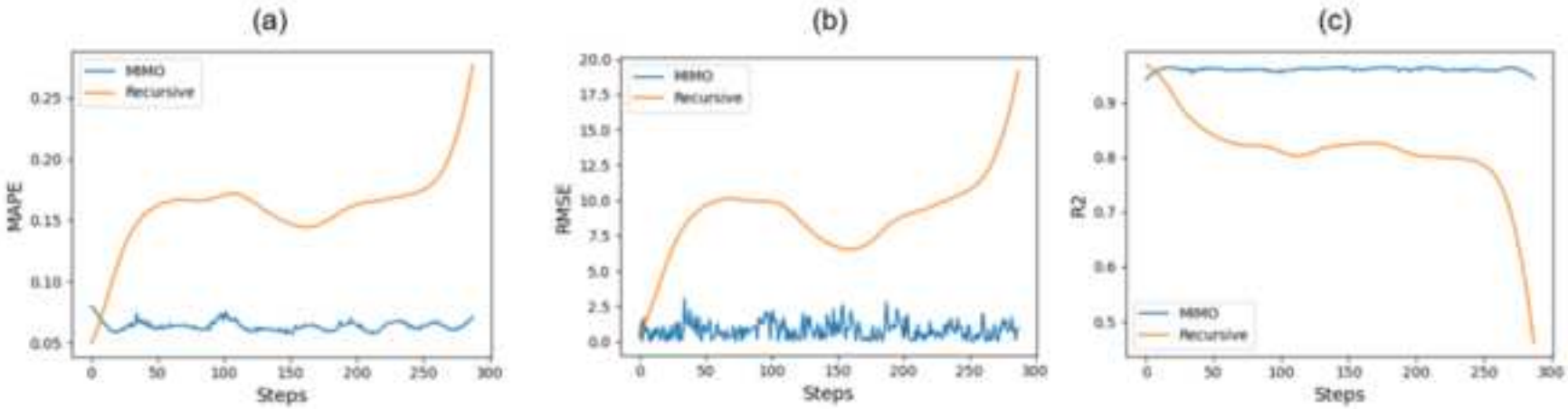


Figure10

Click here to access/download;Figure;Fig. 10..tiff

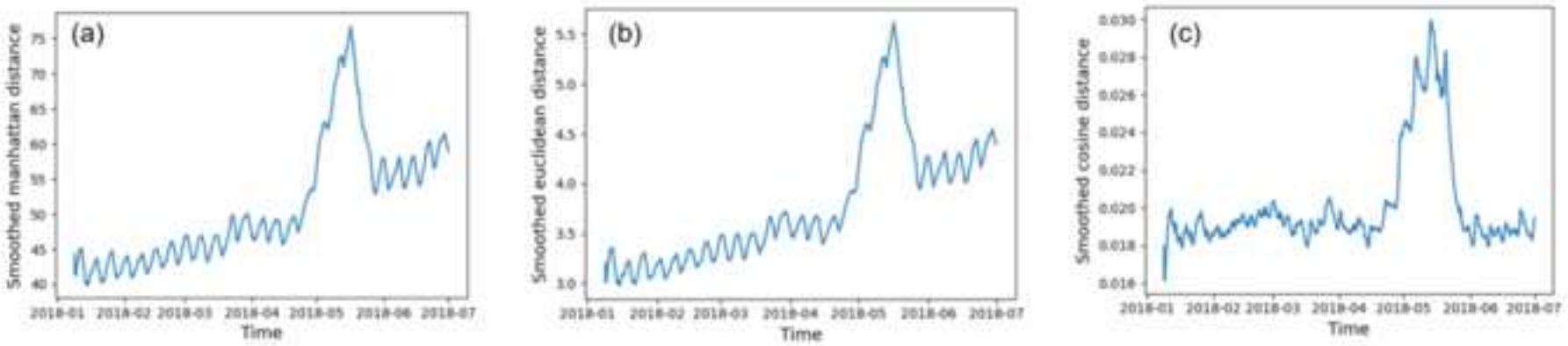


Figure 11

Click here to access/download;Figure;fig11.roc curve2.tif

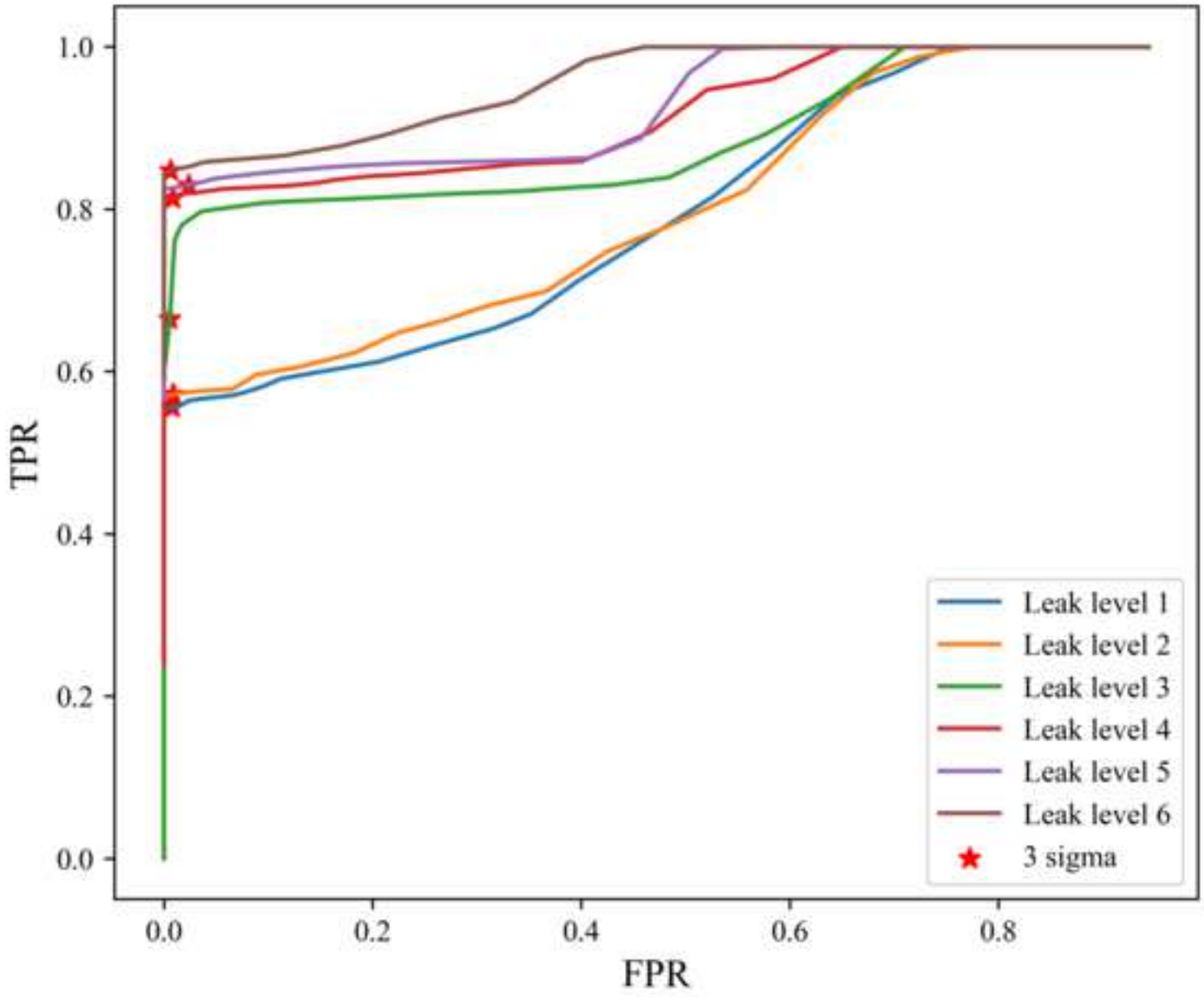


Figure12

Click here to access/download;Figure;fig12 real_data.tif

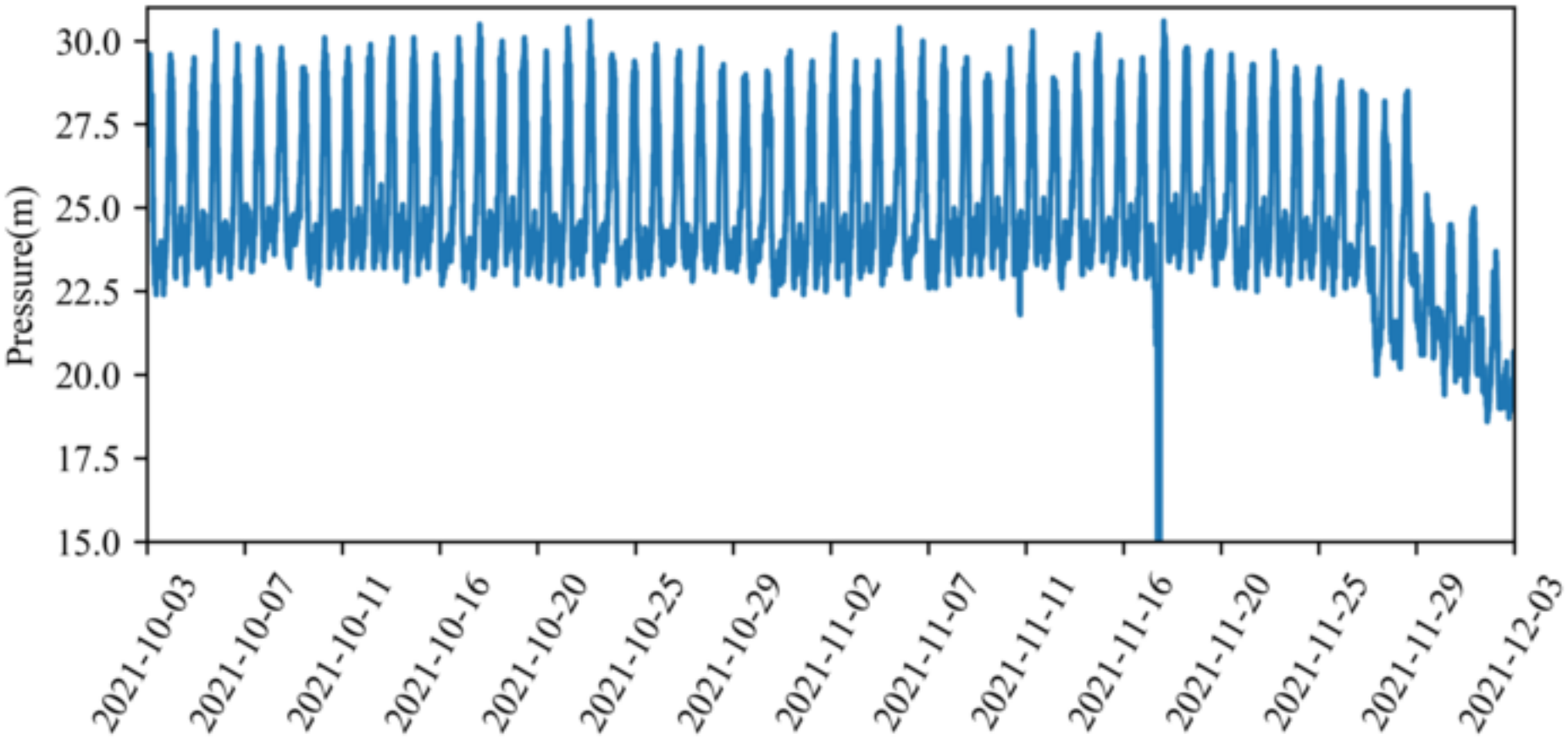
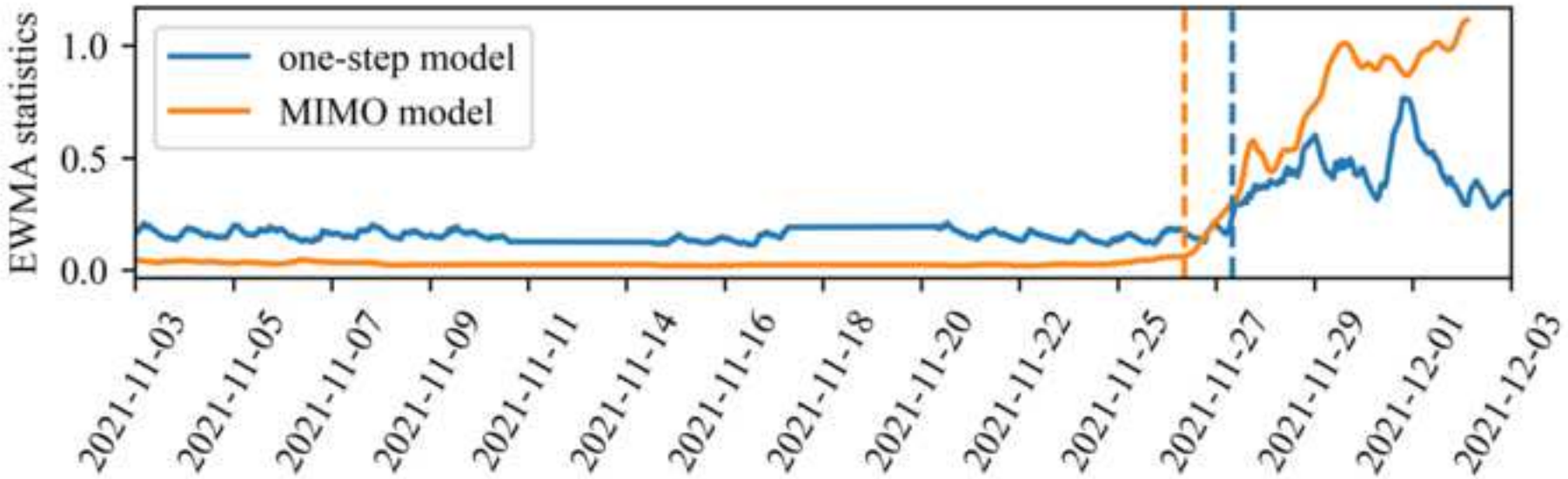


Figure13

Click here to access/download;Figure;fig13 real_results.tif



List of Figures