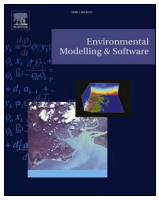
Environmental Modelling and Software 167 (2023) 105772

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Event-based decision support algorithm for real-time flood forecasting in urban drainage systems using machine learning modelling

Farzad Piadeh [a], Kourosh Behzadian [a,b,*], Albert S. Chen [c], Luiza C. Campos [b], Joseph P. Rizzuto [a], Zoran Kapelan [d]

[a] School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, UK
[b] Department of Civil, Environmental and Geomatic Engineering, University College London, Gower St, London, WC1E 6BT, UK
[c] Centre for Water Systems, Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK
[d] Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology (TU Delft), Delft, Netherlands

ARTICLE INFO

Handling Editor: Daniel P Ames

Keywords:
Event identification
Machine learning
Online platform
Real-time flood forecasting
Urban drainage systems

ABSTRACT

Urban flooding is a major problem for cities around the world, with significant socio-economic consequences. Conventional real-time flood forecasting models rely on continuous time-series data and often have limited accuracy, especially for longer lead times than 2 hrs. This study proposes a novel event-based decision support algorithm for real-time flood forecasting using event-based data identification, event-based dataset generation, and a real-time decision tree flowchart using machine learning models. The results of applying the framework to a real-world case study demonstrate higher accuracy in forecasting water level rise, especially for longer lead times (e.g., 2–3 hrs), compared to traditional models. The proposed framework reduces root mean square error by 50%, increases accuracy of flood forecasting by 50%, and improves normalised Nash–Sutcliffe error by 20%. The proposed event-based dataset framework can significantly enhance the accuracy of flood forecasting, reducing the occurrences of both false alarms and flood missing and improving emergency response systems.

Software and data availability

- Programming language: MATLAB 2021b using Machine learning and deep learning toolbox.
- Hardware requirement: Any computer with windows 10 and newer, any intel or AMD x86-64 processor, 4 GB minimum RAM, 8 GB minimum storage, no specific graphics card.
- Written code: 72 KB Modular code contains main file with 5 function files, available at github.com/FarzadPiadeh21452390/Event-based-platform.git
- Dataset: Ruislip water level data, Heathrow, Iver Heath, and RAF Northolt rainfall data (London, UK) used in this study. Real-time data are available and can be directly extracted by using an application programming interface (API) provided by the UK Environment Agency up to the last 28 days. Long-term historic data can also be available as csv file format in "environment.data.gov.uk/flood-monitoring" free of charge for research purposes by the UK Environment Agency upon request.

1. Introduction

Urban flooding is one of the most devastating natural disasters, and its impacts on economic, population, and property loss can be exacerbated by climate change (Xie et al., 2017). However, the use of real-time urban flood forecasting (RTUFF) models can help mitigate these impacts effectively by providing early warning for emergency response, early action, and contingency planning (Ahmed et al., 2021). Urban flooding can be described as a temporary overland flow in urban areas, including pluvial, fluvial, coastal, flash, groundwater, and urban drainage systems (UDS) flooding (Hamil, 2011). UDS flooding is a complex process that happens when excess water escapes from the UDS in one or more parts of the system (Chen et al., 2018). This type of flooding can be caused by a variety of factors, such as high-intensity rainfall, surface runoff in densely built urban areas, and lack of drainage capacity. The consequences of UDS flooding can have a significant impact on sustainability, including economic and human losses, business interruption, and damage to urban infrastructure (Piadeh et al., 2022b).

Urban flooding is a complex process that can be simulated and

* Corresponding author. School of Computing and Engineering, University of West London, St Mary's Rd, London, W5 5RF, UK.
E-mail address: kourosh.behzadian@uwl.ac.uk (K. Behzadian).

https://doi.org/10.1016/j.envsoft.2023.105772
Received 26 February 2023; Received in revised form 23 June 2023; Accepted 3 July 2023
Available online 5 July 2023

forecasted using physically based models (PBM) or RTUFF models based on time-series artificial neural networks (ANN) and machine learning (ML) (Zakaria et al., 2021). In recent years, the use of RTUFF models has gained significant attention from researchers due to the limitations of PBM, such as the high demand for data and the need to adopt a physically consistent model structure. PBM models in RTUFF are typically restricted to limited-dimensional spaces within human cognitive capacity and existing scientific knowledge of modelling in this field (Lin et al., 2021). Compared to PBM, ANN/ML models have been found to offer several advantages, including hyper-flexibility, the ability to require limited data for model parameters, and computational efficiency (Razavi et al., 2022). These advantages make ANN/ML models suitable for modelling complex mappings of real-time urban flood forecasting (RTUFF) with any form of correlations and dimensionality embedded in the dataset. (Razavi et al., 2021). In particular, computationally efficient ML models that require less and smaller types of data could play a crucial role in emergency planning and the real-time operation of urban flood early warning systems. These models can be retrained quickly after each flood occurrence, providing more accurate and timely forecasting in a very short time window e.g., few seconds, which can help decision makers implement proper structural and non-structural solutions, such as flood awareness, flood direction, or evacuation (Piadeh et al., 2023).

While there have been efforts to improve the interpretability of ML models to help stakeholders understand which factors are most influential in the decision-making process, it is important to acknowledge that achieving clear and understandable explanations from ANNs still remains a significant challenge. These models are often interpreted as black box simulations, which can limit their interpretability. Furthermore, while sensitivity analysis aids in comprehending the influence of input variations on model predictions and uncertainty analysis quantifies the potential range of outcomes due to input uncertainty, these analyses alone may not offer the desired level of flexibility for applications involving highly complex and dynamic rainfall or flood data (Razavi et al., 2021; Saltelli et al., 2021). This can lead to a reduction in the trust and preference of decision makers to fully rely on these models.

Furthermore, the developed models have mainly focused on flood susceptibility, forecasting of surface runoff discharging into UDS, flood inundation, flood risk analysis, and flooding alarm systems UDS and a few models have been developed for water level forecasting in combined sewer systems or stormwater collections systems (Luo et al., 2022). Various types of feed-forward neural networks (FFNN), including single-layer and multi-layer, have been employed as a surrogate for physically based models in hydraulic and hydrological modelling (Li et al., 2023; Piadeh et al., 2022a). Additionally, weak learner data mining models have been used for alarm-based flood occurrence detection (Piadeh et al., 2022b). Recurrent neural networks, particularly those with long short-term memory (LSTM), have been shown to improve flood forecasting performance by capturing temporal behaviour through extra memory (Nanda et al., 2019). While LSTMs have been primarily used for rainfall forecasting and satellite-based deep learning (Adikari et al., 2021), Nonlinear Auto-Regressive eXogenous (NARX) is also introduced as a recurrent neural network forecasting model which is suitable for multivariate time series data (Chang et al., 2014; Nanda et al., 2019; Huang et al., 2021). However, one major challenge is the accuracy of flood forecast at various lead times (i.e., the flood forecast period which is defined as the latency between the real/current timestep, and flood forecast timestep). More specifically, the accuracy of these models for long lead times, such as several hours, is quite limiting although it can be acceptable for short lead time e.g., less than 1 hr. For example, accurately forecasted water levels are limited to approximately 120min (Mounce et al., 2014; Garofalo et al., 2017; Abou Rjeily et al., 2018), which hinders generating acceptable key performance indicators (KPIs) and results in a low rate of Nash-Sutcliffe model efficiency coefficient (NSE) and a high range of root mean square error (RMSE) (Zhang et al., 2018).

The accuracy of flood forecasting models may be hindered by the database used for model development, which typically requires long-term records of rainfall and flood events, specifically water levels in UDS. These time-series data are usually continuous and include a large number of observations from non-flood conditions, or dry weather periods, which can lead to bias in model development (Piadeh et al., 2021). While these models use different definitions of flood events to evaluate performance, they inevitably require the entire database, including both dry and wet weather periods (the latter of which can be further divided into flood and non-flood events), for training and validation. Despite the focus on developing advanced or hybrid models, there is often a lack of attention given to classifying the database based on possible events, such as accurate identifying flood and non-flood events.

This may be attributed to the absence of a comprehensive and clear definition for identifying a flood event. Darabi et al. (2020) noted that there has been no universally accepted definition for flood event identification in the concept of RTUFF. ButlerDigman et al. (2018) mentioned that differentiating between the degree and extent of interaction between different events is complex and requires further investigation. Some studies have categorised time-series test data, training and validation data were excluded, based on specific time windows, such as using annual separation of databases as an individual event (Lv et al., 2020; Alizadeh et al., 2021). Alternatively, a few studies have classified flood events based on rainfall characteristics. For example, Adikari et al. (2021) used a three-class chamber flow by considering annual peak runoff and runoff threshold, and Lin et al. (2021) introduced flood events based on rainfall duration and return period. However, all of these research works demonstrate that there is no comprehensive framework for flood event identification and relevant database preparation, including event identification or event-based dataset generation for the RTUFF modelling. They also noted that while training models with distinguishing between flood and non-flood events would be ideal, defining accurate flood events and manipulating relevant data may still be challenging.

The objective of this study is to introduce a novel event-based decision support algorithm for developing RTUFF models using popular time-series ML methods including NARX and FFNN. The framework aims to create ML models that are trained on an event-based dataset. These models are used based on decision support system for real-time forecasting. The main aim of the present study is to provide more accurate forecasting water level in the longest lead time in comparison to current conventional RTUFF. This higher accuracy is crucial for real early warning systems, as both false alarms and missed flooding events can have significant consequences for stakeholders, particularly the public, local authorities, and relevant emergency response systems. False alarms can lead to a loss of confidence in the early warning system and may result in reduced sensitivity to future warnings, while missed flooding events can result in a lack of effective response, thereby disrupting trust between the public and local governments. As such, the proposed event-based dataset framework should be capable of accurately forecasting water level rise and flooding in the UDS in real-time, with longer lead times exceeding 2 hrs in comparison with currently available models. This can significantly improve the accuracy of flood forecasting, reducing both false alarms and missed events, and improving emergency response systems.

2. Methodology

A flooding event in the UDS is typically defined as the rise of water level in the drainage system that exceeds its capacity, causing water to overflow into flood plain areas and urban surfaces such as streets, pavements, and buildings (SEPA, 2021). Flooding in UDS can occur along any waterway, including urban rivers, channels, canals, and conduits that can be included in the present methodology. Besides, the proposed framework has the capability to incorporate various time-series data, including but not limited to, rainfall, air temperature, air moisture, wind speed, and wind direction, based on the level of complexity required by the developed model (Piadeh et al., 2022a). Also, various sources of rainfall data,

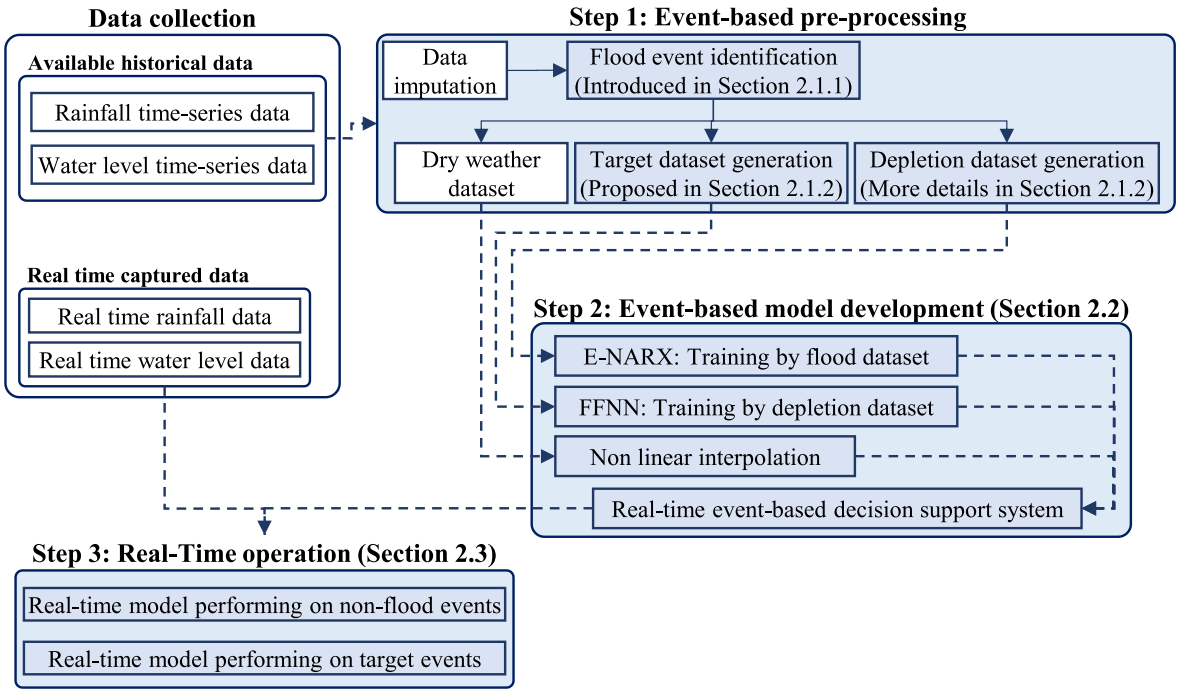


Fig. 1. The proposed framework of the event-based ML modelling for real-time urban flood forecasting.

such as IoT-based rainfall network monitoring stations, real-time data from radar stations, and data extracted from satellite productions may be available. In addition, numerous specifications of urban catchments, including land use, soil moisture, and infiltration rate, may also be accessible (Szeląg et al., 2022). However, this study utilises a limited set of selected data from one rainfall monitoring station and one water level gauging station, as these are the most commonly used devices for measuring rainfall and water level in hydrological practices (ButlerDigman et al., 2018). The primary reason for this is that the proposed framework aims to provide a generalised model that requires a minimum amount of data, making it applicable to other similar cases where data availability may be limited to a single set of rainfall and water level data. Finally, the presented framework focuses on temporal modelling of flood forecasts at a point e.g., gauging station of water level in the UDS. However, the framework can be utilised for several points. The proposed framework, as illustrated in Fig. 1, comprises four key steps: (1) collecting historical and real-time data, (2) generating an event-based dataset, (3) developing event-based ML models, and (4) assessing the model performance by real-time operation.

The continuous time-series data is utilised to distinguish between dry weather flow, target events (i.e., any water level rise due to rainfall occurrence), and depletion events (i.e., UDS flow rate declining after rainfall occurrence until water level returns to the base flow). Based on the data classification described in Section 2.1, new time-series datasets are generated in Step 2 using a new algorithm outlined in Section 2.2. In Step 3, the new databases generated from the original database (i.e., target dataset and depletion dataset in Fig. 1) are used to develop a real-time flood forecasting platform based on an event-based decision support algorithm and pre-trained ML models demonstrated in Section 2.3. Finally, the pre-trained platform is tested with new, unseen original data in a real-time simulation, and its performance is evaluated based on the description in Section 2.4.

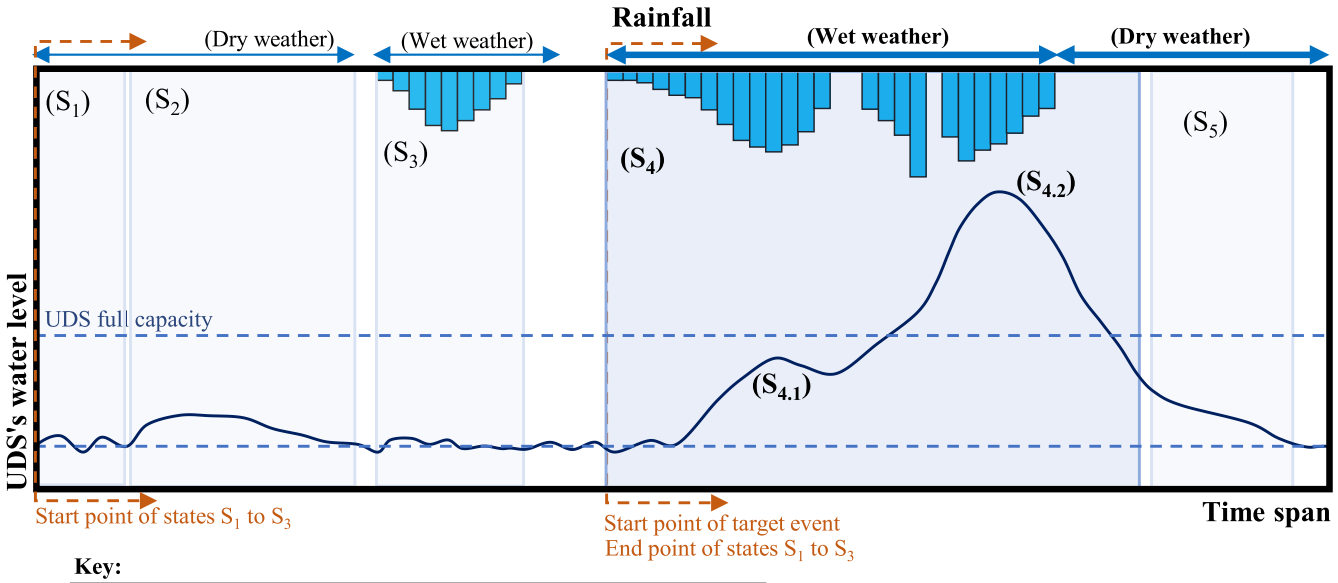
2.1. Step 1: Event-based pre-processing

2.1.1. Event identification method

All the collected data is first analysed to identify any missing rainfall or water level data, which are then imputed using copula-based regression recommended by Ben Aissia et al. (2017). For preparing data from weather and gauging stations, events of rainfall and runoff (denoted here to as water level in UDS) need to be first identified temporally. This can be carried out based on the classification of rainfall (dry or wet weather) and water level in the UDS classified under five states (S_1–S_5) as shown in Fig. 2a (inspired from Piadeh et al., 2021): (S_1- dry weather) with no rainfall and trivial/no change in the water level; (S_2- dry weather) with no rainfall but an increase in the water level due to several reasons such as sanitary sewage discharged into combined sewerage (diurnal pattern of wastewater), leakage/exfiltration, or infiltration; (S_3- wet weather) with rainfall without an increase in the water level due to either evaporation or infiltration into the soil and hence no surface runoff or discharge into the UDS; (S_4- wet and dry weather) with rainfall resulting in an increase in the water level with a lag time; (S_5- dry weather) with no rainfall and decrease in the water level back to S_1 (base flow).

Among these five states, S_4 is considered here as the target event that can result in flooding ($S_{4.2}$) or only rising water level without exceeding the full capacity of the UDS ($S_{4.1}$) as shown in Fig. 2a. The risk of the UDS flooding is defined here as any rising water level due to rainfall occurrence. This is because while $S_{4.2}$ can only result in flooding (high flood risk), the UDS can manage to convey any excessive rainfall and surface runoff without flooding, i.e., $S_{4.1}$, which is classified as a low risk of further flooding. Target events are initiated by the occurrence of rainfall and concluded using the curvature method (Hamil, 2011). Specifically, when rainfall ceases and the water level begins to decline, the inflection point is identified as the endpoint situated between the recession curve and the depletion curve in the falling limb of the water level hydrograph (see Fig. 2b). Details of described algorithm are provided in pseudo-code Algorithm A1 in the appendix.

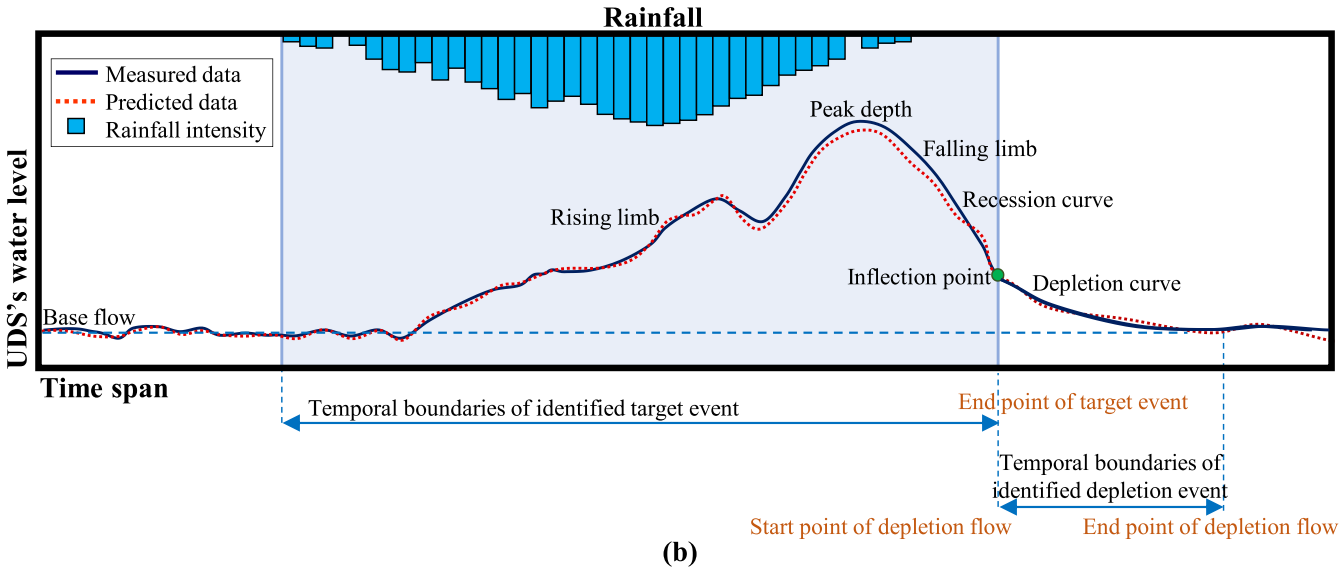
Although all of the generated data and the mentioned data classification will be utilised for real-time flood forecasting in the UDS (see Section 2.3 and the event-based decision support algorithm), only a portion of the data is selected for the training and validation of the ML models. The selected data is then saved in two separate databases: (1) matrix dataset including time-series data of water level depletion events (i.e. S_5 in Fig. 2a) as n samples with m elements, in which n is the number of identified depletion events and m is timesteps of each



Key:

State	Captured data	
	Rainfall	Water level
(S_1): Dry weather	Zero	No change
(S_2): Sudden rising flow	Zero	Increase or decrease
(S_3): Ineffective precipitation	Occurrence	No change
(S_4): Target event		
($S_{4.1}$): No flooding	Occurrence	Increase or decrease
($S_{4.2}$): Flooding	Occurrence	Increase or decrease
(S_5): Depletion flow	Zero	Decrease

(a)



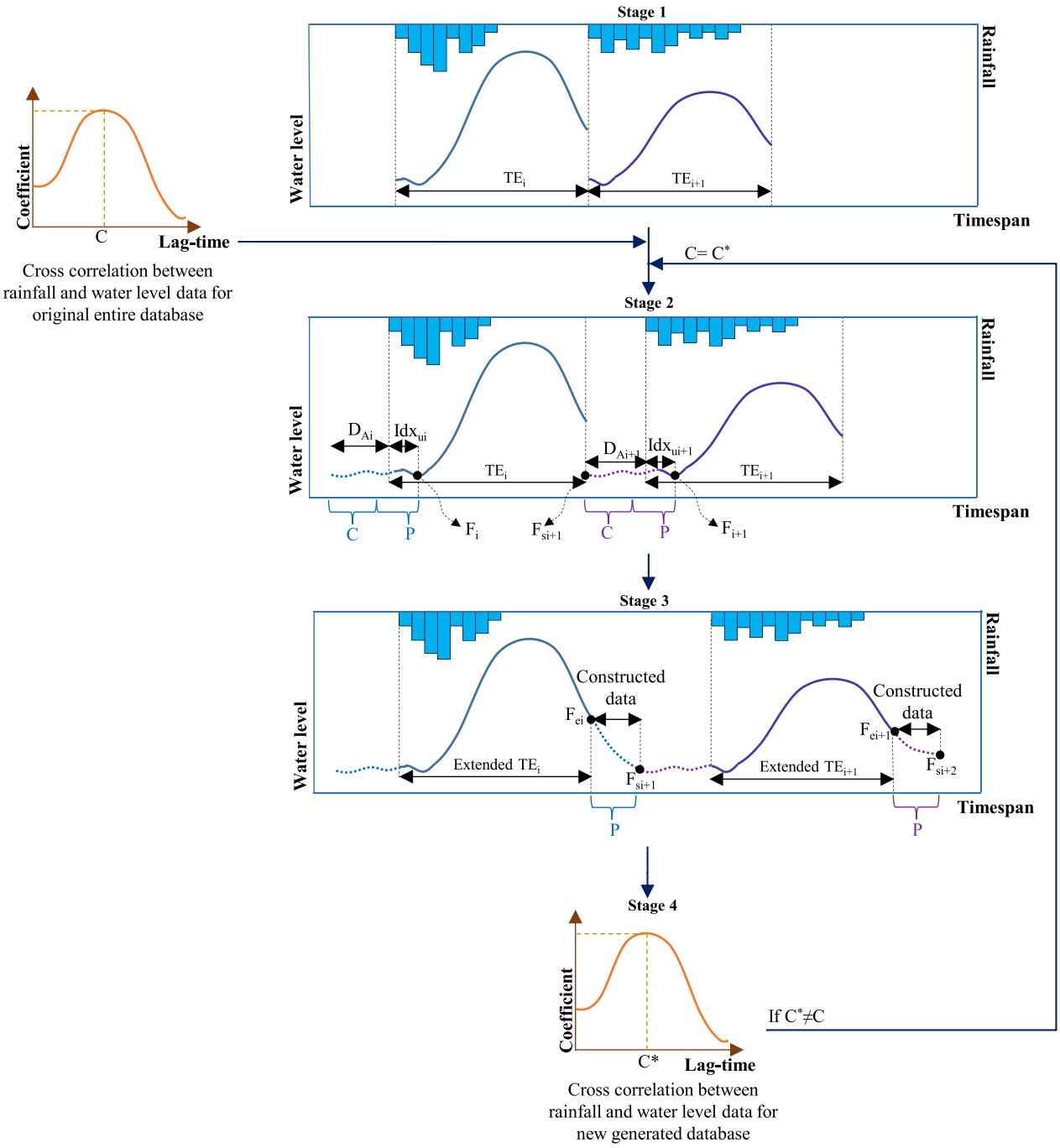
(b)

Fig. 2. Schematic representation of rainfall-runoff: (a) five-state classification of rainfall and the water level in UDS, (b) typical target events with temporal boundaries.

depletion event. This dataset is used to develop ML models which are responsible for water level forecasting in depletion situations. (2) different time-series dataset generated for developing time-series ML models which are responsible for water level forecasting in target event situations. The methodology of this database is described in next section.

2.1.2. *Event-based time-series dataset generation*

Time-series ML models usually require uninterrupted continuous time-series data for their training and validation. However, the output of the previous section includes various identified target events that occurred at different temporal timesteps and on different days or hours



Key

C: Best lag time in which cross coefficient between rainfall data and water level is highest
C*: Best lag time in which cross coefficient between rainfall data and water level in new generated database is highest
D_{Ai}: Added earlier timestep of data transformed from real data existed before i^{th} identified target event
F_i: First temporal point of water level rising in TE_i event
F_{ei}: Last temporal point of water level in TE_i event
F_{si+1}: First temporal point of water level in extended TE_i event
$Idxu_i$: time difference between the start point of the rainfall event and the timestep of water level rise in i^{th} target event
P: Interested timestep of forecasting
TE_i: Rainfall and water level series data extracted as i^{th} target event in the event identification step

Fig. 3. The four stages involved in generating an event-based time-series dataset: Stage 1) arranging all identified events in a consecutive manner; Stage 2) adding earlier timestep data to the beginning of each event; Stage 3) synthesising new data through nonlinear regression to extend the water level of the end point of each event until reaching the start point of the following event; and Stage 4) analysing the cross-correlation between water level and rainfall data in the newly generated dataset.

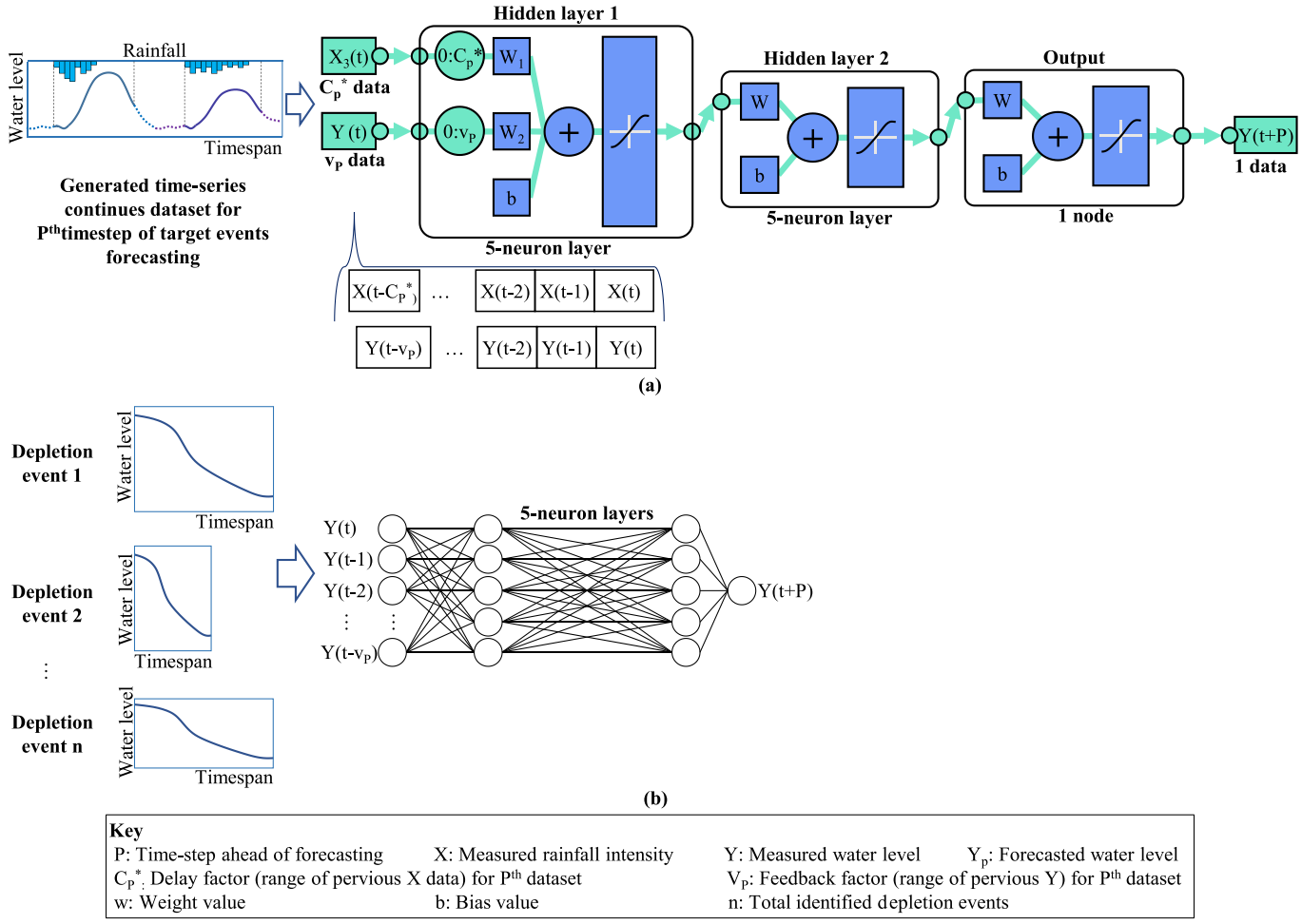


Fig. 4. Structure of the proposed ML models: (a) E-NARX used for water level forecasting of target events, (b) FFNN used for water level forecasting of depletion events.

of the day. Thus, this study introduces a novel event-based time-series dataset generation method for preparing the dataset used for water level forecasting in target event situations. This method involves creating datasets based on the identified target events (i.e., S_4 in Fig. 2a) for each lead time of interest (P). In other words, for each P-timesteps ahead of forecasting, this method is repeated to generate the proper dataset. The method involves four iterative stages proposed in Fig. 3 to generate the required dataset used for developing forecasting models with the desired timesteps ahead. These stages are defined and updated based on two key parameters: P and the lag time (C), which is the highest coefficient of the cross-correlation between rainfall and water level data as shown in Fig. 3 or Eq. (1).

$$C_C = \frac{\sum_1^n [[x(i)-\bar{x}] \times [[y(i-C)-\bar{y}]]]}{\sqrt{\sum_1^n [x(i)-\bar{x}]^2}\sqrt{\sum_1^n [y(i-C)-\bar{y}]^2}} \quad (1)$$

where C_C is the cross-correlation coefficient at the C^{th} lag time, which results in the highest coefficient, x and y are the two sets of data, n is the size of each dataset.

In Stage 1, the identified target events (rainfall and water level) are arranged in a continuous time-series in the order of their temporal occurrence. In Stage 2, the time-series dataset is extended by adding data from earlier time steps in the original database to the beginning of each event to ensure sufficient data for training the ML model. The number of earlier timesteps (D_A) added to the start of each event is calculated using Eq. (2), which depends on three factors: the best lag time corresponding to the highest cross-correlation between rainfall and water level in the entire original database (C), the time difference between the start of the rainfall event and the onset of water level rise in an event (Idx_u), and lead time of interest (P). To train ML models for forecasting the first instance of water level rising in a time series (e.g., F_i in TE_i event as shown in Stage 2 of Fig. 3), data from the preceding P time steps is necessary. For example, if the goal is to forecast water level rising four-time steps into the future, data from preceding four-timesteps need to be used. If some of the preceding-timestep data (i.e., Idx_{ui}) are available from Stage 1, the remaining data (DA_i) are collected from the historic database to be used as the database for training ML models.

$$D_A = C + P - Idx_u \quad (2)$$

To ensure that the ML model understands the nature of the event data completely, including water level decreasing until the inflection point (F_{ei} in Stage 3 of Fig. 3 for TE_i, for instance), it is necessary to add all available data to the training procedure of the ML model. However, these data can only be used for forecasting up to P time steps ahead. Furthermore, between the last water level measured in event i (see F_{ei} in Fig. 3 Stage 3) and the initial water level of extended event $i+1$(F_{si+1} in Fig. 3 Stage 2 and 3) that need to be filled in consistently. To address this issue, new data can be synthesised through nonlinear regression using a simple spline infilling method in Stage 3 (Jones et al., 2015). The start and end points for the spline infilling method are defined as the last water level measured in event i (see F_{ei} in Fig. 3 Stage 3) and the initial

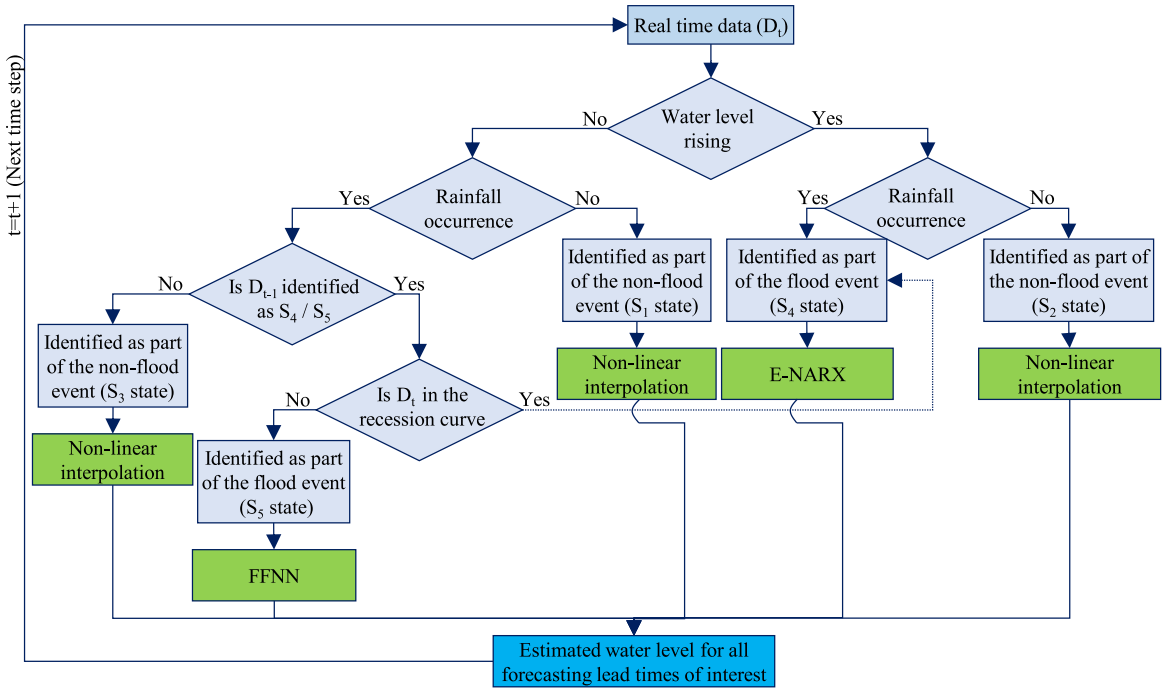


Fig. 5. Event-based decision support flowchart for real-time operation and flood forecasting in UDS.

water level of extended event *i+1* (F_{si+1} in Fig. 3 Stage 2 and 3), respectively and the number of timesteps added to the end of each event is equal to *P*.

Finally, Stage 4 includes analysing the cross-correlation between water level and rainfall of the new dataset and comparing the best lag time (i.e., the highest coefficient) of the new dataset (C^*) with the previous lag time (C). If they are different, a new C is assigned to be equal to C^*, and Stages 2–4 are repeated until C^* converges to C.

2.2. Step 2: Model development and decision tree of real-time forecasting

Two types of commonly-used ML models were developed in MATLAB software tool 2021a for water level forecasting. The nonlinear autoregressive network with exogenous inputs (hereafter E-NARX) model was used for water level forecasting of target events (S_4), while the FFNN model was used for water level forecasting of depletion states (S_5). Regardless of dry weather situations (S_1 to S_3), conventional RTUFF models have been employed to forecast both S_4 and S_5 states in other research works which are usually complicated and time-consuming. Therefore, a simple high speed FFNN model is selected here to take part of responsibility from E-NARX forecasting i.e., S_5 states that usually have decreasing pattern of water level and does not require a RNN model (See Fig. 2a). In previous studies, conventional RTUFF models, particularly complicated RNN models, have been utilised to forecast both S_4 and S_5 states (See these states in Fig. 2a), which can be time-consuming and require more computational efforts. However, in the present study, a simple, high-speed FFNN is selected to forecast S_5 states, which typically exhibit a decreasing pattern of water level (as shown in Fig. 2a), thereby conserving computational efforts and time.

E-NARX and FFNN were chosen for their wide use in previous hydrological applications, as demonstrated in introduction. The NARX model was trained using historical records of rainfall and water level data from the generated time-series dataset introduced in the previous section, while the FFNN model was trained using historical records of water level data from a matrix dataset of all depletion events. Depletion events occur when there is no rainfall within a certain period of time or is passed away time of concentration of the UDS catchment area. In such events, the water level reduction is not affected by rainfall (eighter has zero or non-zero values) and therefore, the rainfall data can be removed from the database of the FFNN modelling. Both models were structured as two hidden 5-neuron layers and are trained using the Levenberg-Marquardt method for the training process, with mean square error and 10 epochs adjusted for training failure. The detailed structure of both models is shown in Fig. 4 and Eq. (3).

$$Y(t+P)=\begin{cases} f\left[Y(t),Y(t-1),\ldots,Y(t-v_P),X(t),X(t-1),\ldots,X\left(t-C_P^*\right)\right] & \textit{for NARX model} \\ f[Y(t),Y(t-1),\ldots,Y(t-v_P)] & \textit{for FFNN model} \end{cases} \tag{3}$$

where $Y(t+P)$ is the forecasted water level at timestep t+i; X and Y are measured rainfall and water level at timestep t and preceding ones, respectively; C_P^* as delay factor corresponds to the best time lag that maximises the cross-correlation coefficient between rainfall and water level data for the P^{th} generated dataset (as described in Stage 4 of Fig. 3); v_P as feedback factor represents that maximises the auto-correlation coefficient in the water level data of the P^{th} generated dataset. Method of delay factor and feedback factor calculation are inspired as previous several works (Mounce et al., 2014; Chang et al., 2014; Abou Rjeily et al., 2018; Nanda et al., 2019; Huang et al., 2021). The original database was divided into three parts: 70% for calibration, 15% for validation, and 15% for test processes (more details will be discussed in Section 3 – case study). All the models were developed on a laptop equipped with an Intel i7-6700 HQ CPU @ 2.60 GHz and 16 GB RAM Memory.

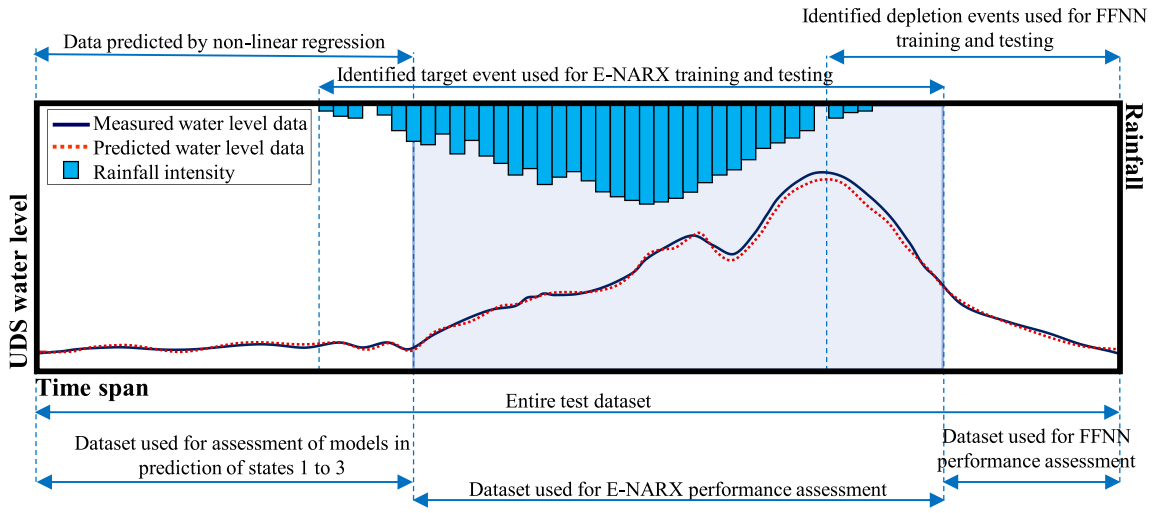


Fig. 6. Schematic presentation of classification of dataset used for training and test (performance assessment) of the proposed decision support algorithm.

This study introduced an event-based decision tree flowchart as a novel online platform for real-time operation of the UDS and flood forecasting, as depicted in Fig. 5. This platform operates based on the event identification method presented in Section 2.1 and pre-trained ML models. While the data resources were divided into training and validation sets for ML modelling, unseen test data were used to simulate real-time operation only. For this purpose, the proposed decision tree analyses the test data. If S_1, S_2, or S_3, indicating dry or wet weather conditions without any water level rise, is recognised, the estimation of water level in real-time flood forecasting involves the use of nonlinear regression techniques applied to previously captured water levels for desired forecasting time steps. This method is recommended when significant water level changes are not expected or causal factors are expected to remain constant (Chen et al., 2023). Historical trends of water level data are used to project the water level at the next time step and sequentially extended up to P time steps ahead. The number of historical data points used for nonlinear regression is determined by the catchment time of concentration, which represents the time it takes for water to travel from the furthest point of the catchment to the UDS. The accuracy and reliability of water level estimates can be enhanced by selecting the appropriate number of historical data points and considering the time of concentration. This improvement can lead to more effective flood management and mitigation strategies. For identifying the two other states, i.e., S_4 or S_5, the event-based identification method described in Section 2.2 is applied. After specifying the state, the associated pre-trained ML model is used to estimate the water level for any forecasting lead time of interest.

2.3. Step 3: Performance assessment for real-time operation

The developed decision support algorithm, which includes a decision support flowchart and pretrained ML models, is tested using new unseen data in a simulated real-time operation. The model performance is evaluated using three widely used KPIs, namely, RMSE, NNSE, and accuracy of flood forecasting, as defined in Eqs. (4)–(6).

$$RMSE\ (mm) = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - Y_{pi}\right)^2}{n}} \tag{4}$$

where n is the total number of forecasted data, Y_i is the ith measurement data, and Y_{pi} is the corresponding ith forecast data.

$$NNSE\ (\%) = 1 - \frac{1}{1 + \left(\frac{\sum_{i=1}^{n}\left(Y_i - Y_{pi}\right)^2}{\sum_{i=1}^{n}\left(\overline{Y_i} - Y_i\right)^2}\right)} \tag{5}$$

where NNSE is the normalised Nash–Sutcliffe model efficiency coefficient with an optimal value of 1.0 and a range between 0.0 and 1.0, and $\overline{Y_i}$ is the mean of measurement data.

$$\text{Accuracy}\ (\%) = 100 \times \frac{TP + TN}{FP + FN + TP + TN} \tag{6}$$

where accuracy is the ratio of the total number of true forecasts for flooding to the total number of forecasts with an optimal value of 100% and a range between 0.0 and 100%, *TP* is the number of true positive forecasts (correctly detected flood conditions), *TN* is the number of true negative forecasts (correctly detected non-flood conditions), *FP* is the number of false positive forecasts (wrongly indicates flooding for non-flood conditions), *FN* is the number of false negative forecasts (wrongly missed flood conditions).

It is important to note that even though the results are obtained in a simulated continuously real-time operation without any data manipulation, they are organised based on different states (S_1 to S_5), as illustrated in Fig. 6 (following the approach of Piadeh et al., 2021). This organisation facilitates a more comprehensive, in-depth, and detailed discussion of the results.

The real-time platform developed in this study was also subjected to sensitivity and uncertainty analyses (inspired by Meles et al., 2021; Razavi et al., 2021; Saltelli et al., 2021). The analyses focused on four following types of analysis: (1) sensitivity analysis of the hyperparameters of the developed models, including delay factors (range of input rainfall data for each training iteration), feedback factors (range of water level data for each training iteration), layers, nodes, and random training seed; (2) performance of the proposed model in forecasting water levels for different events, such as S_1 to S_5; (3) uncertainty analysis of performance on different sizes of the training dataset; and (4) performance of the proposed model based on different rainfall characteristics, such as intensity and duration. Relative RMSE increase and relative error (RE) as defined in Eqs. (8) and (9) were utilised for these analyses.

$$Reltaive\ RMSE\ increase\ (\%) = \frac{RMSE_{initial}}{RMSE_{new}} \times 100 \tag{7}$$

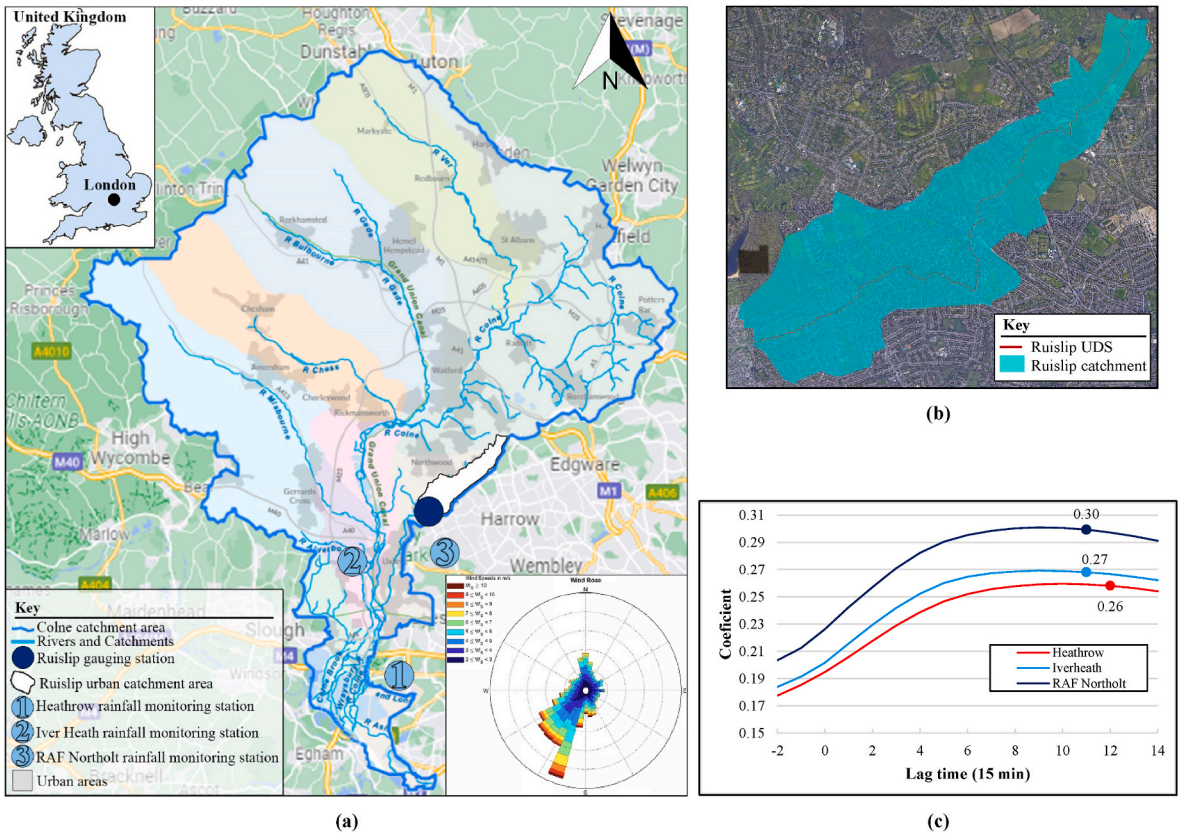


Fig. 7. Geographical location of the pilot study: (a) location of case study catchment and monitoring stations, (b) layout of Ruislip UDS and catchment, (c) cross-correlation between water level at Ruislip gauging station and selected three rainfall stations.

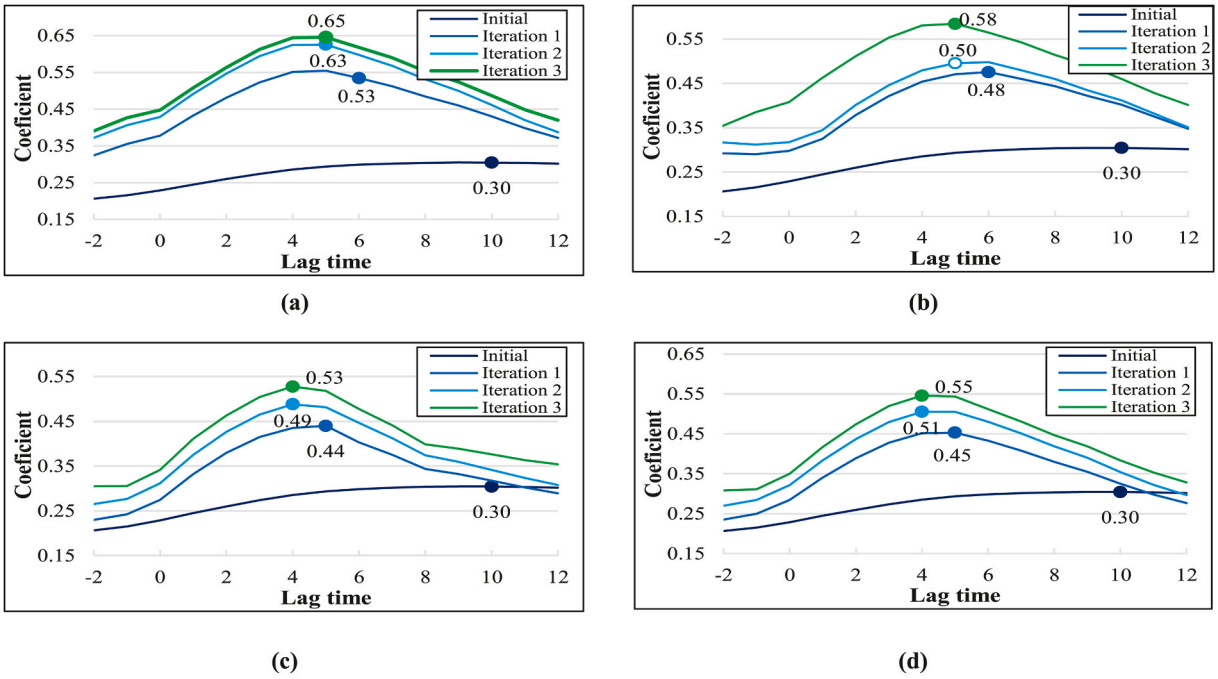


Fig. 8. Coefficient enhancement and lag time reduction for event-based datasets for lead time equal to (a) 1-timestep ahead, (b) 4-timestep ahead, (c) 8-timestep ahead, (d) 12-timestep ahead.

Where $RMSE_{initial}$ is the obtained RMSE for the developed model and $RMSE_{new}$ is new obtained RMSE of developed model for sensitivity analysis.

$$Relative\ error\ (\%) = \frac{Y_{pi}}{Y_i} \times 100 \tag{8}$$

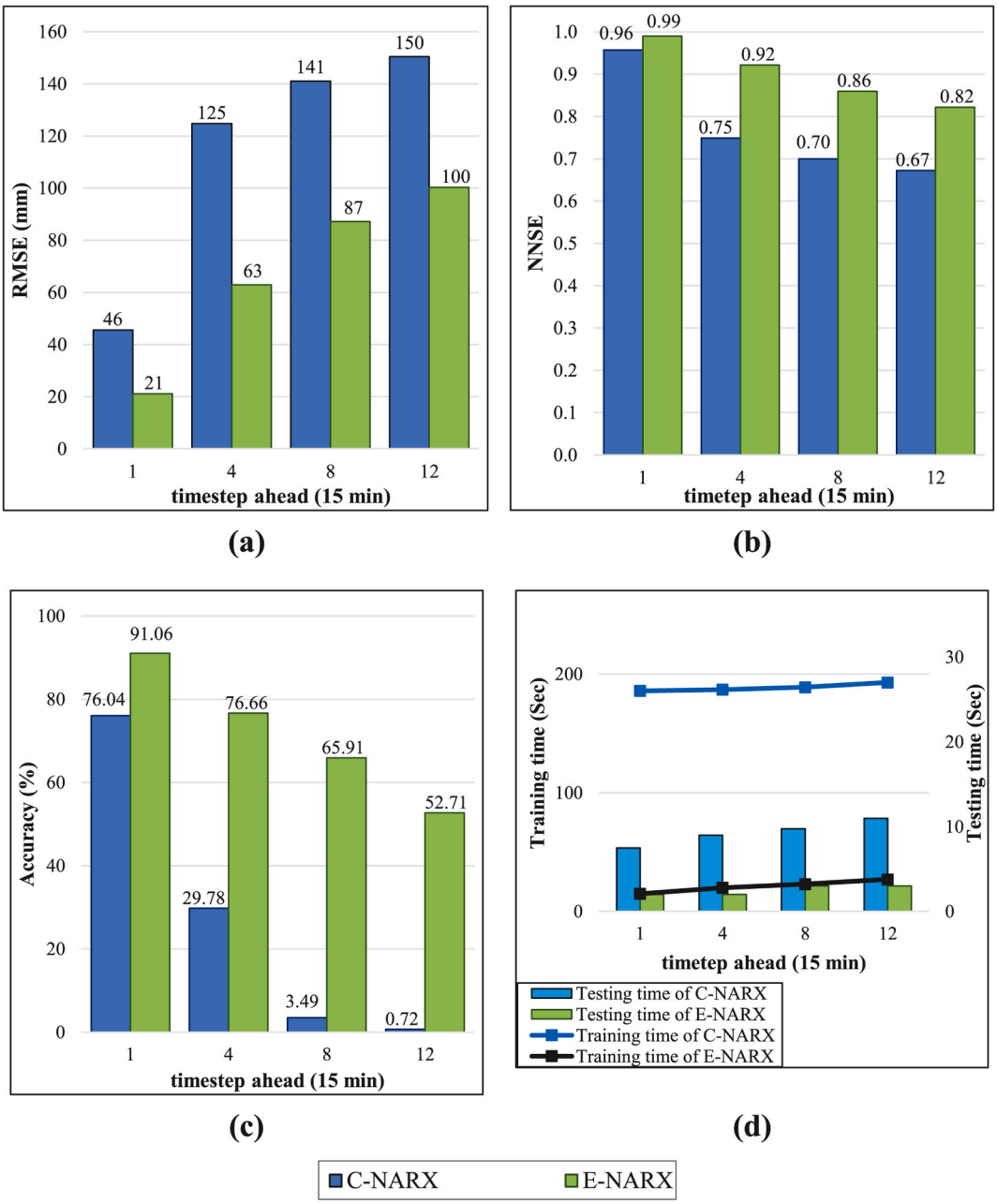


Fig. 9. Performance indicators of the NARX models for (a) RMSE, (b) NNSE, (c) accuracy of flood forecasting, (d) computational time.

3. Results and discussion

The results of the proposed methodology for real-time water level forecasting at the gauging station of a real-world case study are presented for the unseen original test data. For better discussion, it is compared with the conventional NARX model (referred to as C-NARX hereafter), as previously explored by several research works including Nanda et al. (2016) and Abou Rjeily et al. (2018). The C-NARX model is developed using the database comprising all wet and dry periods with the characteristics described for the E-NARX model in the previous section, including training and validation. The performance assessment of water level and flood forecasting is carried out for four lead times namely, the next 15-min timestep ahead (t+1), 1-hr timestep ahead (t+4), 2-hrs timestep ahead (t+8), and 3-hrs timestep ahead (t+12).

3.1. Study area and collected data

The Ruislip urban catchment, as shown in Fig. 7a, which is characterised by a high frequency of fluvial flooding in the Ruislip neighbourhoods, has been selected as a real-world pilot study. This catchment area is located in the London Borough of Hillingdon and drives the Colne catchment surface runoff from south Hertfordshire to a tributary of the River Thames in England. Covering an area of 13 km^2, the catchment area consists mainly of open channels and passes away from densely populated urban areas, with open space parks (see Fig. 7b). The Ruislip gauging station, located at the outlet of the Ruislip UDS in the river Pinn, is responsible for measuring and recording water levels. This gauging station is one of the 55 installed in the Colne catchment area. An ultrasonic IoT-based depth monitor system has been used to record the real-time time-series of water level every 15min at the station since 2009 (DEFRA, 2022). In this case, it is assumed that urban flooding is likely to occur in the Ruislip UDS when the water level at the gauging station

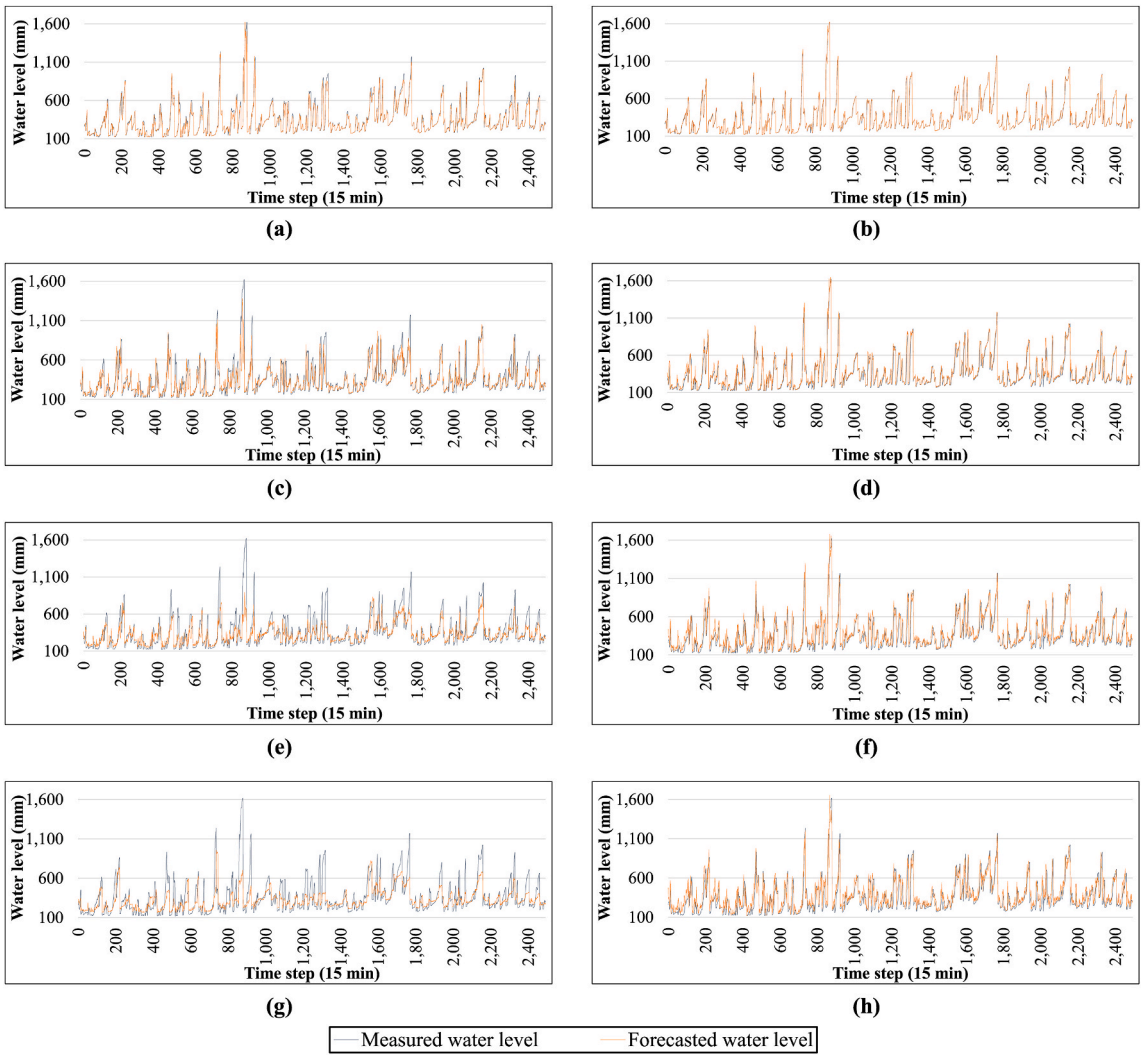


Fig. 10. Comparison of measured with the forecasted water levels by C-NARX model (left) and the E-NARX model (right) for four lead times: (a–b) 1-timestep ahead, (c–d) 4-timestep ahead, (e–f) 8-timestep ahead, (g–h) 12-timestep ahead.

exceeds 850 mm. This threshold value is based on data collected from DEFRA in 2022 and is used to indicate the potential onset of urban flooding in the study area. Rainfall is also measured in the pilot study area using IoT-based tipping bucket method every 15min at Heathrow, Iver Heath, and RAF Northolt rain gauge stations, as shown in Fig. 7a. The selection of these rain gauge stations is based on the direction of the prevailing wind in the pilot study area (i.e., southwest). Among these three rainfall stations, RAF Northolt was selected for model development based on its higher cross-correlation with Ruislip UDS, using the method employed by several research works including Abou Rjeily et al. (2018), and Nanda et al. (2019), as shown in Fig. 7c. The entire database includes 365,233 data samples for both rainfall and water level, with 15-min time intervals and a continuous duration of 12 years (2009–2021), which are accessible through the application programming interface of the UK Environment Agency (DEFRA, 2022).

3.2. Cross-correlation and delays in the dataset creation

Fig. 8 illustrates the trend of improvement and convergence in the coefficients and lag times obtained from the cross-correlation analysis of the generated event-based datasets using the method proposed in section 2.2. This part only demonstrates effect of generating event-based dataset on the rainfall and water level data used only for training and validation processes. Raw data is completely used for testing the framework without any manipulation. Initially, the cross-correlation coefficient between rainfall and water level data of the entire database (including wet and dry periods) was relatively low at 0.30, as also shown in Fig. 7c for the RAF Northolt station. However, the coefficient increased and converged after a few iterations of updating the time-series event-based datasets. The converged coefficients ranged from 0.53 to 0.65 for the four lead times, i.e., 1, 4, 8, and 12 timesteps, corresponding to 15min, 1hr, 2 h s, and 3 h s of forecasting, respectively. Additionally, the lag times were reduced from 10 timesteps in the original database to 4 or 5 timesteps in the converged event-based datasets. These enhancements in the coefficient and reduction in lag time indicate that the event-based dataset generation process can provide better correlation between rainfall and water level data by removing irrelevant or less relevant data from the datasets.

3.3. Model performance of real-time forecasting for target events

The results of the evaluation of the C-NARX and proposed E-NARX models on all target events (S_4 in Fig. 2) are presented in Fig. 9. Additionally, Fig. 10 provides a visual comparison of the forecasted water level of both C-NARX and E-NARX models with the corresponding observations. The E-NARX model exhibits better performance compared to the C-NARX model in all four indicators and lead times, as illustrated in Fig. 9. The models' performance in forecasting one timestep ahead (i.e.,

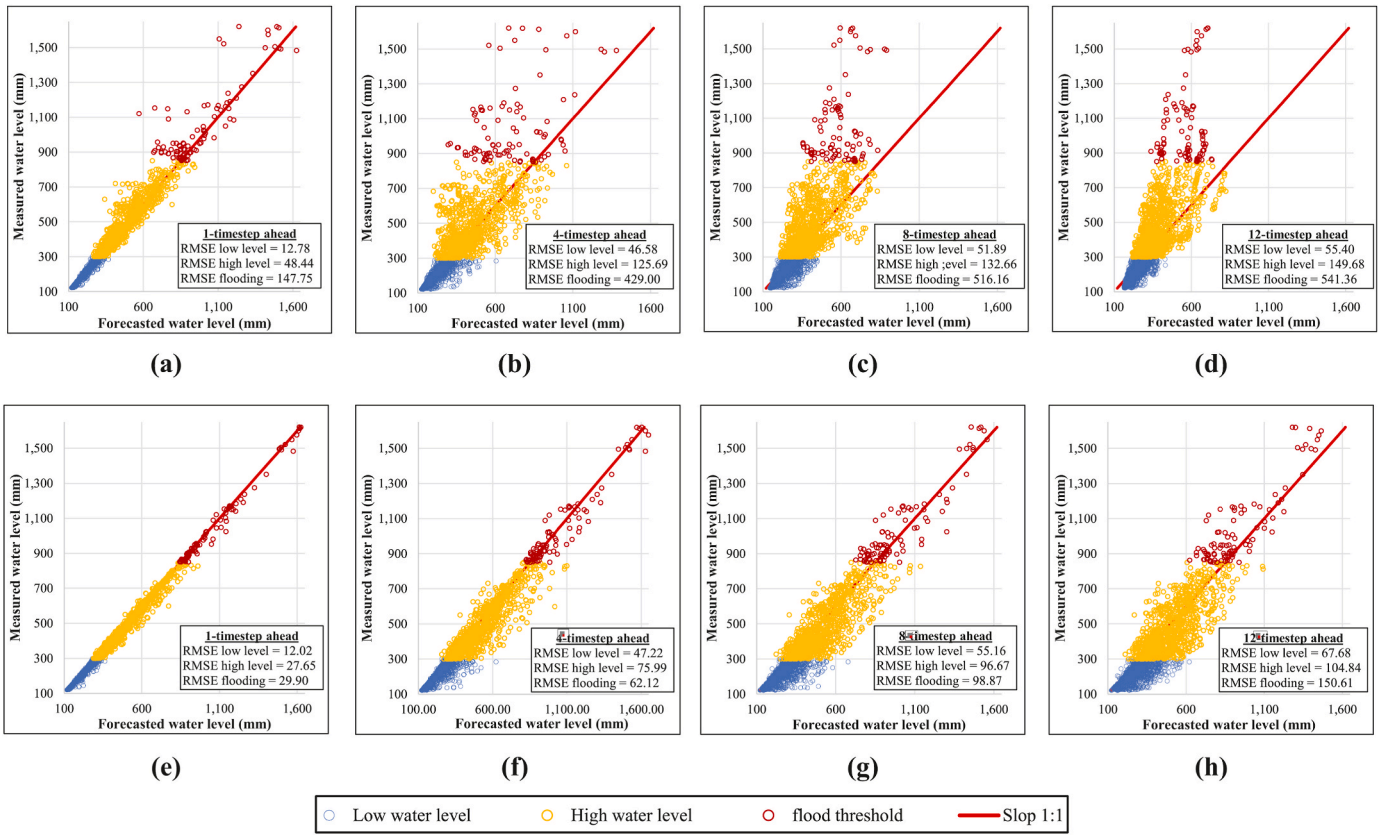


Fig. 11. Scatter plot of forecasted water levels vs corresponding measurements for various time leads and water level classes for (a)–(d) the C-NARX model, and (e)–(h) the E-NARX model.

15-min) is satisfactory with low RMSE, high NNSE, and high accuracy of flood forecasting, although the event-based C-NARX model slightly outperforms the conventional model. However, for longer forecasting horizons (i.e., four or more timesteps), there is a considerable difference between the two models. Specifically, the E-NARX model performance gradually declines, but it still outperforms the C-NARX model significantly.

While different climatic and geographical conditions make it challenging to compare these results with other studies, the C-NARX is relatively confirmed with similar studies developed by NARX models. Several studies have reported the performance of NARX models for flood forecasting. For instance, Nanda et al. (2019) reported NSE values of 0.9, 0.8, and 0.7 for 1, 4, and 8-timestep ahead of forecasting, respectively, which are relatively similar to the results obtained in this study for the corresponding lead times. Abou Rjeily et al. (2018) and Chang et al. (2014) reported NSE values above 0.9 for one 15-min timestep ahead. However, Chang et al. (2014) observed a drop in the NSE performance of their NARX models to 0.67 for 60min forecasting (equivalent to 4-timestep forecasting in this study) and thus limited their work to only 60min lead time. Moreover, Nanda et al. (2016) reported a false alarm ratio of around 30–50% for their improved NARX model for 1 to 3-timestep ahead. While this study measures the accuracy of flood forecasting instead of false alarm ratio, it is worth noting that the developed E-NARX model also shows good performance in terms of false alarm ratio.

The results demonstrate that the C-NARX model is unable to provide accurate forecasts of water levels beyond 8-timestep ahead during heavy rainfall events. In contrast, the event-based model produces relatively acceptable forecasts at all lead times, as indicated by the spikes in Fig. 10. The failure of the C-NARX model is particularly evident for lead times of 8 and 12 timesteps. While the performance indicators of the E-NARX model decrease slightly for longer lead times (4–12 timesteps), they are still significantly better than those of the C-NARX model and fall within acceptable ranges (i.e., 92%–82% for NNSE, 63–100 mm for RMSE, and 77%–53% for accuracy of flood forecasting in Fig. 9). However, the accuracy of flood forecasting for the C-NARX model was deemed unacceptable for longer lead times, particularly for 8 and 12-timesteps ahead.

Besides, The E-NARX model exhibits a much faster computation time than the C-NARX model, taking less than 20 s for each lead time in the test data (as shown in Fig. 9d). The comparison of computational times shows that the E-NARX model can be trained and tested approximately 20 times faster than the C-NARX model. This can be attributed to the framework's ability to remove unnecessary data (i.e., dry periods and non-flood events) from the database, resulting in less data to analyse. While this faster computation time is not surprising, it is a notable advantage, particularly in real-time applications where model retraining time is critical, especially for model development in multiple stations using big data analysis, which would typically be time-consuming.

The performance of the developed NARX models is further analysed by categorising water levels into three main classes: low depth, high depth, and flooding. These classes are defined as follows: low water level corresponds to water levels below 300 mm, high water level corresponds to water levels between 300 mm and 850 mm, and the flood threshold corresponds to water levels above 850 mm. The flood threshold is determined based on the specifications of the UDS conduit, while the values of the other two classes are determined based on a 2-class clustering using the K-nearest neighbourhood (KNN) method, as described by Rahman et al. (2021). Fig. 11 depicts a scatter plot of the model forecasts versus measurements with corresponding RMSE for each class and lead time. For forecasting the low water level in 1 and 4-timestep ahead, the RMSE of both models is relatively similar. However, in the next two lead times (i.e., 8-timestep ahead and 12-timestep ahead), the

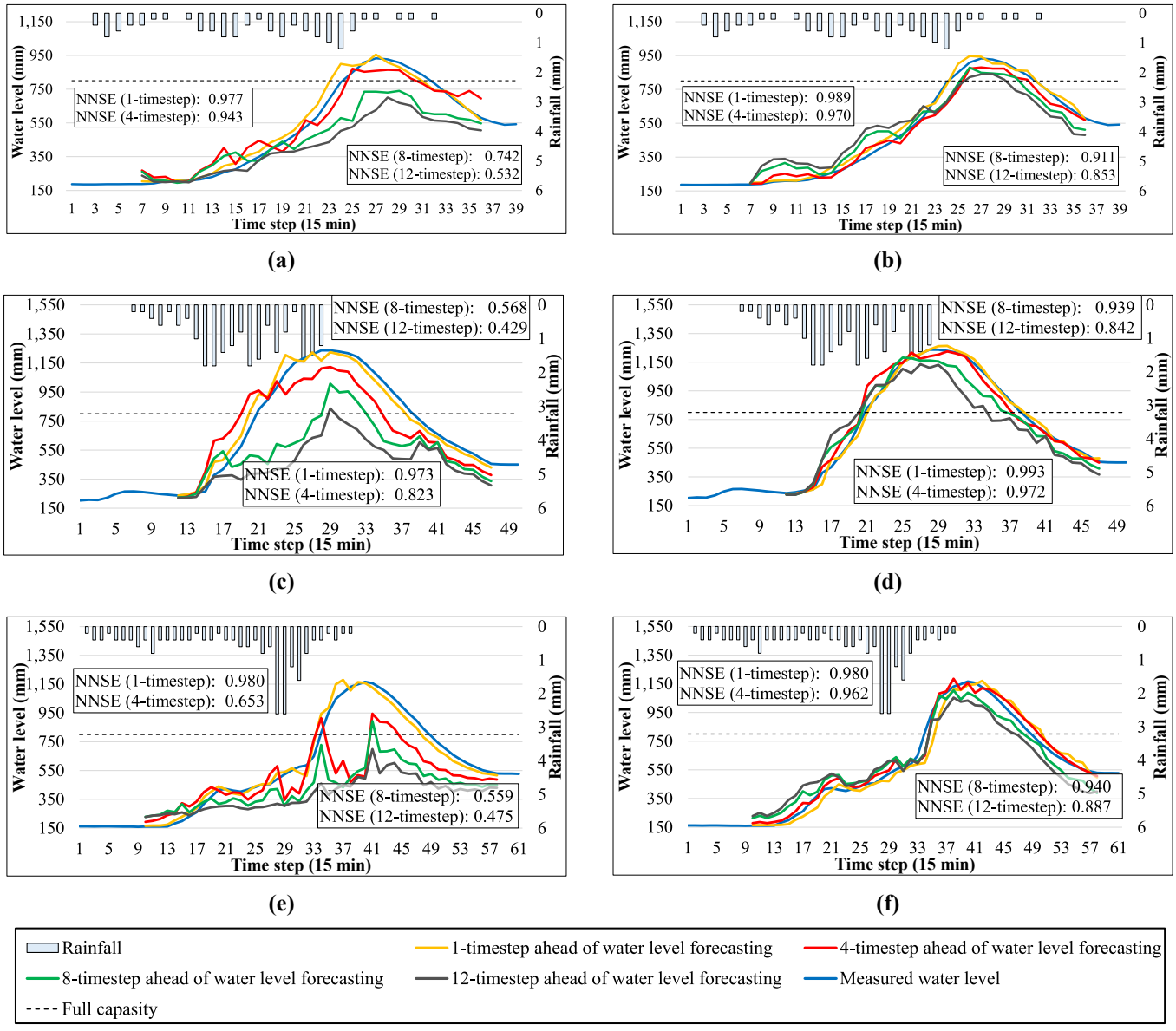


Fig. 12. Water level forecasting of the C-NARX model (left) and the E-NARX model (right) for (a)–(b) event example 1, (c)–(d) event example 2, (e)–(f) event example 3.

C-NARX model outperforms the E-NARX model with a slightly better RMSE (i.e., 51.89 mm versus 55.16 mm and 55.40 versus 67.68 mm, respectively).

In contrast, the RMSE for high depth and flood threshold is significantly worse in the C-NARX model compared to the event-based model. Specifically, the RMSE for flooding in the C-NARX model is almost 10 times worse than the RMSE for low depth in all lead times, while in the E-NARX model, the RMSE for both low depth and flood thresholds is only worsened by a factor of two to three. The poor performance of the C-NARX model in high depth and flood threshold further confirms its unsuitability for moderate to heavy rainfall periods. Moreover, the results indicate that although the E-NARX model performance is affected during higher flood events, the forecasting inaccuracy can be quite tolerable compared to the C-NARX model. This finding can be attributed to the event-based dataset ability to significantly improve the model forecasting performance during heavy rainfalls and flooding.

The aforementioned results are further supported by comparing the scatter plots of model forecasts versus measurements between the two models along the 1:1 line (see Fig. 11). Specifically, for low depth data, the event-based models with larger lead times exhibit slightly more overestimated forecasts. Conversely, the forecasts generated by the C-NARX model are gradually shifted towards the underestimation area from low depth to flooding data and are markedly underestimated, particularly for high depth and flood classes for large lead times (see Fig. 11a–d). As a result, nearly all forecasted C-NARX water levels for flooding data are underestimated. Moreover, the flooding water levels are significantly shifted to this area but with a more uniform pattern. Furthermore, while both NARX models can be used for relatively acceptable performance indicators for the three classes of water level for 1-timestep ahead forecasting, the E-NARX model is more reliable for larger lead times, particularly for flood data with a better distribution of results around the 1:1 line.

The performance of the developed NARX models is further evaluated by investigating their ability to forecast water levels during three selected flood events with distinct hydrological specifications. The results are shown in Fig. 12, and a summary of key hydrological features of the selected events can be found in Table A1 in the appendix. Both models can provide relatively acceptable forecasts when water level

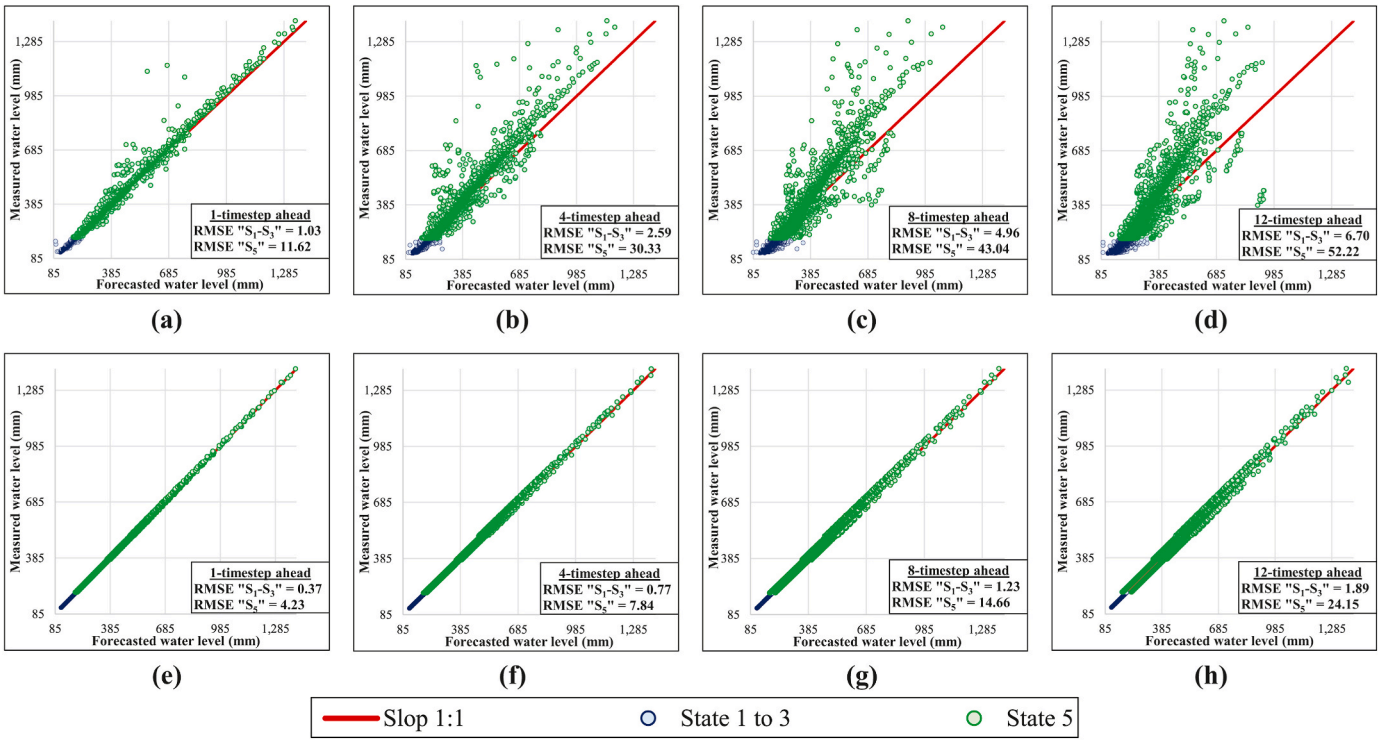


Fig. 13. Scatter plot of forecasted water levels vs corresponding measurements for various lead times and water level classes for (a)–(d) the C-NARX model, and (e)–(h) real-time platform i.e., nonlinear interpolation for S_1 to S_3 and the FFNN for S_5 (See Fig. 2a for defined states).

variation is smooth (i.e., event example 1) or for short lead times (i.e., 1-timestep ahead). However, the C-NARX model fails to provide accurate forecasts for longer lead times, whereas the E-NARX model can provide more reliable forecasts for multi-depth peaks in event example 2 and more accurate forecasts for the sudden rise of water level in event example 3.

However, the event-based model exhibits a minor lag time when tracking the measured water level, which occurs after the first peak (low depth) and before the second peak (high depth) in event example 2. Additionally, the E-NARX model is capable of responding to peak values faster than the C-NARX model, particularly for longer lead times (i.e., 8 and 12-timesteps ahead). The E-NARX model also appears to be more sensitive to slight variations (e.g., event example 2), whereas the C-NARX model seems to have a time-lag issue in responding adequately to incoming storms. These findings suggest that the E-NARX model possesses better memory and cognitive capabilities, enabling it to learn flood events effectively during the training phase and overcome the time-lag issue associated with incoming rainfalls and tracking water levels in the UDS.

3.4. Model performance real-time forecasting for non-flood events

To assess the performance of the water level forecasting models in the UDS during other events (i.e., S_1 to S_3 and S_5 in Fig. 2a), real-time forecasting was conducted for these events and compared with the C-NARX model. The evaluation covers two categories: (1) States S_1 to S_3, which account for 70% of the non-target events data, and (2) depletion events, i.e., S_5, which account for the remaining 30% of the non-target events data. Fig. 13 presents a scatter plot of the forecasted water depth versus the corresponding measurements for both models (i.e., C-NARX and the proposed FFNN) at the Ruislip gauging station. The plots show considerable scatter around the 1:1 slope line (red) for smaller lead times (i.e., 1 and 4 timesteps ahead), although the forecasts are more frequently underestimated, particularly for larger lead times (i.e., 8 and 12 timesteps ahead).

The results also indicate that the RMSE for the depletion events (S_5) varies between 11.62 mm for 1-timestep ahead and 55.22 mm for 12-timesteps ahead in the conventional model, which is much worse than the corresponding values for the proposed model, ranging from only 4.23 mm for 1-timestep ahead to 24.15 mm for 12-timesteps ahead. This demonstrates that the RMSE of the depletion flow is significantly improved by using the pretrained FFNN model, whereas the conventional model is largely impacted by other non-flood events. In longer lead times, the conventional model tends to underestimate measured water levels because it is trained on a large number of S_1 to S_3 events in which the water level is less than 200 mm. Alternatively, the real-time platform could perfectly forecast S_1 to S_3 by simply using non-linear interpolation, whereas the FFNN model can forecast more complicated situations such as the depletion state.

3.5. Sensitivity analysis of real-time platform

The performance of the proposed real-time platform was previously analysed based on different states of water level. To gain a better understanding of the hyperparameters used for the models, sensitivity analysis was conducted, and the results are shown in Fig. 14. The lag times used for rainfall data (based on the proposed method in Section 2.1.2) and water level data (recommended by literature described in section 2.2) were found to be the best parameters, as seen in Fig. 14a and b. Although the delay factor showed more flexibility and less sensitivity to variation, the feedback increasing resulted in an increase in RMSE immediately. This increase was observed for longer timesteps ahead of forecasting, mainly because adding unnecessary input data in the form of time-series data during the training process may mislead the model in understanding the true nature of both target and depletion events, particularly when the model tries to forecast the water level for longer timesteps.

Furthermore, as previously mentioned, the developed ML models were structured with 2 5-node hidden layers based on recommendations from similar research studies (see Section 2.2 and Abou Rjeily et al.,

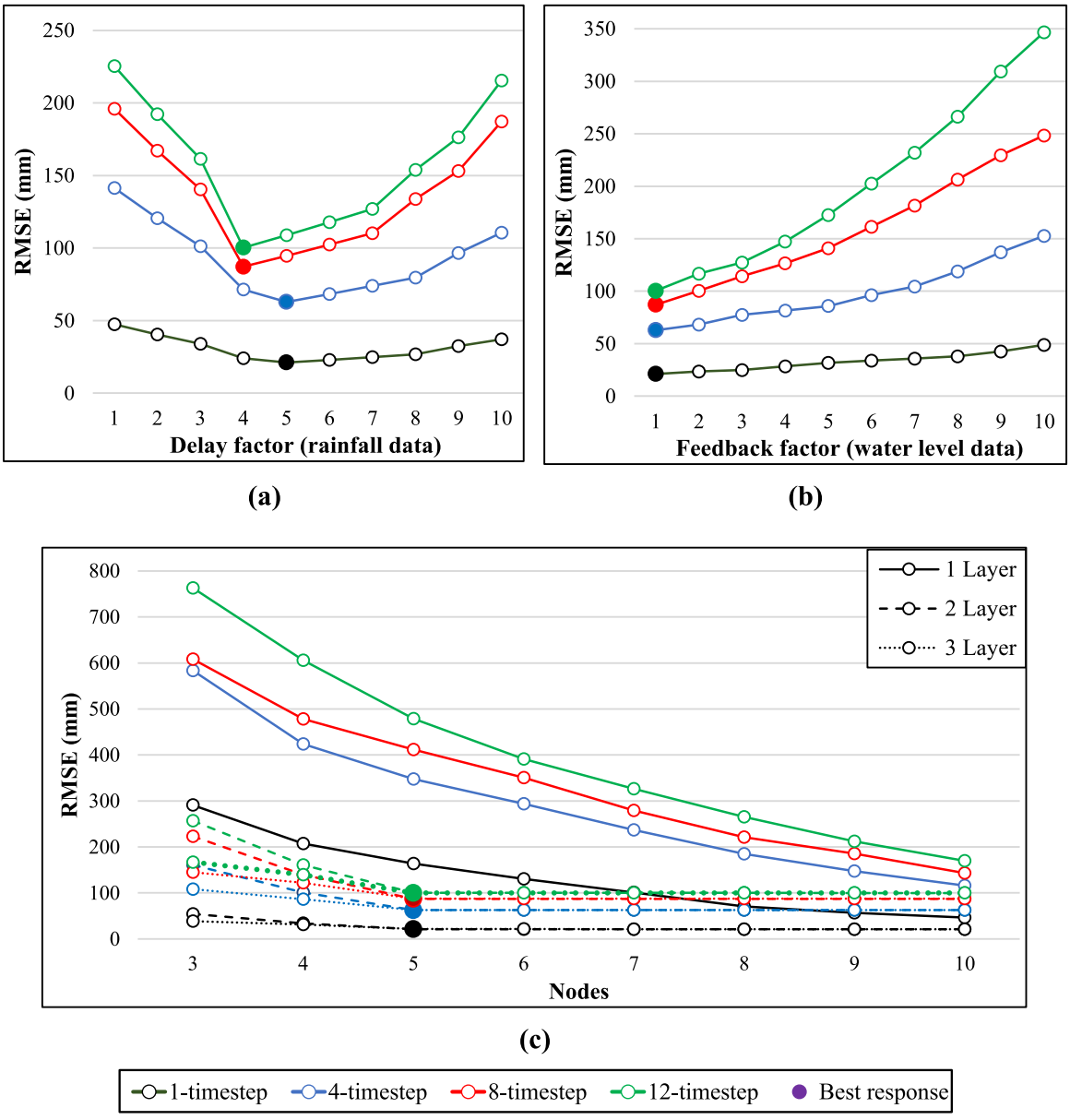


Fig. 14. Sensitivity analysis on hyperparameters of model tuning (a) delay factors, (b) feedback factors, (c) layers and nodes (See Fig. 4 for defined parameters).

2018; Nanda et al., 2019 for reference). However, these numbers were tested as shown in Fig. 14c. While increasing layers and nodes is expected to increase the efficiency of the models, it was observed that RMSE did not significantly increase (less than 5% for the best-performed model) with more complex models beyond 2 layers and 5 nodes for each layer. When comparing this result with the computational time listed for all developed models in Table A2 in the Appendix, it is evident that computation time can increase up to 20% for the best-performed model if the model is structured with more layers than proposed, while RMSE decrease significantly. This can be important for a specifically real-time early warning systems which may need to retrain regularly to keep themselves up to date. Although training time for the model is negligible in this case study, as a general concept, this point should be considered for similar research studies.

To assess the impact of the dataset used for the training processes of the proposed model, multiple models were trained and tested with five different initial seeds. The results of these experiments are presented in Figure A1. The findings indicate that there is less than a 5% difference in both RMS and NNSE between the model with the minimum and maximum performance, respectively. This suggests that the proposed model is robust and is not highly dependent on the initial random seed used for the training process. This is a desirable characteristic for ML models as it indicates that the model is able to generalise well to new data and is not overly sensitive to small variations in the training process.

The models were also evaluated for different dataset sizes to assess their sensitivity to the size of the dataset. As depicted in Fig. 15, the models were able to be trained with a reduced size of the dataset of up to 75% of the total available training dataset while maintaining an RMSE above 90%. In other words, the models were able to perform well, maintain their accuracy with this range of training dataset size, and resist to be impacted of dataset reduction. However, it was observed that with further reduction in dataset size, the models adapted themselves and their accuracy decreased in a near-linear behavior. Finally, for datasets less than 35%, the models completely failed to perform (yielded to the dataset reduction), and the RMSE increased significantly. This finding is significant because although the models were developed in a relatively simple structure, such savings can be useful when applied to

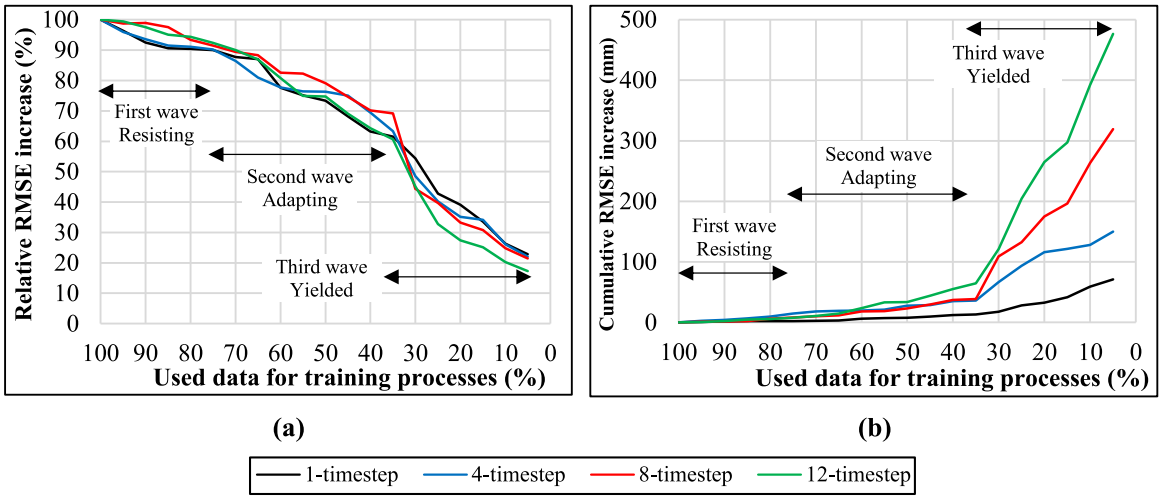


Fig. 15. Uncertainty analysis of performed models based on size of training data (a) cumulative RMSE increase, (b) Relative RMSE loss.

wider applications such as UDS networks. This can conserve energy costs and reduce wasted time for constantly retraining in the context of early warning systems.

The models were also evaluated for different rainfall intensity and duration, as presented in Fig. 16. To better understand the results, all rainfall events were classified into four groups using a 2-class clustering KNN method (as described by Rahman et al., 2021): (group 1) low intensity and short duration, (group 2) low intensity but long duration, (group 3) high intensity and short duration, and (group 4) high intensity and long duration. The results indicate that the relative error (RE) varies differently for each group (Fig. 16a vs 16j). Specifically, the RE of group 1 uniformly increases from 1-timestep to 12-timestep, while this parameter dispersed for group 3. This suggests that the model struggles to understand the nature of sudden rainfall (light blue dots in Fig. 16j) where high-intensity rainfall occurs during a short duration, resulting in imperfect forecasts of associated water level rising. In contrast, for group 3, the model shows more stability mainly due to the long duration which provides the model with the opportunity to comprehend the intensity of the rainfall. Even for higher intensity rainfall, the associated water level rising is distributed gradually during a long duration (yellow dots in Fig. 16j), resulting in the model's improved understanding of the rainfall event. This finding is further confirmed when the second group of rainfall is observed. While the RE was initially shaped as a horizontal cylinder (red dots in Fig. 16a), it gradually shifted to a triangular shape in Fig. 16j, where again the model had more RE for rainfall events with shorter duration. Therefore, it seems that the duration of rainfall has a greater impact on model accuracy compared to rainfall intensity.

To expand on the previous point, heatmap figures in Fig. 16 demonstrate the impact of different characteristics of rainfall on the performance of the model. The figures again approved that the RE of the To elaborate on the previous point, the heatmap figures in Fig. 16 show how different rainfall characteristics affect the model's performance. The figures confirm that the model's error (RE) increases with shorter rainfall duration. More specifically, when looking at rainfall intensity (Fig. 16b vs 16k), the RE growth mostly expands vertically, whereas for rainfall duration (Fig. 16c vs 16l), the RE growth rate shrinks horizontally. These observations suggest that the model is more sensitive to different rainfall durations, as the density of the RE rate has a lower growth rate for longer rainfall durations. However, the model may not be perfectly sensitive to low-intensity rainfall events, particularly for longer forecast horizons such as 12-timesteps ahead, as seen in the reduction of the RE for high-intensity rainfall events. Moreover, the reduction of the RE for rainfall events with longer durations indicates that the model may be more sensitive to rainfall events with shorter durations. These findings offer valuable insights into the limitations of the model, which can help in developing better forecasting models for flood risk management and early warning systems.

4. Conclusions

The paper presents a new RTUFF framework that utilises an event identification method, event-based generated dataset, and event-based ML-based forecasting. The proposed framework in this study focuses on forecasting water level rise in UDS with greater accuracy than the C-NARX model, particularly for longer lead times of forecasting exceeding 120min. A novel approach is presented for flood event identification, which distinguishes between different events in the dataset and generates associated event-based datasets that can significantly enhance the performance of the introduced decision tree-based real-time flood forecasting platform, compared to the widely used conventional NARX model. The following key findings were noted from the pilot study:

- The proposed models generate different datasets based on different identified events for event-based model training, improving the accuracy of flood forecasting in target events.
- The E-NARX model significantly increases the accuracy of flood forecasting, with an accuracy of 91% for one timestep ahead of forecasting (15min) and maintained an acceptable accuracy of 77% for 1hr ahead and 53% for 3 hrs later.
- The FFNN model outperforms the C-NARX model, especially for depletion events. C-NARX suffers from the majority of underestimated forecasts, whereas RMSE dropped 50% and 60% for the longest timestep (i.e., 3 hrs ahead) of dry weather and depletion events, respectively.
- The proposed model significantly improves all KPIs, especially for larger lead times of 2 hrs–3 hrs ahead (i.e., 92%–82% for NNSE, 63–100 mm for RMSE, and 77%–53% for accuracy of flood forecasting and computational time).
- The high performance of the proposed model in forecasting both high depth and flood threshold makes it appropriate for moderate to heavy rainfalls, while the accuracy of flood forecasting in the conventional model was almost unacceptable for longer lead times.
- model is more sensitive to rainfall duration than rainfall intensity. Specifically, the model appears to be more sensitive to low intensity rainfall events with shorter durations, which are considered to be critical types of target events.

The proposed platform enables the development of a faster and more accurate flood forecasting model for UDS, which can be trained based on a smaller yet more accurate dataset. The proposed real-time flood

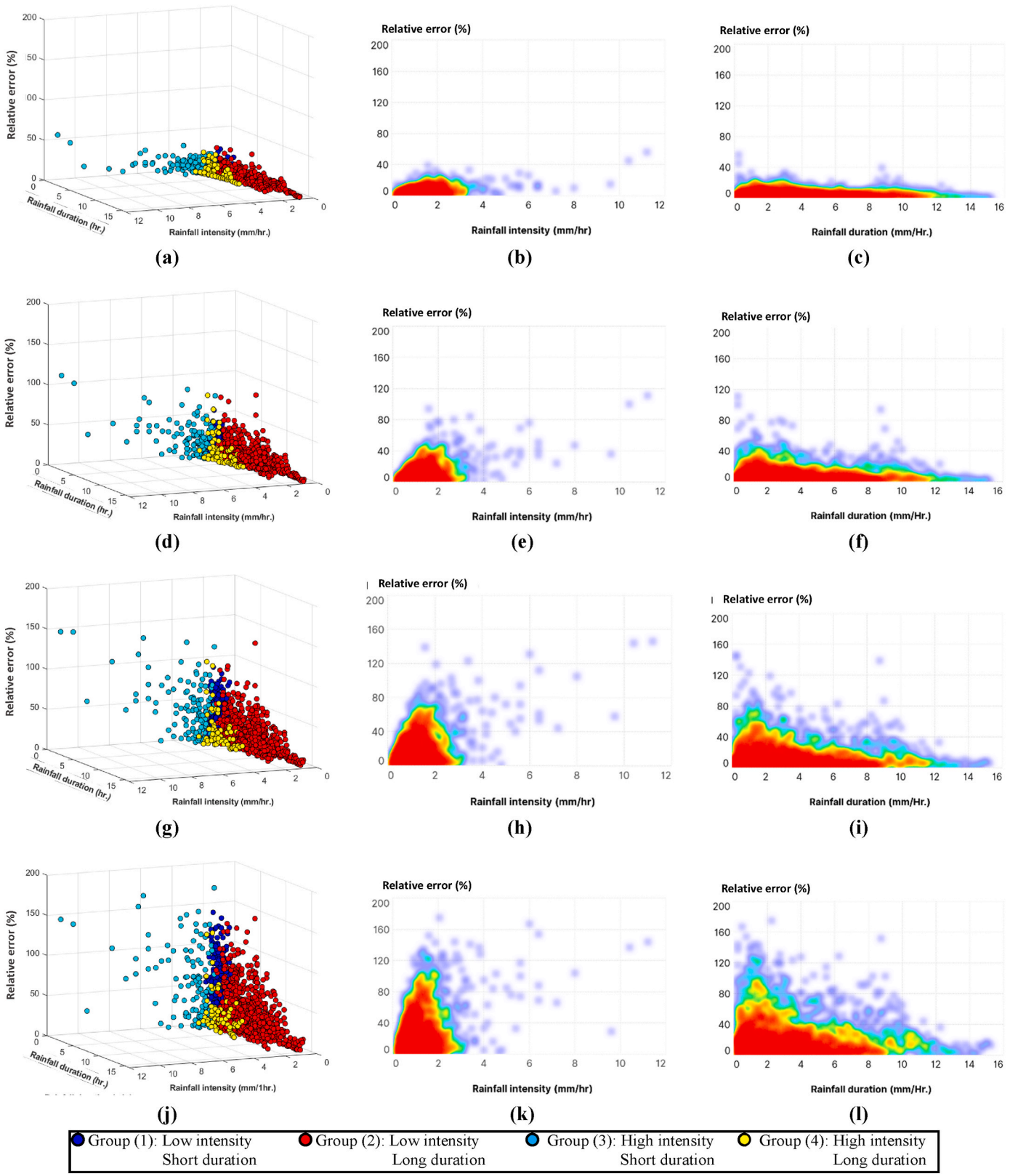


Fig. 16. Relative error based on rainfall intensity and duration: (Left) scatter plot of test data, (middle) heatmap of rainfall intensity, (Right) heatmap of rainfall duration, (a–c) 1-timestep ahead, (d–f) 4-timestep ahead, (g–i) 8-timestep ahead, (j–l) 12-timestep ahead.

forecasting framework successfully distinguishes between different events, such as dry weather flow, non-effective rainfall events, target events, and depletion events, and provides accurate estimations of water levels in UDS based on non-linear regression, using the FFNN and E-NARX models. The proposed platform is defined as a simple yet more efficient system, and the results demonstrate the advantages of using event-based data pre-processing and multiple strategies in utilising ML-based models. The proposed event-based dataset framework's higher

accuracy is of paramount importance for early warning systems, as false alarms and missed flooding events can have significant impacts on stakeholders, especially the public, local authorities, and relevant emergency response systems. False alarms may undermine confidence in the early warning system and lead to reduced responsiveness to future warnings, whereas missed flooding events may cause an ineffective response, leading to a loss of trust between the public and local governments. The proposed event-based dataset framework could accurately forecast water level rise and flooding in the UDS in real-time, with longer lead times exceeding 2 h s compared to currently available models. This can significantly enhance the accuracy of flood forecasting, decrease the occurrences of both false alarms and missed events, and improve emergency response systems.

Although the proposed real-time flood forecasting platform has shown promising results in comparison to other models, further validations by other time-series RTUFF models are required. Moreover, it is worth noting that this model was tested solely on the UDS of the pilot study, and its applicability to other settings, such as river basins and reservoir basins, would need further validation. Additionally, while the proposed real-time flood forecasting framework demonstrated high accuracy, further improvements in KPIs for longer lead times and also Rainfall events characterised by low intensity and short duration. This could involve adjusting or removing unnecessary data in the dataset during model development or incorporating more sophisticated data mining techniques to improve the data-driven model.

Declaration of generative AI in scientific writing

During the preparation of this work the authors used ChatGPT to improve readability and language of the text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work is supported by the PhD scholarship allocated to the first author and the Fellowship allocated to the second author. The authors wish to acknowledge the PhD Vice Chancellor Scholarship supported by the University of West London and the Fellowship supported by the Royal Academy of Engineering under the Leverhulme Trust Research Fellowships scheme. The authors also wish to thank the editor and the three anonymous reviewers for making constructive comments which substantially improved the quality of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2023.105772.

References

Abou Rjeily, Y., Abbas, O., Sadek, M., Shahrour, I., Chehade, F., 2018. Model predictive control for optimising the operation of urban drainage systems. J. Hydrol. 566, 558–565.

Adikari, K., Shrestha, S., Ratnayake, D., Budhathoki, A., Mohanasundaram, S., Dailey, M., 2021. Evaluation of artificial intelligence models for flood and drought forecasting in arid and tropical regions. Environ. Model. Software 144, 105136.

Ahmed, A., Deo, R., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., Yang, L., 2021. Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. J. Hydrol. 599, 126350.

Alizadeh, B., Bafti, A., Kamangir, H., Zhang, Y., Wright, D., Franz, K., 2021. A novel attention-based LSTM cell post-processor coupled with bayesian optimization for streamflow prediction. J. Hydrol. 601, 126526.

Ben Aissia, M., Chebana, F., Bouarda, T., 2017. Multivariate missing data in hydrology – review and applications. Adv. Water Resour. 110, 299–309.

Butler, D., Digman, C., Makropoulos, C., Davies, J., 2018. Urban Drainage, fourth ed. CRC Press, Boca Raton: USA, pp. 78–89.

Chang, F., Chen, P., Lu, Y., Huang, E., Chang, K., 2014. Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control. J. Hydrol. 517, 836–846.

Chen, M., Papadikis, K., Jun, C., Macdonald, N., 2023. Linear, nonlinear, parametric and nonparametric regression models for nonstationary flood frequency analysis. J. Hydrol. 616, 128772.

Chen, S., Garambois, P.-A., Finaud-Guyot, P., Dellinger, G., Mosé, R., Terfous, A., Ghenaim, A., 2018. Variance based sensitivity analysis of 1D and 2D hydraulic models: an experimental urban flood case. Environ. Model. Software 109, 167–181.

Darabi, H., Haghighi, A., Mohamadi, M., Rashidpour, M., Ziegler, A., Hekmatzadeh, A., Kløve, B., 2020. Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. Nord. Hydrol 51 (1), 127–142.

Department for Environment, 2022. Food and Rural Affairs. DEFRA). DEFRA Official Website [Online] Available at: Environment.data.gov.uk. (Accessed 12 January 2022).

Garofalo, G., Giordano, A., Piro, P., Spezzano, G., Vinci, A., 2017. A distributed real-time approach for mitigating CSO and flooding in urban drainage systems. J. Netw. Comput. Appl. 78, 30–42.

Hamil, L., 2011. Understanding Hydraulics, third ed. Macmillan Education, London, pp. 507–585.

Huang, X., Li, Y., Tian, Z., Ye, Q., Ke, Q., Fan, D., Mao, G., Chen, A., Liu, J., 2021. Evaluation of short-term streamflow prediction methods in urban river basins. Phys. Chem. Earth 123, 103027.

Jones, S., Le Quéré, C., Rödenbeck, C., Manning, A., Olsen, A., 2015. A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data. J. Adv. Model. Earth Syst. 7 (4), 1554–1557.

Li, H., Zhang, C., Chen, M., Shen, D., Niu, Y., 2023. Data-driven surrogate modeling: introducing spatial lag to consider spatial autocorrelation of flooding within urban drainage systems. Environ. Model. Software 161, 105623.

Lin, Y., Wang, D., Wang, G., Qiu, J., Long, K., Du, Y., Xie, H., Wei, Z., Shangguan, W., Dai, Y., 2021. A hybrid deep learning algorithm and its application to streamflow prediction. J. Hydrol. 601, 126636.

Luo, P., Luo, M., Li, F., Qi, X., Huo, A., Wang, Z., He, B., Takara, K., Nover, D., Wang, Y., 2022. Urban flood numerical simulation: research, methods and future perspectives. Environ. Model. Software 156, 105478.

Lv, N., Liang, X., Chen, C., Zhou, Y., Li, J., Wei, H., Wang, H., 2020. A long Short-Term memory cyclic model with mutual information for hydrology forecasting: a Case study in the xixian basin. Adv. Water Resour. 141, 103622.

Meles, M., Goodrich, D., Gupta, H., Burns, I., Unkrich, C., Razavi, S., Guertin, D., 2021. Multi-criteria, time dependent sensitivity analysis of an event-oriented, physically-based, distributed sediment and runoff model. J. Hydrol. 598, 126268.

Mounce, S., Shepherd, W., Sailor, G., Shucksmith, J., Saul, A., 2014. Predicting combined sewer overflows chamber depth using artificial neural networks with rainfall radar data. Water Sci. Technol. 69 (6), 1326–1333.

Nanda, T., Sahoo, B., Beria, H., Chatterjee, C., 2016. A wavelet-based non-linear autoregressive with exogenous inputs (WNARX) dynamic neural network model for real-time flood forecasting using satellite-based rainfall products. J. Hydrol. 539, 57–73.

Nanda, T., Sahoo, B., Chatterjee, C., 2019. Enhancing real-time streamflow forecasts with wavelet-neural network based error-updating schemes and ECMWF meteorological predictions in Variable Infiltration Capacity model. J. Hydrol. 575, 890–910.

Piadeh, F., Behzadian, K., Alani, A.M., 2021. The Role of Event Identification in Translating Performance Assessment of Time-Series Urban Flood Forecasting [Online] available at: repository.uwl.ac.uk/id/eprint/9113/3. UWL annual conference, London UK. (Accessed 21 February 2023).

Piadeh, F., Behzadian, K., Alani, A.M., 2022a. A critical review of real-time modelling of flood forecasting in urban drainage systems. J. Hydrol., 127476

Piadeh, F., Behzadian, K., Alani, A.M., 2022b. Multi-step flood forecasting in urban drainage systems using time-series data mining techniques. West Indies: Trinidad and Tobago, [Online] available at: repository.uwl.ac.uk/id/eprint/9690/1. In: Water Efficiency Conference. (Accessed 21 February 2023).

Piadeh, F., Behzadian, K., Chen, A., Campos, L., Rizzuto, J., 2023. Real-time flood overflow forecasting in Urban Drainage Systems by using time-series multi-stacking of data mining techniques. EGU General Assembly 2023. https://doi.org/10.5194/egusphere-egu23-8574. Vienna, Austria, [Online] available at: (Accessed 8 May 2023).

Rahman, M., Chen, N., Elbeltagi, A., Islam, M., Alam, M., Pourghasemi, H., Tao, W., Zhang, J., Shufeng, T., Faiz, H., Baig, M., Dewan, A., 2021. Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. J. Environ. Manag. 295, 113086.

Razavi, S., 2021. Deep learning explained: fundamentals, explainability, and bridgeability to process-based modelling. Environ. Model. Software 144, 105159.

Razavi, S., Hannah, D., Elshorbagy, A., Kumar, S., Marshall, L., Solomatine, D., Dezfuli, A., Sadegh, M., Famiglietti, J., 2022. Coevolution of machine learning and

process-based modelling to revolutionize Earth and environmental sciences. A perspective. Hydrological Processes 36 (6), e14596.
Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H., 2021. The Future of Sensitivity Analysis: an essential discipline for systems modeling and policy support. Environ. Model. Software 137, 104954.
Saltelli, A., Jakeman, A., Razavi, S., Wu, Q., 2021. Sensitivity analysis: a discipline coming of age. Environ. Model. Software 146, 105226.
Scottish Environment Protection Agency (SEPA), 2021. Flood Risk Management Glossary [Online] Available at: consultation.sepa.org.uk. (Accessed 18 April 2022).
Szeląg, B., Suligowski, R., De Paola, F., Siwicki, P., Majerek, D., Łagód, G., 2022. Influence of urban catchment characteristics and rainfall origins on the phenomenon of stormwater flooding: case study. Environ. Model. Software 150, 105335.
Xie, J., Chen, H., Liao, Z., Gu, X., Zhu, D., Zhang, J., 2017. An integrated assessment of urban flooding mitigation strategies for robust decision making. Environ. Model. Software 95, 143–155.
Zakaria, M., Malek, M., Zolkepli, M., Ahmed, A., 2021. Application of artificial intelligence algorithms for hourly river level forecast: a case study of Muda River, Malaysia. Alex. Eng. J. 60 (4), 4015–4028.
Zhang, D., Lindholm, G., Ratnaweera, H., 2018. Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring. J. Hydrol. 556, 409–418.