



Enhancing Flooding Depth Forecasting Accuracy in an Urban Area Using a Novel Trend Forecasting Method

Song-Yue Yang¹ · You-Da Jhong² · Bing-Chen Jhong³ · Yun-Yang Lin¹

Received: 6 August 2023 / Accepted: 27 December 2023 / Published online: 10 January 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Accurate flood runoff and water level predictions are crucial research topics due to their significance for early warning systems, particularly in improving peak flood level forecasts and reducing time lags. This study proposes a novel method, Trend Forecasting Method (TFM), to improve model accuracy and overcome the time lag problem due to data scarcity. The proposed method includes the following steps: (1) select appropriate input factors causing flood events, (2) determine the most suitable AI method as the basis for forecasting models, (3) a forecasting model using a multi-step-ahead approach and a forecasting model with variation in flood depth as input are developed as compared to the selected model in Step 2, and (4) according to the rising limb and falling limb of a flood hydrograph, the maximum and minimum values predicted by the models above are respectively selected as the final outputs. The application to demonstrate the advantages of the proposed method was conducted in the Annan District of Tainan City, Taiwan. Of all the models tested, the Gated Recurrent Unit (GRU) demonstrated superior accuracy in forecasting flood depths, followed by Long Short-Term Memory (LSTM) and Bidirectional LSTM, with the Back Propagation Neural Network falling behind. With a Nash–Sutcliffe efficiency coefficient (NSE) of 0.56 for the next hour's forecast, the GRU model's structure appears particularly fitting for flood depth forecast. However, all four models showed time lag issues. TFM substantially enhanced the GRU model's forecast accuracy, mitigating the time lag. TFM achieved an NSE of 0.82 for forecasting 10-, 20-, 30-, 40-, 50-, and 60-min lead time. The observed flood depths had a 68% probability of consistent rise or fall, validating TFM's underlying hypothesis. Furthermore, including an autoregressive model in TFM reduced the time lag problem.

Keywords BPNN · LSTM · GRU · BiLSTM · Flooding depth

✉ Bing-Chen Jhong
jhongbc0516@gmail.com

¹ Department of Urban Planning and Spatial Information, Feng Chia University, Taichung 407102, Taiwan

² Department of Water Resources Engineering and Conservation, Feng Chia University, Taichung 407102, Taiwan

³ Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taipei 106335, Taiwan

1 Introduction

Numerous coastal cities worldwide grapple with water-related issues due to the interplay between natural and human factors. These issues range from water scarcity and ground subsidence to seawater intrusion into aquifers, coastal erosion, and flooding (Schuetze and Chelleri 2013). One hundred thirty-six largest coastal cities in the world are particularly vulnerable due to a combination of factors, such as climate change, rising sea levels, subsidence from low-lying terrain, ongoing urbanization (Du et al. 2015), high-value assets, and bustling economic activities (Meng et al. 2019). On the other hand, another crucial water-related issue is flooding hazards. Cities with high population density and economic significance are particularly vulnerable to flood disasters, often stemming from anthropogenic factors such as decreased urban permeability and rapid runoff, coupled with the effects of global warming and extreme weather events (Chakraborty et al. 2023; Roy et al. 2020). Therefore, addressing such events' increasing frequency and intensity is crucial to safeguarding cities from future threats (Chowdhuri et al. 2020; Imteaz and Hossain 2023; Ruidas et al. 2022, 2023).

Flood hazard simulations play a crucial role during intense rainfall to address the issues mentioned above, as they aid decision-makers in understanding such disasters. Traditionally, numerous studies have employed physically-based numerical models to simulate the extent of flooding at different temporal and spatial scales during storm events. Applying such models offers several advantages. For instance, they accurately simulate flood depths based on initial and boundary conditions. The output values of these models are rooted in physics and mathematical theories, imparting a certain degree of reference value. However, they also come with several drawbacks, requiring substantial computational time, extensive data storage, and effective management, mainly when focused on flood forecasting rather than mere simulation. Several studies have indicated that flood prediction is crucial in enhancing the effectiveness of early warning systems (e.g., Plate 2007).

A significant advancement is the application of AI technology to urban flood prediction. These methods consider the results of two-dimensional (2D) hydraulic simulations as a database used for training and testing AI models. Compared to the computational resources the 2D hydraulic model demands, the trained AI requires significantly fewer resources and can deliver results rapidly. These characteristics make it well-suited for real-time flood prediction in flood prevention systems (Hofmann and Schüttrumpf 2021). Accurate forecasts provide ample time for both agencies and the public to respond. With the advent of artificial intelligence (AI), a variety of machine learning and deep learning methods for predicting flood runoff and water levels have been proposed in recent years (Sit et al. 2020; Sun et al. 2020). These include Support Vector Machines (Jhong et al. 2017), Back Propagation Neural Network (BPNN) (Berkhahn et al. 2019; Chu et al. 2020), stacked autoencoder with Recurrent Neural Networks (RNN) (Kao et al. 2021), Long Short-Term Memory (LSTM) (Nearing et al. 2022), and Convolutional Neural Network (CNN) (Guo et al. 2021; Hosseiny 2021).

Establishing complex and non-linear interdependencies among various hydrological variables has always been a significant research focus in flood forecasting. For instance, utilizing artificial intelligence (AI) to model the non-linear relationships between input variables (such as precipitation, temperature, and evapotranspiration) and output variables like river discharge or water level enables AI-based forecasting models to predict the future extent of flooding accurately. The BPNN is a supervised learning algorithm that refines the network's performance by adjusting the weights of the input, hidden, and output layers. BPNN has been widely

used in hydrological modeling, particularly for forecasting river discharges and water levels. LSTM, a unique variant of RNN, was introduced by Hochreiter and Schmidhuber (1997) specifically to circumvent the vanishing gradient problem. It has found extensive use in water resources management, including applications in rainfall-runoff simulation (Cui et al. 2021; Kratzert et al. 2018), probabilistic streamflow forecasting (Zhu et al. 2020), flood level forecasting (Dazzi et al. 2021), combined sewer overflow monitoring (Palmitessa et al. 2021), and flooding depth (Yang et al. 2023). While LSTM is extensively utilized, its complexity serves as a significant drawback. Gated Recurrent Unit (GRU) was introduced by Cho et al. (2014), effectively simplifying the LSTM model. Unlike LSTM, which has three gates in each module, GRU only has two: the reset and update gates. Numerous researchers have employed GRU for predicting river flow. The Bidirectional LSTM (BiLSTM), a derivative of the bidirectional Elman neural network (Graves and Schmidhuber 2005), establishes both forward and backward hidden layers. The input layer feeds into these forward and backward hidden layers, which jointly compute the predicted value. Kang et al. (2020) utilized BiLSTM to predict urban wastewater flow. However, despite the emphasis in the literature on applying the AI above methods to various hydrological forecasting studies, the accuracy of most flood prediction models remains limited. Another critical issue is how to further enhance the accuracy of flood prediction models based on the characteristics of flood hydrographs.

Incorporating additional input factors, in addition to rainfall and flood depth variables, is crucial to enhance the accuracy of flood prediction models. Investigating their impact is also essential for improving the performance of these models. In this study, a novel method, the Trade Forecasting Method (TFM), is proposed to improve the accuracy of forecasting models and solve time lag problems. The Annan District of Tainan City, Taiwan, was selected as the study area due to its low-lying terrain, which often experiences flooding during typhoons and heavy rains. This study explored two issues: (1) Compare the accuracy of BPNN, LSTM, GRU, and BiLSTM in forecasting flood depth and (2) Discuss how much our proposed TFM could improve the accuracy of model forecasting. The proposed model can be used for urban flood forecasting. In this paper, the first chapter serves as an introduction. The second chapter outlines the methodology, providing explanations of algorithms and the proposed method. The third chapter details the study area, data used, and model development; the fourth, fifth, and concluding chapters present results, discussions, and conclusions.

2 Methodology

2.1 Back Propagation Neural Network

A BPNN is an artificial neural network that employs a supervised learning algorithm called backpropagation for training the network (Najafabadipour et al. 2022). BPNN consists of an input layer, one or more hidden layers, and an output layer. Each layer comprises interconnected neurons (nodes or units) that process and transfer information. The network learns by adjusting the weights of the connections between neurons to minimize the error between the forecasted outputs and the actual target outputs. The net input (net_j) is calculated for each node in the hidden layer using the formula:

$$net_j = \sum w_{ij} \cdot x_i + b_j \quad (1)$$

where w_{ij} is the weight from the input node i to the hidden node j , x_i is the input value, and b_j is the bias for the hidden node j . An activation function σ is applied to net_j to get the output (y_j) from the hidden nodes:

$$y_j = \sigma(net_j) \quad (2)$$

2.2 Long Short-Term Memory

The LSTM model comprises a forget gate, an input gate, and an output gate, each serving distinct functions (Kao et al. 2020). LSTM networks are constructed from memory blocks, also known as cells. The cell state and hidden state are propagated to the subsequent cell. Initially, the current input at time t , denoted as x_t , and the output from the previous time unit $t-1$, denoted as h_{t-1} , are fed into the activation function σ . This step determines which portion of the previous output should be discarded, a process called the forget gate (f_t). The corresponding formula is as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

where σ is the active function; W_f , U_f are weight matrices; b_f is weight bias. After introducing x_t and h_{t-1} into the network, the input gate i_t employs an activation function to decide whether to disregard or incorporate new information. The cell state \tilde{C}_t , which signifies the content to be updated, is computed using a tanh function.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (5)$$

where W_i , U_i , W_c , and U_c are weight matrices; b_i and b_c are weight bias. The previous cell state C_{t-1} is multiplied by f_t to ascertain the extent of memory retention from the last instance. This outcome is then added to the new memory, derived from i_t multiplied by \tilde{C}_t . Ultimately, the newly updated memory C_t is outputted.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (6)$$

where \odot denotes the Hadamard product. Upon introducing x_t and C_{t-1} into the network, the activation function σ is employed to decide whether to output new information, a process known as the output gate o_t . C_t is input into the tanh function and multiplied by o_t to yield the output h_t at time t .

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (7)$$

$$h_t = o_t \odot \tanh(C_t) \quad (8)$$

where W_o , U_o are weight matrices; b_o is weight bias.

2.3 Gated Recurrent Unit

GRU can also be regarded as a simple variant of LSTM (Xie et al. 2022). GRU has two gating layers: reset gate z_t and update gate r_t . The reset gate determines how much

information to forget from a previous memory. The function of the update gate is similar to the forget gate and input gate of the LSTM unit. It determines how much information from previous memory can be passed to the future. The z_t and r_t formulas are as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{9}$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{10}$$

where σ is the active function; h_{t-1} is the output of the last unit; W_z , U_z , W_r and U_r are weight matrices; b_z and b_r are weight bias. The hidden state candidate (\tilde{h}_t) and hidden state (h_t) at time t can be defined by the following formula:

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{11}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{12}$$

where W_h and U_h are weight matrices; b_h is weight bias.

2.4 Bidirectional Long Short-Term Memory

BiLSTM constructs a forward and a backward hidden layer, linking the input layer to these forward and backward hidden layers, respectively (Wu et al. 2023). Subsequently, the forecasted value is computed collectively. Within the BiLSTM model, the output values of the hidden layer are as follows:

$$\vec{h}^{(t)} = \vec{\sigma}^{(t)} \odot \tanh(\vec{c}^{(t)}) \tag{13}$$

$$\overleftarrow{h}^{(t)} = \overleftarrow{\sigma}^{(t)} \odot \tanh(\overleftarrow{c}^{(t)}) \tag{14}$$

In the above, $\vec{h}^{(t)} \in R^{p_1 \times 1}$ represents the output of the hidden layer calculated by the forward LSTM, while $\overleftarrow{h}^{(t)} \in R^{p_1 \times 1}$ denotes the output of the hidden layer calculated by the backward LSTM. The weight matrix of the hidden layer output, as calculated by the forward and backward LSTM, can be designated as $V \in R^{k \times p_1}$ and $\Lambda \in R^{k \times p_2}$, respectively, and the input bias weight can be set to $b_y \in R^{k \times 1}$. The forecasted value of the BiLSTM model at time t can be expressed as follows:

$$\hat{y}^{(t)} = \sigma_y \left(a_y^{(t)} \right) = \sigma_y \left(V \vec{h}^{(t)} + \Lambda \overleftarrow{h}^{(t)} + b_y \right) \tag{15}$$

where $\sigma_y(\cdot)$ is generally set as the Softmax function.

2.5 Trend Forecasting Method

The process flow of the TFM is illustrated in Fig. 1 and comprises four steps, detailed as follows:

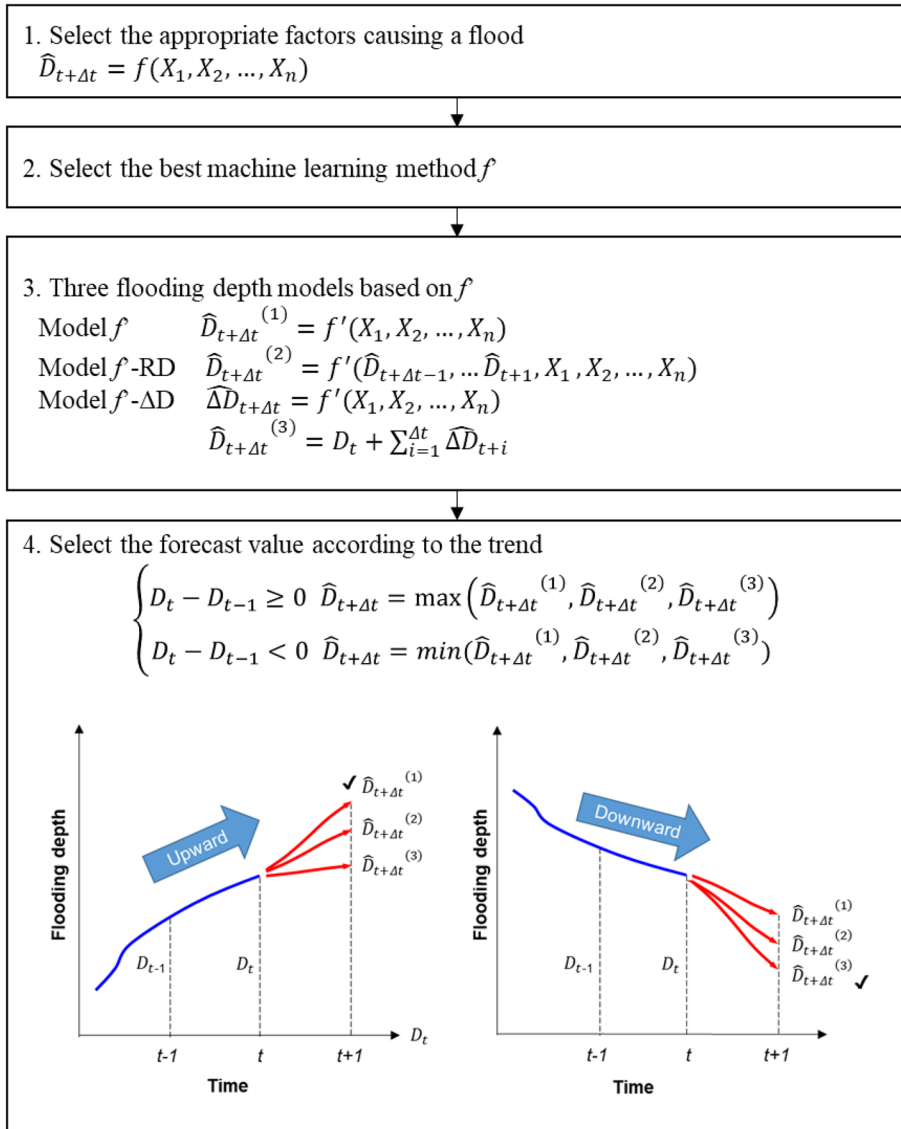


Fig. 1 Flowchart for Trend Forecasting Method (TFM)

2.5.1 Select the Appropriate Factors Causing a Flood

First, the factors that mainly lead to the flooding depth are listed, which the following formula can express:

$$\widehat{D}_{t+\Delta t} = f(X_1, X_2, \dots, X_n) \tag{16}$$

where $\widehat{D}_{t+\Delta t}$ represents the forecasted flood depth at time $t + \Delta t$; Δt is the lead time; and f denotes various machine learning methods; the terms X_1, X_2, \dots, X_n represent input factors

and their lag lengths. For an inundation forecasting model, selecting input factors is crucial, precisely determining the value n in the Eq. (16). In the proposed method, the value n can primarily be determined through optimization algorithms or by assessing the correlation between the input and output factors of the model.

2.5.2 Select the Best Machine Learning Method

Flood depths are forecasted using various machine learning methods. The performance of each method is assessed using a specific evaluation index, leading to the selection of the most suitable machine learning method, denoted as f' . In this study, to choose the most suitable machine learning method, the performance of the flood prediction model can be determined through various evaluation metrics. Commonly used evaluation metrics typically involve calculating errors or correlations between model output values and observed values to assess the quality of the model. Therefore, the more diverse the evaluation metrics used, the more representative the selected optimal machine learning method becomes.

2.5.3 Three Flooding Depth Models Based on the Best Machine Learning Method

(a) Typical Flooding Depth Model (Model f')

Model f' employs the original input and output, which the following formula can represent:

$$\hat{D}_{t+\Delta t}^{(1)} = f'(X_1, X_2, \dots, X_n) \tag{17}$$

where $\hat{D}_{t+\Delta t}^{(1)}$ denotes the forecasted flood depth by Model f' .

(b) Recurrent Flooding Depth Model (Model f' -RD)

Several studies have indicated that incorporating forecast information from each time step as model input data can enhance the model's accuracy (Jhong et al. 2017; Yang et al. 2019). Consequently, Model f' -RD can be represented by the following formula:

$$\hat{D}_{t+\Delta t}^{(2)} = f'(\hat{D}_{t+\Delta t-1}, \dots, \hat{D}_{t+1}, X_1, X_2, \dots, X_n) \tag{18}$$

where $\hat{D}_{t+\Delta t}^{(2)}$ denotes the forecasted flood depth by Model f' -RD.

(c) Delta Flooding Depth Model (Model f' -ΔD)

Variations in rainfall either increase or decrease, resulting in flood depth alterations. Consequently, the model can establish the relationship between rainfall and the variation in flood depth $\Delta\hat{D}_{t+\Delta t}$ at $t + \Delta t$. The term $\Delta\hat{D}_{t+\Delta t}$ represents the variation in flood depth between the time steps $t + \Delta t - 1$ and $t + \Delta t$. The Model f' -ΔD can be expressed as follows:

$$\Delta\hat{D}_{t+\Delta t} = f'(X_1, X_2, \dots, X_n) \tag{19}$$

$\hat{D}_{t+\Delta t}^{(3)}$ represents the current water depth D_t accumulating the changes in flood depth at subsequent time steps. The formula is as follows:

$$\hat{D}_{t+\Delta t}^{(3)} = D_t + \sum_{i=1}^{\Delta t} \Delta \hat{D}_{t+i} \tag{20}$$

where $\hat{D}_{t+\Delta t}^{(3)}$ denotes the forecasted flood depth by Model $f-\Delta D$.

(d) Select the forecast value according to the trend

Three different forecasted results for future flood depth predictions can be obtained: $\hat{D}_{t+\Delta t}^{(1)}$, $\hat{D}_{t+\Delta t}^{(2)}$ and $\hat{D}_{t+\Delta t}^{(3)}$. The concept of trend forecasting (as illustrated in the step 4 in Fig. 1) is introduced. If the flood depth at the current moment is increasing, the likelihood of it rising at the next moment may be high. Consequently, the maximum value among the three is selected. Conversely, if the flood depth at the current moment is decreasing, the chance of it reducing at the next moment is high, leading to the selection of the minimum value among the three. The following equation can represent this:

$$\begin{cases} D_t - D_{t-1} \geq 0 \hat{D}_{t+\Delta t} = \max(\hat{D}_{t+\Delta t}^{(1)}, \hat{D}_{t+\Delta t}^{(2)}, \hat{D}_{t+\Delta t}^{(3)}) \\ D_t - D_{t-1} < 0 \hat{D}_{t+\Delta t} = \min(\hat{D}_{t+\Delta t}^{(1)}, \hat{D}_{t+\Delta t}^{(2)}, \hat{D}_{t+\Delta t}^{(3)}) \end{cases} \tag{21}$$

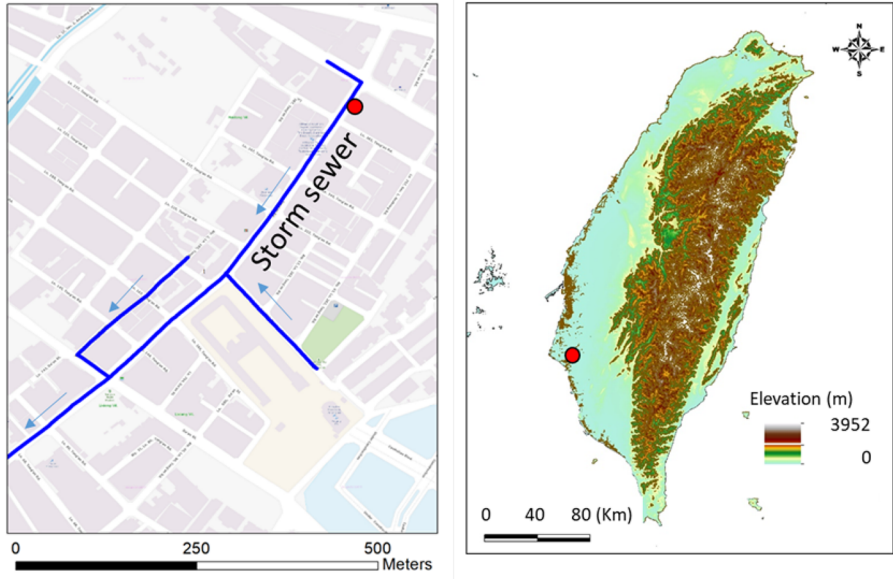
3 Materials

3.1 Study Area

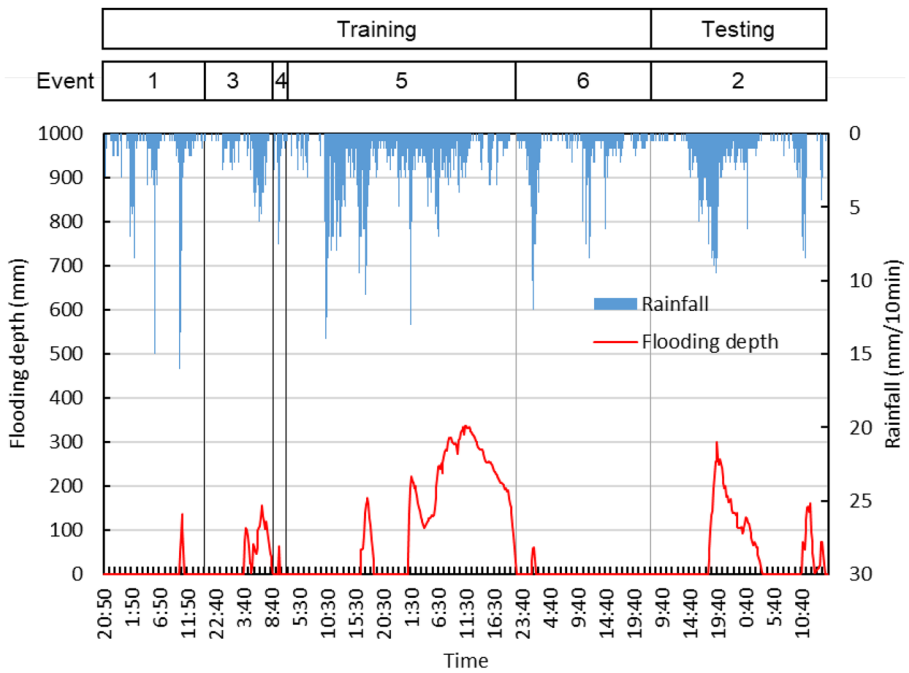
The research area is located on one of the local traffic arteries in Annan District, Tainan City, Taiwan (23°02'38.6 "N 120°11'35.9" E), as shown in Fig. 2a. The site is centrally located and densely populated with buildings. The area is located upstream of the storm sewer system, but the low-lying terrain makes it difficult for surface runoff to flow into the stormwater sewer system. Flooding often occurs during typhoons and heavy rains, resulting in traffic interruptions and flooding buildings.

3.2 Observed Flooding Depths

The rainfall data in this study is sourced from a rainfall station established by the Tainan City Government, with data recorded at ten-minute intervals. The flood depth data, recorded at ten-minute intervals, is derived from a Flooding Depth Gauge (FDG) (Model Anasystem SenSmart WLS) installed by the Water Resources Planning Branch, Water Resources Agency. The FDG measures inundation depth via radio frequency admittance, boasting an accuracy of 0.5% of the sensor length (typically 1.5 to 2.0 m). The observed data is transmitted to a cloud server every 30 s using Long Range technology. The FDG was installed in 2016 and has recorded 781 data across six rainfall events. Three of these events were caused by typhoons and tropical depressions, while the other three resulted from heavy rains. Five events (comprising 592 data) were used for training, and one event (consisting of 189 data) was used for testing, as depicted in Fig. 2b. Event 6 was the test event due to its second-highest maximum flooding depth.



(a)



(b)

Fig. 2 a Location of the study area in Annan District, Tainan City, Taiwan b Observed flooding events for model training and testing

3.3 Model Development

Figure 2b demonstrates a strong correlation between rainfall and flooding depth. Consequently, this study used $D_t, R_t, R_{t-1}, \dots, R_{t-(L_R-1)}$ as X in formula (16). Here, D_t represents the real-time observed flood depth, R_t is the real-time observed rainfall, and $R_{t-1}, \dots, R_{t-(L_R-1)}$ are the observed antecedent rainfalls. L_R is the lag length of rainfall (10 min). The correlation coefficient between flood depth and rainfall for different L_R values was calculated to identify an appropriate L_R . Then, the trial and error method was employed to determine the optimal L_R . Consequently, this study selected L_R as 6. Moreover, the water depth was forecasted for the next 10, 20, 30, 40, 50, and 60 min; hence Δt was set as 6. The initial experiment evaluated the flood depth forecasting ability of the BPNN, LSTM, and BiLSTM models.

This study employed four AI models, specifically BPNN, LSTM, GRU, and BiLSTM, to evaluate flood depth forecasting. The general forms of the LSTM, GRU, and BiLSTM models are presented as follows:

$$\begin{aligned}
 \text{Model BPNN} \quad & \hat{D}_{t+\Delta t} = f_{BPNN}(D_t, R_t, R_{t-1}, \dots, R_{t-(L_R-1)}) \\
 \text{Model LSTM} \quad & \hat{D}_{t+\Delta t} = f_{LSTM}(D_t, R_t, R_{t-1}, \dots, R_{t-(L_R-1)}) \\
 \text{Model GRU} \quad & \hat{D}_{t+\Delta t} = f_{GRU}(D_t, R_t, R_{t-1}, \dots, R_{t-(L_R-1)}) \\
 \text{Model BiLSTM} \quad & \hat{D}_{t+\Delta t} = f_{BiLSTM}(D_t, R_t, R_{t-1}, \dots, R_{t-(L_R-1)})
 \end{aligned} \tag{22}$$

This study computed the evaluation indicators for the four models, with GRU demonstrating superior performance (refer to Sect. 4.1 for details). Based on Eqs. (18) and (19), Models GRU-RD and GRU- ΔD were respectively developed. The general forms of Models GRU-RD and GRU- ΔD are presented as follows:

$$\text{Model GRU - RD} \quad \hat{D}_{t+\Delta t} = f_{GRU}(\hat{D}_{t+\Delta t-1}, \dots, \hat{D}_{t+1}, D_t, R_t, R_{t-1}, \dots, R_{t-(L_{RD}-1)}) \tag{23}$$

$$\text{Model GRU - } \Delta D \quad \Delta \hat{D}_{t+\Delta t} = f_{GRU}(D_t, R_t, R_{t-1}, \dots, R_{t-(L_{RD}-1)}) \tag{24}$$

Because the actual flooding depth must be greater than or equal to zero, if the forecasted flood depth was negative in this study, it was regarded as a 0 value. The Maximum Forecasting Method (MFM) and Average Forecasting Method (AFM) were adopted to assess the improvements provided by the proposed TFM, allowing for a direct comparison with TFM. Based on the MFM, the maximum value among $\hat{D}_{t+\Delta t}^{(1)}, \hat{D}_{t+\Delta t}^{(2)}$, and $\hat{D}_{t+\Delta t}^{(3)}$ is used as the forecast value, which the following equation can represent:

$$\hat{D}_{t+\Delta t} = \max(\hat{D}_{t+\Delta t}^{(1)}, \hat{D}_{t+\Delta t}^{(2)}, \hat{D}_{t+\Delta t}^{(3)}) \tag{25}$$

The concept of AFM involves using the average value of the three forecasted values as the final value. The following equation can represent this:

$$\hat{D}_{t+\Delta t} = \frac{\hat{D}_{t+\Delta t}^{(1)} + \hat{D}_{t+\Delta t}^{(2)} + \hat{D}_{t+\Delta t}^{(3)}}{3} \tag{26}$$

If the warning authorities exhibit a tendency toward safety and conservatism, they are inclined to opt for the maximum forecasted values. In the MFM method, the maximum value among $\hat{D}_{t+\Delta t}^{(1)}, \hat{D}_{t+\Delta t}^{(2)}$, and $\hat{D}_{t+\Delta t}^{(3)}$ is selected as the model output. Additionally, the application of the AFM method is justified by the prevalence of the mean as a statistical method.

In the AFM method, the average value among $\widehat{D}_{t+\Delta t}^{(1)}$, $\widehat{D}_{t+\Delta t}^{(2)}$, and $\widehat{D}_{t+\Delta t}^{(3)}$ is selected as the model output. Finally, the model output values selected by the TFM, MFM, and AFM are compared.

The models were developed using Python 3.8 and the Keras library. Data normalization was achieved using the Max–min Scaler. Hyperparameters for the models were optimized through trial and error, meaning the models have been run several times. All models were assigned the same hyperparameters to facilitate the comparison of structural differences between models. Each model had three hidden layers and 20 neurons. The batch size was set to 10, and the dropout rate was 0.2. The loss function used was Mean Squared Error (MSE), and the activation function was tanh. The program was compiled using the Adam optimizer, and the number of epochs was set to 120.

3.4 Evaluation of Model Performance

In this study, the Root Mean Square Error (RMSE), coefficient of determination (R^2), and Nash–Sutcliffe model efficiency coefficient (NSE) were used to evaluate the overall forecasting performance of the model. ED_p and ET_p were also used to assess the model’s forecasting performance for peak values. RMSE represents the error between forecasted and observed values. A model can forecast more accurately when the RMSE is closer to 0. R^2 is often used to assess the linear correlation between the model’s and the target’s output. R^2 ranges from 0 to 1. A model can forecast more accurately when its R^2 is closer to 1. NSE is commonly used to evaluate hydrological forecasting models. NSE values range from negative infinity to 1. A model with an NSE value closer to 1 can make more accurate forecasts. Models with an NSE value of less than 0 demonstrate poorer performance than those that only produce the mean. ED_p evaluates the error between the forecasted peak and the observed peak, while ET_p assesses the error between the forecasted peak occurrence time and the observed peak occurrence time. The smaller the absolute values of ED_p and ET_p , the better the model’s performance. The evaluation indices can be calculated using the following formulas:

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n [D_{for}(t_i) - D_{obs}(t_i)]^2 \right]^{1/2} \tag{27}$$

$$R^2 = \left[\frac{\sum_{i=1}^N (D_{for}(t_i) - \overline{D}_{for}(t_i))(D_{obs}(t_i) - \overline{D}_{obs}(t_i))}{\sqrt{\sum_{i=1}^N (D_{for}(t_i) - \overline{D}_{for}(t_i))^2 \sum_{i=1}^N (D_{obs}(t_i) - \overline{D}_{obs}(t_i))^2}} \right]^2 \tag{28}$$

$$NSE = 1 - \frac{\sum_{i=1}^n [D_{for}(t_i) - D_{obs}(t_i)]^2}{\sum_{i=1}^n [D_{obs}(t_i) - \overline{D}_{obs}(t_i)]^2} \tag{29}$$

$$ED_p = \frac{D_{p,for} - D_{p,obs}}{D_{p,obs}} \times 100\% \tag{30}$$

$$ET_p = T_{p,for} - T_{p,obs} \quad (31)$$

where $D_{for}(t_i)$ and $D_{obs}(t_i)$ represent the i -th forecasted and observed values, respectively. $\bar{D}_{for}(t_i)$ and $\bar{D}_{obs}(t_i)$ are the mean values of the forecasted and observed values, respectively. $D_{p,for}$ and $D_{p,obs}$ represent the forecasted and observed peak values, respectively, while T_p^{for} and T_p^{obs} denote the occurrence times for the forecasted and observed peak values, respectively.

4 Results and Discussion

4.1 Comparative Analysis of the Flood Forecasting Models Based on the BPNN, LSTM, GRU, and BiLSTM

Table 1 details BPNN, LSTM, GRU, and BiLSTM training and testing outcomes in flood depth forecasting. Figure 3a–c compare these models during the testing phase. As lead time extended, all models' RMSE increased, and their R^2 and NSE decreased, thus reducing forecasting precision. This trend corroborates Yang et al. (2023) findings from forecasting flood depth in Rende District, Tainan City, Taiwan, using BPNN, RNN, and LSTM.

BPNN exhibited notably higher RMSE and significantly lower R^2 and NSE than LSTM. Figure 4c–f demonstrate BPNN's fewer black points compared to LSTM's blue points at water depths below 150 mm, becoming comparable beyond this threshold. Particularly under 150 mm, BPNN's forecasts were substantially inferior to LSTM's. This reflects Yang et al. (2023) similar findings on BPNN versus LSTM. The fluctuation in flood depth corresponds to the hydrological process where initial rainfall runoff is managed by ditches and sewers, causing no initial flooding. However, as runoff exceeds these capacities, flooding ensues. Post-rainfall, as runoff decreases, so does flooding. LSTM, adept at capturing such time series and long-term dependencies, consistently surpasses BPNN, especially over longer forecast lead times.

Figures 3a–c and 4a–f show that GRU, with slightly lower RMSE and marginally higher R^2 and NSE than LSTM, forecasted flood depths more accurately. While LSTM units have three gates (input, forget, output) (Ding et al. 2020), GRU units have only two (update and reset), lacking an output gate and using the hidden state as output (Kao et al. 2020). This makes GRUs simpler and potentially less prone to overfitting in linear systems. Despite the task and data dictating the choice between LSTM and GRU, our experiments suggest a slight superiority of GRU in flood depth forecasting.

Figures 3a–c and 4a–f show BiLSTM's slightly higher RMSE and lower R^2 and NSE than LSTM, indicating LSTM's marginally better forecast accuracy. While both models are tailored for sequential data, BiLSTM, which processes data forward and backward, offers enhanced context understanding, which is beneficial for tasks like language modeling (Vatanchi et al. 2023). However, for applications like flood depth forecasting, where future context is less critical and past rainfall predominantly influences changes, our experiments reveal that LSTM outperforms BiLSTM.

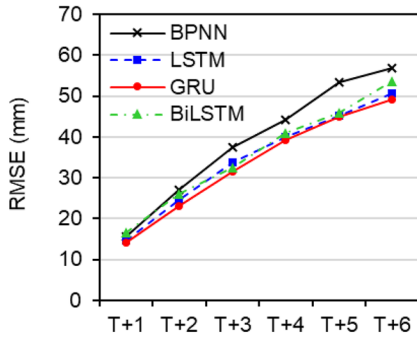
Figure 3d–e show ED_p and ET_p for LSTM, GRU, and BiLSTM, with most ED_p values being negative, indicating underestimated flood peaks. BPNN's average ED_p was -3.11%, slightly outperforming LSTM (-9.32%), GRU (-9.89%), and BiLSTM (-8.36%) in peak prediction. ET_p for the models ranged from 10 to 60 min across six forecasts, showing time lags. Inputs included D_t and past rainfall, with D_t having a more significant impact.

Table 1 Evaluation indexes of the models

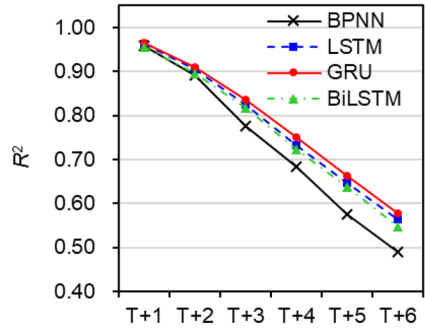
Model	Lead time	Training			Testing				
		RMSE (mm)	R^2	NSE	RMSE (mm)	R^2	NSE	ED_p (%)	ET_p (min)
BPNN	1	0.02	0.99	0.99	16	0.96	0.96	0.36	10
	2	0.04	0.98	0.98	27	0.89	0.87	-3.02	20
	3	0.05	0.97	0.97	38	0.77	0.75	1.24	30
	4	0.07	0.94	0.94	44	0.68	0.65	-2.90	40
	5	0.08	0.93	0.92	54	0.57	0.49	-6.65	50
	6	0.08	0.92	0.92	57	0.49	0.42	-7.69	60
LSTM	1	0.03	0.99	0.99	15	0.96	0.96	-6.45	10
	2	0.05	0.97	0.97	24	0.90	0.89	-8.13	20
	3	0.07	0.94	0.94	34	0.83	0.80	-11.56	30
	4	0.08	0.92	0.92	40	0.73	0.71	-10.39	40
	5	0.09	0.90	0.90	45	0.65	0.63	-10.39	50
	6	0.10	0.88	0.88	51	0.56	0.54	-8.98	60
GRU	1 ^a	0.03	0.99	0.99	14	0.96	0.96	-5.84	10
	2	0.05	0.97	0.97	23	0.91	0.90	-7.88	20
	3	0.07	0.94	0.94	32	0.84	0.82	-12.22	30
	4	0.08	0.92	0.92	39	0.75	0.72	-9.90	40
	5	0.09	0.90	0.90	45	0.66	0.64	-9.90	50
	6	0.10	0.88	0.87	49	0.58	0.56	-13.59	60
BiLSTM	1	0.03	0.99	0.99	17	0.96	0.95	-2.01	10
	2	0.05	0.97	0.97	26	0.90	0.88	-4.28	20
	3	0.07	0.95	0.94	33	0.82	0.81	-4.65	30
	4	0.08	0.92	0.92	41	0.72	0.70	-14.04	40
	5	0.09	0.90	0.89	46	0.64	0.62	-14.04	50
	6	0.10	0.89	0.88	54	0.55	0.48	-11.15	60
GRU-RD	1 ^a	0.03	0.99	0.99	14	0.96	0.96	-5.84	10
	2	0.03	0.99	0.99	15	0.97	0.96	-9.16	10
	3	0.04	0.99	0.98	16	0.95	0.95	-8.68	10
	4	0.04	0.99	0.98	22	0.93	0.91	-18.27	10
	5	0.03	0.99	0.99	24	0.90	0.90	-13.39	20
	6	0.04	0.99	0.98	29	0.88	0.85	-20.30	30
GRU-ΔD	1	6.44	1.00	1.00	15	0.96	0.96	-2.01	10
	2	3.73	1.00	1.00	23	0.91	0.90	-3.94	20
	3	4.43	1.00	1.00	31	0.84	0.83	-5.47	30
	4	3.75	1.00	1.00	38	0.75	0.74	-6.56	40
	5	3.84	1.00	1.00	45	0.66	0.64	-8.20	50
	6	3.16	1.00	1.00	50	0.57	0.54	-9.87	60

^aThe forecasted results of GRU and GRU-RD are the same for \hat{D}_{t+1} forecasting

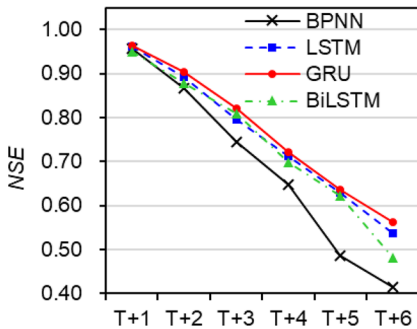
Balancing recent and historical data is essential to prevent overlooking long-term trends and reacting to short-term noise, which can cause forecast lags. While real-time data is critical, incorporating a broader historical context is necessary to capture patterns over



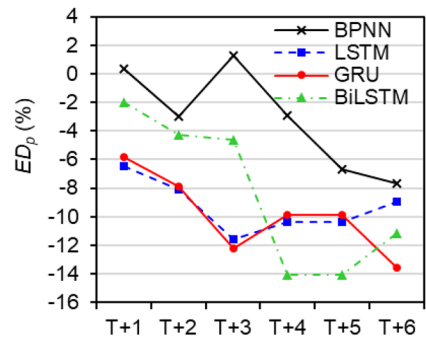
(a)



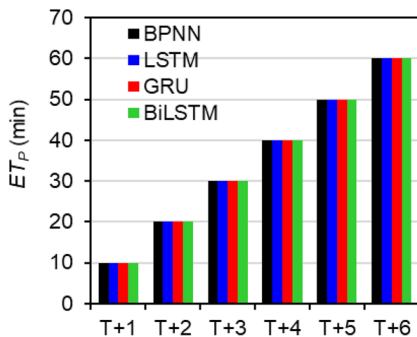
(b)



(c)



(d)



(e)

Fig. 3 **a** RMSE of BPNN, LSTM, GRU, and BiLSTM; **b** R^2 of BPNN, LSTM, GRU, and BiLSTM; **c** NSE of BPNN, LSTM, GRU, and BiLSTM; **d** ED_p of BPNN, LSTM, GRU, and BiLSTM; **e** ET_p of BPNN, LSTM, GRU, and BiLSTM

time. An overly focused model on recent data may miss future trends (Bollerslev et al. 1994; Box et al. 2015; Hyndman and Athanasopoulos 2018; Zhang et al. 1998).

4.2 Improvement of Model Accuracy Due to the Use of GRU, GRU- Δ D, and GRU-RD

Table 1 compares flood depth forecasts for GRU, GRU-RD, and GRU- Δ D. GRU- Δ D had slightly higher RMSE but similar R^2 and NSE values to GRU, indicating comparable accuracy. GRU- Δ D's average ED_p of -6.0% surpassed GRU's -9.2%, making it more effective in peak depth prediction, though both had similar time lags. Meanwhile, GRU-RD showed lower RMSE and higher R^2 and NSE than GRU, enhancing accuracy. Although its -12.6% ED_p suggests slightly reduced peak forecasting precision compared to GRU's -9.2%. However, GRU-RD's smaller ET_p indicates improved time lag. GRU-RD's feedback loops, which pass previous inputs and outputs to the next step, account for this reduced lag. These results are consistent with findings by Nanda et al. (2016) and Yang et al. (2019).

In time series forecasting, autoregressive modeling, which uses previous outputs as inputs, effectively reduces time lag by capturing the time-dependent structure. This responsiveness is crucial for non-stationary series where statistics vary over time, enhancing forecast accuracy by integrating recent information and minimizing error. It also captures non-linear relationships in complex dynamics (Brockwell and Davis 2016; Chatfield and Xing 2019; Hyndman and Athanasopoulos 2018; Zhang et al. 1998).

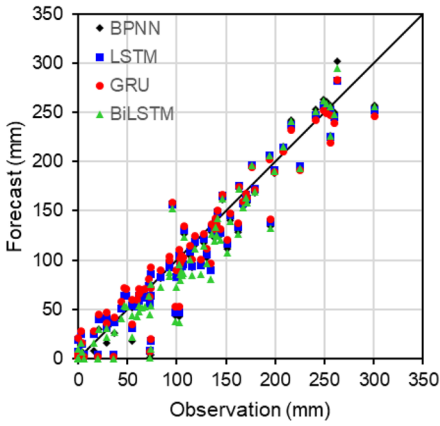
4.3 Improvement of Model Accuracy Due to the Use of the Proposed TFM

Figure 5a–c show GRU's RMSE as notably higher and R^2 and NSE as significantly lower than AFM, MFM, and TFM. GRU's forecasts, as Fig. 6d–f depict, notably underestimated actual values, a shortcoming mitigated by applying AFM, MFM, and TFM. TFM demonstrated smaller RMSE and larger R^2 and NSE than AFM, indicating superior performance as shown in Figs. 5a–c and 6b–f. During the test event, flood depth rose or fell continuously in 49 steps and discontinuously in 23, confirming a 68% probability of continued rise or fall, supporting the TFM hypothesis. While AFM and TFM showed similar forecast biases (ED_p), AFM exhibited a time lag in peak forecasts, unlike TFM. Consequently, TFM more accurately predicts flood peaks without noticeable time lag.

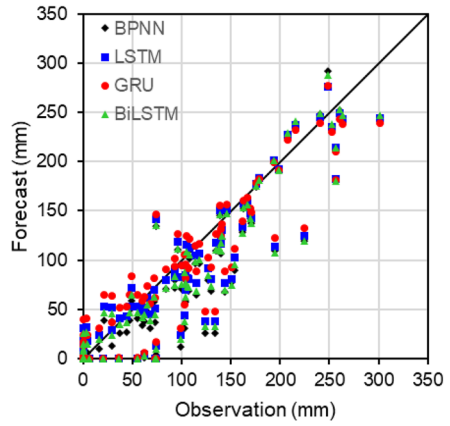
Figure 5a–c show that TFM outperformed MFM in early forecasts (T+1–T+4) with lower RMSE and higher R^2 and NSE, while MFM excelled in later forecasts (T+5 and T+6). Figure 6b–f depict MFM's significant overestimation by selecting maximum values from various forecasts. Despite MFM's smaller ED_p indicating better peak predictions, it had time lags, unlike TFM, which could predict flood peaks without noticeable delays (Fig. 5d–e). Overall, TFM was the most accurate for flood depth, followed by MFM and AFM. While MFM led in peak forecasting, its time lags were a drawback, whereas TFM, though slightly less accurate, predicted flood peaks without significant time lag.

4.4 Limitations of the Work and Future Research

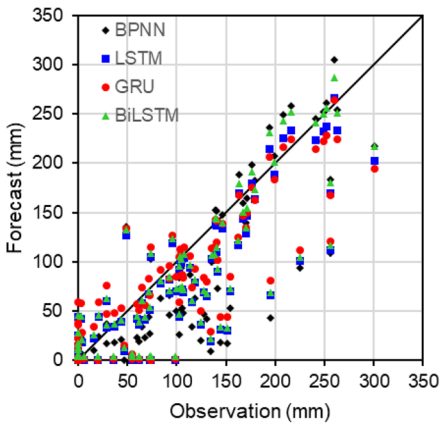
This study's limited observational data includes six rainfall events and 781 records. Enhancements could consist of using additional deep learning models to generate more data. Future work utilizing the GRU model might explore more complex models and extend the current 60-min lead time to longer periods.



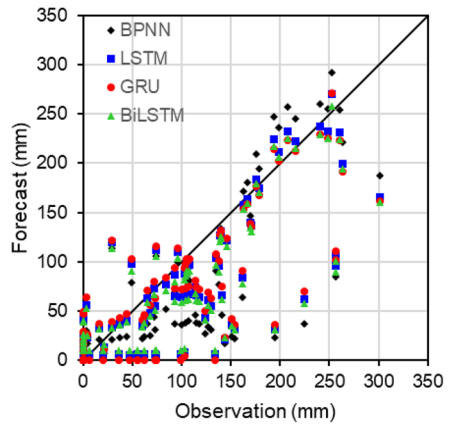
(a) T+1



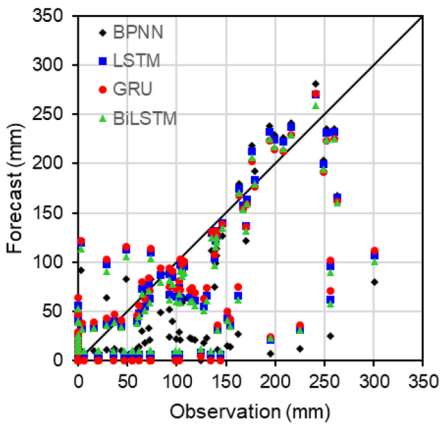
(b) T+2



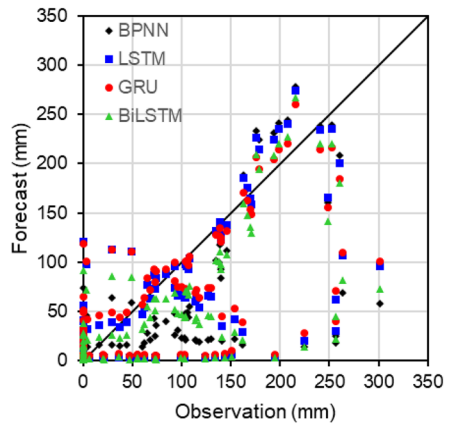
(c) T+3



(d) T+4



(e) T+5



(f) T+6

◀ Fig. 4 Scatter diagrams of BPNN, LSTM, GRU, and BiLSTM for a T+1, b T+2, c T+3, d T+4, e T+5, and f T+6 forecasting

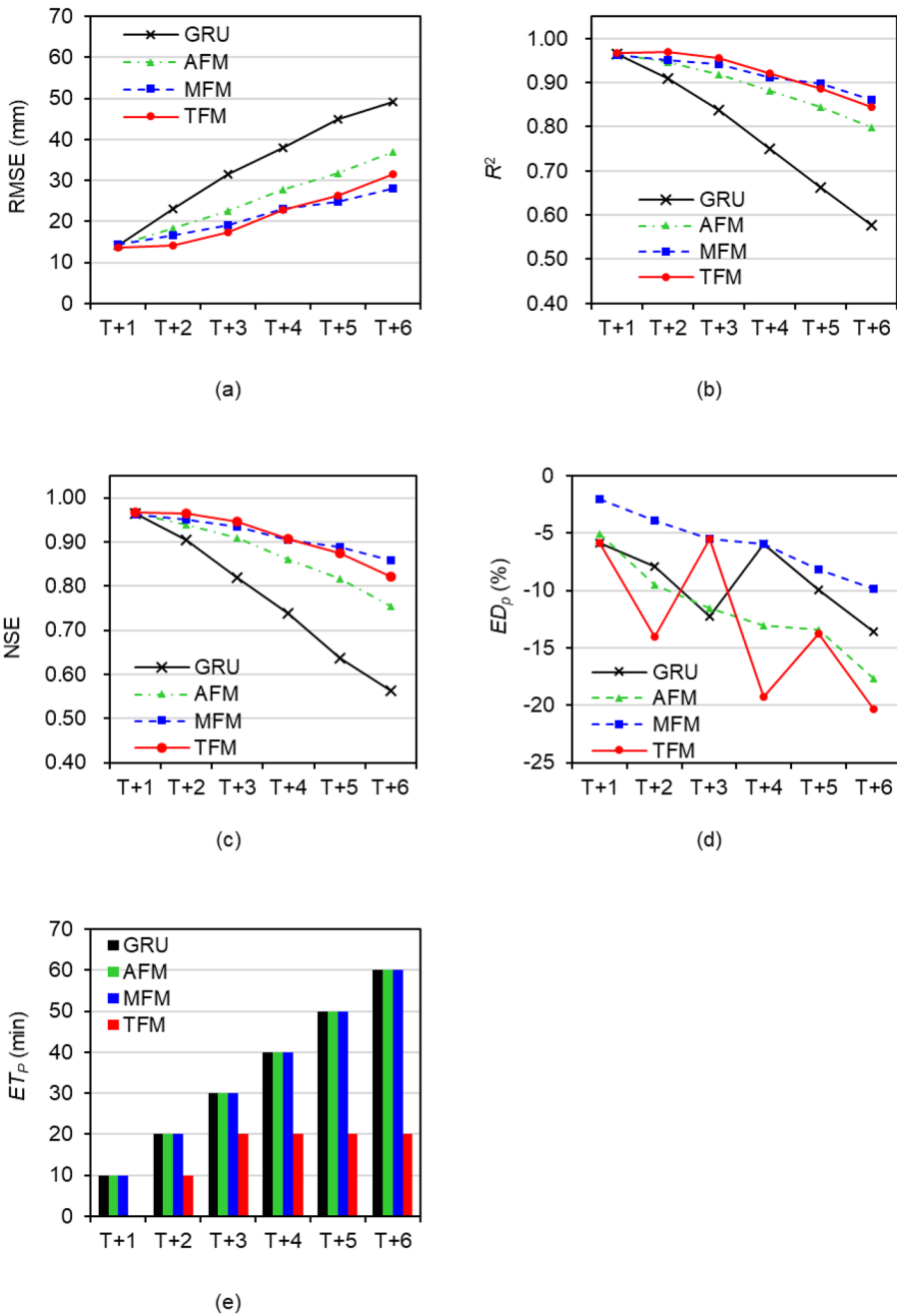
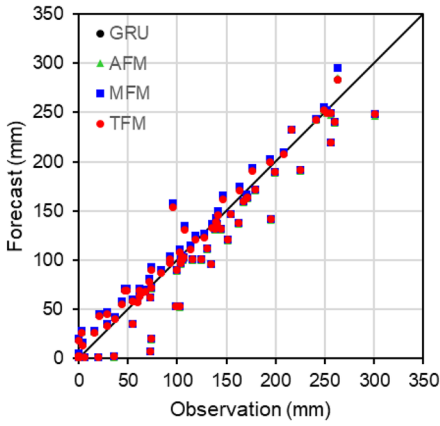
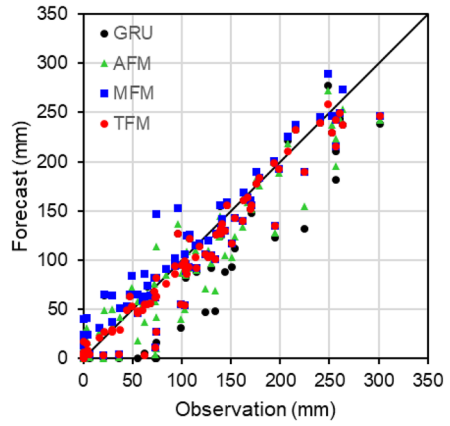


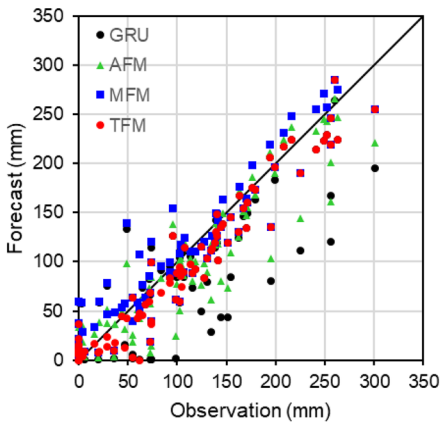
Fig. 5 a RMSE of GRU, AFM, MFM and TFM; b R² of GRU, AFM, MFM and TFM; c NSE of GRU, AFM, MFM and TFM; d ED_p of GRU, AFM, MFM and TFM; e ET_p of GRU, AFM, MFM and TFM



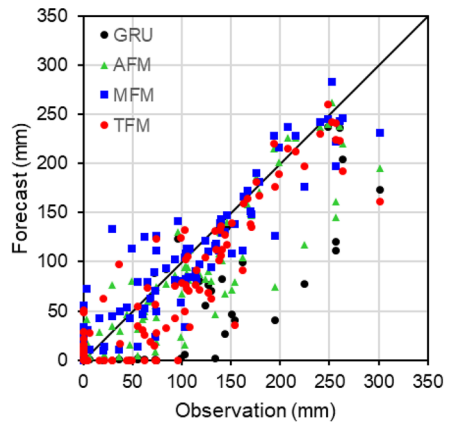
(a) T+1



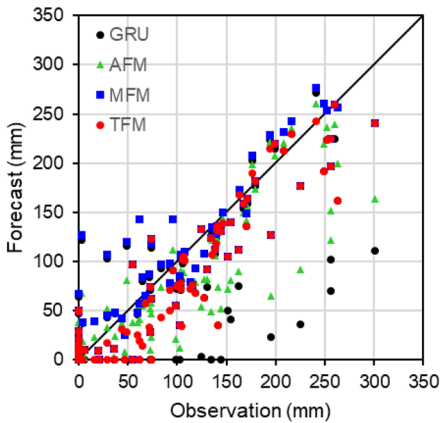
(b) T+2



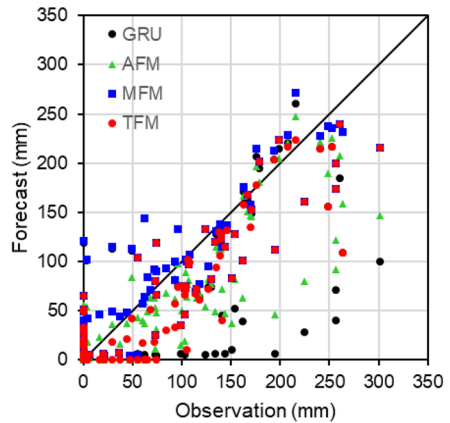
(c) T+3



(d) T+4



(e) T+5



(f) T+6

◀**Fig. 6** Scatter diagrams of GRU, AFM, MFM, and TFM for **a** T + 1, **b** T + 2, **c** T + 3, **d** T + 4, **e** T + 5, and **f** T + 6 forecasting

4.5 Impact and Usefulness of Work Concerning Water Resources Management

The proposed TFM is versatile and applicable to various hydrological forecasts like stream-flow, flood stage, and sewer water depth, and it is especially effective in predicting continuous hydrograph trends. By integrating TFM, multimodal forecasting becomes more accurate. Furthermore, integrating this model into Tainan City's flood response system could preemptively warn residents and authorities, triggering preventive actions like installing barriers, closing roads, and deploying water pumps, thereby enhancing flood management.

5 Conclusions

This study aimed to improve model accuracy and overcome the time lag problem. The proposed Trend Forecasting Method (TFM) has achieved this purpose. First, the appropriate input factors causing flood events and the most suitable AI algorithm were determined to construct the forecasting models. Second, the forecasting models, respectively using the multi-step-ahead approach and the variation in flood depth as input, were developed to investigate the relationships between the input and output variables of models predicting flooding depths at different lead times and the improvement of flood prediction concerning the variation between the current and the previous time step in flooding depth. Then, based on a flood hydrograph's rising and falling limb, the maximum and minimum values predicted by the models above were chosen as the final outputs, respectively.

The benefits of the proposed method were showcased through its implementation in the Annan District of Tainan City, Taiwan. The research results indicated that the GRU exhibited the highest accuracy among all examined models, with LSTM, BiLSTM, and BPNN following that order. Despite all models demonstrating time delay issues, GRU has shown the best empirical performance, while BPNN excels in peak forecast. Based on the selection of the GRU-based forecasting model, the proposed method enhanced the prediction accuracy of the original GRU model, addressing issues such as the time lag commonly encountered in time series forecasting. Moreover, it enabled a more accurate prediction of flood peaks. Based on the findings of this study, the proposed method applies to various hydrologically related time series forecasting domains, particularly those requiring improved accuracy in predictive models for time series or long-term dependencies. Additionally, AI-related models demand substantial-high-quality historical observational data for training and testing. Due to limitations in available observational data, future research could involve updating datasets with more observations and exploring integrating new models to refine the proposed method, thereby enhancing predictive accuracy and extending the lead time of predictions.

Acknowledgements We acknowledge the National Science and Technology Council of Taiwan (Grant numbers: MOST 109-2625-M-035-007-MY3 and NSTC 112-2625-M-011-001 -) for granting support. The authors thank the Water Resources Planning Branch, Water Resources Agency, Ministry of Economic Affairs for providing relevant data and Mr. George Chih-Yu Chen for his assistance in the English editing. The authors also thank Tsung-Tang Tsai for his invaluable work developing AI models. During the preparation of this work the authors used GPT-4 in order to English editing.

Authors Contributions Song-Yue Yang: Conceptualization, Methodology, Visualization, Writing – original draft, Funding acquisition; You-Da Jhong: Writing – review & editing; Bing-Chen Jhong: Conceptualization, Supervision, Writing – review & editing; Yun-Yang Lin: Data curation, Software.

Funding This research project is funded by the National Science and Technology Council, Taiwan (grant numbers: MOST 109–2625-M-035–007-MY3 and NSTC 112–2625-M-011–001 -).

Availability of Data and Materials The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

Declarations

Ethical Approval The authors will comply with all academic norms by the journal of Water Resources Management.

Consent to Participate All authors agreed to join this research.

Consent to Publish All authors agreed with the content and that all have explicit consent to submit.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

- Berkhahn S, Fuchs L, Neuweiler I (2019) An ensemble neural network model for real-time prediction of urban floods. *J Hydrol* 575:743–754
- Bollerslev T, Engle RF, Nelson DB (1994) ARCH models. *Handb Econ* 4:2959–3038
- Box GE, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*. John Wiley & Sons
- Brockwell PJ, Davis RA (2016) *Introduction to time series and forecasting*. Springer, New York
- Chakraborty R, Pal SC, Ruidas D, Roy P, Saha A, Chowdhuri I (2023) Living with floods using state-of-the-art and geospatial techniques: flood mitigation alternatives, management measures, and policy recommendations. *Water* 15:558
- Chatfield C, Xing H (2019) *The analysis of time series: an introduction with R*. CRC Press
- Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078)
- Chowdhuri I, Pal SC, Chakraborty R (2020) Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. *Adv Space Res* 65:1466–1489
- Chu H, Wu W, Wang QJ, Nathan R, Wei J (2020) An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environ Modell Softw* 124:104587
- Cui Z, Zhou Y, Guo S, Wang J, Ba H, He S (2021) A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting. *Hydrol Res* 52:1436–1454
- Dazzi S, Vaccondio R, Mignosa P (2021) Flood stage forecasting using machine-learning methods: a case study on the Parma River (Italy). *Water* 13:1612
- Ding Y, Zhu Y, Feng J, Zhang P, Cheng Z (2020) Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing* 403:348–359
- Du S, Van Rompaey A, Shi P, Ja W (2015) A dual effect of urban expansion on flood risk in the Pearl River Delta (China) revealed by land-use scenarios and direct runoff simulation. *Nat Hazards* 77:111–128
- Graves A, Schmidhuber J (2005) Framework phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw* 18:602–610
- Guo Z, Leitão JP, Simões NE, Moosavi V (2021) Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *J Flood Risk Manag* 14:e12684
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780

- Hofmann J, Schüttrumpf H (2021) floodGAN: Using deep adversarial learning to predict pluvial flooding in real time. *Water* 13:2255
- Hosseiny H (2021) A deep learning model for predicting river flood depth and extent. *Environ Modell Softw* 145:105186
- Hyndman RJ, Athanasopoulos G (2018) *Forecasting: principles and practice*. OTexts
- Inteaz MA, Hossain I (2023) Climate change impacts on 'seasonality index' and its potential implications on rainwater savings. *Water Resour Manag* 37:2593–2606
- Jhong BC, Wang JH, Lin GF (2017) An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. *J Hydrol* 547:236–252
- Kang H, Yang S, Huang J, Oh J (2020) Time series prediction of wastewater flow rate by bidirectional LSTM deep learning. *Int J Control Autom Syst* 18:3023–3030
- Kao IF, Liou JY, Lee MH, Chang FJ (2021) Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts. *J Hydrol* 598:126371
- Kao IF, Zhou Y, Chang LC, Chang FJ (2020) Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. *J Hydrol* 583:124631
- Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrol Earth Syst Sci* 22:6005–6022
- Meng M, Dąbrowski M, Tai Y, Stead D, Chan F (2019) Collaborative spatial planning in the face of flood risk in delta cities: A policy framing perspective. *Environ Sci Policy* 96:95–104
- Najafabadipour A, Kamali G, Nezamabadi-Pour H (2022) Application of artificial intelligence techniques for the determination of groundwater level using spatio-temporal parameters. *ACS Omega* 7:10751–10764
- 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
- Nearing GS, Klotz D, Frame JM et al (2022) Data assimilation and autoregression for using near-real-time streamflow observations in long short-term memory networks. *Hydrol Earth Syst Sci* 26:5493–5513
- Palmitessa R, Mikkelsen PS, Borup M, Law AW (2021) Soft sensing of water depth in combined sewers using LSTM neural networks with missing observations. *J Hydro-Environ Res* 38:106–116
- Plate EJ (2007) Early warning and flood forecasting for large rivers with the lower Mekong as example. *J Hydro-Environ Res* 1:80–94
- Roy P, Pal SC, Chakraborty R, Chowdhuri I, Malik S, Das B (2020) Threats of climate and land use change on future flood susceptibility. *J Clean Prod* 272:122757
- Ruidas D, Chakraborty R, Islam ARMT, Saha A, Pal SC (2022) A novel hybrid of meta-optimization approach for flash flood-susceptibility assessment in a monsoon-dominated watershed. *Eastern India Environ Earth Sci* 81:145
- Ruidas D, Saha A, Islam ARMT, Costache R, Pal SC (2023) Development of geo-environmental factors controlled flash flood hazard map for emergency relief operation in complex hydro-geomorphic environment of tropical river, India. *Environ Sci Pollut Res* 30:106951–106966
- Schuetze T, Chelleri L (2013) Integrating decentralized rainwater management in urban planning and design: Flood resilient and sustainable water management using the example of coastal cities in the Netherlands and Taiwan. *Water* 5:593–616
- Sit M, Demiray BZ, Xiang Z, Ewing GJ, Sermet Y, Demir I (2020) A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci Technol* 82:2635–2670
- Sun W, Bocchini P, Davison BD (2020) Applications of artificial intelligence for disaster management. *Nat Hazards* 103:2631–2689
- Vatanchi SM, Etemadifard H, Maghrebi MF, Shad R (2023) A comparative study on forecasting of long-term daily streamflow using ANN, ANFIS, BiLSTM, and CNN-GRU-LSTM. *Water Resour Manag* 37:4769–4785
- Wu J, Wang Z, Hu Y, Tao S, Dong J (2023) Runoff forecasting using convolutional neural networks and optimized bi-directional long short-term memory. *Water Resour Manag* 37:937–953
- Xie H, Randall M, Chau K-w (2022) Green roof hydrological modelling with GRU and LSTM networks. *Water Resour Manag* 36:1107–1122
- Yang SY, Jhong BC, Jhong YD, Tsai TT, Chen CS (2023) Long short-term memory integrating moving average method for flood inundation depth forecasting based on observed data in urban area. *Nat Hazards* 116:2339–2361
- Yang S, Yang D, Chen J, Zhao B (2019) Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model. *J Hydrol* 579:124229

- Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: The state of the art. *Int J Forecast* 14:35–62
- Zhu S, Luo X, Yuan X, Xu Z (2020) An improved long short-term memory network for streamflow forecasting in the upper Yangtze River. *Stoch Environ Res Risk Assess* 34:1313–1329

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.