

Received August 6, 2020, accepted August 13, 2020, date of publication August 17, 2020, date of current version August 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017277

# Real-Time Prediction of the Water Accumulation Process of Urban Stormy Accumulation Points Based on Deep Learning

ZENING WU, YIHONG ZHOU<sup>✉</sup>, AND HUILIANG WANG

College of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Huiliang Wang (wanghuiliang@126.com)

This work was supported by the Key Project of National Natural Science Foundation of China under Grant 51739009.

**ABSTRACT** Influenced by climate change and urbanization, urban flood frequently occurs and represents a serious challenge for many cities. Therefore, it is necessary to generate refined predictions of urban floods, such as the prediction of water accumulation processes at water accumulation points, which is of great significance for supporting water-related managers to reduce flood losses. In this study, 16 combination schemes of rainfall sensitivity indicators were used to determine the optimal scheme for predicting the depth of accumulated water, and the gradient boosting decision tree (GBDT) algorithm in deep learning was used to build a prediction model of the accumulation process of urban stormy accumulation points. Among the 16 schemes, the relative error of scheme 1 is 15.39%, and the qualified rate is 92.86%. This scheme exhibits the highest accuracy for the prediction results of water accumulation depth. Given this finding, the GBDT algorithm was used to construct a regression prediction model of the water accumulation process based on the collected historical rainfall water accumulation data of 50 water accumulation points. The results demonstrated that the GBDT regression prediction model has a mean relative error of 19.77%, a qualified rate of 82.00%, and a peak average relative error of 5.48%, which verify the validity and applicability of the model for the real-time prediction of the process of water accumulation.

**INDEX TERMS** Urban flood, deep learning, water accumulation, real-time prediction.

## I. INTRODUCTION

In recent years, global warming and urbanization have led to the increasing frequency and influence of urban floods [1], [2], posing severe challenges to urban flood control and drainage. The Louisiana flood in 2016, the Chinese Shouguang and Zhengzhou floods in 2018, and the Iranian “3.25” flood in 2019 serve as examples, and these heavy rains and floods resulted in considerable economic losses and casualties to the city, which has become a prominent bottleneck affecting the healthy development of a city [3], [4]. Recent studies demonstrated that future global warming will lead to significant changes in extreme rainfall intensity and frequency [5], [6].

The heavy losses caused by urban floods have made people attach great importance to urban flood prevention and control [7]. Over recent years, scholars and urban management departments from various countries have adopted a

large number of engineering and nonengineering measures to continuously promote urban flood prevention and control work [8]. For example, a comprehensive investigation assessed the drainage capacity of the urban drainage system, the effective utilization rate of the pipe network, and the blockage of river channels. In addition, some engineering measures, such as improving the design standard of urban drainage pipe networks, increasing artificial lakes, and repairing deep tunnels in the city, were used to improve the flood control capacity of the city to alleviate the losses caused by urban flooding as much as possible. After years of governance, although urban drainage capacity and urban water-logging prevention capacity have been improved to a certain extent, urban floods still frequently occur.

To monitor the city’s water accumulation situation in real time and improve the city’s ability to deal with floods, urban water accumulation monitoring systems have been established in Shanghai, Wuhan, Zhengzhou and other cities in China. The establishment of a water accumulation monitoring system with high temporal and spatial resolution can be used

The associate editor coordinating the review of this manuscript and approving it for publication was Nilanjan Dey.

to monitor the process at water accumulation points, release and display the ponding information of water accumulation points in real time, and provide timely flood control and emergency measures for key water accumulation areas. However, the prediction information of ponding is more valuable than the monitoring information in urban flood prevention and control. Therefore, it is necessary to build a data-driven water accumulation prediction model based on the monitoring data of water accumulation.

The prediction model of water accumulation processes based on data-driven requires tremendous rainfall and water accumulation process data of water accumulation points [9]. Before the emergence of urban water accumulation monitoring systems, it was difficult to obtain these data. The emergence of urban water accumulation monitoring systems provides data support for the construction of a prediction model of water accumulation processes. Therefore, some studies have built a variety of time series models using neural networks to study the prediction methods of ponding process [10], [11]. However, due to the accumulation of errors in the multistep iteration of time series models, this model has a good effect on short-term water accumulation prediction. The prediction effect will gradually decline with the extension of the prediction period [12], which limits the application of the model to a certain extent. Some recent studies have shown that nontime series deep learning methods [13], such as GBDT [14], [15], SVM, random forest [16] and neural network [17], [18], exhibit good forecasting effects in some forecasting fields, but the prediction of water accumulation processes is a continuous process that changes over time. These nontime series models can only predict a certain feature of the water accumulation process and cannot realize the prediction of the water accumulation process. Therefore, there is an urgent need to identify a modeling method that is suitable for the prediction of the water accumulation process in a longer prediction period. As a type of integrated learning algorithm, GBDT performs well in tasks such as classification and regression [19], [20] given its high efficiency, high precision and low deviation, which has attracted increasing attention [21], [22]. However, to the authors' best knowledge, the application of the GBDT algorithm to urban flood research is still rare, and no study has applied the GBDT algorithm to the prediction of water accumulation process.

Therefore, this study attempts to use a deep learning model (GBDT) to propose a new modeling method for real-time prediction of the water accumulation process. The rainfall sensitivity index applicable to the prediction of water accumulation depth was first proposed in this study. On this basis, by splitting and reorganizing the rainfall and water accumulation data, the prediction regression model of the water accumulation process was constructed using the GBDT algorithm. Mean relative error (MRE), qualification rate (QR), deterministic coefficient (DC) and average relative error of the peak value (AREPV) were used to evaluate the performance of the model in the prediction of water

accumulation processes. The research results may be useful for urban flood forecasting and reducing flood losses.

The innovation of this article is that a new modeling method was proposed using the GBDT algorithm, which was suitable for the prediction of water accumulation processes at water accumulation points. This modeling method can be used to predict the water accumulation process with a longer prediction period based on the historical rainfall and water logging data, achieving a longer prediction period of water accumulation processes based on nontime series models. This study provides a new idea and method for the prediction of water accumulation processes and provides a technical reference for urban management personnel to prevent and control urban floods.

## II. LITERATURE REVIEW

Urban flood prediction is an effective means to help urban flood management personnel reduce the losses caused by urban flood. Many scholars have done considerable research on the theory and technology of urban flood prediction in recent years [23], [24]. To the authors' best knowledge, there are two main types of urban flood forecasting research: research based on hydrological and hydrodynamic models and research based on data driven models.

SWMM [25], [26], Mike [27], [28], and Storm [29] are widely used hydrological and hydrodynamic models in flood prediction, among which the SWMM model developed by the U.S. Environmental Protection Agency is one of the most widely used models to simulate urban runoff and drainage [30], [31]. Zhou *et al.* [32] comprehensively considered the land use type, surface impermeability and drainage system to establish a SWMM model and estimated the flood volume and risk under urbanization and climate change. Zhu *et al.* [33] simulated the influence of different pavement structures (drainage surface, permeable pavement and permeable road) on reducing surface runoff by constructing a SWMM model. Kim and Cho [34] employed SWMM and a 2D surface model to simulate the inundation area and range of the city under 320 different rainfall situations. This study provides a scenario-based urban flood forecasting method. These urban flood simulation models have achieved good results in urban flood simulation and risk analysis by building different modules, such as runoff yield, confluence, and channels [35]–[37]. In addition, with the improvement of computer capabilities and the development of geographic information system (GIS) technology, the spatial resolution of surface data has been significantly improved in recent years, and the division of the model spatial scale has become increasingly detailed [38], [39]. However, there is no consensus on how to divide the spatial scale in hydrological simulation [40], [41]. Moreover, studies in recent articles have shown that due to the complex interactions between the drainage system and the surface that are difficult to determine and the lack of validation data in the simulation results, the promotion and use of the model is occasionally limited [42]–[44].

During the past 10 years, with the rapid development of artificial intelligence and deep learning technology, data-driven deep learning models [45] have become very popular and have been widely used in engineering [17], [18] electricity [13], [16] agriculture [11] and many other fields. Chatterjee *et al.* [17] proposed a method based on particle swarm optimization to train the NN (NN-PSO), which can solve the problem of predicting the failure of multistoried reinforced concrete buildings by detecting the failure probability of the multistoried RC building structures in the future. Hu *et al.* [13] designed a low-cost solution for interconnecting electrical and electronic devices using classic machine learning models for smart grid load analysis and forecasting. Singh and Mohanty [10] proposed a hybrid model combining a generalized neuron model and adaptive genetic algorithm and conducted a short-term forecast of electricity prices in the New South Wales electricity market. The results show that the mean absolute percentage error (MAPE) of the hybrid model is significantly lower than that of the conventional neural networks and regression models. In the field of hydrology, neural network model is widely used in water level prediction and urban flood prediction given its ability of approaching nonlinear functions [46], [47]. Chiang *et al.* [48] employed a recurrent neural network (RNN) to construct a model of the relationship between rainfall and the water level in the urban sewage system. The results show that the performance of RNN gradually decreases for 5-, 10-, 15-, and 20-min-ahead water level predictions, but the CC (the correlation coefficient) value is greater than 0.95, which indicates that the RNN can effectively predict the water level in a short time. Chang *et al.* [49] constructed three recurrent neural network (RNN) models for short-term (10-60 min) water level predictions. Abou Rjeily *et al.* [50] employed a nonlinear autoregressive neural network to construct a relationship model between rainfall intensity and water depth in urban drainage wells. The inputs of the model are rainfall intensities and water depth at the previous time step. The results show that it performs well on both minor and severe storm events. In the above research, these neural network models have achieved good results in short-term water accumulation prediction, but the accuracy is occasionally significantly reduced when predicting the depth of water accumulation in a longer foreseeing period. This phenomenon is explained by the fact that the residual at the previous time step will be superimposed [12] for each prediction step. The accumulation of residuals reduces the prediction accuracy of the model for the water accumulation depth in a longer prediction period.

In summary, the time series model based on neural network has a good performance on the short-term prediction of water accumulation process, but the residual at the previous time step will be superimposed for each prediction step, the accuracy of prediction in longer prediction period is occasionally significantly reduced. Therefore, this study attempts to use the GBDT algorithm to propose a new modeling method, which hope that this modeling method

can solve the problem that the prediction accuracy of water accumulation process in a longer foreseeing period will be gradually reduced.

### III. MOTIVATION

A city is an area with a concentrated economy and population. Flood disasters represent significant challenges to the normal operation of the city. Although the urban management department has noticed the threat of urban flood disasters, it has also taken various measures to control and monitor urban flood problems. However, the phenomenon of urban flooding remains prominent and represents a severe challenge to urban flood control. In addition, in the process of using deep learning technology to predict urban floods, given the difficulty of data collection and guaranteeing model accuracy in long-term prediction periods, most of the existing studies focus on short-term water accumulation prediction for water accumulation points, providing a reference for urban flood prevention and control. However, the prediction results of a longer prediction period for water accumulation processes can provide more time for prevent and control flood, which is urgently needed for urban flood prevention and control.

The motivation of this study can be attributed to the rapid development of cities and the construction of modern cities, and urban flood control work has put forward new and urgent needs for scientists. Exclusive real-time monitoring of water accumulation points can no longer meet the requirements of urban flood control work. It is urgent to obtain a prediction method for water accumulation processes with a longer prediction period to control the trend of urban water accumulation in real-time. Therefore, the objective of this study is to propose a modeling method suitable for the prediction of the accumulation of water in a longer prediction period based on the existing water accumulation data and deep learning methods. The following chapters will elaborate on the detailed process of model construction.

### IV. MATERIAL

#### A. STUDY AREA

Zhengzhou is located in the north central part of China (112°42'E to 114°14'E, 34°16'N to 34°58'N), with a total area of 7446.3 km<sup>2</sup>. It is one of the largest cities in Central China (Fig. 1) and an important node-city on the “new silk road” between Europe and Asia. As of the end of 2018, the permanent population of Zhengzhou reached 10.13 million, and the total GDP (Gross Domestic Product) exceeded 1 trillion yuan. The region exhibits a continental monsoon climate with an average annual rainfall of 524.1 mm. However, 60% of the rainfall occurs between June and September, leading to an increased risk of urban flood. Urban management departments have adopted measures, such as dredging drainage pipes, increasing the permeable surface area, and improving the design standards of drainage pipe networks to improve the city's flood control capabilities, but urban floods still occasionally occur. For example, heavy rains on August 19, 2018 and August 1, 2019 caused widespread flooding in

Zhengzhou City, posing a considerable threat to the safe operation of the city and the safety of people's property.

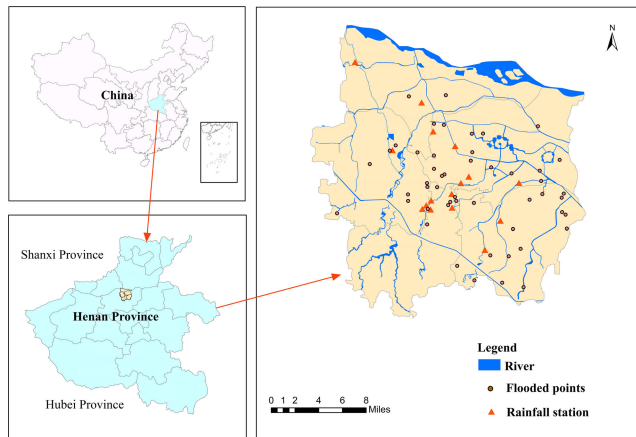


FIGURE 1. Location of study area.

## B. DATA

The data of rainfall and water accumulation process are the basis for constructing the relationship model of rainfall and water accumulation, and these data support the prediction of water accumulation processes of water accumulation points. Based on previous research results, this study used historical rainfall data as the input variable of the model and accumulated water depth as the output variable to construct the relationship model between rainfall and accumulated water. A detailed description of the data is provided as follows:

1) Historical rainfall data: These data are the time distribution data of rainfall obtained from self-recording rain gauges of 16 rainfall stations in the study area (Fig. 1). Nineteen historical rainfall events from 2014 to 2018 with 10-min temporal resolution were used as sample data for the model. These events occurred on August 7, 2014; August 30, 2014; September 14, 2014; September 17, 2014; August 3, 2015; August 26, 2015; September 4, 2015; July 14, 2016; July 19, 2016; August 25, 2016; May 22, 2017; July 18, 2017; August 12, 2017; August 25, 2017; August 25, 2018; May 15, 2018; June 26, 2018; July 13, 2018 and August 19, 2018. In order to obtain the rainfall process data of ponding points, the Kriging method of space interpolation was used to interpolate the rainfall data of 16 rainfall stations, and use GIS to obtain the rainfall process data of each ponding point. All rainfall data were obtained from the Henan Meteorological Service.

2) Flooded data: The locations and inundation process of flooded urban areas were obtained from the historical flooding records, which were collected from the monitoring equipment at each intersection and stored in the urban disaster database. The flooded data were obtained from the Zhengzhou Municipal Urban Management Bureau. In this study, the water accumulation process data of 50 water accumulation points was collected as sample data. It should be noted that the time resolution of water accumulation process

data at the water accumulation point is 1min. In order to unify with the temporal resolution of rainfall, the time resolution of water accumulation process data in the model is also 10min.

## V. SELECTION OF SENSITIVITY INDEX FOR DEPTH PREDICTION OF ACCUMULATED WATER

Urban floods are the result of the comprehensive effect of climate variables and underlying surface conditions (including rainfall, topography, river network, land use and pipe network) [37], in which rainfall is the driving factor of urban flood [51]. Due to the limited change of the underlying surface conditions in the city in the short term, rainfall is the direct reason of urban flooding [52]. For fixed water accumulation points, an internal relationship between rainfall and water accumulation is observed. Therefore, a model of the relationship between rainfall and water accumulation processes was generated in this study. However, the deep learning model based on the GBDT algorithm cannot directly input the rainfall process data into the model. The rainfall sensitivity data must be input to characterize the rainfall process data, which differs from the hydrological model. Therefore, the selection of the rainfall sensitivity index that affects the depth of water accumulation is a key step in constructing the relationship model between rainfall and the water accumulation processes.

Studies in recent years have demonstrated that rainfall characteristics, such as rainfall, rainfall duration, peak rainfall, position coefficient, rainfall intensity variance, and peak multiple, have different effects on water accumulation [53], [54]. However, for fixed water accumulation points, an important factor that affects the depth of water accumulation at a certain moment in the water accumulation process is the location of early larger rainfall intensity. Among the existing sensitivity indicators, rainfall, rainfall duration, peak rainfall, rainfall intensity variance, and peak multiple cannot reflect the location of early larger rainfall intensity. Although the location coefficient reflects the location of peak rainfall, it does not reflect the location of other early larger rainfall intensity. Based on this notion, a sensitive index called concentration skewness was proposed in this study, which is used to reflect the location of early larger rainfall intensity. Its formula is as follows:

$$P_i = \text{Rank}(p_i) \quad (1)$$

$$T_i = \text{Rank}(t_i) \quad (2)$$

$$CS = \frac{\sum_{i=1}^{n/5} (T_{pi} - t_n)}{n/5} \quad (3)$$

where  $n$  is the total number of time intervals of rainfall, which refers to the ratio of rainfall duration to the resolution of rainfall data;  $i$  is the number of the  $i$ -th time interval of the rainfall;  $P_i$  and  $T_i$  are the rainfall intensity and duration of the  $i$ -th time interval, respectively; Rank is the descending order function;  $T_{pi}$  is the location when the rainfall intensity is  $P_i$ ; and CS is the concentration skewness.



**TABLE 1. Combination scheme of sensitivity indexes for prediction of water accumulation depth.**

Combination scheme number	rainfall	rainfall duration	peak rainfall	location coefficient	peak multiple	rainfall intensity variance	concentration skewness	MRE
Scheme 1	✓	✓	✓	✓	✓	✓	✓	15.39%
Scheme 2	✓	✓	✓	✓	✓	✓	-	18.44%
Scheme 3	✓	✓	✓	✓	✓	-	✓	22.42%
Scheme 4	✓	✓	✓	✓	-	✓	✓	21.82%
Scheme 5	✓	✓	✓	-	✓	✓	✓	27.74%
Scheme 6	✓	✓	✓	✓	✓	-	-	27.75%
Scheme 7	✓	✓	✓	✓	-	✓	-	38.25%
Scheme 8	✓	✓	✓	✓	-	-	✓	21.30%
Scheme 9	✓	✓	✓	-	✓	✓	-	37.12%
Scheme 10	✓	✓	✓	-	✓	-	✓	21.69%
Scheme 11	✓	✓	✓	-	-	✓	✓	16.10%
Scheme 12	✓	✓	✓	✓	-	-	-	44.69%
Scheme 13	✓	✓	✓	-	✓	-	-	51.99%
Scheme 14	✓	✓	✓	-	-	✓	-	55.29%
Scheme 15	✓	✓	✓	-	-	-	✓	15.41%
Scheme 16	✓	✓	✓	-	-	-	-	62.76%

Notes: Where ✓ represents contain, - represents none

Different sensitivity indicators and different index combination schemes may have different degrees of impact on the depth of accumulated water. Among them, rainfall, rainfall duration, and peak rainfall are typically considered to be important sensitive indicators that affect the depth of accumulated water [55], [56]. Based on this notion, rainfall, rainfall duration, and peak rainfall were used as common indicators in this study. Sixteen index combination schemes were obtained using the random combination method for the location coefficient, rainfall intensity variance, peak multiple, and concentration skewness. The rainfall-water accumulation relationship model was constructed for each index combination scheme, and the MRE and QR of the prediction results of the water accumulation depth were used to evaluate the performance of each index combination scheme. The detailed indicator selection process is as follows:

1) Data preparation: Calculate the sensitivity index value for rainfall and water accumulation process based on the data collected from the rainfall process and the water accumulation process of the water accumulation point in the urban flooding process. Specifically, the calculation includes rainfall, rainfall duration, peak rainfall, location coefficient, rainfall intensity variance, peak multiple, and concentration skewness.

2) Index combination: Fifteen combination schemes were obtained using the random combination method for location coefficient, rainfall intensity variance, peak multiple and concentration skewness, among which rainfall, rainfall duration

and peak rainfall were common indexes. In addition, rainfall, rainfall duration, and peak rainfall were used as blank control groups, and 16 index combination schemes were obtained in total (Table 1).

3) Index scheme selection method: Logistic regression is a multivariate statistical model with a simple calculation and clear physical meaning that can form a multiple regression relationship between the dependent variable and several independent variables [57], [58]. Among the many statistical analysis methods, the significant advantage of logistic regression is that it can better solve the problem of interdependence between factors in the process of evaluating each impact factor [59], which is useful for the selection and evaluation of indicator schemes. Therefore, this study selected an index combination scheme by constructing a logistic regression model of each index combination scheme. The sensitivity index value in the index scheme was used as the input variable, and the corresponding water accumulation depth was used as the output variable. In total, 70% data were used as training data, and 30% data were used as verification data.

4) Accuracy analysis method of index scheme: The MRE and QR were used to comprehensively evaluate the accuracy difference of the 16 index schemes, and the index scheme with the highest comprehensive accuracy was selected as the sensitive index combination scheme affecting the prediction of water accumulation depth. In this study, the MRE refers to the average of the ratio of the absolute error between the

TABLE 2. Model evaluation index explanation.

Indicator	Full name	Formula	Remarks
MRE	mean relative error	$MRE = \left( \frac{1}{n} \sum_{i=1}^n \left  \frac{y_{ci} - y_{oi}}{y_{oi}} \right  \right) \times 100\%$	Where, $y_c$ is the predicted value; $y_o$ is the measured value
QR	qualified rate	$QR = \frac{c}{n} \times 100\%$	Where, c is the qualified sample number; n is the total number of samples.
DC	deterministic coefficient	$DC = 1 - \frac{\sum_{i=1}^n [y_{ci} - \bar{y}_o]^2}{\sum_{i=1}^n [y_{oi} - \bar{y}_o]^2}$	Where, $y_c$ is the predicted value; $y_o$ is the measured value; $\bar{y}_o$ is the average of the measured values;
APEPV	average relative error of the peak value	$AREPV = \sum_{j=1}^m \frac{h_{cj} - h_{oj}}{h_{oj}}$	Where, $h_{cj}$ is the maximum value of the predicted water accumulation depth of the water accumulation point; $h_{oj}$ is the maximum value of the actual water accumulation depth of the water accumulation point; m is the number of rainfall events.

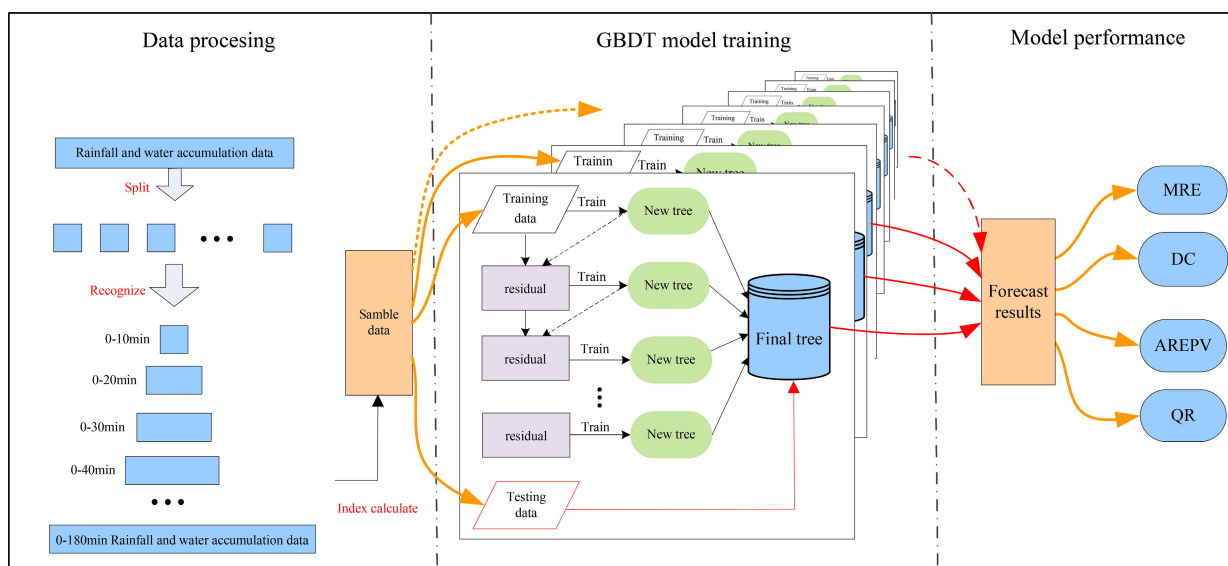


FIGURE 2. Model construction of urban flood water accumulation process prediction based on GBDT.

predicted value and the actual value to the actual value, and QR refers to the percentage of the number of samples that are predicted to be qualified in the total sample size (Table 2). According to the amount and type of data, an absolute value of the relative error of less than 20% is regarded as qualified in this study.

QR and MRE represent the overall error level of prediction results. The greater the QR and the less the MRE, the less the overall error of the prediction result.

VI. MODEL CONSTRUCTION

In this study, the GBDT algorithm was used to build the prediction model of water accumulation processes. The main modeling process can be divided into three steps (Fig. 2). The first step involves data processing and preparation, and the data are split, reorganized and used to calculate the index value of the collected rainfall and water accumulation process data. The second step is to input the sample data into GBDT model for training and input the test data into the trained

model to output the model prediction results. The third step is to evaluate the performance of the model using MRE, QR, DC and AREPV. Since our objective is to develop a real-time prediction model of water accumulation process, we have chosen Dell’s workstation as the experimental platform. The running memory of the device is 32 GB, and the CPU is E3-1505M v6 of Intel Xeon series with 3 GHz working frequency, which can achieve higher computing speed on lightweight equipment. On this basis, the model is trained and tested in Python 3.7 based on Windows 10 system.

A. DATA PROCESSING

Equidistant splitting method was used to split and reorganize the rainfall and water accumulation process data, and data of rainfall and water accumulation processes were divided into several groups of rainfall processes and water accumulation processes. As shown in Fig. 3, the 180-min rainfall process data was divided into 18 segments according to the time resolution of the rainfall data (10 min). By accumulating

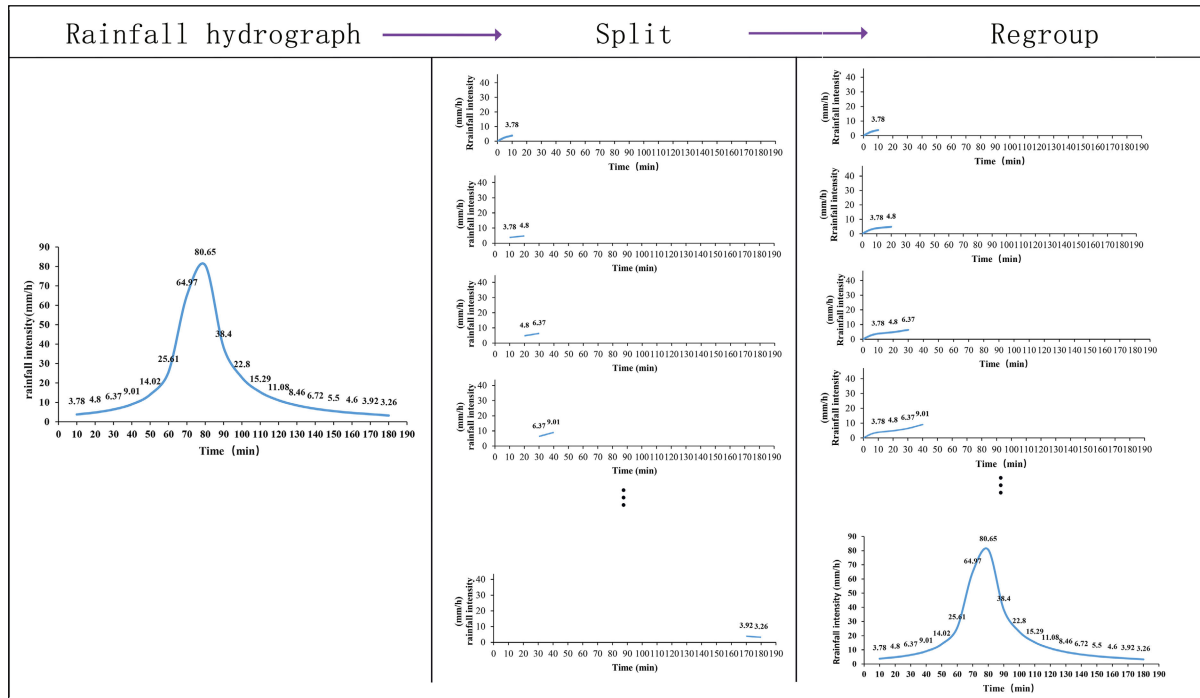


FIGURE 3. Rainfall data processing.

and reorganizing the separated rainfall in sequence, 18 groups (0-10 min, 0-20 min, 0-30 min, ..., 0-180 min) of rainfall processes were obtained (Fig. 3). Similarly, the water accumulation process was also divided into corresponding water accumulation processes. After splitting and reorganizing, a rainfall and water accumulation event with a 180-min rainfall duration become 18 rainfall and water accumulation events. By calculating the index value of each rainfall and water accumulation event, 18 sample data can be obtained from one rainfall and water accumulation event after splitting and reorganizing. Similarly, we split and reorganize the rainfall and water accumulation event of each water accumulation point in 19 rainfall and ponding events and calculate the index values after the split and reorganization. Finally, we obtained 27230 sample data, and each sample data contain seven input variables (i.e., rainfall, rainfall duration, peak rainfall, location coefficient, rainfall intensity variance, peak multiple and concentration skewness) and one output variable (i.e., depth of water accumulation). Of note, we store the sample data set of each water accumulation point separately. Each water accumulation point contains two independent tables: training data table and test data table. Finally, we store 100 independent tables of 50 water accumulation points by CSV file to form the sample data set of the model.

**B. MODEL STRUCTURE**

The relationship between rainfall and water accumulation process differs in different locations. Therefore, this study constructed a relatively independent rainfall accumulation relationship model for each accumulation point. For each water accumulation point, the GBDT algorithm was used

to train the sample data of the first 16 rainfall and water accumulation events, and the sample data of the last three rainfall events were input into the trained model for prediction. Finally, the prediction results of 50 water accumulation points were output, separately. The training part of the model can be regarded as a three-dimensional structure. First, the entire model can be divided into 50 layers. Each layer is the prediction model of each water accumulation point, and the prediction model of each water accumulation point is independent of each other. In addition, the training process of each water accumulation point can be divided into three layers, including data input, GBDT iterative training model, model prediction and result output (Fig. 2).

**C. GBDT MODELING METHOD**

GBDT is an ensemble learning algorithm combining the decision tree and gradient boosting algorithm [19], in which the decision tree generally chooses classification and regression tree (CART). The advantage of the GBDT algorithm is that during each iteration, the decision tree is trained according to the residual of the previous tree, and the classification results of all the trees are ultimately accumulated and output, effectively avoiding over fitting phenomenon [60]. The GBDT algorithm has been demonstrated to be an efficient, high-precision, low-bias model in many practices and has been widely used by data scientists in tasks, such as classification and regression. A complete mathematical and technical description of the GBDT model can be found in [19] and [61].

Using GBDT algorithm, a regression prediction model was constructed to predict the water accumulation process of the water accumulation point. Loading training sample data into

the model is the prerequisite for model training. In this study, the data reading module of Python was used to read the sample data into the model as a CSV file. The rainfall, rainfall duration, peak rainfall, location coefficient, rainfall intensity variance, peak multiple and concentration skewness were taken as input variables, and the depth of water accumulation was taken as the output variable. The expression of the sample data set was  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ .

The core of the GBDT model training is to integrate multiple trees generated by iteration into a final tree. For each iteration, the negative gradient (i.e., residual) of the function is calculated by  $r_{ti} = -[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}]_{f(x)=f_{t-1}(x)}$ , and the CART regression tree was fitted according to the obtained negative gradient. Assuming that the number of leaf nodes of the fitted regression tree is  $J$ , the area corresponding to each leaf node is  $R_{t1}, R_{t2}, \dots, R_{tJ}$ . The best fit value  $c_{ij}$  is calculated for the region corresponding to the leaf node of each decision tree using the following formula:

$$c_{ij} = \underset{c}{\operatorname{argmin}} \sum_{x_i \in R_{tj}} L(y_i, f_{t-1}(x) + c) \quad (4)$$

The model is updated according to the best fit value determined in the previous step to obtain a new tree, which completes an iterative process.

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^J c_{tj} I(x \in R_{tj}) \quad (5)$$

The above iterative steps are repeated, and the trees of each iteration are superimposed to obtain the final tree, which is the GBDT regression prediction model.

$$f(x) = f_T(x) = \sum_{t=1}^T f_t(x) \quad (6)$$

The algorithm flow of GBDT model training is shown in Figure 4. It should be noted that this study constructs an independent GBDT water accumulation process prediction model for each independent water accumulation point. Therefore, for each water accumulation point, the training process of the above model needs to be repeated to obtain a complete GBDT regression prediction model.

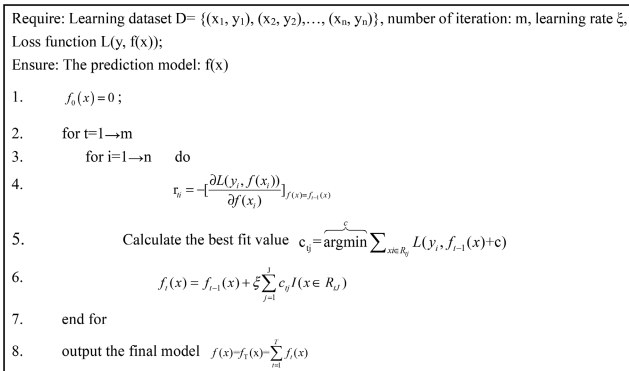


FIGURE 4. The algorithm flow of GBDT model.

#### D. PARAMETER SETTING

Parameter setting is one of the key steps in model training. The key parameters of the GBDT model are the number of iterations, the learning rate, and the complexity of the tree [62], which can effectively avoid over fitting phenomenon [14]. The number of iterations is also called the number of regression trees. Each iteration of the model will generate a new regression tree, but an excessive number of iterations may lead to overfitting. In the process of optimizing the parameters of the GBDT model, we found that when the number of iterations exceeds 1,000, the model occasionally exhibits a partial overfitting phenomenon. When the number exceeds 10,000, the model will exhibit an obvious overfitting phenomenon. Therefore, to ensure the effectiveness of the model, the upper limit of the number of iterations was set to 10,000 times in this study. The learning rate is a parameter used to characterize the contribution of each basic tree model, which can effectively prevent the over fitting phenomenon of the model. However, the decrease in the learning rate will lead to an increase in iterations, and an extremely low learning rate may lead to over fitting. Our previous research shows that better results are obtained when the learning rate is 0.05 [15]. Considering the similarity between the research data and the research content, the learning rate was set to 0.1-0.0001 for parameter optimization in this study. The complexity of the tree (i.e., the number of nodes used to fit each decision tree) reflects the true interaction between variables. To capture the interaction between variables, it is necessary to increase the complexity of the tree. However, excessively high complexity may also lead to model over fitting. Therefore, the complexity of the tree was set to 1-10 for parameter optimization in this study.

#### E. ACCURACY INSPECTION METHODS OF THE MODEL

Four indicators, including MRE, QR, DC and AREPV, were used to evaluate the performance of the GBDT regression prediction model (Table 2). In this study, the MRE and QR reflected the overall error level of the prediction result. The DC (0-1) represented the consistency between the prediction results and the actual water accumulation process. The closer the DC is to 1, the better the consistency of the predicted results. The AREPV represented the prediction accuracy of the maximum water accumulation depth. The smaller the AREPV, the greater the prediction accuracy of the maximum water accumulation depth.

### VII. RESULTS

#### A. ANALYSIS OF THE RESULTS OF DIFFERENT INDEX COMBINATIONS

Based on the collected 19 historical rainfall and water accumulation process data, index values of each rainfall, including rainfall, rainfall duration, peak rainfall, location coefficient, rainfall intensity variance, peak multiple, and concentration skewness, are calculated. The logistic regression model of water depth of 16 different index combination schemes is



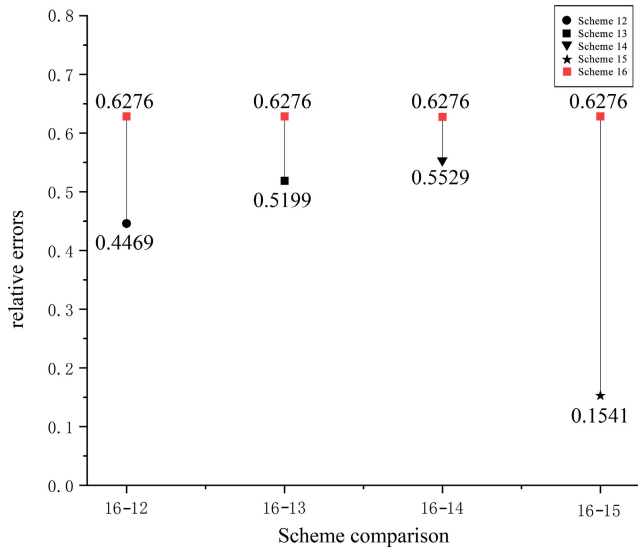


FIGURE 5. Comparison chart of accuracy improvement of water accumulation depth prediction by different indexes.

constructed by using SQL Server Data Tools, which is a data processing and analysis software of Microsoft. Here, 70% data served as training data, and 30% data served as verification data. The water accumulation depth prediction results of each index combination scheme were obtained (Table 1).

As shown in Table 1, the MRE of the prediction results of water accumulation depth of scheme 1, scheme 11 and scheme 15 are the highest at 15.39%, 16.10% and 15.41% respectively. To analyze the influence degree of different sensitive indexes on the prediction of water accumulation depth, a quantitative analysis of the improvement effect of the location coefficient, rainfall intensity variance, peak multiple, and concentration skewness on the prediction accuracy of water accumulation depth was performed, i.e., the improvement effect of schemes 12, 13, 14, and 15 relative to scheme 16 in the prediction accuracy of water depth. As shown in Fig. 5, the concentration skewness (i.e., scheme 15) exhibits the greatest improvement in the prediction accuracy of water accumulation depth at 15.41%. The MRE of water accumulation depth prediction with concentration skewness is reduced by 47.35% relative to no concentration skewness (i.e., scheme 16), indicating that the concentration skewness, a new rainfall characteristic index proposed in this study, has good applicability to the prediction of water accumulation depth.

To a certain extent, the MRE reflects the overall accuracy difference of the prediction results of each index combination scheme, but it is occasionally significantly affected by individual extreme values. The qualified rate is a parameter used to evaluate the overall qualification level of water accumulation depth prediction, which can effectively avoid the impact of individual extreme values on the overall accuracy. Therefore, to more comprehensively analyze and describe the advantages and disadvantages of each index program, the qualified rate was introduced to evaluate the performance

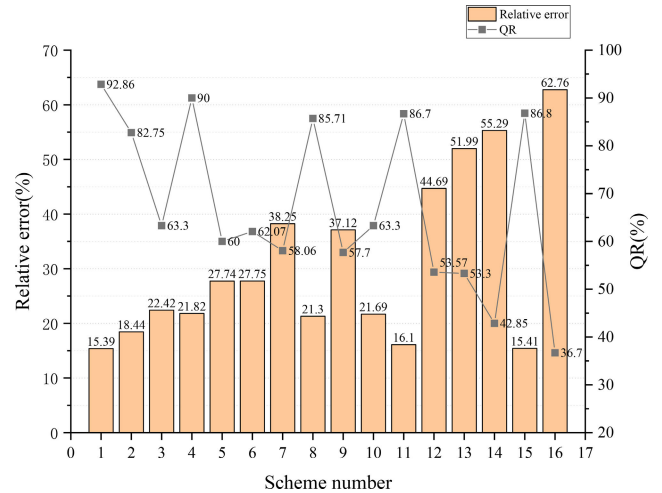


FIGURE 6. Accuracy analysis chart of qualification rate.

of each index scheme on the prediction results of water accumulation depth (Fig. 6). Figure 6 reveals that the MRE of scheme 1 is the lowest and that the qualified rate is the highest. These findings indicate that scheme 1 exhibits the greatest accuracy for prediction results of water accumulation depth and is the most suitable combination scheme for predictions of water accumulation depth.

### B. RESULT ANALYSIS OF THE GBDT REGRESSION PREDICTION MODEL

19 historical rainfall and water accumulation process data of 50 water accumulation points from 2013 to 2018 were split and reorganized as described in Fig. 3, and a total of 27230 sample data were obtained. In this study, Python 3.7 was used to build the GBDT regression prediction model, and 22,730 sample data of the first 11 rainfall and flood data were selected as training data. The parameters of the model were continuously optimized using the control variable method to determine the number of iterations, the learning rate, and the complexity of the tree (Table 3). Based on training the model according to the model parameter values determined in Table 3, the rainfall indicators of 4500 sample data of the last 3 rainfall and flood data were input into the model, and the prediction result of the depth of water accumulation of each test sample was obtained (Table 4). Table 4 demonstrates that the MRE of the GBDT regression prediction model is 19.77%, which is acceptable for the prediction of water accumulation processes.

TABLE 3. Parameters optimization results of the GBDT model.

The number of iterations	Learning rate	Complexity of the tree
150	0.05	6

The fitting degree of the prediction results of the water accumulation process is an important indicator for evaluating the overall prediction performance of the model. To more

**TABLE 4. Prediction results of the GBDT regression prediction model.**

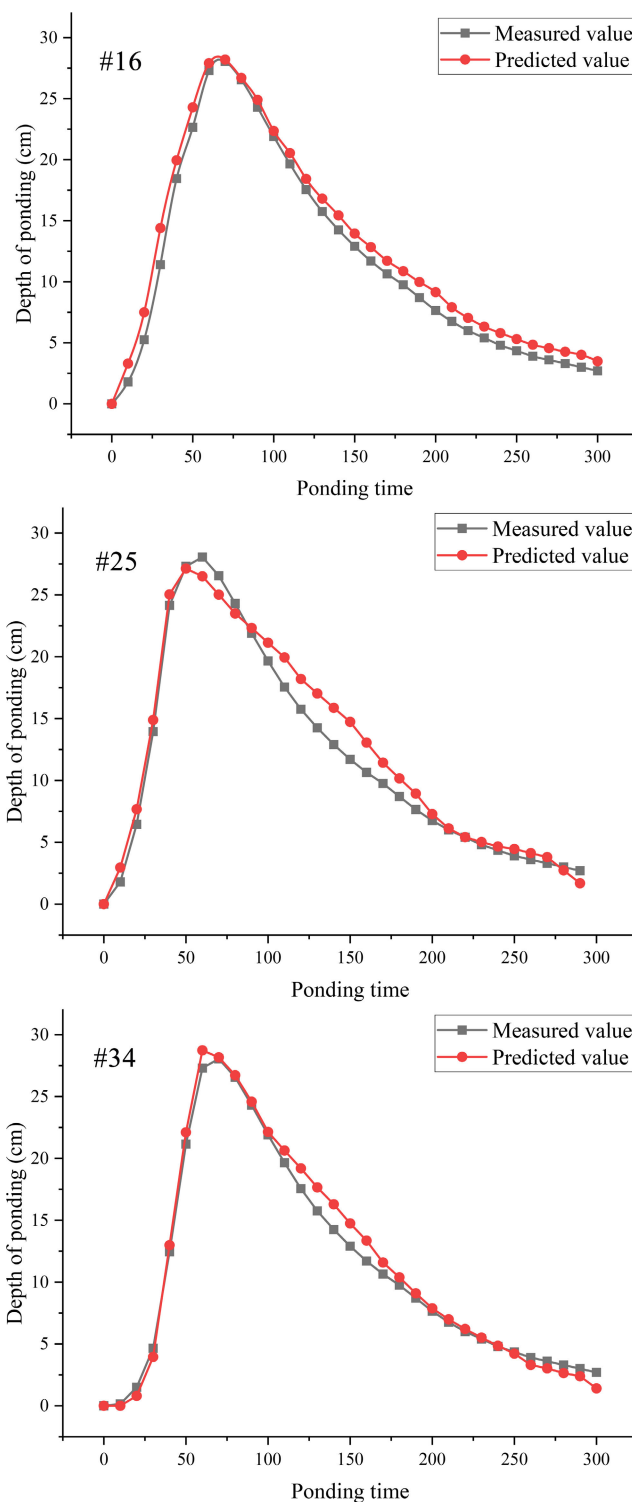
Data number	Prediction depth /cm	Measured depth /cm	Absolute error /cm	Relative error %
22731	0.00	0.00	0.00	0.00
22732	0.15	0.21	-0.06	28.57
22733	1.50	1.81	-0.31	17.13
22734	4.65	4.85	-0.20	4.12
...	/	/	/	/
27228	3.60	3.22	0.38	11.80
27229	3.30	2.75	0.55	20.00
27230	3.00	2.39	0.63	26.36
Mean value	10.82	10.34	1.30	19.77

comprehensively evaluate model performance in the prediction of water accumulation processes, the performance of prediction results of water accumulation process of each rainfall was analyzed quantitatively using QR, DC and AREPV (Table 5). As shown in Table 5, the QR is greater than 80%, the DC is 0.96, and the AREPV is 5.48%. These results demonstrate that the GBDT regression prediction model is feasible in the prediction of water accumulation processes.

**TABLE 5. Performances of the GBDT regression prediction model.**

The number of rainfall events	QR (%)	DC	AREPV
16	80.42	0.9716	8.90%
17	83.65	0.9602	3.84%
18	81.93	0.9607	3.41%
mean value	82.00	0.9642	5.48%

To evaluate the fitting effect of the model more intuitively, the fitting curve of predicted water accumulation and measured water accumulation processes of 3 water accumulation points (#16, #25, #34) was drawn using the random sampling method in this study. As shown in Fig. 7, as the rainfall prediction period is extended, the accuracy of the prediction results of the water accumulation process decreases slightly, but the errors are all within the acceptable range. Among them, the fitting degree of the prediction results performs best in the first 60 min, and the prediction accuracy of the GBDT model does not decrease significantly with the extension of the prediction period. This finding is mainly attributed to the fact that there is no residual accumulation in the training and prediction process of GBDT model, and the prediction results of the next period do not depend on the prediction results of the previous period. In addition, it is remarkable that the prediction results of peak rainfall of three water accumulation points are very close to the measured results, and the absolute errors are less than 2.5 cm, demonstrating that the GBDT regression prediction model exhibits good applicability for the prediction of water accumulation processes in a longer forecast period.



**FIGURE 7. Fitting curve of water accumulation process.**

**C. CONTRIBUTION ANALYSIS OF SENSITIVE FACTORS**

The contribution of sensitive factors was analyzed quantitatively using the GBDT model, which can understand the impact of different rainfall sensitive factors on the water accumulation. As shown in Figure 8, peak rainfall, rainfall, and concentration skewness exhibit the greatest impact on water

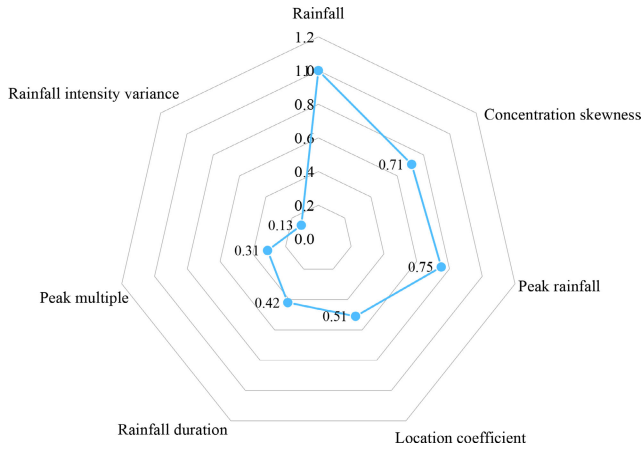


FIGURE 8. Importance of flood sensitive factors based on GBDT model.

accumulation, indicating that urban floods are more sensitive to heavy rainfall. This phenomenon is explained by the fact that the development of urbanization has gradually increased impervious areas (such as roads and buildings). When heavy rainfall reaches the surface, it converges quickly to form water accumulation, which increases the risk of urban floods. Therefore, urban flood forecasting and disaster prevention should pay special attention to short-term heavy rainfall and extreme rainfall events. In addition, it is necessary to reduce impervious surfaces and the speed of rainfall confluence and improve the ability of urban flood control systems to respond to heavy rainfall events.

TABLE 6. Comparison of prediction accuracy of different models.

Prediction step size	RMSE(m)			GBDT
	BPNN	Elman NN	NARX	
10min-ahead forecast	0.10	0.09	0.09	0.12
20min-ahead forecast	0.17	0.15	0.16	0.13
30min-ahead forecast	0.18	0.18	0.18	0.09
40min-ahead forecast	0.21	0.17	0.18	0.08
50min-ahead forecast	0.22	0.20	0.19	0.12
60min-ahead forecast	0.23	0.22	0.19	0.11

D. COMPARISON WITH OTHER MODELS

Since time series models [49] are commonly used methods for urban flood process prediction. Therefore, in order to verify the effectiveness of the GBDT model, we compared the prediction performance of the GBDT model with the three time series models. As shown in Table 6, the prediction accuracy of the GBDT model is not as good as the time series model in a short forecast period, but with the increase of the forecast period, the prediction accuracy of the GBDT model is significantly better than the time series model, which indicates the effectiveness and applicability of the GBDT model in water accumulation process prediction.

VIII. DISCUSSION

In this study, an index combination method suitable for predicting the depth of water accumulation is proposed. Among them, we propose a new sensitivity index (concentration skewness) for depth prediction of accumulated water. As shown in Figure 6, the prediction accuracy of the new index (scheme 1) for the existing index combination scheme (scheme 2) was improved by 10.11%, which verifies its effectiveness and applicability for predicting the depth of water accumulation. In addition, comparisons of the accuracy of 16 index combination schemes revealed that scheme 1 exhibits the highest prediction accuracy for ponding depth. On this basis, the GBDT algorithm is used to construct the prediction model of ponding process, aiming to propose a prediction model suitable for long-term periods.

In our previous study, we used the GBDT algorithm to build a prediction model for the depth of water accumulation. The average relative error of the model for predicting the maximum depth of water accumulation is 11.52% [15]. Although the previous study only provided the maximum depth of water accumulation and not the time distribution, it achieved the first step for early warning purposes. Based on this model, this study splits, reorganizes and calculates the index value of the rainfall accumulation data, which changes the traditional modeling method, and constructs a prediction model of the accumulation process suitable for a longer encounter period by the GBDT algorithm. The average relative error of this model for predicting the maximum depth of stagnant water is 5.48%, representing an improved effect of predicting the depth of water accumulation.

Research on the time distribution of ponding depth is a trend and a major challenge for flood warning. Serval studies have assessed multistep-ahead flood forecasts using time series methods [52], [63]. Chang et al. [49] used the time series method of three neural network models to predict the water level in the floodwater storage pond for 10- to 60-min-ahead forecasts. The results show that the three types of neural networks exhibit good accuracy in one-step-ahead forecast, among which the nonlinear autoregressive with exogenous input (NARX) network has the best prediction effect. However, due to the error accumulation in the multistep prediction of time series, when the prediction period changes from 10 to 60 min, the prediction accuracy of the NARX network gradually decreases, and the root mean square error increases from 0.09 to 0.19. Compared with the study of Chang et al., the prediction accuracy of the new modeling method proposed in this study does not significantly decrease as the prediction period is increased to 180 min in advance (Fig. 7), which indicates that our new modeling method can complement the time series model.

However, this study also encounters some limitations. We build a relationship model between rainfall and water accumulation for each accumulation point. However, for each accumulation point, the relationship model between rainfall and accumulation is not constant. If the underlying surface

conditions change, such as urban subway, pipe gallery and other engineering construction, the rainfall and the water accumulation model of the accumulation point may significantly change. These changes will reduce the accuracy of the model prediction results, which limits the use of the model.

Regarding future work, it should not be limited to using historical rainfall and water accumulation data to build a prediction model of the water accumulation process. With the continuous construction and maintenance of the city, the construction of the project will affect the drainage process of the accumulation point, and the rainfall and water accumulation process of the accumulation point will also change. Therefore, future work will focus on real-time collection of rainfall and water accumulation data to study how to use existing data to update and modify the model in real time.

## IX. CONCLUSION

In this study, based on the selection of sensitive indicators and index combination schemes that were suitable for the depth prediction of the water accumulation, a prediction model for the water accumulation process of the urban flood water accumulation point was constructed using the deep learning algorithm (GBDT). In the process of index scheme selections, by comparing the prediction accuracy of water accumulation depth of different index combination schemes, scheme 1 (i.e., rainfall, rainfall duration, peak rainfall, location coefficient, rainfall intensity variance, peak multiple, and concentration skewness) was identified as a suitable index scheme for water depth prediction. A more mature deep learning algorithm (GBDT) was used to build a prediction model of the water accumulation process of the water accumulation point. The average relative error of the model for the water accumulation process prediction was 19.77%, and the pass rate was 82.00%. These findings demonstrate the validity of the model for the predicting the water accumulation process of the water accumulation point. In addition, the GBDT model was used to quantitatively analyze the contribution of different sensitivity indicators to water accumulation in this study. Rainfall, peak rainfall and concentration skewness were identified as important factors affecting urban flooding. These research results provide effective technical support for urban flood control and forecasting.

This research also has some further directions that should be explored. For instance, in this study, only 50 water accumulation points in the study area were included in the prediction model of water accumulation processes. In the future, a water accumulation process prediction model should be constructed for the entire city to achieve real-time prediction of water accumulation processes.

## ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments. They declare that there is no conflict of interest regarding the publication of this paper.

## REFERENCES

- [1] X. Wang, G. Huang, and J. Liu, "Projected increases in intensity and frequency of rainfall extremes through a regional climate modeling approach," *J. Geophys. Res.*, vol. 119, no. 23, pp. 13271–13286, Dec. 2014.
- [2] *Climate Change 2014 Synthesis Report*. Accessed: Jun. 6, 2020. [Online]. Available: [https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/SYR\\_AR5\\_FINAL\\_full\\_zh.pdf](https://archive.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_zh.pdf)
- [3] L. Ling, Z. Yusop, and M. F. Chow, "Urban flood depth estimate with a new calibrated curve number runoff prediction model," *IEEE Access*, vol. 8, pp. 10915–10923, Jan. 2020.
- [4] M. Nayeib Yazdi, M. Ketabchy, D. J. Sample, D. Scott, and H. Liao, "An evaluation of HSPF and SWMM for simulating streamflow regimes in an urban watershed," *Environ. Model. Softw.*, vol. 118, pp. 211–225, Aug. 2019.
- [5] H. C. Winsemius, J. C. J. H. Aerts, L. P. H. van Beek, M. F. P. Bierkens, A. Bouwman, B. Jongman, J. C. J. Kwadijk, W. Ligtvoet, P. L. Lucas, D. P. van Vuuren, and P. J. Ward, "Global drivers of future river flood risk," *Nature Climate Change*, vol. 6, no. 4, pp. 381–385, Dec. 2015.
- [6] L. Alfieri, L. Feyen, and G. Di Baldassarre, "Increasing flood risk under climate change: A pan-European assessment of the benefits of four adaptation strategies," *Climatic Change*, vol. 136, nos. 3–4, pp. 507–521, Mar. 2016.
- [7] O. M. Rezende, A. B. R. da Cruz de Franco, A. K. B. de Oliveira, A. C. P. Jacob, and M. G. Miguez, "A framework to introduce urban flood resilience into the design of flood control alternatives," *J. Hydrol.*, vol. 576, pp. 478–493, Sep. 2019.
- [8] H. Wang, C. Mei, J. Liu, and W. Shao, "A new strategy for integrated urban water management in China: Sponge city," *Sci. China Technol. Sci.*, vol. 61, no. 3, pp. 317–329, Jan. 2018.
- [9] U. C. Nkwunono, M. Whitworth, and B. Baily, "A review of the current status of flood modelling for urban flood risk management in the developing countries," *Sci. Afr.*, vol. 7, Mar. 2020, Art. no. e00269.
- [10] N. Singh and S. R. Mohanty, "Short term price forecasting using adaptive generalized neuron model," *Int. J. Ambient Comput. Intell.*, vol. 9, no. 3, pp. 419–428, Jul. 2018.
- [11] S. Gupta, M. Khosravy, N. Gupta, H. Darbari, and N. Patel, "Hydraulic system onboard monitoring and fault diagnostic in agricultural machine," *Brazilian Arch. Biol. Technol.*, vol. 62, pp. 1678–4324, Oct. 2019.
- [12] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse, "Methodology for long-term prediction of time series," *Neurocomputing*, vol. 70, nos. 16–18, pp. 2861–2869, Oct. 2007.
- [13] C. Hu, L. Bi, Z. Piao, C. Wen, and L. Hou, "Coordinative optimization control of microgrid based on model predictive control," *Int. J. Ambient Comput. Intell.*, vol. 9, no. 3, pp. 57–75, Jul. 2018.
- [14] C. Ding, X. Cao, and C. Liu, "How does the station-area built environment influence metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds," *J. Transp. Geography*, vol. 77, pp. 70–78, May 2019.
- [15] Z. Wu, Y. Zhou, H. Wang, and Z. Jiang, "Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse," *Sci. Total Environ.*, vol. 716, May 2020, Art. no. 137077.
- [16] A. Mukherjee, P. Mukherjee, N. Dey, D. De, and B. K. Panigrahi, "Lightweight sustainable intelligent load forecasting platform for smart grid applications," *Sustain. Comput., Informat. Syst.*, vol. 25, Mar. 2020, Art. no. 100356.
- [17] S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A. S. Ashour, and V. E. Balas, "Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2005–2016, Aug. 2017.
- [18] S. Hore, S. Chatterjee, S. Sarkar, N. Dey, A. S. Ashour, D. Balas-Timar, and V. E. Balas, "Neural-based prediction of structural failure of multistoried RC buildings," *Struct. Eng. Mech.*, vol. 58, no. 3, pp. 459–473, May 2016.
- [19] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [20] S. Deng, C. Wang, M. Wang, and Z. Sun, "A gradient boosting decision tree approach for insider trading identification: An empirical model evaluation of China stock market," *Appl. Soft Comput.*, vol. 83, Oct. 2019, Art. no. 105652.
- [21] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Sol. Energy*, vol. 150, pp. 423–436, Jul. 2017.



- [22] V. A. Dev and M. R. Eden, "Formation lithology classification using scalable gradient boosted decision trees," *Comput. Chem. Eng.*, vol. 128, pp. 392–404, Sep. 2019.
- [23] L. A. Treinish, A. P. Praino, J. P. Cipriani, U. T. Mello, K. Mantripragada, L. C. V. Real, P. A. Sesini, V. Saxena, T. George, and R. Mittal, "Enabling high-resolution forecasting of severe weather and flooding events in Rio de Janeiro," *IBM J. Res. Develop.*, vol. 57, no. 5, pp. 7:1–7:11, Sep. 2013.
- [24] J. Noymanee and T. Theeramunkong, "Flood forecasting with machine learning technique on hydrological modeling," *Procedia Comput. Sci.*, vol. 156, pp. 377–386, Jun. 2019.
- [25] J. Gironás, L. A. Roesner, L. A. Rossman, and J. Davis, "A new applications manual for the storm water management model (SWMM)," *Environ. Model. Softw.*, vol. 25, no. 6, pp. 813–814, Jun. 2010.
- [26] M. Niazi, C. Nietch, M. Maghrebi, N. Jackson, B. R. Bennett, M. Tryby, and A. Massoudieh, "Storm water management model: Performance review and gap analysis," *J. Sustain. Water Built Environ.*, vol. 3, no. 2, May 2017, Art. no. 04017002.
- [27] M.-A. Sandu and A. Virsta, "Applicability of MIKE SHE to simulate hydrology in Argesel river catchment," *Agricult. Agricult. Sci. Procedia*, vol. 6, pp. 517–524, Sep. 2015.
- [28] C. Doulgeris, P. Georgiou, D. Papadimos, and D. Papamichail, "Ecosystem approach to water resources management using the MIKE 11 modeling system in the Strymonas river and lake Kerkini," *J. Environ. Manage.*, vol. 94, no. 1, pp. 132–143, Feb. 2012.
- [29] J. J. Warwick and J. S. Wilson, "Estimating uncertainty of stormwater runoff computations," *J. Water Resour. Planning Manage.*, vol. 116, no. 2, pp. 187–204, Mar. 1990.
- [30] A. M. Rabori and R. Ghazavi, "Urban flood estimation and evaluation of the performance of an urban drainage system in a semi-arid urban area using SWMM," *Water Environ. Res.*, vol. 90, no. 12, pp. 2075–2082, Dec. 2018.
- [31] M. Shahed Behrouz, Z. Zhu, L. S. Matott, and A. J. Rabideau, "A new tool for automatic calibration of the storm water management model (SWMM)," *J. Hydrol.*, vol. 581, Feb. 2020, Art. no. 124436.
- [32] Q. Zhou, G. Leng, J. Su, and Y. Ren, "Comparison of urbanization and climate change impacts on urban flood volumes: Importance of urban planning and drainage adaptation," *Sci. Total Environ.*, vol. 658, pp. 24–33, Mar. 2019.
- [33] H. Zhu, M. Yu, J. Zhu, H. Lu, and R. Cao, "Simulation study on effect of permeable pavement on reducing flood risk of urban runoff," *Int. J. Transp. Sci. Technol.*, vol. 8, no. 4, pp. 373–382, Dec. 2019.
- [34] J. Kim and H. Cho, "Scenario-based urban flood forecast with flood inundation map," *Tropical Cyclone Res. Rev.*, vol. 8, no. 1, pp. 27–34, Mar. 2019.
- [35] S. Babaei, R. Ghazavi, and M. Erfanian, "Urban flood simulation and prioritization of critical urban sub-catchments using SWMM model and PROMETHEE II approach," *Phys. Chem. Earth*, vol. 105, pp. 3–11, Jun. 2018.
- [36] J. Chen, A. A. Hill, and L. D. Urbano, "A GIS-based model for urban flood inundation," *J. Hydrol.*, vol. 373, nos. 1–2, pp. 184–192, Jun. 2009.
- [37] J. Liu, W. Shao, C. Xiang, C. Mei, and Z. Li, "Uncertainties of urban flood modeling: Influence of parameters for different underlying surfaces," *Environ. Res.*, vol. 182, Mar. 2020, Art. no. 108929.
- [38] Z. Li, L. Wu, W. Zhu, M. Hou, Y. Yang, and J. Zheng, "A new method for urban storm flood inundation simulation with fine CD-TIN surface," *Water*, vol. 6, no. 5, pp. 1151–1171, Jan. 2014.
- [39] Z. Afifi, H.-J. Chu, Y.-L. Kuo, Y.-C. Hsu, H.-K. Wong, and M. Z. Ali, "Residential flood loss assessment and risk mapping from high-resolution simulation," *Water*, vol. 11, no. 4, Apr. 2019, Art. no. 751.
- [40] E. Salvatore, J. Bronders, and O. Batelaan, "Hydrological modelling of urbanized catchments: A review and future directions," *J. Hydrol.*, vol. 529, no. 1, pp. 62–81, Oct. 2015.
- [41] P. Costabile, C. Costanzo, G. De Lorenzo, and F. Macchione, "Is local flood hazard assessment in urban areas significantly influenced by the physical complexity of the hydrodynamic inundation model?" *J. Hydrol.*, vol. 580, Jan. 2020, Art. no. 124231.
- [42] F. Macchione, P. Costabile, C. Costanzo, and G. De Lorenzo, "Extracting quantitative data from non-conventional information for the hydraulic reconstruction of past urban flood events. A case study," *J. Hydrol.*, vol. 576, pp. 443–465, Sep. 2019.
- [43] D. Molinari, K. M. De Bruijn, J. T. Castillo-Rodríguez, G. T. Aronica, and L. M. Bouwer, "Validation of flood risk models: Current practice and possible improvements(Review)," *Int. J. Disaster Risk Reduct.*, vol. 33, pp. 441–448, Feb. 2019.
- [44] J. Leandro, A. Schumann, and A. Pfister, "A step towards considering the spatial heterogeneity of urban key features in urban hydrology flood modelling," *J. Hydrol.*, vol. 535, pp. 356–365, Apr. 2016.
- [45] C. Chen, Q. Hui, Q. Pei, Y. Zhou, B. Wang, N. Lv, and J. Li, "CRML: A convolution regression model with machine learning for hydrology forecasting," *IEEE Access*, vol. 7, pp. 133839–133849, Sep. 2019.
- [46] A. Bomers, B. Meulen, R. M. J. Schielen, and S. J. M. H. Hulscher, "Historic flood reconstruction with the use of an artificial neural network," *Water Resour. Res.*, vol. 55, no. 11, pp. 9673–9688, Nov. 2019.
- [47] X. H. Le, H. V. Ho, G. Lee, and S. Jung, "Application of long short-term memory (LSTM) neural network for flood forecasting," *Water*, vol. 11, no. 7, p. 1387, Jul. 2019.
- [48] Y.-M. Chiang, L.-C. Chang, M.-J. Tsai, Y.-F. Wang, and F.-J. Chang, "Dynamic neural networks for real-time water level predictions of sewerage systems-covering gauged and ungauged sites," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 7, pp. 1309–1319, Mar. 2010.
- [49] F.-J. Chang, P.-A. Chen, Y.-R. Lu, E. Huang, and K.-Y. Chang, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control," *J. Hydrol.*, vol. 517, pp. 836–846, Sep. 2014.
- [50] Y. A. Rjeily, O. Abbas, M. Sadek, I. Shahrou, and F. H. Chehade, "Flood forecasting within urban drainage systems using NARX neural network," *Water Sci. Technol.*, vol. 76, no. 9, pp. 2401–2412, Nov. 2017.
- [51] Y. Chen, R. Liu, D. Barrett, L. Gao, M. Zhou, L. Renzullo, and I. Emelyanova, "A spatial assessment framework for evaluating flood risk under extreme climates," *Sci. Total Environ.*, vol. 538, pp. 512–523, Dec. 2015.
- [52] S. Berkhahn, L. Fuchs, and I. Neuweiler, "An ensemble neural network model for real-time prediction of urban floods," *J. Hydrol.*, vol. 575, pp. 743–754, Aug. 2019.
- [53] W. H. M. Wan Mohtar, J. Abdullah, K. N. Abdul Maulud, and N. S. Muhammad, "Urban flash flood index based on historical rainfall events," *Sustain. Cities Soc.*, vol. 56, May 2020, Art. no. 102088.
- [54] W. Yuan, H. Song, and M. Liu, "An early warning model for flash floods based on random rainfall patterns," *Adv. Water Science.*, vol. 30, no. 4, pp. 515–527, Jul. 2019.
- [55] G. Fu and Z. Kapelan, "Flood analysis of urban drainage systems: Probabilistic dependence structure of rainfall characteristics and fuzzy model parameters," *J. Hydroinform.*, vol. 15, no. 3, pp. 687–699, Jul. 2013.
- [56] S.-C. Kao and R. S. Govindaraju, "Probabilistic structure of storm surface runoff considering the dependence between average intensity and storm duration of rainfall events," *Water Resour. Res.*, vol. 43, no. 6, Jun. 2007, Art. no. W06410.
- [57] P. M. Atkinson and R. Massari, "Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy," *Comput. Geosci.*, vol. 24, no. 4, pp. 373–385, May 1998.
- [58] S.-B. Bai, J. Wang, G.-N. Lü, P.-G. Zhou, S.-S. Hou, and S.-N. Xu, "GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China," *Geomorphology*, vol. 115, nos. 1–2, pp. 23–31, Feb. 2010.
- [59] C. Conoscenti, S. Angileri, C. Cappadonia, E. Rotigliano, V. Agnesi, and M. Märker, "Gully erosion susceptibility assessment by means of GIS-based logistic regression: A case of Sicily (Italy)," *Geomorphology*, vol. 204, pp. 399–411, Jan. 2014.
- [60] J. Cheng, G. Li, and X. Chen, "Research on travel time prediction model of freeway based on gradient boosting decision tree," *IEEE Access*, vol. 7, pp. 7466–7480, Dec. 2019.
- [61] C. Ding, X. Cao, and P. Næss, "Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo," *Transp. Res. A, Policy Pract.*, vol. 110, pp. 107–117, Apr. 2018.
- [62] B. Zhang, J. Ren, Y. Cheng, B. Wang, and Z. Wei, "Health data driven on continuous blood pressure prediction based on gradient boosting decision tree algorithm," *IEEE Access*, vol. 7, pp. 32423–32433, Mar. 2019.
- [63] P.-A. Chen, L.-C. Chang, and F.-J. Chang, "Reinforced recurrent neural networks for multi-step-ahead flood forecasts," *J. Hydrol.*, vol. 497, pp. 71–79, Aug. 2013.





**ZENING WU** was born in Henan, China, in 1962. He received the M.E. degree in water conservancy and hydropower engineering construction from the Zhengzhou Institute of Technology in 1983 and the Ph.D. degree in engineering from Hehai University in 2004. Since 1983, he has been engaged in teaching and scientific research with the Department of Water Conservancy Engineering, School of Water Conservancy and Environmental Engineering, Zhengzhou University of Technology, and the School of Water Conservancy and Environment, Zhengzhou University. He has published more than 90 articles and eight academic books. His research interests include the optimal allocation and planning management of water resources, flood disaster risk, water resource economy, and the optimal scheduling of water engineering systems. He is the academic and technological leader of Henan province, and has successively won the First Prize of Dayu Water Conservancy Science and Technology and the Science and Technology Progress of Henan Province.



**YIHONG ZHOU** was born in Anhui, China, in 1996. He received the B.E. degree in geological engineering from Henan Polytechnic University in 2018. Since 2018, he has been studying with the College of Water Conservancy Engineering, Zhengzhou University. He has published an article in *Science of the Total Environment* in January 2020. His research interests include the application of deep learning in urban flood simulation and waterlogging prediction. He won the National Scholarship and the National Encouragement Scholarship, in 2015 and 2016, respectively.



**HUILIANG WANG** was born in Anyang, China, in 1982. He received the M.S. degree in physical geography from the Institute of Measurement and Geophysics, Chinese Academy of Sciences, in 2010, and the Ph.D. degree in ecology from the Center for Ecological Environment Research, Chinese Academy of Sciences, in 2014. Since 2009, he has been an Associate Professor and a Doctoral Supervisor with the College of Water Conservancy Engineering, Zhengzhou University. In recent five years, he has published 19 articles as a corresponding author or first author. His research interests include hydrological and water resource models and numerical simulation, and water ecological restoration. Dr. Wang is a member of the special committee of ecological environment of the China Dam Engineering Society. In 2018, he won the First Prize of Science and Technology Progress of Water Conservancy of Henan Province, and he was rated as the advanced individual in the graduate education of national water conservancy engineering degree in 2019.

...