



A Framework on Fast Mapping of Urban Flood Based on a Multi-Objective Random Forest Model

Yaoxing Liao¹ · Zhaoli Wang^{1,2} · Chengguang Lai^{1,2} · Chong-Yu Xu³

Accepted: 13 March 2023 / Published online: 11 April 2023
© The Author(s) 2023

Abstract

Fast and accurate prediction of urban flood is of considerable practical importance to mitigate the effects of frequent flood disasters in advance. To improve urban flood prediction efficiency and accuracy, we proposed a framework for fast mapping of urban flood: a coupled model based on physical mechanisms was first constructed, a rainfall-inundation database was generated, and a hybrid flood mapping model was finally proposed using the multi-objective random forest (MORF) method. The results show that the coupled model had good reliability in modelling urban flood, and 48 rainfall-inundation scenarios were then specified. The proposed hybrid MORF model in the framework also demonstrated good performance in predicting inundated depth under the observed and scenario rainfall events. The spatial inundated depths predicted by the MORF model were close to those of the coupled model, with differences typically less than 0.1 m and an average correlation coefficient reaching 0.951. The MORF model, however, achieved a computational speed of 200 times faster than the coupled model. The overall prediction performance of the MORF model was also better than that of the k-nearest neighbor model. Our research provides a novel approach to rapid urban flood mapping and flood early warning.

Keywords Coupled model · Designed rainfall events · Multi-objective random forest (MORF) method · Rainfall-inundation database · Urban flood prediction

1 Introduction

Urban flood disasters are becoming increasingly more frequent against the background of climate change and growing urbanization (Wu et al. 2017; IPCC 2021; Jian et al. 2021; Tellman et al. 2021). As population size and property value continue to expand in cities, damage and loss caused by floods, even if the floods as hazard remain on the same scale as previously encountered, will become more severe than ever before (Lai et al. 2020; Deng et al. 2022). Prediction of the flooding situation is an extremely effective measure for disaster prevention and mitigation in urban areas (Lin et al. 2020). People can take preventive measures after receiving

the information from early-warning and prediction systems, and then the exposure and vulnerability of elements at risk may be greatly reduced.

However, there are two major challenges for urban flood prediction (Teng et al. 2019; Bentivoglio et al. 2022). The first is the timeliness of prediction. Because an urban flood can develop within a few minutes, the time spent on prediction should be as short as possible, and must guarantee enough reaction time for people and relevant government departments to respond. The second challenge is the accuracy of prediction. Accurate prediction results are vital for disaster prevention. Targeted measures that reduce losses can be adopted if the inundated area and depth, occurrence time, and persistent period of a flood are accurately predicted. Therefore, how to improve the timeliness and accuracy of urban flood prediction has become a focus in current studies and urgently needs to be addressed for the real-time or quasi real-time prediction of urban flood.

Hydrological and hydrodynamic models are often applied to simulate and predict the physical processes of urban flood (Chen et al. 2017; Wu et al. 2017; Zeng et al. 2022). Commonly used models include one-dimensional (1D) models

✉ Chengguang Lai
laichg@scut.edu.cn

¹ State Key Laboratory of Subtropical Building Science, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China

² Pazhou Lab, Guangzhou 510335, China

³ Department of Geosciences, University of Oslo, 0316 Oslo, Norway

that are usually solved using the Saint-Venant equations of mass and momentum conservation (Rossman 2015), two-dimensional (2D) models that use the shallow water equations or cellular automata (Guidolin et al. 2016; Zhang et al. 2021), and coupled models that combine 1D and 2D models (Wu et al. 2017; Chen et al. 2018; Wu et al. 2018). These models provide a sufficient representation of flow movement, achieving accurate simulation of urban flood. Yet existing models suffer from problems, such as complex iterative solutions of hydrodynamic equations and excessive simulation time, especially when running on high-resolution grids, even though computing power has greatly improved in recent years (Bhola et al. 2018; Kim and Cho 2019). Under these circumstances, the conventionally physics-based models are difficult to meet the requirements of timeliness at present when they are directly applied to the real-time rapid prediction of urban flood or used in a context that requires a large number of operating results (Teng et al. 2019; Chu et al. 2020; Kabir et al. 2020).

There have been many exploratory developments of novel methods of rapid prediction and simulation of urban floods. Some studies have used digital elevation model (DEM)-based, simplified hydrodynamic models, which are orders of magnitude faster than traditional hydrodynamic models (McGrath et al. 2018; Teng et al. 2019). These models have deficiencies, however, because they cannot simulate dynamic changes of water depth and flow velocity. Others use rainfall-inundation databases generated by hydrodynamic models to extract or call up the flood inundation map in advance (Schulz et al. 2015; Bhola et al. 2018). The speed of this approach is highly advantageous, but direct extraction of an inundation map may lead to unacceptable errors (Bhola et al. 2018; Kim and Cho 2019).

Many studies have verified that machine learning (ML) can provide an alternative and effective solution for urban flood mapping and prediction without directly considering the physical processes involved (Lin et al. 2021; Zouneimat-Kermani et al. 2021). Relative to traditional hydrodynamic models, a ML-based approach has some distinct advantages. First, benefitting from its data-driven characteristics, a ML-based approach can quickly learn the relationship between hydrological elements in historical flooding data generated by traditional physical models (Xu and Liang 2021). Furthermore, it can combine common feature data to form more abstract attributes, or higher-level feature representations, giving a more detailed description of hydrological and hydrodynamic phenomena (Lai et al. 2016; Brunton et al. 2019). Given these advantages, many ML-based methods have been widely applied to flood hazard, susceptibility, and risk analysis (Wang et al. 2015; Lai et al. 2016; Li et al. 2020; Lin et al. 2021). Most of these studies used ML methods to predict the occurrence probability of flooding, or to perform 1D runoff prediction, but few studies

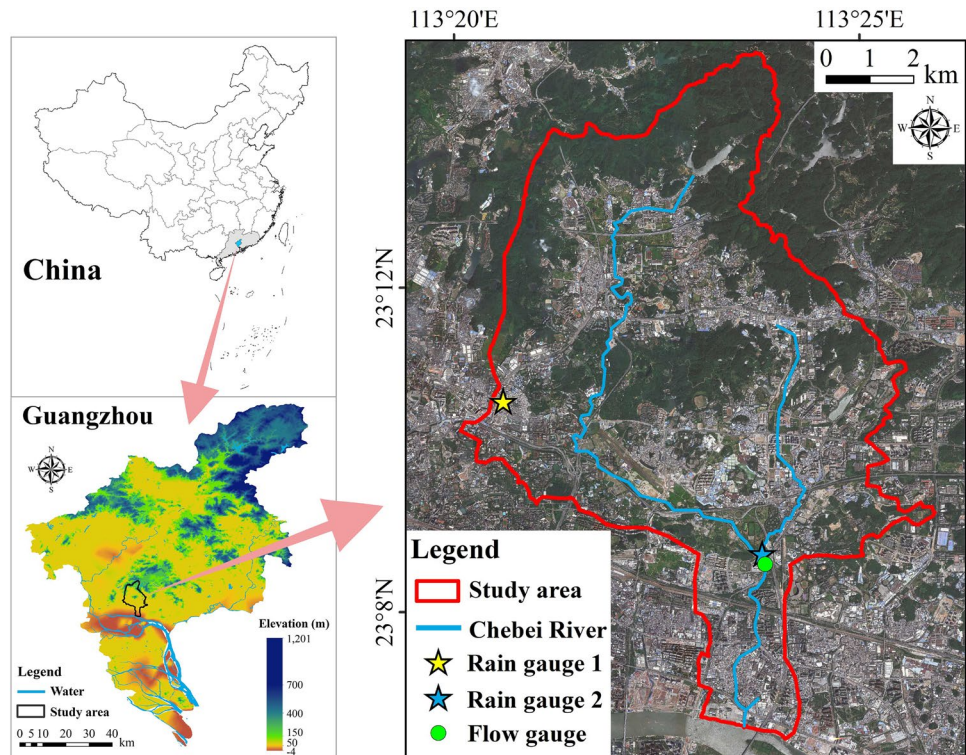
have used such methods to conduct 2D flood inundation prediction (Kabir et al. 2020).

Some studies have explored combinations of ML with hydrological and hydrodynamic models to realize 2D flood prediction, including support vector machine (SVM) (Bermúdez et al. 2019) and artificial neural network (ANN) (Kabir et al. 2020). Machine learning-based 2D flood inundation prediction can generally be carried out using two kinds of approaches. One is the two-stage method, which first predicts some water depth points and then expands these to the 2D space (Lin et al. 2013; Jhong et al. 2017). However, there can be significant errors when expanding the water depth of partial points into spatial inundated depths. Another method is to directly predict the water depth on a spatial grid. For example, Chu et al. (2020) and Lin et al. (2020) constructed 14,227 and 10,000 ANN models with more than 10,000 inundated grid units each in their study areas for rapid prediction of inundated depths. Due to the large number of models and the complex weight optimization and adjustment needed, however, building a large number of neural network models to predict water inundation depths across multiple grids encounters problems of complex parameters, slow training speeds, and the consumption of too much computational time. Hou et al. (2021) attempted to improve these shortcomings by using a random forest (RF) approach, but this was only applied to rainfall data generated by the Chicago¹ rainfall pattern generator, and the method was not verified at the watershed scale using real data.

When selecting ML technology for flood prediction, it is worth noting that most existing algorithms (such as SVM) are not suitable for multi-objective scenarios (Kabir et al. 2020), that is, predicting flood variables (such as depth) in multiple units through a single model. Generally, the prediction of water depth at one point can be solved satisfactorily by constructing a single model using a single-objective regression method (Xu et al. 2020). However, spatial prediction of inundated depths is a multi-objective problem. Multi-objective prediction is the generalization of multi-target regression and classification (Nakano et al. 2022). In this regard, a RF method may be more appropriate in theory as it can solve both single-objective and multi-objective problems. Multi-objective random forest (MORF) does not over-fit the training data, has lower sensitivity to noise in the training sample, and can efficiently process high-dimensional data, high-order interactions, and nonlinear problems of variables compared with other algorithms, such as linear or logistic regressions (Breiman 2001). Compared with other multi-objective algorithms like convolutional neural

¹ A single-peaked rainfall pattern with high peak intensity, proposed by Keifer and Chu (1957).

Fig. 1 Map of the Chebei River Basin case study area in Guangzhou City, China



networks (Kabir et al. 2020), MORF has fewer parameters, avoiding a large number of parameter settings, weight optimizations, and long training times. Recently, MORF methods have made significant progress in solving multi-objective problems in many fields, such as shale gas production forecasting (Xue et al. 2021), vegetation condition prediction (Kocev et al. 2009), and earthquake early warning (Adhaityar et al. 2021). Multi-objective random forest models should also be suitable for 2D flood inundation prediction, providing rapid computational speed and high spatial resolution in theory, but few applications have so far been reported. Hence, the performance of the MORF models that attempt to simulate and predict urban flood requires further discussion.

In summary, data-driven ML methods have lower computational cost and higher efficiency, and possess great potential for overcoming the shortcomings of traditional hydrodynamic models. This study aimed to develop an effective hybrid framework, which combined the advantages of the hydrological model, hydrodynamic model, and ML methods to achieve rapid and accurate prediction of urban flood. The main goal was achieved through: (1) constructing a high-precision hydrological and hydrodynamic coupled model of urban flood; (2) producing a rainfall-inundation database using different designed rainfall events; and (3) developing a prediction model based on MORF to realize rapid and high-precision mapping of urban flood. The study contributes to the advancement in risk management of urban flood disasters and provides a methodological device and analytical

framework for rapid forecasting of urban flood, disaster prevention and mitigation, and flood-related risk management in urban areas.

2 Data and Methods

This section presents an overview of the study area, the main sources of data employed in the project, and the relevant methodology used in this study.

2.1 Study Area

The Chebei River Basin (CRB), with an area of 74 km², is one of the severest flood-prone areas in Guangzhou City, China (Fig. 1). It has a subtropical monsoon climate with an average annual temperature of 20–22 °C and an average annual rainfall of about 1720 mm. The northern and central parts of the CRB are mostly mountainous areas, and the southern part is a relatively low-lying and highly urbanized area. With rapid urbanization development, the areal coverage of impervious surfaces in the Tianhe District, where the CRB is located, increased from 16% in 1990 to 71% in 2013 (Pan et al. 2017).

Due to factors such as heavy rainfall, low-lying terrain, and insufficient drainage capacity, part of the CRB often experiences rainstorm and flood disasters, and there are many flood-prone areas in the basin. For example, the

extreme rainfall in Guangzhou on 22 May 2020 caused severe flooding in the city, resulting in four deaths. During this period, the water level of the Chebei River rose sharply, the maximum water depth over several highways exceeded 1 m, and a large number of buildings were partially inundated, resulting in significant economic losses (Zhang et al. 2021).

2.2 Data

The digital elevation model (DEM) data (floating-point type) with a spatial resolution of $8\text{m} \times 8\text{m}$, used to construct the hydrological and hydrodynamic coupled model were obtained from the Guangzhou Land Resources and Planning Commission. To characterize the blocking effect of buildings on surface flow, Google satellite images were used to depict the outline of buildings, and the building heights of the original DEM were updated to a fixed value of 10 m.

Drainage network data and land-use data were obtained from the Guangzhou Water Affairs Bureau and the National Geographic Center of China, respectively. Hourly rainstorm data from 1954 to 2012 recorded at the Wushan Station near the CRB were provided by the Guangzhou Meteorological Bureau, and these data were used to generate designed rainfall events. The observed rainfall and flow data were used for model calibration and verification, and these were obtained using field monitoring instruments with a temporal resolution of 5 minutes. To process the data, ArcGIS 10.2, Python v.3.8.5, R v.3.5.1, and Microsoft Excel 2016 were the primary software used.

2.3 Methodology

The physical-based coupled model, ML model, method of rainfall design, and evaluation metrics are introduced in detail in this section.

2.3.1 Overall Research Strategy

The main purpose of this study is to create a framework that can predict the maximum inundated area and depth, as well as to map the urban flood zones based on these prediction results. This framework mainly consists of three parts, which are depicted in Fig. 2.

First, a coupled model, based on the storm water management model (SWMM) and a weighted cellular automata 2D (WCA2D) inundation model, was constructed to simulate the inundated area and depth. A rainfall database was generated by identifying the temporal distribution of local historical rainfall events, and then a rainfall-inundation database was further established using the coupled model.

Second, the inundated and noninundated grids were determined with 0.01 m as the water depth threshold, and these water depth raster data were read and converted into

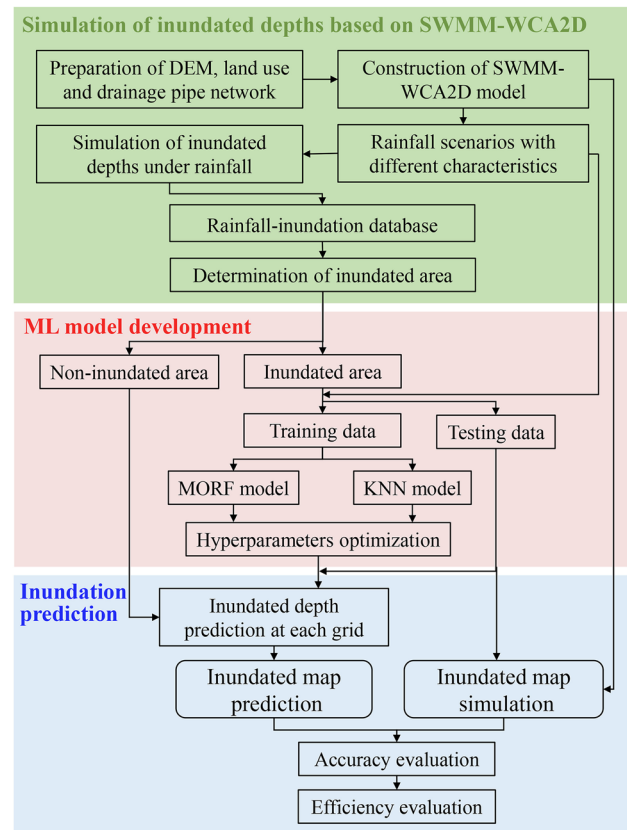


Fig. 2 A schematic summarizing the hybrid modeling approach for rapid flood mapping developed in this study based on the SWMM-WCA2D and ML models

a two-dimensional array and that became input into the ML model. The MORF and k-nearest neighbor (KNN) models were constructed by generating training and test datasets and optimizing hyperparameters.

Finally, flood inundation prediction was achieved by inputting rainfall into the constructed ML model. The effectiveness of ML model prediction, including accuracy and calculation time, was compared with the coupled model simulation.

2.3.2 SWMM-WCA2D Coupled Model

The SWMM provided a 1D model of the drainage network and was used to analyze the stormwater runoff in the drainage pipe system (Rossman 2015). This model is a popular tool for simulating urban hydrology and hydrodynamics using the 1D Saint-Venant equations, and includes modules for ground runoff generation, ground runoff convergence, and pipeline convergence (Wang et al. 2022). Since the SWMM lacks a surface overflow module, coupling SWMM with other 2D models enables rainfall inundation simulation.

The cellular automata (CA) model, first developed by Dottori and Todini (2011), can simulate 2D hydrodynamic processes. It utilizes the Manning equation to calculate interactive flows between cells. Guidolin et al. (2016) further introduced an interactive method to simplify flow conversion rules between cells, and proposed a WCA2D inundation model. The main advantage of the WCA2D model is that it is more efficient in flood simulation and can reflect certain physical mechanism of surface flow movement, and the model was used to simulate inundated area and depth in our study.

We coupled the SWMM and the WCA2D models to simulate the process of urban rainstorms and associated flooding by importing overflow information from the SWMM into the WCA2D model, and finally we simulated the evolution of inundation using a graphics processing unit (GPU) acceleration technology.

2.3.3 Rainfall Event Design

Designed rainfall events were used to generate a rainfall-inundation database for ML training. Rainfall event design that contains the features of local historical rainfall can improve the representativeness of the rainfall-inundation databases (Kim and Cho 2019). The characteristics of rainfall include return periods and rainfall patterns. The return period is related to the rainfall amount, while rainfall pattern refers to the temporal distribution of rainfall intensity during a rainfall event (He et al. 2022).

A total of seven rainfall pattern types were proposed by Molokov and Shtigorin (1956). These are pattern types I, II, and III (unimodal with early, late, and middle peaks, respectively), pattern type IV (uniform), and patterns V, VI, and VII (bimodal with side, early and late peaks, respectively).

In this study, we used a rainfall design method based on local historical rainfall that was proposed in our previous study (Zhang et al. 2021). First, the historical rainstorm events were identified and classified into seven rainfall patterns, the historical rainstorm process for each rainfall pattern was obtained, and then the temporal distribution proportion of different rainfall patterns was deduced. Finally, the total rainfall for the return period and distribution ratios were combined to generate rainfall events that accurately reflect the actual rainfall characteristics of the study area. The Chicago rainfall pattern, a widely used pattern in drainage planning and flood simulation, was also added to the rainfall database. In this case, a total of 48 rainfall scenarios (Fig. 3), each lasting 120 minutes, were designed, featuring eight rainfall patterns (I to VII, and Chicago) and six return periods of 1, 5, 10, 20, 50, and 100 years with total rainfall amounts of 65.88, 95.74, 109.90, 125.97, 144.82, and 157.73 mm, respectively. In this study, rainfall intensity varies over time and is assumed to be uniform spatially.

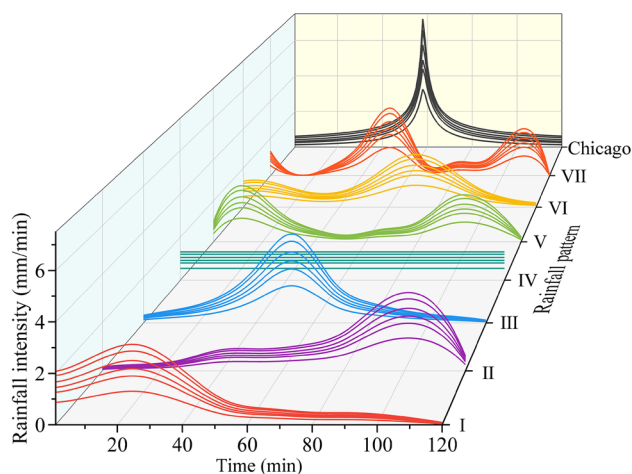


Fig. 3 Designed rainfall events consisting of six return periods (1, 5, 10, 20, 50, and 100 years) and eight rainfall patterns (I to VII, and Chicago)

2.3.4 Multi-objective Random Forest

A multi-objective random forest (MORF) algorithm was used for the rapid prediction of urban flood in this study. The implementation from single-objective to multi-objectives generally includes the problem transformation method and algorithm adaptation method (Borchani et al. 2015). The former method combines the predictions of each single-objective model, ignoring a certain spatial correlation between the individual objectives. A multi-objective model based on algorithm adaptation may have more advantages in improving the prediction accuracy of each spatial grid, because it predicts all targets simultaneously using a single model that captures all dependencies and internal relationships (Borchani et al. 2015; Ling et al. 2022).

Random forest is an ensemble of classification and regression trees (Breiman 2001). The traditional RF is typically employed to solve single objective problems (Xiong et al. 2020; Liao et al. 2021), which are based on univariate regression trees (URT). In URT, the impurity of a node is generally defined as the sum of squares of the single response variable values with respect to the node mean (Shih and Tsai 2004; Saha et al. 2016):

$$i(t) = \frac{1}{N_t} \sum_{y_k \in t} (y_k - \bar{y}(t))^2 \tag{1}$$

where $\bar{y}(t)$ is the mean value of the complete sample of y at node t , y_k is the observed value of the response variable, and N_t is the number of data points at node t .

Starting from a single node at the top of the tree (containing all data), the tree grows by repeatedly binary splitting the data. Splits are typically chosen to minimize the impurity

Fig. 4 Schematic diagram of the multi-objective regression tree (MRT) methodology

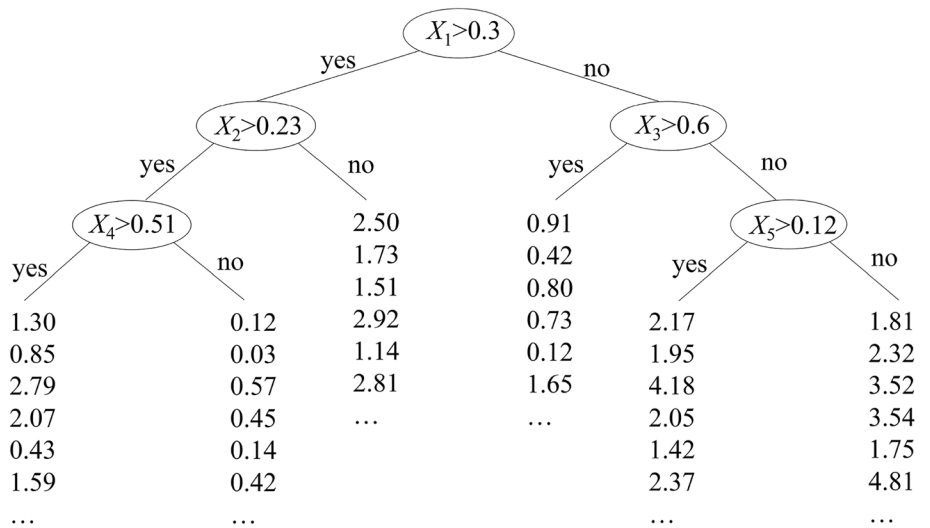
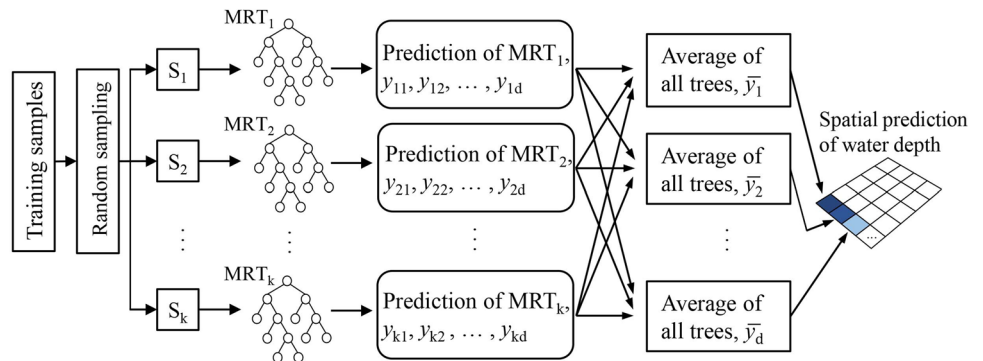


Fig. 5 Process of spatial water depth prediction using a multi-objective random forest (MORF) model Note: MRT = Multi-objective regression tree



of two nodes. Let a predictor variable X_p split the parent node t into two child nodes t_L and t_R at split point c ; the impurity reduction related to node t caused by the splitting of X_p , $\Delta i(c, t)_{X_p}$ is calculated as follows (Saha et al. 2016):

$$\Delta i(c, t)_{X_p} = i(t) - \frac{N_{t_L}}{N_t} i(t_L) - \frac{N_{t_R}}{N_t} i(t_R) \tag{2}$$

where N_t , N_{t_L} and N_{t_R} are the number of data points at node t , left child node t_L , and right child node t_R , respectively.

To realize the simultaneous output and prediction of multiple objectives, MORF was further proposed (Kocev et al. 2007; Xue et al. 2021), and its submodel becomes multi-objective regression trees (MRT) (De'ath 2002; Struyf and Džeroski 2006). When constructing the MRT, the training dataset D with N instances includes m predictor variables (X_1, \dots, X_m) and d response variables (Y_1, \dots, Y_d), that is, $D = \{(x^{(1)}, y^{(1)}), \dots, (x^{(N)}, y^{(N)})\}$. The construction process of MRT is similar to URT, that is, the univariate response of URT is replaced with multivariate response (De'ath 2002). Multi-objective regression trees

redefine the impurity of the node t_d as the sum of squared error, $i(t_d)$, summing the impurity over the multivariate response, as follows (Borchani et al. 2015):

$$i(t_d) = \sum_{l=1}^N \sum_{j=1}^d (y_j^{(l)} - \bar{y}_j)^2 \tag{3}$$

where $y_j^{(l)}$ represents the value of the output variable Y_j of the instance l , and \bar{y}_j represents the mean value of Y_j in the node. The splitting point is determined by selecting the minimizing sum of the squared error. Each leaf of the tree can be characterized by the multi-variable mean of its instances, the number of instances, and its feature values. The leaves of a MRT store vectors instead of individual values. Each component of this vector represents a prediction for one of the targets. To better understand the structure of MRT, an example of an MRT with five predictor variables and six targets is showed in Fig. 4. Finally, MORF can be constructed by a large number of MRT based on training sets through random sampling, and the spatial water depth of d grid cells can be predicted by MORF (Fig. 5).

2.3.5 K-Nearest Neighbor Algorithm

The k-nearest neighbor (KNN) algorithm is a nonparametric regression prediction case-based learning method in the field of data mining, and is a popular method to deal with multi-objective problems (Liu et al. 2019). In a KNN model, if a sample to be predicted is the most similar to the K samples in the training set, the results of the predicted sample can be determined based on these K samples. Similarity is defined in terms of a distance measure between two samples (each contains n data points). A popular choice is the Euclidean distance given by:

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2 + \dots + (x_n - x'_n)^2} \quad (4)$$

Using the nearest neighbor decision rule, an observation is assigned to the group to which most of its k th-nearest neighbors belong, and then the value of the dependent variable corresponding to the observation is predicted. The KNN regression algorithm was used to search the K samples closest to an unknown rainfall event (that is, predictor variables) in the rainfall-inundation database (that is, training set), and then we used these K samples for inundation prediction. The prediction results by KNN were later compared with those of MORF.

2.3.6 Model Performance Indicators

The Nash-Sutcliffe efficiency (NSE) coefficient was adopted to evaluate the SWMM model accuracy (Nash and Sutcliffe 1970), and it can be calculated by Eq. 5. The R^2 score was used to evaluate the training effect of the ML model and represents the proportion of variance in the model that has been explained by independent variables, which can be calculated by Eq. 6. The performance of the flood prediction model was evaluated by analyzing indices including mean absolute error (MAE), root mean square error ($RMSE$), and Pearson correlation coefficient (PCC). The MAE was used to determine the overall accuracy when estimating the mean value of the coupled model with the mean value of the ML model. The $RMSE$ was based on squared error and is suitable for assessing the performance of the ML method on large-scale flooding (Chu et al. 2020; Lin et al. 2020). The PCC was used to measure the consistency between the predicted water depth of ML and the results of the coupled model. The calculation of MAE , $RMSE$, and PCC was performed using Eqs. 7, 8, and 9, respectively:

$$NSE = 1 - \frac{\sum_{t=1}^T [Q_{sim}(t) - Q_{obs}(t)]^2}{\sum_{t=1}^T [Q_{obs}(t) - \bar{Q}_{obs}]^2} \quad (5)$$

where T is the total number of time steps, $Q_{sim}(t)$ is the simulated discharge at time t , $Q_{obs}(t)$ is the observed discharge at time t , and \bar{Q}_{obs} is the mean observed discharge.

$$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{Y})^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - X_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2} \quad (8)$$

$$PCC = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (9)$$

where X_i and Y_i are the i th simulated water depth from the coupled model and the estimated value from the ML method, respectively; n represents the total number of samples; \bar{X} and \bar{Y} are the simulated mean values of water depth from the coupled model and the estimated value from the ML method, respectively.

3 Results

This section introduces the results on validation of the physical-based model, the creation of inundation database, the accuracy and efficiency of the MORF model, and the comparisons with the KNN model.

3.1 Model Calibration and Validation

The SWMM used in this study was improved from our previous study (Zhang et al. 2021), and hence further calibration and verification were carried out. The initial values of the parameters were determined by referring to the SWMM user manual (Rossman 2015) and from values used in other cities in or around Guangzhou City (Wu et al. 2018; Wang et al. 2021; Li et al. 2022).

The model was calibrated and verified using two measured rainfall-flow datasets (see Fig. 1 for location) on 2 June 2021 and 21 June 2021. These two rainfall events had large amounts of total and concentrated rainfall, with total rainfall amounts of 59 mm and 104 mm respectively (Fig. 6). We used the SWMM to simulate the flow of the two rainfall events, and a comparison of simulated and monitored flows is presented in Fig. 6.

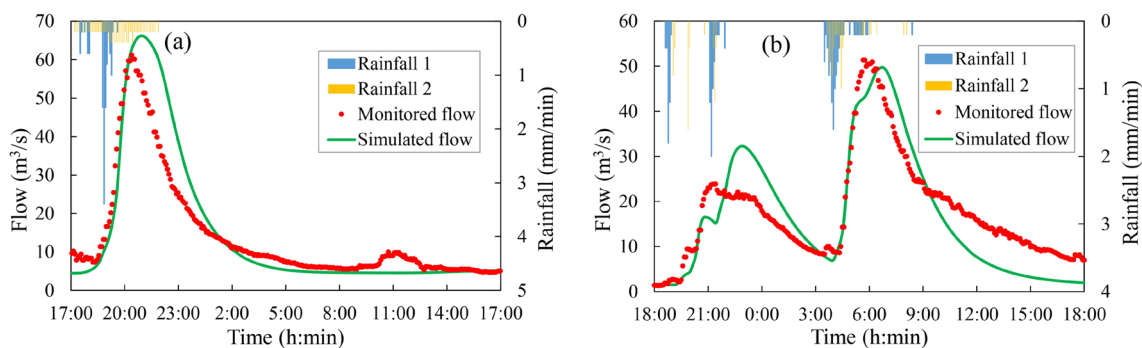


Fig. 6 Details of the validation of the storm water management model (SWMM). Rainfall 1 (blue) and Rainfall 2 (yellow) represent the rainfall monitored by rain gauge 1 and 2 in Fig. 1, respectively

The results show that the simulated peak flow time at the monitoring point is basically in line with the monitored flow. The simulated and monitored flows for the two rainfall events are highly correlated, and the *NSE* coefficient reached 0.78 and 0.73, respectively.

The 2D verification of the WCA2D model was carried out by comparing the simulated inundation range with historical waterlogging points. To simulate the location of the historical waterlogging points, we designed a rainstorm event using the Chicago rainfall pattern and a 100-year return period (Fig. 7). It was found that the internally flooded areas were mainly concentrated in low-lying areas along rivers, which was confirmed with comparison to observations.

The above results show that the coupled model (SWMM-WCA2D) developed for this study provided acceptable simulations, which well described the actual water flow and flooding conditions, suggesting that the coupled model has good applicability and high reliability in the study area.

3.2 Rainfall and Inundation Database

We constructed a rainfall-inundation database by inputting the different designed rainfall events into the coupled model. This database should contain as many different features as possible to improve the learning and generalization ability of the ML method. Table 1 summarizes the maximum inundation areas ($h > 0.15$ m) of the 48 scenarios. There were obvious differences in the inundation produced by the different rainfall events. When the return period increases, the inundated area becomes larger. Under the same return period, the inundation resulting from single-peak rainfall patterns (patterns I, II, III, and Chicago) is significantly larger than that resulting from uniform and double-peak patterns. For example, the inundated area for pattern III rainfall with a 100-year return period is 501.66 ha, which is 15.96% and 15.53% larger than pattern IV (432.61 ha) and pattern V (434.21 ha), respectively. Generally, due to the nonlinear relationship between rainfall and inundation, the areas of

inundation caused by rainfall of different patterns are quite different even for the same return period.

3.3 Urban Flood Mapping Using the MORF Model

This section presents the parameters and optimization results of the MORF model, evaluates its performance in predicting inundation depth, and compares its results with the coupled model as well as the KNN model.

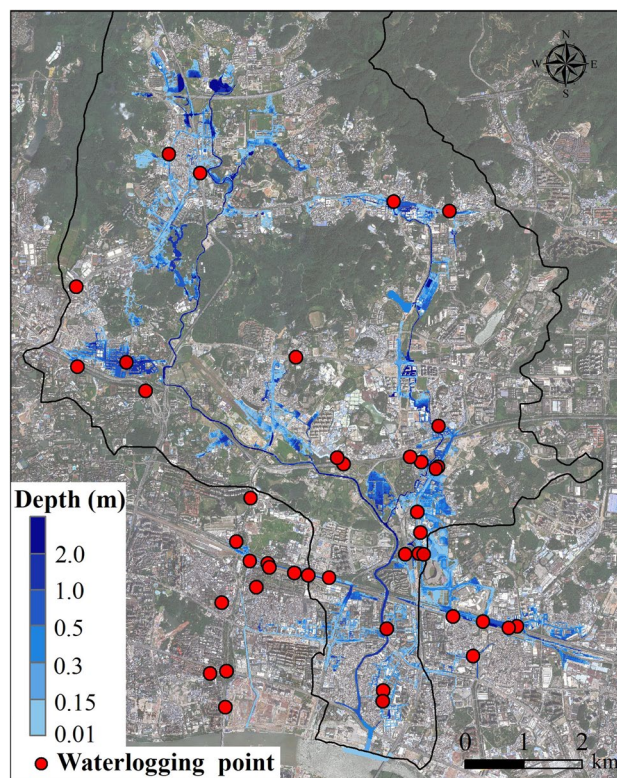
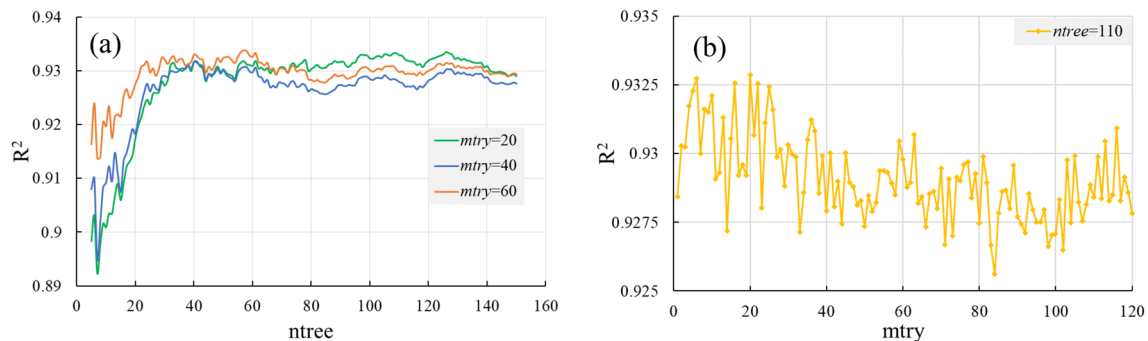


Fig. 7 Details of the validation of the WCA2D model used in this study in Guangzhou City, China

Table 1 Maximum inundation areas (ha) resulting from rainstorms with different return periods and rainfall patterns (see Fig. 3)

Return period ^a	Rainfall pattern							
	I	II	III	IV	V	VI	VII	Chicago
1	92.53	98.32	101.45	73.67	75.33	88.15	90.36	107.59
5	189.91	208.60	210.41	162.27	167.24	182.71	186.85	211.76
10	258.37	277.79	281.39	219.35	229.60	255.14	249.40	280.17
20	326.52	353.04	350.67	289.80	296.99	325.07	328.47	344.49
50	408.91	438.92	440.15	370.41	373.88	409.23	413.38	419.25
100	469.68	499.31	501.66	432.61	434.21	471.55	473.36	477.80

^aReturn time intervals expressed in years**Fig. 8** Parameter optimization process of the MORF model: R^2 changes with $mtry = 20, 40,$ and 60 (a), and with $ntree = 110$ (b)

3.3.1 MORF Model Construction and Optimization

After constructing the rainfall-inundation database, MORF was applied to learn the relationship between rainfall events and inundation scenarios. The independent variable in the model is the actual rainfall process (that is, if the rainfall duration is 120 minutes, there are a total of 120 input variables), and the dependent variable is the water depths at different spatial points (114,812 grid points in total). The 48 simulated rainfall and inundation scenarios were divided into a training and a test dataset with a ratio of 8:2. The test set was randomly divided and fine-tuned to include as many different rainfall features as possible, such as rainfall intensity and peak time, to test the generalization ability of the model.

There are two important parameters when constructing the MORF—the number of trees in the forest ($ntree$), and the number of independent variables randomly selected for splitting at each node in a tree ($mtry$). Liaw and Wiener (2002) suggested generating a forest with an increasing number of trees until the increase in the number does not effectively improve prediction performance. It is generally recommended for RF regression that $mtry$ should equal one-third of the number of prediction variables (Saha et al. 2016). To explore the most suitable parameters, $mtry$ was set to one-sixth, one-third, and one-half of the number of independent variables, corresponding to 20, 40, and 60 respectively in

this study. At the same time, $ntree$ was gradually increased to test model performance Fig. 8. The results suggest that the model performed best when $mtry$ and $ntree$ were set to 20 and 110, respectively, and these values will be used as the final parameter scheme, and all other parameters are set to their default values. The MORF model was built by using the package scikit-learn 1.0.2 of the Python program.

3.3.2 Comparison of Inundation Maps based on the MORF Model and the Coupled Model

The prediction effectiveness of the MORF model was verified by using 12 rainstorm and inundation scenarios, including 10 designed rainfall events and 2 actual rainfall events. The 10 designed rainfalls used rainfall patterns I, IV, and VII with a 50-year return period, patterns II, V, and Chicago with a 20-year return period, patterns III and VI with a 10-year return period, and patterns I and IV with a 5-year return period. Two actual rainfall events from 23 June 2021 (R20210623) and 28 July 2021 (R20210728) were also used (Fig. 9). For these two events, the rainfall amounts in 2 hours were 72 mm (a one-year return period event) and 103 mm (a five-year return period event), respectively. Spatial distributions of water depth predicted by the MORF model and the coupled model for rainfall pattern I with 5-year return period and pattern VII with 50-year return period in the test set were evaluated Fig. 10, and a comparison between the

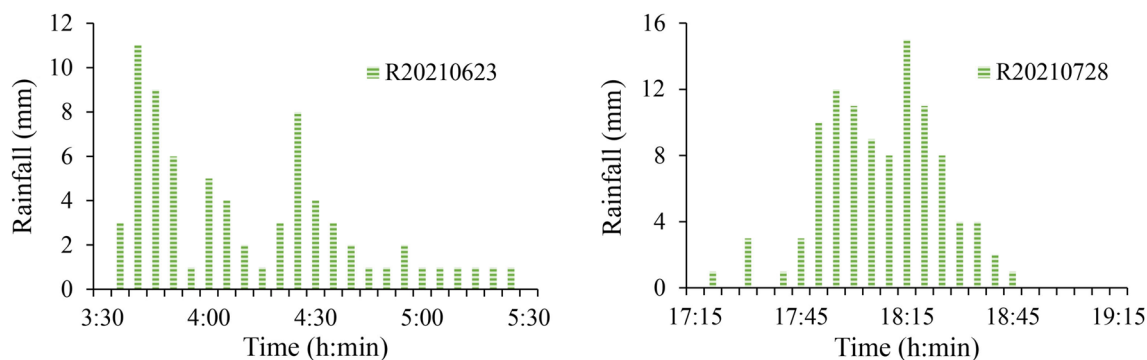


Fig. 9 Details of the two actual rainfall events used to test the effectiveness of the MORF model

MORF model predictions and those of the coupled model under the 12 test scenarios is shown in Figs. 11 and 12.

The spatial distributions of inundated depth predicted by the MORF model were similar to those of the coupled model, and the maximum water depth difference between them was typically less than 0.1 m. The flood-prone points and inundated area were generally consistent. Figure 12 shows the scatter plot of correlation between inundated depth at each grid point predicted by the MORF model and by the coupled model. Nearly all grid points had high correlations between predicted values of the two prediction methods, with significant linear correlations ($P < 0.001$). The performance indicators of the MORF model are presented in Table 2. The *MAE* values were between 3.4 cm and 9.6 cm, with an average of 6.5 cm; the mean *RMSE* was 0.189, and the *PCC* reached an average of 0.951. The MORF model also showed high accuracy for water depth prediction when using the actual rainfall events.

From the perspective of computing time, the coupled model required more computing time, as each unit runs at high surface resolution. In the CRB study area, when the resolution of the topographic data was $8\text{ m} \times 8\text{ m}$, the maximum number of inundated grids in the target area was 114,812. The MORF model was able to predict the maximum inundated range and water depth in the study area within 2 seconds for a given rainfall event, while the simulation time of the coupled model, using SWMM and WCA2D, reached 468–1614 seconds. Hence, the prediction efficiency of the MORF model was approximately 200 times higher than that of the coupled model.

To show the advantages of the MORF model, we further constructed a KNN model and compared its prediction performance of spatial inundated depth with that of the MORF model. The distance metric of the KNN model is the Euclidean distance, and the optimal parameter $K = 1$ was obtained using the 10-fold cross-validation method (Wang et al. 2015). For the KNN model, the mean value of *MAE*, *RMSE*, and *PCC* is 7.9 cm, 0.247, and 0.935, respectively

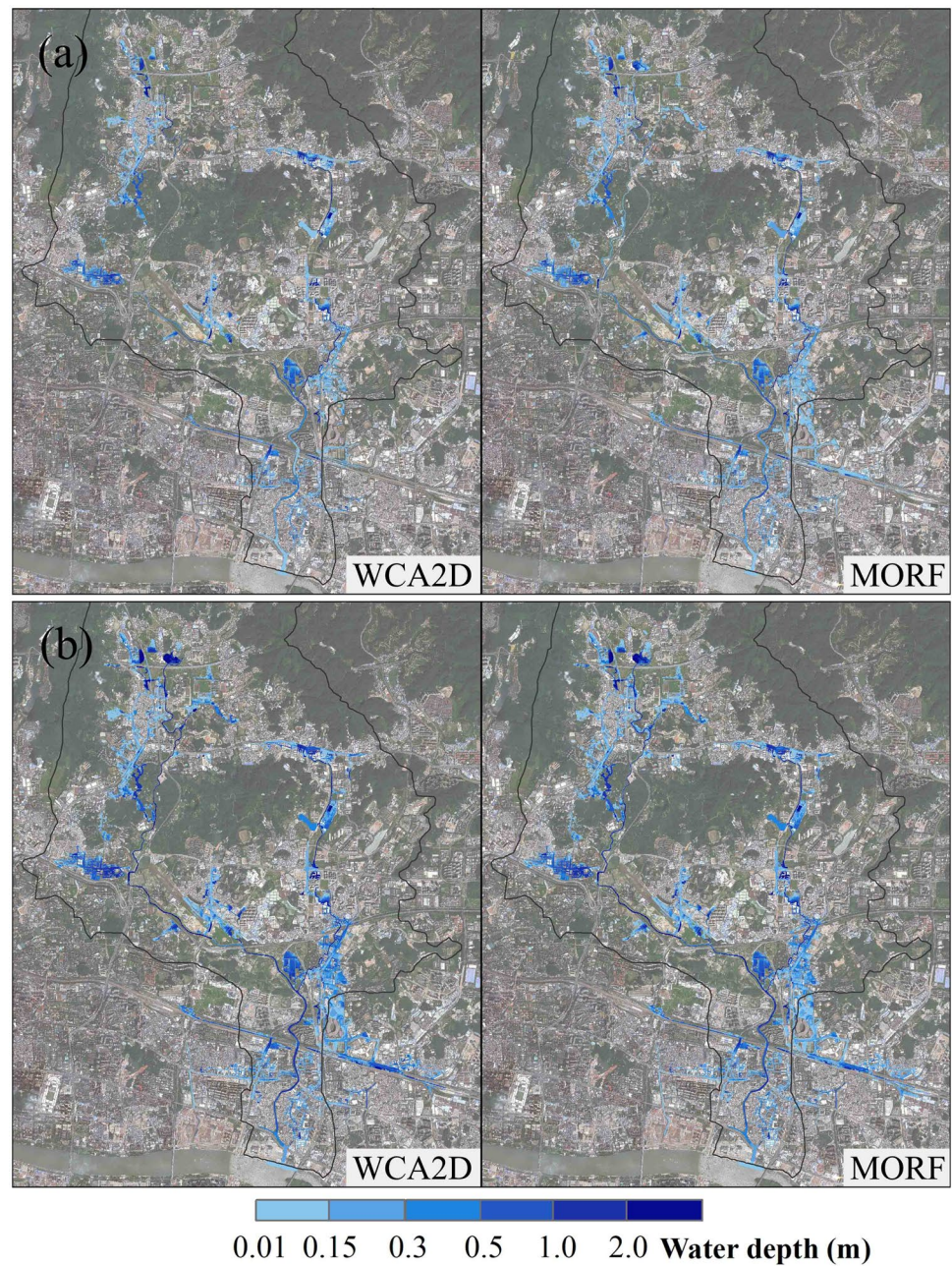
(Table 2). The results confirm that the spatial distributions of the predicted water depth of both the KNN and the MORF models were satisfactory, but the overall prediction accuracy of the MORF model was better. Although the training time of the MORF model and the KNN model reached 10 minutes and 1 second respectively, they used similar amounts of computational time in terms of prediction efficiency.

In conclusion, the simulation results of the MORF and coupled models demonstrated little difference and showed strong correlation. As the computational time required for simulating and predicting water depths using the MORF model is short, and its accuracy meets all of the expected requirements, this study suggests that MORF has great potential for rapid and real-time prediction of flood inundation.

4 Discussion

Rapid and accurate prediction of flood inundation induced by rainstorm is an effective nonengineering measure that can reduce loss of urban flood disasters (Berkhahn et al. 2019). In this study, a MORF algorithm-based fast simulation framework of urban flood prediction was introduced. Only one model is needed to predict the water depth at multiple grids at the same time while a large number of models are required in the previous studies (Chu et al. 2020; Lin et al. 2020). The developed MORF model in the framework is competent for automatically mining the nonlinear relationship between rainfall and inundation response in a rainfall-inundation database generated by a physical model. It can also deal with new inundation scenarios by identifying the characteristics of an unknown (or predicted) rainfall event, rather than simply calling existing similar rainfall-inundation scenarios from a database like the KNN model. This ability is crucial for improving the model's prediction accuracy. In addition, the actual rainstorms and the corresponding urban disasters that occur later can be added to the database and

Fig. 10 Comparison of inundated depth predicted by the MORF model and the coupled model. The images in **a** represent rainfall pattern I with a 5-year return period, and those in **b** present pattern VII with a 50-year return period in the test set, respectively



then the MORF model could be updated continually. In this case, the model would be more robust and accurate.

The framework of the MORF model has the advantages of fast construction, few parameter settings, high efficiency, and accurate urban flood prediction. Moreover, this study showed that the MORF model is at least 200 times faster than the coupled SWMM-WCA2D model with similar accuracy. In addition, our previous study (Zhang et al. 2021) found that the simulation times of another coupled model (SWMM- LISFLOOD-FP), a traditional hydrodynamic model (Coulthard et al. 2013), in the same study area were much longer. Their run times were typically up to 2 hours

(executed in the same operating environment), while the MORF model developed here had typical run time of a few seconds. When the study area is large enough, or the grid spatial accuracy required is high (that is, a greater number of grid points), the high prediction efficiency of the MORF model will be more advantageous in comparison to a hydrodynamic model, without significant loss of accuracy.

In this study, we set the duration of rainfall to 120 minutes. However, the model demonstrated satisfactory predictions under actual input rainfalls, whose duration was less than 120 minutes. In practice, the duration of rainfall may exceed, or be shorter than 120 minutes. To test water depth

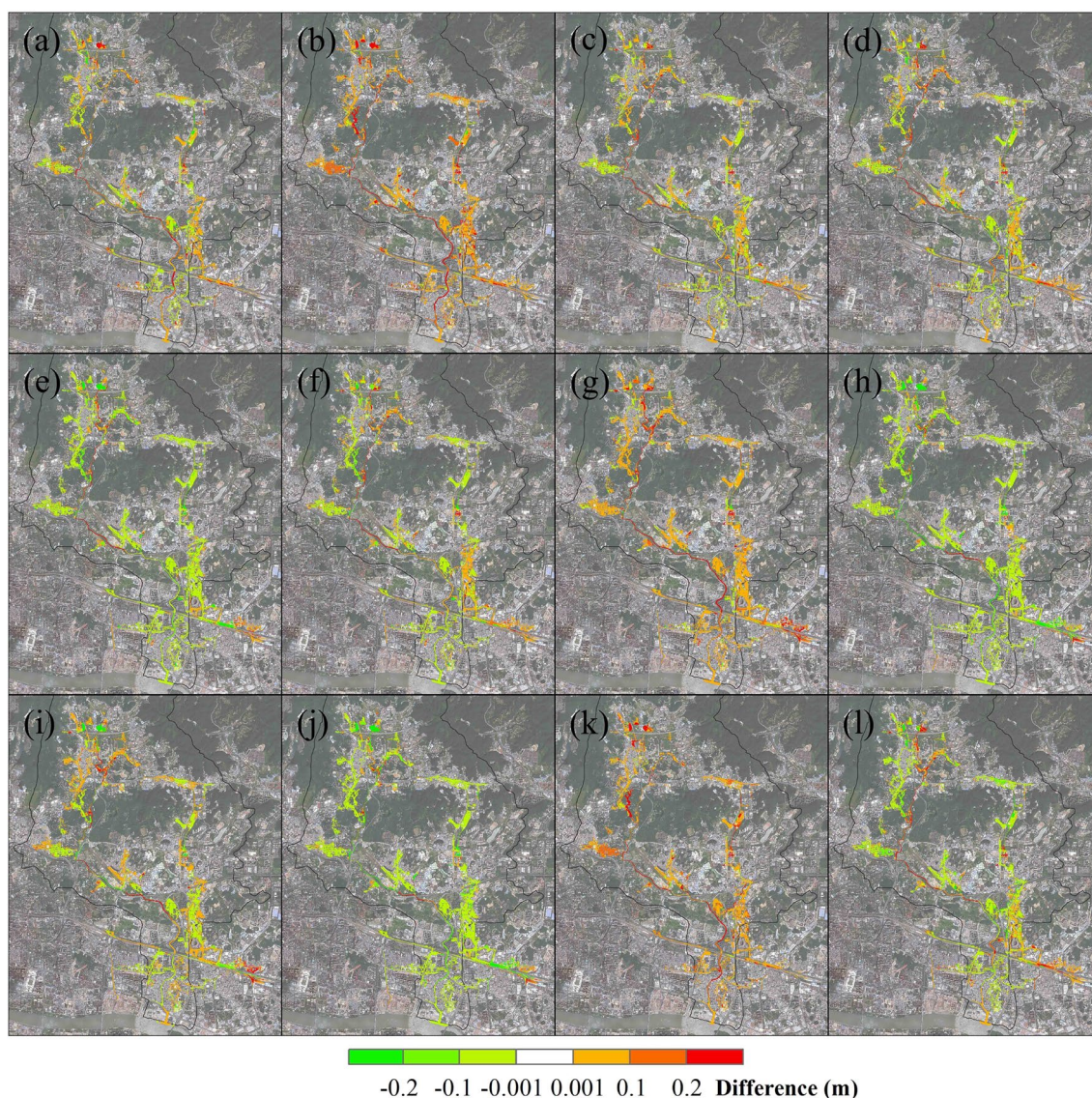


Fig. 11 Spatial differences of inundated depth predicted by the MORF model and the coupled model. (a, b) represent patterns I and IV with 5-year return period, (c, d) represent patterns III and VI with 10-year return period, (e–g) represent patterns II, V, and Chi-

cago with 20-year return period, (h–j) represent patterns I, IV, and VII of 50-year return period, and (k, l) represent actual rainfall events “R20210623” and “R20210728”, respectively. The specific designed rainfall process can be found in Fig. 3

prediction using the MORF model with longer rainfall durations, we extended the rainfall duration to 240 and 360 minutes, and then regenerated the rainfall-inundation database and reconstructed the MORF model. Table 3 shows that the results of the three evaluation indices under different rainfall durations are satisfactory, meaning that the MORF model is still robust under changing rainfall durations.

With current computing power and available algorithms, building an accurate real-time early warning and forecasting system for urban flood based on artificial intelligence (AI) has become particularly important for disaster prevention and mitigation. The importance of AI has increased because rainstorm and urban flood disasters usually occur suddenly

and cause serious disaster losses in a very short time. The proposed framework can provide strong support for early warning and forecasting of such disasters. The framework is a hybrid with the advantage of high-precision and high-efficiency as it carries over the accuracy of the traditional physical model and simultaneously includes the computational efficiency of ML. With the improvement of rainfall “nowcasting” technology, near real-time forecasting of the amount and spatial-temporal distribution of rainfall (for example, 1 to 3 hours) can be done in advance. The predicted rainfall can then be put into the framework, and accurate information of inundated area, water depth, and flood-prone points can be quickly obtained in advance. Solutions can

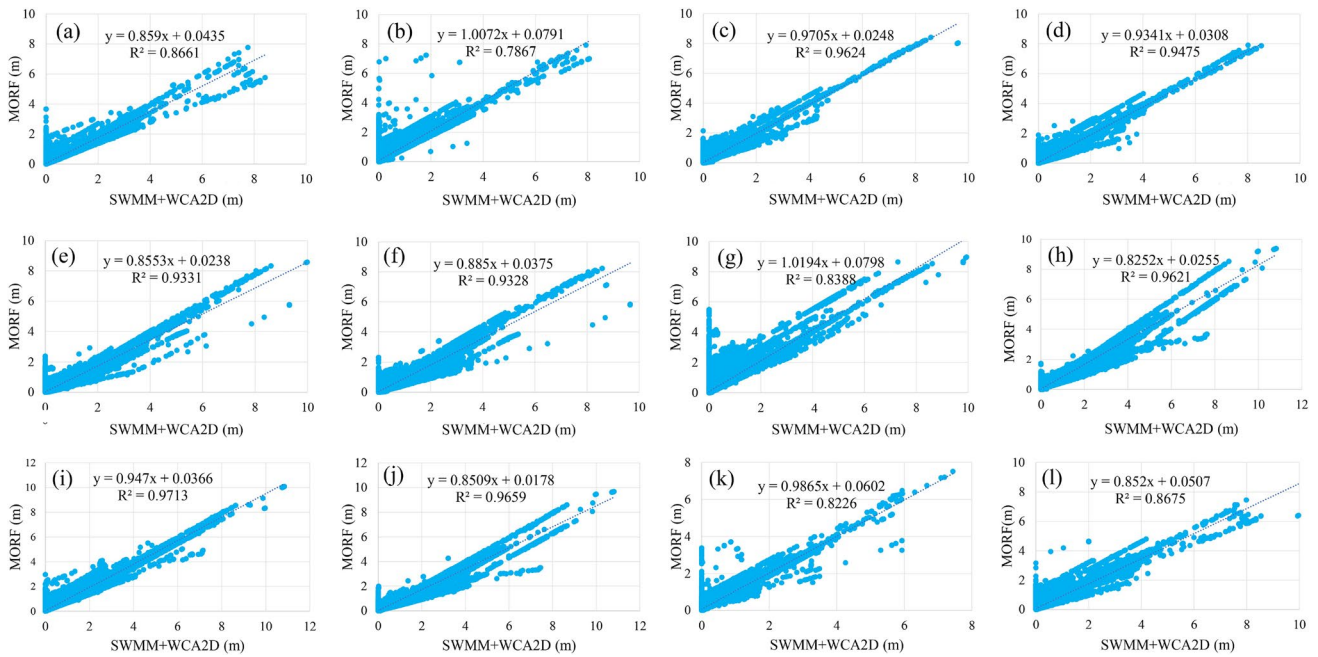


Fig. 12 Correlation of inundated depth predicted by the MORF model and the coupled SWMM-WCA2D model. (a, b) represent patterns I and IV with 5-year return period, (c, d) represent patterns III and VI with 10-year return period, (e–g) represent patterns II, V,

and Chicago with 20-year return period, (h–j) represent patterns I, IV, and VII with 50-year return period, and (k, l) represent the actual rainfall events “R20210623” and “R20210728”, respectively. The specific designed rainfall process can be found in Fig. 3

Table 2 The MAE, RMSE, and PCC values of the test data (2h) with the MORF and KNN approach

No.	Test rainfall ^a	MORF			KNN		
		MAE (m)	RMSE (m)	PCC	MAE (m)	RMSE (m)	PCC
1	50 (I)	0.080	0.223	0.981	0.100	0.272	0.973
2	50 (IV)	0.051	0.144	0.986	0.104	0.293	0.966
3	50 (VII)	0.073	0.198	0.983	0.101	0.282	0.971
4	20 (II)	0.069	0.188	0.966	0.099	0.282	0.931
5	20 (V)	0.050	0.172	0.966	0.091	0.304	0.902
6	20 (Chicago)	0.096	0.321	0.916	0.090	0.269	0.939
7	10 (III)	0.034	0.113	0.981	0.077	0.262	0.902
8	10 (VI)	0.047	0.125	0.973	0.073	0.227	0.920
9	5 (I)	0.048	0.175	0.931	0.073	0.259	0.898
10	5 (IV)	0.084	0.245	0.887	0.061	0.221	0.916
11	R20210623	0.063	0.162	0.907	0.015	0.091	0.965
12	R20210728	0.080	0.200	0.931	0.065	0.198	0.942
Average		0.065	0.189	0.951	0.079	0.247	0.935

^aTest rainfall return time intervals expressed in years

then be rapidly adopted to reduce property and infrastructure damage and loss of life.

Generally, the framework proposed by our study does not seem “smart” enough because a large number of calculations and simulations has to be conducted to construct the database before mapping the urban flood when encountering real-time or predicted rainstorms. But it can generate the flooding map with a few seconds and then in return gain

precious reaction time for the residents. The framework can be considered a feasible approach for real-time prediction until substantial breakthroughs on the coupled model’s algorithm and general computing power take place in the future. The framework is reasonable and suitable for any region, but for the study in the Chebei River Basin (CRB), there are still some limitations. First, our work was limited by data accessibility; only two rainstorm events were used for calibration

Table 3 MAE, RMSE, and PCC values of MORF model results for rainfall durations longer than 120 minutes (4 and 6 hours)

No.	Test rainfall ^a	MAE (m)		RMSE (m)		PCC	
		4h	6h	4h	6h	4h	6h
1	50 (I)	0.017	0.024	0.048	0.075	0.999	0.998
2	50 (IV)	0.020	0.032	0.064	0.122	0.995	0.978
3	50 (VII)	0.012	0.012	0.039	0.053	0.998	0.996
4	20 (II)	0.023	0.031	0.073	0.100	0.997	0.995
5	20 (V)	0.012	0.011	0.029	0.038	1.000	0.998
6	20 (Chicago)	0.009	0.031	0.025	0.081	1.000	0.997
7	10 (III)	0.069	0.074	0.209	0.241	0.972	0.967
8	10 (VI)	0.072	0.180	0.238	0.180	0.952	0.970
9	5 (I)	0.053	0.062	0.181	0.195	0.937	0.928
10	5 (IV)	0.112	0.114	0.343	0.303	0.837	0.864
Average		0.040	0.057	0.125	0.139	0.969	0.969

^aTest rainfall return time intervals expressed in years

and verification of the coupled model. Accuracy could be improved if more events were used. Second, we only used the designed rainfall events and assumed a total of 48 rainfall scenarios. Better effect could be obtained if real rainfall events that cover different return periods and rainfall patterns were used. In addition, our research mainly considered the rapid prediction of inundation under spatially generalized rainfall, and it is worth further research to extract the spatial characteristics of rainfall to achieve a rapid prediction of inundation under spatiotemporal changes of rainfalls. Third, ensemble learning methods can be considered, such as combining the prediction results of multiple models, to improve the accuracy and robustness of the predictions. Finally, we focused on predicting the maximum inundated depth due to its significant impact on flood risk. But flood inundation is a dynamic process, and how to predict the dynamic inundation process is also a subject for further consideration.

5 Conclusion

In this study, we proposed a framework for fast mapping of urban flood inundation. We first constructed a high-accuracy coupled model based on SWMM and WCA2D. Then rainfall events with different characteristics were designed and entered into the coupled model to construct a rainfall-inundation database. Finally, a prediction model, using a data-driven ML method based on MORF, was constructed and its prediction performance was systematically evaluated. The prediction model based on MORF was further compared to that based on KNN. The main conclusions of this study can be summarized as follows:

- (1) Our coupled model can deliver accurate information about inundated areas and water depths. The *NSE* coefficients of the two simulated flow events reach 0.78

and 0.73 respectively. The inundated area simulated by the coupled model is consistent with historical water-logging points. Therefore, the coupled model well described the actual water flow and flooding conditions, and presented good applicability and high reliability for the study area.

- (2) A total of 48 different scenarios were used to construct a rainfall-inundation database. By inputting the rainfall events of eight rainfall patterns and six return periods (1, 5, 10, 20, 50, and 100 years) into the coupled model, inundation scenarios under 48 different kinds of rainfall events were simulated. Due to the nonlinear relationship between rainfall and inundation, there are obvious differences in inundation under the different rainfall events even with the same return period.
- (3) The prediction model based on a MORF method can effectively learn the complex nonlinear relationship between rainfall and inundation, and provide satisfactory prediction and maps of inundated depth under designed and measured rainfall events. The spatial distribution of inundated depth predicted by the MORF model is similar to that simulated by the coupled model, with differences typically less than 0.1 m and an average correlation coefficient of 0.951. The overall prediction performance of the MORF model is also better than a KNN-based method. The prediction efficiency of the MORF model is much higher as the computation time is 200 times faster than a coupled model, suggesting that it shows a good prospect in real-time prediction for urban flood.

Acknowledgements This research acquired financial or data support of the National Key R&D Program of China (2021YFC3001000), the National Natural Science Foundation of China (U1911204, 51879107), the Natural Science Foundation of Guangdong Province

(2023B1515020087, 2022A1515010019), and the Fund of Science and Technology Program of Guangzhou (202102020216).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adhaityar, B. Y., D.P. Sahara, C. Pratama, A. Wibowo, and L.S. Heli-ani. 2021. Multi-target regression using Convolutional Neural Network-Random Forests (CNN-RF) for early earthquake warning system. Paper presented at the 9th International Conference on Information and Communication Technology (ICOICT), 3–5 August 2021, Yogyakarta, Indonesia.
- Bentivoglio, R., E. Isufi, S.N. Jonkman, and R. Taormina. 2022. Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences* 26(16): 4345–4378.
- Berkhahn, S., L. Fuchs, and I. Neuweiler. 2019. An ensemble neural network model for real-time prediction of urban floods. *Journal of Hydrology* 575: 743–754.
- Bermúdez, M., L. Cea, and J. Puertas. 2019. A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. *Journal of Flood Risk Management* 12(1): 1–14.
- Bhola, P.K., J. Leandro, and M. Disse. 2018. Framework for offline flood inundation forecasts for two-dimensional hydrodynamic models. *Geosciences* 8(9): Article 346.
- Borchani, H., G. Varando, C. Bielza, and P. Larrañaga. 2015. A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery* 5(5): 216–233.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5–32.
- Brunton, S.L., B.R. Noack, and P. Koumoutsakos. 2019. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* 52: 477–508.
- Chen, W., G. Huang, and H. Zhang. 2017. Urban stormwater inundation simulation based on SWMM and diffusive overland-flow model. *Water Science & Technology* 76(12): 3392–3403.
- Chen, W., G. Huang, H. Zhang, and W. Wang. 2018. Urban inundation response to rainstorm patterns with a coupled hydrodynamic model: A case study in Haidian Island, China. *Journal of Hydrology* 564: 1022–1035.
- Chu, H., W. Wu, Q.J. Wang, R. Nathan, and J. Wei. 2020. An ANN-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software* 124: Article 104587.
- Coulthard, T.J., J.C. Neal, P.D. Bates, J. Ramirez, G.A.M. de Almeida, and G.R. Hancock. 2013. Integrating the LISFLOOD-FP 2D hydrodynamic model with the CAESAR model: Implications for modelling landscape evolution. *Earth Surface Processes and Landforms* 38(15): 1897–1906.
- De'ath, G. 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology* 83(4): 1105–1117.
- Deng, Z., Z. Wang, X. Wu, C. Lai, and Z. Zeng. 2022. Strengthened tropical cyclones and higher flood risk under compound effect of climate change and urbanization across China's Greater Bay Area. *Urban Climate* 44: Article 101224.
- Dottori, F., and E. Todini. 2011. Developments of a flood inundation model based on the cellular automata approach: Testing different methods to improve model performance. *Physics and Chemistry of the Earth, Parts A/B/C* 36(7): 266–280.
- Guidolin, M., A.S. Chen, B. Ghimire, E.C. Keedwell, S. Djordjević, and D.A. Savić. 2016. A weighted cellular automata 2D inundation model for rapid flood analysis. *Environmental Modelling & Software* 84: 378–394.
- He, S., Z. Wang, D. Wang, W. Liao, X. Wu, and C. Lai. 2022. Spatiotemporal variability of event-based rainstorm: The perspective of rainfall pattern and concentration. *International Journal of Climatology* 42(12): 6258–6276.
- Hou, J., N. Zhou, G. Chen, M. Huang, and G. Bai. 2021. Rapid forecasting of urban flood inundation using multiple machine learning models. *Natural Hazards* 108(2): 2335–2356.
- IPCC (Intergovernmental Panel on Climate Change). 2021. *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, et al. Cambridge, UK and New York, USA: Cambridge University Press.
- Jhong, B.-C., J.-H. Wang, and G.-F. Lin. 2017. An integrated two-stage support vector machine approach to forecast inundation maps during typhoons. *Journal of Hydrology* 547: 236–252.
- Jian, W., S. Li, C. Lai, Z. Wang, X. Cheng, E.Y.-M. Lo, and T.-C. Pan. 2021. Evaluating pluvial flood hazard for highly urbanised cities: A case study of the Pearl River Delta Region in China. *Natural Hazards* 105(2): 1691–1719.
- Kabir, S., S. Patidar, X. Xia, Q. Liang, J. Neal, and G. Pender. 2020. A deep convolutional neural network model for rapid prediction of fluvial flood inundation. *Journal of Hydrology* 590: Article 125481.
- Keifer, C.J., and H.H. Chu. 1957. Synthetic storm pattern for drainage design. *Journal of the Hydraulics Division* 83(4): 1332-1–1332-25.
- Kim, J., and H. Cho. 2019. Scenario-based urban flood forecast with flood inundation map. *Tropical Cyclone Research and Review* 8(1): 27–34.
- Kocev, D., C. Vens, J. Struyf, and S. Džeroski. 2007. Ensembles of multi-objective decision trees. In *Machine learning: ECML 2007*, ed. J.N. Kok, J. Koronacki, R.L. de Mantaras, S. Matwin, D. Mladenić, and A. Skowron, 624–631. Berlin and Heidelberg: Springer.
- Kocev, D., S. Džeroski, M.D. White, G.R. Newell, and P. Griffioen. 2009. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. *Ecological Modelling* 220(8): 1159–1168.
- Lai, C., Q. Shao, X. Chen, Z. Wang, X. Zhou, B. Yang, and L. Zhang. 2016. Flood risk zoning using a rule mining based on ant colony algorithm. *Journal of Hydrology* 542: 268–280.
- Lai, C., X. Chen, Z. Wang, H. Yu, and X. Bai. 2020. Flood risk assessment and regionalization from past and future perspectives at basin scale. *Risk Analysis* 40(7): 1399–1417.
- Li, S., Z. Wang, C. Lai, and G. Lin. 2020. Quantitative assessment of the relative impacts of climate change and human activity on flood susceptibility based on a cloud model. *Journal of Hydrology* 588: Article 125051.

- Li, S., Z. Wang, X. Wu, Z. Zeng, P. Shen, and C. Lai. 2022. A novel spatial optimization approach for the cost-effectiveness improvement of LID practices based on SWMM-FTC. *Journal of Environmental Management* 307: Article 114574.
- Liao, Y., Z. Wang, J. Xiong, and C. Lai. 2021. Dimming in the Pearl River Delta of China and the major influencing factors. *Climate Research* 82: 161–176.
- Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2(3): 18–22.
- Lin, G.-F., H.-Y. Lin, and Y.-C. Chou. 2013. Development of a real-time regional-inundation forecasting model for the inundation warning system. *Journal of Hydroinformatics* 15(4): 1391–1407.
- Lin, Q., J. Leandro, S. Gerber, and M. Disse. 2020. Multistep flood inundation forecasts with resilient backpropagation neural networks: Kulmbach case study. *Water* 12(12): Article 3568.
- Lin, Y., D. Wang, G. Wang, J. Qiu, K. Long, Y. Du, H. Xie, Z. Wei, et al. 2021. A hybrid deep learning algorithm and its application to streamflow prediction. *Journal of Hydrology* 601: Article 126636.
- Ling, F., J.-J. Luo, Y. Li, T. Tang, L. Bai, W. Ouyang, and T. Yamagata. 2022. Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. *Nature Communications* 13(1): Article 7681.
- Liu, W., D. Xu, I.W. Tsang, and W. Zhang. 2019. Metric learning for multi-output tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(2): 408–422.
- McGrath, H., J.-F. Bourgon, J.-S. Proulx-Bourque, M. Nastev, and A.A.E. Ezz. 2018. A comparison of simplified conceptual models for rapid web-based flood inundation mapping. *Natural Hazards* 93(2): 905–920.
- Molokov, M.B., and IT Shtigorin. 1956. *The rain water and confluent channel*. Beijing: Architectural Engineering Press (in Chinese).
- Nakano, F.K., K. Pliakos, and C. Vens. 2022. Deep tree-ensembles for multi-output prediction. *Pattern Recognition* 121: Article 108211.
- Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology* 10(3): 282–290.
- Pan, C., X. Wang, L. Liu, H. Huang, and D. Wang. 2017. Improvement to the Huff Curve for design storms and urban flooding simulations in Guangzhou, China. *Water* 9(6): Article 411.
- Rossman, L.A. 2015. *Storm water management model user's manual version 5.1*. Washington, DC: United States Environmental Protection Agency (USEPA).
- Saha, D., P. Alluri, and A. Gan. 2016. A random forests approach to prioritize Highway Safety Manual (HSM) variables for data collection: Random forests to prioritize HSM variables. *Journal of Advanced Transportation* 50(4): 522–540.
- Schulz, A., J. Kiesel, H. Kling, M. Preishuber, and G. Petersen. 2015. An online system for rapid and simultaneous flood mapping scenario simulations – The Zambezi FloodDSS. In *Proceedings of EGU General Assembly 2015*, 12–17 April 2015, Vienna, Austria.
- Shih, Y.-S., and H.-W. Tsai. 2004. Variable selection bias in regression trees with constant fits. *Computational Statistics & Data Analysis* 45(3): 595–607.
- Struyf, J., and S. Džeroski. 2006. Constraint based induction of multi-objective regression trees. In *Knowledge discovery in inductive databases*, ed. F. Bonchi, and J.-F. Boulicaut, 222–233. Berlin and Heidelberg: Springer.
- Tellman, B., J.A. Sullivan, C. Kuhn, A.J. Kettner, C.S. Doyle, G.R. Brakenridge, T.A. Erickson, and D.A. Slayback. 2021. Satellite imaging reveals increased proportion of population exposed to floods. *Nature* 596(7870): 80–86.
- Teng, J., J. Vaze, S. Kim, D. Dutta, A.J. Jakeman, and B.F.W. Croke. 2019. Enhancing the capability of a simple, computationally efficient, conceptual flood inundation model in hydrologically complex terrain. *Water Resources Management* 33(2): 831–845.
- Wang, Z., C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai. 2015. Flood hazard risk assessment model based on random forest. *Journal of Hydrology* 527: 1130–1141.
- Wang, W., W. Chen, and G. Huang. 2021. Urban stormwater modeling with local inertial approximation form of shallow water equations: A comparative study. *International Journal of Disaster Risk Science* 12(5): 745–763.
- Wang, Z., S. Li, X. Wu, G. Lin, and C. Lai. 2022. Impact of spatial discretization resolution on the hydrological performance of layout optimization of LID practices. *Journal of Hydrology* 612: Article 128113.
- Wu, X., Z. Wang, S. Guo, W. Liao, Z. Zeng, and X. Chen. 2017. Scenario-based projections of future urban inundation within a coupled hydrodynamic model framework: A case study in Dongguan City, China. *Journal of Hydrology* 547: 428–442.
- Wu, X., Z. Wang, S. Guo, C. Lai, and X. Chen. 2018. A simplified approach for flood modeling in urban environments. *Hydrology Research* 49(6): 1804–1816.
- Xiong, J., Z. Wang, C. Lai, Y. Liao, and X. Wu. 2020. Spatiotemporal variability of sunshine duration and influential climatic factors in mainland China during 1959–2017. *International Journal of Climatology* 40(15): 6282–6300.
- Xu, D., Y. Shi, I.W. Tsang, Y.-S. Ong, C. Gong, and X. Shen. 2020. Survey on multi-output learning. *IEEE Transaction on Neural Networks and Learning Systems* 31(7): 2409–2429.
- Xu, T., and F. Liang. 2021. Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews-Water* 8(5): Article e1533.
- Xue, L., Y. Liu, Y. Xiong, Y. Liu, X. Cui, and G. Lei. 2021. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *Journal of Petroleum Science & Engineering* 196: Article 107801.
- Zeng, Z., Z. Wang, and C. Lai. 2022. Simulation performance evaluation and uncertainty analysis on a coupled inundation model combining SWMM and WCA2D. *International Journal of Disaster Risk Science* 13(4): 448–464.
- Zhang, M., M. Xu, Z. Wang, and C. Lai. 2021. Assessment of the vulnerability of road networks to urban waterlogging based on a coupled hydrodynamic model. *Journal of Hydrology* 603: Article 127105.
- Zounemat-Kermani, M., O. Batelaan, M. Fadaee, and R. Hinkelmann. 2021. Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology* 598: Article 126266.