

Article

Applying Machine Learning Methods to Improve Rainfall–Runoff Modeling in Subtropical River Basins

Haoyuan Yu ¹  and Qichun Yang ^{1,2,*} 

¹ Thrust of Earth, Ocean and Atmospheric Sciences, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511400, China; hyu012@connect.hkust-gz.edu.cn

² Center for Ocean Research in Hong Kong and Macau, Hong Kong University of Science and Technology, Hong Kong, China

* Correspondence: qichunyang@hkust-gz.edu.cn

Abstract: Machine learning models' performance in simulating monthly rainfall–runoff in subtropical regions has not been sufficiently investigated. In this study, we evaluate the performance of six widely used machine learning models, including Long Short-Term Memory Networks (LSTMs), Support Vector Machines (SVMs), Gaussian Process Regression (GPR), LASSO Regression (LR), Extreme Gradient Boosting (XGB), and the Light Gradient Boosting Machine (LGBM), against a rainfall–runoff model (WAPABA model) in simulating monthly streamflow across three subtropical sub-basins of the Pearl River Basin (PRB). The results indicate that LSTM generally demonstrates superior capability in simulating monthly streamflow than the other five machine learning models. Using the streamflow of the previous month as an input variable improves the performance of all the machine learning models. When compared with the WAPABA model, LSTM demonstrates better performance in two of the three sub-basins. For simulations in wet seasons, LSTM shows slightly better performance than the WAPABA model. Overall, this study confirms the suitability of machine learning methods in rainfall–runoff modeling at the monthly scale in subtropical basins and proposes an effective strategy for improving their performance.

Keywords: data-driven models; machine learning; rainfall–runoff model; subtropical basins; performance evaluation



Citation: Yu, H.; Yang, Q. Applying Machine Learning Methods to Improve Rainfall–Runoff Modeling in Subtropical River Basins. *Water* **2024**, *16*, 2199. <https://doi.org/10.3390/w16152199>

Academic Editor: Francesco De Paola

Received: 29 June 2024

Revised: 31 July 2024

Accepted: 31 July 2024

Published: 2 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rainfall–runoff modeling is critical for understanding and predicting the transformation of rainfall into runoff [1–4]. Accurately simulating streamflow is essential for flood forecasting, reservoir operations, water supply planning, irrigation scheduling, and the design of hydraulic structures such as dams and reservoirs [5–8]. In the context of climate change, a well-performing runoff model will allow us to assess the potential impacts of changing precipitation patterns on water availability and flood risks and thus formulate effective strategies to efficiently manage precious water resources and mitigate hazards [9–11].

In recent years, machine learning models, as a subset of data-driven models, have become increasingly popular in hydrological investigations. One of the merits of such models is their flexibility, allowing users to train, test, and deploy runoff simulations without an extensive understanding of physical processes regulating the water cycle [12,13]. Furthermore, machine learning models stand out because of their suitability to address the non-linearity and intricate interactions buried in data, making them especially suitable for modeling complex relationships between meteorological variables and hydrological processes [13–15]. However, other researchers raised concerns about the application of machine learning methods in the field of hydrology because these methods have limited interpretability and physical consistency, potentially leading to physically incorrect or unreasonable simulations, especially when the training data have low quality [16].

Despite concern about the interpretability of machine learning models, recent studies have recognized their promising performance in runoff simulation [17,18]. In previous studies, machine learning models have been mainly used to simulate daily or hourly streamflow [19–21], with much less attention paid to monthly streamflow. However, monthly streamflow simulations are equally, if not more, important for water resource management [22,23]. Unlike daily streamflow, climate variables (e.g., precipitation and evapotranspiration), rather than water storages in basins, play a dominant role in affecting streamflow on a monthly scale. Whether machine learning models are capable of capturing such features remains unclear.

As a result, the performance of machine learning models in simulating monthly streamflow warrants further investigation. First, the types of machine learning methods suitable for hydrological modeling need to be elucidated. Commonly used machine learning models in hydrology include regression analysis models, support vector machines, ensemble learning methods, and deep learning models with multi-layer neural network structures [24,25]. Many studies compared the performance of multiple machine learning models and showed that LSTM may be the best model in daily streamflow simulations. For instance, Rahimzad et al. [26] indicated that LSTM demonstrated better performance than linear regression and support vector machines (SVMs) in the Kentucky River in the U.S. In Latif and Ahmed [27], LSTM outperformed Random Forest (RF) and Tree Boost for forecasting the daily streamflow of the Warragamba dam in Australia. Adnan et al. [28] also showed that LSTM had greater accuracy than the Extreme Learning Machine (ELM) and RF techniques. The LSTM model has demonstrated capability in reconstructing streamflow across hundreds of basins in the United States [29], and it has also shown better performance than traditional process-based approaches in applications to ungauged basins [30]. The LSTM's outstanding ability to learn the influence of past events on future outcomes is a critical factor responsible for its superior performance, enabling it to effectively capture hydrological processes, such as snowmelt and groundwater recharge, which are critical for streamflow dynamics in temperate regions [31]. However, whether the performance of such models still holds at the monthly scale needs further investigation.

Recently, hybrid frameworks integrating physical processes with machine learning models have been developed to simulate hydrological processes [16], showing promise in improving the performance of data-driven models [32]. For example, LSTM combined with physically based hydrological models has shown a better capacity in streamflow simulation [33,34]. RF and regression models also achieved improved performance based on a hybrid framework [35,36]. However, these new directions of applying machine learning models are still under development.

Second, strategies for setting up machine learning models in hydrological applications need further investigation. Machine learning models often exhibit varied performance across different river basins, suggesting that their full potential in hydrological simulations has not been realized. In an investigation across a group of river basins in Australia, machine learning methods demonstrated better performance in larger river basins [37]. The dependency of model performance on river basin characteristics indicates the necessity of including additional variables other than meteorological forcing inputs in machine learning simulations. In another investigation across different basins in the U.S., Kratzert et al. [30] found that the LSTM could achieve improved performance in simulating daily streamflow by adding river basin characteristics (e.g., soil and topography) as the input data. However, which river basin property should be included in machine learning-based simulations for subtropical basins remains unclear.

Third, the majority of previous studies focus predominantly on the overall performance of the models, neglecting the variability of model performance under different hydrological conditions (e.g., wet vs. dry periods). Compared with normal flow, water resource managers are often more concerned about large flow events, which could lead to risks such as floods, especially in subtropical and tropical regions. Accurately simulating large flow events has been a challenge in hydrological modeling. Evaluating the perfor-

mance of machine learning models under various hydrological conditions will help to clarify the suitability of using such models in predicting the highly variable streamflow [38].

River basins located in subtropical or tropical regions have special hydrological processes compared with temperate basins. Subtropical and tropical basins often do not have snowfall, and thus lack the responses of streamflow to snowmelting. In addition, abundant rainfall and high evapotranspiration rates make these basins' hydrological response times relatively short [39]. Given these special features of hydrological processes in subtropical/tropical basins, Whether the machine learning models are applicable to those basins needs to be investigated.

This study aims to comprehensively investigate the suitability of machine learning models in streamflow simulations at the monthly scale in a typical region of China. The performance of machine learning models in simulating monthly streamflow is compared with a rainfall–runoff model (e.g., the WAPABA model). The objectives of this study are to answer the following questions: (1) What is the relative performance of machine learning models to a traditional hydrological model for monthly streamflow simulation in subtropical regions? (2) Which type of machine learning models are more suitable for subtropical streamflow simulations at the monthly scale? (3) How can machine learning models be effectively set up to improve their performance?

2. Data and Methods

2.1. Study Area

The Pearl River Basin (PRB), located in southern China, was selected as the study area. We chose the Pearl River Basin's three major sub-basins, namely the North River sub-basin, the East River sub-basin, and the West River sub-basin (Figure 1), to evaluate streamflow simulations by different models. The Pearl River, also known as the Zhujiang River, is one of the largest and most important river systems in China. The main stream of the river travels over 2214 km from its headwaters to the estuary, with a drainage area of about 453,690 km², making it the third-longest river in China [40].

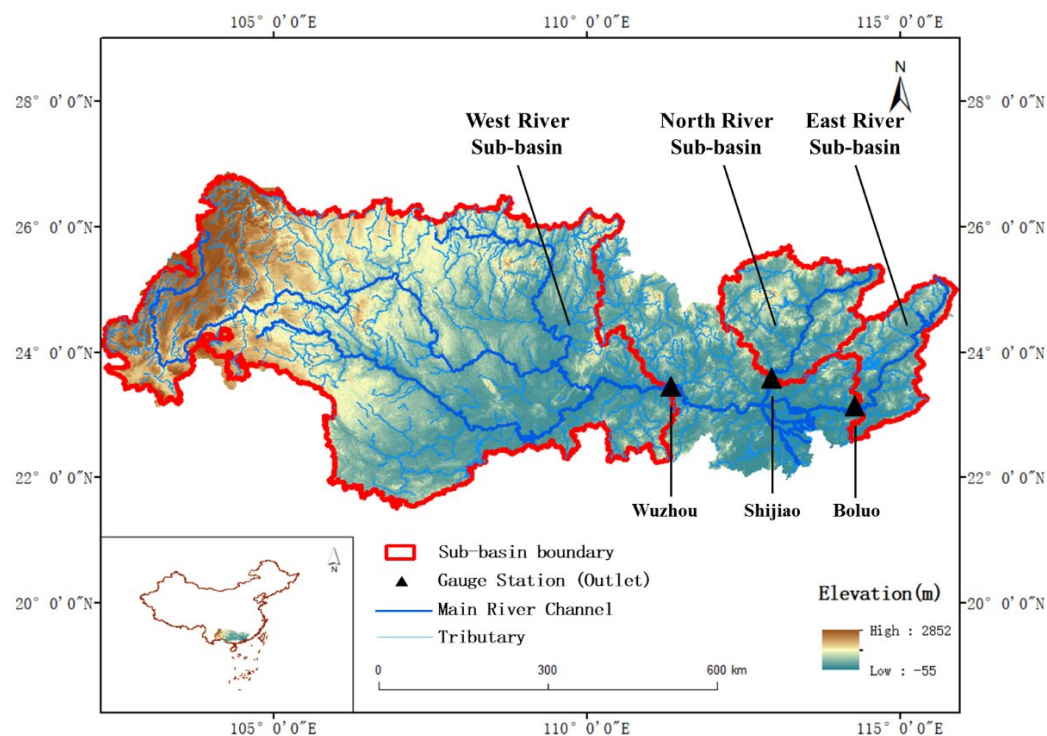


Figure 1. Location of the Pearl River Basin, the three sub-basins, and gauge stations.

The PRB has a tropical and subtropical climate, with hot and wet summers and mild winters, and is free of snow throughout the year. Annual mean temperatures in the PRB

fluctuate between 14 and 22 °C, accompanied by high precipitation averaging between 1200 and 2200 mm, primarily occurring from April to September [40,41]. Such a climate supports a wide range of flora and fauna and thriving agriculture. The PRB experiences heavy precipitation events during the wet season, leading to hazards such as flooding. Existing studies have highlighted the necessity of developing flood mitigation measures for this region [42].

2.2. Data

In this study, three gauge stations, including Shijiao, Boluo, and Wuzhou, were selected for the North River, East River, and West River sub-basins, respectively (Figure 1). The boundaries of the sub-basin were obtained from the Global Runoff Data Centre database [43]. The areas of these three sub-basins are 38,363, 25,325, and 329,705 km², respectively (Figure 1). Gauge records archived by the GRDC [44] database and the Ministry of Water Resources, PRC, were collected for model calibration and evaluation in this study. The depth of runoff in each sub-basin was calculated based on the streamflow observations from these stations and their corresponding drainage areas.

The climate data for hydrological simulations were obtained from the ERA5-Land dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [45]. In this study, all meteorological and hydrological data were aggregated to the monthly scale and averaged for each sub-basin for hydrological simulations. Climatic and hydrological variables used in this study, their units, and multi-year averages are shown in Table 1, while the other statistics are shown in Table S1. The North River sub-basin has the highest rainfall, while the East River sub-basin receives the least among the three sub-basins. The North River sub-basin has the highest runoff, while the West River sub-basin has the lowest.

Table 1. Variables used for model simulations in each sub-basin. All data are monthly averages.

Variables	North River	East River	West River	Unit
Runoff (R)	89.05	75.46	50.46	mm/month
Precipitation (P)	161.04	145.90	150.11	mm/month
Vapor pressure (e_a)	1.81	1.91	1.71	kPa
Wind speed at 2 m (u_2)	0.66	0.81	0.63	m/s
Surface net radiation (R_n)	11.62	12.39	11.25	MJ/(m ² day)
Daily maximum temperature (T_{max})	22.40	23.74	21.72	°C
Daily minimum temperature (T_{min})	12.22	12.92	12.60	°C

2.3. Models

2.3.1. WAPABA Model

The WAPABA (water partition and balance) model is a lumped rainfall–runoff model by Wang et al. [46]. The model has been widely used in monthly streamflow simulations and verified to perform well at this time scale [46–48]. Therefore, WAPABA was used as the benchmark model to compare the performance of multiple machine learning methods in this study.

The WAPABA model requires two fundamental inputs (i.e., the mean monthly rainfall and ET_o) to simulate the monthly runoff of a river basin. ET_o is calculated using the meteorological variables listed in Table 1 and the FAO Penman–Monteith Equation [49]. The schematic diagram of the WAPABA model is shown in Figure S1 of the Supplementary Material. There are five parameters in the WAPABA model, including the catchment consumption curve (α_1), the parameter of the evapotranspiration curve (α_2), the percentage

of groundwater yield from the catchment (β), the soil's maximum water-holding capacity (S_{max}), and the groundwater store time constant in reverse (K). These parameters in our study were calibrated with the Nelder–Mead optimization algorithm [50].

2.3.2. Machine Learning Models

Long Short-Term Memory Networks

The Long Short-Term Memory (LSTM) network is a recurrent neural network (RNN) designed to address the problem of gradient vanishing or explosion that exists in traditional RNNs [51]. A standard LSTM unit consists of a cell with an input gate, an output gate, and a forget gate [52].

In this study, we applied a Sequential Model using the Keras API of the TensorFlow library. The model's architecture was initiated with an LSTM layer equipped with a ReLU activation function containing ten neurons. This layer was then followed by a densely connected layer employing an L2 regularizer and outputting a single real value. During model compilation, we chose the Adam optimizer for its adaptive learning rates, which collaborated with the MSE loss function. Then, the model was trained for 2000 epochs with an initial training rate of 0.01. Additionally, we incorporated learning rate alterations using the 'Reduce LR On Plateau' strategy. This strategy involved multiplying the learning rate by a factor of 0.5 if there was no improvement in the loss after 20 epochs until reaching the minimal learning rate of 0.00001. We also utilized the 'Model Checkpoint' function to save the optimal state and implemented the 'Early Stopping' command with a patience of 100 epochs to prevent overfitting. Such hyperparameters are defined by a commonly used framework for parameter optimization of the neural networks model, as we initially set the model parameters based on empirical values, followed by splitting the data into training and validation sets for hyperparameter optimization to obtain a better hyperparameter setting. The training data were divided into 80% of the training set and the remaining 20% of the validation set to prevent overfitting. Ten distinct LSTM models were trained, and their values were averaged to mitigate the potential errors caused by the randomness. The topological configuration of the LSTM model is shown in the Supplementary Material (Figure S2). A comprehensive introduction to the LSTM algorithm can be found in Sherstinsky [53].

Support Vector Machine

The Support Vector Machine (SVM) is a type of machine learning method performing data regression through supervised learning, in which the decision boundary becomes a hyperplane that fits the data points to minimize the overall error rate. It works by mapping input vectors into a high-dimensional feature space and then searching for the hyperplane that optimally fits the data [54]. SVM was also widely used in hydrological studies [55], especially in rainfall–runoff forecasting [56–58].

In our study, we employed the SVM model with a radial basis function (RBF) kernel with the Scikit-Learn library. The RBF kernel is capable of handling non-linear relationships between the predictors and the response variable [59]. The cost parameter was set to 100, and the gamma coefficient was set to 0.01 by the grid search algorithm within a predefined parameter range. A general topological configuration of SVM is shown in Figure S3 of the Supplementary Material. The specific algorithm and related equations of support vector regression can be found in Raghavendra and Deka [55].

Gaussian Process Regression

Gaussian Process Regression (GPR) is a nonparametric model for data regression analysis using a Gaussian process prior. The modeling assumptions for GPR include both noise (regression residuals) and a Gaussian process prior, which is solved according to Bayesian inference [60]. It is known for providing both predictions and quantifying the predictive uncertainty. This method is suitable for hydrological science due to its capacity

for handling non-linear relationships, managing multiple input features, and its robustness in the face of noisy data [61–63].

In this study, we initialized the GPR model with a radial basis function (RBF) kernel, which is a good general-purpose kernel used for capturing the smoothness of the rainfall–runoff relationship [59]. The hyperparameters of the kernel were then optimized using the built-in optimization methods in the GPy Python library, which flexibly re-adjusts the kernel to best fit the data. The specific equations for the calculation of GRP can be found in Schulz et al. [64].

LASSO Regression

Lasso Regression (LR), also known as the Least Absolute Shrinkage and Selection Operator, is a linear regression model that uses shrinkage, where data values are shrunk towards a central point, typically the mean [65]. LR stands out for its inherent ability to perform automatic feature selection and reduce the dimensional complexity of the model [66]. These characteristics make LR suitable for rainfall–runoff modeling that involves a wide array of variously correlated meteorological variables [67,68]. In our study, LR was implemented through the Scikit-learn library in Python 3.8.18, with an alpha (regularization strength) of 0.1. The algorithm of LR can be found in Roth [69].

Extreme Gradient Boosting

XGBoost (Extreme Gradient Boosting, XGB) is an optimized distributed gradient boosting library designed for substantial scale machine learning problems with a focus on efficiency, flexibility, and portability [70]. In this study, we utilized Python’s XGBoost library to implement an XGBoost Regressor. The process of optimizing hyperparameters involved a grid search within a predefined range of values. The configuration of the XGB model aimed to minimize the squared error loss. Notably, 80% of features were subsampled for each tree. The specific hyperparameter set included a learning rate of 0.001, a maximum tree depth of 3, an alpha (for L1 regularization) of 10, and a limit of 5000 boosting rounds. To avoid overfitting, the training data were split, with 80% treated as the training set and the remaining 20% as the validation set, using a ‘train_test_split’ with a random state set at 42 for reproducibility. During training, an early stopping mechanism was activated if model performance failed to improve consecutively over 100 rounds, thus stopping the training and preserving the performance-optimized model. The specific algorithm of XGBoost can be found in Mitchell and Frank [71].

Light Gradient Boosting Machine

LightGBM (LGBM) is a powerful gradient-boosting framework developed by Microsoft [72]. In our study, we used Python’s LightGBM library to construct a gradient-boosting model. Like our approach with the XGB model, a grid search was conducted to optimize the hyperparameters. The model was configured with the ‘GBDT’ boosting type to address our regression problem. We employed both ‘l2’ (Mean Squared Error) and ‘l1’ (Mean Absolute Error) as metrics for model evaluation. Specifications included 31 leaves, a maximum of 5000 iterations, and a learning rate of 0.001. Each split considered 90% of features (feature fraction = 0.9). For each boosting round, we implemented bagging with 80% of the data (bagging fraction = 0.8) and updated every five iterations (bagging frequency = 5). The methodology for splitting the training data and implementing early stopping was consistent with our approach in the XGB model. The algorithm of the LGBM can be found in Ke et al. [72].

The six machine learning models selected for this study are widely used in hydrology, and the strengths and limitations of each model are summarized in Table 2.

Table 2. Summary of different machine learning models used in this study.

Model	Category	Strengths	Limitations
LSTM	Deep learning	Handle time series data. Capture complex and non-linear relationships.	Structural complexity. Computationally expensive.
SVM	Super vector machine	Handle high-dimensional data. Memory efficient.	Not suitable for larger datasets. Require careful parameter tuning.
GPR	Regression analysis	Provide uncertainty estimates. Less prone to overfitting.	Need to assume Gaussian noise. Relatively computationally intensive.
LR	Regression analysis	Capture linear relationships. Less prone to overfitting.	Weak ability to capture non-linear relationships. Sensitive to noise.
XGB	Ensemble learning	High efficiency and fast speed. Good performance with large datasets.	Prone to overfitting. Larger data requirements.
LGBM	Ensemble learning	High efficiency and fast speed. Good performance with large datasets.	Prone to overfitting. Larger data requirements.

2.4. Model Simulations

For all the machine learning models employed in this study, both meteorological and hydrological data (shown in Table 1) were split into training and evaluation periods for all sub-basins. The training period starts from January 1954 to December 1986, with a total of 396 months. During the training process for LSTM, XGB, and LGBM, we split 20% of the data as the validation dataset for the early stop mechanism to prevent the model from overfitting. Data normalization was not deemed necessary for the model training and testing based on our investigation, except for the SVM model, where normalization is a required step in this model. SVM, GPR, and LR do not involve a process of training through multiple iterations to gradually approach the optimal solution [55,64,69], and therefore we did not utilize a validation set for early stopping in their training methodology. The evaluation period spans from January 2004 to May 2023 (with missing runoff data across different river sub-basins: 20 months of data are missing in the North River and the West River sub-basins, while 23 months are missing in East River sub-basin), with a total of about 210 months. The period of 1987–2003 was not included in model simulations because of missing observations. The periods for the calibration and evaluation of the WAPABA model were aligned with the training and evaluation periods for machine learning models, respectively.

It is worth noting that we used the term ‘evaluation’ for the assessment of model performance against observations that were not used for model training for both WAPABA and machine learning models. This term is equivalent to ‘testing’ used in many machine learning modeling studies.

In this study, we conducted three sets of model simulations to evaluate the impacts of three input data combinations on the performance of machine learning models. We first used the climatic forcings only (Experiment 1) as model input data, following the design of many previous investigations [73–75]. Since the streamflow of the previous month represents overall river basin conditions (e.g., wetness), using this additional variable, other than the meteorological variables, as the input data might add extra skills to streamflow simulations. As a result, we conducted another two sets of simulations to evaluate model performance in response to using simulated (Experiment 2) or observed (Experiment 3) runoff of the previous month as additional model inputs.

We conducted a preliminary test to understand the impacts of different preceding data on the performance of the LSTM model. The results indicated that using preceding data of one month resulted in better performance than using longer preceding data (Figure S4). Consequently, we adopted a one-month time step as the preceding time step for LSTM simulations in this study. Specifically, input data combinations of each type of simulation

are shown in Table 3. Experiment 1 utilizes 6 meteorological variables (i.e., $P(t)$, $e_a(t)$, $u_2(t)$, $R_n(t)$, $T_{max}(t)$, and $T_{min}(t)$) as the models' input. In Experiments 2 and 3, antecedent runoff of the previous month ($R(t-1)$) is added as an extra input variable to drive these machine learning models. Considering the availability of streamflow observations and the applicability of long-term forecasts, the simulated runoff is utilized as the input of Experiment 2, whereas the observed runoff is used as the input in Experiment 3 during the evaluation period.

Table 3. Input data of three machine learning simulations. The full name of all variables is listed in Table 1. Model output data is $R(t)$, where t denotes the data of the current month and $t-1$ the data of the previous month.

	Input Data	$R(t-1)$ Source
Training and validation data (January 1954–December 1986)		
Experiment 1	$P(t), e_a(t), u_2(t), R_n(t), T_{max}(t), T_{min}(t)$	None
Experiment 2	$P(t), e_a(t), u_2(t), R_n(t), T_{max}(t), T_{min}(t), R(t-1)$	Observed runoff
Experiment 3	Same as Experiment 2	Observed runoff
Evaluation data (January 2004–May 2023)		
Experiment 1	$P(t), e_a(t), u_2(t), R_n(t), T_{max}(t), T_{min}(t)$	None
Experiment 2	$P(t), e_a(t), u_2(t), R_n(t), T_{max}(t), T_{min}(t), R(t-1)$	Simulated runoff
Experiment 3	Same as Experiment 2	Observed runoff

2.5. Evaluation Metrics

In this study, Bias, Root Mean Squared Error (RMSE), Correlation Coefficient (r), and Nash Sutcliffe efficiency coefficient (NSE) [76] were used for model performance evaluation. The formulas of these metrics are presented in Equations (1)–(4).

$$Bias = \frac{1}{T} \sum_{t=1}^T Q_s(t) - \frac{1}{T} \sum_{t=1}^T Q_o(t) \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (Q_s(t) - Q_o(t))^2}{T}} \quad (2)$$

$$r = \frac{\sum_{t=1}^T (Q_o(t) - \bar{Q}_o) (Q_s(t) - \bar{Q}_s)}{\sqrt{\sum_{t=1}^T (Q_o(t) - \bar{Q}_o)^2 \sum_{t=1}^T (Q_s(t) - \bar{Q}_s)^2}} \quad (3)$$

$$NSE = 1 - \frac{\sum_{t=1}^T (Q_o(t) - Q_s(t))^2}{\sum_{t=1}^T (Q_o(t) - \bar{Q}_o)^2} \quad (4)$$

where T is the length of the time series data; $Q_s(t)$ and $Q_o(t)$ denote the simulation and observation at time t ; and \bar{Q}_s and \bar{Q}_o are the time averages of simulations and observations. In this research, the units of Bias and RMSE are mm/month, while r and NSE are unitless measures. BIAS is a metric measuring the average of the difference between simulated and observed streamflow, while RMSE evaluates the average errors in model simulations. r assesses the ability of models to reconstruct the overall timing and magnitude of streamflow, while NSE is a common metric in hydrology for assessing the fit of model simulations to observed data. For simulations matching observations perfectly, their bias and RMSE should be 0, while r and NSE should be 1.

3. Results

3.1. Performance of the WAPABA Model

We first evaluated the performance of the WAPABA model in simulating monthly runoff in the PRB. Figure 2 presents the WAPABA simulations and runoff observations in the North, East, and West River sub-basins during the calibration and evaluation periods.

The performance of the WAPABA model varies across the three sub-basins, as suggested by the evaluation metrics (Figure 2). During the calibration period, the North River sub-basin has a low r and NSE of 0.59 and 0.33, respectively, as a result of the underestimated peak flows, whereas the other two sub-basins all show r above 0.8 and NSE above 0.6, with the highest values (0.88 and 0.77 for r and NSE, respectively) found in the West River sub-basin.

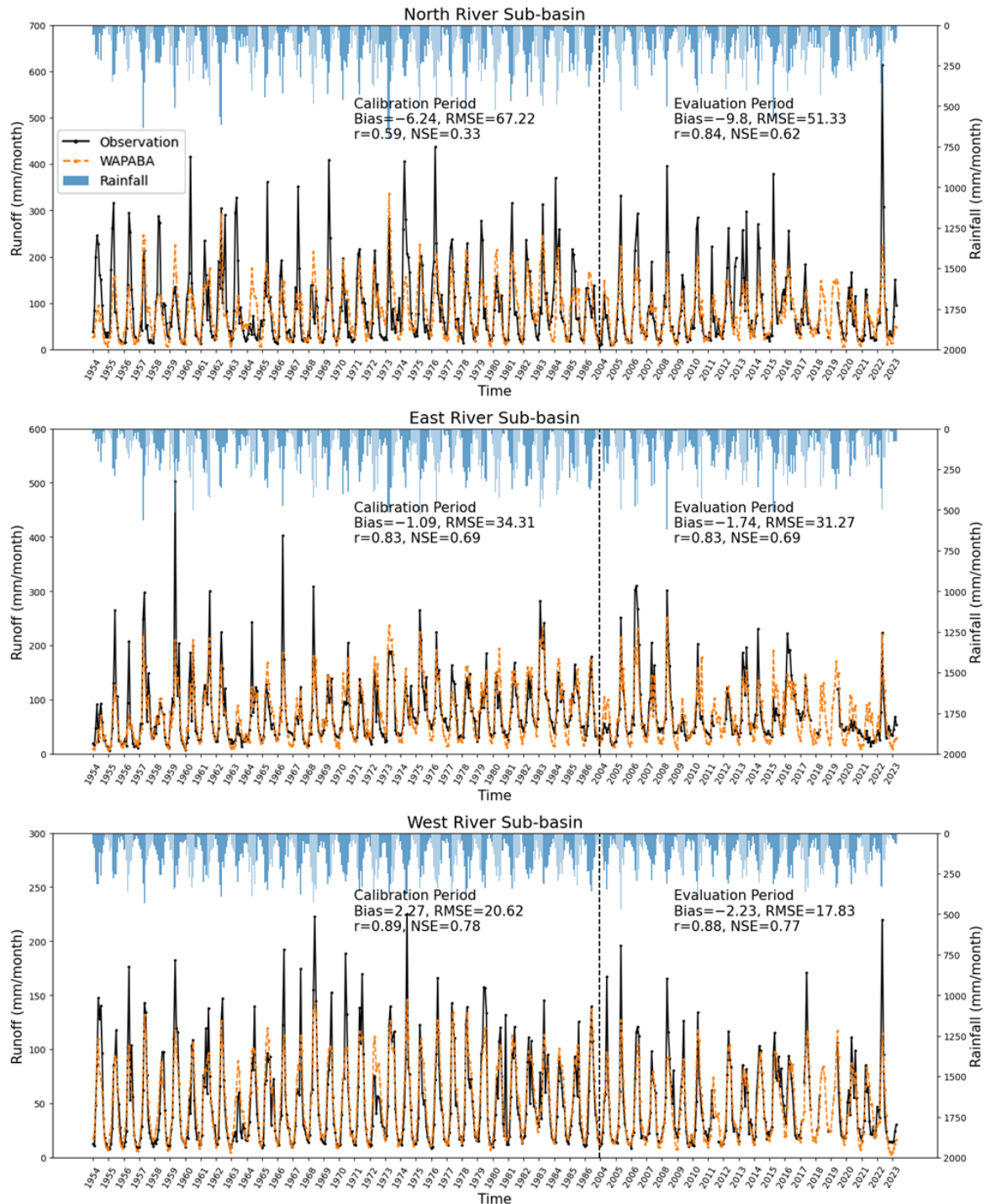


Figure 2. Evaluation of rainfall–runoff simulation by the WAPABA model against observations among different sub-basins in calibration (January 1954–December 1986) and evaluation (January 2004–May 2023) periods.

Compared with the calibration period, the WAPABA model shows comparable or even better performance during the evaluation period. We found marked improvements

in WAPABA simulations during the evaluation period for the North River sub-basin, as suggested by the 42% increases in r and 88% increases in NSE. The bias during the evaluation period is -9.80 mm/month, -1.74 mm/month, and -2.23 mm/month in the North, East, and West River sub-basins, respectively, indicating that the WAPABA model underestimated the streamflow, particularly for peak flows in the three sub-basins. The values of r are higher than 0.8 in all river sub-basins during the evaluation period (0.84, 0.83, and 0.88 in the North, East, and West River sub-basins, respectively), proving that the WAPABA reconstructed seasonal variations in observations well. When using NSE to evaluate the performance of hydrological simulations, it is generally considered that an NSE greater than 0.5 indicates a satisfactory model performance [77]. In this study, the NSEs for three sub-basins during the evaluation period are 0.62, 0.69, and 0.77, respectively, confirming that the WAPABA model performed satisfactorily in the study area in the evaluation period. However, it is worth mentioning that NSE is only 0.33 for the North River sub-basin during the calibration period. This may result from the relatively poorer quality of observation in this sub-basin as well as other factors not considered in WAPABA, such as land use changes and anthropogenic activities, which may have larger impacts in this sub-basin. Overall, considering that the steps during the modeling are correct and that the performance is acceptable during the evaluation period, we think the WAPABA model's performance is acceptable for evaluating the relative performance of machine learning models.

3.2. Simulation of Machine Learning Models Based on Climate Forcings Only

The time series of the observed and simulated runoff from multiple machine learning models in Experiment 1 are shown in Figure 3. Here, we only present results from 2021 to 2022 during the evaluation period to make sure the time series information is readable, while results for the entire training and evaluation periods are shown in Figure S5 of the Supplementary Material. Evaluation metrics for both the training and evaluation periods are shown in Table 4.

In Experiment 1, all machine learning models demonstrate the best performance in the West River sub-basin and the poorest performance in the North River sub-basin, consistent with that of the WAPABA simulation (Figure 2). These results suggest that despite differences in the structures and algorithms of the models, using the same input data can lead to similar simulation outcomes across different river basins. The differences between model simulations and observations might be attributable to unaccounted hydrological processes by the models in the sub-basin, such as anthropogenic disturbances (reservoir operation and water withdrawal) on natural water cycling.

Compared with the WAPABA model, all machine learning models of Experiment 1 exhibit noticeably higher RMSE and lower r and NSE during the evaluation period in the North and the East River sub-basins. Specifically, the average RMSE, r , and NSE values of machine learning models in the North River sub-basin are 63.31, 0.65, and 0.42, respectively, while the WAPABA model shows values of 51.33, 0.84, and 0.62 during the calibration period. In the East River sub-basin, machine learning models produce evaluation metrics of 37.82, 0.75, and 0.54, slightly worse than the WAPABA model's values of 31.27, 0.83, and 0.69, for RMSE, r , and NSE, respectively. Evaluation metrics show slightly worse performance of the machine learning models relative to WAPABA in the West River sub-basin. In this sub-basin, the average RMSE, r , and NSE values of machine learning models are 18.70, 0.87, and 0.75, respectively, and the WAPABA model shows values of 17.83, 0.88, and 0.77. The results indicate that machine learning methods in Experiment 1 underperformed compared to the WAPABA model when the model input included meteorological variables only. It is noteworthy that, except for the LR model, all machine learning models exhibited NSE values less than 0.5 in the North River sub-basin, indicating the limitations of Experiment 1 in simulating the streamflow of this sub-basin.

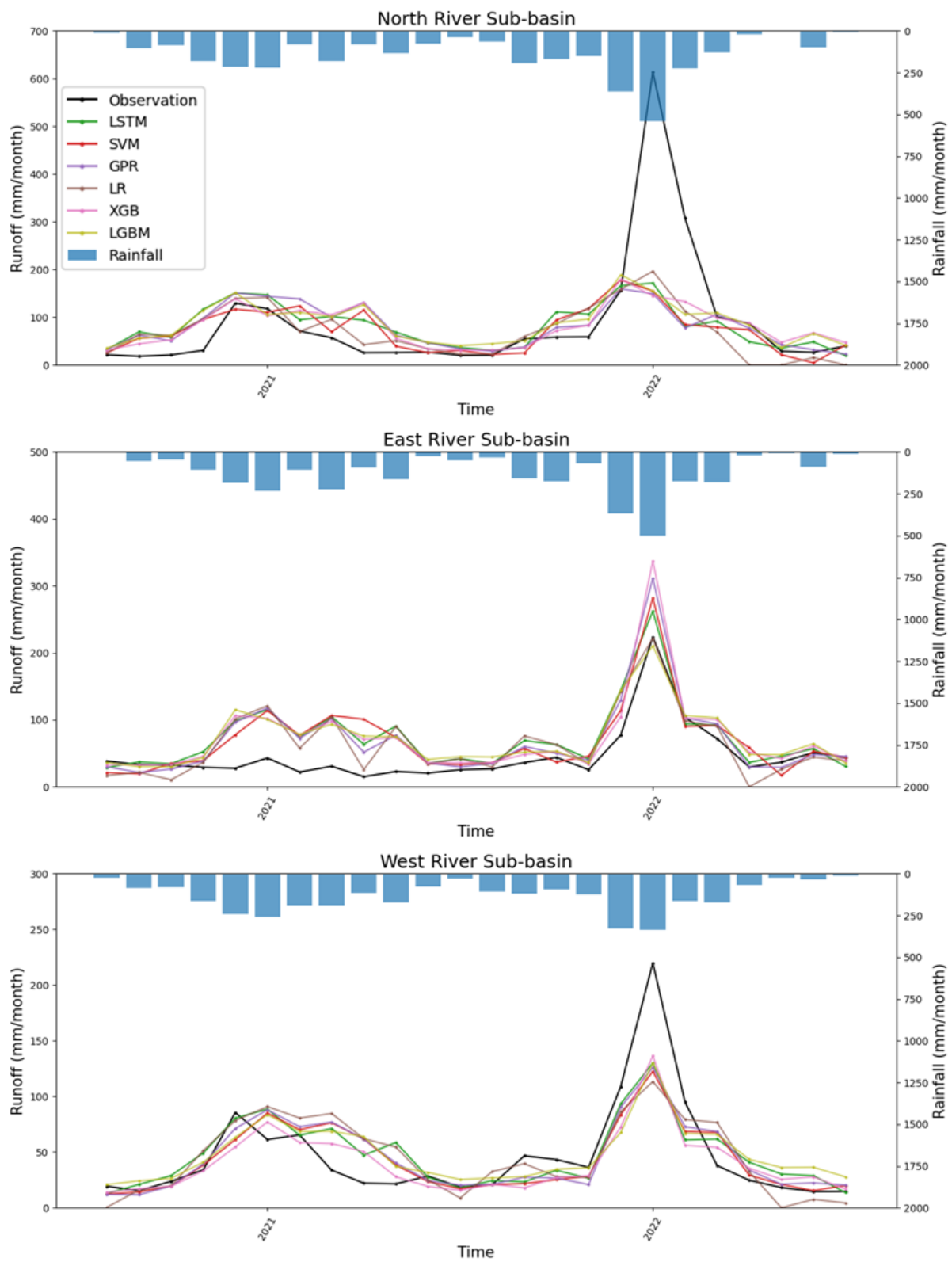


Figure 3. Runoff simulations by multiple machine learning models using meteorological forcings alone as input data (Experiment 1) against observations among different sub-basins during 2021–2022.

Table 4. Performance of machine learning models in Experiment 1 during the training (January 1954 to December 1986) and the evaluation period (January 2004 to May 2023) in different sub-basins. The unit of Bias and RMSE is mm/month, and the other two evaluation metrics (r and NSE) are unitless.

	LSTM	SVM	GPR	LR	XGB	LGBM
North River Sub-Basin						
Training Period						
Bias	1.33	−7.46	−0.10	0.00	−4.86	0.96
RMSE	59.12	59.31	58.72	62.37	47.62	52.01
r	0.70	0.70	0.70	0.65	0.83	0.79
NSE	0.48	0.48	0.49	0.42	0.66	0.60
Evaluation Period						
Bias	2.45	−8.23	−1.04	−4.26	−2.34	3.59
RMSE	62.53	64.68	64.41	57.31	66.20	64.71
r	0.67	0.64	0.64	0.74	0.60	0.63
NSE	0.43	0.39	0.40	0.52	0.36	0.39
East River Sub-Basin						
Training Period						
Bias	1.34	−4.84	0.00	0.01	−2.58	1.24
RMSE	33.85	32.18	31.20	35.80	25.01	31.20
r	0.84	0.86	0.86	0.81	0.92	0.87
NSE	0.70	0.73	0.74	0.66	0.83	0.74
Evaluation Period						
Bias	2.21	−3.77	−2.16	−2.52	0.56	0.92
RMSE	36.14	37.46	37.61	36.24	40.92	38.53
r	0.77	0.75	0.75	0.77	0.71	0.72
NSE	0.58	0.55	0.55	0.58	0.46	0.52
West River Sub-Basin						
Training Period						
Bias	0.77	−2.12	0.00	0.22	−5.75	0.22
RMSE	22.25	20.55	20.60	22.83	19.64	21.04
r	0.86	0.89	0.88	0.85	0.91	0.89
NSE	0.74	0.78	0.78	0.73	0.80	0.77
Evaluation Period						
Bias	0.49	−3.25	−2.83	−2.41	−6.26	1.10
RMSE	17.83	17.99	18.64	19.12	19.36	19.25
r	0.88	0.88	0.87	0.86	0.87	0.86
NSE	0.77	0.76	0.75	0.73	0.73	0.73

3.3. Simulations of Machine Learning Models with Antecedent Runoff Input

Meteorological conditions play a crucial role in determining water entering (e.g., precipitation) and leaving (e.g., evapotranspiration) river basins. However, using these variables as the only input for machine learning modeling could not represent water storage in river basins well. This limitation may have resulted in the worse performance of machine learning models (Experiment 1) than the WAPABA model (Figures 2 and 3, Table 4). To better represent the process of water storage within the river sub-basin, we used runoff of the previous month as an additional input variable to drive machine learning simulations.

The evaluation metrics of three experiments during the evaluation period are shown in Figure 4. Evaluation metrics during the training period are shown in Figure S6 of the Supplementary Material. The simulated streamflow by multiple machine learning models in Experiments 2 and 3 are shown in Figures S7 and S8 of the Supplementary Material.

The evaluation metrics of Experiment 2 are shown in Table S2 of Supplementary Material, while the metrics of Experiment 3 are shown in Table 5.

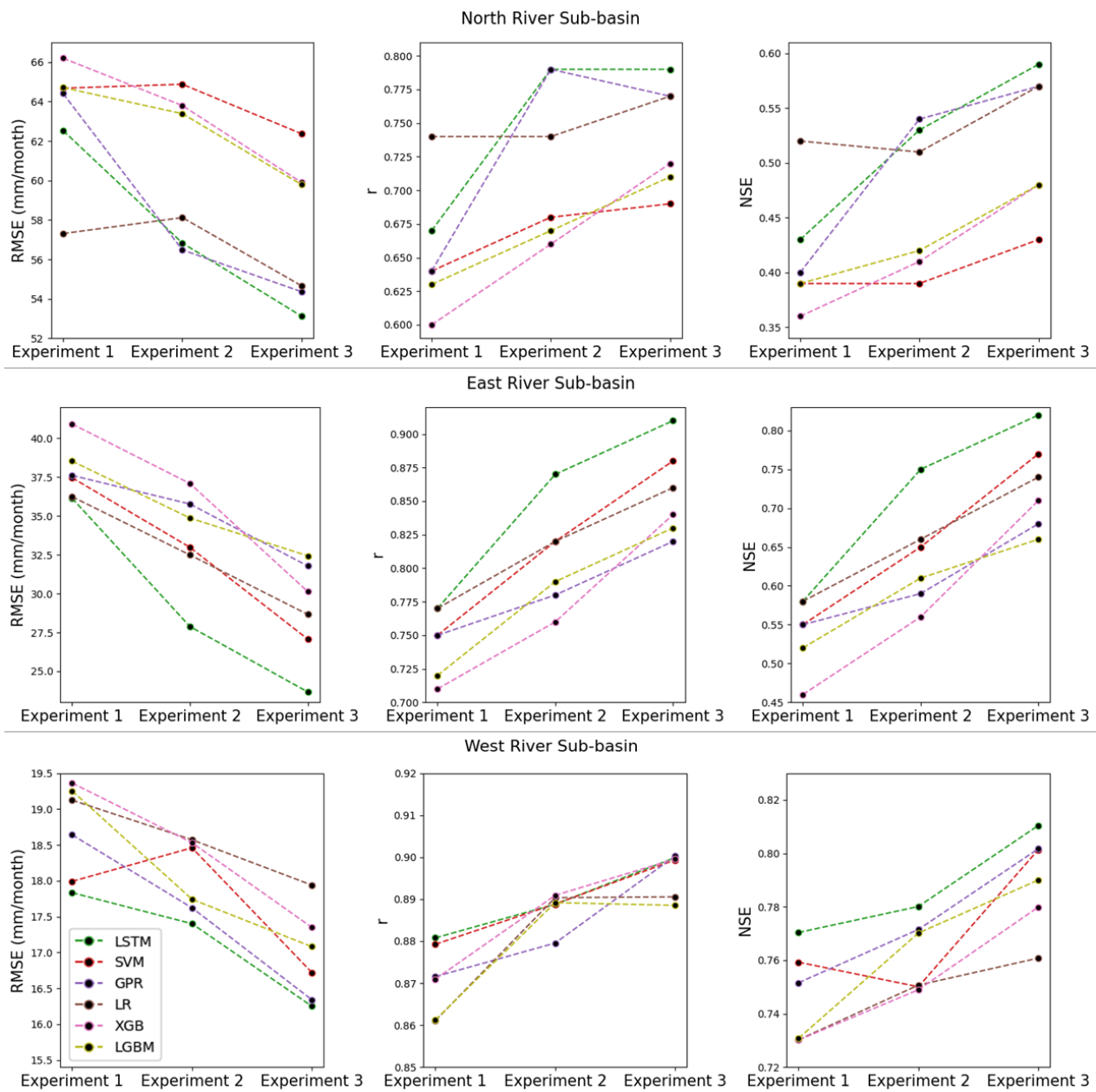


Figure 4. RMSE, r , and NSE of different machine learning models in Experiments 1, 2, and 3 during the evaluation period (legend is shown in the subplot of the first column and the third row).

In each sub-basin, most of the models have lower RMSE as well as higher r and NSE in Experiments 2 and 3 compared with those of Experiment 1, indicating that adding runoff of the previous month as an additional input variable effectively improves the performance of machine learning models (Figure 4). Furthermore, almost all models in Experiment 3 demonstrate better performance than those in Experiment 2, as suggested by the lower RMSE as well as higher r and NSE. Therefore, including the previous month’s runoff could significantly improve the performance of machine learning models, and using observed runoff data could achieve better performance than using simulated runoff.

Table 5. Evaluation metrics for machine learning models in Experiment 3 and the WAPABA model during the evaluation period in different river sub-basins. The unit of Bias and RMSE is mm/month, and the other two evaluation metrics (r and NSE) are unitless. Bold text indicates models with better performance.

	WAPABA	LSTM	SVM	GPR	LR	XGB	LGBM
North River sub-basin							
Bias	−9.80	−3.13	−13.19	−2.90	−7.03	−6.11	−2.52
RMSE	51.33	53.12	62.37	54.36	54.66	59.90	59.79
r	0.84	0.79	0.69	0.77	0.77	0.72	0.71
NSE	0.62	0.59	0.43	0.57	0.57	0.48	0.48
East River sub-basin							
Bias	−1.74	−2.71	−4.77	−2.59	−3.69	−0.21	−0.33
RMSE	31.27	23.67	27.07	31.78	28.66	30.13	32.43
r	0.83	0.91	0.88	0.82	0.86	0.84	0.83
NSE	0.69	0.82	0.77	0.68	0.74	0.71	0.66
West River sub-basin							
Bias	−2.23	−1.85	−4.32	−1.56	−4.32	−4.93	0.10
RMSE	17.83	16.25	16.72	16.34	17.94	17.35	17.08
r	0.88	0.90	0.90	0.90	0.89	0.90	0.89
NSE	0.77	0.81	0.80	0.80	0.76	0.78	0.79

3.4. Comparison between Machine Learning Models and WAPABA

Figure 5 and Table 5 show the comparison of Bias and NSE of simulations by machine learning models and the WAPABA model during the evaluation period. Except for the SVM, XGB, and LGBM models in the North River sub-basin, the NSE of all model simulations is higher than 0.5, showing the acceptable performance of machine learning models from Experiment 3 in simulating runoff. Absolute biases are generally lower than 10 mm/month or even 5 mm/month for most model simulations. Among all machine learning models, LSTM demonstrates the highest NSE, which is also higher than that of the WAPABA model in the East River and West River sub-basins. In the North River sub-basin, although the WAPABA simulation has a larger bias than five of the six machine learning models, the NSE, r , and RMSE suggest that it has slightly better performance than the machine learning models (Figure 5).

We further visualize the performance of the machine learning models and WAPABA using Taylor diagrams (Figure 6). Across the three sub-basins, the standard deviation of observed runoff is higher than that in all model simulations, mainly because of the underestimation of the peak runoff during the wet seasons (e.g., in June 2022). In the North River sub-basin, the LSTM model exhibits the lowest RMSE and highest r among all machine learning models, but it performs slightly worse than the WAPABA model in terms of these two metrics. In the East and West River sub-basins, the WAPABA model shows higher RMSE and lower r than machine learning models, indicating worse performance. LSTM still exhibits the lowest RMSE and highest r in the East River sub-basin, while all machine learning models demonstrate comparable results in these two metrics in the West River sub-basin.

Figure 7 compares the observed and simulated Flow Duration Curves (FDCs) from the WAPABA model as well as the LSTM method, which show better performance than other machine learning models in Experiment 3 (the FDCs for the remaining models are shown in Figure S9 in the Supplementary Material). The FDC depicts the probability that a given runoff quantity is exceeded. Overall, the FDC for the WAPABA and the two best-performing machine learning models demonstrate strong consistency with the observed runoff, particularly during the dry seasons. However, for high-flow events, where the exceedance probability is less than approximately 10%, model simulations in each basin are significantly lower than observations. This suggests that these models underestimate runoff

during the wet seasons. The LSTM model shows runoff values that are approximately 10% higher than those of the WAPABA model at low exceedance probabilities, suggesting that its performance during wet seasons is slightly superior to that of the WAPABA model. For low-flow events with exceedance probabilities higher than approximately 90%, both LSTM and WAPABA are able to accurately reconstruct the magnitude of runoff in the North River sub-basin. However, in the other two sub-basins, WAPABA tends to underestimate streamflow in dry seasons by around 20%, while LSTM demonstrates better performance in capturing the low streamflow.

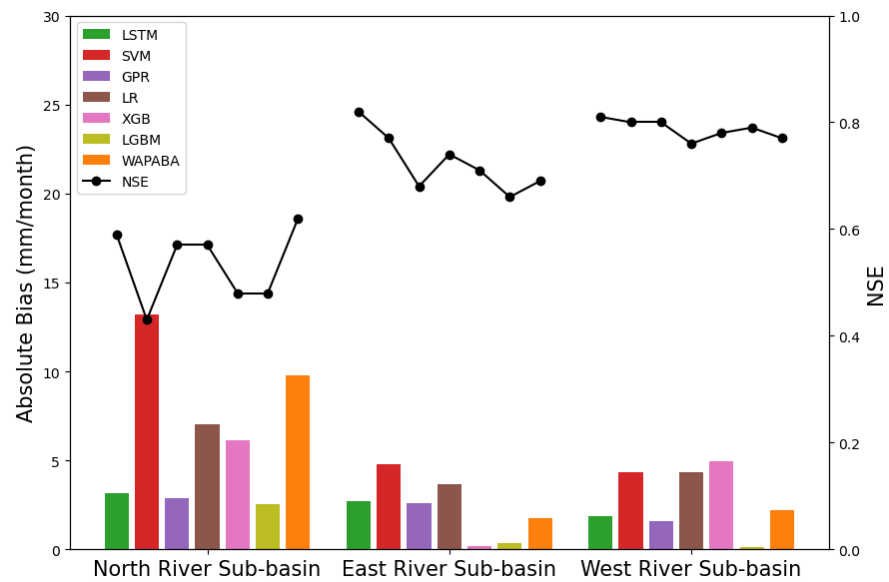


Figure 5. Comparison of runoff simulations by WAPABA and the six machine learning models from Experiment 3. Color bars show absolute bias, and dotted lines indicate the NSE of the simulations.

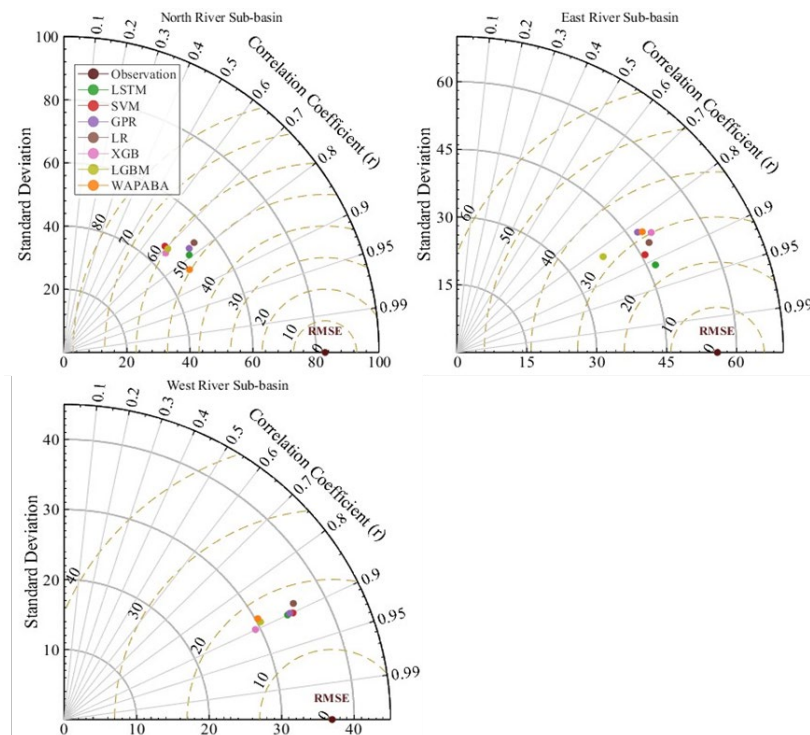


Figure 6. Taylor diagram of machine learning models from Experiment 3 and WAPABA models across three sub-basins during the evaluation period (the units of standard deviation and RMSE are mm/month).

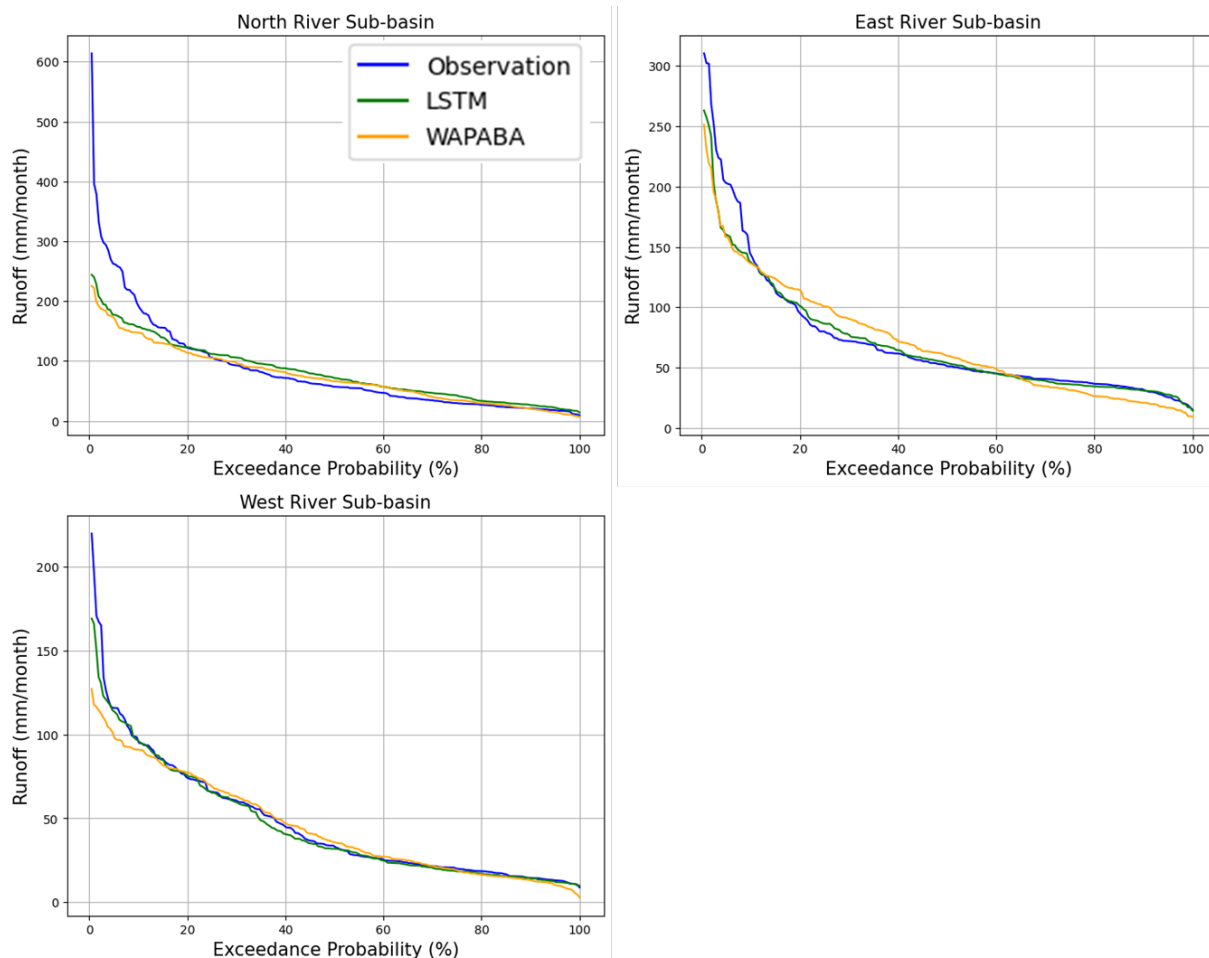


Figure 7. The Flow Duration Curves (FDCs) of observations as well as simulations by LSTM and WAPABA in Experiment 3 in the North, East, and West River sub-basins. The x -axis represents the exceedance probability, indicating the probability of runoff exceeding a specific runoff level shown by the y -axis.

In summary, based on the selected evaluation metrics (RMSE, r , and NSE) and the FDC, LSTM performs the best among all machine learning models tested in this study (Figures 5 and 6). Compared with the WAPABA model, LSTM shows similar performance as that of WAPABA in the North River sub-basin, but it performs better than WAPABA in the East and West River sub-basins. For simulations in wet seasons, The LSTM model shows slightly improved performance relative to the WAPABA model. The East River and West River sub-basins also have better performance with the LSTM model relative to WAPABA.

4. Discussion

4.1. Performance of Monthly Runoff Simulations

This investigation in the subtropical zone indicates that the overall performance of machine learning models is better than or comparable with a widely used rainfall–runoff model. The WAPABA model has been applied in climate change impact assessment and seasonal runoff forecasting [46–48]. This study further confirms the effectiveness of the model in applications in subtropical regions. Compared with the WAPABA model, the overall performance of machine learning models is slightly poorer in the North River sub-basin but better in the East and West River sub-basins. The comparable performance between the LSTM and the WAPABA model is consistent with the conclusions drawn by

Clark et al. [37], who found that the LSTM model performed better in 69% of the catchments in Australia when compared to the WAPABA model.

During the wet seasons, WAPABA demonstrates a larger negative bias in the extreme runoff simulations compared to LSTM. This indicates that machine learning models have greater potential in improving the prediction of extreme events, as reported by Frame et al. [38], who demonstrated that the performance of the LSTM model in simulating streamflow of high return periods was better than a conceptual model and a process-based model across the United States.

With the increasing use of machine learning-based models in hydrological modeling in recent years, concerns are growing about the performance of such data-driven models [18,78]. Our study confirms the suitability of machine learning models for rainfall–runoff simulation in subtropical basins, especially the LSTM approach, which performs comparably or better than the conceptual rainfall–runoff model.

4.2. Deep Learning in Rainfall–Runoff Modeling

Deep learning models provide a powerful method for learning and representing complex, non-linear patterns from data, which is mainly due to their multilayer structure and non-linear activation functions [79]. Models like LSTM consist of multiple hidden layers of neurons, each of which takes the output from the previous layer as input and processes it by utilizing non-linear activation functions, enabling the network to capture a variety of complex processes [80]. Meanwhile, the stacked structure of these models facilitates the understanding and learning of complex interactions between different features [81]. These characteristics equip deep learning models with more robust capability in handling complex and non-linear problems, such as hydrological processes, when compared to linear models (e.g., LR) and models dealing with simple nonlinearities (e.g., SVM and GPR).

Among the six distinct machine learning models investigated in this study, LSTM, which belongs to the category of deep learning, exhibits superior performance in simulating the rainfall–runoff relationship across different river basins. This result suggests that this model is more suitable for monthly streamflow simulations than other machine learning models. The superior performance of the LSTM from this study is in line with the fact that it has been widely applied in hydrology [82]. This finding also provides valuable implications for the selection of machine learning models in future hydrological investigations.

4.3. Strategies for Setting Up Machine Learning Models in Rainfall–Runoff Modeling

Compared with traditional hydrological models, which simulate key components of the terrestrial water cycle (e.g., infiltration, evapotranspiration, and groundwater recharge), insufficient representation of water accumulation in water pools is limiting the performance of machine learning models in simulating streamflow. Our Experiment 1 suggested that when using meteorological variables alone as inputs, machine learning-based simulations could not reconstruct streamflow well. Before runoff generation, precipitation undergoes infiltration, evapotranspiration, and exchanges with basin storages. In these processes, water storage in soil and aquifers significantly modulates the amount of water that eventually becomes runoff and shapes the temporal patterns of streamflow [83]. Traditional hydrological models, such as WAPABA, use state variables (e.g., S_{max} and K) to simulate the accumulation of water in these pools and the subsequent release of water to river channels. However, machine learning methods do not have such variables, limiting this type of model in simulating temporally continuous processes [16]. Using meteorological variables as the only input will not allow machine learning models to simulate water storage in soils and groundwater pools (Figure 3 and Table 4). This could explain the inferior performance of these machine learning models relative to that of the WAPABA model in Experiment 1 (Table 4).

In addition, although including river basin properties as input variables contributes to improving streamflow simulations with machine learning models, these datasets are not always readily available, challenging the inclusion of basin-specific information in hydro-

logical modeling based on machine learning models. To account for the spatial variability of streamflow, watershed characteristics (e.g., plant growth, reservoir operations, soils, and topography) are also treated as predictors of streamflow modeling based on machine learning models across multiple river basins. However, using watershed properties as inputs often may require substantial efforts in data collection [84–86].

Our modeling experiments provide valuable information for dealing with the above two limitations in hydrological modeling using machine learning models. The streamflow of the previous month carries the information on how precipitation is transformed into runoff and thus could be used as a surrogate variable to represent the impacts of both water storage and watershed properties on streamflow [87]. The improvement in model performance with the adoption of the previous month's runoff as model input confirms the necessity of dealing with the intrinsic limitations of machine learning models in modeling continuous hydrological processes and provides a cost-effective solution for considering watershed properties in streamflow modeling. The strategy (Experiments 2 and 3) tested in this study could be used in future applications of machine learning models in streamflow simulation.

Success in improving the performance of machine learning models by incorporating streamflow observation from the previous month (e.g., Experiment 3) indicates the advantage of such data-driven models. Recent studies have demonstrated that applying machine learning techniques can significantly improve the performance of the original model in data assimilation in the Earth's system learning [88,89]. The strategy of incorporating previous observational data information to enhance the performance of data-driven models can be adopted by future applications of machine learning models.

Machine learning methods are often more computationally efficient than process-based models. In addition, using machine learning models does not require expertise and understanding of hydrological processes, making them quickly adopted by a broad range of stakeholders. Such conveniences and their rapidly evolving power in simulating complex processes make machine learning models very promising techniques in helping move hydrologic modeling forward.

Although the WAPABA model exhibited worse performance than the LSTM model in our study, its advantage in simulating the physical processes of the water cycle should not be ignored. The WAPABA model employs simplified physical equations to represent water storage processes within watersheds, providing a degree of physical constraint and interpretability, which are pitfalls of machine learning models. However, the oversimplification of these processes in WAPABA may also introduce uncertainties to runoff simulations. Considering the advantages and disadvantages of both types of models, we believe that developing hybrid modeling frameworks to couple both process-based and data-driven models could be a promising direction in future hydrological modeling.

More broadly, this study further emphasizes the importance of a physics-guided setting in machine learning modeling. Traditionally, many users adopt the machine learning model as a black-box approach, focusing on the input and output only, without learning the physical processes. While such a black-box approach can still provide reasonable simulations in many cases, increasing investigations suggest the necessity of reflecting a certain degree of the physical processes in the applications of such models [90,91]. Our investigation suggested that a good understanding of the physical processes is essential for maximizing machine learning models' capability in simulating complex processes.

4.4. Future Work

Although this investigation confirms the suitability of machine learning models in monthly rainfall–runoff modeling, especially LSTM techniques, further investigations are still required to address a few challenges in the application of machine learning models in hydrology.

First, machine learning modeling processes that map directly from inputs to outputs can obscure basin-specific details, which is unfeasible for fully characterizing the individu-

ality of watersheds [92,93]. Thus, traditional data-driven models should be complemented with physics-informed approaches as they lack physical constraints and interpretability. In recent years, researchers have begun to work on building a hybrid framework that combines physical processes with deep learning structures to address such limitations [16,94]. This framework has been utilized in rainfall–runoff modeling [18], revealing flood mechanisms [34,95], and predicting groundwater levels [96,97]. These works reveal physical mechanisms to some extent while using machine learning structures in hydrologic modeling, which is at the frontier of research in this field.

Second, applications of machine learning models in simulating high flows need to be further investigated. Machine learning models require ample data for optimal performance. However, the mechanism regulating extreme runoff differs from the rainfall–runoff relationship of normal flows [98,99]. As a result, models trained on normal flow periods may not be suitable for predicting extreme flows. A possible solution to this limitation is to generate ‘virtual’ extreme events to increase training samples, thereby improving the model’s prediction of extreme events [100]. In addition, using generative AI, such as the Generative Adversarial Network (GAN), has the potential to simulate large flow events but requires additional testing in future studies [101].

5. Conclusions

In this study, we carried out a comprehensive evaluation of the performance of six machine learning models in simulating monthly runoff across three subtropical sub-basins of the Pearl River Basin. These models’ performance compared to a traditional hydrological model, different strategies for setting up model simulations, and model performance in simulating extreme runoff were thoroughly evaluated. The findings of this study include:

- (1) LSTM performs better in simulating runoff in the PRB relative to the other five machine learning models (SVM, GPR, LR, XGB, LGBM), with about 11.7%, 5.1%, and 10.8% improvements in RMSE, r , and NSE, respectively.
- (2) Adding the previous month’s runoff as an additional input variable can achieve better performance than using meteorological forcing only, and using observed runoff as input can achieve better performance than using simulated runoff of the last month.
- (3) LSTM outperforms the WAPABA model in two out of three sub-basins. Although all models underestimate the peak streamflow, the performance of the LSTM model is slightly better than that of WAPABA in all sub-basins during wet seasons. Additionally, the LSTM performs slightly better than WAPABA in the East River and West River sub-basins.

The findings of this study on the relative performance of different machine learning models to a conventional hydrological model provide evidence of the suitability of using these models in hydrological studies in subtropical regions. The strategies for setting up the model will help guide the future use of machine learning models in simulating runoff and hydrological forecasting. We suggest that future research could expand upon the methodologies used in this study by incorporating more comprehensive input data, such as a longer range of antecedent conditions and basin characteristics, into machine learning models. Such enhancements could potentially resolve the limitations exhibited in the present study and further increase the performance and interpretability of machine learning models.

Supplementary Materials: The following Supplementary Material can be downloaded at: <https://www.mdpi.com/article/10.3390/w16152199/s1>, Figure S1: Schematic diagram of the WAPABA (water partition and balance) model. P means precipitation, and ET means reference crop evapotranspiration; Figure S2: Schematic diagram of the Long Short-Term Memory Network; Figure S3: Schematic diagram of the Support Vector Machine; Figure S4: RMSE, NSE, and r for different input time lags (months ahead) for the LSTM model; Figure S5: Runoff simulations using multiple machine learning models in Experiment 1 against observations among different sub-basins in training (1954–1986) and evaluation (2004–2023) periods; Figure S6: Root Mean Squared Error (RMSE), Correlation Coefficient

(r), and Nash Sutcliffe efficiency coefficient (NSE) of different machine learning models in Experiments 1, 2, and 3 during the training period. Note that Experiments 2 and 3 have the same training results; Figure S7: Runoff simulations using multiple machine learning models in Experiment 2 against observations among different sub-basins in training (1954–1986) and evaluation (2004–2023) periods; Figure S8: Runoff simulations using multiple machine learning models in Experiment 3 against observations among different sub-basins in training (1954–1986) and evaluation (2004–2023) periods; Figure S9: The Flow Duration Curves (FDCs) of observations and simulations by all machine learning models and WAPABA in Experiment 3 in the North, East, and West River sub-basins. The x-axis represents the exceedance probability, indicating the probability that a specific runoff amount equals or exceeds a given runoff level shown on the y-axis; Table S1: Statistical summary of variables used in this study. ‘Mean’ refers to the average value, ‘Std’ refers to the standard deviation. ‘Min’ refers to the minimum value in the data. ‘25%’, ‘50%’, and ‘75%’ refer to the 25th, 50th, and 75th percentiles, indicating that 25%, 50%, and 75% of the data values are lower than this value, respectively. ‘Max’ refers to the maximum value in the data; Table S2: Evaluation metrics results among each machine learning model in Experiment 2 during the training (January 1954 to December 1986) and evaluation periods (January 2004 to May 2023) in different river sub-basins. The unit of Bias and RMSE is mm/month, and the other two evaluation metrics (r and NSE) are unitless. Note that the training results of Experiment 3 are the same as those in Experiment 2.

Author Contributions: Conceptualization, Q.Y.; methodology, H.Y. and Q.Y.; software, H.Y. and Q.Y.; validation, H.Y. and Q.Y.; formal analysis, H.Y. and Q.Y.; investigation, Q.Y.; resources, Q.Y.; data curation, Q.Y.; writing—original draft preparation, H.Y. and Q.Y.; writing—review and editing, Q.Y.; visualization, H.Y.; supervision, Q.Y.; project administration, Q.Y.; funding acquisition, Q.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hongkong-Macau Center of Ocean Research (CORE) 2023 program (CORE is a joint research center for ocean research between Laoshan Laboratory and HKUST), the Guangzhou Technology Bureau and Hongkong University of Science and Technology 2023 joint program (SL2024A03J00999), and the Chinese Academy of Science Earth System simulator program (elpt_2023_000430). The work described in this paper was substantially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/P-601/23-N).

Data Availability Statement: Data in this study are accessible from public resources. The meteorological data are derived from ECMWF ERA5-Land (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land>, accessed on 10 June 2023). The boundaries of river sub-basins and streamflow gauge stations records are derived from GRDC (https://grdc.bafg.de/GRDC/EN/Home/homepage_node.html, accessed on 10 June 2023) and the Ministry of Water Resources, PRC (<http://xxfb.mwr.cn/>, accessed on 10 June 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Geogr. Sci.* **1999**, *4*, 135–143. [[CrossRef](#)]
2. Sajikumar, N.; Remya, R. Impact of land cover and land use change on runoff characteristics. *J. Environ. Manag.* **2015**, *161*, 460–468. [[CrossRef](#)] [[PubMed](#)]
3. Liu, J.; Gao, G.; Wang, S.; Jiao, L.; Wu, X.; Fu, B. The effects of vegetation on runoff and soil loss: Multidimensional structure analysis and scale characteristics. *J. Geogr. Sci.* **2018**, *28*, 59–78. [[CrossRef](#)]
4. Zhang, G.; Chan, K.; Oates, A.; Heenan, D.; Huang, G. Relationship between soil structure and runoff/soil loss after 24 years of conservation tillage. *Soil Tillage Res.* **2007**, *92*, 122–128. [[CrossRef](#)]
5. Dallison, R.J.; Patil, S.D.; Williams, A.P. Impacts of climate change on future water availability for hydropower and public water supply in Wales, UK. *J. Hydrol. Reg. Stud.* **2021**, *36*, 100866. [[CrossRef](#)]
6. Masseroni, D.; Cislighi, A.; Camici, S.; Massari, C.; Brocca, L. A reliable rainfall–runoff model for flood forecasting: Review and application to a semi-urbanized watershed at high flood risk in Italy. *Hydrol. Res.* **2017**, *48*, 726–740. [[CrossRef](#)]
7. Mishra, V.; Aadhar, S.; Shah, H.; Kumar, R.; Pattanaik, D.R.; Tiwari, A.D. The Kerala flood of 2018: Combined impact of extreme rainfall and reservoir storage. *Hydrol. Earth Syst. Sci. Discuss.* **2018**, *2018*, 1–13.
8. He, C.; Chen, F.; Long, A.; Qian, Y.; Tang, H. Improving the precision of monthly runoff prediction using the combined non-stationary methods in an oasis irrigation area. *Agric. Water Manag.* **2023**, *279*, 108161. [[CrossRef](#)]

9. Tabari, H. Climate change impact on flood and extreme precipitation increases with water availability. *Sci. Rep.* **2020**, *10*, 13768. [[CrossRef](#)] [[PubMed](#)]
10. Wasko, C.; Nathan, R.; Stein, L.; O'Shea, D. Evidence of shorter more extreme rainfalls and increased flood variability under climate change. *J. Hydrol.* **2021**, *603*, 126994. [[CrossRef](#)]
11. Pokhrel, Y.; Felfelani, F.; Satoh, Y.; Boulange, J.; Burek, P.; Gädeke, A.; Gerten, D.; Gosling, S.N.; Grillakis, M.; Gudmundsson, L. Global terrestrial water storage and drought severity under climate change. *Nat. Clim. Chang.* **2021**, *11*, 226–233. [[CrossRef](#)]
12. Abraham, R.J.; See, L.M.; Solomatine, D.P. *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008; Volume 68.
13. Nearing, G.S.; Kratzert, F.; Sampson, A.K.; Pelissier, C.S.; Klotz, D.; Frame, J.M.; Prieto, C.; Gupta, H.V. What Role Does Hydrological Science Play in the Age of Machine Learning? *Water Resour. Res.* **2021**, *57*, e2020WR028091. [[CrossRef](#)]
14. Liu, J.; Yuan, X.; Zeng, J.; Jiao, Y.; Li, Y.; Zhong, L.; Yao, L. Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 265–278. [[CrossRef](#)]
15. Gauch, M.; Kratzert, F.; Gilon, O.; Gupta, H.; Mai, J.; Nearing, G.; Tolson, B.; Hochreiter, S.; Klotz, D. In Defense of Metrics: Metrics Sufficiently Encode Typical Human Preferences Regarding Hydrological Model Performance. *Water Resour. Res.* **2023**, *59*, e2022WR033918. [[CrossRef](#)] [[PubMed](#)]
16. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]
17. Adnan, R.M.; Petroselli, A.; Heddam, S.; Santos, C.A.G.; Kisi, O. Comparison of different methodologies for rainfall–runoff modeling: Machine learning vs. conceptual approach. *Nat. Hazards* **2021**, *105*, 2987–3011. [[CrossRef](#)]
18. Herath, H.M.V.V.; Chadalawada, J.; Babovic, V. Hydrologically informed machine learning for rainfall–runoff modelling: Towards distributed modelling. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 4373–4401. [[CrossRef](#)]
19. Kao, I.F.; Zhou, Y.; Chang, L.-C.; Chang, F.-J. Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. *J. Hydrol.* **2020**, *583*, 124631. [[CrossRef](#)]
20. Wang, J.-H.; Lin, G.-F.; Chang, M.-J.; Huang, I.H.; Chen, Y.-R. Real-Time Water-Level Forecasting Using Dilated Causal Convolutional Neural Networks. *Water Resour. Manag.* **2019**, *33*, 3759–3780. [[CrossRef](#)]
21. Bai, Y.; Bezak, N.; Sapač, K.; Klun, M.; Zhang, J. Short-Term Streamflow Forecasting Using the Feature-Enhanced Regression Model. *Water Resour. Manag.* **2019**, *33*, 4783–4797. [[CrossRef](#)]
22. Dams, J.; Nossent, J.; Senbeta, T.B.; Willems, P.; Batelaan, O. Multi-model approach to assess the impact of climate change on runoff. *J. Hydrol.* **2015**, *529*, 1601–1616. [[CrossRef](#)]
23. Roudier, P.; Ducharne, A.; Feyen, L. Climate change impacts on runoff in West Africa: A review. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 2789–2801. [[CrossRef](#)]
24. Ray, S. A quick review of machine learning algorithms. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 35–39.
25. Mahesh, B. Machine learning algorithms-a review. *Int. J. Sci. Res.* **2020**, *9*, 381–386. [[CrossRef](#)]
26. Rahimzad, M.; Moghaddam Nia, A.; Zolfonoon, H.; Soltani, J.; Danandeh Mehr, A.; Kwon, H.-H. Performance Comparison of an LSTM-based Deep Learning Model versus Conventional Machine Learning Algorithms for Streamflow Forecasting. *Water Resour. Manag.* **2021**, *35*, 4167–4187. [[CrossRef](#)]
27. Latif, S.D.; Ahmed, A.N. Application of deep learning method for daily streamflow time-series prediction: A case study of the Kowmung River at Cedar Ford, Australia. *Int. J. Sustain. Dev. Plan.* **2021**, *16*, 497–501. [[CrossRef](#)]
28. Adnan, R.M.; Petroselli, A.; Heddam, S.; Santos, C.A.G.; Kisi, O. Short term rainfall-runoff modelling using several machine learning methods and a conceptual event-based model. *Stoch. Environ. Res. Risk Assess.* **2020**, *35*, 597–616. [[CrossRef](#)]
29. Kratzert, F.; Klotz, D.; Brenner, C.; Schulz, K.; Herrnegger, M. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 6005–6022. [[CrossRef](#)]
30. Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A.K.; Hochreiter, S.; Nearing, G.S. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. *Water Resour. Res.* **2019**, *55*, 11344–11354. [[CrossRef](#)]
31. Kratzert, F.; Herrnegger, M.; Klotz, D.; Hochreiter, S.; Klambauer, G. NeuralHydrology–interpreting LSTMs in hydrology. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer: Cham, Switzerland, 2019; pp. 347–362.
32. Jiang, S.; Sweet, L.b.; Blougouras, G.; Brenning, A.; Li, W.; Reichstein, M.; Denzler, J.; Shangquan, W.; Yu, G.; Huang, F. How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences. *Earth's Future* **2024**, *12*, e2024EF004540. [[CrossRef](#)]
33. Liu, J.; Koch, J.; Stisen, S.; Troldborg, L.; Schneider, R.J. A national-scale hybrid model for enhanced streamflow estimation–consolidating a physically based hydrological model with long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 2871–2893. [[CrossRef](#)]
34. Jiang, S.; Tarasova, L.; Yu, G.; Zscheischler, J. Compounding effects in flood drivers challenge estimates of extreme river floods. *Sci. Adv.* **2024**, *10*, ead14005. [[CrossRef](#)] [[PubMed](#)]
35. Sezen, C.; Šraj, M. Improving the simulations of the hydrological model in the karst catchment by integrating the conceptual model with machine learning models. *Sci. Total Environ.* **2024**, *926*, 171684. [[CrossRef](#)] [[PubMed](#)]

36. Mimeau, L.; Künné, A.; Branger, F.; Kralisch, S.; Devers, A.; Vidal, J.-P. Flow intermittence prediction using a hybrid hydrological modelling approach: Influence of observed intermittence data on the training of a random forest model. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 851–871. [[CrossRef](#)]
37. Clark, S.R.; Lerat, J.; Perraud, J.-M.; Fitch, P. Deep learning for monthly rainfall-runoff modelling: A comparison with classical rainfall-runoff modelling across Australia. *Hydrol. Earth Syst. Sci. Discuss.* **2023**, *2023*, 1–34.
38. Frame, J.M.; Kratzert, F.; Klotz, D.; Gauch, M.; Shalev, G.; Gilon, O.; Qualls, L.M.; Gupta, H.V.; Nearing, G.S. Deep learning rainfall-runoff predictions of extreme events. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 3377–3392. [[CrossRef](#)]
39. Soulsby, C.; Tetzlaff, D.; Rodgers, P.; Dunn, S.; Waldron, S. Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: An initial evaluation. *J. Hydrol.* **2006**, *325*, 197–221. [[CrossRef](#)]
40. Pearl River Water Resources Committee (PRWRC). *The Zhujiang Archive*; Guangdong Science and Technology Press: Guangzhou, China, 1991; Volume 1.
41. Zhang, Q.; Xu, C.Y.; Becker, S.; Zhang, Z.; Chen, Y.; Coulibaly, M. Trends and abrupt changes of precipitation maxima in the Pearl River basin, China. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2009**, *10*, 132–144. [[CrossRef](#)]
42. Zhang, Q.; Gu, X.; Singh, V.P.; Shi, P.; Sun, P. More frequent flooding? Changes in flood frequency in the Pearl River basin, China, since 1951 and over the past 1000 years. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2637–2653. [[CrossRef](#)]
43. GRDC. Watershed Boundaries of GRDC Stations/Global Runoff Data Centre. 2011. Available online: https://grdc.bafg.de/GRDC/EN/02_srvcs/22_gslrs/222_WSB/watershedBoundaries_node.html (accessed on 10 June 2023).
44. Global Runoff Data Centre. The Global Runoff Data Centre, 56068 Koblenz, Germany. 2019. Available online: https://grdc.bafg.de/GRDC/EN/02_srvcs/21_tmrs/210_prtl/prtl_node.html;jsessionid=EB00BD8AE9F95C552C2FEB91E12DF962.live11313 (accessed on 10 June 2023).
45. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [[CrossRef](#)]
46. Wang, Q.J.; Pagano, T.C.; Zhou, S.L.; Hapuarachchi, H.A.P.; Zhang, L.; Robertson, D.E. Monthly versus daily water balance models in simulating monthly runoff. *J. Hydrol.* **2011**, *404*, 166–175. [[CrossRef](#)]
47. Bennett, J.C.; Wang, Q.J.; Li, M.; Robertson, D.E.; Schepen, A. Reliable long-range ensemble streamflow forecasts: Combining calibrated climate forecasts with a conceptual runoff model and a staged error model. *Water Resour. Res.* **2016**, *52*, 8238–8259. [[CrossRef](#)]
48. Bennett, J.C.; Wang, Q.J.; Robertson, D.E.; Schepen, A.; Li, M.; Michael, K. Assessment of an ensemble seasonal streamflow forecasting system for Australia. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6007–6030. [[CrossRef](#)]
49. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. Crop evapotranspiration—Guidelines for computing crop water requirements—FAO Irrigation and drainage paper 56. *Fao Rome* **1998**, *300*, D05109.
50. Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [[CrossRef](#)]
51. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
52. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [[CrossRef](#)]
53. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
54. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
55. Raghavendra, N.S.; Deka, P.C. Support vector machine applications in the field of hydrology: A review. *Appl. Soft Comput.* **2014**, *19*, 372–386. [[CrossRef](#)]
56. Zhang, D.; Lin, J.; Peng, Q.; Wang, D.; Yang, T.; Sorooshian, S.; Liu, X.; Zhuang, J. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* **2018**, *565*, 720–736. [[CrossRef](#)]
57. Hosseini, S.M.; Mahjouri, N. Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. *Appl. Soft Comput.* **2016**, *38*, 329–345. [[CrossRef](#)]
58. Granata, F.; Gargano, R.; De Marinis, G. Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA’s storm water management model. *Water* **2016**, *8*, 69. [[CrossRef](#)]
59. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 4.
60. Williams, C.K.; Rasmussen, C.E. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006; Volume 2.
61. Sun, A.Y.; Wang, D.; Xu, X. Monthly streamflow forecasting using Gaussian Process Regression. *J. Hydrol.* **2014**, *511*, 72–81. [[CrossRef](#)]
62. Yang, J.; Jakeman, A.; Fang, G.; Chen, X. Uncertainty analysis of a semi-distributed hydrologic model based on a Gaussian Process emulator. *Environ. Model. Softw.* **2018**, *101*, 289–300. [[CrossRef](#)]
63. Karbasi, M. Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model. *Water Resour. Manag.* **2018**, *32*, 1035–1052. [[CrossRef](#)]
64. Schulz, E.; Speekenbrink, M.; Krause, A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *J. Math. Psychol.* **2018**, *85*, 1–16. [[CrossRef](#)]
65. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]

66. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [[CrossRef](#)]
67. Xiang, Z.; Yan, J.; Demir, I. A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [[CrossRef](#)]
68. Kang, Y.; Cheng, X.; Chen, P.; Zhang, S.; Yang, Q. Monthly runoff prediction by a multivariate hybrid model based on decomposition-normality and Lasso regression. *Environ. Sci. Pollut. Res.* **2023**, *30*, 27743–27762. [[CrossRef](#)] [[PubMed](#)]
69. Roth, V. The generalized LASSO. *IEEE Trans. Neural Netw.* **2004**, *15*, 16–28. [[CrossRef](#)] [[PubMed](#)]
70. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
71. Mitchell, R.; Frank, E. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci.* **2017**, *3*, e127. [[CrossRef](#)]
72. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
73. Machado, F.; Mine, M.; Kaviski, E.; Fill, H. Monthly rainfall-runoff modelling using artificial neural networks. *Hydrol. Sci. J.* **2011**, *56*, 349–361. [[CrossRef](#)]
74. Damavandi, H.G.; Shah, R.; Stampoulis, D.; Wei, Y.; Boscovic, D.; Sabo, J. Accurate Prediction of Streamflow Using Long Short-Term Memory Network: A Case Study in the Brazos River Basin in Texas. *Int. J. Environ. Sci. Dev.* **2019**, *10*, 294–300. [[CrossRef](#)]
75. Sankaranarayanan, S.; Prabhakar, M.; Satish, S.; Jain, P.; Ramprasad, A.; Krishnan, A. Flood prediction based on weather parameters using deep learning. *J. Water Clim. Chang.* **2020**, *11*, 1766–1783. [[CrossRef](#)]
76. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
77. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [[CrossRef](#)]
78. Karpatne, A.; Atluri, G.; Faghmous, J.H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29*, 2318–2331. [[CrossRef](#)]
79. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
80. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
81. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3320–3328.
82. Sit, M.; Demiray, B.Z.; Xiang, Z.; Ewing, G.J.; Sermet, Y.; Demir, I. A comprehensive review of deep learning applications in hydrology and water resources. *Water Sci. Technol.* **2020**, *82*, 2635–2670. [[CrossRef](#)] [[PubMed](#)]
83. Dingman, S.L. *Physical Hydrology*; Waveland Press: Long Grove, IL, USA, 2015.
84. Kratzert, F.; Klotz, D.; Shalev, G.; Klambauer, G.; Hochreiter, S.; Nearing, G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 5089–5110. [[CrossRef](#)]
85. Lees, T.; Buechel, M.; Anderson, B.; Slater, L.; Reece, S.; Coxon, G.; Dadson, S.J. Benchmarking data-driven rainfall-runoff models in Great Britain: A comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 5517–5534. [[CrossRef](#)]
86. Gauch, M.; Kratzert, F.; Klotz, D.; Nearing, G.; Lin, J.; Hochreiter, S. Rainfall-runoff prediction at multiple timescales with a single Long Short-Term Memory network. *Hydrol. Earth Syst. Sci.* **2021**, *25*, 2045–2062. [[CrossRef](#)]
87. Penna, D.; Tromp-van Meerveld, H.; Gobbi, A.; Borga, M.; Dalla Fontana, G. The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 689–702. [[CrossRef](#)]
88. Farchi, A.; Laloyaux, P.; Bonavita, M.; Bocquet, M. Using machine learning to correct model error in data assimilation and forecast applications. *Q. J. R. Meteorol. Soc.* **2021**, *147*, 3067–3084. [[CrossRef](#)]
89. Buizza, C.; Casas, C.Q.; Nadler, P.; Mack, J.; Marrone, S.; Titus, Z.; Le Cornec, C.; Heylen, E.; Dur, T.; Ruiz, L.B. Data learning: Integrating data assimilation and machine learning. *J. Comput. Sci.* **2022**, *58*, 101525. [[CrossRef](#)]
90. Lees, T.; Reece, S.; Kratzert, F.; Klotz, D.; Gauch, M.; De Bruijn, J.; Kumar Sahu, R.; Greve, P.; Slater, L.; Dadson, S.J. Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 3079–3101. [[CrossRef](#)]
91. Frame, J.M.; Kratzert, F.; Gupta, H.V.; Ullrich, P.; Nearing, G.S. On strictly enforced mass conservation constraints for modelling the Rainfall-Runoff process. *Hydrol. Process.* **2023**, *37*, e14847. [[CrossRef](#)]
92. Beven, K.J. Uniqueness of place and process representations in hydrological modelling. *Hydrol. Earth Syst. Sci.* **2000**, *4*, 203–213. [[CrossRef](#)]
93. Herath, H.M.V.V.; Chadalawada, J.; Babovic, V. Genetic programming for hydrological applications: To model or to forecast that is the question. *J. Hydroinform.* **2021**, *23*, 740–763. [[CrossRef](#)]
94. Jiang, S.; Zheng, Y.; Solomatine, D. Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophys. Res. Lett.* **2020**, *47*, e2020GL088229. [[CrossRef](#)]
95. Jiang, S.; Bevacqua, E.; Zscheischler, J. River flooding mechanisms and their changes in Europe revealed by explainable machine learning. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 6339–6359. [[CrossRef](#)]

96. Cai, H.; Liu, S.; Shi, H.; Zhou, Z.; Jiang, S.; Babovic, V. Toward improved lumped groundwater level predictions at catchment scale: Mutual integration of water balance mechanism and deep learning method. *J. Hydrol.* **2022**, *613*, 128495. [[CrossRef](#)]
97. Cai, H.; Shi, H.; Liu, S.; Babovic, V. Impacts of regional characteristics on improving the accuracy of groundwater level prediction using machine learning: The case of central eastern continental United States. *J. Hydrol. Reg. Stud.* **2021**, *37*, 100930. [[CrossRef](#)]
98. Zhang, J.; Chen, H.; Fu, Z.; Luo, Z.; Wang, F.; Wang, K. Effect of soil thickness on rainfall infiltration and runoff generation from karst hillslopes during rainstorms. *Eur. J. Soil Sci.* **2022**, *73*, e13288. [[CrossRef](#)]
99. Berghuijs, W.R.; Woods, R.A.; Hutton, C.J.; Sivapalan, M. Dominant flood generating mechanisms across the United States. *Geophys. Res. Lett.* **2016**, *43*, 4382–4390. [[CrossRef](#)]
100. Xie, K.; Liu, P.; Zhang, J.; Han, D.; Wang, G.; Shen, C. Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *J. Hydrol.* **2021**, *603*, 127043. [[CrossRef](#)]
101. Weng, P.; Tian, Y.; Liu, Y.; Zheng, Y. Time-series generative adversarial networks for flood forecasting. *J. Hydrol.* **2023**, *622*, 129702. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.