

Article

Research on Runoff Simulations Using Deep-Learning Methods

Yan Liu ¹, Ting Zhang ^{1,*} , Aiqing Kang ², Jianzhu Li ¹ and Xiaohui Lei ²

¹ State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China; liuy_an@outlook.com (Y.L.); lijianzhu@tju.edu.cn (J.L.)

² State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; kaqing@iwhr.com (A.K.); lxh@iwhr.com (X.L.)

* Correspondence: zhangting_hydro@tju.edu.cn

Abstract: Runoff simulations are of great significance to the planning management of water resources. Here, we discussed the influence of the model component, model parameters and model input on runoff modeling, taking Hanjiang River Basin as the research area. Convolution kernel and attention mechanism were introduced into an LSTM network, and a new data-driven model Conv-TALSTM was developed. The model parameters were analyzed based on the Conv-TALSTM, and the results suggested that the optimal parameters were greatly affected by the correlation between the input data and output data. We compared the performance of Conv-TALSTM and variant models (TALSTM, Conv-LSTM, LSTM), and found that Conv-TALSTM can reproduce high flow more accurately. Moreover, the results were comparable when the model was trained with meteorological or hydrological variables, whereas the peak values with hydrological data were closer to the observations. When the two datasets were combined, the performance of the model was better. Additionally, Conv-TALSTM was also compared with an ANN (artificial neural network) and Wetspa (a distributed model for Water and Energy Transfer between Soil, Plants and Atmosphere), which verified the advantages of Conv-TALSTM in peak simulations. This study provides a direction for improving the accuracy, simplifying model structure and shortening calculation time in runoff simulations.

Keywords: deep learning; ANN; WetSpa; runoff simulation; HanJiang basin



Citation: Liu, Y.; Zhang, T.; Kang, A.; Li, J.; Lei, X. Research on Runoff Simulations Using Deep-Learning Methods. *Sustainability* **2021**, *13*, 1336. <https://doi.org/10.3390/su13031336>

Academic Editor: Andrzej Wałęga

Received: 5 December 2020

Accepted: 22 January 2021

Published: 27 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Runoff simulations are of great significance to the planning management and rational utilization of water resources [1–4]. Simulations based on hydrological models are a hot topic in hydrological research [5,6]. Hydrological models are classified into physical models and data-driven models [7]. Physical models are centered on structure and parameters. Parameters describe the effect of surface and underground conditions on model input, while structure reflects the physical relationship between model input and output. The model simulates the physical process of watershed runoff formation through the coupling of structure and parameters [8]. With the development of computer technology, geographic information and remote-sensing technology, rich basin spatial information (such as topography, soil, and vegetation types) is gradually integrated into the structure of the physical model, and the physical meaning of model parameters is clarified. At the same time, the theoretical research of producing and conflux of runoff gradually matures, and the model structure improves. A physical model has been developed and applied rapidly from lumped type to distributed type [9–15]. Although the distributed hydrological model based on complex physical mechanism can truly reflect the spatial variability of runoff and concentration process, the models have their applicable physical background, which limits the generality in all basins [16]. Due to the high degree of non-linearity, uncertainty and variability of the hydrological process, even if the model is improved, the runoff simulation may not meet expectations. It will also encounter other problems, such as the same effect with different parameters, difficulty in obtaining data and expensive calculations [17,18].

Data-driven models make predictions by mining the relevant information between input and output variables without studying physical processes. Many studies have applied data-driven models to water-quality simulations, runoff forecasting, water level forecasting, wind speed forecasting, etc. [19–23]. Maniquiz et al. [24] used multiple linear regression (MLR) to establish an equation for estimating pollutant load with rainfall as a variable, indicating that total rainfall and average rainfall intensity can be used as predictors of pollutant load. Ouyang et al. [25] preprocessed the precipitation data based on ensemble empirical mode decomposition (EEMD), and then applied support vector regression (SVR) technology to forecast monthly rainfall. The performance of the EEMD-SVR model was satisfactory. Okkan and Serbes [26] suggested that a discrete wavelet transform-feed-forward neural network (DWT-FFNN) model is better than other models in simulating reservoir runoff. Chua and Wong [27] compared the runoff prediction performance of artificial neural network (ANN), kinematic wave (KW) and autoregressive moving average models (ARMA). The prediction results of the ANN model are more in line with the observations. From the above examples, we can see that the data-driven models have achieved good simulation results.

In recent years, artificial intelligence and big data-driven technology have provided new ideas and technical methods to hydrological study. The new generation of artificial neural networks represented by deep learning has begun to explore applications in rainfall forecasts, flood forecasts, etc. [28]. Deep-learning methods have evolved from simple linear networks to classic generative adversarial networks (GANs) [29]. The field has experienced the fast iterations of the deep belief network (DBN) [30], sparse coding, the convolutional neural network (CNN) [31] and the recurrent neural network (RNN) [32]. RNN has the ability to transfer historical information and is suitable for processing time series [33,34]. The long short-term memory network (LSTM) introduces a memory-gated unit to alleviate the disappearance of the gradient, which is more advantageous than the original RNN simulation with long-term dependent hydrological data. Bowes et al. [35] demonstrated that LSTMs perform better than traditional RNNs by predicting the response of groundwater level to flood events. Zhang et al. [36] established a multilayer perceptron (MLP), a wavelet neural network (WNN), a long short-term memory (LSTM) and a gated recurrent unit (GRU) model to simulate the water level of an urban drainage pipeline, which showed that LSTM had good multi-step prediction ability of time series. In runoff prediction, LSTM models have attracted more attention. Kratzert et al. [37] confirmed that the regional model based on LSTM has a higher forecasting ability in different basins, indicating that the model has good general applicability. Yin et al. [38] showed that the LSTM model performed better than the Xinanjiang model in different forecast periods. They also discussed the hyperparameters of the LSTM. The results suggested that the number of hidden layer neurons significantly affects the prediction accuracy and training speed. Yuan et al. [39] optimized the parameters of the LSTM through the antlion optimization algorithm (ALO), and proposed a high-precision monthly runoff forecast method. Jiang et al. [40] analyzed the simulation effect of LSTM driven by daily scale rainfall data of weather stations and monthly scale TRMM data. Xiang et al. [41] used the LSTM-based seq2seq model to predict the 24-h runoff, indicating that the model outperforms LSTM. Similarly, Liu et al. [42] demonstrated that the LSTM coupled with the k-nearest neighbor algorithm (LSTM-KNN) is more superior than pure LSTM in real-time flood forecasting under different climatic conditions.

The above studies provide new ideas for runoff simulation, and improve the simulation accuracy compared with physical models. Because of its advantages of simulating time series, the LSTM network has become the first choice of deep learning in the hydrological field. Also, its combination with other networks has been widely used in text classification, behavior prediction and other fields [43–45]. Sun et al. [44] made forecasts of the soybean yield in-season and at the end of the season based on a CNN-LSTM model. And the results were better than the pure CNN or LSTM models. Kim and Cho [45] trained a CNN-LSTM model to predict housing energy consumption, and achieved an almost perfect prediction

performance. The spatial feature vector was extracted by CNN and then predicted as the input of LSTM in CNN-LSTM. Although the combination of LSTM with CNN has demonstrated good performance in these studies, its application in hydrological field is rarely seen. Besides, these combined models have not highlighted the input of key time points. Therefore, it is worth discussing how the LSTM network and its combination with deep-learning networks perform in the hydrological field. In addition, input data is the key to the data-driven model, and previous studies have shown that more meteorological variables can improve the performance of the model [46]. Therefore, it is necessary to study the model effect of meteorological data and hydrological data as inputs.

In this study, we constructed several deep-learning models of LSTM combined with CNN and attention mechanism, and first applied them on the runoff simulation in Hanjiang River Basin. Separate and combined input methods for meteorological data and hydrological data were adopted. The purpose of this article is to analyze model performance from the perspective of model component, parameters and inputs. The following work was carried out: (1) the convolution kernel and attention mechanism were introduced into LSTM to establish the Conv-TALSTM model, and the comparisons between Conv-TALSTM and its variants (LSTM, TALSTM, Conv-LSTM) were conducted; (2) the influence of different inputs on the optimization of key parameters was analyzed; (3) the performance of the model under different input data was compared; (4) the performance of the deep-learning model (Conv-TALSTM), data-driven model(ANN) and physical model (Wetspa) was compared.

2. Materials and Methods

2.1. Study Area

The Hanjiang River Basin was selected as the research area. The Hanjiang River Basin is the second-largest watershed in Guangdong Province behind the Pearl River Basin, and is located between $115^{\circ}13' \sim 117^{\circ}09' \text{ E}$ and $23^{\circ}17' \sim 26^{\circ}05' \text{ N}$. The drainage area is $30,112 \text{ km}^2$, and the outlet is the Chaoan hydrological station (Figure 1). The Meijiang and Tingjiang rivers are called the Hanjiang after meeting. The Hanjiang River flows into the Hanjiang Delta from north to south, and then flows into the South China Sea through Shantou City. The terrain of the Hanjiang River Basin slopes from northwest and northeast to southeast. The landform is dominated by mountains, accounting for 70% of the total area of the basin. The Hanjiang River Basin is located in the subtropical and Southeast Asian monsoon climate zone. The climate is hot and humid with abundant rainfall. The average annual rainfall is approximately 1600 mm, but the annual distribution is uneven, mainly concentrated in April to September. The runoff during this period accounts for 80% of the annual runoff. The mean annual flow is approximately 24.5 billion m^3 and the recorded maximum peak flow is $13,300 \text{ m}^3/\text{s}$.

2.2. Data Introduction

The data of four meteorological stations and three hydrological stations in the Hanjiang River Basin from 2005 to 2018 were used in this study. They are all on a daily scale. Meteorological data include daily precipitation, maximum temperature, minimum temperature, average temperature, average wind speed, relative humidity and sunshine duration. Hydrological data include daily flow data. The Shanghang station is a hydrological station for both meteorological and hydrological observation. The training period was from 2005 to 2016, and the verification period was from 2017 to 2018. The Z-score standardized algorithm was used to normalize the data input, and inverse normalization was used for output [38].

We analyzed the correlation between the daily rainfall series of the meteorological stations and the daily flow series of the hydrological stations. The result is provided in Figure 2. Shanghang1 and Shanghang2 represent rainfall and discharge of the Shanghang station, respectively. Figure 2 shows a clearer correlation between the same types of variables. For example, there is a strong correlation between the runoff of Chaoan and Shanghang (Shanghang2 in Figure 2), and a weak correlation between the runoff of Chaoan

and the rainfall of Shanghang (Shanghang1 in Figure 2). Correlation analysis can offer basic information for the following analysis.

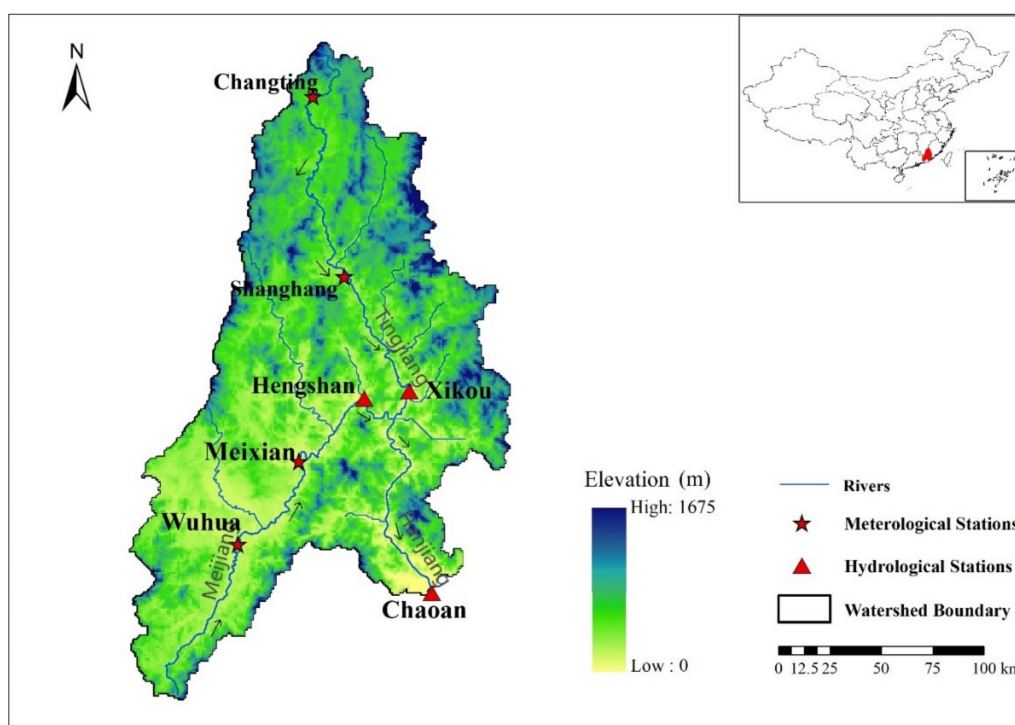


Figure 1. Topography, river networks and observation stations of the Hanjiang River Basin.

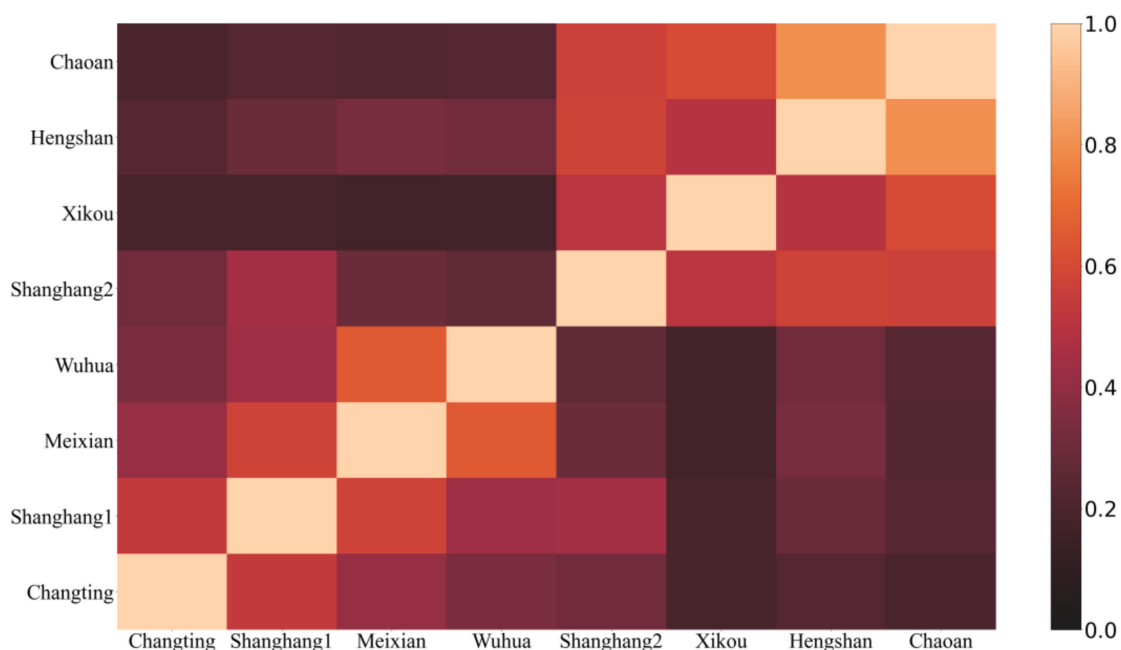


Figure 2. The correlation between the meteorological variables and the hydrological variables. Changting, Shanghang1, Meixian and Wuhua represent the rainfall of the meteorological stations; Shanghang2, Xikou, Hengshan and Chaoan represent the flow of the hydrological stations.

2.3. Models

2.3.1. Deep Learning

1. Convolution Neural Network

A convolution neural network (CNN) is a kind of multilayer feed-forward neural network that can express raw data in a more abstract way [47]. Sparse connection and weight-sharing greatly reduce the number of weights and improve the efficiency of the model [48]. The basic structure of a CNN generally includes a convolution layer, a pooling layer and a full-connection layer. Figure 3 shows the convolution operation process. Its output feature C can be expressed as:

$$C = \sigma(X \otimes W + b) \quad (1)$$

where X is the input data; \otimes is the convolution operation; W is the weight vector of the convolution kernel; b is the offset; σ is the activation function; and relu, sigmoid, tanh, etc. are commonly used.

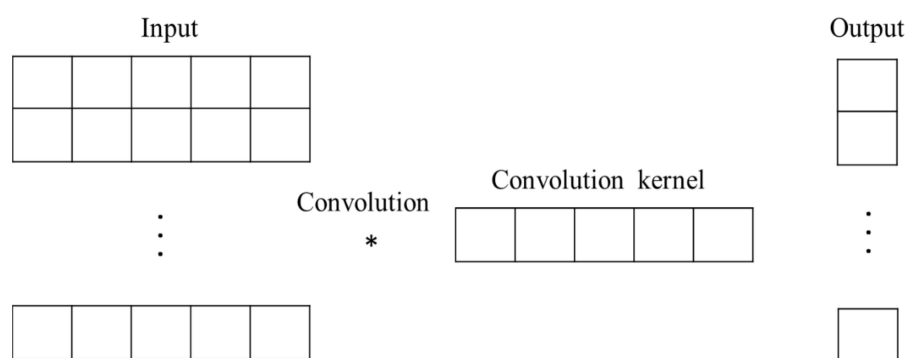


Figure 3. The convolution operation process. * represents convolution operation.

2. Long Short-Term Memory Network

The original RNN connects the historical information with the current task, and can learn the inherent characteristics of time series. However, with the increase of training time and network layers, the disappearance of the gradient prevents it from transmitting information of long-distance data. To overcome this problem, a gate unit is introduced into the LSTM, which is an adapted version of an RNN [49,50]. The LSTM consists of an input gate, a forget gate and an output gate [51]; its internal structure is shown in Figure 4.

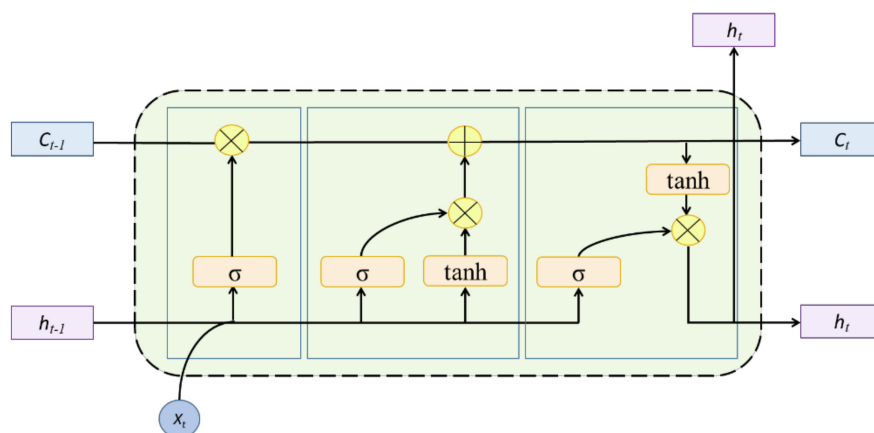


Figure 4. The architecture of the Long Short-Term Memory (LSTM) cell.

The input at time t includes current input x_t and historical information of the hidden layer h_{t-1} and the gate control unit c_{t-1} . First, the forget gate selectively discards cell c_{t-1} information. Next, the input gate determines how much current external information x_t is retained, and generates candidate cell \bar{c}_t . Then, the cell c_t is updated. Finally, the output gate decides which features of the cell c_t to output, and generates the hidden layer variable h_t . The corresponding formula of the above process is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\bar{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$c_t = c_{t-1} \odot f_t + \bar{c}_t \odot i_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

where W_f , W_i , W_c and W_o are the weight vectors of forget gate, input gate, output gate and gate unit, respectively; b_f , b_i , b_c and b_o are the bias vectors of the forget gate, input gate, output gate and gate unit, respectively; σ is sigmoid activation function; and \tanh is hyperbolic tangent activation function.

3. Attention Mechanism

An attention mechanism is an efficient information-processing method inspired by human vision [52]. There are two types, hard attention and soft attention. Hard attention only takes the focus-position information as input and ignores other meaningless information. The existing attention models are mainly based on soft attention, which selectively ignores part of the information to update the weight of the rest of the information. The calculation process can be divided into two steps. One is to calculate a score s_i for each input information x_i , and then to obtain the attention weight α_i of x_i by normalizing s_i using softmax function [53]. The other is to weight the original input and merge it into the intermediate semantic c , a new expression of information. The corresponding formula for the above process is as follows:

$$s_i = \sigma(W^T x_i + b) \quad (8)$$

$$\alpha_i = \text{softmax}(s_i) \quad (9)$$

$$c = \sum_{i=1}^k \alpha_i x_i \quad (10)$$

where W^T and b are trainable parameters; and σ is activation function.

2.3.2. Modeling Process

1. Model Framework

The runoff time series is $Y = (y_1, y_2, \dots, y_T) \in R^T$, and the input data time series is $X = (x_1, x_2, \dots, x_T) = (x^1, x^2, \dots, x^N)^T$. Matrix X includes two dimensions, time dimension and space dimension, which can be expressed as the following formula:

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^N \\ x_2^1 & x_2^2 & \cdots & x_2^N \\ \vdots & \vdots & \vdots & \vdots \\ x_T^1 & x_T^2 & \cdots & x_T^N \end{pmatrix} \in R^{T \times N} \quad (11)$$

where $x_t = (x_t^1, x_t^2, \dots, x_t^N)$ is the set of observation of N variables at time t , and $x^n = (x_1^n, x_2^n, \dots, x_T^n)$ is the sequence of observation of the n th variable during historical time.

In the Hanjiang River Basin, meteorological stations include Changting, Shanghang, Wuhua and Meixian, which are represented by M1, M2, M3 and M4, respectively. Hydrological stations include Shanghang, Xikou, Hengshan and Chaoan, which are represented by H1, H2, H3 and H4, respectively. The Chaoan station is the target station for the simulation. In order to compare the impact of hydrological variables and meteorological variables on the simulation results, we adopted three different input matrixes (A1~A3) to simulate the runoff of the target station. A1 includes the meteorological variables of four meteorological stations and the historical flow of the target station. A2 includes the flow of four hydrological stations, three upstream hydrological stations and the target station (using the historical flow data). A3 includes all data of the meteorological stations and hydrological stations. The input data contains the information of simulation time t and history time $t-i$. The input details of three matrixes are summarized in Table 1. All of the meteorological variables, including daily precipitation, maximum temperature, minimum temperature, average temperature, average wind speed, relative humidity and sunshine duration were used as the input meteorological data. All the input data were normalized first, and then the input matrix was formed. In order to keep the same length for all input variables in the matrix, the flow of H4 at time t was set to 0. Similar methods can be found in previous studies [38,54]. When the simulation was completed, anti-normalization was used for the output sequence.

Table 1. Details of Input Data.

Input Data	No. of Inputs	Detailed Inputs	
		Data Information at Time t	History Information at Time $t-i$ ($i > 1$)
A1	29	M1 _{<i>t</i>} ; M2 _{<i>t</i>} ; M3 _{<i>t</i>} ; M4 _{<i>t</i>}	M1 _{<i>t-i</i>} ; M2 _{<i>t-i</i>} ; M3 _{<i>t-i</i>} ; M4 _{<i>t-i</i>} ; H4 _{<i>t-i</i>}
A2	4	H1 _{<i>t</i>} ; H2 _{<i>t</i>} ; H3 _{<i>t</i>}	H1 _{<i>t-i</i>} ; H2 _{<i>t-i</i>} ; H3 _{<i>t-i</i>} ; H4 _{<i>t-i</i>}
A3	32	M1 _{<i>t</i>} ; M2 _{<i>t</i>} ; M3 _{<i>t</i>} ; M4 _{<i>t</i>} ; H1 _{<i>t</i>} ; H2 _{<i>t</i>} ; H3 _{<i>t</i>}	M1 _{<i>t-i</i>} ; M2 _{<i>t-i</i>} ; M3 _{<i>t-i</i>} ; M4 _{<i>t-i</i>} ; H1 _{<i>t-i</i>} ; H2 _{<i>t-i</i>} ; H3 _{<i>t-i</i>} ; H4 _{<i>t-i</i>}

In this study, a convolution kernel and an attention mechanism were introduced into the LSTM model, and the Conv-TALSTM model was proposed. As shown in Section 2.2, there is a correlation between meteorological variables and hydrological variables. The one-dimensional convolution layer of the CNN was used to express the correlation between input variables in a higher level and more abstract way. The information after processing was transferred into the LSTM network. The attention mechanism was based on the temporal dimension. It can enhance the influence of key time points, and reduce the influence of other time points. This effectively solves the problem of the model being unable to distinguish the difference between the importance of time series. The framework of Conv-TALSTM is shown in Figure 5; it is mainly composed of the input layer, convolution layer, LSTM layer, attention-mechanism layer and full-connection layer.

The details are as follows:

- (1) Convolution layer: The preprocessed data was input into the convolution layer, and a convolution kernel with the size of $1 * k$ was selected to extract more abstract feature structures of different variables in space. The number of convolution kernels was n and the time steps were T . Then, we output the $T * n$ dimensional feature vector, W_{CNN} .
- (2) LSTM layer: W_{CNN} was used as the input for LSTM, and the significant features of time dimension were extracted by LSTM. The number of hidden layer units in LSTM was m . Specifically, x_t in formula (2) is W_{CNN} , and the output of W_{CNN} was W_{LSTM} .
- (3) Attention-mechanism layer: We took W_{LSTM} as the input of the attention-mechanism layer. The influence degree of different time points on the model was expressed as “weight.” The weight was normalized by softmax function [30], and the numerical

value was restricted to 0~1. The weight output was $W_{\text{attention}}$. We performed a weighted summation of $W_{\text{attention}}$ and WLSTM to obtain the final comprehensive timing information. Specifically, x_i is W_{LSTM} and α_i is $W_{\text{attention}}$ in formula (10).

- (4) Full-connection layer. A full connection layer was set up as the output layer.

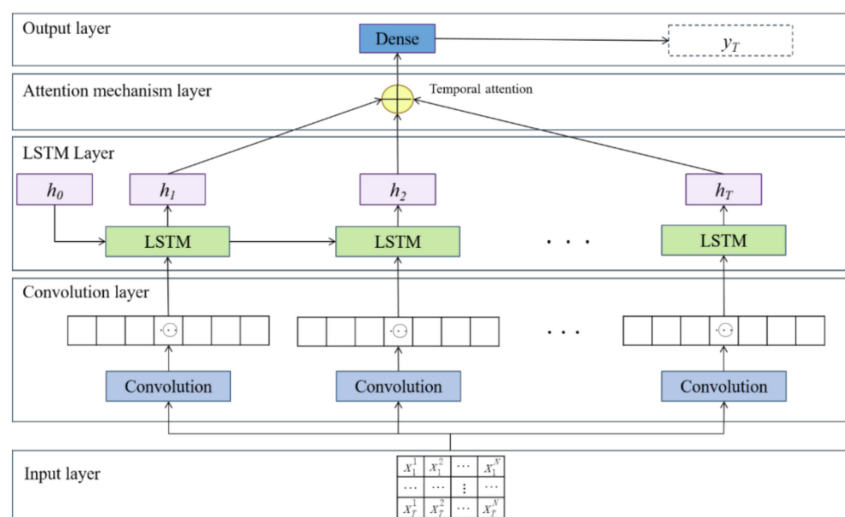


Figure 5. The model framework.

2. Model Experiments

In this study, the initial parameters of the model were set with reference to the existing studies. The window size was 7, and the number of convolution kernels was 32. The kernel_size was 1. One LSTM layer with 32 LSTM neurons was set. The loss function was the mean squared error function. An Adam algorithm was adopted to optimize the loss function with an initial learning rate of 0.0001. Furthermore, we set the maximum number of epochs and the number of batch size to 500 and 64, respectively. A rectified linear unit was used.

The window size, the number of convolution kernels and the number of neurons are the key parameters of each layer in the model framework. The window size represents the length of historical information, and the number of convolution kernels and the number of neurons represent the depth of the model. In this study, we analyzed the influence of these parameters on the simulation effect under three inputs based on the Conv-TALSTM model. The number of convolution kernels and the number of neurons were set to 2, 4, 8, 16, 32, 64, 128, 256, and 512, and the window size was from 2 to 10 days. Other parameters were set as initial parameters.

The different variant models of the Conv-TALSTM model were established, including a pure LSTM model (LSTM), an LSTM model with convolution kernels (Conv-LSTM) and an LSTM model with a time attention mechanism (TALSTM). We compared the Conv-TALSTM with its variants to analyze the influence of the model component on model performance.

Based on the four models above, the influence of input data on the simulation results was considered. The comparison of three different inputs was used to analyze whether meteorological variables or hydrological variables can better simulate runoff, and whether the combined input had a positive impact on the results.

2.3.3. Artificial Neural Network

An artificial neural network (ANN) is inspired by the biological neural system. The core of the artificial neural network is the artificial neuron, which uses large-scale inter-connection and parallel processing to form a complex network [54]. Each neuron receives input from other neurons and then converts it into output according to certain rules. The

most common artificial neural network consists of an input layer, a hidden layer and an output layer [46]. The model structure is shown in Figure 6. The input layer receives the data from the external source, and the output layer outputs the prediction target. They are connected by one or more hidden layers. In this study, a simple feed-forward network with three layers was established to simulate the runoff of the Chaoan station under A3 input. After parameter optimization, a fully connected layer with 16 neurons was set as the hidden layer, and a fully connected layer with one neuron was set as the output layer.

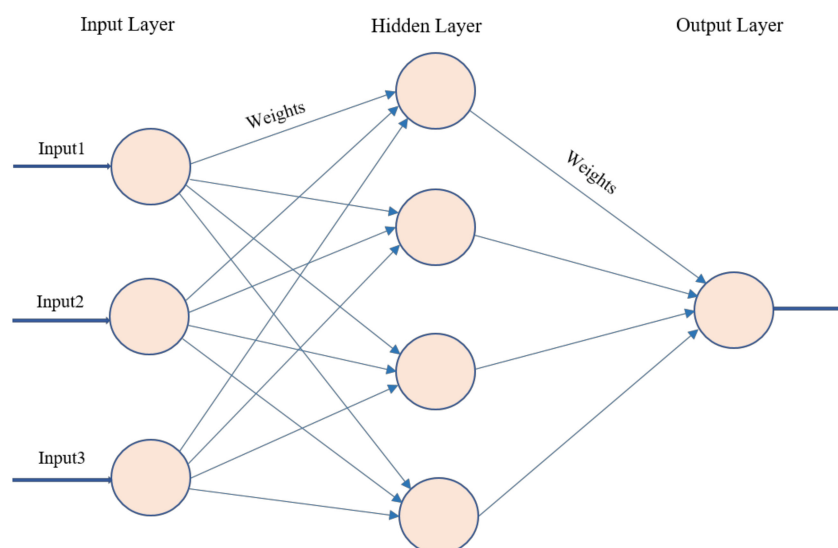


Figure 6. The architecture of the artificial neural network (ANN).

2.3.4. Physical Model

To compare the performance of the deep-learning model and the physical model, the Wetspa (a distributed model for Water and Energy Transfer between Soil, Plants and Atmosphere) model was selected to simulate the runoff of the Hanjiang River Basin. The Wetspa model is a distributed watershed hydrological model that was proposed by Wang et al. of the Free University of Brussels, Belgium in 1996 [55]. Bahremand et al. [56] improved the time step of the Wetspa model from fixed-day to optional-day, hour and minute. The improved Wetspa model discretizes the entire study area into grids, in which the water and energy balance are simulated in layers. Rainfall first meets the interception of forest canopy. Part of the rainwater falling on the ground fills the depression and produces surface runoff, while the other part infiltrates into the soil. According to the different soil water content, the water entering the soil is stored in the root zone in the form of soil water, or flows along the horizontal direction to form interflow, or continues to move downward to form groundwater recharge. Evapotranspiration mainly includes vegetation transpiration, evaporation of rainwater intercepted or filled by plants and evaporation of soil moisture. In addition to the meteorological data mentioned in Section 2.2, the input of the model includes terrain, land use, soil type and other digital data. The input data are transformed into surface runoff, interflow and groundwater flow. The routing of runoff from different cells to the watershed outlet depends on flow velocity and the wave-damping coefficient using the method of diffusive wave approximation. The detailed calculation formulas are shown in [57].

The improved Wetspa model is a distributed physical hydrological model based on GIS technology. The 90-m digital elevation model (DEM) dataset was provided by Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences. Taking ArcView3.2 as the operation platform, the DEM was used to extract the digital features of the watershed, which provided the input of underlying surface data. It mainly included flow direction, cumulative flow, confluence network, slope,

hydraulic radius and boundary division of sub-basins. The model soil data obtained from the Harmonized World Soil Database (HWSD) constructed by FAO and IIASA were classified according to the soil triangle method proposed by the USDA (U.S. Department of Agriculture), and there were four main soil types (clay, clay loam, loam and sandy clay loam.) in the HanJiang basin. The land-use data, provided by Resource and Environment Science and Data Center, Chinese Academy of Sciences, was categorized into eight types (evergreen needleleaf forest, evergreen broadleaf forest, deciduous needleleaf forest, closed shrublands, savannahs, croplands, urban and built-up, and water bodies) according to the IGBP (International Geosphere Biosphere Program) land-use classification standard. The land use and soil types are shown in Figure 7. All spatial distribution parameters in the Wetspa model were derived from terrain, land use and soil type data [57]. Therefore, the model can not only simulate the dynamic change of runoff process at each point, but also can be used to analyze the impact of changing environment on hydrological processes. Global parameters need to be set to run the model [18]. We used daily-scale simulation results of runoff to compare with the Conv-TALSTM model. The scaling factor for interflow computation (C_i), groundwater recession coefficient (C_g), correction factor for potential evapotranspiration (K_{ep}), surface runoff exponent for a near-zero rainfall intensity (K_{run}) and rainfall intensity corresponding to a surface runoff exponent of 1 (P_{max}) were selected to calibrate using the SCE-UA (shuffled complex evolution) algorithm, which is widely used in parameter optimization of distributed hydrological models [58,59].

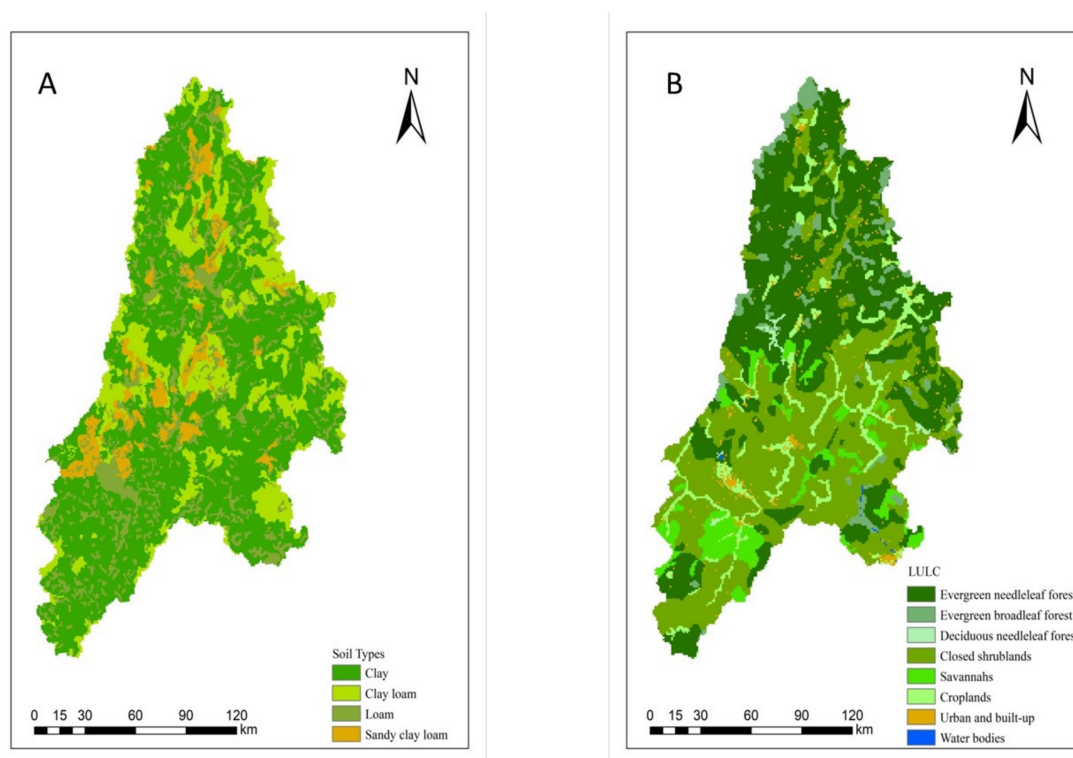


Figure 7. Characteristics of the Hanjiang River Basin: (A) soil types; (B) land use.

2.3.5. Model Evaluation Criteria

In this study, three indices were adopted to quantitatively evaluate the performance of the model: the root mean square error (RMSE), the R-squared score (R^2) and the Nash–Sutcliffe efficiency (NSE). The specific formulas are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_t - y'_t)^2}{n}} \quad (12)$$

$$R^2 = \frac{\left(\sum_{i=1}^n (y_t - \bar{y}_t) (y'_t - \bar{y}'_t) \right)^2}{\sum_{i=1}^n (y_t - \bar{y}_t)^2 \sum_{i=1}^n (y'_t - \bar{y}'_t)^2} \quad (13)$$

$$R^2 = \frac{\left(\sum_{i=1}^n (y_t - \bar{y}_t) (y'_t - \bar{y}'_t) \right)^2}{\sum_{i=1}^n (y_t - \bar{y}_t)^2 \sum_{i=1}^n (y'_t - \bar{y}'_t)^2} \quad (14)$$

where y_t and y'_t represent observed and simulated runoff at time t , \bar{y}_t and \bar{y}'_t represent the average of observed and simulated runoff at time t , and n is the total number of samples.

The RMSE is used to measure the deviation between a simulation and an observation. The range of the RMSE is from 0 to $+\infty$. The closer to 0, the better the overall simulation effect is. R^2 is the square of sample correlation coefficient between 0 and 1 to evaluate the size of model variance. The NSE is often used to evaluate the simulation results in hydrology fields. The variation range of the NSE is from $-\infty$ to 1. A value approximating to 1 means that the simulation process is perfect and the credibility of the model is high.

3. Results

3.1. Optimization of Parameters under Different Inputs

In this study, we analyzed the influence of input data on the optimization of Conv-TALSTM model parameters, including the window size, the number of convolution kernels and the number of hidden layer neurons. Taking NSE, R^2 and RMSE as evaluation indices, the corresponding evaluation results are shown in Figure 8. The common feature was that when there were many input variables, the increase or decrease in the evaluation indexes was relatively gentle.

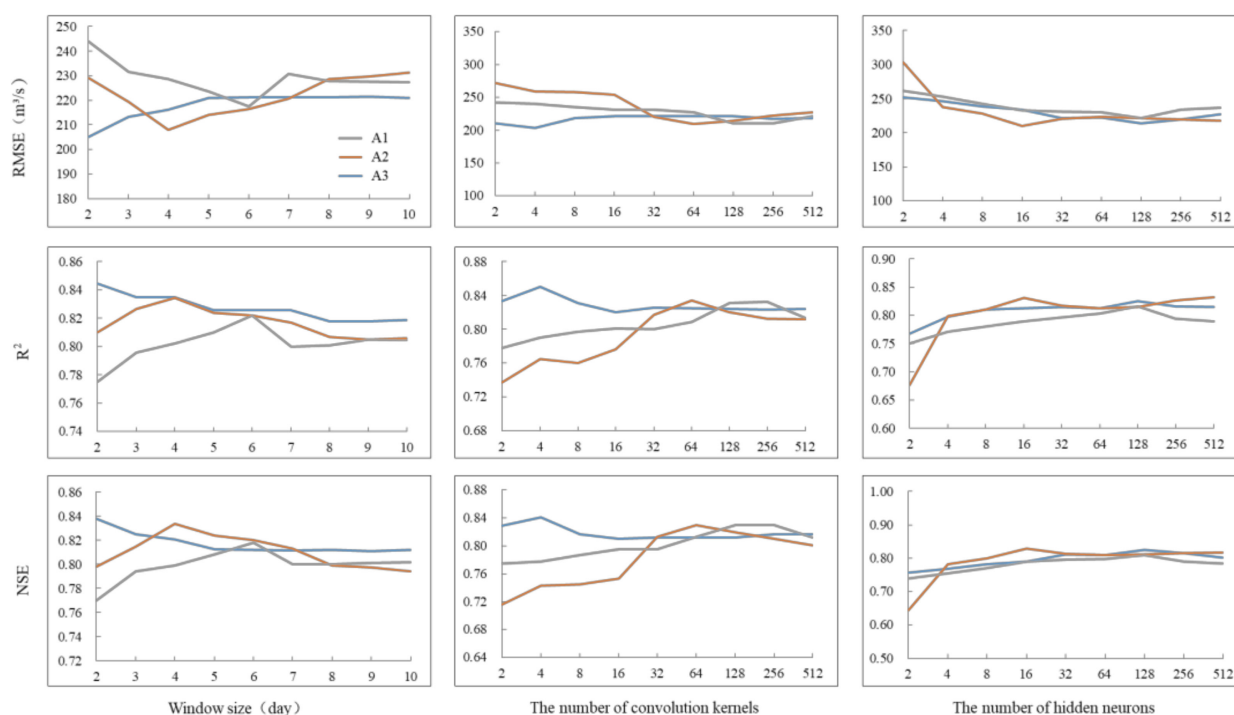


Figure 8. Influence of the Conv-TALSTM model input on parameter optimization.

When the window size changed from 2 to 6 days, the RMSE under the A1 input decreased quickly in an approximately linear trend as the window size increased. When the window size was longer than 6 days, the RMSE became larger than that of 6 days. The NSE and R^2 reached the highest point when the window size was 6 days. Under the A2

input, the RMSE was the lowest when the window size was 4 days. When the window size was longer than 4 days, the effect of each evaluation index became worse. The window size was shorter (2 days) under the A3 input when the evaluation indices achieved the optimal value.

According to the analysis above, the correlation of the A3 input and the target series was stronger than that of the A1 input and target series. The optimal window size decreased as the correlation between input data and target value increased. There was a similar trend in optimizing the number of convolution kernels. When the correlation between the input and output data was weaker, the optimal number of convolution kernels was smaller. The corresponding numbers were 256, 64 and 4 from the A1 input to the A3 input. When the number of convolution kernels was less than the optimal value, the performance of the model became better with the increase of convolution kernels. When the number of convolution kernels was larger than 16, the effect of evaluation indices remained unchanged to a certain extent under A3 input.

With the increase of the neurons in a certain range, the accuracy of the model was significantly improved. When there were more input variables, more neurons were needed to attain good results. The best performance was achieved when the number of neurons was 128, 16 and 128 from the A1 input to the A3 input. When the number of neurons was larger than the optimal value, there was no obvious change in all evaluation indices. It should be noted that the optimization of a hyperparameter was based on the initial value of other parameters.

3.2. Comparison of Different Model Components

To evaluate the effectiveness of the convolution network and the attention mechanism in improving the performance of the model, we compared Conv-TALSTM with its variant models. Table 2 shows the statistical indices of comparison. The performance difference of the four models in the training period was less than that in the verification period. The Conv-TALSTM produced satisfactory results, and the RMSE of flow simulation was less than 210 m³/s in the verification period. Taking A3 input as an example, the influence of the model components on the simulation effect was analyzed.

Table 2. Statistics of the runoff simulations.

Model	Input Data	Time Length	Calibration Period			Validation Period		
			RMSE	R ²	NSE	RMSE	R ²	NSE
WetSpa	A3	1	542.22	0.73	0.68	341.79	0.70	0.65
ANN	A3	6	324.02	0.85	0.85	237.69	0.81	0.78
LSTM	A1	6	293.87	0.87	0.87	232.91	0.81	0.79
	A2	4	290.03	0.88	0.87	231.33	0.81	0.79
	A3	2	289.79	0.88	0.88	217.14	0.82	0.81
Conv-LSTM	A1	6	292.41	0.88	0.87	228.56	0.81	0.80
	A2	4	288.30	0.88	0.88	219.16	0.83	0.81
	A3	4	284.04	0.89	0.88	215.64	0.84	0.82
TALSTM	A1	6	286.29	0.88	0.88	220.55	0.81	0.81
	A2	3	281.79	0.88	0.88	216.72	0.82	0.82
	A3	3	279.06	0.89	0.88	205.70	0.84	0.83
Conv-TALSTM	A1	7	283.71	0.88	0.88	210.00	0.83	0.83
	A2	4	280.74	0.89	0.88	208.06	0.84	0.83
	A3	2	277.70	0.90	0.89	205.11	0.85	0.84

LSTM and Conv-LSTM were regarded as the baseline models. TALSTM had a higher R² (0.84) and NSE (0.83) than the LSTM in the validation period. The R² and NSE of Conv-TALSTM were 0.1 and 0.2 higher than that of Conv-LSTM, respectively. The RMSE of Conv-TALSTM was 0.1 lower than that of Conv-LSTM. Similarly, LSTM and TALSTM

were used as baseline models. Conv-LSTM performed better than LSTM. For example, the R^2 of Conv-LSTM was 0.84, while the R^2 of LSTM was 0.82. Compared with the TALSTM, the three indicators of Conv-TALSTM are superior. Similar trends were witnessed under the A1 and A2 inputs.

The comparison of daily runoff simulated and observed in the verification period is displayed in Figure 9. It can be seen that four models can basically reproduce the runoff process under any input. The daily runoff process simulated by Conv-TALSTM was very close to the real runoff process. There was no significant difference between TALSTM and Conv-LSTM. However, LSTM tends to underestimate high flow, except for individual peak values.

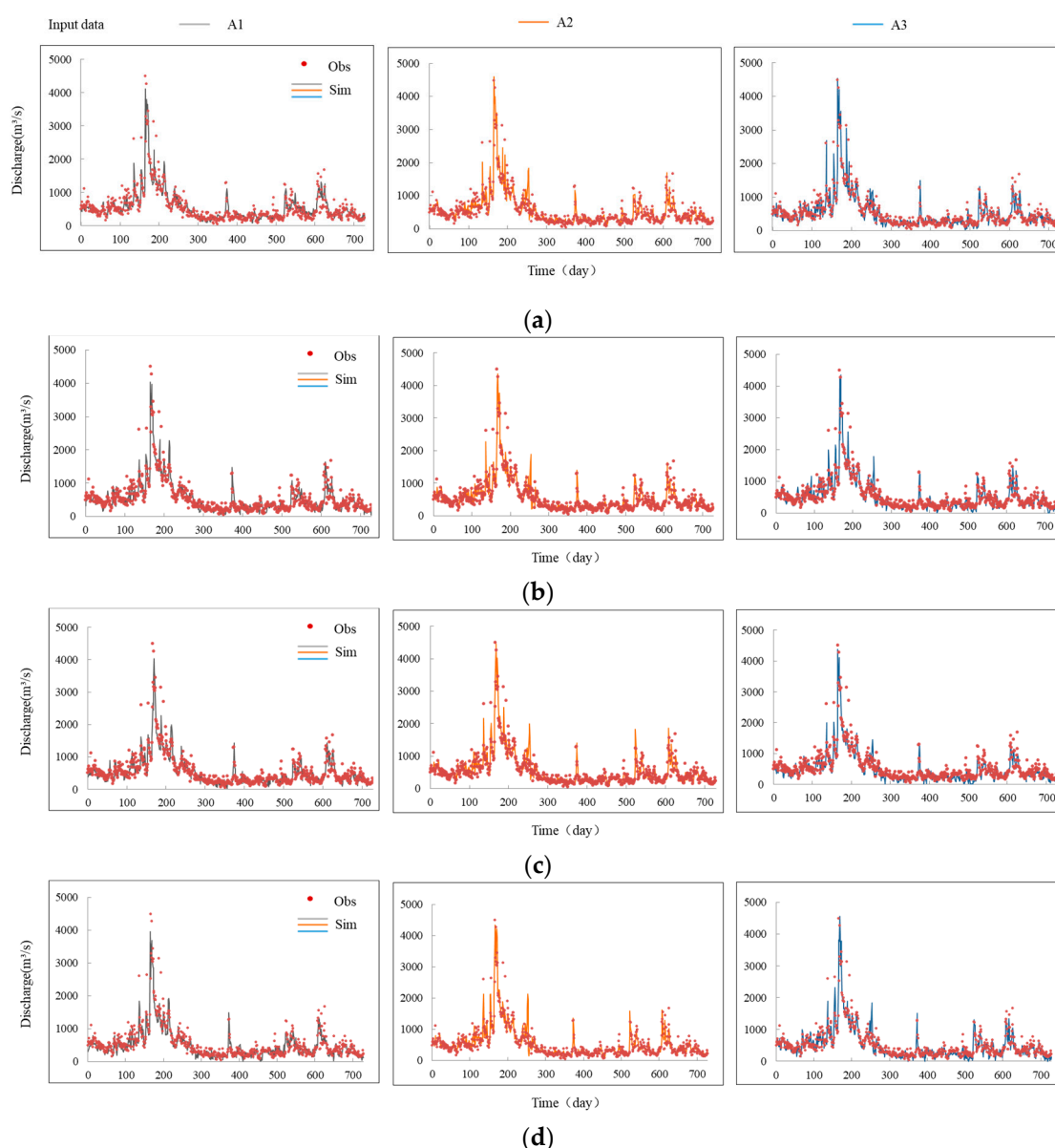


Figure 9. Performance of the four models during the validation period under different inputs. (a) Conv-TALSTM; (b) TALSTM; (c) Conv-LSTM; (d) LSTM.

To further analyze the performance of the four models, Figure 10 shows boxplots of them during the verification period. The median values of the four models are almost at the same level as the actual values under A3 input. And the simulated runoff from the first 25% to 75% is similar to the observed runoff. All the outliers are located on the larger side,

which indicates that the simulated distributions are right-skewed. However, the models differ greatly on the identification of outliers. All of them can capture the largest outliers, though they are slightly higher or lower. Conv-TALSTM is the closest to the observation. Under the A1 and A2 inputs, the results are basically similar. Therefore, Conv-TALSTM had the best performance, which is consistent with the previous conclusion.

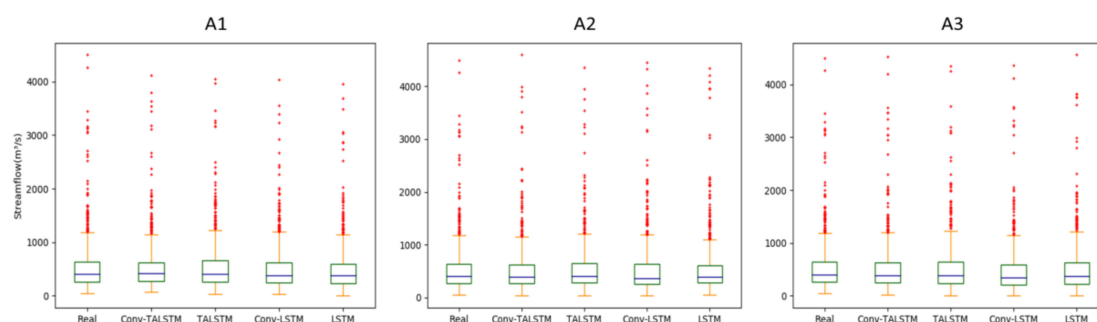


Figure 10. Bar graphs of RMSE, R^2 and NSE for three different inputs (A1–A3) using four models.

3.3. Comparison of Different Inputs

It can be seen in Table 2 that the average values of R^2 were 0.88 and 0.82 for the training period and verification period under the A1 input. The NSE of models except LSTM during the validation period was greater than 0.80. In addition, the RMSE ranged from 210 m^3/s to 232.91 m^3/s . Runoff is formed by precipitation, and other meteorological factors also play an important role in its formation process. Therefore, the simulation results under the A1 input are reliable. Figure 11 shows the comparison of evaluation indicators of all models under three different inputs.

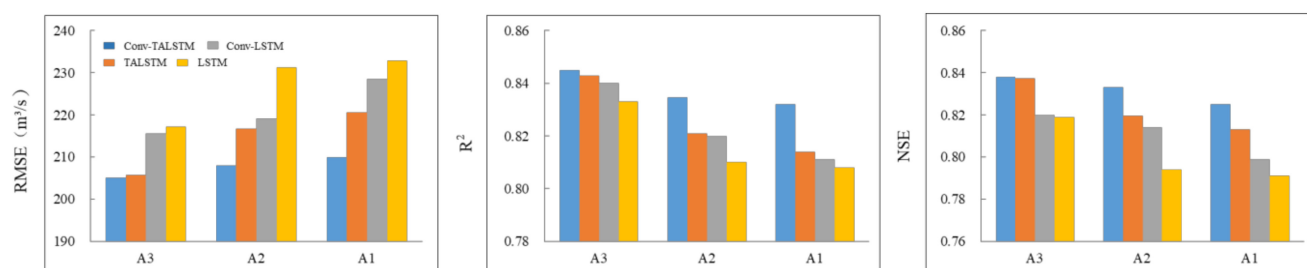


Figure 11. Bar graphs of RMSE, R^2 and NSE for three different inputs (A1–A3) using four models.

Compared with that under the A1 input, the performance of each evaluation index under the A2 input was slightly improved, but it was still at a similar level. The evaluation indices of Conv-LSTM were significantly different under these two inputs, and the RMSE during the verification period was reduced by about 10 m^3/s . Conv-LSTM convolutes the input data, which can be understood as giving each variable a certain weight according to the correlation between the variable and the target value. The correlation between upstream and downstream flow is greater than that flow and meteorological variables, so Conv-LSTM is greatly influenced by the input data.

As can be seen in Figure 9, all models could capture the time pattern of runoff during the validation period under the A1 and A2 inputs. However, the accuracy of the peak value under A2 input was better than that of A1, where the peak value was underestimated. Whether in the training period or validation period, the performance of each model under the A3 input was significantly better than that under the A1 and A2 inputs. R^2 exceeded 0.80 under any input, but R^2 under the A3 input was higher than others, and the highest values in the training and verification periods were 0.90 and 0.85, respectively. The NSE

was also the highest under the A3 input, while the RMSE was much smaller than that under A1 and A2, and the maximum difference was $15.77 \text{ m}^3/\text{s}$ and $14.19 \text{ m}^3/\text{s}$, respectively. It can be seen in Figure 8 that each model could simulate the peak value and runoff process more accurately under the A3 input.

3.4. Comparison of Simulation Capability with Other Models

The ANN and Wetspa model were used to simulate the runoff of the Chaoan station. The statistical results of different models are shown in Table 2. As seen in Table 2, the ANN and Wetspa model performed well. But under the same input, the evaluation indices of Conv-TALSTM were greatly improved compared with the ANN and Wetspa. In order to analyze the simulation potential in detail, especially the accuracy of the flood peak simulation, we selected representative years for comparison. During the training period and the verification period, 2007 (the worst rainstorm flood in Hanjiang River since 1997) and 2017 (the higher peak value during the verification period) were selected, respectively.

The errors of runoff simulation with the three models in representative years are shown in Figure 12. If the error is positive, it means the simulation is high. It can be seen that the deviation of the ANN and Wetspa was often negative at high flow, which illustrates that they tended to underestimate high flow. The error of Conv-TALSTM fluctuated between positive and negative, and the values were small. The simulation is in good agreement with the observation. High errors often occur with the peak flows. But in most cases, the error of the ANN and WetSpa was larger than that of the Conv-TALSTM.

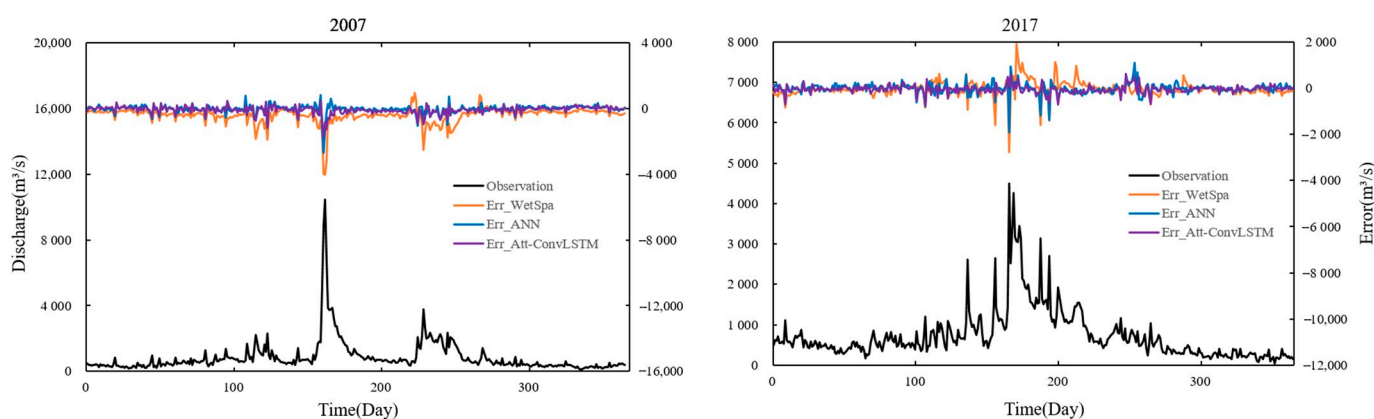


Figure 12. Comparison of the error using the Conv-TALSTM model, ANN model and Wetspa model in representative years.

The results for Conv-TALSTM, ANN and Wetspa in non-flood periods were close, which indicates that the three models had a similar simulation ability, and all of them could reflect runoff processes well. During the flood period, the simulation errors of the ANN and Wetspa were relatively large, while Conv-TALSTM could reproduce the large flood process precisely. In 2007, the discharge of the catastrophic flood process decreased rapidly after reaching the peak value, and the flow difference before and after the flood peak was about $6000 \text{ m}^3/\text{s}$. The error of Wetspa changes from negative value to positive value before and after the peak value, indicating that the simulated flood had a large flow in the process of water-lowering. The ANN also showed a similar trend, but the error at the peak value was smaller than that of Wetspa. In addition, the ANN and Wetspa had a large deviation for the multi-peak flow process in 2017. It was found that the simulation at the largest flood peak was relatively low, while the latter two peaks were relatively high. This can be considered to be caused by the peak time-lag. The simulation process of the ANN was closer to the measured process compared with Wetspa, but it could not accurately reproduce the magnitude of flood peak either. Nevertheless, Conv-TALSTM performed better than both of them.

4. Discussion

Many studies have shown that deep-learning models, especially LSTM, have great potential in hydrological simulations. Chollet and Allaire [60] pointed out that it is necessary to choose the right model structure through practice. In this study, we combined a convolution kernel and an attention mechanism with an LSTM to compare the effect of the model components on the simulation results. The convolution kernel was used to extract the features of each dimension at the same time point, and the temporal attention mechanism learned the influence of different time points. Compared with a single LSTM, the combined model took both data abstraction and time importance into account. It has been shown that we can improve the simulation accuracy by changing the model component. It was found that the improvement effect of the model with an attention mechanism is better than that of a one-dimensional CNN, which may be due to the CNN's limitations in learning spatial-position information. The input form of the two-dimensional grid will be explored.

A deep-learning model consists of several layers [60], and is much simpler than physical model. In addition, the evaluation indices of all the deep-learning models are better than that of the physical model. The deep-learning method for runoff simulations has the advantages of simple feasibility and high accuracy. The trained model can be used for real-time prediction, which can provide important information for flood control. Despite such good results, such models have no physical base. Kratzert et al. [37] suggested that the basin dynamics can be reflected by LSTM internal variables, but its rationality needs to be proved. On one hand, we can use the data-driven model to modify the results of the physical model in real-time. On the other hand, we can improve the data-driven model by adding the input data.

The method of selecting the model parameters is also a problem to be considered when the model framework is determined. The number of hyperparameters of deep-learning models is less than that of physical models, but the value range of each hyperparameter is huge. Artificial intelligence is often required to determine some parameter values. Most researchers analyzed the influence of other parameters on the premise of determining the initial values of some parameters and certain input data. In our study, we discussed the effect of window size, the number of convolution kernels and the number of neurons on the simulation results under different input conditions. The correlation between the target station and other stations was considered, which provides useful information for parameter optimization. Due to the shorter time required for single operation, it takes less time to optimize hyperparameters than to calibrate a physical model. However, we cannot exhaust all the parameter values, and the final parameter combination might not be the optimal combination of the whole hyperparametric space. This also limits the further improvement of the model performance. Therefore, it is necessary to develop an effective algorithm for automatic optimization of parameters.

Data characteristics also determine the performance of the model. We analyzed the influence of meteorological variables and hydrological variables on the results in detail. Runoff in the Hanjiang River Basin is formed by rainfall. The process of rainfall-runoff is affected by other meteorological variables. In this study, the results of meteorological variables and hydrological variables were comparable, indicating that they contain similar information. The peak value with hydrological variables is more accurate, which results from the stronger correlation between the same type of stations. The amount of data is also an important factor affecting the data-driven model other than data relevance. The A3 input, including the source (rainfall) and intermediate process (other meteorological variables and upstream flow) of outlet runoff formation, had more abundant data information than A1 and A2. There is no doubt that the simulation effect was the best under the A3 input. Compared with other research, the performance of the deep-learning model in this study was not the best, although its accuracy was higher than that of the hydrological model. The main reason was the input condition with sparse data and short time series. Tian et al. [61] showed that whether a neural network model or a hydrological model

is used for hydrological simulation, the basin with high station density will have more abundant data information and more accurate simulation results. With meteorological variables as input, the highest value of the NSE exceeded 0.9 during the validation period in the study of Fan et al. [46], while it was close to 0.83 in our study. It was found that the station density studied by them was 1248 km²/station, which is about six times that of our study (7528 km²/station). Hu et al. [62] obtained great simulation results when the density of rain-gauging stations was 200 km²/station. The data conditions and research results of Yin et al. [38] were similar to those of Hu et al. [62]. Jiang et al. [40] simulated runoff based on long series data of 50 years, and the model performance was satisfactory. In future research, we will consider replacing monitored data with other meteorological products such as TRMM (Tropical Rainfall Measuring Mission). Converting the site location information and underlying surface conditions into more abundant input data is also a further research direction. In addition, the deep-learning models proposed in this paper are also suitable for the prediction of water quality, groundwater and other factors, which is of great significance to realize the sustainable development of river basins.

5. Conclusions

This study investigated the application of deep learning in runoff simulations. Several deep-learning models were developed to discuss the effects of model component, model parameters and model input on model performance. Additionally, the results for the Conv-TALSTM model were compared with the data-driven model (ANN) and the distributed hydrological model (Wetspa). A convolution kernel and a temporal attention mechanism were introduced to the Conv-TALSTM model, which can extract spatial data correlation and highlight key time-point information. Compared with different variant models (ANN and Wetspa), the Conv-TALSTM model showed a much better performance. Therefore, the simulation accuracy can be improved by changing the model composition. The optimal parameters were strongly influenced by different input data. When the input data had a strong correlation with the target value, the optimal window size and the number of convolution kernels was always small. When the input data had more information, more hidden layer units were needed. Moreover, the overall difference among the simulation results was small with meteorological data or hydrological data as the model input, but the peak value could be captured more accurately with hydrological data. The accuracy of the model was improved when both of them are input. Therefore, enriching input data is another effective method to improve the performance of the model.

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

Funding: This research is supported by the National Key Research and Development Program of China (No. 2018YFC0407902).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from National real time water and rain database, Ministry of Water Resources of the People's Republic of China and are available from the authors with the permission of Ministry of Water Resources of the People's Republic of China.

Acknowledgments: We are grateful to the Data Center for Resources and Environment Sciences, Chinese Academy of Sciences for providing data.

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

References

1. Luo, P.; Sun, Y.; Wang, S.; Wang, S.; Lyu, J.; Zhou, M.; Nakagami, K.; Takara, K.; Nover, D. Historical Assessment and Future Sustainability Challenges of Egyptian Water Resources Management. *J. Clean. Prod.* **2020**, *263*, 121154. [\[CrossRef\]](#)
2. Zhang, Y.; Luo, P.; Zhao, S.; Kang, S.; Wang, P.; Zhou, M.; Lyu, J. Control and Remediation Methods for Eutrophic Lakes in Recent 30 years. *Water Sci. Technol.* **2020**, *81*, 1099–1113. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Zhu, Y.; Luo, P.; Su, F.; Zhang, S.; Sun, B. Spatiotemporal Analysis of Hydrological Variations and Their Impacts on Vegetation in Semiarid Areas from Multiple Satellite Data. *Remote Sens.* **2020**, *12*, 4177. [\[CrossRef\]](#)
4. Luo, P.; Kang, S.; Apip, A.; Zhou, M.; Lyu, J.; Aisyah, S.; Mishra, B.; Regmi, R.K.; Nover, D. Water quality trend assessment in Jakarta: A rapidly growing Asian megacity. *PLoS ONE* **2019**, *14*, e0219009. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Song, X.M.; Kong, F.Z.; Zhan, C.S.; Han, J.W. Hybrid optimization rainfall-runoff simulation based on xinanjiang model and artificial neural network. *J. Hydrol. Eng.* **2012**, *17*, 1033–1041. [\[CrossRef\]](#)
6. Niemi, T.J.; Warsta, L.; Taka, M.; Hickman, B.; Pulkkinen, S.; Krebs, G.; Moiseev, D.N.; Koivusalo, H.; Kokkonen, T. Applicability of open rainfall data to event-scale urban rainfall-runoff modelling. *J. Hydrol.* **2017**, *547*, 143–155. [\[CrossRef\]](#)
7. Kan, G.; Li, J.; Zhang, X.; Ding, L.; He, X.; Liang, K.; Jiang, X.; Ren, M.; Li, H.; Wang, F.; et al. A new hybrid data-driven model for event-based rainfall-runoff simulation. *Neural Comput. Appl.* **2017**, *28*, 2519–2534. [\[CrossRef\]](#)
8. Rui, X. Discussion of watershed hydrological model. *Adv. Sci. Technol. Water Resour.* **2017**, *37*, 1–7. [\[CrossRef\]](#)
9. Wang, Y.; Shao, J.; Su, C.; Cui, Y.; Zhang, Q. The Application of Improved SWAT Model to Hydrological Cycle Study in Karst Area of South China. *Sustainability* **2019**, *11*, 5024. [\[CrossRef\]](#)
10. Muleta, M.K.; Nicklow, J.W. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **2005**, *306*, 127–145. [\[CrossRef\]](#)
11. Meng, X.; Zhang, M.; Wen, J.; Du, S.; Xu, H.; Wang, L.; Yang, Y. A Simple GIS-Based Model for Urban Rainstorm Inundation Simulation. *Sustainability* **2019**, *11*, 2830. [\[CrossRef\]](#)
12. Huo, A.; Peng, J.; Cheng, Y.; Luo, P.; Zhao, Z.; Zheng, C. Hydrological Analysis of Loess Plateau Highland Control Schemes in Dongzhi Plateau. *Front. Earth Sci.* **2020**, *8*, 528632. [\[CrossRef\]](#)
13. Mu, D.; Luo, P.; Lyu, J.; Zhou, M.; Huo, A.; Duan, W.; Nover, D.; He, B.; Zhao, X. Impact of temporal rainfall patterns on flash floods in Hue City, Vietnam. *J. Flood Risk Manag.* **2020**, e12668. [\[CrossRef\]](#)
14. Luo, P.; Mu, D.; Xue, H.; Ngo-Duc, T.; Dang-Dinh, K.; Takara, K.; Nover, D.; Schladow, G. Flood inundation assessment for the Hanoi Central Area, Vietnam under historical and extreme rainfall conditions. *Sci. Rep.* **2018**, *8*, 12623. [\[CrossRef\]](#)
15. Huo, A.; Yang, L.; Luo, P.; Cheng, Y.; Peng, J.; Daniel, N. Influence of Landfill and land use scenario on runoff, evapotranspiration, and sediment yield over the Chinese Loess Plateau. *Ecol. Indic.* **2020**. [\[CrossRef\]](#)
16. Wu, X.; Liu, C. Progress in watershed hydrological models. *Progr. Geogr.* **2002**, *21*, 341–348. [\[CrossRef\]](#)
17. Wood, E.F.; Roundy, J.K.; Troy, T.J. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resour. Res.* **2011**, *47*. [\[CrossRef\]](#)
18. Yin, Z.; Liao, W.; Lei, X.; Wang, H.; Wang, R. Comparing the Hydrological Responses of Conceptual and Process-Based Models with Varying Rain Gauge Density and Distribution. *Sustainability* **2018**, *10*, 3209. [\[CrossRef\]](#)
19. Parkin, G.; Birkinshaw, S.J.; Younger, P.L.; Rao, Z.; Kirk, S. A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows. *J. Hydrol.* **2007**, *339*, 15–28. [\[CrossRef\]](#)
20. Guo, Z.; Chi, D.; Wu, J.; Zhang, W. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the Apriori algorithm. *Energy Convers. Manag.* **2014**, *84*, 140–151. [\[CrossRef\]](#)
21. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **2015**, *520*, 224–243. [\[CrossRef\]](#)
22. Chiamsathit, C.; Adeyoye, A.J.; Bankaru-Swamy, S. Inflow forecasting using artificial neural networks for reservoir operation. *Proc. Int. Ass. Hydrol. Sci.* **2016**, *373*, 209–214. [\[CrossRef\]](#)
23. Shoaib, M.; Shamseldin, A.Y.; Khan, S.; Khan, M.M.; Khan, Z.M.; Sultan, T.; Melville, B.W. A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. *Water Resour. Manag.* **2018**, *32*, 83–103. [\[CrossRef\]](#)
24. Maniquiz, M.C.; Lee, S.; Kim, L. Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables. *J. Environ. Sci.* **2010**, *22*, 946–952. [\[CrossRef\]](#)
25. Ouyang, Q.; Lu, W.; Xin, X.; Zhang, Y.; Cheng, W.; Yu, T. Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. *Water Resour. Manag.* **2016**, *30*, 2311–2325. [\[CrossRef\]](#)
26. Okkan, U.; Serbes, Z.A. The combined use of wavelet transform and black box models in reservoir inflow modeling. *J. Hydrol. Hydromech.* **2013**, *61*, 112–119. [\[CrossRef\]](#)
27. Chua, L.H.C.; Wong, T.S.W. Runoff forecasting for an asphalt plane by artificial neural networks and comparisons with kinematic wave and autoregressive moving average models. *J. Hydrol.* **2011**, *397*, 191–201. [\[CrossRef\]](#)
28. Liu, C. New generation hydrological model based on artificial intelligence and big data and its application in flood forecasting and early warning. *China Flood Drought Manag.* **2019**, *29*, 11–22. Available online: <https://www.ixueshu.com/document/035f7dd25a73b911a0a9d37b30b561f6318947a18e7f9386.html> (accessed on 27 March 2020).
29. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [\[CrossRef\]](#)

30. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]
31. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 278–2324. [CrossRef]
32. Bengio, Y.; Imard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef] [PubMed]
33. Han, M.; Xi, J.; Xu, S.; Yin, F.L. Prediction of chaotic time series based on the recurrent predictor neural network. *IEEE Trans. Signal Process.* **2004**, *52*, 3409–3416. [CrossRef]
34. Yang, L.; Wu, Y.; Wang, J.; Liu, Y. Research on recurrent neural network. *J. Comput. Appl.* **2018**, *38*, 1–6.
35. Bowes, B.D.; Sadler, J.M.; Morsy, M.M.; Behl, M.; Goodall, J.L. Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks. *Water* **2019**, *11*, 1098. [CrossRef]
36. Zhang, D.; Lindholm, G.; Ratnaweera, H. Use long short-term memory to enhance internet of things for combined sewer overflow monitoring. *J. Hydrol.* **2018**, *556*, 409–418. [CrossRef]
37. 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]
38. Yin, Z.; Liao, W.; Wang, R.; Lei, X. Rainfall-runoff modelling and forecasting based on long short-term memory(LSTM). *S. N. Water Transf. Water Sci. Technol.* **2019**, *6*, 1–9. [CrossRef]
39. Yuan, X.; Chen, C.; Lei, X.; Yuan, Y.; Adnan, R.M. Monthly runoff forecasting based on LSTM-ALO model. *Stoch. Environ. Res. Risk A* **2018**, *32*, 2199–2212. [CrossRef]
40. Jiang, S.; Lu, J.; Chen, X.; Liu, Z. The research of stream flow simulation using Long and Short Term Memory (LSTM) network in Fuhe River Basin of Poyang Lake. *J. Cent. China Norm. Univ.* **2020**, *54*, 128–139. [CrossRef]
41. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*. [CrossRef]
42. Liu, M.; Huang, Y.; Li, Z.; Tong, B.; Zhang, H. The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. *Water* **2020**, *12*, 440. [CrossRef]
43. Ding, L.; Fang, W.; Luo, H.; Love, P.E.D.; Zhong, B.; Ouyang, X. A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory. *Automat. Constr.* **2018**, *86*, 118–124. [CrossRef]
44. Sun, J.; Di, L.; Sun, Z.; Shen, Y.; Lai, Z. County-level soybean yield prediction using deep CNN-LSTM model. *Sensors* **2019**, *19*, 4363. [CrossRef]
45. Kim, T.Y.; Cho, S.B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]
46. Fan, H.; Jiang, M.; Xu, L.; Zhu, H.; Jiang, J. Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water* **2020**, *12*, 175. [CrossRef]
47. Kontogiannis, D.; Bargiotas, D.; Daskalopulu, A. Minutely Active Power Forecasting Models Using Neural Networks. *Sustainability* **2020**, *12*, 3177. [CrossRef]
48. Huang, C.J.; Kuo, P.H. A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. *Sensors* **2018**, *18*, 2220. [CrossRef]
49. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
50. Shoaib, M.; Shamseldin, A.Y.; Melville, B.W.; Khan, M.M. A comparison between wavelet based static and dynamic neural network approaches for runoff prediction. *J. Hydrol.* **2016**, *535*, 211–225. [CrossRef]
51. Son, H.; Kim, C. A Deep Learning Approach to Forecasting Monthly Demand for Residential-Sector Electricity. *Sustainability* **2020**, *12*, 3103. [CrossRef]
52. Abedinia, O.; Amjady, N.; Zareipour, H. A new feature selection technique for load and price forecast of electrical power systems. *IEEE T. Power Syst.* **2016**, *32*, 62–74. [CrossRef]
53. Hinton, G.E.; Salakhutdinov, R.R. Replicated softmax: An undirected topic model. In *Advances in Neural Information Processing Systems 22 (NIPS 2009)*; Curran Associates Inc.: Red Hook, NY, USA, 2009; Available online: <https://proceedings.neurips.cc/paper/2009/file/31839b036f63806cba3f47b93af8ccb5-Paper.pdf> (accessed on 1 May 2020).
54. Fuente, A.D.L.; Meruane, V.; Meruane, C. Hydrological Early Warning System Based on a Deep Learning Runoff Model Coupled with a Meteorological Forecast. *Water* **2019**, *11*, 1808. [CrossRef]
55. Wang, Z.M.; Batelaan, O.; De Smedt, F. A distributed model for water and energy transfer between soil, plants and atmosphere (wetspa). *Phys. Chem. Earth* **1996**, *21*, 189–193. [CrossRef]
56. Bahremand, A.; De Smedt, F.; Corluy, J.; Liu, Y.B.; Poorova, J.; Velcicka, L.; Kunikova, E. WetSpa model application for assessing reforestation impacts on floods in mategany-Hornad Watershed, Slovakia. *Water Resour. Manag.* **2007**, *21*, 1373–1391. [CrossRef]
57. Liu, Y.B.; De Smedt, F. WetSpa Extension, A GIS-based Hydrologic Model for Flood Prediction and Watershed Management Documentation and User Manual. *Vrije Univ. Bruss. Belgium* **2004**, *1*, e108. [CrossRef]
58. Ma, H.; Dong, Z.; Zhang, W.; Liang, Z. Application of SCE-UA algorithm to optimization of TOPMODEL parameters. *J. Hohai Univ. Nat. Sci.* **2006**, *4*, 361–365. [CrossRef]
59. Lei, X.; Jiang, Y.; Wang, H.; Tian, Y. Distributed hydrological model EasyDHM II. Application. *J. Hydrol. Eng.* **2010**, *41*, 893–907. [CrossRef]

-
60. Chollet, F.; Allaire, J.J. *Deep Learning with R*; Manning Publications: New York, NY, USA, 2018; pp. 24–50.
 61. Tian, Y.; Xu, Y.P.; Yang, Z.; Wang, G.; Zhu, Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water* **2018**, *10*, 1655. [[CrossRef](#)]
 62. Hu, C.; Wu, Q.; Li, H.; Jian, S.; Li, N.; Lou, Z. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* **2018**, *10*, 1543. [[CrossRef](#)]