Effects of Training Data on the Learning Performance of LSTM Network for Runoff Simulation

Anbang Peng1 · Xiaoli Zhang2 · Wei Xu3 · Yuanyang Tian1,3

Received: 13 January 2022 / Accepted: 10 April 2022 / Published online: 20 April 2022 © The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
With the rapid development of Artificial Intelligence (AI) technology, the Long Short-Term Memory (LSTM) network has been widely used for forecasting hydrological process. To evaluate the effect of training data amount on the performance of LSTM, the study proposed an experiment scheme. First, K-Nearest Neighbour (KNN) algorithm is employed for generating the meteorological data series of 130 years based on the observed data, and the Soil and Water Assessment Tool (SWAT) model is used to obtain the corresponding runoff series with the generated meteorological data series. Then, the 130 years of rainfall and runoff data is divided into two parts: the first 80 years of data for model training and the remaining 50 years of data for model verification. Finally, the LSTM models are developed and evaluated, with the first 5-year, 10-year, 20-year, 40-year and 80-year data series as training data respectively. The results obtained in Yalong River, Minjiang River and Jialing River show that increasing the training data amount can effectively reduce the over-fittings of LSTM network and improve the prediction accuracy and stability of LSTM network.

Keywords LSTM · Rainfall runoff · Data amount · Over-fitting

1 Introduction

The reliability of runoff prediction is of great significance to water resource management, but the evident non-linear nature and randomness of runoff process makes it challenging to predict runoff (Wang et al. 2021a, b; Rahimzad et al. 2021). The models available for runoff prediction fall into three categories: (1) conceptual model, which is simple and easy to use, but gives a less detailed description of runoff generation and flow concentration; (2) physical model, which enjoys a good physical foundation but involves complicated calculation...
structure and large amount of basic data to build model. (3) data-driven model, which just needs to create the relationship between input and output data, without considering the hydrological mechanism of the basin and having a good simulation. Among them, the data-driven model, e.g., the convergence of the conventional back propagation (BP) neural network is too slow and thus has limited learning ability. With the rapid development of deep learning over the years (Hinton and Salakhutdinov 2006), the learning abilities of the Deep Neural Networks (DNN) are advancing by leaps and bounds, demonstrating super learning abilities in learning knowledge of various industries. However, the traditional DNN network cannot be used for time series modelling.

In order to learn the changing rule of the time series, Recurrent Neural Network (RNN) came into being. Based on conventional RNN, Long-Short Term Memory (LSTM) Network was first proposed by Hochreiter and Schmidhuber (1997). Due to the strong abilities in learning the time series data by digging out the long-term dependence relationship in time series, LSTM has been widely used in many fields, such as, image processing (Jiang and Liu 2021; Lin et al. 2021), transportation (Huang et al. 2021a, b), geological disasters relief (Dikshit et al. 2021), reference evapotranspiration financial economy (Dami and Esterabi 2021) and so on. Moreover, its applicability for hydrological prediction has been demonstrated by many researchers. For example, Xu et al. (2020) analyzed the performance of the LSTM network, the Soil and Water Assessment Tool (SWAT) model, Xin’anjiang Model, Multiple Regression Model and Back-propagation Neural Network (BPNN), and concluded that LSTM network has better learning abilities except for the problem of over-fitting. (Rahimzad et al. 2021) compared the accuracy of Linear Regression (LR), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and LSTM network in daily streamflow forecasting, and the results indicated that the LSTM is a robust data-driven technique to characterize the time series behaviors in hydrological modeling applications. All the aforementioned researches indicate that the LSTM network has strong capability in simulating runoff. In order to improve the learning ability of LSTM network, some scholars have conducted in-depth research (Wang et al. 2021a, b; Yin et al. 2021; Dilip 2021). However, there are little studies focusing on the issue of over-fitting and the effect of training data amount on the performance of LSTM model.

At present, the Deep Learning Network training needs to take big data as the foundation (Hu et al. 2018; Mao et al. 2021), however, the hydrological data available often fail to meet the needs. It is necessary to find out the appropriate training data amount for developing a stable model. Therefore, studying the influence of training data amount on the LSTM and finding the future breakthrough point is the key to the application of deep learning in the field of hydrology.

2 Research Methodologies

2.1 Experiment Scheme

In order to study the effect of training data amount on the LSTM network performance, an experiment scheme is proposed in this study. Firstly, the K-Nearest Neighbour (KNN) algorithm (Lall and Sharma 1996) is used to derive rainfall data with long series. Secondly, the rainfall data is used as the input for SWAT model to simulate runoff data (served as experimental real data). Rainfall and runoff data are divided into training data and verification data. Then, the effect of the training data amount on the LSTM network performance
are analyzed by varying the training data amount. Finally, the better parameters choices and data volume are found out, and the learning ability of LSTM network are tested based on this. The flow chart of this experiment scheme is shown in Fig. 1.

2.2 LSTM Network

The key parts of the LSTM network are comprised of fully connected layer and LSTM cells. As shown in Fig. 2a, the LSTM network includes four kinds of layers: (1) input layer, which is to receive input series data; (2) fully connected layer, which acts as a bridge
between input layer and the LSTM cell layer to transmit the dimension of the input data to the dimension of LSTM cell; (3) LSTM cell layer, which provides \( n \) cells with different memory capacity; (4) output layer, which is to pass the output of LSTM cells.

LSTM cell structure is shown in Fig. 2b. There are two key states in calculation of LSTM cell, that is, Cell State \( (S_{t-1}) \) and Hidden State \((y_{t-1}, \gamma_{t-1})\). In the hidden state, information can be added or deleted from the cell state, which is controlled by “Forget Gate”, “Input Gate” and “Output Gate”. LSTM cell uses all of these gates to control memory progress so as to avoid long-term dependence (Yu et al. 2019). The details of these gates are shown below.

The “Forget Gate” \( f_t \) decides what cell states are forgotten, and the forgotten extent is determined by sigmoid network output whose value ranges \([0, 1]\), where 1 represents “Keeping all” and 0 “Forgetting all”. The “Forget Gate” makes it possible to recall past information in current time step, see the Formula (1); “Input Gate” \( i_t \) determines the update of cell state, consisting of sigmoid network and a tanh network. The sigmoid network here determines how much “Hidden State” information is involved in updating, see the Eq. (2); the output of the tanh network is one-dimensional vector, which determines how much “Hidden State” information needs updating, see the Eq. (3); based on the output from “Forget Gate” and “Input Gate”, the cell state can be updated, see the Eq. (4); in the “Output Gate” \( \alpha_t \), the sigmoid network is used to determine which information of the hidden state is output, see the Eq. (5); finally, the output of LSTM cells is as shown in Eq. (6).

\[
f_t = \sigma(W_f \times [y_{t-1}, x_t] + b_f)
\]

\[
i_t = \sigma(W_i \times [y_{t-1}, x_t] + b_i)
\]

\[
\overline{C_t} = \tanh(W_c \times [y_{t-1}, x_t] + b_c)
\]
\[ S_t = f_t \times S_{t-1} + i_t \times C_t \]  (4)

\[ o_t = \sigma(W_o \times [y_{t-1}, x_t] + b_o) \]  (5)

\[ A \times y_t = o_t \times \tanh(S_t) \]  (6)

where, \( \delta \) represents \( s \)-network; \( W_f \) and \( b_f \) represent weight matrix and bias of sigmoid network in “Forget Gate”, respectively; \( W_i \) and \( b_i \) represent weight matrix and bias of sigmoid network in “Input Gate”, respectively; \( \tanh \) is one-dimensional matrix, with value range \([0,1]\); \( W_c \) and \( b_c \) represent weight matrix and bias of tanh network, respectively; \( f_t \times S_{t-1} \) indicates the information transmitted from “Forget Gate”, and \( i_t \times C_t \) indicates the updated information from “Input Gate”; \( W_o \) and \( b_o \) represents weight matrix and bias of “Output Gate”, with \( o_t \) being the output of LSTM cell.

The LSTM network has two categories of parameters: neuron-parameters and super-parameters. The neuron-parameters such as neuron weight and bias are updated in the model iterations, while super-parameters namely time-step, batch-size, and cell-size vary in their sensitivity and need adjustment based on experience. In particular, the parameter time-step reflects the length of data series. It can extract and learn the periodic relationship within time-step by learning data series. The parameter batch-size represents the number of data series in the training data packet extracted at one time; the product of the parameters time-step and batch-size represents the training data amount extracted for a single training. And parameter cell-size represents the number of neurons in a single hidden layer.

In rainfall-runoff simulation application, the rainfall data is used as the input data of the LSTM network, the number of input nodes equals the number of rainfall stations in study basin. The runoff data is the output data of the LSTM network, and the number of output nodes is equal to the number of hydrological stations in study basin.

### 2.3 Performance Metrics

Nash–Sutcliffe Efficiency Coefficient (NSE) and Correlation coefficient \( R^2 \) are used to evaluate the simulation effect of the LSTM network. The calculation formula is given below:

\[ NSE = 1 - \frac{\sum_{t=1}^{T} (Q'_t - Q'_m)^2}{\sum_{t=1}^{T} (Q'_t - Q'_0)^2} \]  (7)

where, \( Q'_0 \) represents the measured value, \( Q'_m \) the simulation value, \( Q'_t \) the value at time \( t \), and \( Q'_0 \) mean value of the measured values. The closer the NSE value is to 1, the higher the reliability of the model.

\[ R^2 = \frac{\left[ \sum_{i=1}^{\text{lag}} \left( Q_{m,i} - \bar{Q}_m \right) \left( Q_{s,i} - \bar{Q}_s \right) \right]^2}{\sum_{i=1}^{\text{lag}} \left( Q_{m,i} - \bar{Q}_m \right)^2 \sum_{i=1}^{\text{lag}} \left( Q_{s,i} - \bar{Q}_s \right)^2} \]  (8)

where, \( Q_{m,i} \) represents measured data, \( Q_{s,i} \) simulation data, \( \bar{Q}_m \) mean value of measured data, and \( \bar{Q}_s \) mean value of simulation data. The value range of \( R^2 \) is \([0, 1]\). The closer the \( R^2 \) is to 1, the better the simulation effect.
3 Study Area and Model Building

3.1 Overview of the Study Area

Three watersheds including Minjiang River, Jialing River and Yalong River, are selected as study area. The study details of the three watersheds are shown below.

The Minjiang River, an important tributary of the upper Yangtze River, stretching from latitude 28°13′N to 33°39′N and from longitude 99°36′E to 104°37′E, covers a total drainage area of 135,400 km² (Yin et al. 2022). Except for a very small amount of snow water replenishment in Zhenjiangguan Town and beyond of its trunk stream in Spring, the seasonal variation of Minjiang River’s runoff in a year corresponds to its rainfall season. The annual average flow at the Minjiang Estuary is 2850 m³/s, with the annual runoff of 90 billion m³. The rainfall data from 8 meteorological stations and runoff data from Xiangjiaba hydrological station in the watershed are used in the paper.

Jialing River is the largest tributary of the Yangtze River, stretching from latitude 29°18′N to 34°32′N and from longitude 102°31′E to 109°16′E, with a length of 1345 km and a drainage area of 160,000 km² (Yin et al. 2022). The main runoff sources is rainfall and groundwater. The flood period is June to September each year. The rainfall data from 12 meteorological stations and runoff data from Shapingba hydrological station in the watershed are used in the paper.

Yalong River, located at 26°32′~33°58′N and 96°52′~102°48′E (Liu et al. 2019), covers a total drainage area of approximately 136,000 km², with a length of 1571 km. The average annual flow is 1890 m³/s and the annual runoff is 59.6 billion m³. Half of the runoff comes from rainfall, with the other from underground water and snow melt water. Its annual runoff is abundant and stable, with little annual variability. The rainfall data from 7 meteorological stations and runoff data from Panzhihua hydrological station in the watershed are used in the paper.

The location of the three river basins and the distribution of their meteorological stations are shown in Fig. 3 below.

3.2 Generation of Training Data

In this study, the training data of Minjiang River, Jialing River and Yalong River basins are generated using the SWAT model and KNN algorithm. For the three basins, first, the observed daily meteorological data and runoff data during 1991 to 2016 are used to calibrate SWAT model. Then, KNN algorithm, featured by being non-parametric and inert, is used to derive 130 years of daily rainfall data. Finally, the calibrated SWAT model is used to simulate 130 years of daily runoff process in the three river basins.

The procedure to derive prediction data using KNN model included the following steps: (1) prepare data (i.e., observed daily meteorological data during 1991 to 2016); (2) select parameter K value to predict number of neighboring days; (3) traverse the data set to extract the meteorological data of the neighbor days; (4) calculate the distance between the extracted meteorological data and their mean value; (5) select a probability depending on the distance to generate random meteorological data.

As a distributed watershed hydrological model, SWAT is often used to simulate the long-term hydrological process. The data required by the SWAT model mainly includes the Digital Elevation Model (DEM), the digital river network, the hydrometeorological...
data (temperature, precipitation, solar radiation, weed speed, relative humidity and stream flow), the land use and soil data. The model was calibrated by using Sufi2 given in SWAT-CUP software, the key parameters are same to that listed in the Table 3 of the paper of Xu et al. (2020).

Besides, the 130 years of rainfall and runoff data is divided into two parts: the first 80 years of data and the remaining 50 years of data. The former is used for model training and the latter for model verification. Among the training data, the first 5, 10, 20, 40 and 80 years of rainfall data (input factor) and runoff data (output factor) are taken respectively to train the LSTM network, so as to analyze the effect of training data amount on the runoff simulation performance of the LSTM network.

### 3.3 Selection of Training Parameters

Drawing lessons from the parameter selection experience in Xu et al. (2020) in this study, the parameters time-step, batch-size and cell-size are discretized into multiple values, while the parameters Num_layer, Learning_rate and epoch are set as a fixed value. Therefore, there are 36 different parameter combinations in total. The detailed parameter value ranges and meanings are shown in Table 1. Based on these 36 combinations, the learning performance was analyzed according to the schemes of 5, 10, 20, 40 and 80 years of training data.

---

**Fig. 3** The location of three river basins and distribution of their meteorological stations

---

![Map of three river basins and meteorological stations](image-url)
4 Results and Analysis

4.1 Analysis of KNN Simulation Results

To evaluate the rationality of KNN simulation results, the coefficient of variable (CV) and the coefficient of skew (CS) of the simulated meteorological data series of 130 years are compared to those of measured series. The results of monthly rainfall and daily maximum temperature at Yushu rainfall station in Minjiang basin are shown in Fig. 4. It can be seen that the hydrological characteristics of simulated data was similar to that of measured data, which indicates that the simulated data can be used for the building of the LSTM network.

| Parameter name       | Parameter meaning                              | Value range |
|----------------------|------------------------------------------------|-------------|
| time-step            | Maximum time series stage length               | 80, 120, 160|
| batch-size           | Size of training data packet                   | 10, 20, 30  |
| cell-size            | Number of neurons in single hidden layer       | 10, 20, 30, 40|
| Num_layer            | Number of hidden layers in LSTM cell layer     | 1           |
| Learning_rate        | Learning rate                                  | 0.01        |
| epoch                | Iterations                                     | 600         |

![Fig. 4](image.png) Verification of KNN simulation results for Yushu rainfall station in Minjiang River basin
4.2 Effect of Parameter Combinations

To test the effect of the parameter combinations, 900 simulations, consisted of 36 available parameter combinations, 5 training data sets and 5 simulating times, were conducted for each of the three basins. The results are shown in Fig. 5. Figure 5 (YL1)–(YL5) shows the NSEs of the verification phase in the Yalong River Basin. It can be seen that the accuracy
of the LSTM network at the verification phase is relatively high, and the change of time-step, batch-size and cell-size leads to obvious and regular changes of the accuracy.

The NSEs at the verification phase for the Minjiang River basin is shown in Fig. 5 (MJ1)–(MJ5). It can be seen that, as the parameter combination changes from 1 to 4, time-step and batch-size are set as 80 and 10 respectively, and cell-size increases from 10 to 40, the NSE shows a downward trend, indicating that increasing the cell-size value can reduce the performance of the LSTM network.

The simulation results of Jialing River, shown in Fig. 5 (JL1)–(JL5), are similar to that of Yalong River and Minjiang River. It also can be observed that parameter changes may lead to the obvious change in the performance of LSTM network. In summary, the prediction accuracy of the LSTM network fluctuates as the parameters change.

In order to further study the effect of each parameter on the prediction accuracy, Fig. 5 (YL1) is enlarged and divided by crest and trough, as shown in Fig. 6. And the corresponding parameter values are shown in Table 2. From Fig. 6, it can be seen that, Zone 1, Zone 2 and Zone 3 are located at the crest, and the corresponding LSTM network has a higher prediction accuracy; Zone 4, Zone 5, and Zone 6 are located at the trough, and the corresponding LSTM network has a lower prediction accuracy. From Table 2, it can be observed that, for the parameter combinations in the crest zone, most of the batch-size and cell-size values are 10 and 20, while for the parameter combinations in the trough zone, most of the batch-size and cell-size values are 30 and 40. Therefore, it can be said that, the LSTM network can obtain higher prediction accuracy when the batch-size and cell-size values are smaller.

### 4.3 Effect of Training Data Amount

To evaluate the effect of training data amount on the performance of LSTM network in the three basins, for each training data amount, the mean NSE of 5-time simulations for the

#### Table 2 Parameter values for parameter combinations in crest and trough zone

| Parameters | Parameter combination |
|------------|-----------------------|
|            | Crest Zone            | Trough Zone         |
| time-step  | 80 120 120 160 160    | 80 80 120 120 160   |
| batch-size | 10 20 20 10 20        | 30 30 30 30 30      |
| cell-size  | 10 10 20 40 10        | 30 40 30 40 40      |
training phase and the verification phase with the same parameters is calculated, and the results are shown in Fig. 7.

Figure 7(YL1)–(YL5) are the results of Yalong River Basin. It can be seen that, in Fig. 7(YL1), the NSE at the training phase is much larger than that at the verification phase. The reason is that, the over-fitting, caused by the inadequate hydrological data at the training phase, leads to very poor extension capabilities of the LSTM network at the verification phase. Compared to Fig. 7(YL1), Fig. 7(YL2) shows that increasing the training data amount can lead to the decrease of the NSE at the training phase, and the raising of the NSE at the verification phase, as well as the reduction of the over-fitting phenomenon. In Fig. 7(YL3), the decline of NSE at the training phase and the increase of the
NSE at the verification phase results in the smaller gap between them, which proves that the trained network is highly stable. Figure 7(YL4) and (YL5) shows that the increase of the data amount could further weaken the over-fitting phenomenon of the LSTM network, but the improvement of the NSE at the verification phase is limited. All the above results show that, as the data amount at the training phase increases, the LSTM network tends to become more stable.

The simulation results of Minjiang River basin are shown in Fig. 7(MJ1)–(MJ5). In Fig. 7(MJ1), the NSE in the training phase remains around 0.9, and it is only around 0.7 in the verification phase. In Fig. 7(MJ2)–(MJ5), the NSE at the training phase declines obviously while it increases at the verification phase. It also corroborates that the increase of training data amount weakens the over-fitting of LSTM network.

The simulation results of Jialing River basin is shown in Fig. 7(JL1)–(JL5). It can be seen that, when the training data amount increases from 5 years of data to 10, 20 and 40 years of data, the NSE at the training phase and verification phase gradually decreases. But when the training data amount increases to 80 years of data, the NSE at the training phase and verification phase stands at the same level, indicating that the simulation performance of LSTM network keeps stable.

In order to further analyze the effect of training data amount on performance of the LSTM network, the NSE and $R^2$ of 180 simulations for the verification phase with the same training data amount for the three basins is averaged. The mean NSE and $R^2$ are shown in Fig. 8. It can be seen that, as the training data increases, the value of NSE and $R^2$ increases too, which indicates that the increase of training data amount can considerably reduce the over-fittings in the LSTM network and improve its prediction accuracy. But it also can be observed that when the training data amount exceeds a certain value or value range, the performance of LSTM network can’t be improved remarkably, and the critical value varies for different basins. That is, increasing training data amount can improve the forecast performance, but over-increase of training data amount cannot necessarily lead to better results. In other words, choosing an appropriate amount of training data can better exert the learning ability of the LSTM network.

### 4.4 Analysis of Runoff Process

In Sects. 4.2 and 4.3, it is illustrated that LSTM network can achieve better performance when the amount of training data is set as 20 years, and the parameters time-step, batch-size, and cell-size are set as 120, 10, and 20, respectively. The results of the corresponding

![Fig. 8 NSE and $R^2$ average value at the verification phase](image-url)
LSTM network for the three basins are analyzed in this section to observe the learning ability of the LSTM network.

For the 50-year verification phase, the simulated runoff of the three basins are compared with the observed runoff in Fig. 9. It can be seen that the NSEs for Minjiang River, Jialing River and Yalong River at the verification phase are up to 0.91, 0.94 and 0.96 respectively, which indicates that the LSTM network has a strong learning ability and thus can be used for runoff simulation.

5 Conclusion

In order to analyze the influence of training data amount on the learning performance of the LSTM network in runoff simulation, the LSTM network is used for the runoff simulation of Minjiang River, Jialing River and Yalong River, and its performance under different parameters and training data amount schemes are evaluated in this study. The following conclusions can be drawn from the study:

1. When the parameters change, the prediction accuracy of the LSTM network changes too. When the batch-size and cell-size values are set as 10 or 20, the LSTM network has a better performance, while it has a lower performance when the batch-size and cell-size values are set as 30 or 40. The results shows that the LSTM network could provide higher prediction accuracy when the parameter batch-size and cell-size are smaller.
2. As the training data amount increases, the over-fitting phenomena can be reduced and the learning performance of the LSTM network improves significantly. But when the training data amount increases to a certain extent, the prediction accuracy of the LSTM network no longer improves much. The LSTM network in this study achieves better prediction accuracy and stability when more than 20-year training data amount are used.
3. When the parameters of time-step, batch-size, and cell-size are set as 120, 10, and 20, respectively, and the amount of training data set as 20 years, the NSEs for Minjiang River, Jialing River and Yalong River at the verification phase are up to 0.91, 0.94 and 0.96 respectively, which shows that the LSTM network can be used for runoff simulation.

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection were performed by Wei Xu, model built by Anbang Peng, data analysis were performed by Xiaoli Zhang, figure drawn by Yuanyang Tian. The first draft of the manuscript was written by Anbang Peng and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.
Funding This study was funded by the National Natural Science Foundation of China (Grant No. 51709177, 51709108).

Declarations

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

References

Dami S, Esterabi M (2021) Predicting stock returns of Tehran exchange using LSTM neural network and feature engineering technique. Multimed Tools App 80(13):19947–19970

Dikshit A, Pradhan B, Huete A (2021) An improved SPEI drought forecasting approach using the long short-term memory neural network. J Environ Manage 283:111979

Dilip KR (2021) Long short-term memory networks to predict one-step ahead reference evapotranspiration in a subtropical climatic zone. Environ Process 8(2):911–941

Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504–507

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hu C, Wu Q, Li H, Jian S, Li N, Lou Z (2018) Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. Water 10(11)

Huang Z, Hasan A, Shin K, Li R, Driggs-Campbell K (2021a) Long-term pedestrian trajectory prediction using mutable intention filter and warp LSTM. IEEE Robot Autom Lett 6(2):542–549

Huang Z, Wang J, Pi L, Song X, Yang L (2021b) LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment. Pattern Recognit 112:107800

Jiang Y, Liu X (2021) Image/video aesthetic quality management based on fusing 3D CNN and shallow features. Futur Gener Comput Syst 118:118–123

Lall U, Sharma A (1996) A nearest neighbor bootstrap for resampling hydrologic time series. Water Resour Res 32(3):679–693

Lin Z, Chen AY, Hsieh S (2021) Temporal image analytics for abnormal construction activity identification. Automat Construct 124:103572

Liu X, Yang M, Meng X, Wen F, Sun G (2019) Assessing the impact of reservoir parameters on runoff in the Yalong River Basin using the SWAT model. Water 11(4):643

Mao G, Wang M, Liu J, Wang Z, Wang K, Meng Y, Zhong R, Wang H, Li Y (2021) Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation. Phys Chem Earth Parts A/B/C 123:103026

Rahimzad M, Moghaddam Nia A, Zolfonoon H, Soltani J, Danandeh Mehr A, Kwon H (2021) Performance comparison of an LSTM-based deep learning model versus conventional machine learning algorithms for streamflow forecasting. Water Resour Manage 35(12):4167–4187

Wang W, Du Y, Chau K, Xu D, Liu C, Ma Q (2021a) An ensemble hybrid forecasting model for annual runoff based on sample entropy, secondary decomposition, and long short-term memory neural network. Water Resour Manage 35(14):4695–4726

Wang X, Wang Y, Yuan P, Wang L, Cheng D (2021b) An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. Hydrol Sci J 66(9):1488–1502

Xu W, Jiang Y, Zhang X, Li Y, Zhang R, Fu G (2020) Using long short-term memory networks for river flow prediction. Hydrol Res 51(6):1358–1376

Yin H, Zhang X, Wang F, Zhang Y, Xia R, Jin J (2021) Rainfall-runoff modeling using LSTM-based multi-state-vector sequence-to-sequence model. J Hydro 598:126378

Yin S, Xu W, Xiong Y, Tian Y, Zhao S, Chen S (2022) Hydrological model based on long short-term memory neural network and transfer learning. J Hydroelectr Eng (in Chinese)

Yu Y, Si X, Hu C, Zhang J (2019) A review of recurrent neural networks: LSTM cells and network architectures. Neural Comput 31(7):1235–1270

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