


Article

Runoff Simulation in Data-Scarce Alpine Regions: Comparative Analysis Based on LSTM and Physically Based Models

Jiajia Yue ¹, Li Zhou ^{1,2,*} , Juan Du ¹, Chun Zhou ¹, Silang Nimai ¹, Lingling Wu ³ and Tianqi Ao ^{1,2,*}

¹ State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu 610065, China; yuejiajia@stu.scu.edu.cn (J.Y.); dujuan@stu.scu.edu.cn (J.D.); zhouchun@stu.scu.edu.cn (C.Z.); nimaisilang@163.com (S.N.)

² Institute for Disaster Management and Reconstruction, Sichuan University-Hong Kong Polytechnic University, Chengdu 610065, China

³ Sichuan Hydrological and Water Resources Survey Center, Chengdu 610065, China; djy-wulingling@163.com

* Correspondence: li_zhou@scu.edu.cn (L.Z.); aotianqi@scu.edu.cn (T.A.)

Abstract: Runoff simulation is essential for effective water resource management and plays a pivotal role in hydrological forecasting. Improving the quality of runoff simulation and forecasting continues to be a highly relevant research area. The complexity of the terrain and the scarcity of long-term runoff observation data have significantly limited the application of Physically Based Models (PBMs) in the Qinghai–Tibet Plateau (QTP). Recently, the Long Short-Term Memory (LSTM) network has been found to be effective in learning the dynamic hydrological characteristics of watersheds and outperforming some traditional PBMs in runoff simulation. However, the extent to which the LSTM works in data-scarce alpine regions remains unclear. This study aims to evaluate the applicability of LSTM in alpine basins in QTP, as well as the simulation performance of transfer-based LSTM (T-LSTM) in data-scarce alpine regions. The Lhasa River Basin (LRB) and Nyang River Basin (NRB) were the study areas, and the performance of the LSTM model was compared to that of PBMs by relying solely on the meteorological inputs. The results show that the average values of Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE), and Relative Bias (RBias) for B-LSTM were 0.80, 0.85, and 4.21%, respectively, while the corresponding values for G-LSTM were 0.81, 0.84, and 3.19%. In comparison to a PBM– the Block-Wise use of TOPMEDEL (BTOP), LSTM has an average enhancement of 0.23, 0.36, and –18.36%, respectively. In both basins, LSTM significantly outperforms the BTOP model. Furthermore, the transfer learning-based LSTM model (T-LSTM) at the multi-watershed scale demonstrates that, when the input data are somewhat representative, even if the amount of data are limited, T-LSTM can obtain more accurate results than hydrological models specifically calibrated for individual watersheds. This result indicates that LSTM can effectively improve the runoff simulation performance in alpine regions and can be applied to runoff simulation in data-scarce regions.



Citation: Yue, J.; Zhou, L.; Du, J.; Zhou, C.; Nimai, S.; Wu, L.; Ao, T. Runoff Simulation in Data-Scarce Alpine Regions: Comparative Analysis Based on LSTM and Physically Based Models. *Water* **2024**, *16*, 2161. <https://doi.org/10.3390/w16152161>

Academic Editor: Bommanna Krishnappan

Received: 21 June 2024

Revised: 26 July 2024

Accepted: 28 July 2024

Published: 31 July 2024



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

Keywords: runoff simulation; LSTM; prediction in ungauged basin; BTOP; transfer learning

1. Introduction

Runoff simulation is a key element in hydrological forecasting, addressing pivotal scientific challenges within the fields of hydrology and water resource research [1,2]. The rainfall runoff models are instrumental in the accurate simulation and prediction of flood events, thereby playing an indispensable role in flood mitigation, disaster preparedness, the stewardship of water resources, and the oversight of ecological and environmental health [3–7]. Due to global climate warming and human activities, extreme weather events have increased, exacerbating flood and drought disasters. These events have caused severe economic and social losses to countries worldwide [3]. The Qinghai–Tibet Plateau (QTP) is one of the most sensitive regions to climate change, with amplified temperature and precipitation changes significantly affecting hydrological cycles and future freshwater availability [8–10]. As the world’s primary water tower, runoff from the QTP is crucial

to regional hydrological processes and water supply for downstream populations, and runoff changes in alpine areas are also closely related to natural disasters such as floods and droughts [2,11]. However, there is insufficient spatial density in many parts of the world for adequate long-term river flow observations, especially in the QTP, which is rich in water resources but has few meteorological and hydrological stations. The lack of long-term runoff data poses significant challenges for physically based hydrological models (PBMs), and conventional black-box models often struggle to accurately simulate runoff [12,13]. The scarcity of monitoring stations leads to a significant data deficit, severely limiting the accuracy and effectiveness of forecasting and early-warning systems for various meteorological and hydrological disasters in Tibet [14,15]. The challenge posed by flow simulation in QTP underscores its importance as a pivotal research area in hydrology [16]. Therefore, developing tailored runoff simulation methods for data-scarce basins to enhance the accuracy of runoff simulation holds significant scientific importance and practical application value [17].

Prediction in Ungauged Basins (PUB), launched in 2003, was the decadal project of the International Association of Hydrological Sciences (IAHS) [18]. Over the past two decades, numerous approaches have been proposed to predict runoff in ungauged basins, ranging from parameter regionalization, data assimilation, catchment similarity, and surrogate modeling techniques. A diverse range of innovative data acquisition techniques and models have been developed to improve the confidence of estimations [19–21]. The PUB initiative has resulted in considerable progress. One of the core advances was a relatively widespread awareness of the need to understand flexible modeling methods, which can be adapted to the specific environmental conditions of different basins and bring significant advantages to simulation. Focusing on site-specific dominant processes has been shown to reduce the uncertainty in predictions [22]. Samaniego et al. [23,24] suggested that with the increasing amount of model computation, surrogate models, which can reduce computational costs and achieve good reasonable estimates, are becoming more and more popular in the field of hydrology and water resources [25,26]. However, PUB remains a significant challenge because the majority of streams in the world are either ungauged or poorly gauged, while notable achievements have been realized in gauged basins, rather than ungauged ones [19,20,27,28].

With the advancement of artificial intelligence, machine learning has been increasingly applied to water resource monitoring and management. Its ability to handle diverse inputs simultaneously, combined with lower costs and reduced time requirements, makes it a valuable tool in this field [4]. Common flow simulation models, such as artificial neural networks, have made significant advances in predicting and simulating hydrological processes and perform well in capturing non-stationarity, dynamics, and nonlinearity within datasets [29]. Deep learning (DL) based on artificial intelligence has demonstrated its effectiveness in revealing relationships between hydrological variables. Specifically designed for time series forecasting [30], the Long Short-Term Memory (LSTM) network has consistently provided reliable results in rainfall runoff applications in various localities and regions worldwide [31]. Numerous studies have shown that LSTM can learn the dynamic hydrological characteristics of watersheds from long time series data and outperform some traditional PBMs in flow prediction [32–35]. Furthermore, it has demonstrated good performance in predicting ungauged basins [36]. Kratzert et al. [19] conducted a comprehensive study on flow prediction across 531 basins, wherein the LSTM exhibited superior performance with an average Nash–Sutcliffe efficiency (NSE) value of 0.69 compared to the Sacramento Soil Moisture Accounting (SAC-SMA) conceptual hydrological model and the National Water Model. Ouyang et al. [37] evaluated LSTM in 3557 small watersheds in the United States and found that LSTM performed well in watersheds with minimal human activity. The application of LSTM for runoff simulation facilitates the elucidation of underlying hydrological processes, providing insights not possible with process-based models. Furthermore, in contrast to the PBMs that necessitate a wide range of data such as meteorology, the Digital Elevation Model (DEM), evapotranspiration, soil, and land use

during modeling, the DL model, being a black model, prioritizes the independence of input data, making even a relatively small source dataset valuable. These advantages are crucial in mountainous watersheds where data are scarce [11,38].

Meanwhile, transfer learning (TL) has also emerged as a promising alternative method for transferring hydrological knowledge learned from gauged basins to these ungauged basins [11,39,40]. Kratzert et al. [19,35] trained a kind of Entity-Aware LSTM (EA-LSTM) on many watersheds and applied it to ungauged watersheds using only meteorological and geographical characteristics. The results indicated that EA-LSTM generally provided better flow estimates at ungauged locations. Ma et al. [38] introduced a TL approach using rich data from the CAMELS dataset to train the LSTM model, and then the model was transferred to data-scarce regions; the runoff simulations accuracy was enhanced in Asia, Europe, and South America but were not well in areas where glaciers predominate. Khoshkalam et al. [41] investigated the runoff predictions of a TL model which based on LSTM and compared the results with the HYDROTEL model in the United States. These findings suggest that the unique structure of LSTM can enhance the simulation accuracy of data-scarce basins through TL, but this advantage is not as pronounced in regions where glaciers are prevalent [38,42]. As the third pole of the world, the QTP presents a complex natural environment characterized by the widespread presence of snow, permafrost and glaciers, along with constraints in data availability, which make the TL seldomly realized; thus, there is a pressing need to further explore the applicability and effectiveness of TL in data-scarce basins in QTP.

Previous studies have increased our knowledge of the application of LSTM in runoff simulation, but due to the unique geographical environment of the QTP area and its importance for downstream water resources management and disaster prevention, accurate simulation of runoff in this area is of great importance. In this study, we aim to evaluate the potential of LSTM based on TL in daily streamflow simulation and data gap filling in this region. We applied Block-Wise use of TOPMEDEL (BTOP) and LSTM for streamflow simulations over the Lhasa River Basin (LRB) and Nyang River Basin (NRB) in QTP. We compared the performance of the BTOP and the LSTM in simulating runoff based on the China Meteorological Forcing Dataset (CMFD) dataset and the China Meteorological Data Service Center (CMA) dataset. Then, we integrated TL with LSTM to find out if a transfer-based LSTM (T-LSTM) trained on data-rich sites can be transferred to a data-scarce site, and to assess its effectiveness in simulating runoff at a multi-watershed scale across two basins, which will give us the opportunity to accurately address significant data deficits.

This article is structured as follows: Section 2 outlines the study area and data required for model construction. Section 3 presents the introduction and its setup, along with the design of multiple experimental schemes under two different experiment types (single-watershed and multi-watershed), encompassing data processing, model training and testing, and the model evaluation criteria. Section 4 introduces and discusses our experimental results and offers perspectives on future research directions. This paper concludes with a summary of our conclusions.

2. Study Area and Data

2.1. Study Area

The study areas of this research are shown in Figure 1.

The Lhasa River Basin, one of the highest rivers in the world, is located in the central-southern part of the QTP (Figure 1c). The total river length is approximately 495 km, covering an area of 32,526 km². The average annual precipitation in the basin is about 500 mm, which is mainly concentrated from June to September [43]. Over 80% of the average annual precipitation in Tibet occurs during the four wettest months. The annual runoff varies greatly. The Lhasa River is the largest tributary of the Yarlung Tsangpo River basin, and it flows through the political, economic, and cultural center of the Tibet Autonomous Region. It serves as one of the most important water sources in the region,

making research on its streamflow essential for more accurate water resource prediction and management [44].

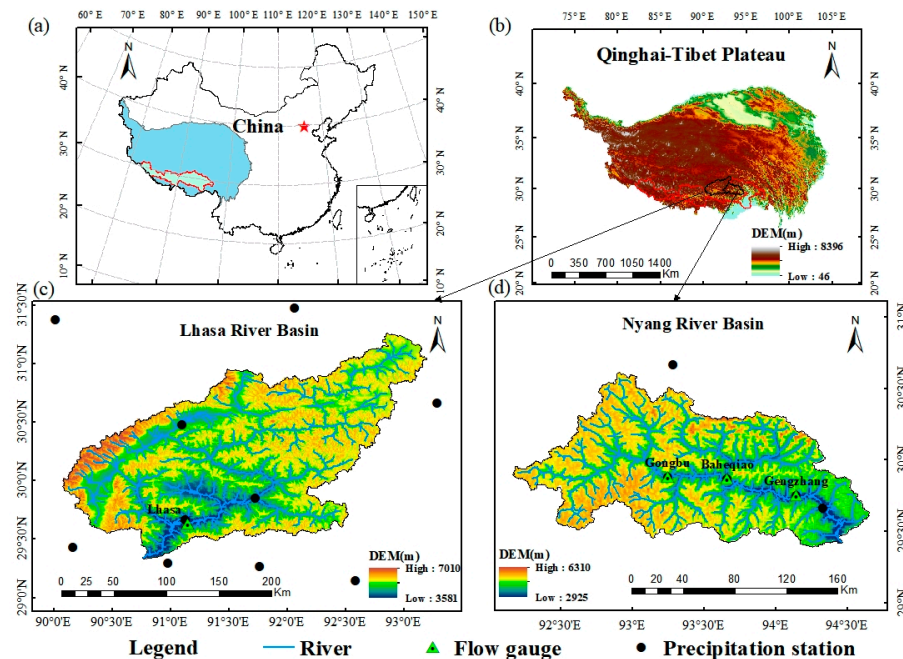


Figure 1. Location, dem, observed station of the study areas. (a) Location of the Qinghai-Tibet Plateau (QTP) and the Yarlung Tsangpo River basin (red boundary) in China, with the red star indicating the location of China's capital city, Beijing. (b) Dem of the QTP and the location of the Lhasa River Basin (LRB) and Nyang River Basin (NRB) in the Yarlung Tsangpo River basin. (c,d) DEM, river network, flow gauges and precipitation stations in the LRB and NRB, respectively.

The headwaters of the NRB exhibit extensive ancient glacial topography, with many areas still containing modern glaciers. The glacier and permanent snow cover is 5.47% of the total basin area, and the contribution of glacial meltwater to streamflow is particularly significant [45]. NRB covers an area of 17,535 km² (Figure 1d). The mainstream is 307 km long, with an average gradient of 7.39%. As a mountain river in the southern QTP, NRB has unique geographical conditions and falls within a plateau temperate monsoon climate. The average annual precipitation ranges from 700 mm to 1100 mm. The Nyang River is the second-largest tributary of the Yarlung Tsangpo River and one of the regions in Tibet with substantial precipitation. The ecological environment of NRB is highly fragile, with frequent natural disasters and increasing human impacts on the environment. Therefore, runoff research in NRB is crucial for disaster prevention and ecological security in QTP [46].

2.2. Data

2.2.1. Forcing Data

Observed discharge from Lhasa, Baheqiao (BHQ), Gengzhang (GZ), and Gongbu (GB) stations was obtained from the Tibet Hydrology Bureau. Precipitation station data, including observed daily precipitation and average daily temperature, were obtained from the CMA (<https://data.cma.cn/en/>, accessed on 26 January 2024). LRB has a total of nine precipitation stations, while the NRB has a total of five. The potential evaporation (PET₀) and potential evapotranspiration (PET) required to drive the models were employed from the ERA5-Land dataset, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) (<https://cds.climate.copernicus.eu/>, accessed on 9 March 2024).

When the actual density of precipitation observation networks is significantly lower than the values recommended by the World Meteorological Organization of 100–250 km²/station, the regional differences in precipitation are not well represented, and the performance of runoff simulation models may be affected [47,48]. To overcome the

limitations due to data scarcity or poor observation quality, reanalysis meteorological products may be an option for flow prediction [3]. The use of the China Meteorological Forcing Dataset (CMFD) dataset can effectively reduce uncertainties caused by data scarcity [49]. This dataset is a composite of multiple data sources, including observations from Chinese meteorological stations, TRMM satellite precipitation analysis data, GLDAS data, GEWEX-SRB data and MERRA data. This dataset covers the period from 1979 to 2018, with a daily temporal resolution and a spatial resolution of 0.1° . The BTOP model was constructed based on the CMFD and CMA datasets, and the model performance was compared.

2.2.2. Static Data

The 30 m DEM data were sourced from the Shuttle Radar Topography Mission (SRTM), jointly acquired by the National Aeronautics and Space Administration (NASA) and the National Imagery and Mapping Agency (NIMA) [50]. Land cover data, classified into 17 categories based on the International Geosphere-Biosphere Programme (IGBP), were obtained from the United States Geological Survey (USGS) [51]. Soil distribution data for sand, silt, and clay were gathered from the Food and Agriculture Organization (FAO) of the United Nations [52]. Leaf Area Index (LAI) data were acquired from the NOAA Climate Data Record (CDR) (<https://www.ncei.noaa.gov/>).

Achieving high accuracy in climate inputs, particularly precipitation data, is crucial for the precision of hydrological simulations [53]. To maintain stability and efficiency in model calibration, the spatial resolution of all grid data was unified to 0.01° in both LRB and NRB. Additionally, area-weighted spatial means were calculated for each meteorological variable in each basin.

3. Methodology

3.1. BTOP Model

The Block-Wise use of the TOPMEDEL (BTOP) model is a physically based semi-distributed hydrological model initially developed for runoff simulations in large river basins [54]. It divides the whole basin into BTOP grids (several natural sub-basins) based on topographic heterogeneity and assumes that regions within the basin that share the same topographic index exhibit identical hydrological response [55]. It has been widely used in more than two thousand basins for varied objectives, such as parameter regionalization, water management, flood forecast, and water environment simulation [56–59]. The BTOP model has also achieved effective simulations in many basins worldwide, such as the Min River Basin in China, the Mekong River Basin, the Fuji River Basin in Japan, and ungauged basins of Jialing River, southwest of China [57,60,61]. These wide application results indicate that the BTOP model is a useful rainfall runoff simulation tool in many river basins. In addition, compared with other distributed hydrological model (such as SWAT and VIC), the BTOP model has fewer tuning parameters, and the parameter interaction and uncertainties are fewer, which can also effectively diminish both computational load and uncertainty [62,63]. In this study, the SCE-UA algorithm (Shuffled Complex Evolution Algorithm with Uncertainty Analysis) [64] has been employed for parameter optimization, and the Nash–Sutcliffe efficiency (NSE) serves as the objective function to evaluate the model performance [63,65].

3.2. Long Short-Term Memory Network

Long Short-Term Memory (LSTM) networks were originally proposed by Hochreiter and Schmidhuber [66] and belong to the recurrent neural network (RNN) family, which can effectively solve the challenges posed by gradient vanishing and explosion [67–70]. The core structure of the LSTM is divided into four parts: the input gate i_t , the cell state c_t , the forget gate f_t , and the output gate o_t [17], shown in Figure 2. In runoff simulation, the input gate can capture the impact of new precipitation data (such as data from periods of concentrated rainfall) and effectively integrate these into the model. The forget gate helps forget historical information that is unimportant for current runoff predictions,

reducing noise. The cell state plays a role in storing and transmitting long-term hydrological information in runoff simulation. The output gate dynamically adjusts the predicted runoff values based on the current cell state and input information. This ensures that the model considers both the impact of historical hydrological information and real-time data, such as current precipitation, thus improving the accuracy and reliability of predictions. The four core structures of LSTM work together to effectively handle long-term dependencies and short-term variations in time series data, thereby enhancing the effectiveness of runoff simulation. In other words, as has been well demonstrated, the LSTM architecture is highly suitable for generating predictions of complex dynamic hydrological system behaviors [71].

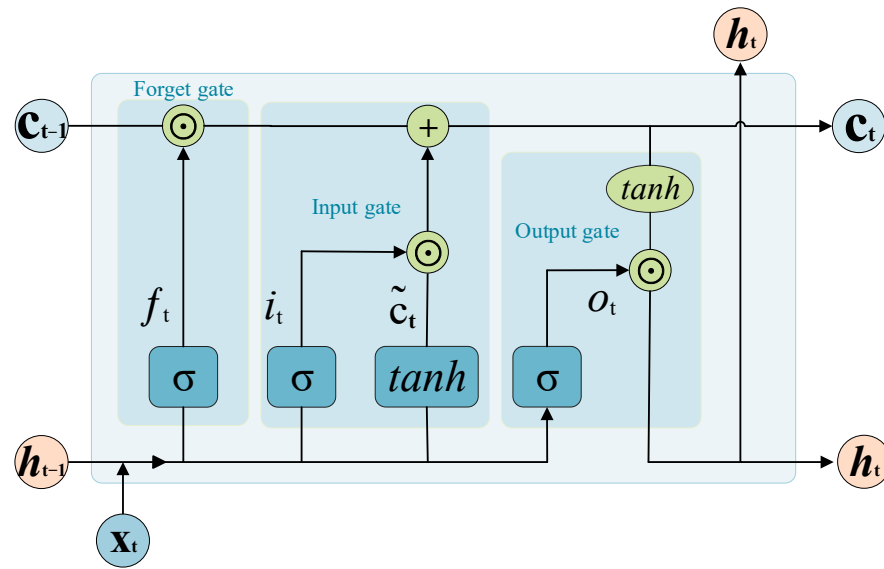


Figure 2. The core structure of the LSTM unit. σ denotes the sigmoid function and \tanh is the hyperbolic tangent activation functions. \odot represents element-wise multiplication. H_{t-1} is the hidden state from the previous time step, h_t is the hidden state, x_t is the input at the current time step, c_t is the cell state, \tilde{c}_t is the candidate memory cell state.

At each time step, the gates are updated by the following equations:

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

$$i_t = \sigma(W_i[h_{t-1}, x_t]) + b_i \quad (2)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t]) + b_o \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

$$h_t = \tanh(c_t) \odot o_t \quad (5)$$

In these equations, σ denotes the sigmoid function, W_f , W_i , W_o are weighted parameters, and b_f , b_i , b_o are the bias parameters. The cell state c_t is updated using the forget gate and the input gate, and the hidden state h_t is updated using the output gate and the new cell state c_t .

These gates are trained simultaneously and automatically, using input data to predict target variables [72–74].

Finding the optimal values for model hyperparameters is a critical task because it has a great impact on the accuracy and reliability of the model [41,75]. In this study, the optimized hyperparameters include batch size, number of hidden layer units, gradient descent method, learning rate, and window size. The root mean square error (RMSE) was used as the loss function in this study [76]. The gradient descent methods employed were Adam and RMSprop. The batch size controls the number of training examples used to

compute the loss function before updating the weight matrix, and the number of hidden layer units represents the complexity of the network.

3.3. Experimental Design

This study focuses on runoff simulation in data-scarce catchments characterized by sparse station networks and short, hard-to-obtain time series of observed hydrological data, where existing simulations from distributed hydrological physical models show certain discrepancies when compared to observed flow data [45]. Compared to PBMs, deep-learning models are less prone to structural errors [77]. To improve the accuracy of streamflow simulations in LRB and NRB, and to evaluate the potential of LSTM based on TL in these basins, we compared the results obtained from the LSTM with the BTOP model and transferred the LSTM trained on data-rich sites to a data-scarce site. The detailed experimental procedure is as follows. It is important to note that the modeling of hydrological models is typically divided into a calibration period and a validation period. In DL-parlance, they normally have a training–validation–testing period. Although the training–validation process of DL models shares some similarities, there are significant differences. Therefore, in the following discussions, the modeling results of hydrological models will include the calibration and the validation period, while the results of the LSTM will focus on the testing period.

3.3.1. Pre-Experiment of LSTM

Data preprocessing: The input data were categorized into dynamic, static, and semi-static data (whose changes are relatively slow and can be considered an intermediate state between static and dynamic, such as leaf area index). Validity encompasses checking the consistency of the data, calculating regional means, and compiling all the data into the recognizable format.

Hyperparameter tuning: Due to computational resource limitations, we determined the range of optimal hyperparameters by experimenting with increasing or decreasing these values [78]. We first adjusted the batch size, followed by simultaneously adjusting the window size, number of hidden layer units, and learning rate. We had 10–20 values for each hyperparameter, which were sequentially taken to compute the simulation results, and observe the influence of these hyperparameter on the calculation speed and results. After that, we listed multiple sets of hyperparameters whose simulation results were within a certain range as the optimal hyperparameter sets.

Model pre-training: Excessive input data may not improve the simulation performance of the model and may even degrade it [60]. To eliminate invalid data and reduce input data redundancy, different input data were compared and their effects on model simulation performance were tested. After this step, we identified the input data for the LSTM model.

Hyperparameter determination: To avoid the impact of network overfitting on the training data, different training epochs were used for different experimental designs. For the B-LSTM and G-LSTM, the model was trained on 100 epoch. For the T-LSTM model, there were 300 epochs.

3.3.2. Basin-Based LSTM and Gauge-Based LSTM

We filtered out redundant and static data through preliminary experiments and used only CMA daily precipitation, CMFD daily precipitation, and CMA daily temperature as inputs. We set up the BTOP and LSTM models separately in LRB and NRB.

The hyperparameter sets selected in the previous step were set one by one, all available data from each watershed were input into the pre-trained LSTM model, the training period was set as 2010–2012, the validation period as 2013, and the testing period as 2014–2015, and the NSE metrics of the simulation results were calculated.

The parameters' regionalization of the hydrological model can be used to reflect the differences in hydrological processes across different regions. However, the LSTM model

does not explicitly perform regionalization during the training period. In situations where only a few years of data are available, hydrological models may outperform LSTM models in simulating individual watersheds [36]. Therefore, we designed two types of LSTM models: basin-based LSTM (B-LSTM) and gauge-based LSTM (G-LSTM). The B-LSTM model treats the entire watershed as a single unit, simultaneously inputting data from all stations within the watershed into the model. Two B-LSTM models were constructed and trained–validated separately for the LRB and NRB watersheds, yielding overall simulation results for each watershed. In contrast, G-LSTM considers each of the three stations in the NRB watershed an individual unit, modeling each station separately. This involves training the model three times, once for one station, to obtain independent simulation results for each station.

3.3.3. Transfer-Based LSTM

The transfer-based LSTM (T-LSTM) is trained without the use of observed flow data of the target station; instead, it solely relies on limited local observational meteorological data. For clarity, the station used for training and validation is called the source station, and the station to be simulated through transfer is called the target station. In previous studies on transfer learning, the research periods typically follow a sequential order, with the training period preceding the testing period, using the same dataset but at different time [38,79]. However, in this study, we propose a unique temporal setup that not only addresses the essence of transfer learning—leveraging knowledge from the source station to enhance performance in the target station—but also delves into the effectiveness of transfer learning within the same temporal framework.

For our T-LSTM model, four transfer cases were carried out to evaluate its performance. The specific model settings are shown in Table 1. In each case, the data from the three source stations were used as the training and validation data; the fourth station is the target station and it was assumed that we do not have the observed discharge. The training period was set as 2010–2013, the validation period as 2014–2015, and the testing period as 2010–2015.

Table 1. The detailed information of experiment schemes.

Abbreviation	Case	Source Station	Training Periods	Validation Period	Target Station	Testing Periods
T-LSTM	①	GB, BHQ, GZ	2010–2013	2014–2015	Lhasa	2010–2015
	②	GB, BHQ, Lhasa	2010–2013	2014–2015	GZ	2010–2015
	③	GB, GZ, Lhasa	2010–2013	2014–2015	BHQ	2010–2015
	④	BHQ, GZ, Lhasa	2010–2013	2014–2015	GB	2010–2015

Due to the change in the amount of input data, we needed to repeat the steps in Section 3.3.1 to reselect the optimal set of hyperparameters for the T-LSTM. After that, different hydrological stations were sequentially selected as targets for training the T-LSTM model. By adjusting the hyperparameters and evaluating each model on test data, the set of parameters corresponding to the best results is referred to as the optimal parameter set. The detailed hyperparameters are listed in Section 4.1.

3.4. Model Evaluation Criteria

The following widely applied performance evaluation criteria were used to evaluate the accuracy of the model developed: Nash–Sutcliffe efficiency (NSE) [80], Relative Bias (RBias), and Kling–Gupta efficiency [81,82] (KGE). The formulas are as follows:

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2} \quad (6)$$

$$RBias = \frac{\sum_{i=1}^n (Q_{sim,i} - Q_{obs,i})}{\sum_{i=1}^n Q_{obs,i}} \times 100\% \quad (7)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \quad (8)$$

$$r = \frac{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})(Q_{sim,i} - \bar{Q}_{sim})}{\sqrt{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2 \sum_{i=1}^n (Q_{sim,i} - \bar{Q}_{sim})^2}}, \beta = \frac{\bar{Q}_{sim}}{\bar{Q}_{obs}}, \gamma = \frac{Cv_{sim}}{Cv_{obs}} \quad (9)$$

where n is the total time steps; $Q_{obs,i}$ and $Q_{sim,i}$ are the observed and the simulated river discharge at the time step i , and the mean values are denoted as \bar{Q}_{obs} and \bar{Q}_{sim} ; r is the Pearson's correlation coefficient, β is the bias ration, and γ is the variability ration, as defined in Equation (9).

For the hydrological simulation, our primary focus lies in replicating temporal variations and maintaining the runoff distribution patterns. Consequently, employing the KGE as a model metric becomes a logical choice as it effectively captures these aspects. Furthermore, the NSE can be viewed as a measure of a model's capability to replicate the observed mean, and an NSE value greater than 0 signifies that the model outperforms the naïve mean predictor [63,81]. Notably, this achievement is comparable to achieving a KGE value above -0.41 [83]. The RBias serves as a valuable indicator of the overestimation or underestimation of the model simulation.

The flowchart of this study is shown in Figure 3. In this study, the distributed hydrological model BTOP and LSTM model were modeled on NRB and LRB. Based on CMFD and CMA dataset, the performance of the BTOP model was compared with basin-based LSTM (B-LSTM) and gauge-based LSTM (G-LSTM)-based model on three model evaluation criteria. Four transfer cases were carried out with T-LSTM to evaluate its performance in simulating runoff in data-scarce areas.

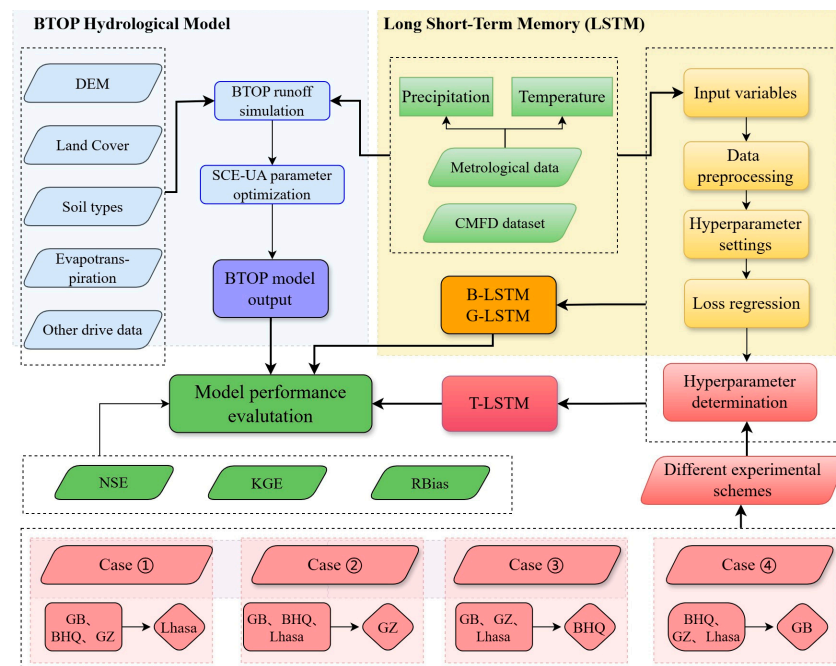


Figure 3. The flowchart of this study. (The station names in this graph are abbreviated as follows: Gongbu-GB; Baheqiao-BHQ; Gengzhang-GZ; Basin-based LSTM: B-LSTM; Gauge-based LSTM: G-LSTM; Transfer-based LSTM: T-LSTM.) In the four cases of the T-LSTM, the sites in the box are the source sites, and the site in the diamond box is the target site.

4. Results and Discussion

4.1. Hyperparameters of LSTM

The modeling process of the BTOP model is somewhat complex, and its parameter optimization process is time-consuming, with the computation efficiency of the model being highly dependent on the number of subregions. Conversely, the LSTM model has a simpler modeling process and a faster optimization process, and it can achieve good simulation results even with minimal input data [77]. While adding static data and other terrain feature data does not improve the model's performance, the LSTM model can still produce satisfactory results using only meteorological forcing data. However, a potential issue with LSTM is its structural design and hyperparameterization, as the choice of model structure and hyperparameters can significantly affect its performance [84]. Preliminary experiments explored different hyperparameter settings for the LSTM model. Despite limited training data, appropriate hyperparameter settings can effectively optimize the pre-trained model.

Table 2 shows the information of the test and optimal hyperparameters of the T-LSTM model under different experimental designs.

Table 2. Hyperparameters of LSTM under different experimental designs.

Model		Batch Size	Window Size	Hidden Layer Units	Gradient Descent Method	Learning Rate
Pretraining	Tested	2, 4, 8, 16, 32, 64, 128, 256, 512	10, 20, 30, 40, 50, 60, 70, 80, 90, 120, 150, 180	10, 20, 30, 40, 50, 60, 70, 80, 90, 100	RMSprop Adam	0.001, 0.002, 0.003, 0.004, 0.005, 0.008, 0.01, 0.015, 0.02, 0.03, 0.05
	Selected	32	30	40	RMSprop	0.002
G-LSTM	Tested	32, 64	30, 60, 90	40, 60, 80	RMSprop Adam	0.001, 0.002, 0.003
B-LSTM	Selected	32	30	40	RMSprop	0.002
T-LSTM	Tested	32, 64	30, 60, 90	40, 60, 80	RMSprop Adam	0.001, 0.002, 0.003
	Selected	64	60	60	Adam	0.003

Note(s): Bold font is the optimal parameter for the model.

It can be observed from Table 2 that under different experimental settings, the hyperparameters of the LSTM vary. For B-LSTM and G-LSTM, the optimal values for batch size, window size, number of hidden layer units, gradient descent method and learning rate are 32, 30, 40, RMSprop, and 0.002, respectively. For T-LSTM, the optimal values for these hyperparameters are 64, 60, 60, Adam, and 0.003. The B-LSTM and G-LSTM models have a training period of 1096 days, a validation period of 365 days, and a testing period of 730 days. The T-LSTM model has a total training data length of 3288 days, a validation data length of 1095 days, and a testing period sequence length of 4381 days; the length of the training data significantly influences the values of the hyperparameters. Increasing the number of units in hidden layers significantly affects the training speed of the model, while increasing the batch size prolongs the learning time as each step processes more samples. Nevertheless, it reduces the time for each training iteration. The choice of gradient descent method influences the convergence speed of the loss function during model training and affects the training outcomes [78]. Through manual tuning, we found that the RMSprop learning algorithm can train small sample data to optimal conditions faster. In contrast, the Adam algorithm provides more stable simulation results when dealing with larger datasets. Neither of these learning algorithms is used in conjunction with Dropout because preliminary testing suggests that using Dropout would impair the network's performance [85].

4.2. Performance of BTOP Model

The performance metrics of the BTOP model, as shown in Figure 4, indicate that while the model performed well in the Lhasa River Basin (LRB), its accuracy was lower in the Nyang River Basin (NRB).

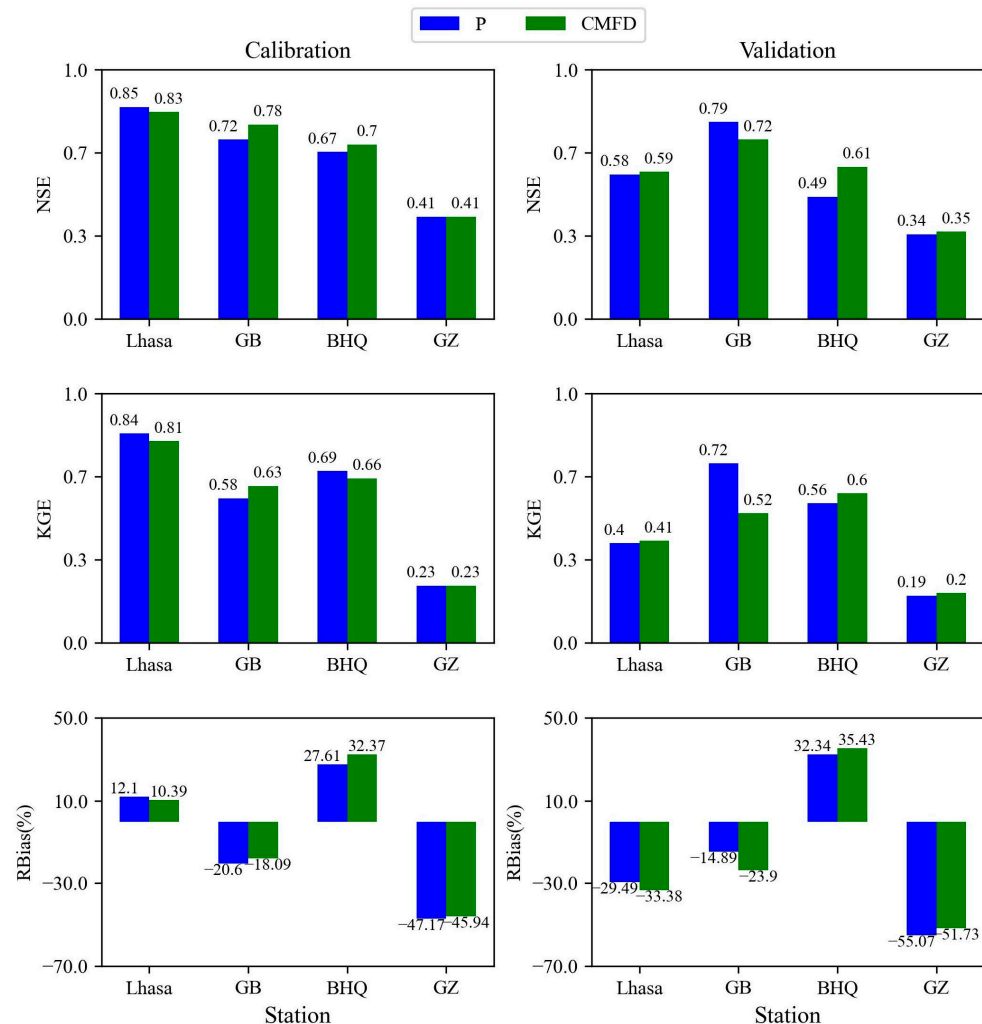


Figure 4. The runoff performance of BTOP-simulated results during calibration and validation at four gauges with two precipitation datasets. (P is for CMA dataset precipitation, CMFD is for CMFD dataset precipitation).

At Lhasa station, the NSE reached 0.85 in the calibration period and 0.58 during the validation period. The KGE was 0.84 in the calibration period but dropped to 0.4 in the validation period. The RBias was 12.1% in the calibration period and -29.49% in the validation period. Notably, the performance of the CMFD dataset was almost identical to that of the precipitation data from meteorological stations.

The results varied significantly among the three stations in NRB. At the GB station, the model achieved an NSE greater than 0.72 in both the calibration and validation periods. The KGE and RBias showed similar performance. At BHQ station, the CMFD dataset outperforms the station precipitation data in calibration and validation periods. This advantage was most pronounced during the validation period compared to all other stations. For other metrics, the results of the two datasets were nearly the same. At the GZ station, the NSE, KGE, and RBias of the CMA dataset in the validation period were only 0.34, 0.19 and -55.07% , which were the worst among all sites.

The BTOP model demonstrates considerable variability in the simulation results across different basins. The difference in the runoff magnitude of the GB, BHQ, GZ stations poses a

challenge for the BTOP. The flow at BHQ is overestimated, and at GZ, underestimated to an unacceptable level. In Figure 5a, we can find that during the second peak of the validation period, neither the station precipitation nor the CMFD dataset can have similar simulated results. This suggests that the model's ability to accurately simulate runoff during this period may be limited by the input data's accuracy or the model's sensitivity to specific meteorological conditions. Figure 5b demonstrates a more favorable performance at the GB station, where the simulated runoff closely mirrors the observed discharge. However, at the BHQ station, the predicted flow exceeds the actual observations, particularly during the dry season and around peak runoff periods. Conversely, the simulated discharge at the GZ station is conspicuously lower than the observed flow, particularly during the wet season.

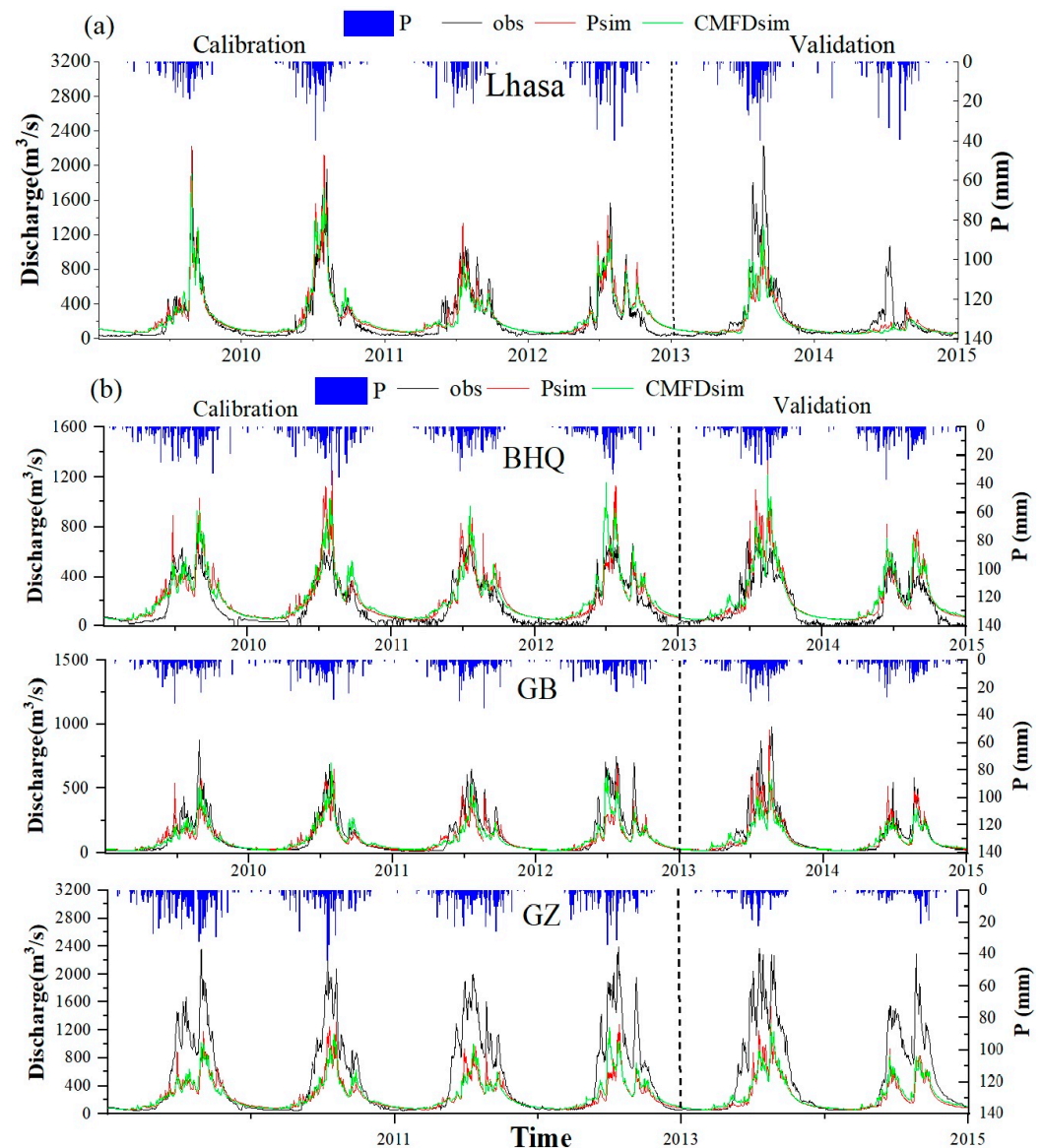


Figure 5. Observed and simulated discharge of BTOP in LRB and NRB during calibration and validation with two precipitation datasets. (a) for Lhasa River Basin, (b) for Nyang River Basin. (P is for CMA dataset precipitation, CMFD is for CMFD dataset precipitation).

In Figure 5b, the CMA precipitation dataset simulated a small flood peak before the year's maximum flood peak in the 2010 runoff simulation, but this situation did not appear in the CMFD dataset simulation. Although the BTOP model uses interpolation methods to distribute station precipitation data across the entire watershed, the sparse number of stations means that the interpolation results do not accurately represent the spatial

precipitation conditions of the watershed. This is especially problematic when a single station controls a large precipitation area; an increase in precipitation at a single station can potentially lead to an increase in flood volume for the entire watershed, resulting in incorrect flood risk assessment [58]. In contrast, the CMFD dataset does not show abnormal flood peaks during the simulation process, indicating that the CMFD dataset contains more useful spatial information and can be used to simulate daily flow in watersheds with limited data [86].

As can be seen from Figures 4 and 5b, the BTOP model exhibits overestimation at BHQ station, while the calculation results revealed a substantial underestimation in the BTOP simulation results at GZ stations, where the underestimation amounted to approximately 50%. This underscores the inadequacy of the BTOP model in accurately capturing the specific hydrological conditions of GZ station within the NRB. It implies that while the model parameters may be suitable for GB and BHQ stations, they are not effectively applicable to GZ, highlighting the need for tailored adjustments to better represent the unique characteristics of this station.

The BTOP model undergoes parameter calibration on a block-by-block basis. Once the parameters of the upstream blocks in the watershed are determined, the downstream blocks can only adjust parameters within the range between upstream and downstream stations. Consequently, when there are significant differences in the flow processes and magnitudes across these stations, the upstream parameter settings limit the adjustable range of simulated flows for the downstream blocks, thereby affecting the simulation performance at the downstream stations. This limitation is particularly evident when the area between GZ and BHQ accounts for a relatively small proportion of the total basin but the differences in flow magnitudes are substantial.

4.3. Performance of B-LSTM and G-LSTM Models

In LRB, the B-LSTM model was constructed based on the only station—Lhasa station—while in the NRB watershed, both the B-LSTM and G-LSTM models were constructed.

At the GZ station, the NSE coefficient increased from 0.34 with the BTOP model to 0.95 with the B-LSTM model, which is highly significant for accurate simulation of peak flows. The KGE coefficient increased to more than 0.86 and the RBias decreased to -0.76% . At the BHQ station, the B-LSTM increased the NSE and KGE to 0.78 and 0.87, respectively, and reduced the RBias to 3.26% . The three indicators of GB station were also improved. As can also be seen from the scatter plot of Figure 6, the scatter values simulated by the B-LSTM are closer to the $y = x$ line than those simulated by the BTOP model, especially at the GZ station, where the scatter values are almost evenly distributed around the $y = x$ line. Figure 6 further underscores the B-LSTM model's capability to accurately fit observed streamflow in NRB, and there is no overestimation at the BHQ station and obvious underestimation at the GZ station as in the BTOP model. Although the water volume of the GZ station is about three times that of BHQ station, the LSTM model can effectively capture this feature. By solely utilizing areal average rainfall as input, it proficiently differentiates runoff conditions, enabling runoff simulations that closely align with actual observations.

The hydrographs in Figure 6a show that the runoff process in 2014 was relatively well-matched, whereas significant discrepancies emerged in 2015. Consequently, we sought to explore whether limiting the testing period exclusively to 2014 would lead to an improvement in model metrics. When analyzing the performance metrics specifically for the year 2014 as the sole testing period, the BTOP model registered NSE, KGE, and RBias values of 0.62, 0.41, and -28.58% , respectively, whereas the B-LSTM model achieved 0.76, 0.80, and 0.31% , respectively. Notably, when compared to the two-year testing period results, the NSE and KGE of the B-LSTM model showed significant improvement, indicating that the relationship between precipitation and runoff in 2015 deviated significantly from previous years, thereby hindering the model's ability to achieve optimal efficiency during the validation (testing) period.

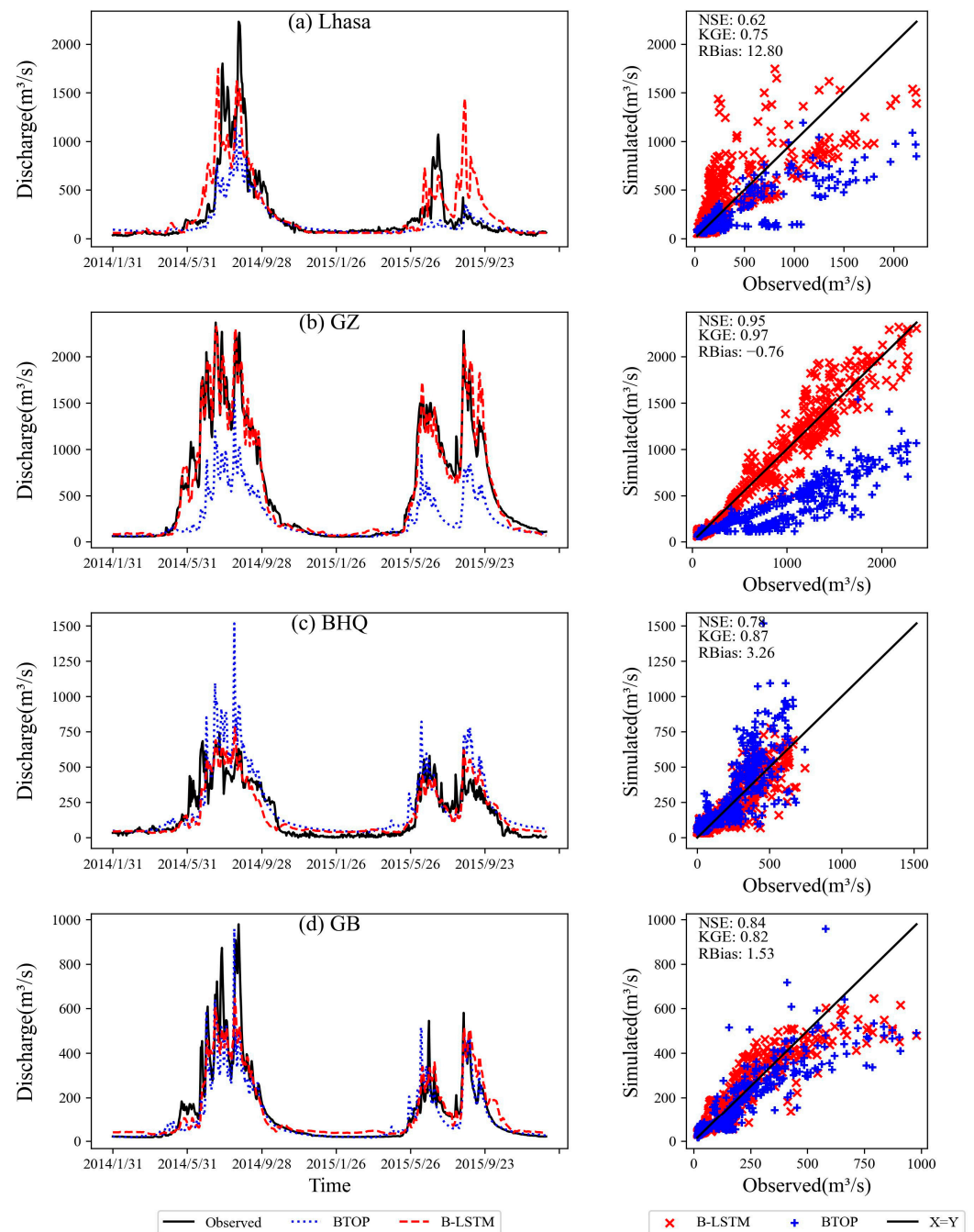


Figure 6. Comparison of observed and simulated streamflow of B-LSTM for each gauge in the testing period as time series and scatter, respectively. (a) for Lhasa station; (b) for GZ station; (c) for BHQ station; (d) for GB station.

Figure 7 depicts the comparison between observed and simulated runoff by G-LSTM during the testing period in NRB. The figure reveals that the trend of the simulated hydrograph aligns well with that of the observed, particularly at the GZ station where the simulated and observed values are nearly identical. However, the model exhibits deviations in the peak runoff regions, with the smallest deviation occurring at the GZ station, followed by BHQ, and the largest deviation observed at the GB station. Despite the RBias value being close to 0 for GB, the scatter plot indicates that the scatter points exhibit greater deviation under high-flow conditions compared to low-flow conditions.

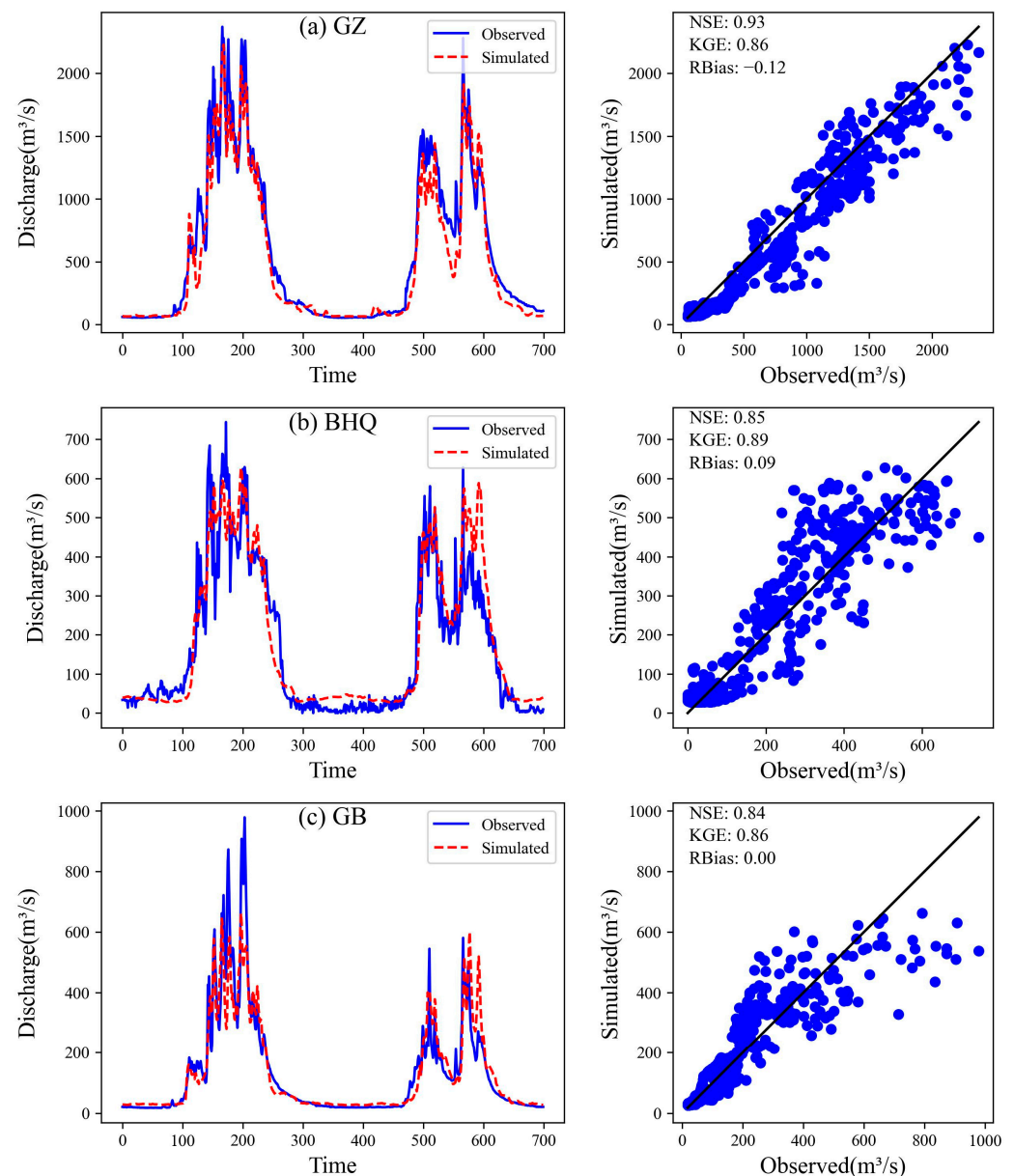


Figure 7. Comparison of observed and simulated streamflow of G-LSTM for gauges in NRB in the testing period as time series and scatter, respectively. (a) for GZ station; (b) for BHQ station; (c) for GB station.

Table 3 shows the evaluation criteria values of B-LSTM and G-LSTM in the test period. It is evident that compared to the BTOP model, the G-LSTM and B-LSTM models significantly improved the NSE and KGE values and reduced the RBias, especially at the GB station, where the RBias of G-LSTM was generally reduced to 0. The G-LSTM model is trained separately based on data from each station, resulting in a more targeted network structure, which explains why the simulation results outperform those of B-LSTM to a certain extent, particularly at the GB and BHQ stations. Nevertheless, at the GZ station, the B-LSTM model yields slightly better results than G-LSTM, suggesting that the inclusion of inputs from other stations enhances the runoff simulation performance at the GZ station.

Table 3. Evaluation criteria of different models.

Station	Model	Testing Period	NSE	KGE	RBias (%)
Lhasa	BTOP	2014–2015	0.58	0.4	−29.49
	BTOP	2014	0.62	0.41	−28.58
	B-LSTM	2014	0.76	0.80	0.31
	B-LSTM	2014–2015	0.62	0.75	12.8
GB	BTOP	2014–2015	0.79	0.72	−14.89
	B-LSTM	2014–2015	0.84	0.82	1.53
	G-LSTM	2014–2015	0.84	0.86	0
BHQ	BTOP	2014–2015	0.49	0.56	32.34
	B-LSTM	2014–2015	0.78	0.87	3.26
	G-LSTM	2014–2015	0.85	0.89	0.09
GZ	BTOP	2014–2015	0.34	0.19	−55.07
	B-LSTM	2014–2015	0.95	0.97	−0.76
	G-LSTM	2014–2015	0.93	0.86	−0.12

The significant improvement in runoff simulation performance of LSTM shows that the structure of LSTM is well suited to converting meteorological data into runoff. When driven solely by meteorological observation data, the LSTM model has better runoff simulations for NRB and LRB than those produced by a physically based hydrological model. By changing the modeling objective from watershed to gauge, the LSTM model still surpasses the hydrological model. Previous studies have also reported improvements in simulation performance with LSTM [35,39], and some studies have compared the performance of localized models to regional models [35,87]. The results obtained are similar to our findings in the data-scarce regions of QTP, indicating that LSTM significantly outperforms traditional hydrological models in runoff simulation for alpine, data-scarce regions [36,71].

4.4. Performance of T-LSTM Model

For the detailed experiment schemes of four transfer cases of T-LSTM, please refer to Table 1 and Section 3.3.3. The T-LSTM model was trained and validated using data from 10–15 years at the source station, and then the model was transferred to simulate the runoff at the target station for the same period. Therefore, the simulated runoff during the testing period covers 10–15 years. The hydrograph of the testing period of the four cases is shown in Figure 8.

From Figure 8, in Case ① with Lhasa station as the target, it can be observed that the red dashed line (simulated values) and the blue solid line (observed values) show significant differences at multiple peaks and troughs. The T-LSTM model's simulation performance at the Lhasa station is poor, failing to accurately capture the fluctuations and trends of the observed flow. In Case ②–Case ④, the simulation results at the GZ, BHQ, and GB stations are relatively good, especially at the BHQ and GB stations where the model's performance is particularly outstanding. The NSE and KGE values are high, and the RBias values are close to 0, indicating that the model accurately captures the fluctuations in and trends of the observed flow.

A scatter plot comparison of the results of T-LSTM and BTOP model is shown in Figure 9. From the results of the four experimental cases of T-LSTM, it can be seen that at both GZ and GB stations, T-LSTM improves the simulation runoff of high discharge, and the fitting lines are closer to the line $y = x$. At the BHQ station, BTOP has a slight overestimation of low discharge, which is corrected by T-LSTM. In Lhasa station, the overestimation of T-LSTM keeps the fitting line away from the line $y = x$.

Figure 10 shows the evaluation criteria values of BTOP, B-LSTM, G-LSTM, and T-LSTM in the testing period at four gauges.

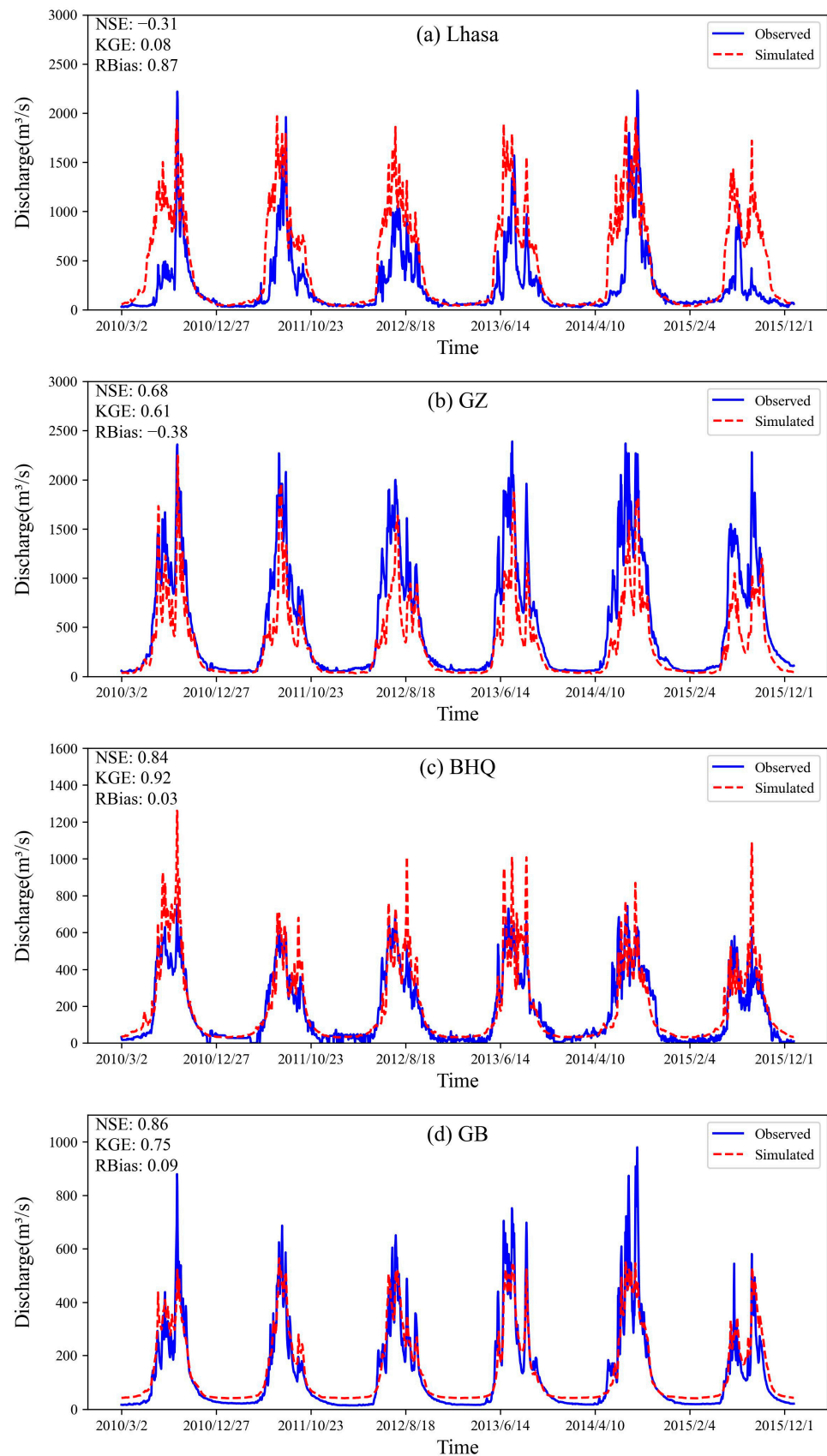


Figure 8. Hydrographs of the training and simulated streamflow for T-LSTM: (a) Case ①; (b) Case ②; (c) Case ③; (d) Case ④.

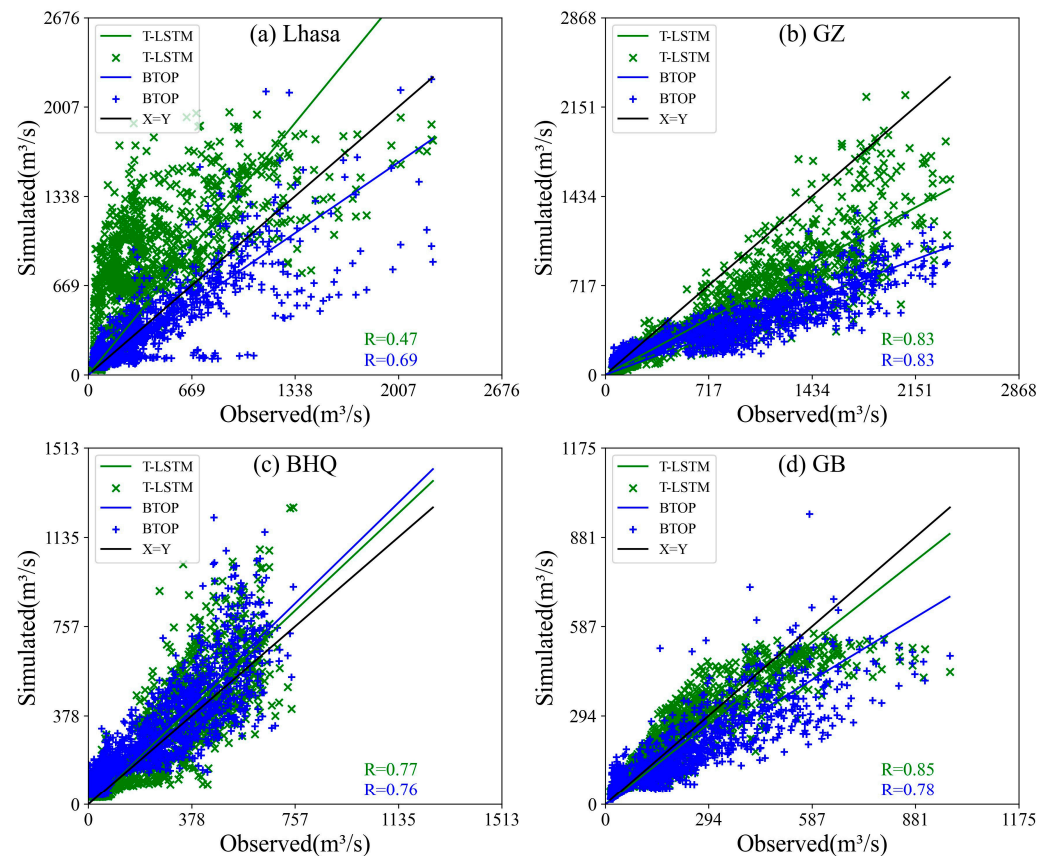


Figure 9. Scatter plot comparison of the results of the T-LSTM and the BTOP model at different stations. (a) for Lhasa station; (b) for GZ station; (c) for BHQ station; (d) for GB station.

As can be seen from Figure 10, the NSE of T-LSTM is only -0.31 , KGE is only 0.08 , and RBias is 0.87% , respectively, at Lhasa station. These three indicators were lower than the results of the BTOP and G-LSTM models. The schemes transferring to BHQ station (Case ③) and GB station (Case ④) yield better results than the BTOP model. Specifically, at the BHQ station, T-LSTM achieves the same NSE coefficient as G-LSTM, and the KGE is superior to G-LSTM; at the GB station, the NSE of T-LSTM is slightly better than that of the G-LSTM model, but the KGE of T-LSTM is slightly lower. And at these three stations, the RBias of T-LSTM comes the lowest among all methods. The results of the GZ station are quite different. The BTOP model struggled to accurately simulate the runoff at the GZ station, achieving an NSE of only 0.34 . In contrast, the G-LSTM model demonstrated a significant improvement, achieving an NSE of 0.93 . Notably, the best result was provided by the B-LSTM model, with an NSE of 0.95 . Even the T-LSTM model performed admirably, achieving an NSE of 0.68 , outperforming the BTOP model despite its reliance on extensive data. The trend in the KGE results was consistent with that of the NSE.

In the modeling process of T-LSTM, only meteorological data from the GZ station needed to be fed into the pre-trained T-LSTM model, eliminating the requirement for additional data such as terrain physics and evapotranspiration. Despite this, T-LSTM was able to produce results superior to those of the BTOP model, demonstrating its effectiveness in the NRB watershed. This improvement in runoff simulation for data-scarce basins is crucial for water management [88]. Even in the absence of observed data, the transferability of the LSTM model surpasses that of the calibrated PBMs. The high predictive accuracy of LSTM in data-scarce basins further demonstrates that machine learning can be a powerful and practical tool for runoff analysis and flood management [4].

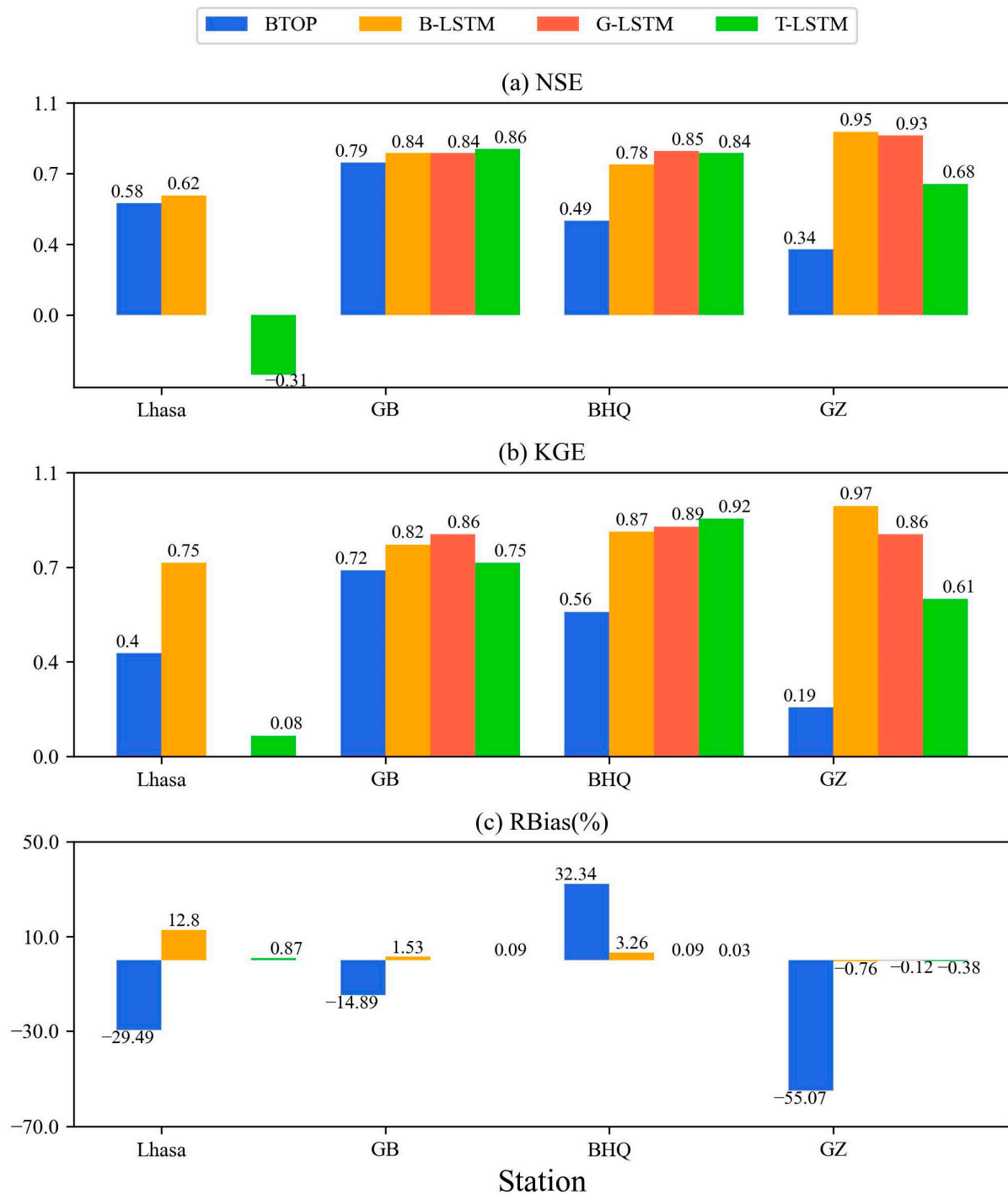


Figure 10. Evaluation criteria values of BTOP, B-LSTM, G-LSTM, and T-LSTM in the testing period at four gauges. (a) NSE, (b) KGE, (c) RBias.

4.5. Discussion about GZ Station

GZ station, as the downstream control station of the NRB, has simulation results that are crucial for flood management across the entire basin. In T-LSTM Case ②, the model achieved results significantly better than the physically based BTOP model without learning the watershed runoff and convergence process specific to the controlled area of the GZ station. To explore the reasons behind this, we added two comparative schemes:

- Scheme a: The model was trained only with data from GB and BHQ stations and then transferred to the GZ station (Case ②-a).

- Scheme b: The model was trained only with data from Lhasa station and then transferred to the GZ station (Case ②-b).

The hydrograph as depicted in Figure 11 shows that for Case ②-a, the NSE coefficient of T-LSTM was only 0.36, which is a little lower than the BTOP model. For Case ②-b, the NSE coefficient was 0.62. Case ②-b performed better than Case ②-a. This means that although the GB, BHQ and GZ stations are in the same watershed, there are still differences in the mechanisms of runoff and confluence within the watershed areas controlled by each station. This is also the reason why the BTOP model struggles to achieve satisfactory simulation results at GZ station.

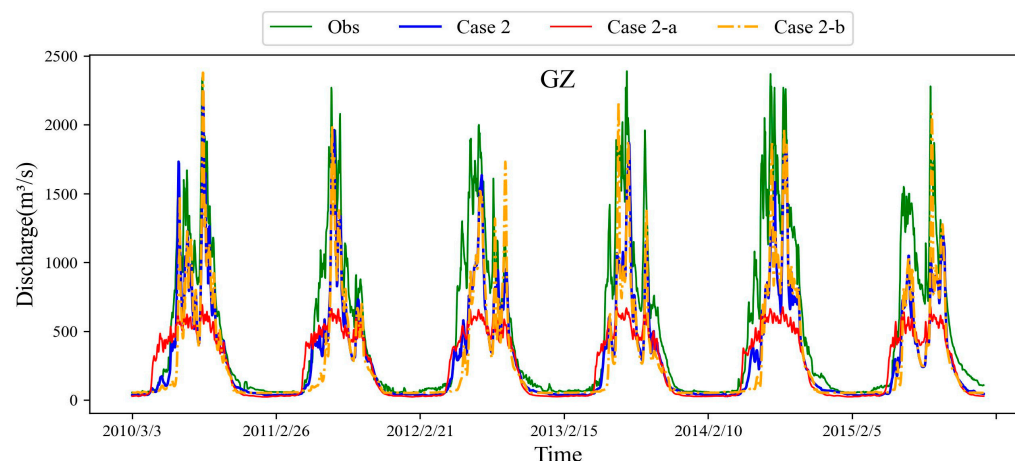


Figure 11. Simulated streamflow at the GZ station in Case ②, Case ②-a, Case ②-b.

Figure 11 shows that the inclusion of data from the Lhasa station provided the LSTM model with an opportunity to learn from larger peak flows, thereby improving its simulation of peak flows. However, for medium and low flows, the model primarily learns from the various stations within the NRB. Thus, when data from Lhasa, GB, and BHQ stations are all used as training inputs, the T-LSTM simulating performance at GZ station significantly improves. These results indicate that the T-LSTM model can achieve a significant performance even with limited training data. However, the consistency and representativeness of the training data have a substantial impact on the modeling [89].

LRB and NRB both exhibit distinct seasonal patterns of precipitation variation, which are clearly illustrated in Figure 12a. In NRB, monthly precipitation steadily rises from March onwards, peaking in July, reflecting a gradual accumulation of rainfall from spring to summer. In contrast, precipitation in LRB experiences a relatively moderate increase between April and May, followed by a sharp surge to its annual maximum in June. However, it is noteworthy that despite similar precipitation levels in certain months, the runoff response varies significantly among different stations, such as GZ station and Lhasa station. As depicted in Figure 12b, under similar average precipitation, the runoff volume at the GZ station increases markedly, with its peak flow rate nearly double that of Lhasa station and almost triple that of both GZ and GB station. This phenomenon underscores the critical importance of accurately capturing such peak variations for a profound understanding and reflection of actual hydrological conditions within the basin. Given the direct causality between seasonal precipitation changes and runoff variations, models employed in simulations must meticulously depict the intricate effects of concentrated precipitation inputs on the runoff process. These models must not only be capable of handling seasonal fluctuations in precipitation volumes but also possess the ability to capture changes in key factors such as precipitation intensity and duration, thereby enabling accurate predictions of dynamic runoff variations. Furthermore, differences in precipitation distribution can significantly impact the setting of model parameters. As precipitation characteristics (e.g., concentration and intensity) vary across stations or basins, it may be necessary to adjust model parameters accordingly when applying the same model for simulations. This

explains the notable differences in simulation outcomes observed when using the BTOP model for different stations within NRB. Consequently, in the process of LSTM modeling, the model fully considers the seasonal distribution characteristics of precipitation and its influence on model parameters, which is the key to achieving accurate simulation of seasonal runoff changes.

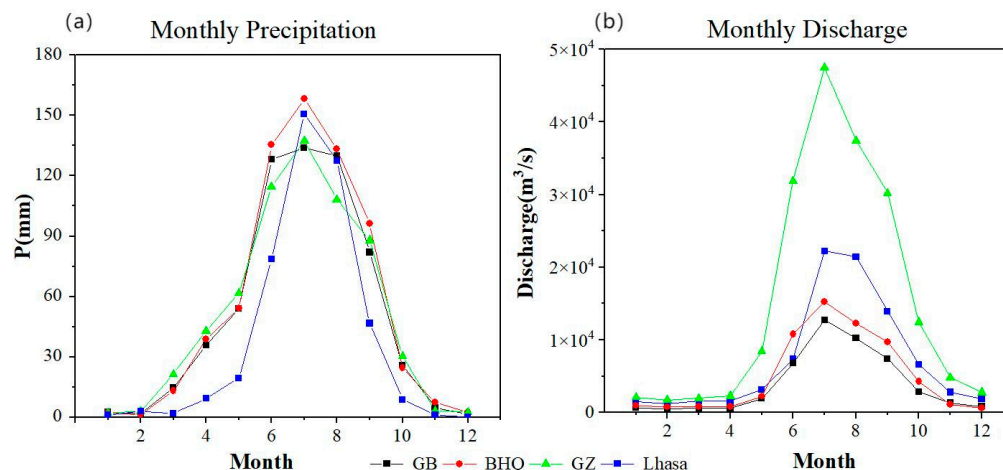


Figure 12. Distribution characteristics of monthly precipitation and discharge in LRB and NRB.

When the transfer model lacks the characteristic information of the LRB, the T-LSTM model exhibits suboptimal performance. As seen in Case ① of T-LSTM, the model is solely reliant on runoff characteristics learned from NRB input data, resulting in unsatisfactory results. However, when the training data incorporate information from additional gauges within the data-scarce basin, even a small convergence area (such as GB station, which covers only about 50% of the NRB), a significantly enhancement in the transfer results can be achieved. Therefore, when transferring the LSTM model to a basin with no observed data, it is imperative to meticulously analyze the characteristics of the basin and consider both spatial proximity and physical similarity. By selecting the optimal data sources for model training, an improved performance can be achieved within a shorter timeframe [88].

4.6. Limitations and Future Research Directions

While LSTM has shown tremendous potential in simulating runoff in data-scarce alpine regions, it cannot solve all the hydrological problems. This study has several limitations. First, climate conditions in the QTP are highly variable, and the influence of glaciers and permafrost on regional runoff is complex. Although the BTOP model used in this study accounts for snowmelt, it has not been improved to simulate glacier and permafrost dynamics. Second, the implementation of TL techniques can be an effective method for simulating runoff in data-scarce regions [31]. However, in the design of the model transfer scheme in this study, the LSTM model was not improved before the transfer, and the model results were not combined with those of physical models. In the future, developing hybrid models based on both physics and DL may be a reliable solution to improve the performance of TL models. In addition, due to the limitations of observed data, this study did not conduct further analysis on the runoff characteristics of the watershed, and the quality and availability of the dataset are also crucial factors influencing the performance of DL models [90]. In the next step of our work, we plan to utilize globally available large datasets to obtain more extended time series of runoff data, enabling a more in-depth and comprehensive study of data-scarce regions in the Qinghai–Tibet Plateau.

5. Conclusions

The scarcity of observations and stations in the Qinghai–Tibet Plateau presents a formidable obstacle to precise runoff simulation. To tackle this challenge and evaluate the feasibility of using LSTM for runoff modeling in data-scarce alpine regions, we conducted

a comprehensive comparison between the BTOP and LSTM, leveraging the CMFD dataset and CMA data from the Lhasa and Nyang River Basin within the QTP. Additionally, we integrated TL with LSTM to create T-LSTM, and migrated the T-LSTM model trained at multiple sites to data-scarce sites to evaluate its simulation accuracy at data-scarce sites.

Before comparing the models, we first emphasize the utility of the CMFD dataset, which offers richer spatial precipitation information than traditional meteorological data. This enhanced information minimizes model uncertainties due to sparse station data, making it suitable for simulating runoff in data-scarce regions of QTP. When it comes to model performance, the average values of NSE, KGE, and RBias for B-LSTM at each station are 0.80, 0.85, and 4.21%, respectively. The corresponding values for G-LSTM are 0.81, 0.84, and 3.19%. The LSTM model outperforms the physically based BTOP model when relying solely meteorological inputs and it improves NSE, KGE, and RBias by 0.23, 0.36, and -18.36% , respectively. LSTM effectively simulates runoff for both entire basins and single stations and performs well in data-scarce regions of the QTP. Furthermore, the integration of Transfer Learning (TL) with LSTM, resulting in T-LSTM, enables the transfer of knowledge from data-rich source domains to data-scarce target domains. This allows T-LSTM to accurately simulate runoff variations even in the absence of direct runoff observations. When trained with precipitation–runoff relationships from parts of the target watershed, T-LSTM outperforms hydrological models specifically calibrated for those basins, highlighting its versatility. Notably, even with limited training data, selecting appropriate data sources can significantly boost the model's performance.

In summary, LSTM models have firmly demonstrated their value in runoff simulation, particularly in the context of the QTP's data-scarce alpine basins. Their ability to simulate runoff using solely precipitation and temperature data sets them apart from physically based distributed hydrological models, which often struggle in such data-limited settings. Furthermore, the efficacy of LSTM in transfer learning underscores its potential as a versatile tool for enhancing hydrological forecasting and water resources management not only in the QTP but also in other regions.

Author Contributions: Conceptualization, L.Z., J.Y. and J.D.; methodology, J.Y. and L.Z.; software, S.N. and J.Y.; validation, J.Y. and L.Z.; formal analysis, J.Y. and C.Z.; data curation, L.W.; writing—original draft preparation, J.Y.; writing—review and editing, L.Z., J.D. and T.A.; funding acquisition, L.Z. and T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Project from the Science and Technology Department of Tibet (XZ202101ZY0007G); The Youth Project of Natural Science Foundation from Science and Technology Department of Sichuan Province (2024NSFSC0984); and the Postdoctoral Interdisciplinary Innovation Fund (JCXK2239) from Sichuan University.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors acknowledge the anonymous reviewers for their comments and suggestions for this manuscript.

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

References

1. Niu, W.; Feng, Z. Evaluating the Performances of Several Artificial Intelligence Methods in Forecasting Daily Streamflow Time Series for Sustainable Water Resources Management. *Sustain. Cities Soc.* **2021**, *64*, 102562. [\[CrossRef\]](#)
2. Wang, Y.; Ye, A.; Zhang, Y.; Yang, F. The Quantitative Attribution of Climate Change to Runoff Increase over the Qinghai-Tibetan Plateau. *Sci. Total Environ.* **2023**, *897*, 165326. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Nevo, S.; Morin, E.; Gerzi Rosenthal, A.; Metzger, A.; Barshai, C.; Weitzner, D.; Voloshin, D.; Kratzert, F.; Elidan, G.; Dror, G.; et al. Flood Forecasting with Machine Learning Models in an Operational Framework. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 4013–4032. [\[CrossRef\]](#)
4. Ahmed, A.A.; Sayed, S.; Abdoulhalik, A.; Moutari, S.; Oyedele, L. Applications of Machine Learning to Water Resources Management: A Review of Present Status and Future Opportunities. *J. Clean. Prod.* **2024**, *441*, 140715. [\[CrossRef\]](#)
5. He, Y.; Bárdossy, A.; Zehe, E. A Review of Regionalisation for Continuous Streamflow Simulation. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3539–3553. [\[CrossRef\]](#)

6. Liu, C.; Shan, Y.; He, L.; Li, F.; Liu, X.; Nepf, H. Plant Morphology Impacts Bedload Sediment Transport. *Geophys. Res. Lett.* **2024**, *51*, e2024GL108800. [[CrossRef](#)]
7. Shan, Y.; Yan, C.; Liu, J.; Liu, C. Predicting Velocity and Turbulent Kinetic Energy inside an Emergent Phragmites Australis Canopy with Real Morphology. *Environ. Fluid Mech.* **2023**, *23*, 943–963. [[CrossRef](#)]
8. Immerzeel, W.W.; Lutz, A.F.; Andrade, M.; Bahl, A.; Biemans, H.; Bolch, T.; Hyde, S.; Brumby, S.; Davies, B.J.; Elmore, A.C.; et al. Importance and Vulnerability of the World's Water Towers. *Nature* **2020**, *577*, 364–369. [[CrossRef](#)]
9. Yao, Y.; Khan, Z.A. Predicting Pakistan's next. *Science* **2022**, *378*, 483. [[CrossRef](#)]
10. Zhang, Z.; Zhang, L.; Liu, Y.; Jin, M. Responses of Annual Streamflow Variability to Annual Precipitation, Extreme Climate Events and Large-Scale Climate Phenomena in the Qinghai-Tibet Plateau. *J. Hydrol.* **2024**, *632*, 130969. [[CrossRef](#)]
11. Yao, Y.; Zhao, Y.; Li, X.; Feng, D.; Shen, C.; Liu, C.; Kuang, X.; Zheng, C. Can Transfer Learning Improve Hydrological Predictions in the Alpine Regions? *J. Hydrol.* **2023**, *625*, 130038. [[CrossRef](#)]
12. Yokoo, K.; Ishida, K.; Ercan, A.; Tu, T.; Nagasato, T.; Kiyama, M.; Amagasaki, M. Capabilities of Deep Learning Models on Learning Physical Relationships: Case of Rainfall-Runoff Modeling with LSTM. *Sci. Total Environ.* **2022**, *802*, 149876. [[CrossRef](#)] [[PubMed](#)]
13. Wei, X.; Wang, G.; Schmalz, B.; Hagan, D.F.T.; Duan, Z. Evaluation of Transformer Model and Self-Attention Mechanism in the Yangtze River Basin Runoff Prediction. *J. Hydrol. Reg. Stud.* **2023**, *47*, 101438. [[CrossRef](#)]
14. van Emmerik, T.; Mulder, G.; Eilander, D.; Piet, M.; Savenije, H. Predicting the Ungauged Basin: Model Validation and Realism Assessment. *Front. Earth Sci.* **2015**, *3*, 62. [[CrossRef](#)]
15. Kumar, V.; Sen, S. Hydrometeorological Field Instrumentation in Lesser Himalaya to Advance Research for Future Water and Food Security. *Environ. Monit. Assess.* **2023**, *195*, 1162. [[CrossRef](#)] [[PubMed](#)]
16. Kratzert, F.; Nearing, G.; Addor, N.; Erickson, T.; Gauch, M.; Gilon, O.; Gudmundsson, L.; Hassidim, A.; Klotz, D.; Nevo, S.; et al. Caravan—A Global Community Dataset for Large-Sample Hydrology. *Sci. Data* **2023**, *10*, 61. [[CrossRef](#)] [[PubMed](#)]
17. Deng, C.; Chen, C.; Yin, X.; Wang, W.; Zhang, Y. Catchment Runoff Simulation by Coupling Data Assimilation and Machine Learning Methods. *Adv. Water Sci.* **2023**, *34*, 839–849. (In Chinese) [[CrossRef](#)]
18. Sivapalan, M.; Takeuchi, K.; Franks, S.W.; Gupta, V.K.; Karambiri, H.; Lakshmi, V.; Liang, X.; McDONNELL, J.J.; Mendiola, E.M.; O'Connell, P.E.; et al. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an Exciting Future for the Hydrological Sciences. *Hydrol. Sci. J.* **2003**, *48*, 857–880. [[CrossRef](#)]
19. Kratzert, F.; Klotz, D.; Herrnegger, M.; Sampson, A.K.; Hochreiter, S.; Nearing, G.S. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. *Water Resour. Res.* **2019**, *55*, 11344–11354. [[CrossRef](#)]
20. Pool, S.; Vis, M.; Seibert, J. Regionalization for Ungauged Catchments—Lessons Learned From a Comparative Large-Sample Study. *Water Resour. Res.* **2021**, *57*, e2021WR030437. [[CrossRef](#)]
21. Du, T.L.T.; Lee, H.; Bui, D.D.; Arheimer, B.; Li, H.-Y.; Olsson, J.; Darby, S.E.; Sheffield, J.; Kim, D.; Hwang, E. Streamflow Prediction in “Geopolitically Ungauged” Basins Using Satellite Observations and Regionalization at Subcontinental Scale. *J. Hydrol.* **2020**, *588*, 125016. [[CrossRef](#)]
22. Hrachowitz, M.; Savenije, H.H.G.; Blöschl, G.; McDonnell, J.J.; Sivapalan, M.; Pomeroy, J.W.; Arheimer, B.; Blume, T.; Clark, M.P.; Ehret, U.; et al. A Decade of Predictions in Ungauged Basins (PUB)—A Review. *Hydrol. Sci. J.* **2013**, *58*, 1198–1255. [[CrossRef](#)]
23. Samaniego, L.; Kumar, R.; Thober, S.; Rakovec, O.; Zink, M.; Wanders, N.; Eisner, S.; Müller Schmied, H.; Sutanudjaja, E.H.; Warrach-Sagi, K.; et al. Toward Seamless Hydrologic Predictions across Spatial Scales. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 4323–4346. [[CrossRef](#)]
24. Samaniego, L.; Kumar, R.; Attinger, S. Multiscale Parameter Regionalization of a Grid-based Hydrologic Model at the Mesoscale. *Water Resour. Res.* **2010**, *46*, 2008WR007327. [[CrossRef](#)]
25. Sun, R.; Pan, B.; Duan, Q. A Surrogate Modeling Method for Distributed Land Surface Hydrological Models Based on Deep Learning. *J. Hydrol.* **2023**, *624*, 129944. [[CrossRef](#)]
26. Williams, B.; Cremaschi, S. Selection of Surrogate Modeling Techniques for Surface Approximation and Surrogate-Based Optimization. *Chem. Eng. Res. Des.* **2021**, *170*, 76–89. [[CrossRef](#)]
27. Pokorny, S.; Stadnyk, T.A.; Ali, G.; Tefs, A.A.G.; Déry, S.J. From Threat to Opportunity: Hydrologic Uncertainty Regionalization across Large Domains. *J. Hydrol. Reg. Stud.* **2024**, *53*, 101819. [[CrossRef](#)]
28. Kumar, V.; Sen, S. Rating Curve Development and Uncertainty Analysis in Mountainous Watersheds for Informed Hydrology and Resource Management. *Front. Water* **2024**, *5*, 1323139. [[CrossRef](#)]
29. Toth, E.; Brath, A.; Montanari, A. Comparison of Short-Term Rainfall Prediction Models for Real-Time Flood Forecasting. *J. Hydrol.* **2000**, *239*, 132–147. [[CrossRef](#)]
30. Nearing, G.S.; Kratzert, F.; Sampson, A.K.; Pelissier, C.S.; Klotz, D.; Frame, J.M.; Prieto, C.; Gupta, H.V. What Role Does Hydrological Science Play in the Age of Machine Learning? *Water Resour. Res.* **2021**, *57*, e2020WR028091. [[CrossRef](#)]
31. Clark, S.R.; Lerat, J.; Perraud, J.-M.; Fitch, P. Deep Learning for Monthly Rainfall–Runoff Modelling: A Large-Sample Comparison with Conceptual Models across Australia. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 1191–1213. [[CrossRef](#)]
32. Feng, D.; Fang, K.; Shen, C. Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales. *Water Resour. Res.* **2020**, *56*, e2019WR026793. [[CrossRef](#)]
33. Xie, J.; Liu, X.; Bai, P.; Liu, C. Rapid Watershed Delineation Using an Automatic Outlet Relocation Algorithm. *Water Resour. Res.* **2022**, *58*, e2021WR031129. [[CrossRef](#)]

34. Xie, J.; Liu, X.; Tian, W.; Wang, K.; Bai, P.; Liu, C. Estimating Gridded Monthly Baseflow From 1981 to 2020 for the Contiguous US Using Long Short-Term Memory (LSTM) Networks. *Water Resour. Res.* **2022**, *58*, e2021WR031663. [\[CrossRef\]](#)
35. 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. [\[CrossRef\]](#)
36. Arsenault, R.; Martel, J.-L.; Brunet, F.; Brissette, F.; Mai, J. Continuous Streamflow Prediction in Ungauged Basins: Long Short-Term Memory Neural Networks Clearly Outperform Traditional Hydrological Models. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 139–157. [\[CrossRef\]](#)
37. Ouyang, W.; Lawson, K.; Feng, D.; Ye, L.; Zhang, C.; Shen, C. Continental-Scale Streamflow Modeling of Basins with Reservoirs: Towards a Coherent Deep-Learning-Based Strategy. *J. Hydrol.* **2021**, *599*, 126455. [\[CrossRef\]](#)
38. Ma, K.; Feng, D.; Lawson, K.; Tsai, W.-P.; Liang, C.; Huang, X.; Sharma, A.; Shen, C. Transferring Hydrologic Data Across Continents—Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions. *Water Resour. Res.* **2021**, *57*, e2020WR028600. [\[CrossRef\]](#)
39. Gauch, M.; Mai, J.; Lin, J. The Proper Care and Feeding of CAMELS: How Limited Training Data Affects Streamflow Prediction. *Environ. Model. Softw.* **2021**, *135*, 104926. [\[CrossRef\]](#)
40. Ghosh, R.; Li, B.; Tayal, K.; Kumar, V.; Jia, X. Meta-Transfer Learning: An Application to Streamflow Modeling in River-Streams. In Proceedings of the 2022 IEEE International Conference on Data Mining (ICDM), Orlando, FL, USA, 28 November–1 December 2022; pp. 161–170.
41. Khoshkalam, Y.; Rousseau, A.N.; Rahmani, F.; Shen, C.; Abbasnezhadi, K. Applying Transfer Learning Techniques to Enhance the Accuracy of Streamflow Prediction Produced by Long Short-Term Memory Networks with Data Integration. *J. Hydrol.* **2023**, *622*, 129682. [\[CrossRef\]](#)
42. Huang, K.; Wang, G.; Song, C.; Yu, Q. Runoff simulation and prediction of a typical small watershed in permafrost region of the Qinghai-Tibet Plateau based on LSTM. *J. Glaciol. Geocryol.* **2021**, *43*, 1144–1156. (In Chinese)
43. Lin, X.; Zhang, Y.; Yao, Z.; Gong, T.; Wang, H.; Chu, D.; Liu, L.; Zhang, F. The Trend on Runoff Variations in the Lhasa River Basin. *J. Geogr. Sci.* **2008**, *18*, 95–106. [\[CrossRef\]](#)
44. Guédé, K.G.; Yu, Z.; Gu, H.; Badji, O.; Ahmed, N.; Sika, B.; Oga, Y.M.S. Sensitivities of Hydrological Processes under Climate Warming and Landuse/Landcover Change in the Lhasa Basin, Tibetan Plateau. *J. Hydrol. Reg. Stud.* **2024**, *52*, 101731. [\[CrossRef\]](#)
45. Zhang, M.; Ren, Q.; Wei, X.; Wang, J.; Yang, X.; Jiang, Z. Climate Change, Glacier Melting and Streamflow in the Niyang River Basin, Southeast Tibet, China. *Ecohydrology* **2011**, *4*, 288–298. [\[CrossRef\]](#)
46. Cuo, L.; Li, N.; Liu, Z.; Ding, J.; Liang, L.; Zhang, Y.; Gong, T. Warming and Human Activities Induced Changes in the Yarlung Tsangpo Basin of the Tibetan Plateau and Their Influences on Streamflow. *J. Hydrol. Reg. Stud.* **2019**, *25*, 100625. [\[CrossRef\]](#)
47. Senent-Aparicio, J.; Blanco-Gómez, P.; López-Ballesteros, A.; Jimeno-Sáez, P.; Pérez-Sánchez, J. Evaluating the Potential of GloFAS-ERA5 River Discharge Reanalysis Data for Calibrating the SWAT Model in the Grande San Miguel River Basin (El Salvador). *Remote Sens.* **2021**, *13*, 3299. [\[CrossRef\]](#)
48. World Climate Research Programme (WCRP); World Meteorological Organization (WMO). *Report of the Workshop on Global Large-Scale Precipitation Data Sets for the World Climate Research Programme*; WCP-111; WMO/TD—No. 94; WMO: Geneva, Switzerland, 1986; p. 45.
49. He, J.; Yang, K.; Tang, W.; Lu, H.; Qin, J.; Chen, Y.; Li, X. The First High-Resolution Meteorological Forcing Dataset for Land Process Studies over China. *Sci. Data* **2020**, *7*, 25. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, RG2004. [\[CrossRef\]](#)
51. Friedl, M.A.; Sulla-Menashe, D. *MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid*; NASA: Washington, DC, USA, 2015.
52. Chesworth, W.; Camps Arbestain, M.; Macías, F.; Spaargaren, O.; Spaargaren, O.; Mualem, Y.; Morel-Seytoux, H.J.; Horwath, W.R.; Almendros, G.; Chesworth, W.; et al. Classification of Soils: FAO. In *Encyclopedia of Soil Science*; Chesworth, W., Ed.; Springer: Dordrecht, The Netherlands, 2008; pp. 111–113. ISBN 978-1-4020-3995-9.
53. Liu, L.; Zhou, L.; Gusyev, M.; Ren, Y. Unravelling and Improving the Potential of Global Discharge Reanalysis Dataset in Streamflow Estimation in Ungauged Basins. *J. Clean. Prod.* **2023**, *419*, 138282. [\[CrossRef\]](#)
54. Ao, T.; Takeuchi, K.; Ishidaira, H.; Yoshitani, J.; Fukami, K. Development and Application of a New Algorithm for Automated Pit Removal for Grid DEMs. *Hydrol. Sci. J.* **2003**, *48*, 985–997.
55. Takeuchi, K.; Hapuarachchi, P.; Zhou, M.; Ishidaira, H.; Magome, J. A BTOP Model to Extend TOPMODEL for Distributed Hydrological Simulation of Large Basins. *Hydrol. Process.* **2008**, *22*, 3236–3251. [\[CrossRef\]](#)
56. Zhou, L.; Koike, T.; Takeuchi, K.; Rasmy, M.; Onuma, K.; Ito, H.; Selvarajah, H.; Liu, L.; Li, X.; Ao, T. A Study on Availability of Ground Observations and Its Impacts on Bias Correction of Satellite Precipitation Products and Hydrologic Simulation Efficiency. *J. Hydrol.* **2022**, *610*, 127595. [\[CrossRef\]](#)
57. Zhu, Y.; Liu, L.; Qin, F.; Zhou, L.; Zhang, X.; Chen, T.; Li, X.; Ao, T. Application of the Regression-Augmented Regionalization Approach for BTOP Model in Ungauged Basins. *Water* **2021**, *13*, 2294. [\[CrossRef\]](#)
58. Zhou, L.; Rasmy, M.; Takeuchi, K.; Koike, T.; Selvarajah, H.; Ao, T. Adequacy of Near Real-Time Satellite Precipitation Products in Driving Flood Discharge Simulation in the Fuji River Basin, Japan. *Appl. Sci.* **2021**, *11*, 1087. [\[CrossRef\]](#)

59. Zhang, H.; Ao, T.; Gusyev, M.; Ishidaira, H.; Magome, J.; Takeuchi, K. Distributed Source Pollutant Transport Module Based on BTOPMC: A Case Study of the Laixi River Basin in the Sichuan Province of Southwest China. *Proc. IAHS* **2018**, *379*, 323–333. [\[CrossRef\]](#)
60. Nimai, S.; Ren, Y.; Ao, T.; Zhou, L.; Liang, H.; Cui, Y. Enhancing Runoff Simulation Using BTOP-LSTM Hybrid Model in the Shinano River Basin. *Water* **2023**, *15*, 3758. [\[CrossRef\]](#)
61. Xiao, Q.; Zhou, L.; Xiang, X.; Liu, L.; Liu, X.; Li, X.; Ao, T. Integration of Hydrological Model and Time Series Model for Improving the Runoff Simulation: A Case Study on BTOP Model in Zhou River Basin, China. *Appl. Sci.* **2022**, *12*, 6883. [\[CrossRef\]](#)
62. Ao, T.; Ishidaira, H.; Takeuchi, K.; Kiem, A.S.; Yoshitani, J.; Fukami, K.; Magome, J. Relating BTOPMC Model Parameters to Physical Features of MOPEX Basins. *J. Hydrol.* **2006**, *320*, 84–102. [\[CrossRef\]](#)
63. Liu, L.; Ao, T.; Zhou, L.; Takeuchi, K.; Gusyev, M.; Zhang, X.; Wang, W.; Ren, Y. Comprehensive Evaluation of Parameter Importance and Optimization Based on the Integrated Sensitivity Analysis System: A Case Study of the BTOP Model in the Upper Min River Basin, China. *J. Hydrol.* **2022**, *610*, 127819. [\[CrossRef\]](#)
64. Duan, Q.Y.; Gupta, V.K.; Sorooshian, S. Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization. *J. Optim. Theory Appl.* **1993**, *76*, 501–521. [\[CrossRef\]](#)
65. Yoshida, T.; Hanasaki, N.; Nishina, K.; Boulange, J.; Okada, M.; Troch, P.A. Inference of Parameters for a Global Hydrological Model: Identifiability and Predictive Uncertainties of Climate-Based Parameters. *Water Resour. Res.* **2022**, *58*, e2021WR030660. [\[CrossRef\]](#)
66. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Anderson, S.; Radić, V. Evaluation and Interpretation of Convolutional Long Short-Term Memory Networks for Regional Hydrological Modelling. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 795–825. [\[CrossRef\]](#)
68. Klotz, D.; Kratzert, F.; Gauch, M.; Keefe Sampson, A.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty Estimation with Deep Learning for Rainfall–Runoff Modeling. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 1673–1693. [\[CrossRef\]](#)
69. Li, Q.; Zhu, Y.; Shanguan, W.; Wang, X.; Li, L.; Yu, F. An Attention-Aware LSTM Model for Soil Moisture and Soil Temperature Prediction. *Geoderma* **2022**, *409*, 115651. [\[CrossRef\]](#)
70. Li, H.; Ke, C.-Q.; Zhu, Q.; Li, M.; Shen, X. A Deep Learning Approach to Retrieve Cold-Season Snow Depth over Arctic Sea Ice from AMSR2 Measurements. *Remote Sens. Environ.* **2022**, *269*, 112840. [\[CrossRef\]](#)
71. De la Fuente, L.A.; Ehsani, M.R.; Gupta, H.V.; Condon, L.E. Toward Interpretable LSTM-Based Modeling of Hydrological Systems. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 945–971. [\[CrossRef\]](#)
72. Wunsch, A.; Liesch, T.; Cinkus, G.; Ravbar, N.; Chen, Z.; Mazzilli, N.; Jourde, H.; Goldscheider, N. Karst Spring Discharge Modeling Based on Deep Learning Using Spatially Distributed Input Data. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 2405–2430. [\[CrossRef\]](#)
73. Frame, J.M.; Kratzert, F.; Klotz, D.; Gauch, M.; Shalev, G.; Gilon, O.; Qualls, L.M.; Gupta, H.V.; Nearing, G.S. Deep Learning Rainfall–Runoff Predictions of Extreme Events. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 3377–3392. [\[CrossRef\]](#)
74. Yin, H.; Guo, Z.; Zhang, X.; Chen, J.; Zhang, Y. RR-Former: Rainfall–Runoff Modeling Based on Transformer. *J. Hydrol.* **2022**, *609*, 127781. [\[CrossRef\]](#)
75. Hutter, F.; Lücke, J.; Schmidt-Thieme, L. Beyond Manual Tuning of Hyperparameters. *KI—Künstl. Intell.* **2015**, *29*, 329–337. [\[CrossRef\]](#)
76. Hunt, K.M.R.; Matthews, G.R.; Pappenberger, F.; Prudhomme, C. Using a Long Short-Term Memory (LSTM) Neural Network to Boost River Streamflow Forecasts over the Western United States. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 5449–5472. [\[CrossRef\]](#)
77. Jin, J.; Zhang, Y.; Hao, Z.; Xia, R.; Yang, W.; Yin, H.; Zhang, X. Benchmarking Data-Driven Rainfall–Runoff Modeling across 54 Catchments in the Yellow River Basin: Overfitting, Calibration Length, Dry Frequency. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101119. [\[CrossRef\]](#)
78. Zuo, G.; Luo, J.; Wang, N.; Lian, Y.; He, X. Decomposition Ensemble Model Based on Variational Mode Decomposition and Long Short-Term Memory for Streamflow Forecasting. *J. Hydrol.* **2020**, *585*, 124776. [\[CrossRef\]](#)
79. Aneja, S.; Aneja, N.; Abas, P.E.; Naim, A.G. Transfer Learning for Cancer Diagnosis in Histopathological Images. *IAES Int. J. Artif. Intell.* **2022**, *11*, 129–136. [\[CrossRef\]](#)
80. Nash, J.E.; Sutcliffe, J.V. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles—ScienceDirect. *J. Hydrol.* **1970**, *10*, 282–290. [\[CrossRef\]](#)
81. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez-Baquero, G.F. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling. *J. Hydrol.* **2009**, *377*, 80–91. [\[CrossRef\]](#)
82. Kling, H.; Fuchs, M.; Paulin, M. Runoff Conditions in the Upper Danube Basin under an Ensemble of Climate Change Scenarios. *J. Hydrol.* **2012**, *424–425*, 264–277. [\[CrossRef\]](#)
83. Knoben, W.J.M.; Freer, J.E.; Woods, R.A. Technical Note: Inherent Benchmark or Not? Comparing Nash–Sutcliffe and Kling–Gupta Efficiency Scores. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 4323–4331. [\[CrossRef\]](#)
84. Tursun, A.; Xie, X.; Wang, Y.; Liu, Y.; Peng, D.; Rusuli, Y.; Zheng, B. Reconstruction of Missing Streamflow Series in Human-Regulated Catchments Using a Data Integration LSTM Model. *J. Hydrol. Reg. Stud.* **2024**, *52*, 101744. [\[CrossRef\]](#)
85. Nogueira Filho, F.J.M.; Souza Filho, F.d.A.; Porto, V.C.; Vieira Rocha, R.; Sousa Estácio, Á.B.; Martins, E.S.P.R. Deep Learning for Streamflow Regionalization for Ungauged Basins: Application of Long-Short-Term-Memory Cells in Semiarid Regions. *Water* **2022**, *14*, 1318. [\[CrossRef\]](#)

86. Tursun, A.; Xie, X.; Wang, Y.; Liu, Y.; Peng, D.; Zheng, B. Enhancing Streamflow Simulation in Large and Human-Regulated Basins: Long Short-Term Memory with Multiscale Attributes. *J. Hydrol.* **2024**, *630*, 130771. [[CrossRef](#)]
87. Khand, K.; Senay, G.B. Evaluation of Streamflow Predictions from LSTM Models in Water- and Energy-Limited Regions in the United States. *Mach. Learn. Appl.* **2024**, *16*, 100551. [[CrossRef](#)]
88. Chen, S.; Huang, J.; Huang, J.-C. Improving Daily Streamflow Simulations for Data-Scarce Watersheds Using the Coupled SWAT-LSTM Approach. *J. Hydrol.* **2023**, *622*, 129734. [[CrossRef](#)]
89. Xu, W.; Chen, J.; Zhang, X.J.; Xiong, L.; Chen, H. A Framework of Integrating Heterogeneous Data Sources for Monthly Streamflow Prediction Using a State-of-the-Art Deep Learning Model. *J. Hydrol.* **2022**, *614*, 128599. [[CrossRef](#)]
90. Ng, K.W.; Huang, Y.F.; Koo, C.H.; Chong, K.L.; El-Shafie, A.; Ahmed, A.N. A Review of Hybrid Deep Learning Applications for Streamflow Forecasting. *J. Hydrol.* **2023**, *625*, 130141. [[CrossRef](#)]

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