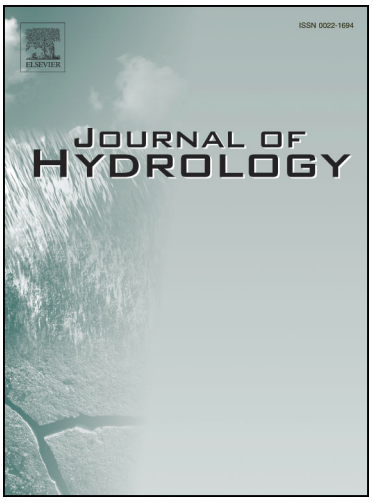
Journal Pre-proofs



Research papers

Effective improvement of multi-step-ahead flood forecasting accuracy through Encoder-Decoder with an exogenous input structure

Zhen Cui, Yanlai Zhou, Shenglian Guo, Jun Wang, Chong-Yu Xu

PII: S0022-1694(22)00339-0
DOI: https://doi.org/10.1016/j.jhydrol.2022.127764
Reference: HYDROL 127764

To appear in: *Journal of Hydrology*

Received Date: 27 December 2021
Revised Date: 9 March 2022
Accepted Date: 22 March 2022

Please cite this article as: Cui, Z., Zhou, Y., Guo, S., Wang, J., Xu, C-Y., Effective improvement of multi-step-ahead flood forecasting accuracy through Encoder-Decoder with an exogenous input structure, *Journal of Hydrology* (2022), doi: https://doi.org/10.1016/j.jhydrol.2022.127764

Effective improvement of multi-step-ahead flood forecasting accuracy through Encoder-Decoder with an exogenous input structure

Zhen Cui[1], Yanlai Zhou[1], Shenglian Guo[1], Jun Wang[1], Chong-Yu Xu[2]

[1.]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, 430072, China

[2.]Department of Geosciences, University of Oslo, P O Box 1047 Blindern, N-0316, Oslo, Norway

Correspondence to slguo@whu.edu.cn (Shenglian Guo).

Abstract: Accurate and reliable multi-step-ahead flood forecasting is beneficial for reservoir operation and water resources management. The Encoder-Decoder (ED) that can tackle sequence-to-sequence problems is suitable for multi-step-ahead flood forecasting. This study proposes a novel ED with an exogenous input (EDE) structure for multi-step-ahead flood forecasting. The exogenous input can be the outputs of process-based hydrological models. This study constructs four multi-step-ahead flood forecasting approaches, including the Xinanjiang (XAJ) hydrological model, the single-output long short-term memory (LSTM) neural network with recursive strategies, the recursive ED combined with the LSTM neural network (LSTM-RED), and the LSTM-EDE models. The performance of these four models is evaluated and compared by the long-term 3h hydrologic data series of the Lushui and Jianxi basins in China. The results show that the LSTM-RED model that integrates recursive strategies into the training process of neural networks is more advantageous than the LSTM model. The proposed LSTM-EDE model can overcome the exposure bias problem, simplify its model

structure, increase the computational efficiency in the validation process, and improve the multi-step-ahead flood forecasting accuracy, as compared to the LSTM-RED model.

Keywords: Flood forecasting; Multi-step-ahead; Neural network; Deep learning; Encoder-Decoder; Exogenous input

1. Introduction

Reliable and accurate flood forecasts are essential for reservoir operations, especially for flood control, power generation, and navigation (Zhu et al., 2020; Lin et al., 2020; Chen et al., 2020; Gauch et al., 2021; Zhang et al., 2022; Zhou et al., 2022a). Many methods or models have been developed to simulate rainfall-runoff processes for flood forecasting, which are broadly classified into two categories: process-based models and data-driven models (Samaniego et al., 2010; Birkel and Soulsby, 2015; Eslamian et al., 2018b; Seibert et al., 2018; Chen et al., 2020; Fatahi Nafchi et al., 2021; Filipova et al., 2022; Ridolfi et al., 2021; Li et al., 2021a). Process-based models consider the causal mechanisms of rainfall-runoff transformation and generalize complex hydrological phenomena to a certain extent, and therefore play a dominant role in real-time flood forecasting (Zhao, 1992; Guo et al., 2004; Chen et al., 2020; Eslamian et al., 2018a; Beylich et al., 2021). However, the rainfall-runoff transformation process is accompanied by nonlinear and non-stationary characteristics, and it is difficult for process-based models to simulate the complex hydrologic process comprehensively. Data-driven models, the deep learning model based on the long short-term memory (LSTM) neural network, benefit from their adeptness in extracting

intrinsic connections from big data to effectively simulate complex systems with nonlinear and non-stationary characteristics (Shortridge et al., 2016; Granata et al., 2016; Ostad-Ali-Askari et al., 2017; Yan et al., 2018; Hu et al., 2018; Kratzert et al., 2018; Zhou et al., 2019).

Artificial neural networks (ANNs), one of the representative data-driven models, have been successfully applied to rainfall-runoff simulation for many years (Young et al., 2017; Ostad-Ali-Askari et al., 2021a). ANNs are able to extract useful information from observed data and have proved their effectiveness in dealing with nonlinear, non-smooth series problems (Ha & Stenstrom, 2003; Chang et al., 2014 &2015; Chang & Tsai, 2016; Wang et al., 2021; Ostad-Ali-Askari et al., 2021b). With the maturity of computer technology and hardware facilities, the deep neural network (DNN) technology based on ANNs has gained broad attention (Tennant et al., 2020). Compared with the simple ANN, one of DNNs' obvious advantages is that they can consider the correlation between input variables at adjacent moments in one, two, or even higher dimensions and capture the dependencies among variables (Zhou, 2020). This characteristic is beneficial for tackling time series prediction with first-order or higher-order Markov-property (Kratzert et al., 2018). One of the most representative DNN techniques is the deep learning technique based on LSTM neural networks.

The LSTM neural network proposed by Hochreiter & Schmidhuber (1997) is a special type of recurrent neural networks (RNNs), which overcomes the problems such as gradient explosion or disappearance occurring in RNN (Zhou, 2020). It not only effectively memorizes information from previous time steps, but also achieves

parameter sharing in neural networks of different time steps, which enables a reduction in the computational cost and is very advantageous to process time-correlated data series. Thus, deep learning models based on LSTM neural networks have been applied to hydrologic forecasting. Kratzert et al. (2018) used the Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) dataset to validate the effectiveness of LSTM in rainfall-runoff modeling and found the LSTM neural network could obtain better performance when used as the regional hydrological model. Hu et al. (2018) investigated the performance of LSTM in rainfall-runoff simulation and found that the LSTM outperformed the ANN with Nash-Sutcliffe Efficiency (NSE) exceeding 0.9. Gao et al. (2020) studied the effect of the time-step length on forecast accuracy and found that the LSTM neural network could improve forecast accuracy as the time step increases. However, an LSTM model depends on a large amount of data, easily leading to overfitting. In particular, flood forecast accuracy decreases as the forecast horizon increases due to insufficient input variables with long periods. Although there are some studies that take the output of the previous forecast horizon as the input of the following forecast horizon (i.e., recursive strategy), this measure tends to cause the transmission and accumulation of forecast errors (Young et al., 2017; Zhou et al., 2019; Kurian et al., 2020).

In order to improve the accuracy and applicability of the LSTM neural network in multi-step-ahead flood forecasting, some studies fused a process-based model with the LSTM neural network in an attempt to compensate each other's inherent limitations (Zhao et al., 2019; Zhou et al., 2022b). The fusion approach generally takes the output

of the process-based model as an additional feature input to the LSTM neural network. Some studies have conducted the runoff simulation to explore the effectiveness of this fusion approach. For instance, Yang et al. (2019) constructed a hybrid physics-machine learning model by taking temperature, precipitation, wind speed, and the simulated flow of the GHMs-CaMa-Flood model chain as inputs to the LSTM neural network for simulating global floods. They concluded that the hybrid model could significantly improve the accuracy of global flood simulations. Konapala et al. (2020) proposed a hybrid model using the simulated flows from the Sacramento soil moisture accounting (SAC) model as inputs of the LSTM neural network for simulating runoff of 531 watersheds under different conditions. They concluded that the hybrid model could combine the advantages of SAC and LSTM models and improve the simulation accuracy. Cui et al. (2021) constructed a hybrid XAJ-LSTM model that integrated the Xinanjiang (XAJ) model into the LSTM neural network for multi-step-ahead flood forecasting. They found that the hybrid model could obtain more accurate long lead-time forecasts than the XAJ and LSTM models, which promotes the transformation of this fusion approach from the theoretical simulation to multi-step-ahead flood forecasting in practice.

With the development of artificial intelligence technology, the Encoder-Decoder (ED) structure is developed to solve the sequence-to-sequence problem emerging in the natural language processing. The ED structure combined with a deep learning model enables multi-step output under the premise of considering the correlation of time series and achieves the best translation results (Sutskever et al., 2014; Cho et al., 2014). The

encoding process extracts critical information from the input features in multiple time steps and information compression (Kao et al., 2020; Yin et al., 2021). The medium vector connecting the encoder and decoder processes is the hidden layer output of the last time step in the encoding process, which contains the critical information extracted from the input features. The decoding process is to transform the medium vector into the target output. The ED structure is sensitive to the order of the input sequence and can consider the temporal association among sequences. Therefore, the ED structure is suitable for tackling sequence-to-sequence problems and can directly produce multiple outputs corresponding to multiple time steps (Ha et al., 2021). The streamflow sequences are continuously varying time series with strong autocorrelation, and therefore the ED structure is suitable for handling multi-step-ahead flood forecasting. The common ED structures in hydrological forecasting can be roughly divided into two types. The first type is the recursive Encoder-Decoder (RED) structure proposed by Cho et al. (2014), where the input of each time step in the decoding process consists of the medium vector and the output value of the previous time step. The second type is the static Encoder-Decoder (SED) structure, where only the medium vector is used as the input to each time step of the decoding process (Xiang et al., 2020a; Han et al., 2021).

Xiang et al. (2020b) applied the SED structure coupled with the LSTM neural network (LSTM-SED model) to hourly flood forecasting in the Clear Creek and Upper Wapsipinicon River watersheds in Iowa, USA, with a comparison of common machine learning models, such as linear regression, lasso regression, and ridge regression. The

observed and forecasted precipitation, previously observed flow, monthly evapotranspiration, and upstream forecast data were used as inputs to the ED structure. They confirmed that the proposed LSTM-SED model could improve the accuracy of short-term flood forecasts. Kao et al. (2020) used the RED structure combined with the LSTM neural network (LSTM-RED model) for 1~6 h flood forecasting in the Shimen Reservoir watershed. The inputs to the ED structure included the previously observed flow discharges and precipitation only. They demonstrated that the LSTM-RED model could obtain accurate multi-step-ahead flood forecasts and reduce the root-mean-square error (RMSE) by 3%~38% during the testing period, compared with an ED model based on feed forward neural network (FFNN-RED) model. Han et al. (2021) constructed an LSTM-SED model to forecast floods in the Russian River basin of northern California, USA. Observed flow data at the outlet point and four upstream stations were used as inputs to the ED structure. Their results show that the LSTM-SED model performs well with NSE values ranging from 0.97 to 0.99 and correlation coefficient (R) metrics ranging from 0.98 to 0.99 at 1~6 h forecast horizons. Ha et al. (2021) developed three DNN models (i.e., LSTM, Conv LSTM, and Gate recurrent unit (GRU) neural networks) based on the RED structure and applied them to the Yangtze River basin based on El Niño-Southern Oscillation and monthly flood data. They found that the Conv LSTM neural network could improve the stability of the ED structure and produce the best forecasting results with the R index exceeding 0.8 and the RMSE index below 8000 m^3/s. Meanwhile, the introduction of the oscillation factor could improve the forecast accuracy of flood peaks and flood occurrence times. Yin et al. (2021) improved the

SED structure and developed an LSTM-based multi-state-vector sequence-to-sequence rainfall-runoff model as well as compared it with the models proposed by Xiang et al. (2020b) and Kao et al. (2020). The model used multi-state vectors for multi-step-ahead runoff forecasting, where observed and forecasted meteorological materials (such as daily maximum and minimum temperature, precipitation, and vapor pressure) and previously observed runoff were employed as inputs to the ED structure. Their experiments in 673 watersheds showed that the proposed model could improve the forecast accuracy for 7d forecast horizon of the benchmark model, and therefore is suitable for multi-day-ahead runoff forecasting.

Although the above studies demonstrated that neural networks combining ED structures could achieve more accurate multi-step-ahead flood forecasts, the recursive process problems and limitations of the RED structure have not been discussed. The following research questions need to be answered, which motivated the current study and formed the research objectives. First, what is the difference between the LSTM-RED model and the single-output LSTM model with a recursive process? Second, what are the advantages and disadvantages of the LSTM-RED model considering recursive processes for multi-step-ahead flood forecasting? Finally, can the disadvantages of the LSTM-RED model be solved by the above-mentioned fusion approach?

The following research work is carried out to achieve the objectives. (1) Explore the differences between the LSTM-RED model and the LSTM model with a recursive process, as well as analyze the impact of the recursive process within the LSTM-RED model on multi-step-ahead flood forecast accuracy from different perspectives. (2) Fuse

the flow forecasts from the XAJ model into the LSTM-ED model in an attempt to solve the disadvantages of the LSTM-RED model. To the best of the authors' knowledge, this fusion approach, focusing on combining process-based models with the single-output LSTM neural network, has not been considered so far in the ED structure models. Therefore, the novelty of this study is that an innovative ED structure with an exogenous input (EDE) is proposed for the first time by considering the forecasted flow of the XAJ model as the exogenous input variable. The LSTM neural network is integrated into the EDE structure to build the LSTM-EDE model. The Lushui and Jianxi basins in China are selected as the case studies to conduct the multi-step-ahead flood forecasting and to demonstrate the improvement in the applicability of emerging artificial intelligence techniques.

The rest of the paper is organized as follows. Section 2 presents the proposed methods and approaches. Section 3 briefly introduces the case study and materials. Section 4 evaluates and analyzes the results of the models adopted for multi-step-ahead flood forecasting. Section 5 provides a discussion based on the forecast results. Conclusions are then given in Section 6.

2. Methods

This study proposes a novel LSTM-EDE model, which is compared with the XAJ, LSTM, and LSTM-RED models. The frameworks of the RED and EDE structures is presented in **Fig.1**.

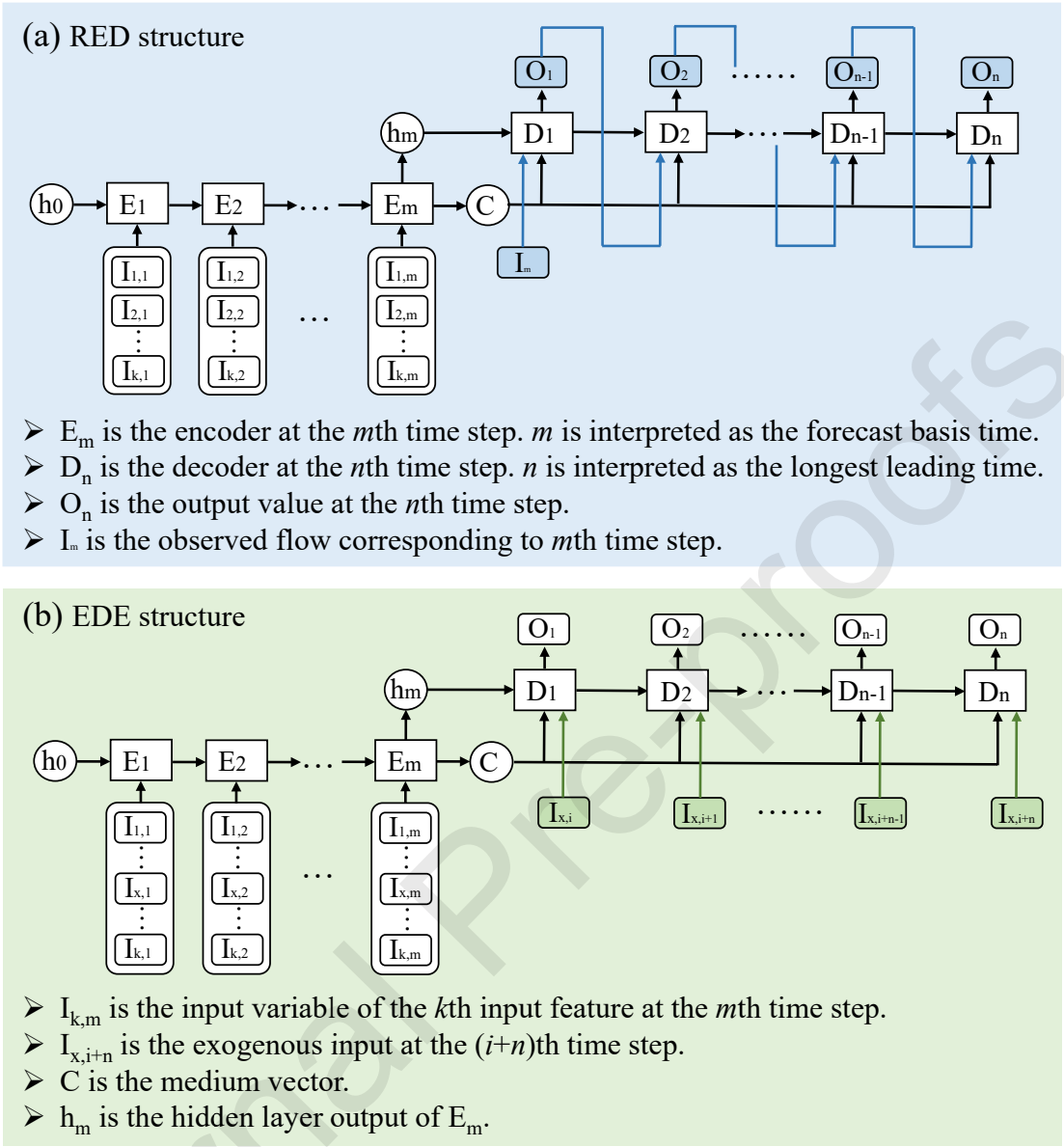


Fig.1 The frameworks of the RED and EDE structures

2.1 Recursive Encoder-Decoder (RED) structure

The RED structure in **Fig.1(a)** was proposed by Cho et al. (2014), where the input of each time step of the decoding process consists of a medium vector and the output value of the previous moment, like the strategy of recursive feedback input. The highlighted arrow in the **Fig.1(a)** indicates the recursive feedback input process.

Through the recursive process, the decoding process in the RED structure can receive not only the important information extracted from the encoding process but also the changing information from the previous moment.

During the training period, the decoding process of the RED structure does not employ a recursive input strategy, but employs the target output of the previous moment, i.e., the target output (observed flow) at t+1 moment is used as the input at t+2 moment (also known as the teacher forcing mechanism (Toomarian & Barhen, 1992)). During the validation period (also known as the inference process (Bengio et al., 2015)), the decoding process of the RED structure adopts the recursive input strategy because of the unavailability of the observed flows in the forecast horizon, i.e., the output (forecasted flow) at t+1 moment is used as the input at t+2 moment. This is the reason for the inconsistency between training and validation processes of the RED structure (the problem is called exposure bias in natural language processing (Zhang et al., 2019)). Meanwhile, it is found from **Fig. 1(a)** that the recursive input strategy complicates the inference process. We will analyze the characteristics of the RED structure from different perspectives in the subsequent sections.

2.2 Encoder-Decoder with an Exogenous input (EDE) structure

We propose the Encoder-Decoder with an Exogenous input (EDE) structure and specify the exogenous input from the forecasted flow sequence of the process-based model. The EDE structure allows the decoding process to receive information associated with the target output and helps the decoder capture the connection between the input and the target output. The length of the exogenous input should be equal to

the maximum forecasting period. The EDE structure takes the medium vector and the exogenous input as the inputs for each time step of the decoding process, and the structure is shown in **Fig.1(b)**.

The input of the decoder in the EDE structure is always the forecasted flows of the process-based model and the medium vector during training and validation processes. Therefore, the training and validation processes remain consistent. The EDE structure has another advantage of simplifying the recursive process and is beneficial to the computational efficiency of the inference process. The most critical difference between RED and EDE structures is the input factor of the decoding process, i.e., the input in the decoding process of the EDE structure is changed to the exogenous input instead of the output value of the previous moment.

2.3 Long short-term memory (LSTM) neural network unit

The LSTM neural network consists of a forgetting gate, an input gate, and an output gate. Each gate of the LSTM neural network connects the information of the input layer and the output value (h_t) of the hidden layer. The principle of the LSTM cell is shown in Equations (1) -(6).

$$f_t = \sigma(W_{x_f} \cdot x_t + W_{h_f} \cdot h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{x_i} \cdot x_t + W_{h_i} \cdot h_{t-1} + b_i) \quad (2)$$

$$\hat{c}_t = \tanh(W_{x_c} \cdot x_t + W_{h_c} \cdot h_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{x_o} \cdot x_t + W_{h_o} \cdot h_{t-1} + b_o) \quad (4)$$

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

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

where *f* denotes the output value of the forgetting gate, *i* denotes the output value of the input gate, *h* denotes the state value of the hidden layer, *W* denotes the weight, *b* denotes the bias value, *c* denotes the state of the memory unit in the hidden layer, *o* denotes the output value of the output gate, *x* denotes the input feature, $\sigma(\cdot)$ denotes the Sigmoid activation function, $\tanh(\cdot)$ denotes the tangent *S*-curve activation function, $\odot$ denotes the Hadamard product.

The single-output LSTM model adopts a recursive input strategy (Zhou et al., 2019; Xu et al., 2021) to calculate the multi-step-ahead flood forecasts in training and validation periods, i.e., the forecast results of the previous forecast horizon are used as the input to the neural network of the following forecast horizon. The LSTM model trains the neural network structure by means of multiple inputs and single outputs. After acquiring an excellent network structure, the recursive input strategy is employed for the multi-step-ahead flood forecasting.

LSTM neural network units are integrated into RED and EDE structures to construct LSTM-RED and LSTM-EDE models, respectively. We employ a fault testing method to determine the relevant hyperparameters of the neural network models. One layer with 64 hidden units (neurons) of the LSTM neural network is adopted in both the encoder and the decoder. The single-output LSTM model adopts a two-layer structure, with 64 hidden units in each layer. The loss function is the mean square error. Meanwhile, the Adam optimization method in Tensorflow.keras.optimizers is applied to training the model, and its parameters are set to be default values (learning_rate=0.001, beta_1=0.9, beta_2=0.999, epsilon=1e-07) (Li et al., 2021b). The

batch size and epochs are set to be 100 and 500, respectively.

2.4 Evaluation of model performance

The Nash-Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970), the relative error of total streamflow (RE) and the root mean square error (RMSE) are adopted to evaluate the model performance, as shown in Equations (7) -(9).

$$NSE = 1 - \frac{\sum_{i=1}^{N}\left(Q_{o,i} - Q_{f,i}\right)^2}{\sum_{i=1}^{N}\left(Q_{o,i} - \overline{Q_o}\right)^2} \tag{7}$$

$$RE = \frac{\sum_{i=1}^{N} Q_{f,i} - \sum_{i=1}^{N} Q_{o,i}}{\sum_{i=1}^{N} Q_{o,i}} \times 100\% \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_{f,i} - Q_{o,i})^2}{N}} \tag{9}$$

where N is the number of sample, Q_o and $\overline{Q_o}$ denote the observed inflows and their mean value, $Q_f$and $\overline{Q_f}$ denote the simulated/forecasted inflows and their mean value.

3. Case study and materials

3.1 Lushui basin

The Lushui River, with a watershed area of about 3950 km², is a first-class tributary of the middle reaches of the Yangtze River, as shown in **Fig.2(b)**. The topography of the basin is high in the southeast and low in the northwest. This basin belongs to the subtropical monsoon climate zone with a warm climate, and its annual average temperature, rainfall, and runoff volume are about 15.5°C, 1,550 mm, and 3.03 billion m³, respectively. The rainy season generally concentrates in April-September,

accounting for 70% of the annual rainfall. The Lushui Reservoir is located at the outlet of the valley of the mainstream. The effective storage capacity of the Lushui Reservoir is about 408 million m^3, with a flood prevention storage of only about 163 million m^3. Since the small and medium-sized mountainous basins are often characterized by concentrated precipitation and fast runoff production, as well as easy to cause flood disasters, the accurate flood forecasts can provide the reservoir important guidance for controlling river water levels and preventing flood. Due to the limitation of a small flood prevention storage capacity, accurate multi-step-ahead inflow forecasts are crucial for flood control and water resources management of the Lushui Reservoir.

This study collected data series during the wet period spanning from May 1 to October 31 during 2012 and 2019, which included 3 h precipitation data from 17-gauging stations, 3 h water surface evaporation, and inflow data of the Lushui Reservoir. All data are provided by the Hydrological Bureau of the Yangtze River Water Resources Commission. The areal mean precipitation was obtained by averaging the recorded data from 17 rainfall stations using the Thiessen polygon method. In this study, the dataset collected from 2012 to 2016 was used to train the model (training period), and the dataset collected from 2017 to 2019 was used to validate the model (validation period).

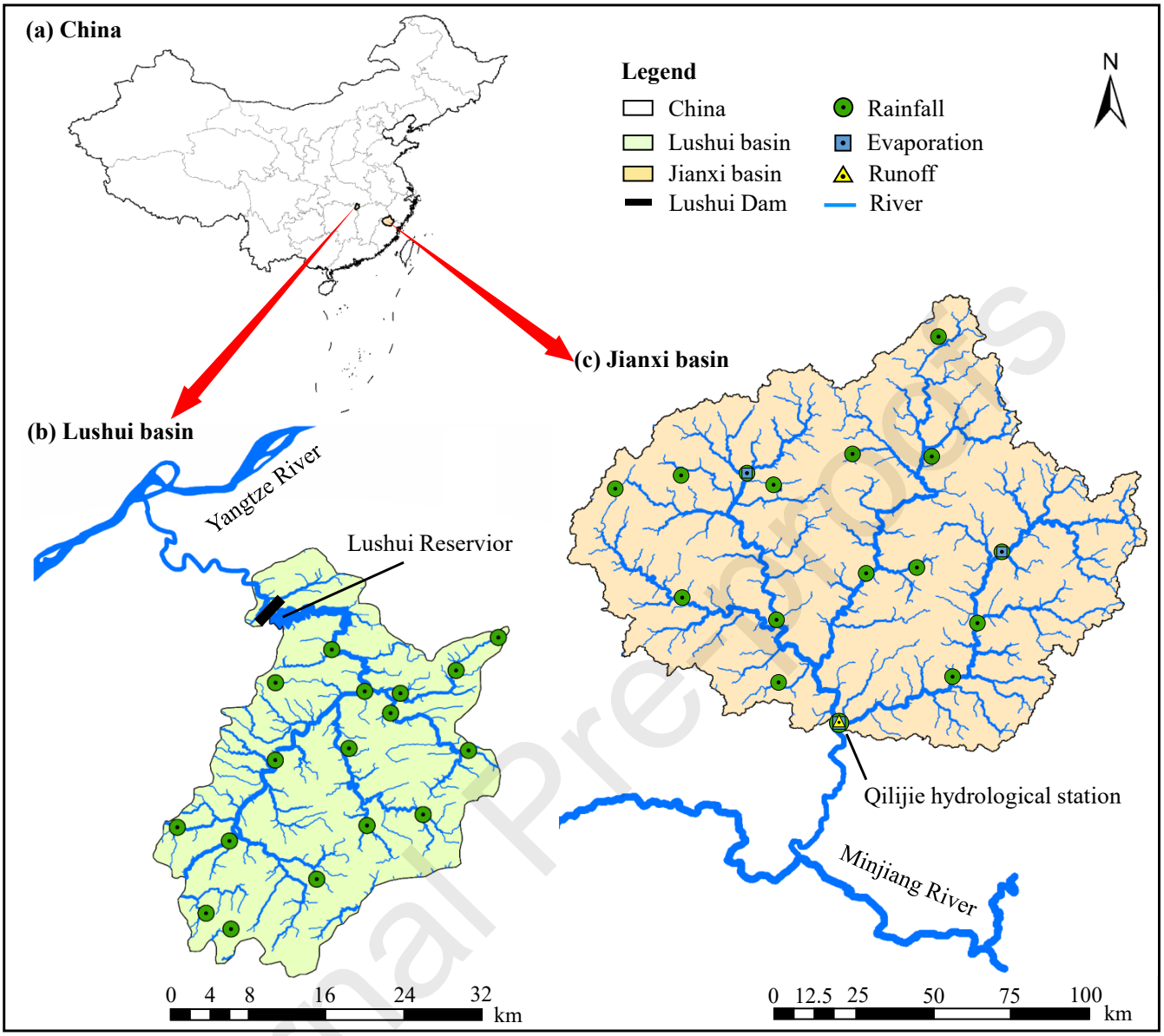


Fig.2 The Lushui basin network and rainfall gauge stations

3.2 Jianxi basin

The Jianxi River basin, with an area of 14,787 km^2, a tributary of the Minjiang River in Fujian province, China, was also selected as the case study. The Jianxi basin belongs to the subtropical monsoon climate zone, and its topography is featured by low hills and mountains. The annual average temperature, rainfall, and runoff volume are about 18 °C, 2,000 mm, and 15.8 billion m^3, respectively. Rainfall is mainly

concentrated from April to September, accounting for 75% up to 80% of the annual precipitation. **Fig.2(c)** shows the river network and rainfall gauge stations in the Jianxi basin, referring to Zhou et al. (2022b) for the specific description.

For the Jianxi basin, this study collected data series of the flood period spanning from April 1 to September 30 during 2009 and 2013, which included 3 h precipitation data from 16-gauge stations, 3 h evaporation data from 3-gauge stations, and 3 h flow data of the Qilijie hydrological station. The areal average precipitation and evaporation data are calculated using the Thiessen polygon method. The training and validation periods are 2009-2011 and 2012-2013, respectively.

3.3 Input variable selection

In this study, we choose rainfall and inflow data as the input features of the neural network. Our previous study used the cross-correlation coefficient (CCF) between the rainfall and runoff series to estimate the flow concentration time in the Lushui basin and found that the flow confluence time was about 12 hours (Cui et al., 2021). The correlation coefficients between basin mean precipitation and observed flow for different time lags in the Lushui and Jianxi basins were plotted in **Fig.3**, which shows the runoff generation and flow confluence times of the Lushui and Jianxi basins are about 12 h and 21 h, respectively.

According to the runoff generation and the flow confluence time of the Lushui and Jianxi basins, the observed flow and areal mean rainfall for the last 12 h (4 time-steps) and 21 h (7 time-steps) are adopted as the inputs of the LSTM model, respectively. All the encoder inputs are the same as those of the LSTM model. In the LSTM-RED model,

the encoder input is the input of the entire structure, and the decoder input comes from its internal output. The decoder input in the LSTM-EDE model is an exogenous input, and the target outputs for the Lushui and Jianxi basins are the observed flow corresponding to the 3~12 h forecast horizons.

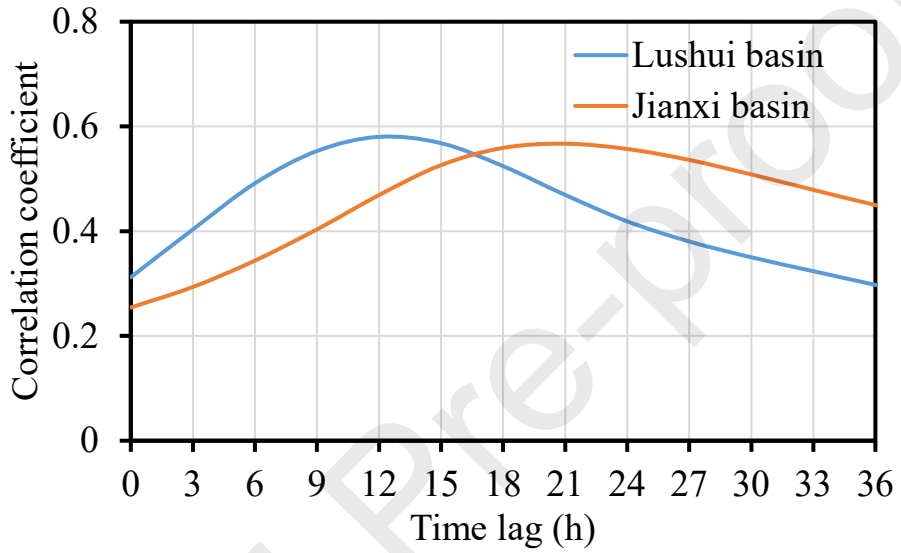


Fig.3 Correlation coefficients between basin mean rainfall and flow discharge at basin outlet for different time lags

Considering the stochastic uncertainty of some internal parameters in the neural network models, the LSTM, LSTM-RED, and LSTM-EDE models are run 30 times consecutively. The network structures with the best forecast performance are selected for analysis.

4. Results evaluation

Four models (the XAJ, LSTM, LSTM-RED, and LSTM-EDE models) were established. The structure, detailed calibration procedures, and methods of the XAJ model can be referred to in our previous studies (Cui et al., 2021; Zhou et al., 2022b).

The evaluation metrics of the four models for flood forecasting during training periods regarding Lushui and Jianxi basins are listed in Table 1. It can be observed that the forecast accuracy decreases significantly with the increase of the forecast horizon. The LSTM-EDE model is superior to the other models in the RE metric. Taking the 12 h forecast horizon in the Lushui basin as an example, the RE metrics of XAJ, LSTM, and LSTM-EDE models are -30.75%, -7.32%, and -4.17%, respectively, while the LSTM-EDE model has the lowest RE value of -0.75%.

Table 1 Evaluation metrics of four models for flood forecasting during training periods

Model	Basin	Evaluation metric	Forecast horizon			
			3 h	6 h	9 h	12 h
XAJ	Lushui	NSE	0.89	0.85	0.78	0.67
		RE (%)	-8.95	-16.40	-23.85	-30.75
		RMSE (m^3/s)	78	88	110	134
	Jianxi	NSE	0.92	0.92	0.92	0.90
		RE (%)	-0.32	-1.03	-2.39	-4.96
		RMSE (m^3/s)	297	297	303	328
LSTM	Lushui	NSE	0.99	0.98	0.93	0.82
		RE (%)	0.35	-0.27	-3.10	-7.32
		RMSE (m^3/s)	16	34	63	97
	Jianxi	NSE	0.99	0.99	0.98	0.97
		RE (%)	-0.67	-1.71	-2.50	-4.06
		RMSE (m^3/s)	65	115	145	192
LSTM-RED	Lushui	NSE	0.98	0.98	0.95	0.89
		RE (%)	0.68	-0.04	-1.59	-4.17
		RMSE (m^3/s)	15	33	61	83
	Jianxi	NSE	0.99	0.99	0.99	0.98
		RE (%)	0.28	0.23	-0.81	-2.12
		RMSE (m^3/s)	66	98	111	118
LSTM-EDE	Lushui	NSE	0.99	0.98	0.94	0.87
		RE (%)	-0.05	0.49	-0.31	-0.75
		RMSE (m^3/s)	19	35	58	84
	Jianxi	NSE	0.99	0.99	0.98	0.98
		RE (%)	1.42	2.05	2.10	2.09
		RMSE (m^3/s)	89	123	133	142

Table 2 summarizes the evaluation metrics of the four models for flood forecasting during validation periods regarding Lushui and Jianxi basins. It can be observed that

the XAJ model has the worst performance. At the 12 h forecast horizon, the NSE, RE, and RMSE metrics are 0.62, -28.06%, and 169 m^3/s, respectively, for the Lushui basin and are 0.85, -12.20%, and 361 m^3/s, respectively, for the Jianxi basin.

Table 2 Evaluation metrics of four models for flood forecasting during validation periods

Model	Basin	Evaluation metric	Forecast horizon			
			3 h	6 h	9 h	12 h
XAJ	Lushui	NSE	0.87	0.83	0.73	0.62
		RE (%)	-5.54	-13.26	-20.97	-28.06
		RMSE (m^3/s)	97	114	141	169
	Jianxi	NSE	0.87	0.87	0.87	0.85
		RE (%)	-8.21	-8.86	-10.05	-12.20
		RMSE (m^3/s)	330	332	339	361
LSTM	Lushui	NSE	0.98	0.95	0.88	0.75
		RE (%)	-0.04	-0.35	-3.23	-7.81
		RMSE (m^3/s)	32	58	96	136
	Jianxi	NSE	0.97	0.93	0.91	0.88
		RE (%)	-1.70	-4.18	-6.31	-8.66
		RMSE (m^3/s)	152	241	286	323
LSTM-RED	Lushui	NSE	0.97	0.95	0.89	0.77
		RE (%)	0.40	-0.38	-2.61	-6.22
		RMSE (m^3/s)	33	58	93	132
	Jianxi	NSE	0.97	0.94	0.92	0.90
		RE (%)	-1.45	-3.36	-5.48	-7.43
		RMSE (m^3/s)	156	235	270	293
LSTM-EDE	Lushui	NSE	0.97	0.95	0.90	0.81
		RE (%)	-0.14	0.32	-0.40	-0.97
		RMSE (m^3/s)	39	55	85	119
	Jianxi	NSE	0.98	0.95	0.94	0.93
		RE (%)	0.26	-0.15	-1.42	-2.10
		RMSE (m^3/s)	136	199	232	254

The LSTM, LSTM-RED, and LSTM-EDE models have no significant difference in 3 h forecast horizon. Starting from the 6 h forecast horizon, the evaluation metrics of the LSTM model decline significantly, especially for the RMSE metric. At the 12 h forecast horizon, the RMSE metrics of the LSTM and LSTM-RED models are equal to 136 m^3/s and 132 m^3/s, respectively, in the Lushui basin and equal to 323 m^3/s and 293 m^3/s, respectively, in the Jianxi basin. Therefore, the performance of the LSTM-RED

model is better than those of the LSTM and XAJ models.

Compared with LSTM and LSTM-RED models, the LSTM-EDE model has the least degradation in forecast performance. For 12 h forecast horizon, in the Lushui and Jianxi basins, the NSE metric reaches as high as 0.81 and 0.93, respectively, the RE metric is as low as -0.97% and -2.10%, respectively, and the RMSE metric is as small as 119 m^3/s and 254 m^3/s, respectively. It can be observed that the advantages of the LSTM-EDE model are obvious in the 12 h forecast horizon, especially in the evaluation metrics of water volume. Therefore, it is concluded that the LSTM-EDE model has the best forecast performance.

Combining Tables 1 and 2, it is found that the LSTM-EDE model can significantly improve the forecast accuracy even though the forecast accuracy of the XAJ model is relatively poor. At the same time, the LSTM-EDE model can substantially improve the computational efficiency in the validation process, as compared with the LSTM-RED model. This is a significant improvement necessary for promoting the ED structure in real-time flood forecasting.

The scatter plots of the XAJ, LSTM, LSTM-RED, and LSTM-EDE models' results for the Lushui and Jianxi basins are shown in **Figs. 4** and **5**, respectively. It can be visually observed that in the Lushui basin, the scattering points of all models under the high-flow conditions (>2500 m^3/s) are significantly lower than the 1:1 ideal line during the validation period. The reason may be that the training data lack high flows, causing all models to underestimate high flows in the validation period. The scatters of all models for the Lushui and Jianxi basins gradually disperse as the forecast horizon

increases, which indicates a gradual decrease in forecast performance. The differences in the scatter plots of different models in the Lushui basin are relatively obvious. The XAJ model produces the most scattered points and severely underestimates the observed flows, as compared to the other models, especially in the 12 h forecast horizon, as shown in **Fig4. (e)**. All the forecasts obtained from the LSTM, LSTM-RED, and LSTM-EDE models stay close to the 1:1 ideal line in the scatter plots on **Figs.4 (b)**, **(c),** and **(d)** for the 6 h forecast horizon, whereas the scatter plots of the three models show different degrees of deterioration for the 12 h forecast horizon (**Figs.4 (f)**, **(g),** and **(h)**). The scattered points of the LSTM-EDE model are the closest to the 1:1 ideal line in the 12 h forecast horizon, especially under high flow conditions (**Fig.4 (h)**). The differences in the scatter plots of different models in the Jianxi basin are small. In the 12 h forecast horizon, the scattered points of the XAJ, LSTM, and LSTM-RED models are dispersed in general, and the scattered points are generally below the 1:1 ideal line under high-flow conditions (>6000 m^3/s) (**Figs.5 (e)**, **(f)** and **(g)**). The scatter plot of the LSTM-EDE model has the narrowest band, with the scattered points the closest to the 1:1 ideal line (**Fig.5 (h)**). Therefore, it can be concluded that the LSTM-EDE model has the best forecast accuracy and can effectively reduce the forecast error in the long forecast horizon, followed by the LSTM-RED and LSTM models, and the XAJ model has the worst performance.

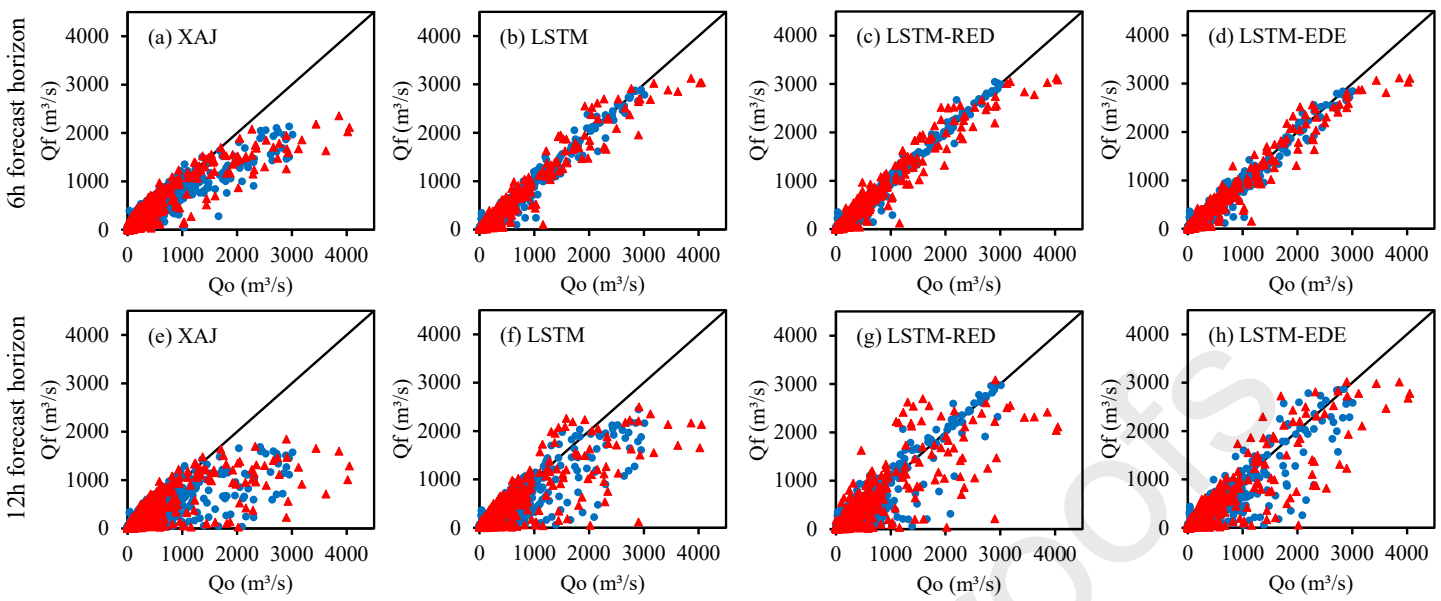


Fig. 4 Scatter plots of XAJ, LSTM, LSTM-RED, and LSTM-EDE models for the 6 h and 12 h forecast horizons in the Lushui basin, where Qo and Q_f denote the observed and forecasted flow, respectively. The blue circles and the red triangles indicate the scattered points in the training and validation periods, respectively.

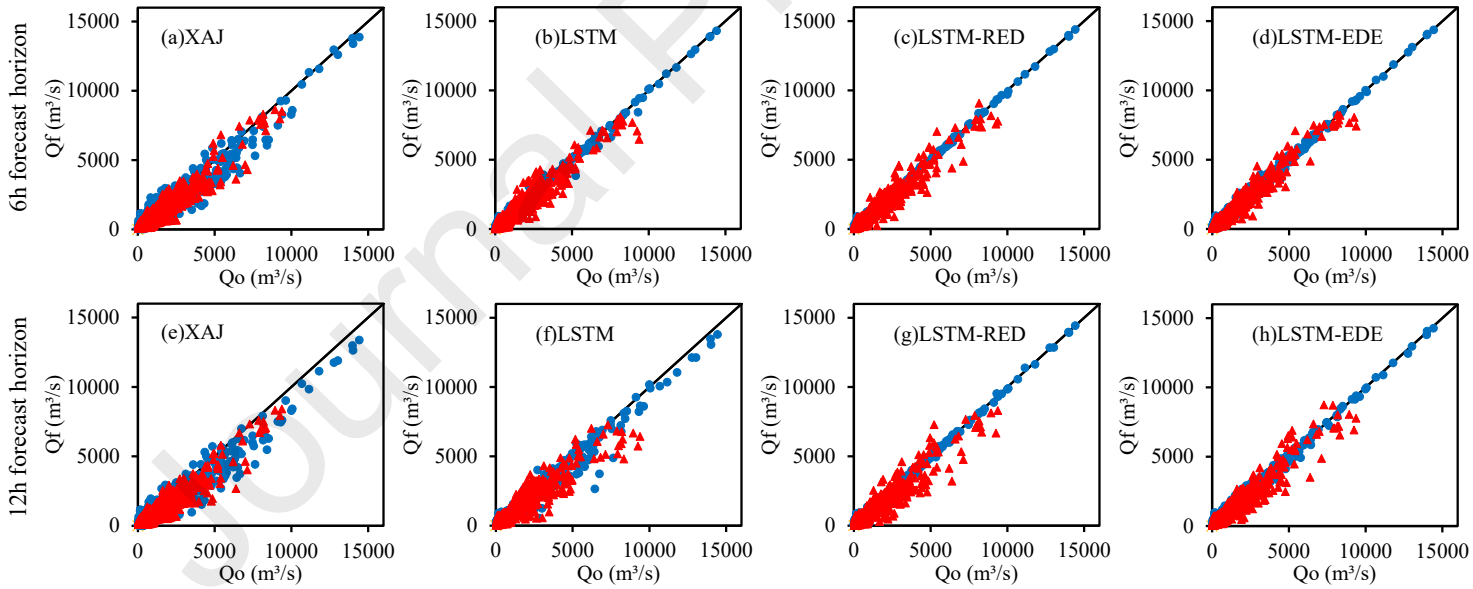


Fig. 5 Scatter plots of XAJ, LSTM, LSTM-RED, and LSTM-EDE models for the 6 h and 12 h lead-times in the Jianxi basin, where Qo and Qf denote the observed and forecasted flow, respectively. The blue circles and the red triangles indicate the scattered points in the training and validation periods, respectively.

We randomly selected three flood events in validation periods to further verify the

model performance in Lushui and Jianxi basins, respectively. A relative error of flood peak (PRE) metric is employed to evaluate the forecast performance, and it is calculated by

$$PRE = \frac{Q_{f,peak} - Q_{o,peak}}{Q_{o,peak}} \times 100\% \quad (10)$$

where $Q_{f,peak}$ and $Q_{o,peak}$ are the forecasted and observed flood peak flow, respectively.

The forecasted flood hydrographs of the Lushui basin are plotted in **Figs.6**, **7**, and **8**, respectively. And those of the Jianxi basin are plotted in **Figs.9**, **10**, and **11**, respectively. Tables 3 and 4 summarize the evaluation metrics of the forecasted hydrographs during the 12 h forecast horizon in the Lushui and Jianxi basins, respectively.

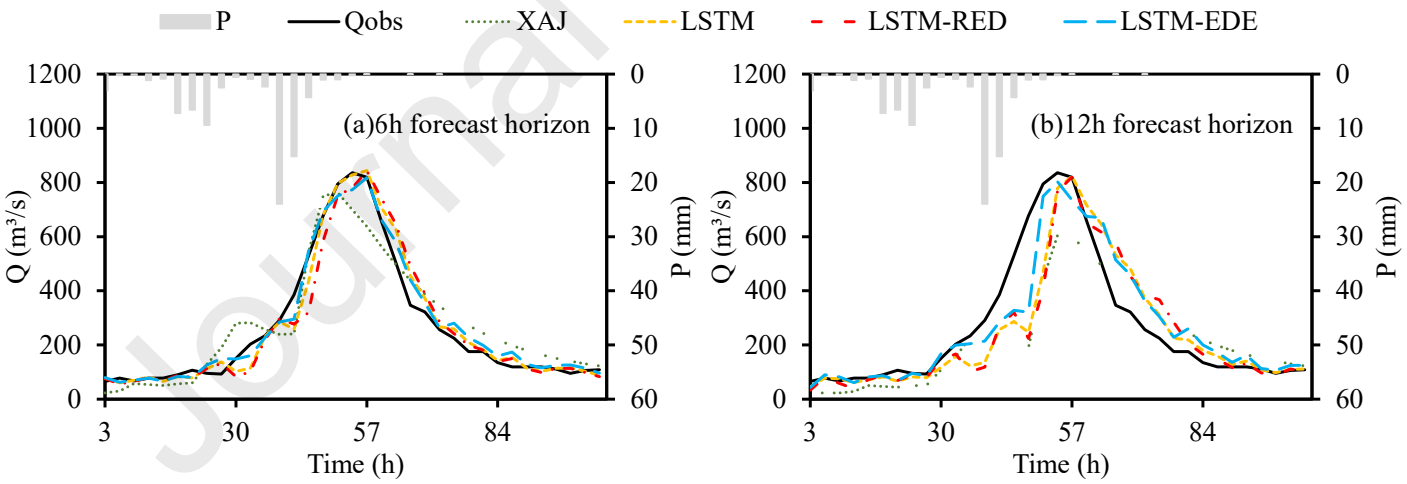


Fig. 6 Comparison of observed and forecasted flood hydrographs of the Lushui basin by four models (the flood event lasted from May 21 at 17:00 to May 25 at 23:00 in 2017).

Fig. 6 shows the observed and forecasted hydrographs of the Lushui basin by the

four models for the flood event lasting from May 21 at 17:00 to May 25 at 23:00 in 2017. The observed precipitation occurs in the rising limb and is nearly 0 mm in the recession limb. As the forecast horizon increases, the agreement between observed and forecasted data gradually worsens. The irregular fluctuations in the rising limbs of the flood hydrographs may be related to the presence of multiple rainfall peaks. The XAJ model forecasted an earlier occurrence of flood peaks for the 6 h forecast horizon and severely underestimated the peak volumes in the 12 h forecast horizon. The LSTM, LSTM-RED, and LSTM-EDE models have better performance than the XAJ model. Among them, the LSTM-EDE model has an advantage in forecasting the rising limb and the occurrence time of the flood peak.

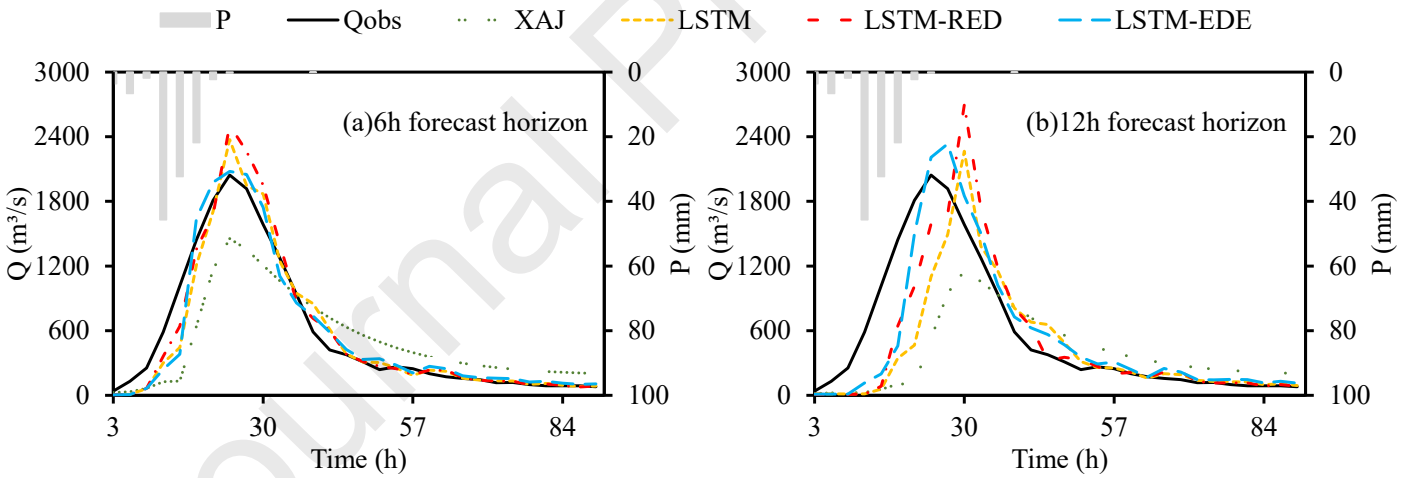


Fig. 7 Comparison of observed and forecasted flood hydrographs of the Lushui basin by four models (the flood event lasted from from May 25 at 17:00 to May 29 at 8:00 in 2019).

Fig. 7 shows the observed and forecasted hydrographs of the Lushui basin by the four models for the flood event lasting from May 25 at 17:00 to May 29 at 8:00 in 2019. It can be seen that the observed flood hydrograph rises rapidly due to the relatively

concentrated precipitation. The XAJ model underestimated the flood peak, while the LSTM and LSTM-RED models overestimated the flood peak. The LSTM-EDE model outperforms other models in terms of occurrence timing and magnitude of flood peaks.

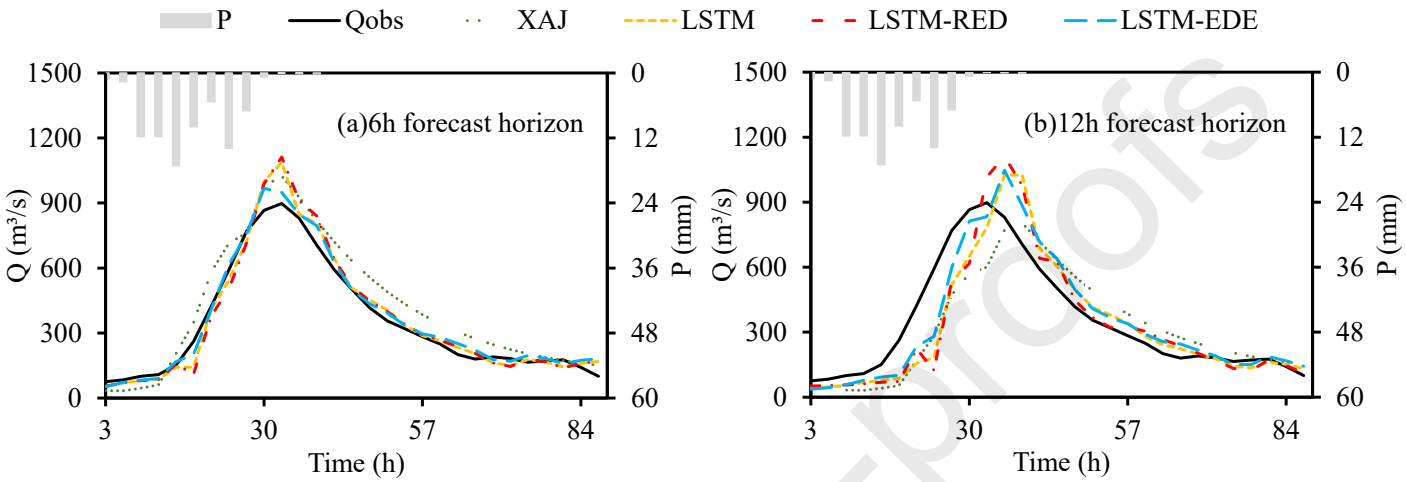


Fig. 8 Comparison of observed and forecasted flood hydrographs of the Lushui basin by four models (the flood event lasted from July 12 at 2:00 to July 15 at 14:00 in 2019).

Fig.8 shows the observed and forecasted hydrographs by the four models for the flood event lasting from July 12 at 2:00 to July 15 at 14:00 in 2019. The XAJ model overestimated the rising limb and the flood peak for the 6 h forecast horizon but underestimated the rising limb and the flood peak in the 12 h forecast horizon. The LSTM-RED and LSTM-EDE models performed reasonably well with the lagged occurrence time of the flood peak.

Table 3 reveals that the LSTM model has the lowest peak error (PRE=-1.55%), while the XAJ model has the largest flood peak error (PRE=-43.54%). According to the NSE, RE, RMSE, and PRE metrics, the LSTM-EDE model has the best forecast performance in these flood events, followed by the LSTM and LSTM-RED models, and the XAJ model has the lowest forecast accuracy.

Table 3 Evaluation metrics of four models for three flood events in 12 h forecast horizon forecasting

Model	Flood events	Evaluation metric			
		NSE	RE (%)	RMSE (m^3/s)	PRE (%)
XAJ	2017/5/21-25	0.62	-17.58	143	-27.86
	2019/5/25-29	0.22	-29.55	579	-43.54
	2019/7/12-15	0.66	-6.98	148	-10.57
LSTM	2017/5/21-25	0.72	-7.30	122	-1.55
	2019/5/25-29	0.45	-20.69	453	10.81
	2019/7/12-15	0.68	-5.95	144	14.26
LSTM-RED	2017/5/21-25	0.68	-8.69	132	-1.88
	2019/5/25-29	0.61	-10.06	385	31.84
	2019/7/12-15	0.67	-5.19	146	23.76
LSTM-EDE	2017/5/21-25	0.84	0.71	94	-3.92
	2019/5/25-29	0.78	-2.72	290	14.14
	2019/7/12-15	0.82	-0.21	109	16.68

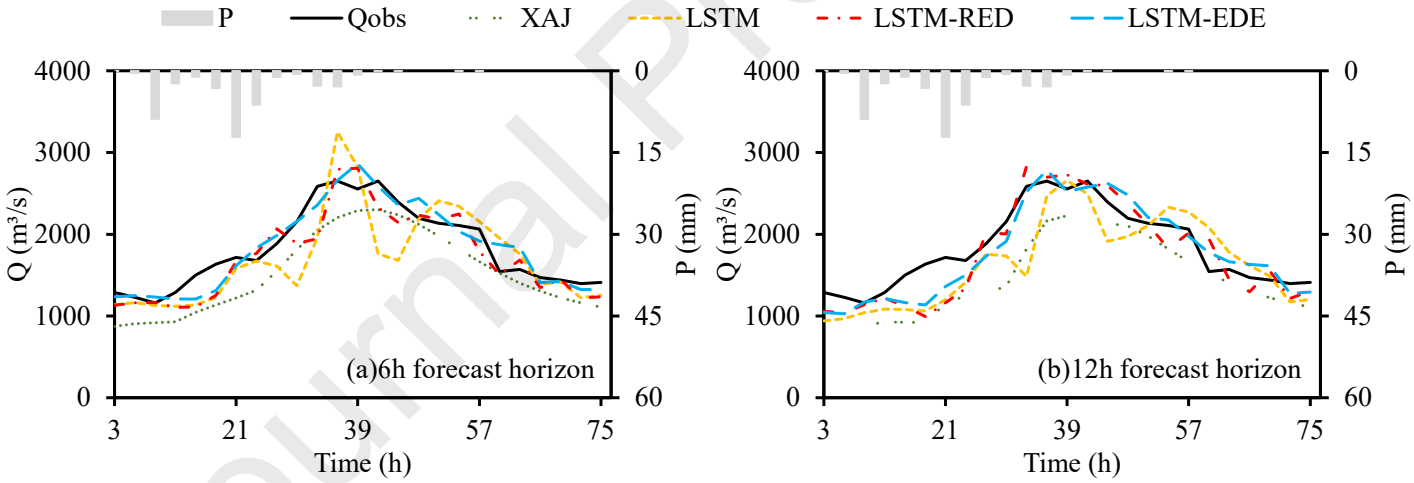


Fig. 9 Comparison of observed and forecasted flood hydrographs of the Jianxi basin by four models (the flood event lasted from May 9 at 0:00 to May 12 at 0:00 in 2012).

As shown in **Fig. 9**, the XAJ model underestimated the peak volume for the forecast horizons of the 6 h and 12 h. The LSTM model showed irregular fluctuations at the flood peak and could not maintain the reasonable shape of the flooding process. The LSTM-RED and LSTM-EDE models performed reasonably well in both cases, and

the LSTM-EDE model has a smaller peak volume error and a more accurate occurrence time of the peak.

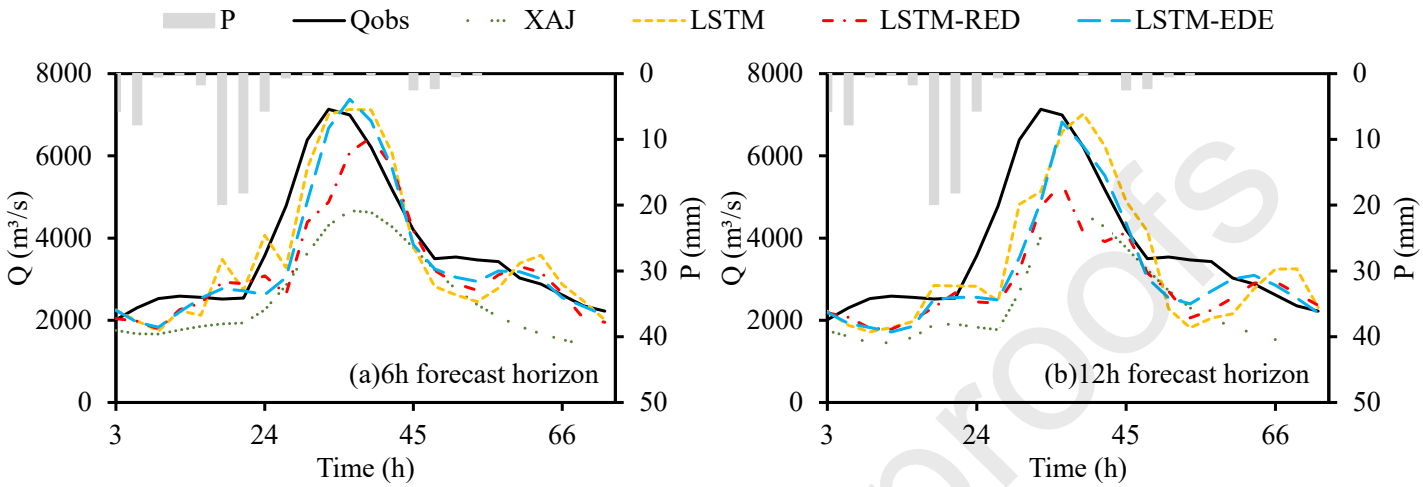


Fig. 10 Comparison of observed and forecasted flood hydrographs of the Jianxi basin by four models (the flood event lasted from April 30 at 9:00 to May 3 at 6:00 in 2012).

As shown in **Fig. 10**, the XAJ model underestimated the flood volume and had the largest volume error in both cases. In the 12 h forecast horizon, the LSTM, LSTM-RED, and LSTM-EDE models underestimated the flood volume in the rising limb. Meanwhile, the LSTM-RED model significantly underestimated the peak volume. The LSTM and the LSTM-EDE models produced reasonable flood processes. And the LSTM-EDE model performs better in forecasting the occurrence time of the flood peak and the recession limb.

As shown in **Fig. 11**, the LSTM model showed large fluctuations and had the worst forecast performance. The XAJ, LSTM-RED, and LSTM-EDE models generated relatively small volume errors, where the XAJ and LSTM-EDE models had the best forecast performance.

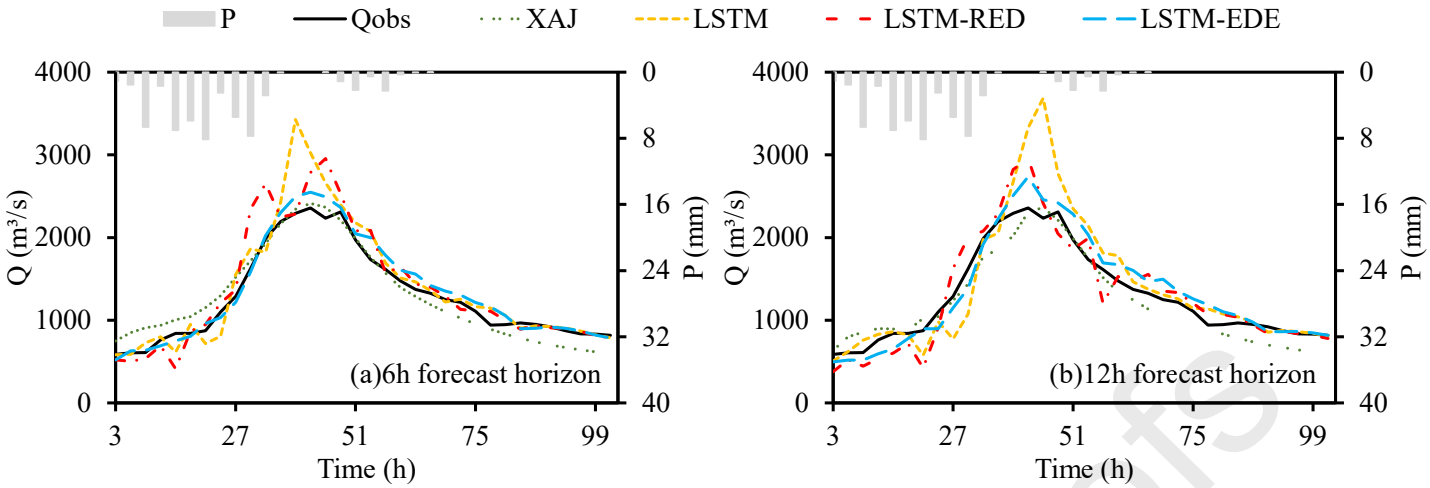


Fig. 11 Comparison of observed and forecasted flood hydrographs of the Jianxi basin by four models (the flood event lasted from May 8 at 15:00 to May 12 at 18:00 in 2013).

Table 4 shows that the LSTM model has the largest PRE value of 56.31% in these flood events. And the XAJ model has the smallest PRE value of 0.25%. According to the comprehensive analysis of these three flood events using NSE, RE, RMSE, and PRE metrics, the LSTM-EDE model has the best forecast performance, followed by the LSTM-RED, LSTM, and XAJ models.

Table 4 Evaluation metrics of four models for three flood events in 12 h forecast horizon forecasting

Model	Flood events	Evaluation metric			
		NSE	RE (%)	RMSE (m³/s)	PRE (%)
XAJ	2012/5/9-12	0.17	-20.76	430	-15.98
	2012/4/30-5/3	-0.03	-35.30	1570	-35.07
	2013/5/8-12	0.91	-6.60	165	0.25
LSTM	2012/5/9-12	0.39	-9.59	368	0.28
	2012/4/30-5/3	0.56	-10.42	1030	-1.57
	2013/5/8-12	0.54	8.55	370	56.31
LSTM-RED	2012/5/9-12	0.71	-4.51	255	6.62
	2012/4/30-5/3	0.35	-22.34	1250	-25.25
	2013/5/8-12	0.82	1.98	233	25.44
LSTM-EDE	2012/5/9-12	0.82	-2.89	204	5.01
	2012/4/30-5/3	0.57	-14.96	1014	-4.34
	2013/5/8-12	0.90	4.43	169	15.99

5. Discussion

This study analyzes and compares the multi-step-ahead flood forecasting of the XAJ model, the single-output LSTM model, and the multi-output LSTM model based on the RED and EDE structures.

The single-output LSTM model adopts a recursive input strategy to achieve multi-step-ahead flood forecasting. Although it can rapidly produce flood forecasts at multiple time-steps and save the computational cost, the training process of the LSTM model excludes the recursive process. This would cause the parameters such as weights to be obtained only based on the input and output of the first forecast horizon. The LSTM model can easily convey the forecast error from the previous forecast horizon to the following forecast horizon through the recursive process, which is likely to cause a rapid accumulation of forecast errors (Tran et al., 2016; Kurian et al., 2020). As the forecast horizon increases, the LSTM model shows substantial deterioration in the NSE, RE, and RMSE metrics, as shown in Tables 1 and 2. The NSE values during the validation period decreased from 0.98 (3 h) to 0.75 (12 h) in the Lushui basin, while from 0.97 (3 h) to 0.88 (12 h) in the Jianxi basin (Table 2).

According to Tables 1 and 2, it can be observed that the evaluation metrics of the LSTM-RED model decreased less in the 12 h forecast horizon, as compared to those of the LSTM model. This may be attributed to the fact that the LSTM-RED model can complement the input within the forecast horizon through the internal recursive process. Differing from the single-output LSTM model, the LSTM-RED model can capture the special recurrence relationship in the decoder. During the training period, the LSTM-

RED model is able to consider the total loss function of multiple forecast horizon outputs simultaneously in the single batch and iteratively updates the parameters of the neural network by error back propagation and gradient descent algorithms. In other words, the LSTM-RED model considers the influence of inputs and outputs of multiple forecast horizons on the network parameters simultaneously, which produces more extraordinary generalization performance than the LSTM model. This makes the model more reasonable and fairer when forecasting in different forecast horizons. Afterwards, the LSTM-RED model can adaptively resolve the relationship between the input and output during different forecast horizons. Therefore, the LSTM-RED model has more advantages than the LSTM model for multi-step-ahead flood forecasting.

The LSTM-RED model employs a recursive input strategy in the decoding process, which is suitable for classification issues, such as number recognition and text translation, because the output values only need to exist within the target output category, and we are not concerned with the true values of the output values. However, according to the evaluation metrics in Tables 1 and 2, the decrease in forecast performance of the LSTM-RED model is greater than that of the LSTM-EDE model. This proves that the exposure bias problem within the LSTM-RED model increases the instability of the model and has a degree of negative impact on multi-step-ahead flood forecasting. To better understand the hazards arising from the exposure bias problem, we assume that the observed and forecasted flows at t+1 moment are ‘true’ and ‘true+error’, respectively. The input at t+2 moment is ‘true’ when the neural network is trained, but the input at t+2 moment becomes ‘true+error’ during the validation

period. The increases in the forecast horizon inevitably led to the accumulation of errors. In other words, the LSTM-RED model is unable to consider the relationship between input error and target output when training the neural network. This problem may become progressively more pronounced as the forecast horizon extends. It can be observed from Tables 1 and 2 that the flood error metrics (e.g., RE and RMSE) for the LSTM-RED model become rapidly large with the increase in the forecast horizon, which leads to poor forecast accuracy for long forecast horizons (e.g., 12 h).

As shown in the previous section, the LSTM-EDE model outperforms the XAJ, LSTM, and LSTM-RED models according to the evaluation metrics, especially for the flood volume error metrics. This may be because the LSTM-EDE model can consider the relationship between the forecast error of the XAJ model and the target output during the training period and is able to make a positive response in parameters such as the weights within the neural network. Furthermore, the LSTM-EDE model can rely on the forecasted flow of the XAJ model to construct the connection between the input and the target output. Therefore, it avoids the recursive process in the decoder of the LSTM-RED model and overcomes the inconsistency between the training and validation processes, which can guarantee or even improve forecast accuracy. Meanwhile, the LSTM-EDE model simplifies the decoding process and improves the computational efficiency. In summary, the LSTM-EDE model is more suitable for long-horizon flood forecasting.

Nevertheless, the LSTM-EDE model is influenced by the accuracy of the exogenous input. As shown in Tables 1 and 2, the forecast accuracy of the XAJ model

gradually decreases as the forecast horizon increases, and the forecast performance of the LSTM-EDE model gradually deteriorates. In this case, an attempt can be made to improve the forecast accuracy of the exogenous input series by considering the numerical precipitation forecast in the XAJ model. Meanwhile, due to the limited availability of hourly observation data, the applicability of the LSTM-EDE model is only discussed in two basins in China. We will consider the LSTM-EDE model's effectiveness at a large temporal-spatial scale in subsequent studies.

6. Conclusions

In this study, a novel Encoder-Decoder structure with an exogenous input (EDE) structure was proposed, and the framework for coupling forecasted flow discharge from the process-based model (e.g., the XAJ model) was also investigated. The proposed LSTM-EDE model was compared with XAJ, LSTM, and LSTM-RED models and verified in the Lushui and Jianxi basins. The main findings were summarized as follows.

(1) Through the evaluation of the forecast performance of the XAJ, LSTM, LSTM-RED, and LSTM-EDE models, we conclude that the LSTM-EDE model has the best forecast performance, followed by the LSTM-RED, LSTM, and XAJ models. The difference between the single output LSTM model and the LSTM-RED model is that the LSTM-RED model can integrate a recursive strategy into the training process.

(2) The LSTM-RED model gradually deteriorates in the flood volume error metrics (RE and RMSE) as the forecast horizon increases. The exposure bias problem of the LSTM-RED model has a negative impact on forecasting with long forecast horizons and increases the computation time in the validation process.

(3) The LSTM-EDE model fuses the forecasted flow of the XAJ model in the decoder, which overcomes the exposure bias problem and simplifies the decoding process of the LSTM-RED model. As a result, it achieves better forecast results and effectively improves the computational efficiency in the validation process. Therefore, it is more suitable for multi-step-ahead flood forecasting.

Future research can be carried out to explore the far-reaching effects of medium- and long-term numerical weather forecast products or other effective meteorological forcing factors on the performance of the proposed approach for improving reservoir operation and water resources management (Cao et al., 2017; Ha et al., 2021). And it calls for further evaluation of the applicability of the LSTM-EDE model in more basins (e.g., CAMELS dataset) in the globe. Furthermore, the advanced uncertainty analysis method can further enhance the reliability and accuracy of the proposed approach for multi-step-ahead flood forecasting.

Acknowledgments

This study is financially supported by the National Key Research and Development Program of China (2021YFC3200301), National Natural Science Foundation of China (No. 51879192), China Three Gorges Corporation (0799254), and the Research Council of Norway (FRINATEK Project No. 274310). The authors would like to thank the Editors and anonymous Reviewers for their constructive comments that are greatly contributive to enriching the manuscript.

References

Bengio, S., Vinyals, O., Jaitly, N, Shazeer, N., 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In Advances in Neural Information Processing Systems, 28, 1171–1179. https://arxiv.org/abs/1506.03099

Beylich, M., Haberlandt, U., Reinstorf, F., 2021. Daily vs. hourly simulation for estimating future flood peaks in mesoscale catchments. Hydrology Research, 52(4), 821-833, https://doi.org/10.2166/nh.2021.152

Birkel, C., Soulsby, C., 2015. Advancing tracer-aided rainfall–runoff modelling: a review of progress, problems and unrealised potential. Hydrological Processes, 29(25), 5227–5240. https://doi.org/10.1002/hyp.10594

Cao, Q., Hao, Z., Yuan, F., Su, Z., Berndtsson, R., Hao, J., Nyima, T., 2017. Impact of ENSO regimes on developing-and decaying-phase precipitation during rainy season in China. Hydrology and Earth System Sciences, 21(11), 5415-5426. https://doi.org/10.5194/hess-21-5415-2017

Chang, F., Chiang, Y., Tsai, M., Shieh, M., Hsu, K., Sorooshian, S., 2014. Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. Journal of Hydrology, 508, 374-384. https://doi.org/10.1016/j.jhydrol.2013.11.011

Chang, F., Tsai, Y., Chen, P., Coynel, A., Vachaud, G., 2015. Modeling water quality in an urban river using hydrological factors- data driven approaches. Journal of Environmental Management, 151, 87-96. https://doi.org/10.1016/j.jenvman.2014.12.014

Chang, F., Tsai, M., 2016. A nonlinear spatiotemporal lumping of radar rainfall for modeling multi-step-ahead inflow forecasts by data-driven techniques. Journal of Hydrology, 535, 256-269. https://doi.org/10.1016/j.jhydrol.2016.01.056

Chen, X., Huang, J., Han, Z., Gao, H., Liu, M., Li, Z., Li, Z., Liu, X., Li, Q., Qi, H., Huang, Y., 2020. The importance of short lag-time in the runoff forecasting model based on long short-term memory. Journal of Hydrology, 589, 125359. https://doi.org/10.1016/j.jhydrol.2020.125359

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.,

2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. https://arxiv.org/abs/1406.1078

Cui, Z., Zhou, Y., Guo, S., Wang, J., Ba, H., He, S., 2021. A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting. Hydrology Research 52(6), 1436-1454, https://doi.org/10.2166/nh.2021.016

Eslamian, S., Parvizi, S., Ostad-Ali-Askari, K., Talebmorad, H., 2018a. Water. In: Bobrowsky P., Marker B. (eds) Encyclopedia of Engineering Geology. Encyclopedia of Earth Sciences Series. Springer, Cham. https://doi.org/10.1007/978-3-319-12127-7_295-1

Eslamian, S., Sayahi, M., Ostad-Ali-Askari, K., Basirat, S., Ghane, M., Matouq, M., 2018b. Saturation. In: Bobrowsky P., Marker B. (eds) Encyclopedia of Engineering Geology. Encyclopedia of Earth Sciences Series. Springer, Cham. https://doi.org/10.1007/978-3-319-12127-7_251-1

Fatahi Nafchi, R., Yaghoobi, P., Reaisi Vanani, H., Ostad-Ali-Askari, K., Nouri, J., Maghsoudlou, B., 2021. Eco-hydrologic stability zonation of dams and power plants using the combined models of SMCE and CEQUALW2. Applied Water Science, 11(7), 1-7. https://doi.org/10.1007/s13201-021-01427-z

Filipova, V., Hammond, A., Leedal, D., Lamp, R., 2022. Prediction of flood quantiles at ungauged catchments for the contiguous USA using Artificial Neural Networks. Hydrology Research, 53(1), 107-123, https://doi.org/10.2166/nh.2021.082.

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., Lin, Q., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589, 125188. https://doi.org/10.1016/j.jhydrol.2020.125188

Gauch, M., Mai, J., Lin, J., 2021. The proper care and feeding of CAMELS: how limited training data affects streamflow prediction. Environmental Modelling & Software, 135, 104926. https://doi.org/10.1016/j.envsoft.2020.104926

Granata, F., Gargano, R., De Marinis, G., 2016. Support vector regression for rainfall-runoff modeling in urban drainage: A comparison with the EPA's storm water management model. Water, 8(3), 69. https://doi.org/10.3390/w8030069

Guo, S., Zhang, H., Chen, H., Peng, D., Liu, P., Pang, B., 2004. A reservoir flood forecasting and

control system for China. Hydrological Sciences Journal, 49(6): 959-972. https://doi.org/10.1623/hysj.49.6.959.55728

Ha, S., Liu, D., Mu, L., 2021. Prediction of Yangtze River streamflow based on deep learning neural network with El Niño–Southern Oscillation. Scientific Reports, 11(1), 1-23. https://doi.org/10.1038/s41598-021-90964-3

Ha, H., Stenstrom, M., 2003. Identification of land use with water quality data in stormwater using a neural network. Water Research, 37(17), 4222-4230. https://doi.org/10.1016/S0043-1354(03)00344-0

Han, H., Choi, C., Jung, J., Kim, H., 2021. Deep learning with long short-term memory-based sequence-to-sequence model for rainfall-runoff simulation. Water, 13(4), 437. https://doi.org/10.3390/w13040437

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

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), 1543. https://doi.org/10.3390/w10111543

Kao, I., Zhou, Y., Chang, L., Chang, F., 2020. Exploring a long short-term memory based encoder-decoder framework for multi-step-ahead flood forecasting. Journal of Hydrology, 583, 124631. https://doi.org/10.1016/j.jhydrol.2020.124631

Konapala, G., Kao, S., Painter, S., Lu, D., 2020. Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. Environmental Research Letters, 15(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11), 6005-6022. https://doi.org/10.5194/hess-22-6005-2018

Kurian, C., Sudheer, K., Vema, V., Sahoo, D., 2020. Effective flood forecasting at higher lead times through hybrid modelling framework. Journal of Hydrology, 587, 124945. https://doi.org/10.1016/j.jhydrol.2020.124945

Li, X., Hu, Q., Wang, R., Zhang, D., Zhang, Q., 2021a. Influences of the timing of extreme precipitation on floods in Poyang Lake, China. Hydrology Research, 52 (1), 26–42, https://doi.org/10.2166/nh.2021.078

Li, D., Marshall, L., Liang, Z., Sharma, A., Zhou, Y., 2021b. Bayesian LSTM with stochastic variational inference for estimating model uncertainty in process‐based hydrological models. Water Resources Research, 57(9), e2021WR029772. https://doi.org/10.1029/2021WR029772

Lin, K., Sheng, S., Zhou, Y., Liu, F., Li, Z., Chen, H., Xu, C., Chen, J., Guo, S., 2020. The exploration of a temporal convolutional network combined with encoder-decoder framework for runoff forecasting. Hydrology Research, 51(5), 1136-1149. https://doi.org/10.2166/nh.2020.100

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology, 10(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6

Ostad-Ali-Askari, K., Shayannejad, M., Ghorbanizadeh-Kharazi, H., 2017. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan, Iran. KSCE Journal of Civil Engineering, 21(1), 134-140. https://doi.org/10.1007/s12205-016-0572-8

Ostad-Ali-Askari, K., Shayan, M., 2021a. Subsurface drain spacing in the unsteady conditions by HYDRUS-3D and artificial neural networks. Arabian Journal of Geosciences, 14(18), 1-14. https://doi.org/10.1007/s12517-021-08336-0

Ostad-Ali-Askari, K., Shayannejad, M., 2021b. Computation of subsurface drain spacing in the unsteady conditions using Artificial Neural Networks (ANN). Applied Water Science, 11(2), 1-9. https://doi.org/10.1007/s13201-020-01356-3

Ridolfi, E., Mondino, E., Di Baldassarre, G., 2021. Hydrological risk: modeling flood memory and human proximity to rivers. Hydrology Research, 52 (1), 241–252, https://doi.org/10.2166/nh.2020.195

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5), W05523. https://doi.org/10.1029/2008WR007327

Seibert, J., Vis, M., Kohn, I., Weiler, M., Stahl, K., 2018. Representing glacier geometry changes in a semi-distributed hydrological model. Hydrology and Earth System Sciences, 22(4), 2211-2224. https://doi.org/10.5194/hess-22-2211-2018

Shortridge, J., Guikema, S., Zaitchik, B., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences, 20(7), 2611-2628. https://doi.org/10.5194/hess-20-2611-2016

Sutskever, I., Vinyals, O., Le, Q., 2014. Sequence to sequence learning with neural networks. Advances in neural information processing systems.

Tennant, C., Larsen, L., Bellugi, D., Moges, E., Zhang, L., Ma, H., 2020. The utility of information flow in formulating discharge forecast models: a case study from an arid snow-dominated catchment. Water Resources Research, 56(8), e2019WR024908. https://doi.org/10.1029/2019WR024908

Toomarian, N., Barhen, J. 1992. Learning a trajectory using adjoint functions and teacher forcing. Neural networks, 5(3), 473-484. https://doi.org/10.1016/0893-6080(92)90009-8

Tran, H., Muttil, N., Perera, B., 2016. Enhancing accuracy of autoregressive time series forecasting with input selection and wavelet transformation. Journal of Hydroinformatics, 18(5), 791-802. https://doi.org/10.2166/hydro.2016.145

Wang, R., Gentine, P., Yin, J., Chen, L., Chen, J., Li, L., 2021. Long-term relative decline in evapotranspiration with increasing runoff on fractional land surfaces. Hydrology and Earth System Sciences, 25(7), 3805-3818. https://doi.org/10.5194/hess-25-3805-2021

Xu, Y., Hu, C., Wu, Q., Li, Z., Jian, S., Chen, Y., 2021. Application of temporal convolutional network for flood forecasting. Hydrology Research 52(6), 1455-1468, nh2021021. https://doi.org/10.2166/nh.2021.021

Xiang, Z., & Demir, I., 2020a. Distributed long-term hourly streamflow predictions using deep learning–a case study for State of Iowa. Environmental Modelling & Software, 131, 104761. https://doi.org/10.31223/osf.io/vs4x8

Xiang, Z., Yan, J., Demir, I., 2020b. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resources Research, 56(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

Yan, J., Jin, J., Chen, F., Yu, G., Yin, H., Wang, W., 2018. Urban flash flood forecast using support vector machine and numerical simulation. Journal of Hydroinformatics, 20(1), 221-231. https://doi.org/10.2166/hydro.2017.175

Yang, T., Sun, F., Gentine, P., Liu, W., Wang, H., Yin, J., Du, M., Liu, C., 2019. Evaluation and machine learning improvement of global hydrological model-based flood simulations. Environmental Research Letters, 14(11), 114027. https://doi.org/10.1088/1748-9326/ab4d5e

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. Journal of Hydrology, 598, 126378. https://doi.org/10.1016/j.jhydrol.2021.126378

Young, C., Liu, W., Wu, M., 2017. A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. Applied Soft Computing, 53, 205-216. https://doi.org/10.1016/j.asoc.2016.12.052

Zhang, W., Feng, Y., Meng, F., You, D., Liu, Q., 2019. Bridging the gap between training and inference for neural machine translation. 57th Annual Meeting of The Association for Computational Linguistics, 4334-4343. https://arxiv.org/abs/1906.02448v2

Zhang, H., Luo, J., Wu, J., Yu, M., 2022. Spatial–temporal characteristics and driving factors of flash floods in Shaanxi Province considering regional differentiation. Hydrology Research, 53(1), 156-174, https://doi.org/10.2166/nh.2021.103

Zhao, R. 1992. The Xinanjiang model applied in China. Journal of Hydrology, 135(1-4), 371-381. https://doi.org/10.1016/0022-1694(92)90096-E

Zhao, W., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., Qiu, G., 2019. Physics‐constrained machine learning of evapotranspiration. Geophysical Research Letters, 46(24), 14496-14507. https://doi.org/10.1029/2019GL085291

Zhou, Y. 2020. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. Journal of Hydrology, 589, 125164. https://doi.org/10.1016/j.jhydrol.2020.125164

Zhou, Y., Cui, Z., Lin, K., Sheng, S., Chen, H., Guo, S., Xu, C., 2022b. Short-term flood probability density forecasting using a conceptual hydrological model with machine learning techniques. Journal of Hydrology, 604, 127255. https://doi.org/10.1016/j.jhydrol.2021.127255

Zhou, Y., Guo, S., Chang, F., 2019. Explore an evolutionary recurrent ANFIS for modelling multi-

step-ahead flood forecasts. Journal of Hydrology, 570, 343-355. https://doi.org/10.1016/j.jhydrol.2018.12.040

Zhou, Y., Guo, S., Xu, C., Xiong, L., Chen, H., Ngongondo, C., Li, L., 2022a. Probabilistic interval estimation of design floods under non-stationary conditions by an integrated approach. Hydrology Research, 53(2), 259-278, https://doi.org/10.2166/nh.2021.007.

Zhu, S., Luo, X., Yuan, X., Xu, Z., 2020. An improved long short-term memory network for streamflow forecasting in the upper Yangtze River. Stochastic Environmental Research and Risk Assessment, 34(9), 1313-1329. https://doi.org/10.1007/s00477-020-01766-4

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Credit Author Statement

Zhen Cui: Conceptualization, Formal analysis, Methodology, Resources, Software, Writing - original draft.

Yanlai Zhou: Conceptualization, Data curation, Investigation, Software, Validation, Visualization.

Shenglian Guo: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing-review & editing.

Jun Wang: Data curation, Formal analysis, Investigation, Validation, Visualization.

Chong-Yu Xu: Funding acquisition, Methodology, Project administration, Resources, Supervision.

Highlights

- A novel Encoder-Decoder with an Exogenous input (EDE) structure is proposed
- Four models are evaluated and compared from different perspectives
- The EDE structure is more suitable for long lead-time flood forecasting
- The LSTM-EDE model improves the multi-step-ahead flood forecasting accuracy