Research on real-time correction of flood forecasts in the middle reaches of the Yellow River using AR、ARMAX and LSTM models

Running head: Research on real-time correction of flood forecast

Authors: Caihong Hu [1,2], Zhichao Li [1,2], Qiang Wu [1,2*], Jinxin Wang，Changqing Li [1,2] and Yuanhao Xu

[1] School of Water Conservancy Science and Engineering, Zhengzhou University, 450001 Zhengzhou, China

[2] Yellow River Institute for Ecological Protection & Regional Coordinated Development, Zhengzhou 450001, China[1]

*Corresponding author: wuqianghlj@163.com

Acknowledgments

We are very grateful to potential reviewers for their constructive comments and suggestions. We would also like to thank other colleagues for their valuable comments and suggestions, which are of great help to improving the manuscript.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Correspondence
Qiang Wu, School of Water Conservancy Science and Engineering, Zhengzhou University, 450001 Zhengzhou, China
E-mail: wuqianghlj@163.com
Funding information
National key research priorities program of China (2019YFC1510703), National Natural Science Foundation of China (51979250), Key Research and Promotion Projects (technological development) in Henan Province (202102310588), Key projects of National Natural Science Foundation of China (51739009)

Research on real-time correction of flood forecasts in the middle reaches of the Yellow River using AR、ARMAX and LSTM models

Abstract: The development of flood forecasting technology is crucial to flood control. Therefore, it is very essential to use the method of real-time error correction to improve the accuracy and reliability of the flood forecasting model. For flood forecasting, this study evaluated the performance of a single Excess Infiltration and Excess Storage (EIES) flood forecast model and the forecast model after error correction using the linear Auto Regressive, Auto Regressive Moving Average with exogenous inputs (ARMAX), and Long Short-term Memory Network (LSTM), and then compared the performance of each model forced with historical flood data in the upper reaches of Jingle station of the Fen River in China. These EIES-standalone, EIES-AR, EIES-ARMAX, and EIES-LSTM frameworks are field-tested for 1- to 6-hours lead-time flood forecasting with historical flood data. The capability of the four models are compared using the mean absolute error (MAE), Nash–Sutcliffe efficiency (NSE), Pearson's correlation coefficients (r), and Percent error in volume (Evol). The evaluation measures analysis reveal that EIES-AR and EIES-ARMAX perform acceptable when the lead time is 1 hour(NSE>0.7), but poorly when the lead time is 2-6 hours; EIES-LSTM model performs well and is the best approach of these models for short to medium range flood forecasting with up to 6 hours lead-time(NSE⩾0.75).

Keywords: Error correction; EIES; Flood forecast; AR; ARMAX; LSTM

1. Introduction

As people pay more and more attention to the flood disasters in many river basins in the world, real-time flood forecasting is regarded as a non-structural measure to develop flood early warning system. The premise of flood control is not only reliable flow, but also sufficient lead time. Generally speaking, the reliability of the advance time of flood forecasting is assessed by the minimum early warning time required for disaster management preparatory action. If predictions are to be made which are in view of existing historical discharge or water level records, using these records to extrapolate a period of time is acceptable.

In the literature on flood forecasting, distributed hydrological model is not only an effective means to explore and understand the process and mechanism of complex hydrological cycle, but also an effective tool to solve the important problems in the field of hydrology. It has played an important role in the research of climate change, Land-Use/Cover Change (LUCC), lack of data areas, eco-hydrology, water resources management and other fields (Hwang et al., 2018; Ko, et al.,2019; Krogh & Pomeroy,2019). Physical (distributed) and experience-based (concept) hydrological models are used to calculate watershed runoff (Beven,1989； Berstrom,1991； Singh,1995； Refsgaard,1996). The benefits of using distributed hydrological models retain the watershed characteristics (Vieux et al., 2002). In order to simulate runoff in the flow domain, an appropriate hydrological model must be selected. According to the watershed hydrological profile and the existing data, an appropriate distributed hydrological model is selected to forecast the watershed runoff in real time. Through the event-based rainfall-runoff model, the discharge forecast in the basin and the warning to flood events are effective (Ajmal et al., 2015; Reddy et al., 2007; Zhang et al., 2015). When planning and managing water resources, accurate measurement of runoff and flood peaks plays a vital role (Athanasios & Lampros, 2014). In the light of the complexity and importance of the runoff process, the runoff process must be simulated according to the physical law that controls the runoff phenomenon. The observed runoff and meteorological data were primarily used for understanding hydrological processes and to calculate runoff in the basins. The hydrological process depends on watershed parameters, such as slope, drainage and canal characteristics (Kishor, et al. 2014; Jong, et al. 2015).

Due to the complexity of hydrological phenomena, there are uncertainties in the structure of the model, model parameters, and model input, which affect the prediction accuracy of the hydrological model and reduce the accuracy of the prediction results. The development of real-time correction technology has greatly improved the accuracy of hydrological model in flood forecasting. At present, there are many real-time correction methods applied to flood forecasting, which can be roughly divided into two categories: the first is the error correction of forecasting process. Its essence is to correct the parameters and state variables of the flood forecasting model, and then forecast by using the corrected model, so as to improve the forecasting accuracy, such as recursive least square method, dynamic system response method (Si, et al., 2015), Kalman filter algorithm (Zhou, et al., 2020; Lee, et al., 2019) and so on. However, this kind of method needs complete intermediate real-time monitoring data, so it is difficult to apply in watersheds where hydrological monitoring data are scarce. The second is the error correction of the forecast results, that is, the error of the model forecast is directly corrected without directly considering the flood forecasting process, so as to effectively reduce the forecast error, update the original forecast value and improve the forecast accuracy, such as error autoregressive(AR) method, autoregressive moving average(ARMA) method, ARMA with exogenous input (ARMAX) models (Bogner & Kalas, 2008) and back propagation neural network(BPNN) correction method(Thirumalaiah & Deo, 2000). Among the many correction schemes, the "forecast result error correction method" has been widely used. it uses flow observations to directly modify the forecast discharge in real time without interfering with the operation of the flood forecasting model and does not need to re-run the flood forecasting model. there is no need to modify the model parameters. This kind of method and forecasting model are not only related to each other, but also operate independently to a certain extent. it can be used in conjunction with a variety of flood forecasting models with low requirements for real-time data, and can be applied to a wider range of river basins.

At present, linear regression methods are widely used in real-time correction of flood forecasting. such as autoregressive moving average (ARMA) or ARMA with exogenous inputs, introduced by Box and Jenkins (1976). AR has been used to forecast discharge (Shamseldin & O'Connor, 1999; Abrahart & See, 2000). According to the World Meteorological Organization (WMO, 1992), there are generally four types of programs for model update, including the update of model state variables, parameter variables, input variables, and output variables. The output variable update method of the model is not related to the structure of the simulation model and has a wide range of applications, also is the most extensive. As usual, linear regressive time series models have been used such as AR (Serban & Askew, 1991; Xiong & O'Connor, 2002) or ARMA (Shamseldin & O'Connor, 2001; Broersen, 2007) models. For this purpose, the AR or ARMA model is calibrated using the error existing in the model prediction result, and then the calibrated model is used for error prediction, and the predicted error is added to the prediction result of the previous forecast model to provide an updated forecast value. The success of the AR or ARMA model update process mainly depends on the correlation degree of the time series of the forecast error obtained by the forecast model (Serban & Askew, 1991; Shamseldin, 1997). However, these linear regressive models can not represent the highly nonlinear dynamics inherent in flooding processes well and therefore may not always perform adequately, which is the defect of the linear regression method used in flood forecasting (Hsu et al., 1995).

Data-driven models are becoming more and more popular in the field of hydrological modeling based on big data and computational intelligence tools such as artificial neural networks (ANNs) in the last two decades. These computational intelligence tools are based on a limited understanding of system physics, using only state variables as the input and output of the model, analyzing the characteristics of system data, and establishing the correspondence between variables. Therefore, when using ANNs for forecasting, by establishing mathematical analysis on time series, learning given data samples, and discovering statistics or causal relationships between variables (Dawson & Wilby, 1998). The long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) is a modern type of recurrent neural network (RNN) that involves feedback links in the architecture of the network. In the RNN, the output from the previous step is a plied as the input of the current step, and the input and output of ANNs are unconstrained. However, one drawback of the RNN is that it can struggle to learn long sequences; hence, training can be extremely problematic and lead to the vanishing/exploding gradient problem (Hochreiter & Schmidhuber, 1997). LSTM can solve this problem in RNN training. By considering the short-term state and long-term state, LSTM network can identify valuable inputs, save them in the long-term state, and extract this information whenever it is needed (Lipton et al., 2015). There are many studies on ANNs for data-driven hydrological modelling, whereas the number of studies using LSTM for this purpose is relatively fewer. In addition, there are few researches on combining traditional hydrological forecasting models with machine learning models for flood forecasting.

For these reasons, this study is dedicated to (1) test and verify effectiveness of LSTM in real-time correction for flood forecast with multiannual flood events; (2) test the applicability of the LSTM for different lead times (1-6 hours); and (3) compare the performance of the AR and ARMAX model to verify the advantages of the model's correction performance.

2. Case study

2.1. Study area

In the past 50 years, due to the combined effects of climate change and human activities, the water cycle of the Yellow River Basin has undergone significant changes, and the watershed situation has become more complicated. Especially in the middle reaches of the Yellow River, two runoff mechanisms coexist, namely, full-storage runoff and over-permeability runoff, making hydrological forecasting more and more difficult. For this reason, we choose the object of composition research in Fen River Basin for experiment.

The Fen River which is located in the middle reaches of the Yellow River, is the second largest tributary of the Yellow River and is situated in the central and southwestern parts of Shanxi Province, China. The upper reaches of the Fen River is about 216km long and the drainage area is about 7705km^2. It belongs to the mid-latitude semi-arid, semi-humid temperate continental monsoon climate. The annual average precipitation is 503mm, and the precipitation decreases from southeast to northwest. The precipitation from June to September accounts for more than 70% of the total annual precipitation. The amount of precipitation varies greatly from year to year, and there is continuous low water. The average water surface evaporation for many years is 1567-2063mm. The controlled area of Jingle Station occupies about one-third of the upper reaches of the Fen River. It is located at 111°55′ east longitude and 38°20′ north latitude. The rainfall distribution in the basin controlled by Jingle station is uneven, showing from south to north and east to West is gradually increasing. The average annual runoff of Jingle Station is 152.52 $m^3{\cdot}s^{-1}$.

2.2. Dataset

In this study, Select the rainfall data of the 14 rainfall stations upstream of Jingle Station during the flood season from 1956 to 2014, and the corresponding flood data at Jingle Station. The data interval is 1h, divided by floods, and a total of 98 floods are selected. The first 78 floods are selected as the training set, The time span is 1956-2003, the last 20 floods are used as the verification and test set: the first 10 floods are used as a verification set, and the last 10 floods are used as a test set, and the time span is 2003-2014.

3. Methodology

The detailed framework of the prediction method proposed here is given in figures 2 and the detailed structure of the standalone model is shown in figure 4. First of all, the Excess Infiltration and Excess Storage model(EIES) is established by using the time series data of Jingle. Then, an independent EIES model is used to predict the early runoff. Since then, this square method has been named EIES-standalone method. The other three methods are to combine EIES model with AR, ARMAX and LSTM error correction or update models for flood forecasting, which are named EIES-AR, EIESARMAX and EIES-LSTM models respectively. Therefore, the EIES independent model works in the on-line simulation mode, and simulates the flood forecast in advance in the EIES model without any error correction. On the contrary, the variants of EIES-AR, EIESARMAX, and EIES-LSTM models work in the on-line error update mode, using the error time series generated by the EIES model corresponding to the observed data, and then using the error series to establish an error update model to correct the deviation in the lead time prediction. A brief description of the modeling framework and error update process is given below.

3.1. Flood forecast model

In view of the characteristics of runoff generation and confluence in the semi-arid and semi-humid areas of the middle reaches of the Yellow River, this study chose the Excess Infiltration and Excess Storage model (EIES) to conduct flood forecasting experiments in the Jingle Basin. This model combines the characteristics of the three-water source model and the tank model on the basis of the Xinanjiang model (Zhao 1980,1992), and also considers the combined mechanism of super-permeable runoff and full-filled runoff, which is conducive to the simulation of the actual situation and analysis The main runoff generation mode during rainfall makes the model more applicable in semi-arid and semi-humid regions.

The infiltration excess runoff generated over the watershed is due to the variability of soil heterogeneity. This is not considered in the original Xinanjiang model. The basic structure of the modified Xinanjiang model is shown in Figure 3. In the figure, $W'{\sim}\alpha$ is the distribution curve of water storage capacity in the basin, where W_0 is the actual soil water storage capacity of the basin at the beginning of the period of rainfall P, and its corresponding maximum ordinate value is W_0'; $F_{\Delta_t}'{\sim}\beta$ is placed on W_0' and corresponds to W_0. The distribution curve of infiltration capacity during the watershed period, $F_{m\Delta_t}'$ is the maximum ordinate of $F_{\Delta_t}'{\sim}\beta$; x is the distance from the intersection of the two curves of $W'{\sim}\alpha$ and $F_{\Delta_t}'{\sim}\beta$ to the origin.

3.2. Error forecasting models

It is completely feasible to develop an error prediction model using the input error timeseries of $\varepsilon(t), \varepsilon(t-1), \ldots, \varepsilon(t-d)$, obtained during the model calibration phase, so as to achieve the purpose of updating the EIES model to predict the discharges; where $\varepsilon(t)$ is the simulation error between the observed(Q_{obs}) and EIES-simulated (Q_{sim}) discharges at any time, t, estimated as: $\varepsilon(t) = Q_{obs}(t) - Q_{sim}(t)$; and d is the effective correlation time lag, which is determined by the autocorrelation function analysis of the error time series. Subsequently, the historical error sequence can be used as a forecast to predict the error at the forecast time level α

$$\hat{\varepsilon}(t+\alpha) = f\{\varepsilon_{sim}(t), \varepsilon_{sim}(t-1), \varepsilon_{sim}(t-2), \ldots \varepsilon_{sim}(t-d)\} \quad (1)$$

where $f\{\cdot\}$ represents a linear or nonlinear error prediction model selected based on the complexity of the hierarchical model. The details of these error correction models are as follows.

3.2.1. AR error-correction model

The AR model correction algorithm assumes that the prediction errors are dependent on each other, and according to the discovery law of the historical prediction error series, it is used to predict the future errors, so as to correct the original prediction results. In the operation forecast, an autoregressive model (correction model) based on the error is usually constructed according to the error between the measured value and the predicted value in the first few periods of the forecast, and then according to the correction model, the error of the prediction time is calculated and added to the predicted value, which is the predicted value after the time is corrected. AR model is widely used in practical production because of its simple algorithm and less data requirements. The mathematical expression of the AR model is:

$$\hat{\varepsilon}(t+\alpha) = \left[a_i \sum_{i=1}^{d} \varepsilon(t+1-i)\right] + e(t+\alpha) \quad (2)$$

where $\varepsilon(t) = Q_{obs}(t) - Q_{sim}(t) =$ flow deviation simulated by EIES model, used as exogenous inputs;$\hat{\varepsilon} =$ error simulated by the Aru model; e= white noise; $d =$ time delay of inputs, it determines the order of the AR model; and $\alpha =$ lead-time of forecast.

3.2.2. ARMAX error-correction model

The autoregressive model can be effectively matched with the moving average model to form the autoregressive moving average model (ARMA), can eliminate the trend of the non-stationary time series by differential processing, and then model. The model can be expressed as

$$A(q) \cdot \hat{\varepsilon}(t+\alpha) = B(q)\left[a_i \sum_{i=1}^{d} \varepsilon(t+1-i) + \sum_{j=1}^{\tau} \varepsilon(t+\alpha-j)\right] + C(q) \cdot e(t+\alpha) \quad (3)$$

where $d =$ time delay for exogenous inputs used in the ARMAX model; $\tau =$ the time delay of the internal source inputs; $A(q)$, $B(q)$, and $C(q)$ are the polynomials of the regression equation. The order of the polynomial is determined by trial and error.

3.2.3. LSTM error-correction model

The LSTM controls the information flowing into the cell through the input gate (Input gate), the output gate (output gate) and the forgetting gate (forget gate). The output of the sigmoid layer indicates that the information is all passed, and the output is 0 means that the information is completely blocked. Among them, the forgetting gate can be understood as a selective forgetting strategy, which determines how much information of the cell unit from the previous moment can be retained to the next moment. The input gate determines how much input information currently can be saved in the cell unit, and how much the state of the control unit is output to the number of current output value of the LSTM and

LSTM error forecasting model could be saw in Figure 5.). The general memory block of LSTM model can be expressed by the formula:

$$C(n+1) = \sigma\left[w_f X(n+1) + W_f Y(n) + b_f\right] \otimes C(n) + \sigma[w_i X(n+1) + W_i Y(n) + b_i] \otimes tanh[w_c X(n+1) + W_c Y(n) + b_c] \tag{4}$$

$$Y(n+1) = \sigma[W_o X(n+1) + W_o Y(n) + b_o] \otimes tanh[C(n+1)] \tag{5}$$

where C is the cell state, and w and W are the weights of the connections between gates and layers. b_i, b_f and b_o are the biases. The symbol σ and $tanh$ are activation functions.

3.3. Performance evaluation of the models

The Chinese flood forecasting guidelines recommend using the pass rate (a) to estimate flood forecasting performance for flood events. Therefore, during training and testing, the size of the flood peak and the pass rate of the occurrence time were calculated. when the difference between the predicted peak and the recorded (observed) value is within ±20% of the recorded (observed) value, the predicted peak emissions are called "qualified". (Li et al., 2010). The qualified flood rate α_P can be calculated by the following formula:

$$\alpha_P = \frac{1}{n}\sum_{i=1}^{n} l\left(\frac{\hat{Q}_P - Q_P}{Q_P} \leq 0.2\right) \tag{6}$$

where l(·) is the index function; $\hat{Q}_P$ is the predicted size of flood peak; Q_P is the recorded or observed value of flood peak; n is the number of flood events.

When the difference between the flood peak forecast and the recorded (observed) occurrence time is within the allowable error (equal to 30% of the observed flood peak occurrence time), the flood peak forecast occurrence time is called "qualified". The pass rate of flood occurrence time α_T can be calculated by

$$\alpha_T = \frac{1}{n}\sum_{i=1}^{n} l\left(\frac{\hat{T}_P - T_P}{T_P} \leq 0.3\right) \tag{7}$$

where T_P and $\hat{T}_P$ are the observed and predicted occurrence times of flood peak, respectively, which are equal to the time of the flood peak minus the time of flood forecasting.

The average of the two pass rates $\bar{\alpha}$ is calculated as follows:

$$\bar{\alpha} = (\alpha_P + \alpha_T)/2 \tag{8}$$

During the test, we use the mean absolute error (MAE) and Pearson's correlation coefficient (r) to evaluate the performance of the model. The volume percentage error (Evol) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) are also selected to describe the accuracy of the hydrological model, and they are also used to describe the statistical characteristics of the model. This research (Moriasi et al., 2015). r and NSE are two standardized statistical indicators. The value of r close to 1 indicates that the simulation effect is better, the value of NSE is from negative infinity to 1, and the value of NSE close to 1 indicates that the model quality is good and the model reliability is high; NSE is close to 0, which means that the simulation result is close to the observed value Average level, that is, the overall result is credible, but the process simulation error is very large; NSE is far less than 0, the model is not credible. As recommended by previous studies, if r is greater than 0.84 per day, month or year, NSE is greater than 0.50, and the values of MAE and Evol are smaller, the model performance can be "satisfactory" for flow simulation. Forecast model. (Moriasi et al., 2007; Yang et al., 2016; Zhang et al., 2018). The calculation formulas for these evaluation indicators are as follows.

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_0^t - Q_m^t)^2}{\Sigma_{t=1}^{T}(Q_0^t - \bar{Q}_0)^2} \tag{9}$$

$$MAE = \frac{\sum_{t=1}^{T}|Q_o^t - Q_m^t|}{T} \tag{10}$$

where Q_0 refers to the observed value, Q_m refers to the simulated value, Q^trefers to a value at time t, and $\bar{Q}_0$ refers to the total average of the observations.

4. Results and Discussion

4.1. Performance of the standalone EIES model for streamflow simulation

The standalone EIES model was calibrated separately with the historical rainfall timeseries of JINGLE considering 78 floods of 1956-2003 as the calibration period. The calibrated EIES model parameters while using the rainfall data of JINGLE inputs are listed in Table 2. The model was validated for 10 floods of the period 1993–1998 and tested for 10 floods of the period 1998–2003.

Fig. 6 illustrates the timeseries and scatter plots of the streamflow forecasts at 1- to 6-hours lead-times by the standalone EIES model using the historical rainfall timeseries of JINGLE. It can be surmised from Fig.6 and Table2 that, in forecasting mode, the standalone EIES model underestimated the discharge forecasts with NSE ≤ 0.40, MAE > 29m^3/s, |Evol| > 30%, and r < 0.7 for 1–6 hours lead-time forecasts. Hence, the lead-time stream-flow forecasting using the EIES-standalone model is unacceptable.

4.2. Lead-time inflow forecasting by the EIES model in the error-updating model

First, we calculated the qualified rate of the flood peak discharge and peak appearance time of the three models after correction during the training period and the test period, and then calculated the indicators of r, NSE, MAE and Evol during the test period to show the performance of the three models.

4.2.1. Performance of AR error-correction model

It can be seen from Table 3 and Table 4 that in the training period, when the lead time is 1-6 hours, α_P is below 80%, but when the lead time is 1 hour, α_P is above 75%, which is 78.3%. α_T is above 85%, which is 85.4%, $\bar{\alpha}$ is 81.85%, more than 80%, the performance is good; when the lead time is 2-4 hours, α_P≤70%, $\bar{\alpha}$ is below 80%, and only when the lead time is 2 hours,

α_P>80%. When the lead time is 5-6 hours, α_P reaches below 60%, and $\bar{\alpha}$ is below 70%. In the test period, all indicators are better than those in the training period, α_P, α_T, and $\bar{\alpha}$ are all above 80%. Other performance indicators: r=0.85, NSE=0.79, MAE=15.47 m^3/s, |Evol|=24.36, which shows good performance; In other forecast periods, the indicators of α_P, α_T and $\bar{\alpha}$ are roughly the same as those of the training period, and NSE<0.7. Figure 7 shows the corrected effect of the AR model to the forecast result of the EIES model. On the whole, when the lead time is 1 hour, the results of the AR correction model can be trusted, but the performance results in other lead times are not acceptable.

4.2.2. Performance of ARMAX error-correction model

Compared with the performance of the AR model, the performance shown by ARMAX is even worse. It can be seen from the ARMAX model prediction results and various indicators shown in Table 3 and Table 5 that in the training period, when the lead time is 1-6 hours, α_P is below 80%, but the same as the AR model is that when the lead time is 1 hour, α_P is above 75%, which is 75.2%. α_T is above 85%, which is 85.1%, $\bar{\alpha}$ is 80.15%, more than 80%, the performance is good; when the lead time is 2-3 hours, α_P≤70%, $\bar{\alpha}$ is below 80%, When the lead time is 4-6 hours, even α_P reaches below 60%, and $\bar{\alpha}$ is below 70%. In the test period, most indicators are better than those in the training period. When the lead time is 1 hour, α_P>75%, α_T, and $\bar{\alpha}$ are all above 80%. Other performance indicators: r=0.82, NSE=0.72, MAE=18.25 m^3/s, |Evol|=26.35, which shows good performance; In other forecast periods, the indicators of α_P, α_T and $\bar{\alpha}$ are roughly the same as those of the training period, and NSE<0.7. Figure 8 shows the corrected effect of the ARMAX model to the forecast result of the EIES model On the whole, when the lead time is 1 hour, the results of the ARMAX correction model can be trusted, but the performance results in other lead times are not acceptable.

4.2.2. Performance of LSTM error-correction model

Different from the linear model, LSTM has the function of feedforward and feedback co-regulation, which can eliminate useless information between various data, so that it has stronger learning ability. It can be seen from the LSTM model prediction results and various indicators shown in Table 3 and Table 6 that in the training period, when the lead time is 1-5 hours, α_P is above 80%(α_P=80.1%~92.4%), and α_T is all above 85%,$\bar{\alpha}$ = 81.6%~94.35%; In the test period, most indicators are better than those in the training period. When the lead time is 1-6 hours, α_T>85%, α_P, and $\bar{\alpha}$ are all above 80%. Other performance indicators: r=0.87-0.93, NSE=0.75-0.87, MAE=11.34-14.25 m^3/s, |Evol|=5.65-10.88%, which shows pretty good performance; Figure 10 shows the corrected effect of the LSTM model to the forecast result of the EIES model. On the whole, when the lead time is 1-6 hours, the results of the LSTM correction model can be trusted.

Overall, the performance of the EIES-standalone and the three error-updating model variants improved in the order: EIES-standalone < EIES-ARMAX < EIES-AR < EIES-LSTM signifying the superiority of the LSTM error forecasting model.

In order to clearly show the model's performance when predicting flood peaks, we also evaluated the peak flow forecasting ability of the EIES-standalone, EIES -AR, EIES -ARMAX, and EIES -LSTM models for the 10 floods of test period. With the decrease of the forecast period, the distribution range of E_{peak} of the AR correction method and LSTM correction method are gradually reduced, and the forecast accuracy is gradually improved. The ARMAX correction method performs well when the forecast period is 1h, and when the forecast period is 3h and 6h The correction effect is not ideal, and the correction result is even worse than the prediction result of EIES; for the treatment of abnormal values, the three correction methods all eliminate the abnormal value of the flood peak prediction of the EIES model. In general, as the forecast period decreases, the correction effect of the AR method tends to stabilize, and the correction effect of the ARMAX method fluctuates, and the effect is worse than the AR method. The best correction effect is the LSTM method, especially in the forecast period of 1h At time, most of the E_{peak} are distributed between 10%-20%, and the peak errors of other forecast periods have been significantly improved. In addition, taking the flood number 20030729 as an example ,timely prevention and early warning of extreme flood events are extremely important in flood forecasting. Among the advocated error-updating models, the peak flood forecasting capability was improved in the order: EIES -ARMAX → EIES -AR → EIES -LSTM (e.g., Fig. 11, Fig12 and see Table 7).

4. Conclusion

For flood management, under the limited availability of hydrometeorological data, there must be an efficient real-time forecast system. Therefore, in this study, a new method based on the LSTM-based external error update model is integrated with the EIES model to predict the reservoir inflow 1 to 6 hours in advance. By using the proposed error update scheme of the 1-hour EIES-AR model and the 6-hour lead time EIES-LSTM model, the shortcomings of independent EIES model predictions forced by historical runoff data can be effectively solved. The dynamic neural network structure of the LSTM error model has short-term autoregressive characteristics, which can effectively reduce the uncertainty in the high lead time flow forecast. The three joint real-time calibration models provided in this article overestimate low flows, but underestimate high flows, especially peak flows. In general, as the complexity of the model increases, the accuracy of flood forecasting is significantly improved: EIES-standalone < EIES-

ARMAX < EIES-AR < EIES-LSTM. The AR model and ARMAX model have the best correction effect when the lead time is 1 hour. The correction effect of other forecast periods is not credible. Take the experiment with a forecast period of 1 hour as an example. Compared with the AR model (NSE=0.79) and the ARMAX model (NSE=0.72), the correction effect of the LSTM model is greatly improved (NSE=0.87), and the performance is much better than the other two models. A model, and when the lead time is 2-6 hours, the LSTM model still has a good correction effect (NSE=0.75-0.83). In addition, the proposed EIES-LSTM model also could address the issue of flood forecasting with sufficient lead-times to be deployed in flood warning system.

Essentially, the terminal error of flood forecasting is the accumulation of errors in the intermediate process of flood forecasting. The process error correction method can correct the error of each process to reduce the final error. This is the theoretical advantage, but it is currently in practical applications. It is difficult to obtain complete intermediate monitoring data (such as detailed monitoring data of rainfall, runoff, and confluence), which limits the use of this method.

Although big data analysis technology has not yet made substantial progress in the real-time correction of flood forecasting, this represents the most promising and breakthrough research direction in the future. Based on massive actual observations and derived data (including previous precipitation, soil moisture, reservoir storage state, river bottom water; the location of the rainstorm center of the current rainfall, temporal and spatial distribution, rainfall (strong), more early climate background, Atmospheric circulation factors; historically observed rainfall and flood series; forecast results data of different model schemes, etc.), using various machine learning algorithms, such as random forest, support vector machine, convolutional neural network (CNN), Deep Neural Network (DNN), RNN, LSTM, etc., to find the association rules between terminal errors or process errors and data, and then establish a big data correction model for forecast errors.

References

Abrahart, R. J., & See L. (2000). Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes* **14**(11-12): 2157–2172. https:// doi.org/10.1002/1099-1085(20000815/30)14:11/12<2157::AID-HYP57>3.0.CO;2-S

Ajmal, M., Waseem, M., Ahn, J. H., & Kim, T. W. (2015). Improved runoff estimation using event-based rainfall-runoff models. *Water Resources Management* **29**(6):1995–2010. https:// doi.org/ 10.1007/s11269-015-0924-z

Athanasios, L., & Lampros, V. (2014). Streamflow simulation methods for ungauged and poorly gauged watershed. *Natural Hazards and Earth System Sciences* **14**(7):1641–1661. https:// doi.org/ 10.5194/nhess-14-1641-2014

Bai, P., & Liu, X. M. (2018). Evaluation of Five Satellite-Based Precipitation Products in Two Gauge-Scarce Basins on the Tibetan Plateau. *Journal of Remote Sensing* **10**(8),112-139. https:// doi.org/ 10.3390/rs10081316

Bergstrom, S. (1991). Principles and confidence in hydrological modeling. *Nordic Hydrology* **22**(2), 123 – 136. https:// doi.org/ 10.2166/nh.1991.009

Beven, K. J. (1989). Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology* **105**(1-2):157–172. https:// doi.org/ 10.1016/0022-1694(89)90101-7

Broersen, P. M. T. (2007). Error correction of rainfall–runoff models with the ARMAsel program. *IEEE Transactions on Instrumentation and Measurement* **56**(6): 2212–2219. https:// doi.org/ 10.1109/TIM.2007.908252

Dawson, C. W, & Wilby, R. (1998). An artificial neural network approach to rainfall–runoff modelling. *Hydrological Sciences Journal* **43**(1): 47–66. https:// doi.org/ 10.1080/02626669809492102

Hochreiter, S, & Urgen, S. J. (1997). Long short-term memory. *Neural Computing* **9**(8):1735–1780. https:// doi.org/ 10.1162/neco.1997.9.8.1735

Hsu, K., Gupta, V. H., & Sorroshian, S. (1995). Artificial neural network modeling of the rainfall–runoff process. *Water Resources Research* **31**(10): 2517–2530. https:// doi.org/ 10.1029/95WR01955

Hwang, T., Martin, K. L., & Vose, J. M. (2018). Non-Stationary Hydrologic Behavior in Forested Watersheds is Mediated by Climate-Induced Changes in Growing Season Length and Subsequent Vegetation Growth. *Water Resource Research* **38**：1681−1696. https:// doi.org/ 10.1029/2017WR022279

Jong, P. K, Won, K., Wong, J., Gwangseob K. (2015). Runoff prediction in Ungauged watersheds using remote sensor datasets. *Journal of Water Resource and Hydraulic Engineering* **4**(3):257–264. https:// doi.org/ 10.5963/JWRHE0403007

Kishor, C., Balram, P., & Jagadish, C. P. (2014). Simulation of rainfall-runoff process using HEC-HMS model for Balijore Nala Watershed, Odisha, India. *International Journal of Geomatics and Geosciences* **5**(2):1203-1219.

Ko, A., Mascaro, G., & Vivoni, E. R. (2019). Strategies to Improve and Evaluate Physics-Based Hyperresolution Hydrologic Simulations at Regional Basin Scales. *Water Resource Research* **55**(2), 1129-1152. https:// doi.org/ 10.1029/2018WR023521

Krogh, S. A, & Pomeroy, J. W. (2019). Impact of Future Climate and Vegetation on the Hydrology of an Arctic Headwater Basin at the Tundra-Taiga Transition. *Journal of Hydrometeorology* **20**(2), 197-215. https:// doi.org/ 10.1175/JHM-D-18-0187.1

Lee, H., Shen, H., Noh, S. J., Kim, S., Seo, D., & Zhang, Y. (2019). Improving flood forecasting using conditional bias-penalized ensemble Kalman filter. *Journal of Hydrology* **575**:596-611. https:// doi.org/ 10.1016/j.jhydrol.2019.05.072

Li, X., Guo, S., Liu, P., & Chen, G. (2010). Dynamic control of flood limited water level for reservoir operation by considering inflow uncertainty. *Journal of Hydrology* **391**(1-2):124–132. https:// doi.org/ 10.1016/j.jhydrol.2010.07.011

Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *Statistics* **4**:1–35.

Moriasi, D. N. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE* **50**(3):885–900.

Moriasi, D. N, Gitau, M. W, Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE* **58**(6):1763–1785. https:// doi.org/ 10.13031/trans.58.10715

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—a discussion of principles. *Journal of Hydrology* **10**(3):282–290.

Reddy, K. V, Eldho, T. I, & Rao, E. P. (2007). Simulation of runoff from watershed using finite element based kinematic wave model. ISH *Journal of Hydraulic Engineering* **13**(2):15–30. https:// doi.org/ 10.1080/09715010.2007.10514868

Refsgaard, J. C. (1996). Terminology, modeling protocol and classification of hydrological model codes. In Distributed Hydrological Modeling, M.B. Abbott and J.C. Refsgaard (eds.), Kluwer, Dordrecht, pp. 17 – 39.

Xiong, L., & O'Connor, K. M. (2002). Comparison of four updating models for real-time river flow forecasting. *Hydrological Science Journal* **47**(4):621–640. https:// doi.org/ 10.1080/02626660209492964

Serban, P., & Askew, A. J. (1991). Hydrological forecasting and updating procedures. IAHS (*International Association of Hydrological Sciences*) Publication: Vienna, Austria.

Shamseldin, A. Y. (1997). Application of a neural network technique to rainfall–runoff modeling. *Journal of Hydrology* **199**(3–4): 272–294. https:// doi.org/ 10.1016/S0022-1694(96)03330-6

Shamseldin, A. Y. (2010). Artificial neural network model for river flow forecasting in a developing country. *Journal of Hydroinformatics* **12**(1):22–35. https:// doi.org/ 10.2166/hydro.2010.027

Shamseldin, A. Y, & O'Connor, K. M. (1999). A real-time combination Method for the outputs of different rainfall–runoff models. *Hydrological Sciences Journal* **44**(6): 895–912. https:// doi.org/ 10.1080/02626669909492288

Shamseldin, A. Y, & O'Connor, K. M. (2001). A Non-Linear Neural Network Technique for Updating of River Flow Forecasts. *Hydrology and Earth System Sciences* **5**(4): 577–97. https:// doi.org/ 10.5194/hess-5-577-2001

Si, W., Bao, W., & Gupta, H. (2015). Updating real-time flood forecasts via the dynamic system response curve method. *Water Resource Research* **51**(7):5128-5144. https:// doi.org/ 10.1002/2015WR017234

Singh, V. P. (1995). Computer Models of Watershed Hydrology *Water Resources Publications*, Highlands Ranch, CO.

Thirumalaiah, K., & Deo M. C. (2000). Hydrological forecasting using neural network *Journal of Hydrologic Engineering* **5**(2): 180-189. https:// doi.org/ 10.1061/(ASCE)1084-0699(2000)5:2(180)

Vieux, B. E, Ester, C., Dempsey, C., & Vieux, J. E. (2002). Prospects for a real-time flood warning system in Arizona. In: Proceedings of the 26th Annual Conference, Breaking the Cycle of Repetitive Flood Loss, Association of State Floodplain Managers, Phoenix.

Yang, T., Gao, X., Sorooshian, S., & Li, X. (2016). Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme. *Water Resource Research* 52 (3):1626–1651. https:// doi.org/ 10.1002/2015WR017394

Zhao, R. (1980). The Xinanjiang model. Proceedings Oxford Symposium. IAHS Publ. 129:351 – 356.

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

Zhang, D. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology* **565**:720–736. https:// doi.org/ 10.1016/j.jhydrol.2018.08.050

Zhang, J., Chen, L., Singh, V. P, Cao, H., & Wang, D. (2015). Determination of the distribution of flood forecasting error. *Natural Hazards* **75**(2):1389–1402. https:// doi.org/ 10.1007/s11069-014-1385-z

Zhou, Y., Guo, S., Xu, C., & Chang, F. (2020). Improving the Reliability of Probabilistic Multi-Step-Ahead Flood Forecasting by Fusing Unscented Kalman Filter with Recurrent Neural Network. *Water* **12**(2):578-593. https:// doi.org/ 10.3390/w12020578