

Post-processing of hydrological model simulations using the convolutional neural network and support vector regression

Songnan Liu^{a,b}, Jun Wang^{ib a,c,*}, Huijun Wang^{a,b} and Yuetao Wu^c

^a Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China

^b Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China

^c Nansen-Zhu International Research Centre, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

*Corresponding author. E-mail: wangjun@mail.iap.ac.cn

 JW, 0000-0001-5783-6990

ABSTRACT

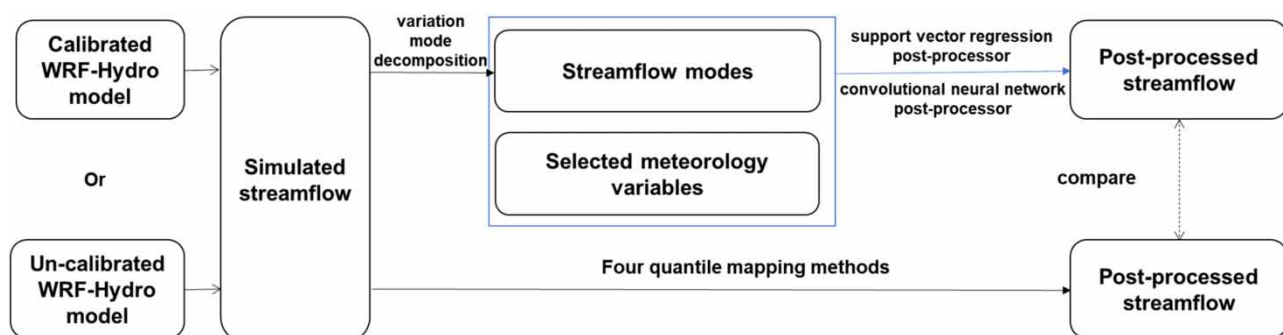
Post-processing methods can be used to reduce the biases of hydrological models. In this research, six post-processing methods are compared: quantile mapping (QM) methods, which include four kinds of transformations, and two newly established machine learning frameworks [support vector regression (SVR) and convolutional neural network (CNN)] based on meteorological data and variation mode decomposition (VMD)-decomposed streamflow. These post-processing methods are applied to a distributed model (WRF-Hydro), and the evaluation is carried out over five watersheds with different areas in South China. The post-processing methods are separately applied to calibrated and uncalibrated models. The results show that the SVR- and CNN-based post-processing methods perform better than the QM methods in terms of daily streamflow simulations in different areas with different topographies in the Xijiang River basin. There are large uncertainties in the QM post-processing methods. The CNN-based post-processing performs slightly better than the SVR-based post-processing, but both methods can markedly improve the simulated streamflow. The CNN- and SVR-based post-processing frameworks are suitable for both calibration and test periods. The differences between post-processing with uncalibrated and calibrated models are quite small for SVR- and CNN-based post-processing, but large for QM post-processing. For WRF-Hydro, the CNN- and SVR-based post-processing methods consume much less time and computational resources than model calibration.

Key words: convolutional neural network, machine learning, post-processing, support vector regression, WRF-Hydro

HIGHLIGHTS

- Two machine learning post-processing methods [convolutional neural network (CNN) and support vector regression (SVR)] are established and tested.
- CNN- or SVR-based post-processing is superior to quantile mapping post-processing in the Xijiang River basin.
- CNN- or SVR-based post-processing is suitable for uncalibrated and calibrated models in the Xijiang River basin.

GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Hydrological modeling benefits water resources management and can help reduce the impacts of floods and droughts. Therefore, it is a field of study that has continually captured the attention of many researchers. Various dynamic nonlinear processes in the hydrological process present challenges for simulating streamflow (Rezaie-Balf *et al.* 2019). Hydrological models can be categorized into data- and physics-driven models. The former type only considers the statistical relationship between the streamflow and the predictors, while the latter considers the physical processes and the mathematical reasoning (Solomatine & Ostfeld 2008). Compared with data-driven models, physics-driven models consider the nonstationary climate change, can generate more reliable results under climate change scenarios using physical constraints, and have strengths for hydrological forecasting (Carpenter & Georgakakos 2006). Therefore, it is reasonable to use physics-driven models to simulate and forecast streamflow.

Physics-driven models contain numerous physical and nonlinear processes, and it is complicated to transfer the rainfall to streamflow in these models (Hsu *et al.* 1995). The uncertainties in physics-driven models arise from the errors of meteorological inputs, inappropriate model frameworks, and incorrect model parameters (Shi *et al.* 2008). Therefore, the calibration of model parameters is essential for applying physics-driven models in hydrological studies. However, because of the complexity and nonlinearity of hydrological processes, a perfectly calibrated model is almost impossible to achieve (Vrugt *et al.* 2008), and even if the model has been perfectly calibrated, the errors of the model framework can still lead to inconsistencies between the observed and simulated streamflow (Li & Sankarasubramanian 2012). Thus, an additional post-processing method of those models is needed to reduce the model errors. In general, post-processing methods are used to improve the model prediction skill and quantify the predictive uncertainty of ensemble streamflow predictions (Robertson *et al.* 2013; Madadgar *et al.* 2012; Xu *et al.* 2019). Except for the ensemble streamflow predictions, post-processing methods can also reduce the biases of a single model (Kim *et al.* 2021). And statistical post-processing methods have been introduced to improve model performance by reducing the biases between the simulated and observed streamflow (Bogner & Kalas 2008; Kim *et al.* 2021). Among the methods, quantile mapping (QM) post-processing is a commonly used post-processing method, which can be employed to reduce the biases of streamflow and other hydrological variables (Shi *et al.* 2008; Madadgar *et al.* 2012). Lucatero *et al.* (2018b) applied the QM method to the model predictions of hydrologically relevant variables such as precipitation, temperature, and evapotranspiration, and found that QM can reduce the biases in the model simulation. Hashino *et al.* (2007) evaluated three bias-correction methods (including QM) for ensemble streamflow volume forecasts, and the results showed that all three methods significantly improved the forecast quality by eliminating unconditional biases and enhancing the potential skill. Lucatero *et al.* (2018a) used QM to reduce the biases of simulated streamflow and concluded that the QM method can improve model performance. It has also been reported that post-processing can enhance the calibration in an efficient way (Yuan & Wood 2012; Ye *et al.* 2015).

With the advancement of technology and computing power, machine learning (ML) techniques have been developed. ML has gained attention in hydrological modeling, and combining ML with hydrological models is gaining popularity (Reichstein *et al.* 2019). ML has been used in hydrological simulations and hydrological post-processing. Wu *et al.* (2019) used ML techniques to post-process a hydrological model by using model inputs and model errors. Frame *et al.* (2021) used long short-term memory (LSTM) methods to post-process the United States National Water Model (NWM) for 531 basins across the contiguous United States and found that the LSTM-based post-processing can improve the performance of NWM, with the average Nash–Sutcliffe efficiency (NSE) increasing from 0.62 to 0.73. Cho & Kim (2022) established an LSTM-based post-processing method by using the hydrometeorological inputs and simulated errors for Soyangho Lake in South Korea, and they concluded that the LSTM-based post-processing method can improve model performance, as measured by a higher NSE and correlation coefficient (CC). Among the ML methods, the support vector machine (SVM) (Cortes & Vapnik 1995) has attracted a great deal of interest (Han *et al.* 2007). Furthermore, the support vector regression (SVR) derived from the SVM was used to solve regression problems. The SVR algorithm has been improved and applied in many fields, and it is also popular in the hydrological field (Zhang *et al.* 2018). It has been used for streamflow forecasting in many regions (Wang *et al.* 2009; Liang *et al.* 2018). As the deep network approaches enable more reliable forecasting results, one deep learning method, i.e. convolutional neural network (CNN), is adopted in this work. A CNN is originally developed to analyze images and videos (Lecun *et al.* 1998). It has been applied in varied fields, including floods, landslides, remote sensing, and voice recognition (Sameen *et al.* 2020). To the best of our knowledge, the application of CNN is still lacking in post-processing.

Due to the complexity and resemble chaotic motion of the hydrological processes, there are nonlinear characteristics in streamflow. Thus, it is necessary to decompose the streamflow before applying it in post-processing. Decomposition is used to extract the trends of the streamflow. At present, decomposition preprocessing has been widely used in streamflow forecasting (Fang *et al.* 2019). The methods used for this preprocessing include empirical model decomposition (EMD), ensemble empirical mode decomposition (EEMD), and variation mode decomposition (VMD). Some researchers have found that the application of decomposed streamflow can generate more accurate streamflow forecasting than the application of the original streamflow (Zuo *et al.* 2020). As a new multi-resolution technique, VMD has better noise and sampling robustness than EMD and EEMD and can overcome the shortcomings of EMD (Fang *et al.* 2019). It has been used successfully in hydrological studies (He *et al.* 2020). Therefore, VMD is used here to decompose the simulated streamflow into different streamflow modes at different frequencies.

In this study, four traditional statistical post-processing methods based on QM and two proposed machine learning (SVR and CNN) methods are applied to a distributed model (WRF-Hydro) to reduce systematic biases in streamflow simulation, and the results are compared and evaluated. To the best of our knowledge, there are few studies that apply the post-processing method to calibrated and uncalibrated models separately. In this research, the CNN- and SVR-based post-processing methods are established based on the decomposed simulated streamflow and meteorological variables. These six post-processing methods are applied to five basins across South China. The post-processing framework is applied to both calibrated and uncalibrated models, and the differences between these two cases are evaluated. The differences between the traditional QM methods and the new machine learning (SVR and CNN) based methods are also compared.

STUDY AREA AND DATA

Study area

The study region is the Xijiang River basin, which is the largest watershed of the Pearl River basin. It contributes to the economic development of South China and meets most of the water supply demands of the Pearl River Delta. The climate in the basin is hot and humid, with the annual mean temperature ranging from 19 to 22 °C and the annual mean precipitation ranging from 1,200 to 2,200 mm. The precipitation and streamflow of this basin exhibit seasonal variability due to the monsoon rainfall. The period from April to September is the flood period, with 65% of the annual rainfall and 75% of the annual streamflow.

The river originates from Maxiong Mountain, travels through the provinces of Yunnan, Guizhou, Guangxi, and Guangdong, and finally flows into the South Sea. It is 2,214 km long with an area of 353,100 km². The topography of the basin is shown in Figure 1. This study covers the mainstream of the Xijiang River basin (where Gaoyao station is located) and its four sub-basins, including the Beipanjiang River (Gaoche station), the Mabie River (Maling station), the Liujiang River (Shihuichang and Liuzhou stations), and the Guijiang River (Guilin and Pingle stations).

Datasets

The Digital Elevation Model (DEM) and the China Meteorological Forcing Dataset (CMFD) are used to drive the offline WRF-Hydro model. Reanalysis data (ERA5) are used to provide the initial conditions for WRF-Hydro. The precipitation, temperature, evaporation, relative humidity, and wind from CN05.1 are used in the predictor input selection for the post-processing.

The DEM data are from the HydroSHEDS (Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales) dataset (Lehner *et al.* 2008), which was developed based on the high-resolution elevation data of the Shuttle Radar Topography Mission. This dataset has been widely used in hydrological and landslide studies (Wang *et al.* 2016). The resolution of 15" (about 500 m) is used in this research. The CMFD, developed by the Institute of Tibetan Plateau Research, shows reasonable consistency with observational data (Chen *et al.* 2011) and has been applied in hydrological and land models as well as in assessing the impacts of climate change in China. The 3-hourly CMFD data start from 1979 with a spatial resolution of 0.1° × 0.1°. The ERA5 reanalysis dataset, developed by the European Center for Medium Range Weather Forecasts, is capable of reasonably representing meteorological and land-surface variabilities (Hersbach *et al.* 2020). The 6-hourly ERA5 data with a spatial resolution of 30 km are employed in this research. The CN05.1 dataset is gridded observational data developed from over 2,400 observation stations in China and has shown good performance (Wu & Gao 2013). The temporal resolution of the data is 1 day and the spatial resolution is 0.25°.

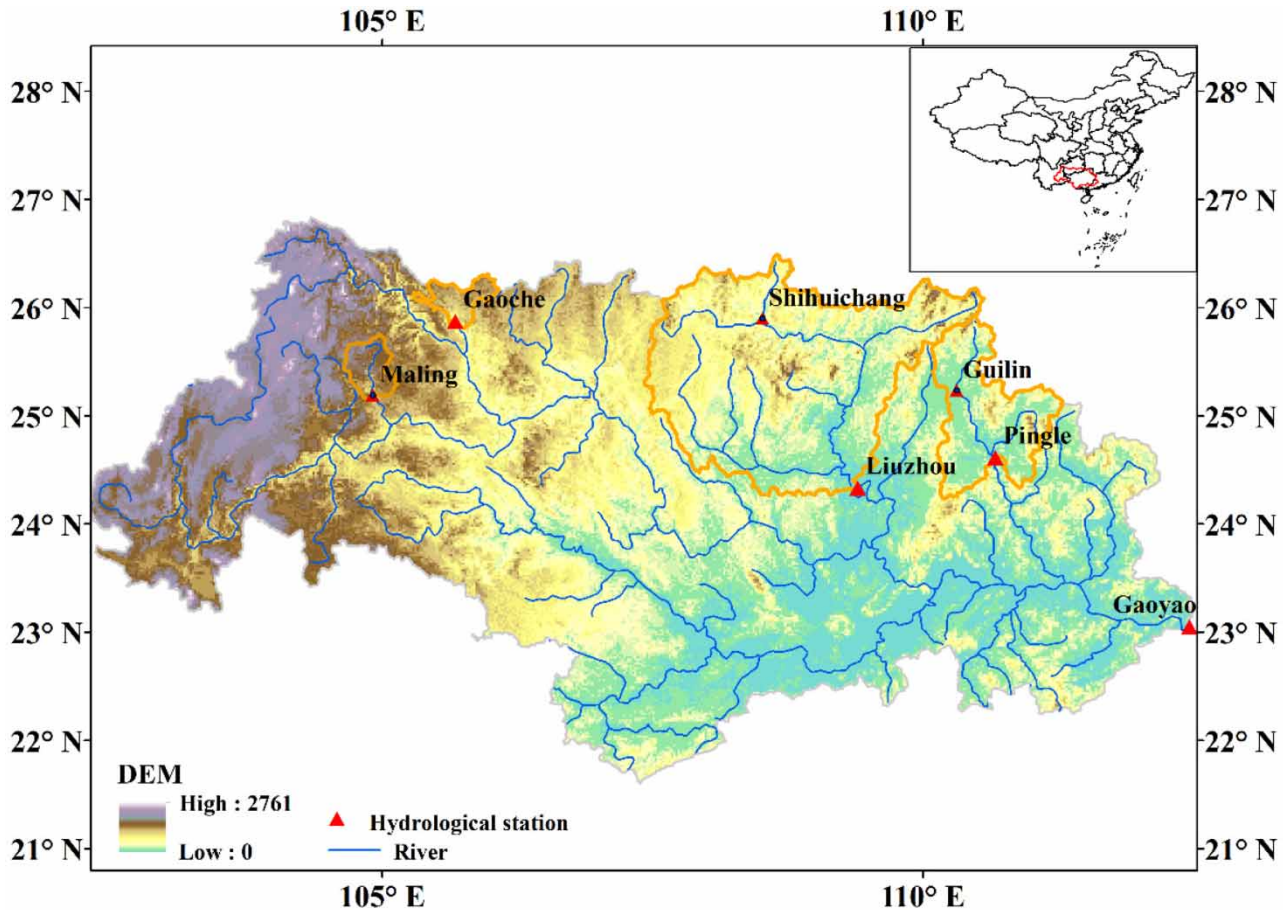


Figure 1 | Location and topography of the Xijiang River basin above the Gaoyao station, where red triangles represent hydrological stations, orange lines delineate sub-basins, and blue lines denote the river network. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2022.004>.

The daily streamflow data during 2006–2014 used in this study are from the published National Hydrological Yearbook of China. The data during the periods 2006–2010 and 2011–2014 are used for model calibration and model verification, respectively.

METHODS

Model

WRF-Hydro is a distributed hydrological model that couples routing modules (baseflow, saturated subsurface flow, overland flow, and channel) into the Noah land-surface model (Noah-LSM) with kinetic and thermal processes (Gochis *et al.* 2018). The resolution between Noah-LSM and the routing modules is flexible, which allows a high resolution for the routing processes. The model has been used for a wide range of projects, including flood prediction, water resources forecasting, and land-atmosphere coupling studies (Kerandi *et al.* 2018). WRF-Hydro is both a hydrological model (a so-called offline model) and a coupling architecture for coupling hydrological models with atmospheric models. Due to the complexity of the model, its calibration consumes a lot of time and computational resources. The offline WRF-Hydro (version 5) model is used in our research. The spatial resolution of Noah-LSM is 5,000 m, and in the routing processes it is 500 m in this research. The surface routing, subsurface routing, channel routing, and baseflow routing modules are all activated. The meteorological forcing data are mainly from the CMFD, while U and V wind data are from ERA5 as there is no wind direction data in the CMFD. The initial condition file is generated by WRF, and the high-resolution condition file is generated by ArcGIS. The generation details can be found in Gochis *et al.* (2018). The model is manually calibrated, and the details of the calibrated parameters are given in Table 1. For each sub-basin, one hydrological station is chosen for the calibration

Table 1 | Calibration parameters and their relevant details

Parameters	Description	Main hydrological response	Range
REFKDT	Parameter in surface runoff	Partitioning of total runoff into surface and subsurface runoff	0.1–5
LKSATFAC	Multiplier on saturated soil lateral conductivity	Routing/interflow process	10–10,000
RETDEPRTFAC	Maximum retention depth	Routing/interflow process	0.1–10
SLOPE	Slope index	Aquifer recharge	0.1–1
OVROUGHRTFAC	Manning's roughness coefficient	Routing/interflow process	0–1
Manning	Manning's roughness coefficient	Routing/interflow process	Depend on the channel order
Expon	Parameters of the GW model	Baseflow	
Z_{\max} C			
SMCMAX	Saturated soil moisture	Infiltration	Depend on the soil type
DKSAT	Saturated soil hydraulic conductivity	Infiltration	Depend on the soil type

and a set of parameters is used. The Gaoche station is used for the calibration of the Beipanjiang River basin and the Mabie River basin; the Liuzhou station is used for the calibration of the Liujiang River basin; the Pingle station is used for the calibration of the Guijiang River basin.

Post-processing methods

In general, the QM method is used as a post-processing method to correct model biases. The method is based on the cumulative distribution function (CDF), and it uses the CDF of observation to correct the CDF of the simulation. QM is formulated as follows:

$$x_{bc} = F(x_m) \quad (1)$$

where F is the transfer function, x_m is the original output of the model, and x_{bc} is the bias-corrected value.

QM employs transformations to correct the simulations. A series of transformation functions are used in QM, including parametric and nonparametric transformation functions. Two parametric transformation functions [PTF1 (1) and PTF2 (2)] are used in this research, the formulae of which are as follows:

$$P_o = a + bP_m \quad (2)$$

$$P_o = (a + bP_m) \left[1 - e^{-\frac{P_m}{\tau}} \right] \quad (3)$$

Here, P_o and P_m are the empirical cumulative distributions of the observation and model simulation during the training period; and a , b , and τ are coefficients.

Besides these two parametric transformation functions, two nonparametric transformation functions are used: RQUANT and SSPLIN. The principle of these methods is to make the empirical CDF of the model simulation close to the empirical CDF of the observation in the training period.

In this study, CNN- and SVR-based streamflow post-processing frameworks are developed, which consist of the following steps: (1) streamflow preprocessing. The simulated daily streamflow is decomposed into several streamflow modes at different frequencies. (2) Predictor input selection. By comparing the observed streamflow and other variables using the cross-correlation calculation, the predictors that have also passed the t -test at the 0.01 significance level are selected. (3) Streamflow post-processing analysis. The SVR-based post-processing or CNN-based post-processing method is performed by using the selected predictors as inputs. The outputs are assessed by hydrography and statistical metrics. The post-processing framework is shown in Figure 2.

VMD is a signal processing technique proposed in 2014, which can decompose a complicated signal into several intrinsic model functions (IMFs) (Dragomiretskiy & Zosso 2014). The purpose of VMD is to decompose a signal into a discrete

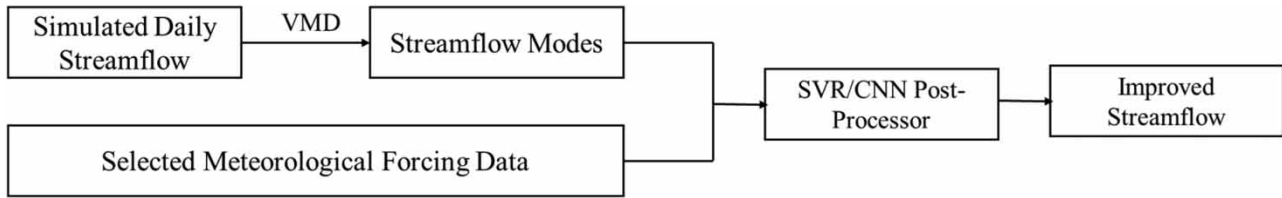


Figure 2 | Flow chart of the post-processing framework.

number of band-limited IMFs. VMD is an adaptive, completely nonrecursive modal variational and signal processing method. This technology has the advantage of being able to determine the number of modal decompositions. Its adaptability lies in determining the number of modal decompositions of a given sequence according to the actual situation. The subsequent search and solution process can adaptively match the optimal center frequency and limited bandwidth of the number of modal decompositions. And it can achieve effective separation of IMFs, signal frequency domain division, and then obtain effective decomposition components of a given signal, and finally obtain the optimal solution of the variational problem. The details of the VMD process can be found in [Dragomiretskiy & Zosso \(2014\)](#).

The SVR used for regression is based on the SVM. The methodology between SVR and SVM has a slight difference. The SVM belongs to a supervised classification technique which is a linear classifier ([Vapnik 1999](#)), while the SVR projects the inputs to a high dimension and then constructs the linear decision function in a high-dimensional space to achieve linear regression. Essentially, the SVR tries to find a function $f(x)$ that can combine the relationship between the outputs (y) and inputs (x), as follows:

$$f(x) = (w \cdot \varphi(x)) + b \quad (4)$$

where w is the weight vector, b is the offset factor, and $\varphi(x)$ is the nonlinear mapping function that can transform the input to a high-dimensional feature. There are several kernel functions in SVR, including linear, polynomial, and radial basis function (RBF). In our research, RBF is used.

The CNN is a neural network first created by [Lecun et al. \(1998\)](#). Compared with traditional neural networks, CNN benefits from the properties of natural signals, including local connections, shared weights, pooling, and the usage of multiple layers. There are a series of layers in the CNN framework, i.e., the first convolutional layers, the second pooling layers, and then the fully connected layers. In the convolutional layers, feature maps are structured by units, and then a nonlinearity activation function (ReLU) is used. The feature extracted from the convolutional layers is passed to the pooling layers. The pooling function statistically summarizes the feature values of a certain position of the plane and its adjacent position, and then uses the summarized result as the value of this position in the plane. In this study, the features are relatively small, so the pooling layers are not used.

Statistical metrics

In this study, four statistical metrics are used to evaluate the performance of the post-processing methods: CC, NSE, root mean square error (RMSE), and relative error (Bias) [Equations (5)–(8), respectively]. CC measures the linear relation between two variables and ranges from -1 to 1 , in which a CC of 1 (-1) shows a perfect positive (negative) correlation. The NSE is used to evaluate the biases of the hydrological model. It ranges from $-\infty$ to 1 , in which a value of 1 corresponds to a perfect match of the modeled discharge to the observation. The closer the value is to 1 , the more accurate the model. A value closes to 0 suggests that the model can simulate the average condition without giving details. A negative value indicates a poor model performance. RMSE measures the error between the predicted and observed variables and ranges from 0 to ∞ . A value of 0 indicates a perfect fit to the data. Bias refers to either overestimation (positive bias) or underestimation (negative bias) of a parameter and ranges from $-\infty$ to ∞ .

$$CC = \frac{n \sum Q_{obs,i} Q_{mod,i} - \sum Q_{obs,i} \sum Q_{mod,i}}{\sqrt{n \sum Q_{obs,i}^2 - (\sum Q_{obs,i})^2} \sqrt{n \sum Q_{mod,i}^2 - (\sum Q_{mod,i})^2}} \quad (5)$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (Q_{\text{obs},i} - Q_{\text{mod},i})^2}{\sum_{i=1}^n (Q_{\text{obs},i} - Q_{\text{obs},i}^*)^2} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Q_{\text{obs},i} - Q_{\text{mod},i})^2}{n}} \quad (7)$$

$$\text{Bias} = \frac{\sum (Q_{\text{mod},i} - Q_{\text{obs},i})}{\sum Q_{\text{obs},i}} \times 100\% \quad (8)$$

Here, $Q_{\text{obs},i}$ is the i th observed streamflow, $Q_{\text{mod},i}$ is the i th simulated streamflow, $Q_{\text{obs},i}^*$ is the temporal average of the observed streamflow, and n is the number of days.

RESULTS

Selection of predictors

The input data are essential for SVR and CNN. In this section, the inputs (predictors) are chosen for the SVR- and CNN-based post-processing methods. The streamflow is the result of both hydrological processes and meteorological forcing. It is closely related to meteorological data, such as temperature and precipitation. Therefore, several meteorological variables are considered, including precipitation, temperature, evaporation, relative humidity, and wind. The meteorological data for selection in this study are from CN05.1 during 2006–2010. Precipitation plays an important role in streamflow forecasting and simulation. As the upstream precipitation can also influence the downstream streamflow, the better one out of the upstream accumulated precipitation and the station precipitation is selected as the predictor. The streamflow time-series contains temporal information, so different time-lags of different predictors are considered. Therefore, the lag correlations between the observed streamflow and other predictors are calculated over the previous 1–30 days.

Figure 3 shows the CC between the observed streamflow and meteorological data at the Liuzhou station corresponding to the lag lengths (1–30 days). The black line indicates the critical CC value at the 0.01 significance level. A CC value between two black lines means that it does not pass the 1% significance test. Both the precipitation at hydrological stations (Pre_PN) and the accumulated precipitation over the upstream of the hydrological stations (Pre_US) are tested. Figure 3(a) shows that the CC of Pre_US is higher than that of Pre_PN, which suggests a larger influence of upstream accumulated precipitation on the station streamflow. Thus, Pre_US is chosen for the streamflow post-processing. Similar results are found at other stations, and the differences between Pre_US and Pre_PN for small watersheds are less significant than for large watersheds. At the Liuzhou station, the 1-day lag presents the largest CC, so the 1-day lag precipitation is chosen. Temperature and relative humidity are selected because of their high linear relations with the observed streamflow, while wind and evaporation are excluded due to the opposite situation. Different time-lags are chosen for different inputs and stations. The CC between the Pre_US and the observed streamflow for lags of 1–10 days at all stations is shown in Figure 3(b). The sequence of the

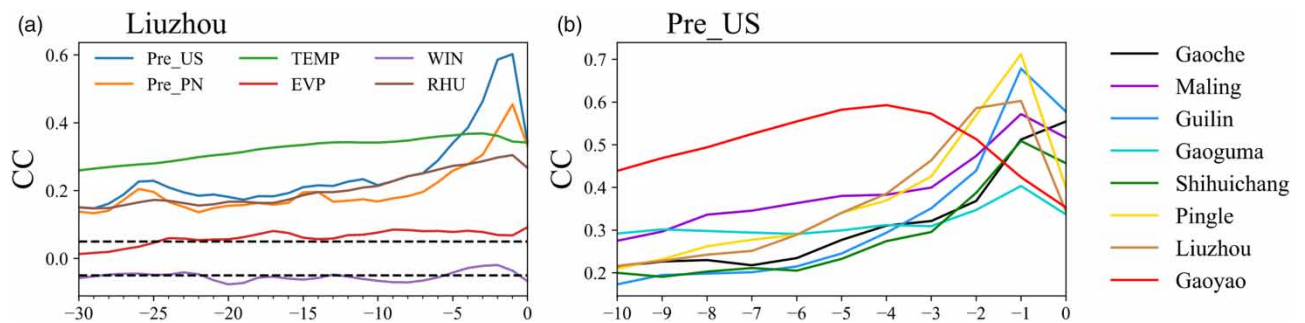


Figure 3 | (a) CC between the observed streamflow and the meteorological data at the Liuzhou station with lag lengths of 0–30 days. (b) CC between the observed streamflow and upstream precipitation at all stations with lag lengths of 0–10 days.

stations in the legend corresponds to the area of the watershed in an ascending order. The time-lag is smaller for small watersheds and larger for large watersheds. The best time-lag is 1 day for all stations except Gaoche and Gaoyao. As the area of the basin in which the Gaoche station is located is small, the best time-lag for this station is 0, which means that its streamflow is influenced by the precipitation that day. The best lag for the Gaoyao station is as long as 4 days due to the large watershed.

Through the decomposition techniques, the extracted periodic oscillations and trends are helpful in revealing the inside driving forces of streamflow. Therefore, a decomposition technique, VMD, is used to decompose the series of simulated streamflow. All time-series are decomposed by VMD into three IMFs that produce the best results in our tests.

In post-processing, 2006–2010 is used as the training/calibration period and 2011–2014 as the test/verification period. The establishment of the SVR- and CNN-based post-processing methods is based on the selected predictors and simulated streamflow. The four QM post-processing methods are based on the simulated streamflow. The influences of the six post-processing methods on uncalibrated and calibrated models are evaluated separately.

Post-processing the uncalibrated model

Both the calibration and post-processing can show the temporal change in streamflow correctly during calibration and verification periods that are high in summer and low in winter. Figure 4 shows the streamflow hydrography of the uncalibrated model, calibrated model, CNN-based post-processing method, SVR-based post-processing method, two parametric QM post-processing methods, two nonparametric QM post-processing methods, and the observation in 2006 at Maling, Liuzhou, Pingle, and Gaoyao stations. These four stations represent the different parts of the Xijiang River basin (Figure 1). From the figure, it can be seen that both calibration and post-processing can reduce the biases of the uncalibrated model, and show the temporal change in streamflow correctly, albeit with differences in details. The streamflow of the post-processing is smaller than that of the calibrated model in most flood periods, and this phenomenon can be found at most stations. The differences between these six post-processing methods are large, especially for streamflow in winter, when the variability of streamflow is small. CNN- and SVR-based post-processing methods can show the small fluctuations in winter at the Maling station, but the QM post-processing methods cannot, and this phenomenon is also apparent in other years. The differences

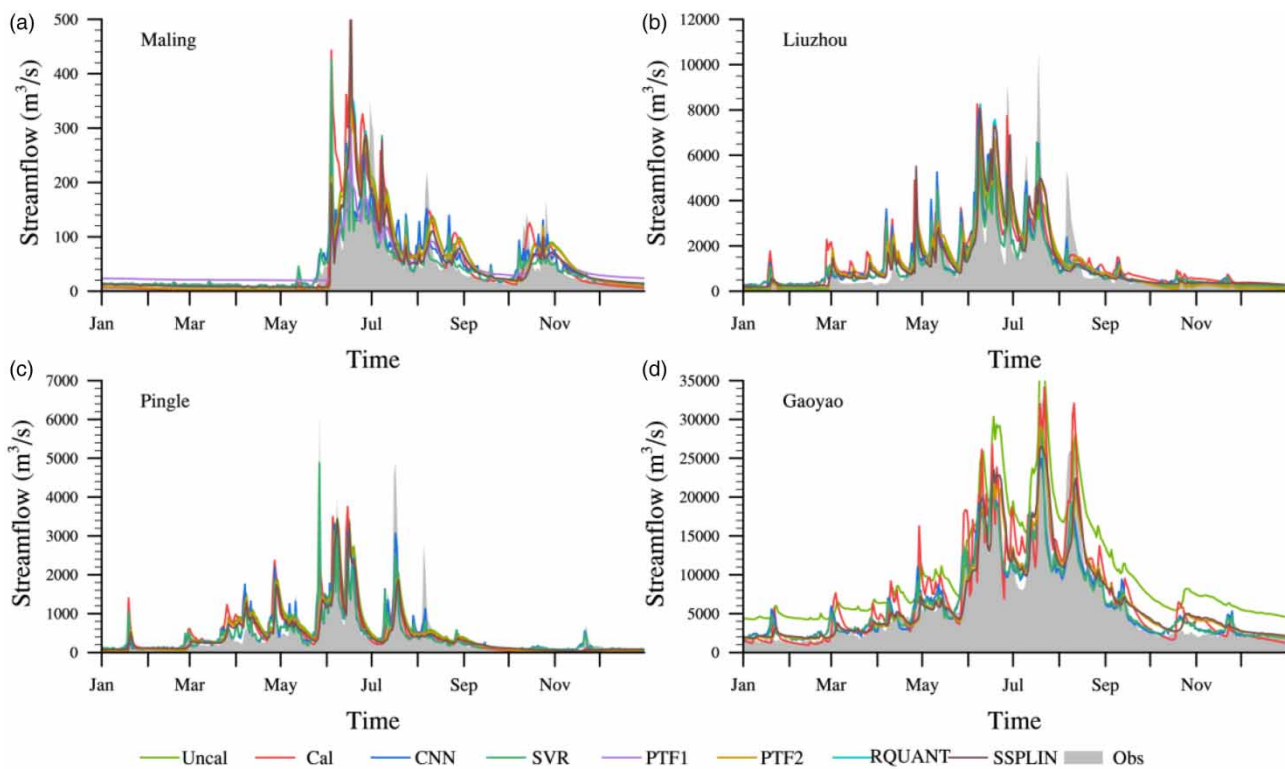


Figure 4 | Hydrography of streamflow from the observation (Obs), calibrated model (Cal), uncalibrated model (Uncal), and six post-processing methods (CNN, SVR, PTF1, PTF2, RQUANT, and SSPLIN) at Maling, Liuzhou, Pingle, and Gaoyao stations in 2006.

between the CNN- and SVR-based post-processing are smaller compared with those between post-processing and calibration. The differences between the four QM post-processing methods are small. Calibration and post-processing reduce the biases of the model in different ways, and for different periods they show different performances. To summarize, both calibration and post-processing can simulate the seasonal change in streamflow; however, there are large differences in adjusting the magnitude of streamflow. As for the magnitude of streamflow, the SVR- and CNN-based post-processing methods perform better than the QM post-processing methods, especially for adjusting the streamflow in winter.

Figure 5 compares the streamflow residual distributions over different seasons among the different methods. To highlight the useful information, the residual values outside the range of 10–90% of all residual values are not shown in this figure. From this figure, it can be seen that the residuals have a temporal pattern at all stations for all methods. The residuals are large in summer, while they are small in other seasons. This may be influenced by precipitation, as it is higher in summer,

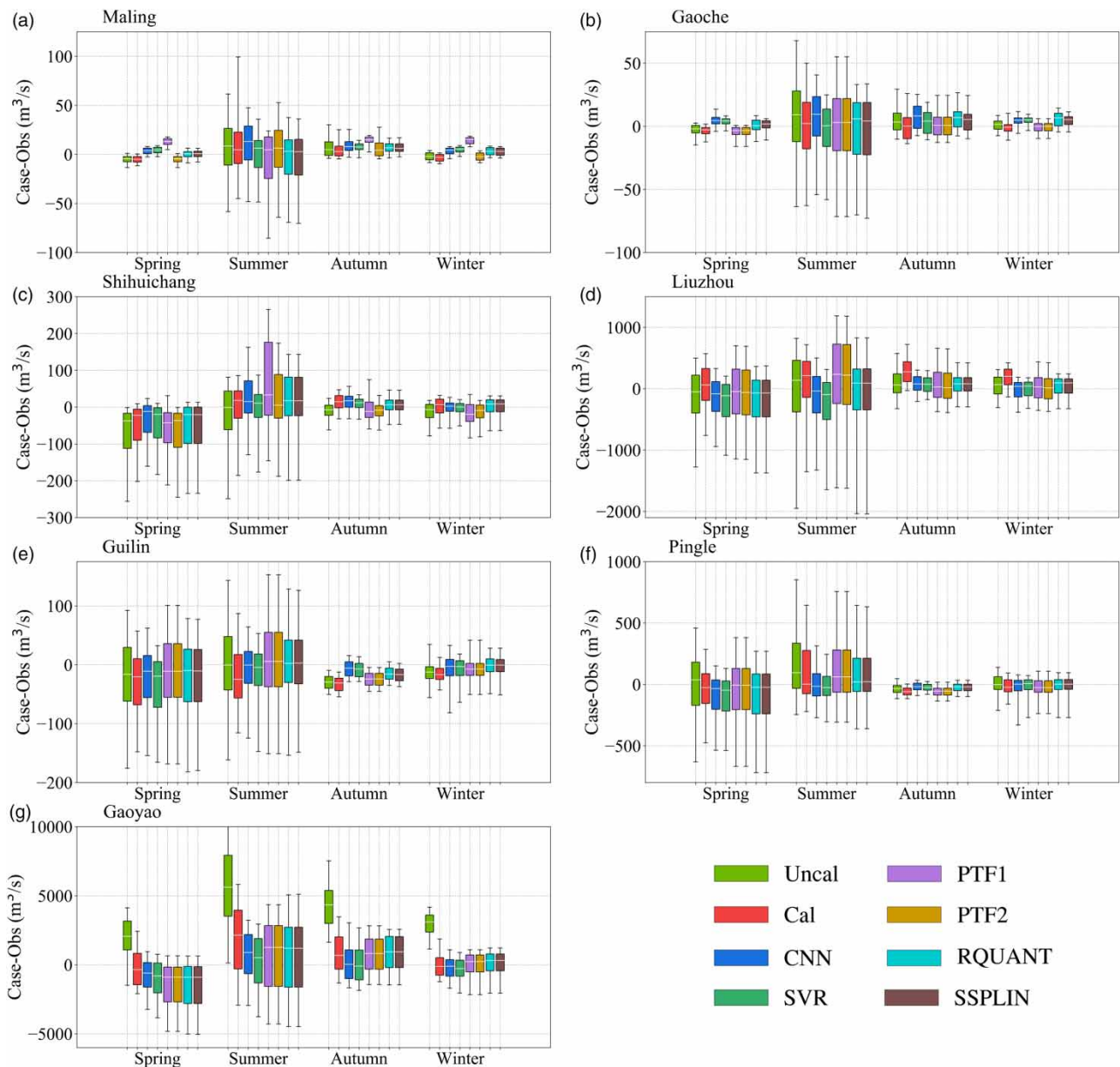


Figure 5 | Comparison of seasonal residuals for each case during the verification period. The boxplots of each season are, from left to right, the result from the uncalibrated model, the calibrated model, and the six post-processing methods (CNN, SVR, PTF1, PTF2, RQUANT, and SSPLIN).

which leads to more streamflow in this season. The residuals are smaller after calibration or post-processing. The abilities of the CNN- and SVR-based post-processing methods in reducing the residuals are similar to that of calibration, but even better than calibration at some stations, such as Pingle. However, the QM post-processing methods are good at some stations, and they are worse than calibration at most stations. Also, the performance of QM using parametric transformations is worse than that of nonparametric QM, and the difference between the two nonparametric QM methods is small.

The statistical parameters at seven stations for the uncalibrated model, the calibrated model, and the six post-processing methods in the verification period are shown in Figure 6. From the figure, it can be seen that calibration, CNN-based post-processing, and SVR-based post-processing can improve the model performance, with higher CC and NSE, and lower Bias and RMSE. The calibration and the SVR- and CNN-based post-processing methods improve the CC at all stations, which means that these methods enhance the model's ability in simulating the temporal change in streamflow. Among them, the CNN-based post-processing method is better than the SVR-based post-processing method at most stations. The NSE of the calibrated model and the SVR- and CNN-based post-processing methods are higher compared with the uncalibrated model at all stations. In terms of Bias, the CNN- and SVR-based post-processing methods are better than calibration at most stations, with the CNN-based post-processing method generating the lowest Bias. Calibration and post-processing produce lower RMSE compared with the uncalibrated model at all stations, and the RMSE of CNN-based post-processing method is the lowest at most stations. The differences in the statistical metrics between the CNN- and SVR-based post-processing methods are small, and the former is better than the latter at most stations. However, QM post-processing shows worse CC and NSE than the uncalibrated model at some stations, but the Bias of these methods is lower. The performances of RQUANT- and SSPLIN-based QM show little differences at almost all stations, and these methods are slightly worse than the PTF1- and PTF2-based QM methods at most stations. The CC and NSE of these four QM methods are lower than those of the calibrated model at most stations, which means that the performances of these methods are worse than model calibration. The performances of SVR- and CNN-based post-processing methods are better

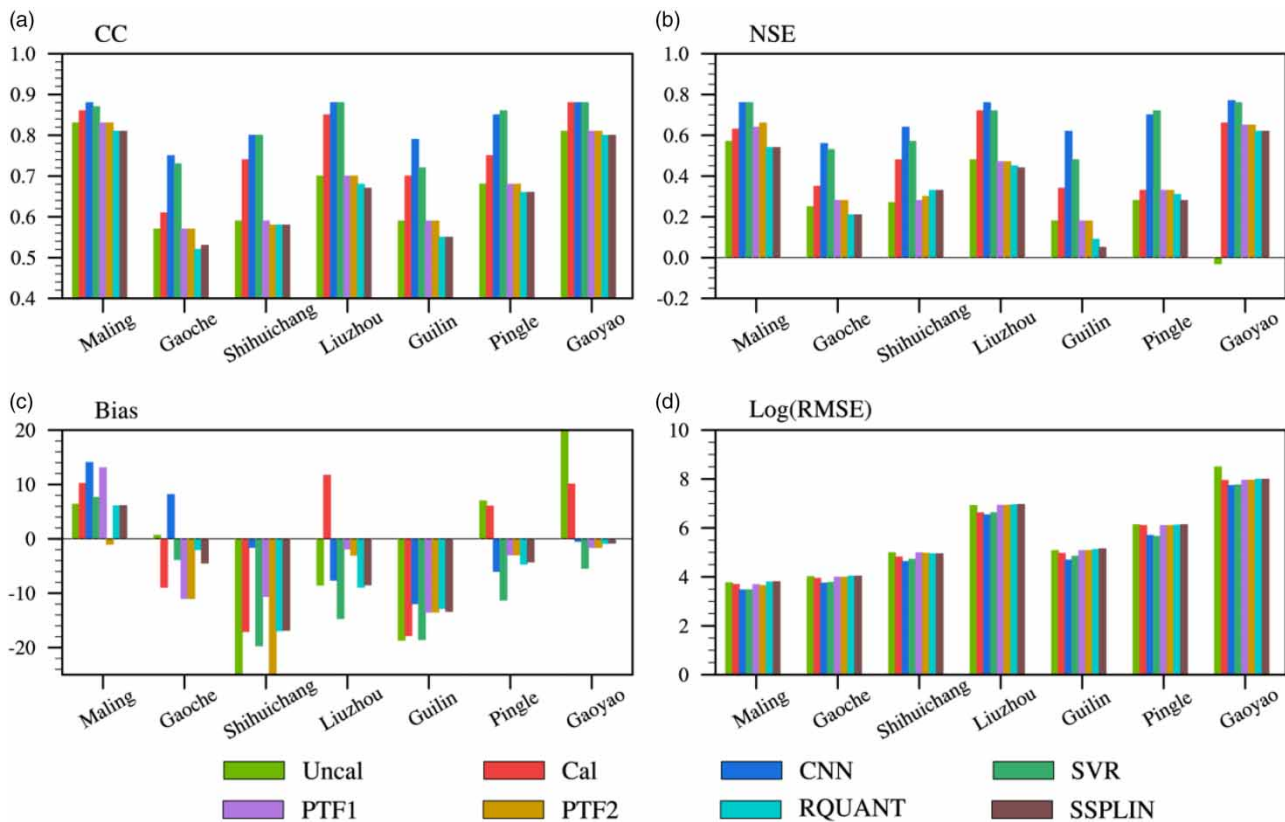


Figure 6 | Bar charts of the CC (a), NSE (b), Bias (c), and RMSE (d) of the uncalibrated model, the calibrated model, and the six post-processing methods at seven stations in the verification period. The different methods are represented by different colors. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2022.004>.

than those of the QM post-processing methods. Among them, the CNN-based post-processing method is the best one at almost all stations.

Figure 7 shows percentile plots of the uncalibrated streamflow, calibrated streamflow, and streamflow after post-processing. The figure shows the one-to-one match between the percentile of the observation and the percentile of the post-processed streamflow. The black line is the 45-degree line. The closer to the black line, the better the performance. A line above the black line means an overestimation of the streamflow, while a line below the black line means an underestimation of the streamflow. The CNN- and SVR-based post-processing methods tend to overestimate low streamflow at the Maling station, while it is underestimated by the calibrated model, the uncalibrated model, and the PTF2-based QM method. The PTF2-based QM post-processing method is the closest to the black line at the Gaoche station. At the Shihuichang station, the post-processing methods (except PTF1- and PTF2-based QM) and the calibrated model overestimate the streamflow that is lower than the 60% percentile of the observed streamflow. At the Liuzhou station, the post-processing methods (except PTF1- and PTF2-based QM) overestimate the streamflow that is lower than the 60% percentile of the observation, but are closer to the observation when the streamflow is higher than this percentile. Post-processing and calibration underestimate the streamflow that is lower than the 80% percentile of the observation at the Guilin station. At the Pingle station, when the streamflow is lower than the 60% percentile of the observation, uncalibrated and calibrated models underestimate the streamflow, whereas, when the streamflow is higher than this percentile of the observation, they overestimate the streamflow. The CNN- and SVR-based post-processing methods reduce the errors, bringing the result closer to the black line. At the Gaoyao station, the CNN- and SVR-based post-processing methods are closer to the black line than the calibrated model. In the percentile plots of the streamflow, the SVR- and CNN-based post-processing methods are better for some stations, such as Shihuichang, Liuzhou, Guilin, Pingle, and Gaoyao, while calibration is better at Maling and Gaoche stations. The QM method is established based on the CDF in the calibration period, and the results of these methods are away from the black line in the verification period.

This section evaluates the performance of all methods in improving the high streamflow (higher than 75% of the observed streamflow) and the low streamflow (lower than 25% of the observed streamflow). The NSE and RMSE of all methods are shown in Figure 8. It can be seen that calibration and post-processing (SVR and CNN) perform well for high streamflow and

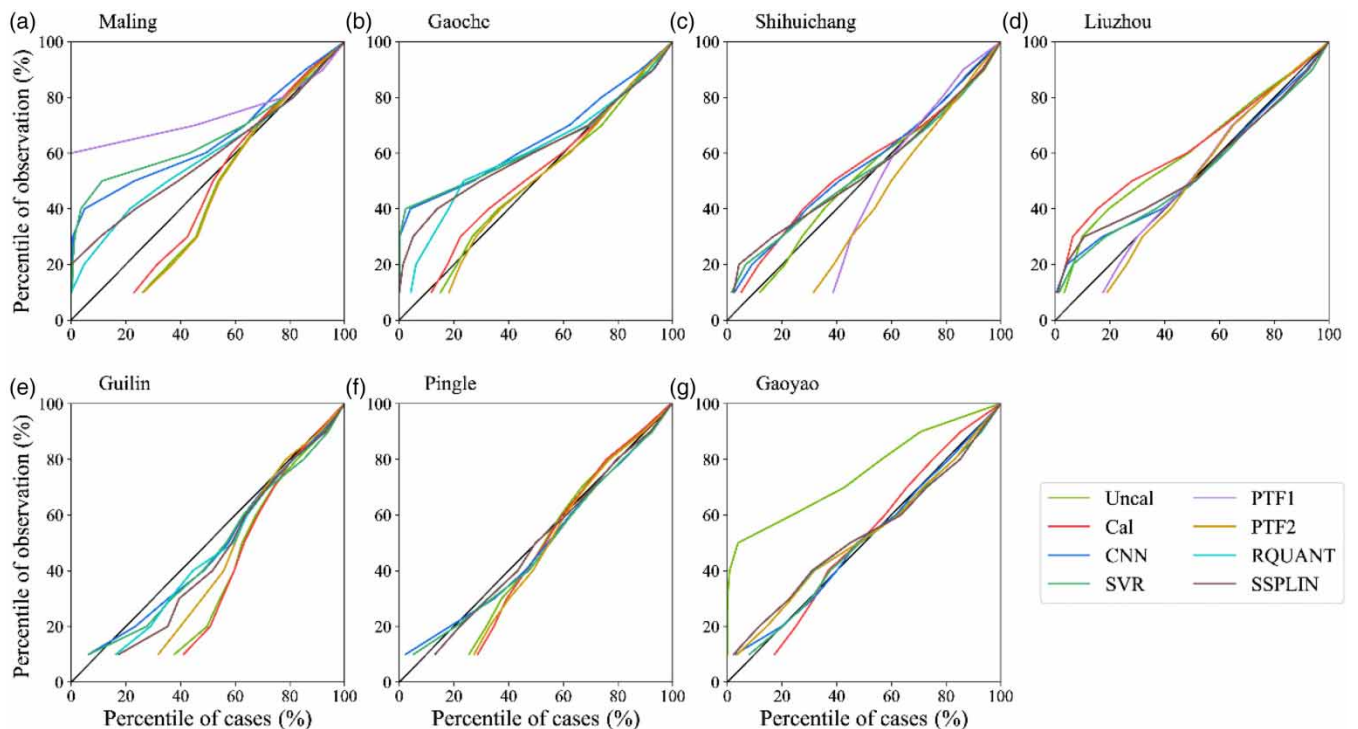


Figure 7 | Percentile plots of the uncalibrated model (Uncal), the calibrated model (Cal), and the six post-processing methods (CNN, SVR, PTF1, PTF2, RQUANT, and SSPLIN) in the verification period.

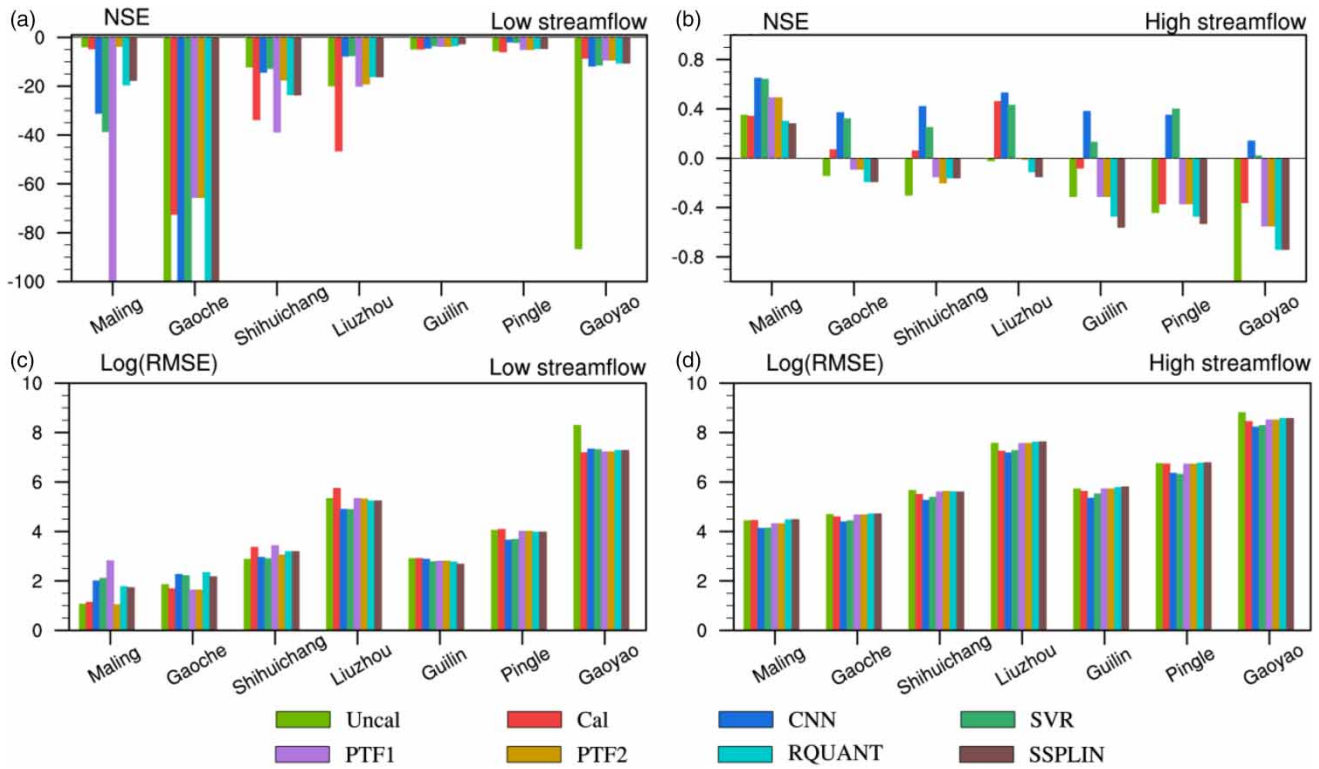


Figure 8 | Bar charts of NSE and RMSE for low streamflow (a, c) and high streamflow (b, d) simulated by the uncalibrated model, the calibrated model, and the six post-processing methods at seven stations in the verification period. The different methods are represented by different colors. Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/nh.2022.004>.

perform poorly in their simulation of low streamflow. For low streamflow, the CNN-based post-processing method is better than the calibration at Shihuichang, Liuzhou, Guilin, and Pingle stations. The CNN-based post-processing method is better than the calibration for high streamflow at all stations and is better than the SVR-based post-processing method at most stations. The performances of QM post-processing methods are poor. The SVR- and CNN-based post-processing methods may have the potential to correct the simulation and forecasting of floods.

Aside from the uncertainties of incorrect model parameters and the model framework, uncertainties in hydrological modeling also stem from the meteorological inputs. Therefore, this study also tests the differences in post-processing (SVR and CNN) with and without meteorological data inputs. From the statistical results (Table 2), it can be seen that the results of post-processing (SVR and CNN) without meteorological data inputs are worse compared with model calibration at some stations. The situation (the calibration performance is good in the calibration period, but unsatisfactory in the verification

Table 2 | Statistical metrics at different stations for the uncalibrated model

	CC			NSE			Bias (%)			RMSE (m ² /s)		
	Cal	CNN	SVR	Cal	CNN	SVR	Cal	CNN	SVR	Cal	CNN	SVR
Maling	0.86	0.87	0.86	0.63	0.72	0.71	10.19	3.36	-5.01	40.13	34.81	35.62
Gaoche	0.61	0.70	0.62	0.35	0.47	0.37	-8.9	4.61	-13.80	51.28	46.36	50.27
Shihuichang	0.74	0.74	0.73	0.48	0.54	0.43	-17.05	-5.89	-26.23	123.39	116.87	129.63
Liuzhou	0.85	0.84	0.81	0.72	0.70	0.63	11.69	-6.21	-14.95	748.44	772.82	855.88
Guilin	0.7	0.68	0.69	0.34	0.44	0.44	-17.8	-15.51	-26.19	142.86	131.55	131.63
Pingle	0.75	0.65	0.67	0.33	0.14	0.36	6.02	-7.03	-12.35	446.48	505.72	434.60
Gaoyao	0.88	0.87	0.85	0.66	0.74	0.72	10.09	-0.51	-6.58	2,509.4	2,435.80	2,534.77

period) in the Guijiang River basin is not improved. Therefore, the meteorological inputs are important for SVR- and CNN-based post-processing. Considering different sources of meteorological data is also important in post-processing, as it might reduce the model uncertainties of the meteorological inputs. This study also compares the results of post-processing with CMFD and CN05.1 datasets. It is found that the performance is better when using the CN05.1 dataset, which is different with the model input.

Post-processing the calibrated model

Six post-processing methods are applied to the calibrated model, and the statistical parameters of these six post-processing methods and the calibration are shown in Figure 9 [Methods with (w) represent that the post-processing is applied to the

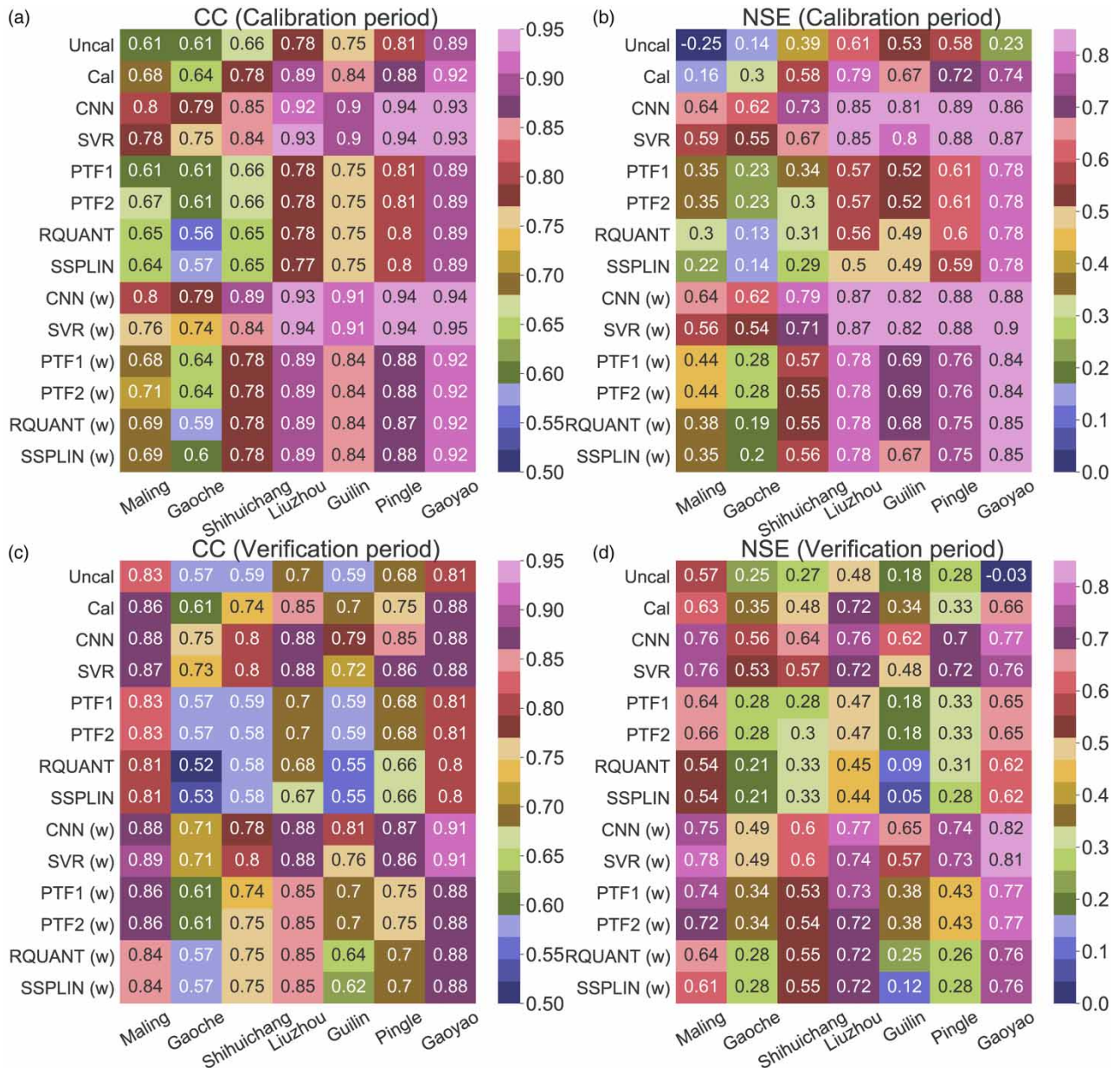


Figure 9 | CC and NSE of the uncalibrated model, the calibrated model, and the six post-processing methods [methods with (w) represent that the post-processing is applied to the calibrated model, while the others are applied to the uncalibrated model].

calibrated model]. From this figure, it can be seen that the QM methods show little influence on the CC, and the RQUANT- and SSPLIN-based QM methods reduce the CC at the Gaoche station. The four QM methods can improve the NSE for most stations, especially in the test period, and these methods can also reduce the Bias. The differences between these four kinds of QM methods are small, and the parametric QM methods are superior to the nonparametric QM methods in terms of CC and NSE. The nonparametric QM methods are unable to improve the calibrated model at Gaoche, Guilin, and Pingle stations. The CNN- and SVR-based post-processing methods have higher CC and NSE, and lower RMSE compared with the calibrated model. The CNN-based post-processing method is better than the SVR one at most stations, while both methods improve the performance of the calibrated model. Compared with the CNN-based post-processing method, the QM methods are worse. For the calibration, the NSE of the Guijiang basin (Guilin and Pingle stations) is satisfactory in the calibration period, but they are worse in the test period. And the QM methods show the same phenomenon. The CNN- and SVR-based post-processing methods are less affected by calibration results than the QM method. The CNN- and SVR-based post-processing methods are suitable for both the calibration and the test period.

The differences between the post-processing methods when applied to the uncalibrated model (post-processing without calibration) and the calibrated model (post-processing with calibration) are also evaluated. For the CC and NSE, the CNN- and SVR-based post-processing methods with calibration are superior to the same methods but without calibration at most stations. The CNN- and SVR-based post-processing methods with calibration are slightly better than without calibration, but the differences between them are small, which means that only using these methods is also acceptable. The results are suitable for different areas with different topographies in the Xijiang River basin, such as high elevation (Maling station) and low elevation (Liuzhou station). From the statistical parameters, it can be seen that the CNN- and SVR-based post-processing methods are adapted to both uncalibrated and calibrated models, but QM is unsuitable at some stations.

DISCUSSIONS

The performances of CNN- and SVR-based post-processing are good in both the training and test periods. Meanwhile, calibration is good in the training period but gets worse in the test period at Guilin and Pingle stations. In fact, another two stations in this basin were also tested, and the same results were found. However, the phenomenon does not exist in the CNN- and SVR-based post-processing.

Calibrating a hydrological model can be costly in terms of computational expense. For large basins where there is large variation in land use, the calibration performance may be good in some places but poor elsewhere. The calibrated parameters may be suitable in one basin, but when they are applied to other basins they may not be appropriate. Hydrological post-processing is an approach that can avoid these calibration problems across large watersheds. Compared with model calibration, post-processing consumes less time and computational resource. A number of parameters are considered during model calibration, and thus the model needs to be run many times. For a large basin (around 304,900 km²), it takes 14 h to run the WRF-Hydro model (land-surface module resolution: 5,000 m; routing module resolution: 500 m) once (5 years) with 48 CPUs. CNN- and SVR-based post-processing, meanwhile, only uses 2 min with 1 CPU to produce results. The selection of inputs and the post-processing construction may need some time, but the demand for computational resource and time is much less than the model calibration. Also, the CNN- and SVR-based post-processing methods are superior to traditional QM post-processing methods.

This study shows that CNN- and SVR-based post-processing methods can reduce model errors in the Xijiang River basin. However, these methods may perform differently in other basins. Therefore, further testing in more basins is needed before CNN- and SVR-based post-processing methods can be applied in practice.

CONCLUSIONS

In this study, two machine learning (SVR and CNN) post-processing methods were established based on simulated streamflow and meteorological data. The performances of these post-processing methods were compared with traditional QM methods and model calibration. Also, the post-processing methods were separately applied to calibrated and uncalibrated models. The training/calibration period was 2006–2010, and the test/verification period was 2011–2014. The results were

tested qualitatively and quantitatively by using hydrography, percentile plots, and statistical metrics. The main conclusions can be summarized as follows.

The inputs (meteorological predictors) of the CNN- and SVR-based post-processing methods were determined by using the time-lagged CC. Besides, the VMD-decomposed streamflow was also used to explain the nonlinear characteristics of the streamflow. Precipitation, temperature, and relative humidity were chosen as the inputs for the post-processing because of their high CC. The contribution of the accumulated upstream precipitation to the station streamflow was found to be larger than that of station precipitation, especially for large basins. The time-lag was different for different inputs and stations, and generally, it was longer for larger watersheds.

For the uncalibrated model, the performances of the CNN- and SVR-based post-processing methods were equivalent to the performance of model calibration, and they performed better than the QM post-processing methods. The CNN- and SVR-based post-processing methods performed well in the training and test periods. They improved the temporal variation and magnitude of the simulated streamflow. The differences between the SVR-based post-processing methods and the CNN-based post-processing methods were small, and the latter performed better in most areas. The CNN- and SVR-based post-processing methods also improved the results of the calibrated model, with higher CC and NSE, and lower RMSE. The CNN- and SVR-based post-processing methods were found to be suitable for calibrated and uncalibrated models. Compared with the QM methods, the CNN- and SVR-based methods were found to be more stable. The QM methods were able to reduce the biases effectively, but they could not improve the CC or NSE at some stations.

Overall, the CNN- and SVR-based post-processing methods improved the uncalibrated and calibrated model simulations for both small and large watersheds in the Xijiang River basin with the WRF-Hydro model. The performances of CNN- and SVR-based post-processing methods when applied to the uncalibrated model were found to be similar to that of model calibration in the Xijiang River basin. These post-processing methods have the potential to be used in flood simulation and forecasting, but its performance in modifying low streamflow needs to be improved. CNN- and SVR-based post-processing methods have been applied here to the Xijiang River basin, but should be tested further in other basins. Although post-processing may not give a clear physical meaning, it can be used in practice to provide reliable results.

ACKNOWLEDGEMENTS

This work was supported by the National Nature Science Foundation of China (Grant Nos 42088101 and 42175170).

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

- Bogner, K. & Kalas, M. 2008 Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmospheric Science Letters* **9** (2), 95–102.
- Carpenter, T. M. & Georgakakos, K. P. 2006 Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology* **329** (1), 174–185.
- Chen, Y. Y., Yang, K., He, J., Qin, J., Shi, J. C., Du, J. Y. & He, Q. 2011 Improving land surface temperature modeling for dry land of China. *Journal of Geophysical Research Atmospheres* **116** (20), 1–15.
- Cho, K. & Kim, Y. 2022 Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *Journal of Hydrology* **605**, 127297.
- Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297.
- Dragomiretskiy, K. & Zosso, D. 2014 Variational mode decomposition. *IEEE Transactions on Signal Processing* **62** (3), 531–544.
- Fang, W., Huang, S. Z., Ren, K., Huang, Q., Huang, G. H., Cheng, G. H. & Li, K. L. 2019 Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *Journal of Hydrology* **568** (1–2), 534–550.
- Frame, J. M., Kratzert, F., Raney, A., Rahman, M., Salas, F. R. & Nearing, G. S. 2021 Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. *Journal of the American Water Resources Association* **57** (6), 885–905.
- Gochis, D. J., Barlage, M., Dugger, A., Fitzgerald, K., Karsten, L., Mcallister, M., Mccreight, J., Mills, J., Rafieeiniasab, A., Read, L., Sampson, K., Yates, D. & Yu, W. 2018 WRF-Hydro V5 technical description. In: NCAR Technical Note. Center for Atmospheric Research (NCAR), Boulder, CO, USA.
- Han, D., Chan, L. & Zhu, N. 2007 Flood forecasting using support vector machines. *Journal of Hydroinformatics* **9** (4), 267–276.

- Hashino, T., Bradley, A. A. & Schwartz, S. S. 2007 Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences* **11** (2), 939–950.
- He, X. X., Luo, J. G., Li, P., Zuo, G. G. & Xie, J. C. 2020 A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. *Water Resources Management* **34** (2), 865–884.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J. N. 2020 The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146** (730), 1999–2049.
- Hsu, K.-L., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research* **31** (10), 2517–2530.
- Kerandi, N., Arnault, J., Laux, P., Wagner, S., Kitheka, J. & Kunstmann, H. 2018 Joint atmospheric-terrestrial water balances for East Africa: a WRF-Hydro case study for the upper Tana River basin. *Theoretical and Applied Climatology* **131** (3), 1337–1355.
- Kim, K. B., Kwon, H.-H. & Han, D. 2021 Bias-correction schemes for calibrated flow in a conceptual hydrological model. *Hydrology Research* **52** (1), 196–211.
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. 1998 Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86** (11), 2278–2324.
- Lehner, B., Verdin, K. L. & Jarvis, A. 2008 New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union* **89** (10), 93–94.
- Li, W. H. & Sankarasubramanian, A. 2012 Reducing hydrologic model uncertainty in monthly streamflow predictions using multimodel combination. *Water Resources Research* **48** (12), W12516.
- Liang, Z. M., Li, Y. J., Hu, Y. M., Li, B. Q. & Wang, J. 2018 A data-driven SVR model for long-term runoff prediction and uncertainty analysis based on the Bayesian framework. *Theoretical and Applied Climatology* **133** (1–2), 137–149.
- Lucatero, D., Madsen, H., Refsgaard, J. C., Kidmose, J. & Jensen, K. H. 2018a Seasonal streamflow forecasts in the Ahlgergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical consistency. *Hydrology and Earth System Sciences* **22** (7), 3601–3617.
- Lucatero, D., Madsen, H., Refsgaard, J. C., Kidmose, J. & Jensen, K. H. 2018b On the skill of raw and post-processed ensemble seasonal meteorological forecasts in Denmark. *Hydrology and Earth System Sciences* **22** (12), 6591–6609.
- Madadgar, S., Moradkhani, H. & Garen, D. 2012 Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes* **28** (1), 104–122.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. & Prabhat 2019 Deep learning and process understanding for data-driven Earth system science. *Nature* **566** (7), 195–204.
- Rezaie-Balf, M., Nowbandegani, S. F., Samadi, S. Z., Fallah, H. & Alaghmand, S. 2019 An ensemble decomposition-based artificial intelligence approach for daily streamflow prediction. *Water* **11** (4), 709–738.
- Robertson, D. E., Shrestha, D. L. & Wang, Q. J. 2013 Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrology and Earth System Sciences* **17** (9), 3587–3603.
- Sameen, M. I., Pradhan, B. & Lee, S. 2020 Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. *Catena* **186** (8), 104249–104261.
- Shi, X. G., Wood, A. W. & Lettenmaier, D. P. 2008 How essential hydrologic model calibration to seasonal stream flow forecasting? *Journal of Hydrometeorology* **9** (6), 1350–1363.
- Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics* **10** (1), 3–22.
- Vapnik, V. N. 1999 An overview of statistical learning theory. *IEEE Transactions on Neural Networks* **10** (5), 988–999.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. & Robinson, B. A. 2008 Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research* **44** (12), W00B09.
- Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374** (3–4), 294–306.
- Wang, J., Wang, H. J. & Hong, Y. 2016 A high-resolution flood forecasting and monitoring system for China using satellite remote sensing data (in Chinese). *Chinese Science Bulletin* **61** (4–5), 518–528.
- Wu, J. & Gao, X. J. 2013 A gridded daily observation dataset over China region and comparison with the other datasets (in Chinese). *Chinese Journal of Geophysics* **56** (4), 1102–1111.
- Wu, R., Yang, L., Chen, C., Ahmad, S., Dascalu, S. M. & Harris Jr., F. C. 2019 MELPF version 1: modeling error learning based post-processor framework for hydrologic models accuracy improvement. *Geoscientific Model Development* **12**, 4115–4131.
- Xu, J., Ancil, F. & Boucher, M. A. 2019 Hydrological post-processing of streamflow forecasts issued from multimodel ensemble prediction systems. *Journal of Hydrology* **578**, 124002.
- Ye, A. Z., Duan, Q. Y., Schaake, J., Xu, J., Deng, X. X., Di, Z. H., Miao, C. Y. & Gong, W. 2015 Post-processing of ensemble forecasts in low-flow period. *Hydrological Processes* **29** (10), 2438–2453.

- Yuan, X. & Wood, E. F. 2012 Downscaling precipitation or bias-correcting streamflow? Some implications for coupled general circulation model (CGCM)-based ensemble seasonal hydrologic forecast. *Water Resources Research* **48** (12), W12519.
- Zhang, D., Lin, J. Q., Peng, Q. D., Wang, D. S., Yang, T. T., Sorooshian, S., Liu, X. F. & Zhuang, J. B. 2018 Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology* **565**, 720–736.
- Zuo, G. G., Luo, J. G., Wang, N., Lian, Y. N. & He, X. X. 2020 Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. *Journal of Hydrology* **585** (11–12), 124776.

First received 11 January 2022; accepted in revised form 10 March 2022. Available online 23 March 2022