Long-term streamflow forecasting using SWAT through the integration of the random forests precipitation generator: case study of Danjiangkou Reservoir

Zhongmin Liang, Tiantian Tang, Binquan Li, Tian Liu, Jun Wang and Yiming Hu

ABSTRACT

Long-term streamflow forecasting is of great significance to the optimal management of water resources. However, the forecast lead time of long-term streamflow forecasting is relatively long and the forecasted precipitation within the forecast lead time has inherent uncertainty, so long-term streamflow forecasting has major challenges. In this paper, a hybrid forecasting model is developed to improve accuracy of long-term streamflow forecasting by combining random forests (RF) and the Soil and Water Assessment Tool (SWAT). The RF model is used to forecast monthly precipitation which is further downscaled to a daily dataset according to the hydrological similarity principle for use in the SWAT model of the Danjiangkou Reservoir basin, China. Performance of this hybrid model is compared to that of seasonal autoregressive (SAR (P)) model. Results show the RF precipitation generator yields accurate predictions at the monthly scale and the hybrid model produces acceptable streamflow series in long-term forecasting cases. In addition, the comparison shows that in the Danjiangkou Reservoir basin, the hybrid model performs better than the SAR (P) model, with average Nash–Sutcliffe efficiency (NSE) values of 0.94 and 0.51, which is better when it is closer to 1. This study provides a method of improving accuracy of long-term streamflow forecasting.

Key words | Danjiangkou Reservoir basin, hydrological similarity principle, long-term streamflow forecasting, random forests, SAR (P) model, SWAT

Zhongmin Liang
Tiantian Tang
Binquan Li (corresponding author)
Tian Liu
Jun Wang
College of Hydrology and Water Resources,
Hohai University,
Nanjing 210098,
China
E-mail: *libinquan@hhu.edu.cn*

Zhongmin Liang
National Cooperative Innovation Center for Water Safety & Hydro-Science,
Nanjing 210024
China

Yiming Hu
Research Institute of Management Science,
Business School,
Hohai University,
Nanjing, Jiangsu,
China

INTRODUCTION

Early and accurate long-term streamflow forecasting is of great significance to the optimal management and effective utilization of water resources (Pagano *et al.* 2014; Seo *et al.* 2015). Scheduling schemes for efficient water uses should be developed to ensure the safety of water conservancy projects and economic benefits (Xiao *et al.* 2016). However, with the extension of the forecasting lead time, the uncertainty of forecasted climatic inputs (e.g., precipitation) increases, and the accuracy of streamflow forecasting decreases considerably. Recently, various methods have been developed to improve the accuracy of long-term streamflow forecasting. Zhou *et al.* (2008) developed a wavelet predictor-corrector model for the simulation and prediction of monthly discharge time series at Yichang station on the Yangtze River. Westra *et al.* (2008) used independent component analysis for seasonal runoff forecasting. Wang *et al.* (2009) presented a model for forecasting seasonal runoff based on Bayesian joint probability (BJP). Pagano *et al.* (2009) used Z-score regression to predict the seasonal runoff volume. Yu *et al.* (2010) established the rainfall–runoff correlation model and applied it to the short-term flood forecast system in Danjiangkou Reservoir which used the correction technique based on the variable forgetting factor recursive least square algorithm to correct the flow

doi: 10.2166/nh.2017.085

process of the forecast in real time. Ran *et al.* (2010) established a forecast model based on the theory of weighted average and least-square method in Danjiangkou Reservoir. Feng *et al.* (2011) presented correlation between the monthly runoff in Danjiangkou Reservoir and the North Pacific sea surface temperature, the northern hemisphere 100 and 500 hPa height fields and the 74 circulation characteristic values in the previous year, and selected the stable meteorological factors with high correlation as the key factors. Then, they established a monthly runoff forecast model by means of the stepwise regression methods. This model was employed to simulate the monthly runoff of Danjiangkou Reservoir during 1957–2000 and to predict it during 2001–2006. Valipour *et al.* (2013) used the autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models to forecast monthly discharge. Terzi & Ergin (2014) forecasted monthly runoff using an autoregressive (AR) model. Huang *et al.* (2014) investigated a modified model for monthly runoff forecasting that combined empirical mode decomposition and a support vector machine. Wang *et al.* (2015a) combined an artificial neural network (ANN) model and ensemble empirical mode decomposition (EEMD) to forecast medium- and long-term runoff time series. Zhang *et al.* (2016) proposed the CEREF model to forecast annual runoff, and the results suggested that the CEREF model performed better than other methods.

Streamflow is a spatial and temporal hydrological response to precipitation. The volume of precipitation determines the change in streamflow and the associated hydrograph. In streamflow forecasting, the prediction of precipitation during the forecast lead time is significant; however, it is not considered in many models, which can result in unsatisfactory forecasting results. This issue may be overcome using numerical weather prediction (NWP) products. NWP products can provide forecasts of future precipitation. Thus, precipitation during the forecast lead time can be treated as an influential factor when forecasting streamflow (Gobena & Gan 2010; Bennett *et al.* 2014; Shi *et al.* 2015; Yu *et al.* 2016). Generally, NWP products provide a certain degree of accuracy for precipitation forecasting over 7–10 days. Nevertheless, the length of the forecast lead time must be further expanded to meet the requirements of production.

This paper aims to improve the accuracy of long-term streamflow forecasting and to develop a hybrid model. The hybrid model is based on a statistical model: random forests (RF), that forecasts precipitation series. These series are then used as input to the hydrological model, which is used to forecast monthly streamflow with a relatively long forecast lead time. We utilized a teleconnection analysis of precipitation and meteorological factors to screen the forecast factors and construct a RF model for forecasting monthly precipitation. Then, the forecasted monthly precipitation was used to screen the typical year through a hydrological similarity measure, and the forecasted monthly precipitation was downscaled into daily precipitation by the ratios of the forecasted year and typical year. The daily precipitation series was used as input of the Soil and Water Assessment Tool (SWAT) distributed hydrological model streamflow forecasting. In the past few decades, RF methods have been widely used in many engineering applications, including hydrology. Dhungel *et al.* (2016) used an RF model to predict site-specific changes in flow regime classes with downscaled climate projections. He *et al.* (2016) used an adaptable RF for spatial precipitation downscaling. Wang *et al.* (2015b) proposed a flood hazard risk assessment model based on RF. Additionally, RF was used to predict nitrate concentrations in shallow groundwater in the Central Valley (Nolan *et al.* 2015). Booker & Woods (2014) found that the RF method provided the best estimates of all hydrological indexes except average flow. Moreover, an RF model was used to predict base flow electrical conductivity (Olson & Hawkins 2012). Zhao *et al.* (2012) introduced an RF model for selecting a predictor set from measured runoff and predicted seasonal low flow in the upper Yangtze. Carlisle *et al.* (2010) used natural watershed characteristics to predict the value of each runoff metric using an RF approach.

The remainder of this paper is organized as follows. The next section introduces the study area and the data used. This is followed by a section describing in detail the RF and SWAT methods, as well as demonstrating the processes of model construction. Then, the results of model calibration and verification are presented and also a comparison of the results of the hybrid model and seasonal autoregressive model (SAR (P)) model. The final section states the conclusions.

STUDY AREA AND DATA

Study area

The Danjiangkou basin, with an area of 95,217 km^2, is located in the upper reaches of the Han River in the Yangtze River basin, China. The basin accounts for 60% of the catchment area of the Han River basin (Figure 1). The terrain is mainly mountainous valleys, high mountains, and hills, and it ranges from high in the northwest to low in the southeast in a horseshoe shape with the opening to the east. There are many water conservancy projects in the basin, and the Danjiangkou Reservoir is the middle route water source project of the South-to-North Water Transfer Project of China. The reservoir's comprehensive functions include flood control, water supply, power generation, irrigation, shipping, and aquaculture. Accurate medium- and long-term streamflow forecasting in the basin is the technical basis for the scientific water resources management of the basin and the south-to-north water diversion. The annual average temperature is 15–17 °C, and the average evaporation in the basin is 900–1,500 mm. The average annual precipitation is approximately 700–1,100 mm, and more than 80% occurs from May to October.

Data

The hydrometerological data we used are monthly precipitation series from 1981 to 2013, provided by the Bureau of Hydrology and values of daily average rainfall, temperature, wind speed, solar radiation and relative humidity from 1995 to 2011 from seven meteorological stations at Ankang, Foping, Xixia, Shangzhou, Shiquan, Hanzhong and Zhenan (Figure 1). Atmospheric circulation data consisting of monthly mean values of 74 atmospheric circulation indexes (ACI) from 1980 to 2012 were downloaded from the National Climate Center, China. Sea surface temperature

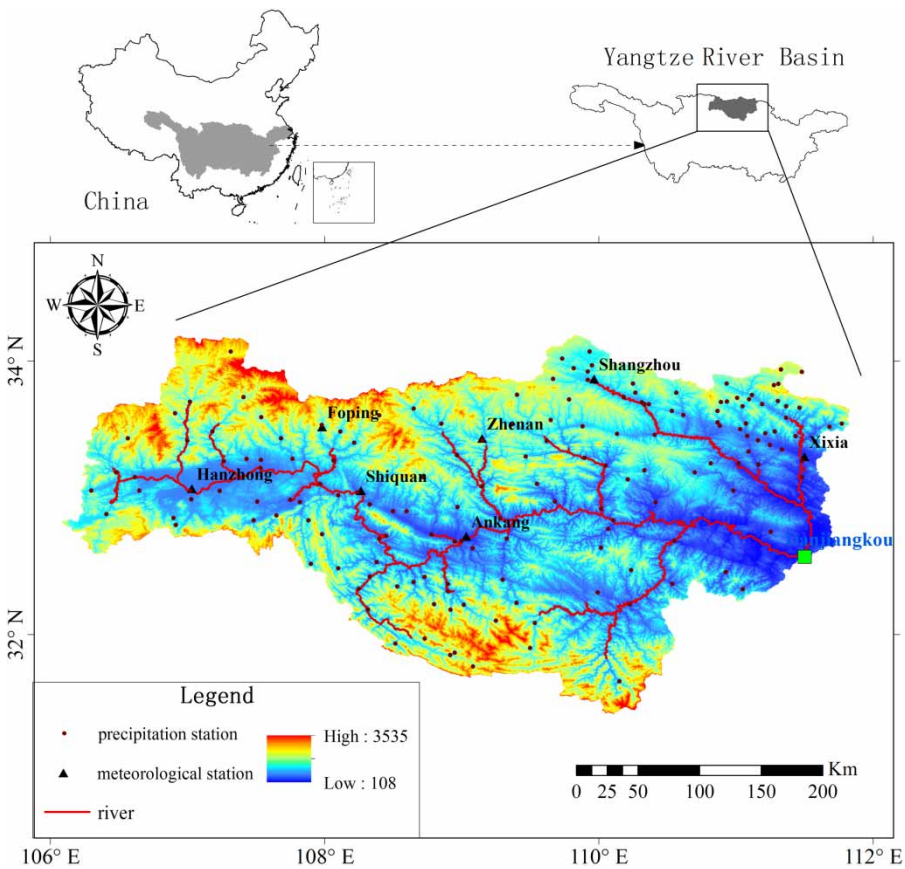

Figure 1 | Map of the Danjiangkou basin.

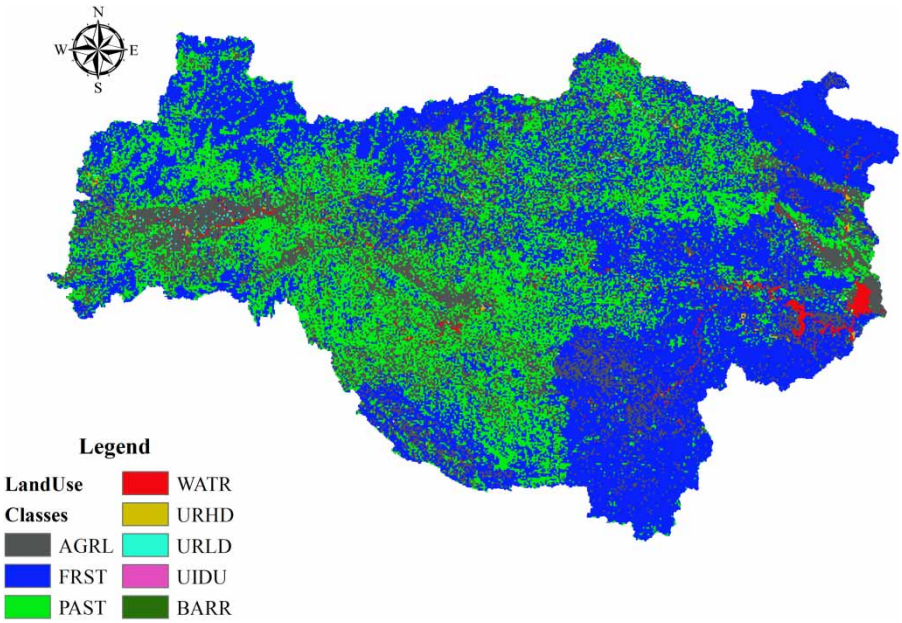

Figure 2 | Land use data.

(SST) data and 500 HPa geopotential height field (500Hpa) data, both consisting of monthly mean values from 1980 to 2012, were provided by the National Centers for Environmental Prediction, USA. The spatial grid resolution of SST is $2^\circ \times 2^\circ$, and the range spans from 52.5°N to 12.5°S and from 117.5°E to 77.5°W. The spatial grid resolution of 500Hpa is $2.5^\circ \times 2.5^\circ$, and the range spans from 80°N to 10°S and from 0°E to 360°E. The spatial and attribute data we used are digital elevation model (DEM), land use data in 2005 (Figure 2) and soil data in 1995 (Figure 3).

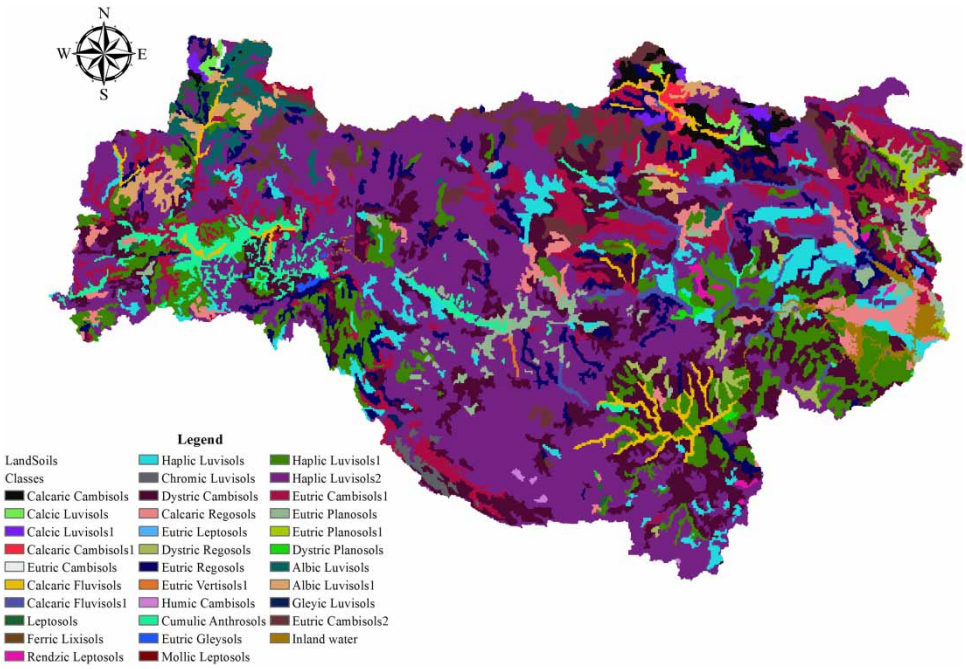

Figure 3 | Soil data.

METHODOLOGY

Random forests

Principle of random forests

RF (Breiman 2001) is a machine learning algorithm that combines bagging and ensemble learning theory (Breiman 1996a) with the random subspace method (Ho 1998). RF is a classifier consisting of a collection of tree-structured classifiers: $\{h(x, \Theta_k), k = 1, \ldots\}$, where $\{\Theta_k\}$ are independent and identically distributed random vectors and each tree casts a unit vote for the most popular class of input x (Breiman 2001). The method utilizes bootstrap re-sampling technology to sample the original data and generate a number of training samples, each of which randomly selects the feature attributes through random subspace methods to construct the decision tree. Finally, the optimal result is obtained by a voting or averaging method. Previous studies have found that RF can effectively overcome the problems of noise and over fitting, and it provides high precision for prediction (Wang *et al.* 2015b).

The main technological aspects of RF are as follows:

1. Bootstrap re-sampling technology (Breiman 1996a) and out-of-band (OOB) error estimation: Different training samples are obtained from the original samples through bootstrap re-sampling. The size of the original sample is assumed to be N. Repeat sampling with replacement is performed, and the size of the training samples is also N. The probability that each sample is not included is $(1 - 1/N)^N$. This probability is 0.368 in the case that N is sufficiently large. Thus, approximately 37% of the original samples are not included and are considered OOB data. These data can be used to estimate the error of each decision tree, and the average value can be used to calculate the generalization error of the RF. The study found that OOB estimation had the same accuracy as the test set with the same sample size (Breiman 1996b).
2. Decision tree and random subspace theory: A decision tree algorithm is a top-down recursive method, and conclusions are obtained at the leaf nodes of the decision tree (Breiman *et al.* 1984). A key step in the decision tree is to select the attributes of node splitting. Based on the random subspace theory, a subset of attributes is extracted from all attributes with equal probability. Then, the Gini impurity level index is used to select an optimal attribute from the subset to split the node (Zhu & Pierskalla 2016).

Model construction

Step 1: The causal relationship between precipitation and hydrometeorological factors is established based on the correlation coefficient method $D = \{\{\times_i, y_i),\ \times_I \varepsilon\ X,\ y_i \varepsilon\ Y, i = 1,2,\ldots N\}$, which is used to selected M forecast factors. These factors are used to construct the training sample set together with the precipitation series, where X is the M-dimensional explanatory variable vector composed of predictors, Y is the target variable of the precipitation series and N is the sample capacity.

Step 2: k training sample subsets are randomly taken from the training sample set D through bootstrap re-sampling, and the size of the training sample subset is N.

Step 3: k decision trees are constructed for the k training sample subsets. According to the random subspace theory, m indexes (generally, $m = \sqrt{M}$) are randomly selected from the M indexes based on the node attribute values of the decision trees. Then, the optimal value is selected based on the Gini impurity level, and this value is the final node attribute value. Liu *et al.* (2015) found that for most data sets, when the number of trees is 100, the RF accuracy can meet the requirements of the method. Thus, k in this research is selected as 100.

Step 4: Each decision tree is executed based on top-down recursive growth to obtain a predicted flow value. The results of k decision trees are then voted on to obtain the ultimate classification (regression) results, namely, the final predicted value of precipitation.

The main structure of the model is shown in Figure 4.

Soil and water assessment tool

Introduction of SWAT

SWAT is a distributed hydrological model developed by the United States Department of Agriculture (USDA). It has been used to predict the influence of long-term land

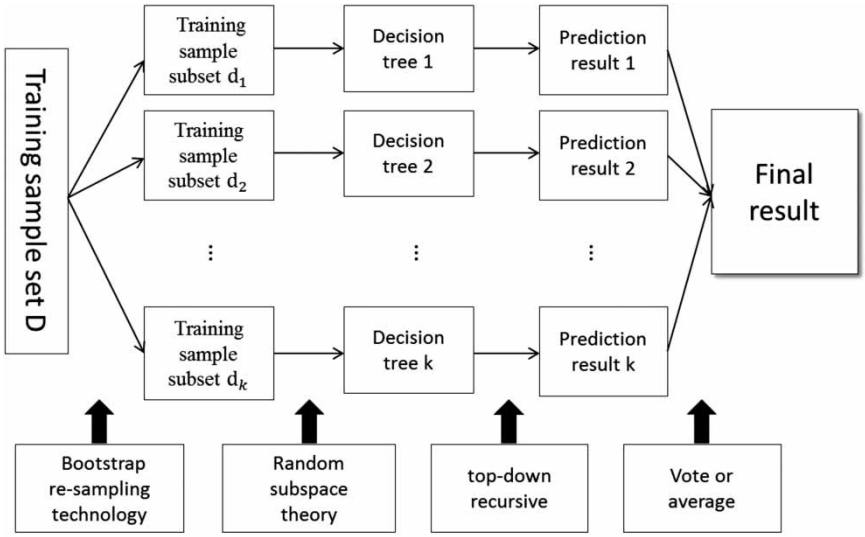

Figure 4 | The main structure of the random forest algorithm.

management measures on various soil types, land uses, and management conditions over large areas of complex watershed affected by runoff, sediment loading, and nutrient flow loss. To diminish the influence of the underlying watershed surface and the space-time variability of climate factors on the runoff simulation of the model, SWAT divides a river basin into multiple sub-basins. Additionally, the sub-basins are subdivided into different hydrologic response units (HRU) according to the type of soil and land use. The calculation of runoff yield after basin subdivision can improve the accuracy of runoff simulation. The ability of SWAT to predict streamflow has been widely verified around the world (Arnold *et al.* 1998; Chanasyk *et al.* 2003; Bouraoui *et al.* 2005; Jayakrishnan *et al.* 2005). Notably, it has yielded high accuracy for long-term simulations of yearly and monthly mean streamflow (Chen *et al.* 2011; Li *et al.* 2013; Lin *et al.* 2013). Therefore, SWAT can be used for medium- to long-term monthly mean streamflow prediction.

SWAT model construction

Step 1: Input basin DEM data. Using ArcGIS spatial analysis tools to analyse the flow directions in the basin, the cumulants of confluences are calculated and sub-basins (15 in this case) defined, as shown in Figure 5.

Step 2: Input the land use data and soil data. The land use and soil are classified according to the database in SWAT and the re-classification index table to define gradients. Then, the hydrological response units in the basin are determined.

Step 3: Load the precipitation station and meteorological station data. The corresponding precipitation and meteorological data are read according to the index name of each station to complete the construction of the model.

Hydrological similarity measure

Predicted precipitation from the RF model is provided at a monthly time step, while SWAT uses daily inputs. Therefore, monthly precipitation data are downscaled to daily series according to a hydrological similarity measure in this study. Based on similarity principle, a Euclidean distance metric function, which measures the difference between predicted and observed monthly precipitation vectors, is constructed:

$$\|D_i\| = \left\|\vec{X} - \vec{Y}_i\right\| = \sqrt{\sum_{j}^{12} (x_j - y_{i,j})^2} \quad (1)$$

where, $\vec{X} = (x_1, x_2, \ldots, x_{12})$ and $\vec{Y}_i = (y_{i,1}, y_{i,2}, \ldots, y_{i,12})$ are the predicted and observed monthly precipitation vectors, respectively. $\|D_i\|$ is the Euclidean distance between the

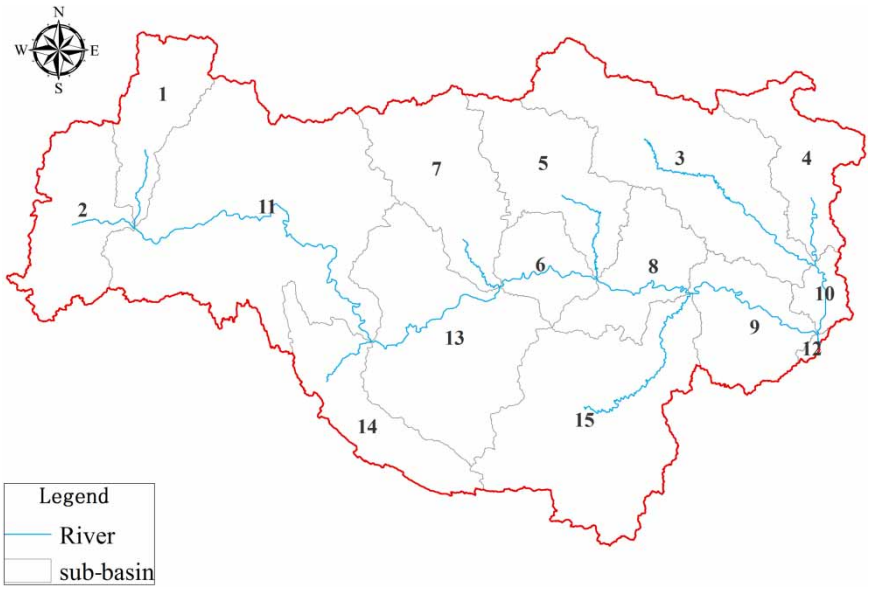

Figure 5 | Sub-basins of the study area.

predicted and the observed values in the ith year, and $i = 1 \sim n$, n is the total number of years.

According to the principle of minimum distance, the most similar historical data series are selected to predict monthly precipitation, and the ratio of observed to predicted precipitation can be calculated. Furthermore, monthly precipitation can be downscaled to obtain the predicted daily precipitation at each station. Additionally, historical meteorological data are used as input data in the SWAT model to calculate the daily evaporation capacity.

Evaluation measures

To evaluate the forecasting ability of the models, the simulation accuracy of each monthly forecast model is summarized. The coefficient of correlation (R), relative error (RE), mean absolute percentage error (MAPE) and Nash–Sutcliffe efficiency (NSE) are used as evaluation measures. They are defined as follows:

1. Coefficient of correlation (R):

$$R = \frac{(1/N)\sum_{i=1}^{N}(Q_o(i) - \overline{Q_o})(Q_f(i) - \overline{Q_f})}{\sqrt{(1/N)\sum_{i=1}^{N}(Q_o(i) - \overline{Q_o})^2}\sqrt{(1/N)\sum_{i=1}^{N}(Q_f(i) - \overline{Q_f})^2}} \quad (2)$$

2. Relative error (RE):

$$RE = \frac{Q_f(i) - Q_o(i)}{Q_o(i)} \times 100\% \quad (3)$$

3. Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(Q_o(i) - Q_f(i)}{Q_o(i)}\right| \times 100\% \quad (4)$$

4. Nash–Sutcliffe efficiency (NSE):

$$NSE = 1 - \frac{\sum_{i=1}^{N}(Q_o(i) - Q_f(i))^2}{\sum_{i=1}^{N}(Q_o(i) - \overline{Q_o})^2} \quad (5)$$

where $Q_o(i)$ and $Q_f(i)$ are the observed and simulated streamflow series, respectively; $\overline{Q_o}$ and $\overline{Q_f}$ are the mean observed and simulated streamflow series, respectively; and N is the length of the time series considered.

RESULTS AND DISCUSSION

Screening of predictors

The key predictors of monthly precipitation forecasts are selected from the 74 atmospheric circulation factors, the

North Pacific Ocean temperature field and the 500 HPa height field and further used to construct the RF model. A correlation analysis between these potential predictors (1980–2012) and monthly precipitation (1981–2013) is conducted to select the key influential predictors (at a significance level of 0.05) in each month of a year. Table 1 shows the selected predictors of precipitation in August as an example.

Forecasting model of monthly mean precipitation based on RF

The calibration and validation periods of the RF model are from 1981 to 2008 and 2009 to 2011, respectively.

Figure 6(a) and 6(b) illustrate the comparison of simulated and observed monthly precipitation during the calibration and verification periods, respectively. The figures show that the RF precipitation generator results exhibit good agreement with the observations, with R^2 values of 0.92 and 0.93 in the calibration and validation periods, respectively. In addition, when the values of observed precipitation are high, the simulated values are generally smaller than the observed values. At other times, these values are roughly distributed near the 1:1 line.

Moreover, according to the accuracy requirement of the quantitative prediction of medium- and long-term precipitation in the Standard for Hydrological Information and Hydrological Forecasting in China (GB/T22482-2008), a 20% variation in the amplitude in multiple years is taken as the permissible error, and the qualified rate (QR) is calculated.

The values of RE for simulated and observed precipitation from 1981 to 2008 are shown in Figure 7. The R, MAPE, NSE and QR values of the predictive simulation model from 1981 to 2008 are summarized, and the results are shown in Table 2.

Figure 7 shows that the values of RE in November and December are relatively large, and those in other months are comparatively well distributed. From the results of the four evaluation measures, the monthly rainfall forecasting model based on the RF model provides good simulation results. The correlation coefficient is above 0.9 in all 12 months, and the mean value is 0.94. The MAPE values in January, February, November and December are relatively large, potentially because the observed rainfall amount is

Table 1 | Selected precipitation predictors in August in the Danjiangkou basin

Number	Predictor	R	Description (previous year)
1	ACI_2_26	0.526	The subtropical high ridge line in India (65 E–95 E) in February
2	ACI _7_36	0.485	North Atlantic subtropical high on the north side of North America (110 W–60 E) in July
3	ACI _2_56	−0.483	Northern hemisphere polar vortex centre (JW) in February
4	ACI _3_64	−0.472	Asian meridional circulation index (IM, 60 E–150 E) in March
5	ACI _11_63	−0.453	Asian zonal circulation index (IZ, 60 E–150 E) in November
6	500 hpa_1_17.42	0.531	500 Hpa in January at (17.42)
7	500 hpa_11_5.46	0.528	500 Hpa in November at (5.46)
8	500 hpa_5_26.66	0.526	500 Hpa in May at (26.66)
9	500 hpa_3_15.136	−0.520	500 Hpa in March at (15.136)
10	500 hpa_8_17.69	0.518	500 Hpa in August at (17.69)
11	SST_7_17.10	0.512	SST in July at (17.10)
12	SST_5_21.33	0.510	SST in May at (21.33)
13	SST_10_7.60	−0.507	SST in October at (7.60)
14	SST_9_6.59	−0.487	SST in September at (6.59)
15	SST_8_5.58	−0.474	SST in August at (5.58)

Note: ACI_2_26 reflects that the forecasted factor is the 26th factor (the subtropical high ridge line in India (65E–95E)) of the 74 atmospheric circulation factors in February one year before the forecasted year.

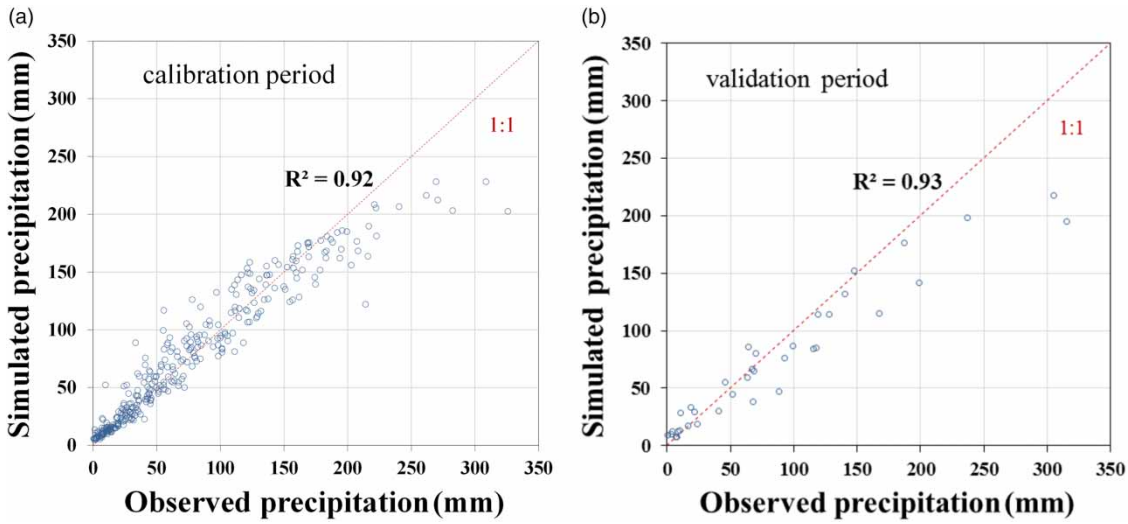

Figure 6 | Observed and simulated precipitation in the calibration and validation periods.

low during the non-flood season, and a slight difference between the simulated precipitation and observed precipitation will lead to a significant RE. However, the QR in January, February, November and December are 92.9%, 82.1%, 96.4% and 89.3%. Moreover, the NSE coefficients are all above 0.7 and roughly meet the accuracy requirements. Excluding January, February, November and December, the average value of MAPE is 21.35% in the other eight months, and the average NSE coefficient is greater than 0.75. Based on the proportion of the simulation results that meet the acceptance criteria to the total number of simulations, the QR is above 90%.

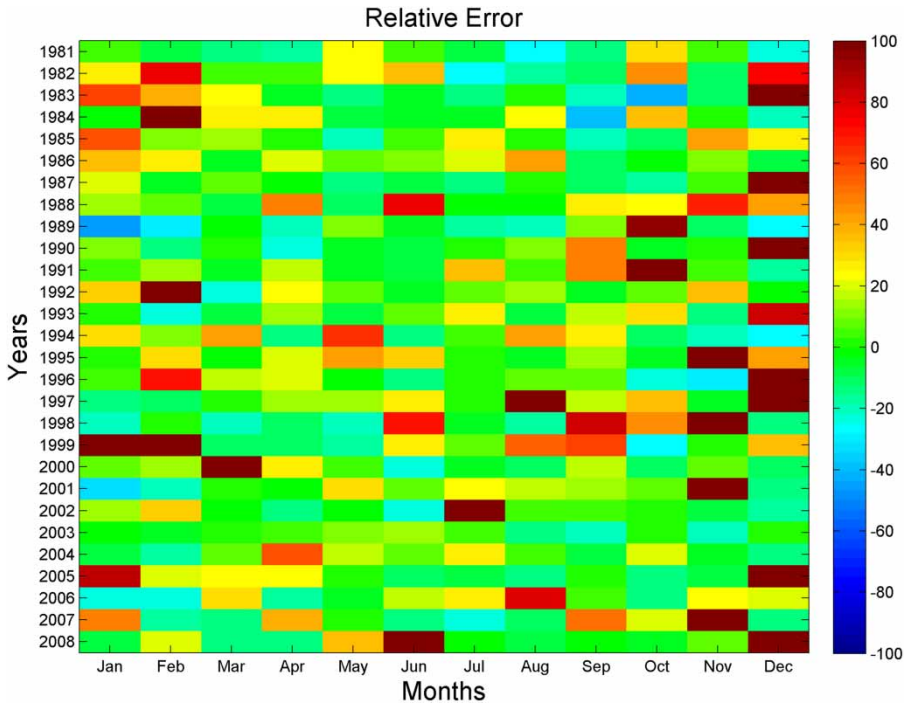

Figure 7 | RE of observed and simulated precipitation in the calibration period.

Table 2 | Values of evaluation measures from 1981 to 2008

Month	R	MAPE (%)	NSE	QR (%)
1	0.90	39.14	0.71	92.9
2	0.97	45.99	0.80	82.1
3	0.94	19.28	0.80	89.3
4	0.95	18.28	0.80	89.3
5	0.96	15.06	0.78	92.9
6	0.95	32.64	0.82	89.3
7	0.96	16.36	0.71	89.3
8	0.95	22.54	0.76	92.9
9	0.92	21.66	0.72	96.4
10	0.91	24.98	0.65	96.4
11	0.96	57.84	0.83	96.4
12	0.97	80.72	0.80	89.3

In addition, the RE is summarized, and a 20% variation in amplitude in multiple years is taken as the permissible error, deemed the accuracy assessment (AA), from 2009 to 2011. Values of RE and AA in the verification period are summarized in Table 3.

Based on the accuracy of the statistical results in the three-year validation period, 9 of the 12 months in 2009 meet the precision requirements. Additionally, 9 of the 12 months in 2010 and 8 of the 12 months in 2011 meet the accuracy requirements. Notably, the values of RE in January are relatively large in these three years. However, the overall accuracy is satisfactory.

Table 3 | Values of RE and AA from 2009 to 2011

	2009		2010		2011	
Month	RE (%)	AA	RE (%)	AA	RE (%)	AA
1	204.14	Y	1082.01	N	209.14	N
2	−21.41	Y	35.37	Y	3.04	Y
3	−26.44	Y	−12.68	Y	32.53	Y
4	−6.18	Y	−6.59	Y	74.24	Y
5	−31.34	N	−4.41	Y	−26.79	N
6	15.24	Y	−17.87	Y	−27.45	Y
7	3.01	Y	−28.60	N	−16.41	Y
8	−28.84	N	−6.06	Y	−5.58	Y
9	−12.69	Y	−10.85	Y	−38.24	N
10	33.62	Y	21.09	Y	−1.60	Y
11	−43.41	N	172.80	N	−46.41	N
12	3.67	Y	−7.24	Y	50.99	Y

Note: If the absolute error between the simulated and observed values is less than the permissible error, AA is labelled as 'Y'. If the absolute error between the simulated and observed values is greater than the permissible error, AA is labelled as 'N'.

Monthly streamflow forecasting using SWAT

In streamflow forecasting, numerous factors affect streamflow generation and routing. The calibration parameters should be related to surface water, soil water, groundwater and the flow process. Considering the above factors and combined with the parameter sensitivity analysis module of the model, the parameter sensitivity order is obtained through the LH-OAT sensitivity analysis method (Griensven *et al.* 2006). The SUFI-2 algorithm in SWAT-CUP is utilized to calibrate and verify the parameters. The calibration period is from 1995 to 2008, and the validation period is from 2010 to 2011. The initial range of parameter calibration and the final calibration results are shown in Table 4.

Table 4 also shows the parameter sensitivity order. The surface runoff lag time coefficient (SURLAG) is the most sensitive parameter, as shown in Table 4. In fact, SURLAG is related to variations in basin parameters. The SCS runoff curve number (CN2) is another notably sensitive parameter. It is a direct index of variation in infiltration. In addition, other parameters are also considered sensitive based on evaluation. These parameters include the baseflow factor (ALPHA_BF), threshold water depth in the shallow aquifer for return flow to occur (GWQMN), groundwater delay time (GW_DELAY), groundwater revap coefficient (GW_REVAP) and threshold water depth in the shallow aquifer for revap (REVAPMN), which are related to variations in groundwater parameters. Additionally, the soil saturated hydraulic conductivity (SOL_K), volume weight of the soil (SOL_BD) and available water content of the soil (SOL_AWC) are related to changes in soil parameters.

The calibration results exhibit good agreement with the observed monthly streamflow (Figure 8), and most RE values are small. However, RE is generally high in January and February. Figure 8 also shows a comparison between observed and simulated monthly streamflow in the validation period. It suggests that the simulated streamflow is generally larger than the observations.

Table 4 | Results of parameter calibration using the SWAT model

Parameter	Description	Optimal value	Range	Rank
SURLAG	Surface runoff lag time coefficient	18.132	(0.05, 24)	1
CN2	SCS runoff curve number	0.155	(− 0.2, 0.2)	2
SOL_K	Soil saturated hydraulic conductivity	0.648	(− 0.8, 0.8)	3
SOL_BD	Volume weight of soil	0.521	(− 0.5, 0.6)	4
SOL_AWC	Available water content of soil	−0.131	(− 0.2, 0.4)	5
GW_DELAY	Groundwater delay time (days)	115.833	(0, 500)	6
ALPHA_BF	Baseflow factor (days)	0.598	(0, 1)	7
GWQMN	Threshold water depth in the shallow aquifer for return flow to occur (mm)	925.000	(0, 5000)	8
GW_REVAP	Groundwater revap. coefficient	0.050	(0.02, 0.2)	9
REVAPMN	Threshold water depth in the shallow aquifer for revap. (mm)	745.000	(0, 1000)	10
ALPHA_BNK	Bank regulation base flow factor	0.622	(0, 1)	11
CH_N2	Manning coefficient of the main channel	0.155	(0, 0.3)	12
CH_K2	Channel effective hydraulic conductivity (mm/h)	0.170	(− 0.01, 0.3)	13
CH_N1	Manning coefficient of branches	8.057	(0.01, 30)	14
CH_K1	Effective permeability coefficient of tributary alluvium (mm/h)	214.500	(0, 300)	15
LAT_TTIME	Lateral flow time	8.100	(0, 180)	16
CANMX	Maximum canopy water storage	19.500	(0, 100)	17
ESCO	Soil evaporation compensation factor	0.259	(0.01, 1)	18
EPCO	Plant transpiration compensation coefficient	0.480	(0, 1)	19

Statistical analyses of the accuracy of the calibration period sequence and the accuracy of the validation period sequence (discussed in the preceding section) are conducted, and the specific results are shown in Table 5.

During the calibration period, the R, MAPE and NSE values are 0.96, 27.02% and 0.87, respectively. Then, the values of R, MAPE and NSE slightly decrease to 0.94, 31.02% and 0.83, respectively, in the validation period. These statistics indicate that the model accuracy in the calibration period is slightly superior to the accuracy in the validation period. Overall, the model provided acceptable accuracy in both the calibration and verification periods, with R values greater than 0.9 and NSE values above 0.8.

Comparison of the hybrid model and SAR (P)

First, we set the monthly precipitation data predicted by the RF model in 2012 as $\vec{X} = (x_1, x_2, \ldots, x_{12})$ and the historically observed monthly precipitation data series as $\vec{Y_i} = (y_{i,1}, y_{i,2}, \ldots, y_{i,12})$. Then, we calculate the Euclidean

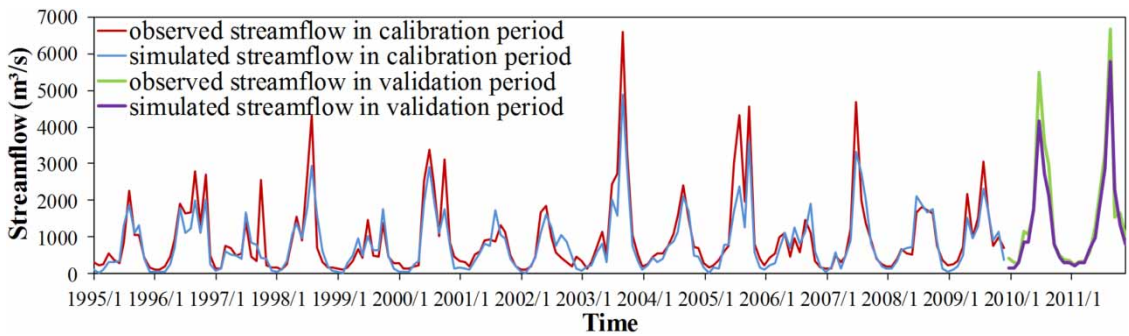

Figure 8 | Observed and simulated monthly streamflow in the calibration period and validation period.

Table 5 | Values of the evaluation measures in the calibration period and validation period

	Calibration period			Validation period		
Accuracy indexes	**R**	**MAPE (%)**	**NSE**	**R**	**MAPE (%)**	**NSE**
	0.96	27.02	0.87	0.94	31.02	0.83

Table 6 | Values of evaluation indicators for the hybrid model and SAR (P) model

	R	MAPE (%)	NSE
Hybrid model	0.97	26.49	0.94
SAR (P)	0.78	74.44	0.51

distance $\|D_i\|$. The monthly precipitation series in 1972 are selected by calculating the minimum distance $\min\|D_i\|$ between the predicted monthly precipitation vectors and historically observed monthly precipitation vectors. The values in 1972 are the most similar to those predicted in 2012. Next, the ratios of observed monthly precipitation in 1972 to predicted precipitation in 2012 are calculated. According to the ratios of each month, the daily rainfall at each precipitation station in 1972 is downscaled to obtain the predicted daily rainfall at seven meteorological stations (Hanzhong, Foping, Shangzhou, Zhenan, Xishan, Shiquan and Ankang) required in the SWAT model in 2012. Meanwhile, other meteorological data in 1972 are used as the input data in the 2012 SWAT model to calculate the daily evaporation capacity. Finally, the results of the hybrid model are compared with those of the widely used SAR model in the study basin.

The simulated and observed precipitation and values of R, MAPE and NSE are obtained and are shown in Figure 9. The results show that there is no significant difference between the simulated and observed precipitation. In addition, the values of R, MAPE and NSE are 0.98, 55.28% and 0.92, respectively, which indicates that the precipitation forecasts are accurate.

Table 6 illustrates that the hybrid model provides better precision than, and generally outperforms, the SAR (P) model. The R and NSE values of the hybrid model are 0.97 and 0.94, respectively, while the corresponding values for the SAR (P) model are 0.19 and 0.43, respectively. Meanwhile, MAPE is 26.49% and 47.95% for the hybrid model and SAR (P) model, respectively.

The comparison suggests that the hybrid model provides a better fit of monthly streamflow based on the observations compared to that of the SAR (P) model, except in the first two months (January and February). In these two months, the SAR (P) model has higher RE values of 11.65% and 44.64%, respectively, while the corresponding values for the hybrid model are −39.3% and −51.51%. However, the accuracy of SAR (P) in the next ten months decreases because it uses the rolling forecast method with the extended forecast lead time. Due to the lack of observed data, the SAR (P) model predicts streamflow values using the P-month data prior to the forecasting month; thus, bias gradually accumulates and progressively decreases the forecasting accuracy.

Moreover, Table 7 shows that, based on the 20% criterion of qualified values, 9 of 12 months meet the requirements, and the QR of the hybrid model reaches 75%. However, the QR of the SAR (P) model is only 50%.

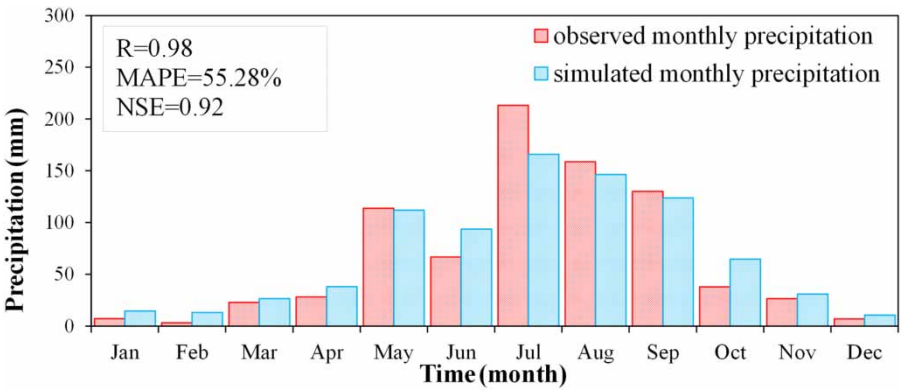

Figure 9 | Simulated and observed monthly precipitation in 2012.

Table 7 | Values of runoff and AA for the hybrid model and SAR (P) model in 2012

	Hybrid model Streamflow (m^3/s)			SAR (P) Streamflow (m^3/s)		
Month	Observed	Simulated	AA	Observed	Simulated	AA
1	515	313	N	515	575	Y
2	477	231	N	477	690	N
3	572	459	Y	572	1,160	N
4	469	330	Y	469	1,050	N
5	1,119	925	Y	1,119	1,840	N
6	599	547	Y	599	1,260	N
7	2,750	2,357	Y	2,750	1,900	Y
8	1,508	1,841	Y	1,508	925	Y
9	2,802	2,840	Y	2,802	2,000	Y
10	560	738	Y	560	1,290	Y
11	369	467	Y	369	710	Y
12	384	171	N	384	826	N

CONCLUSION

In this study, a hybrid model that combined SWAT and RF precipitation generator is developed to forecast long-term streamflow. First, the RF model is used to forecast monthly rainfall. Then, the SWAT model is constructed to forecast monthly streamflow in the Danjiangkou Reservoir basin. According to the hydrological similarity principle, the precipitation data generated by the RF model are downscaled and used as input data in the SWAT model to forecast monthly streamflow in 2012. These forecasted results are then compared to those of the SAR (P) model. The specific findings of this study are as follows:

1. An RF model is constructed based on observed monthly precipitation series from 1981 to 2008 and 15 predictors, which are selected using the correlation coefficient method. The RF model possesses a satisfactory capability in forecasting monthly mean rainfall in the Danjiangkou River basin. For example, during the calibration period, the values of R in 12 months are all above 0.9, and the NSE values are generally greater than 0.7. Thus, these values meet the standard accuracy requirements. For the entire forecasting sequence, the QR is above 90%. In addition, the QR reaches 70% in the verification period. The results show that the forecasting model of monthly mean rainfall is suitable for studies in the chosen river basin.
2. A SWAT model is constructed to forecast monthly streamflow. The SUFI-2 algorithm is applied in the SWAT-CUP software to calibrate the SWAT model. According to the results of model calibration and validation, the values of R, MAPE and NSE are 0.96, 27.02% and 0.87, respectively, in the calibration period and 0.94, 31.02% and 0.83 in the validation period. In general, the SWAT model exhibits good performance in forecasting monthly streamflow.
3. The RF model is used to forecast monthly precipitation in 2012, and the predicted values are compared with observational data to determine the accuracy of the forecasting model. Then, the Euclidean distance between simulated monthly mean rainfall and observed rainfall is calculated to select a typical precipitation year based on a hydrological similarity principle analysis. Daily precipitation data from seven meteorological stations are obtained by scaling using a proportionality coefficient. These data are used as input data in the SWAT model to forecast monthly mean runoff in 2012. The hybrid model obtains R, MAPE and NSE values of 0.97, 26.49% and 0.94, respectively, and the QR reaches 75%. Comparatively, R, MAPE and NSE values of 0.78, 74.44% and 0.51 are obtained using the SAR model. The results of the comparison and analysis demonstrate that the hybrid model is more accurate and applicable in the Danjiangkou basin.
4. By analysing important meteorological factors, we forecast monthly rainfall and monthly mean runoff in combination with a distributed hydrological model. Accurate, reliable and stable precipitation forecasting is the key to the model, which has a direct bearing on the accuracy of runoff forecasting. Therefore, methods of improving the accuracy of precipitation forecasting must be investigated.

ACKNOWLEDGEMENTS

This study is supported by the National Key Research and Development Program of China (2016YFC0402707, 2016YFC0402706, 2016YFC0402709), Special Scientific

Research Fund of Public Welfare Industry of Ministry of Water Resources, China (201501004), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX17_0415).

REFERENCES

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. R. 1998 Large area hydrologic modeling and assessment part I: model development. *Journal of the American Water Resources Association* **34** (1), 91–101.

Bennett, J. C., Robertson, D. E., Shrestha, D. L., Wang, Q. J., Enever, D., Hapuarachchi, P. & Tuteja, N. K. 2014 A system for continuous hydrological ensemble to lead times of 9 days forecasting (SCHEF). *Journal of Hydrology* **519**, 2832–2846.

Booker, D. J. & Woods, R. A. 2014 Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. *Journal of Hydrology* **508**, 227–239.

Bouraoui, F., Benabdallah, S., Jrad, A. & Bidoglio, G. 2005 Application of the SWAT model on the Medjerda river basin (Tunisia). *Physics & Chemistry of the Earth Parts A/B/C* **30** (8), 497–507.

Breiman, L. 1996a *Out-Of-Bag Estimation*. https://www.stat.berkeley.edu/ ~ breiman/OOBestimation.pdf.

Breiman, L. 1996b Bagging predictors. *Machine Learning* **24**, 123–140.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Breiman, L. I., Friedman, J. H., Olshen, R. A. & Stone, C. J. 1984 Classification and regression trees (CART). *Biometrics* **40** (3), 358.

Carlisle, D. M., Falcone, J., Wolock, D. M., Meador, M. R. & Norris, R. H. 2010 Predicting the natural flow regime: models for assessing hydrological alteration in streams. *River Research and Applications* **26** (2), 118–136.

Chanasyk, D. S., Mapfumo, E. & Willms, W. 2003 Quantification and simulation of surface runoff from fescue grassland watersheds. *Agricultural Water Management* **59** (2), 137–153.

Chen, L.-X., Liang, Z.-M. & Zhu, J.-F. 2011 Application of SWAT model to runoff simulation in Huangyahe Basin. *Water Resource and Power* **10**, 8–11.

Dhungel, S., Tarboton, D. G., Jin, J. & Hawkins, C. P. 2016 Potential effects of climate change on ecologically relevant streamflow regimes. *River Research and Applications* **32** (9), 1827–1840.

Feng, X. C., Wang, Y. T., Liu, Y. & Hu, Q. F. 2011 Monthly runoff forecast for Danjiangkou Reservoir based on physical statistical methods. *Journal of Hohai University* **39** (3), 242–247.

Gobena, A. K. & Gan, T. Y. 2010 Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology* **385** (1–4), 336–352.

Griensven, A. V., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M. & Srinivasan, R. 2006 A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology* **324** (1–4), 10–23.

He, X. G., Chaney, N. W., Schleiss, M. & Sheffield, J. 2016 Spatial downscaling of precipitation using adaptable random forests. *Water Resources Research* **52** (10), 8217–8237.

Ho, T. K. 1998 The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **20** (8), 832–844.

Huang, S., Chang, J., Huang, Q. & Chen, Y. 2014 Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology* **511** (7), 764–775.

Jayakrishnan, R., Srinivasan, R., Santhi, C. & Arnold, J. G. 2005 Advances in the application of the SWAT model for water resources management. *Hydrological Processes* **19** (3), 749–762.

Li, L., Dong, X., Yu, D., Liu, J. & Zhou, Q. 2013 Study on runoff simulations on Qingjiang River Basin by SWAT model. *Yangtze River* **44** (22), 25–29.

Lin, K.-R., Wei, X.-P., Huang, S.-X. & He, Y.-H. 2013 Application of SWAT model in Dongjiang River Basin. *Journal of China Hydrology* **33** (4), 32–36.

Liu, M., Lang, R. & Cao, Y. 2015 Number of trees in random forest. *Computer Engineering and Applications.* **51** (5), 126–131.

Nolan, B. T., Fienen, M. N. & Lorenz, D. L. 2015 A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology* **531**, 902–911.

Olson, J. R. & Hawkins, C. P. 2012 Predicting natural base-flow stream water chemistry in the western United States. *Water Resources Research* **48** (2), W02504.

Pagano, T. C., Garen, D. C., Perkins, T. R. & Pasteris, P. A. 2009 Daily updating of operational statistical seasonal water supply forecasts for the western U.S. *Journal of the American Water Resources Association* **45** (3), 767–778.

Pagano, T. C., Wood, A. W., Ramos, M. H., Cloke, H. L., Pappenberger, F., Clark, M. P., Cranston, M., Kavetski, D., Mathevet, T. & Sorooshian, S. 2014 Challenges of operational river forecasting. *Journal of Hydrometeorology.* **15** (4), 1692–1707.

Ran, D. K., Li, M., Wu, S. & Xie, J. C. 2010 Research on multi-model forecasts in mid-long term runoff in Danjiangkou Reservoir. *Journal of Hydraulic Engineering* **41** (9), 1069–1073.

Seo, Y., Kim, S., Kisi, O. & Singh, V. P. 2015 Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *Journal of Hydrology* **520**, 224–243.

Shi, H. Y., Li, T. J., Liu, R. H., Chen, J., Li, J. Y., Zhang, A. & Wang, G. Q. 2015 A service-oriented architecture for ensemble flood forecast from numerical weather prediction. *Journal of Hydrology* **527**, 933–942.

Terzi, O. & Ergin, G. 2014 Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. *Neural Computing & Applications* **25** (1), 179–188.

Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly

inflow of Dez dam reservoir. *Journal of Hydrology* **476**, 433–441.

Wang, Q. J., Robertson, D. E. & Chiew, F. H. S. 2009 A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research* **45** (5), 641–648.

Wang, W. C., Chau, K. W., Qiu, L. & Chen, Y. B. 2015a Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environmental Research* **139**, 46–54.

Wang, Z. L., Lai, C. G., Chen, X. H., Yang, B., Zhao, S. W. & Bai, X. Y. 2015b Flood hazard risk assessment model based on random forest. *Journal of Hydrology* **527**, 1130–1141.

Westra, S., Sharma, A., Brown, C. & Lall, U. 2008 Multivariate streamflow forecasting using independent component analysis. *Water Resources Research* **44** (2), 339–356.

Xiao, M., Zhang, Q., Singh, V. P. & Chen, X. 2016 Probabilistic forecasting of seasonal drought behaviors in the Huai River basin, China. *Theoretical & Applied Climatology* **128**, 667–677.

Yu, S., Ji, C. M., Zhao, B. K., Huang, X. F. & Dong, F. Q. 2010 Research on rainfall-runoff correlation model in application of short-term flood forecast for the Danjiangkou Reservoir. *China Rural Water and Hydropower* **9**, 145–148.

Yu, W., Nakakita, E., Kim, S. & Yamaguchi, K. 2016 Improving the accuracy of flood forecasting with transpositions of ensemble NWP rainfall fields considering orographic effects. *Journal of Hydrology* **539**, 345–357.

Zhang, H. B., Singh, V. P., Bin Wang, B. & Yu, Y. H. 2016 CEREF: A hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. *Journal of Hydrology* **540**, 246–256.

Zhao, T., Yang, D., Cai, X. & Cao, Y. 2012 Predict seasonal low flows in the upper Yangtze River using random forests model. *Journal of Hydroelectric Engineering* **3**, 18–24.

Zhou, H. C., Peng, Y. & Liang, G. H. 2008 The research of monthly discharge predictor-corrector model based on wavelet decomposition. *Water Resources Management* **22** (2), 217–227.

Zhu, J. F. & Pierskalla, W. P. 2016 Applying a weighted random forests method to extract karst sinkholes from LiDAR data. *Journal of Hydrology* **533**, 343–352.

First received 1 May 2017; accepted in revised form 10 October 2017. Available online 20 November 2017