

Article

Improved Medium- and Long-Term Runoff Forecasting Using a Multimodel Approach in the Yellow River Headwaters Region Based on Large-Scale and Local-Scale Climate Information

Haibo Chu ^{1,2}, Jiahua Wei ^{1,2,*}, Jiaye Li ², Zhen Qiao ¹ and Jiongwei Cao ¹

¹ State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China; haibochu0613@163.com (H.C.); 584851559@qq.com (Z.Q.); 498458263@qq.com (J.C.)

² State Key Laboratory of Hydrosience & Engineering, Tsinghua University, Beijing 100084, China; lijaye16@mails.tsinghua.edu.cn

* Correspondence: weijiahua@tsinghua.edu.cn

Received: 21 June 2017; Accepted: 11 August 2017; Published: 15 August 2017

Abstract: Medium- and long-term runoff forecasting is essential for hydropower generation and water resources coordinated regulation in the Yellow River headwaters region. Climate change has a great impact on runoff within basins, and incorporating different climate information into runoff forecasting can assist in creating longer lead-times in planning periods. In this paper, a multimodel approach was developed to further improve the accuracy and reliability of runoff forecasting fully considering of large-scale and local-scale climatic factors. First, with four large-scale atmospheric oscillations, sea surface temperature, precipitation, and temperature as the predictors, multiple linear regression (MLR), radial basis function neural network (RBFNN), and support vector regression (SVR) models were built. Next, a Bayesian model averaging (BMA)-based multimodel was developed using weighted MLR, RBFNN, and SVR models, and the performance of the BMA-based multimodel was compared to those of the MLR, RBFNN, and SVR models. Finally, the high-runoff performance of these four models was further analyzed to prove the effectiveness of each model. The BMA-based multimodel performed better than those of the other models, as well as high-runoff forecasting. The results also revealed that the performance of the forecasting models with multiple climatic factors were generally superior to that without climatic factors. The BMA-based multimodel with climatic factors not only provides a promising, reliable method for medium- and long-term runoff forecasting, but also facilitates uncertainty estimation under different confidence intervals.

Keywords: runoff forecasting; support vector regression; multiple climatic factors; Bayesian model averaging; multimodel; Yellow River headwaters

1. Introduction

The headwaters of China's Yellow River comprise an important source of freshwater resources, contributing nearly 40% of the total amount available in the Yellow River basin [1,2]. Reliable and accurate medium- and long-term runoff forecasting plays an important role in hydropower generation and in the coordinated regulation of water resources so that decision-makers know the quantity of water available in the basin over long periods of time, and can facilitate efficient management of water resources for a multitude of competitive applications in the region [3,4]. According to different forecast periods, runoff forecasting can be divided into short-term forecasting (e.g., hourly or daily), medium- and long-term forecasting (e.g., weekly, monthly, and seasonal), and long-term forecasting (e.g., annual), so medium- and long-term forecasting refers to runoff forecasting at the monthly scale in this paper. The Yellow River headwaters region is located in the northeast of the Tibetan Plateau,

which is very sensitive to climate change and rarely influenced by human activities. Climate change has a great impact on runoff within basins, especially in climate-sensitive areas [5]. There is a complex non-linear relationship between different climatic factors and runoff, and it is difficult to understand such complex relationships and build an accurate runoff forecasting model at medium- and long-term time scales [6,7].

Multiple climatic factors as the input variables are different for runoff forecasting in different regions. Abudu et al. [8] used a stochastic hybrid modeling approach for forecasting monthly streamflow in the Rio Grande headwaters basin, and the input variables were antecedent runoff, precipitation and snow water equivalent. Evsukoff et al. [9] presented the development of a recurrent fuzzy system model for the Iguacu River basin in southern Brazil, and the input variable was only rainfall. Talei et al. [10] applied a Takagi-Sugeno neuro-fuzzy model with online learning for runoff forecasting for three different catchments, and the input variable was the antecedent rainfall. Elsanabary and Gan [11] developed artificial neural networks with a genetic algorithm for forecasting weekly streamflow of the Upper Blue Nile basin, and the input variables were monthly precipitation and sea surface temperature. Djibo et al. [12] used a probabilistic approach statistical seasonal streamflow forecasting over West African Sahel; the input variables were sea level pressure, relative humidity, and air temperature. Biondi and De Luca [13,14] considered a simple lumped and conceptual rainfall-runoff model for design flood estimation in gauged and ungauged catchments in southern Italy, by considering as input 500-years of 20-min synthetic rainfall data, derived from a daily rainfall generator and a specific downscaling procedure for Southern Italy. Previous studies primarily focused on precipitation and antecedent runoff as the inputs for runoff forecasting, however, less attention was paid to fully incorporate large-scale climate variables and other local-scale climate variables as the inputs.

Runoff forecasting has always been a tremendous challenge for water resource management engineers and decision-makers, and a wide variety of models have been applied to runoff forecasting. Runoff forecasting approaches can be divided into two categories: one is based on physical processes, which require an accurate description of runoff processes and, thus, its applicability is limited by large variations in space and time. The other is the data-driven approach, which attempts to use historical climate factors and runoff data features to forecast runoff in the future. Although data-driven approaches may lack the ability to provide physical interpretation, they are becoming increasingly popular for providing relatively the accurate flow forecasts with their rapid development and less information requirements than hydrological models need. Different data-driven models have the ability to capture the randomness, periodicity, and volatility of runoff processes, and multimodel techniques can obtain accurate and comprehensive information on different characteristic and improve forecasting accuracy by linearly combining two or more models according to different weighting strategies. A number of multimodel methods have been developed for runoff forecasting. Block et al. [15] presented the coupling of global climate models (GCMs), multiple regional climate models, and numerous water balance models to improve stream flow forecasting compared to any individual model. Zhang et al. [16] proposed singular spectrum analysis and the autoregressive integrated moving average (ARIMA) hybrid model for annual runoff forecasting. It was shown that the hybrid model exhibited the best predictive performance compared to ARIMA and singular spectrum analysis-linear recurrent formulae (SSA-LRF) models. Wang et al. [17] presented an artificial neural network (ANN) model coupled with ensemble empirical mode decomposition (EEMD) for forecasting medium- and long-term runoff time series, and the proposed EEMD-ANN model attained a significant improvement over the ANN approach alone in medium- and long-term runoff time series forecasting. Recently, the Bayesian model averaging (BMA)-based multimodel has gained popularity as a multimodel because it can provide a more realistic forecast that considers both between-model variances and in-model variances.

As the main water source for Yellow River, the variations in Yellow River headwater runoff greatly affect downstream discharge and available water supply and, therefore, it has attracted more

and more attention from researchers. Zheng et al. [18] have investigated changes in the stream flow regime in the headwater catchments of the Yellow River basin since the 1950s. The results showed no significant trend for the period 1956–2000, but reported that a significant change in annual stream flow occurred around 1990. Lan et al. [19] analyzed the response of the runoff in the headwater region to climate change and reasonableness of the response based on the data measured during the period 1959–2008. Recently, sporadic explorations have been conducted of medium- and long-term runoff forecasting in the Yellow River headwaters region. Wang et al. [20] developed three forms of hybrid artificial neural networks (ANNs) to forecast stream flows for the upper Yellow River. Zhang et al. [21] combined wavelet analysis (WA) and an ANN for runoff time series prediction in the headwaters area. However, few studies have adopted the BMA-based multimodel forecasting model for runoff forecasting, especially with a number of large-scale climatic indices, so it is an attempt to improve the accuracy of runoff forecasting in Yellow River headwaters region.

In this paper, the monthly runoff data from Tangnaihai station in the Yellow River headwaters region was analyzed as the case study. Four atmospheric oscillations, sea surface temperature in different areas, precipitation and temperature were fully considered as the inputs for runoff forecasting. First, multiple linear regression (MLR), radial basis function neural network (RBFNN), and support vector regression (SVR) models were built based on multiple climatic factors, and an investigation of how the main parameters influenced the performance of the SVR model was undertaken, and the best parameters were selected for medium- and long-term runoff forecasting. Then a BMA-based multimodel was developed using weighted models (MLR, RBFNN, and SVR) to further improve medium- and long-term runoff forecasting, and provide uncertainty estimation under different confidence intervals. The performance of the four models for high runoff was further analyzed to prove the effectiveness of each model. Finally, the multimodel, with and without climatic factors, were also built for comparison.

2. Materials and Methods

In this paper, the data-driven models, including multiple linear regression, radial basis function neural network, support vector regression, and a Bayesian model averaging-based multimodel were developed, and data-driven models are actually black box models, which extract the input-output relationship from the historical record of climate data and runoff to provide simplified representations of the complex nonstationary hydrological systems. Thus, the length and quality of the observed climate factors and runoff time series will determine the accuracy of the model, but once the length and quality of the observed runoff series are fixed, the parameters of models are dominant on the accuracy of the model.

2.1. Multiple Linear Regression

The multiple linear regression (MLR) model as a popular statistical time series model, is often used to predict values in the future; and it is built as follows:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \beta_0 \quad (1)$$

where k is the number of the predictors, β_i are regression coefficients calculated by the least-squares method, and x is the climatic factor [22].

2.2. Radial Basis Function Neural Network

The radial basis function neural network (RBFNN) is a typical feed-forward neural network often used for strict interpolation in multidimensional space. The RBFNN consists of one input layer, one hidden layer with a nonlinear RBF activation function, and one output layer with entirely different roles [23–25].

From the input layer to the hidden layer, a Gaussian transfer function is used for the hidden neurons, so the transformation is nonlinear; it can be expressed as follows:

$$\varphi_i(x) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma_i^2}\right) \quad i = 1, 2, \dots, N \quad (2)$$

where $\varphi_i(x)$ is the center of the basis function, σ_i is the spread of the radial basis function in the i th hidden node, N is the number of hidden nodes, and $\|x - x_i\|$ is the radial distance between x and the RBF function center. The nonlinear nodes of the hidden layer are centered so that each of them is specialized on a particular zone of the input space [26,27].

However, the transformation from the hidden layer to the output layer is linear, and it is denoted as the weighted summation of the outputs of all hidden nodes connected to the output nodes:

$$f(x) = \sum_{i=1}^N \omega_i \varphi_i(x) \quad (3)$$

where ω_i are the weights of the linear output nodes [28]. The aim of an RBF network is to determine centers, widths, and the linear output weights linking the RBFs to the output nodes layer.

2.3. Support Vector Regression

Support vector regression (SVR), developed from support vector machines (SVMs), is a promising technique for dealing with forecasting problems based on Vapnik-Chervonenkis dimension theory and the structure risk minimization principle. Compared to conventional artificial neural networks, it has better generalization ability with structural risk minimization instead of training error minimization [29,30].

The linear regression estimating function can be written as follows:

$$f(x) = \omega \phi(x) + b \quad (4)$$

where $\phi(x)$ is a nonlinear mapping from the input space to a high-dimensional feature space, b is a threshold value, and ω is a weight vector, which can be estimated by minimizing the following regularized risk function:

$$R(C) = \left(C \sum_{i=1}^N L_\varepsilon(d_i, y_i) \right) + \frac{1}{2} \|\omega\|^2 \quad (5)$$

The function $L_\varepsilon(d, y)$, called the ε -insensitive loss function, is given by the formula:

$$L_\varepsilon(d, y) = \begin{cases} 0 & |d - y| \leq \varepsilon \\ |d - y| - \varepsilon & |d - y| > \varepsilon \end{cases} \quad (6)$$

C determines the tradeoff between flatness and the empirical risk; ε is considered to measure empirical error with Vapnik's linear loss function.

Two slack variables ξ and ξ' can be incorporated into the regularized risk function to yield the following formulation:

$$R(\omega, \xi, \xi', \varepsilon) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi + \xi') \quad (7)$$

subject to:

$$\begin{cases} (\omega^T \phi(x_i) + b) - y_i \leq \varepsilon + \xi \\ y_i - (\omega^T \phi(x_i) + b) \leq \varepsilon + \xi' \\ \xi, \xi' \geq 0 \end{cases} \quad (8)$$

This constrained optimization problem can be written as a Lagrangian function:

$$\nu(\alpha_i, \alpha_i^*) = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) + \sum_{i=1}^N (\alpha_i - \alpha_i^*)d_i \quad (9)$$

subject to:

$$\begin{aligned} \sum_{i=1}^N (\alpha_i - \alpha_i^*) &= 0 \\ \sum_{i=1}^N (\alpha_i - \alpha_i^*) &\leq C \\ \alpha_i &\geq 0 \quad \alpha_i^* \geq 0 \quad i = 1, 2, \dots, N \end{aligned} \quad (10)$$

and an optimal weight vector of the regression model is:

$$\omega = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\phi(x) \quad (11)$$

and so the linear regression in Equation (11) becomes:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)k(x_i, x) + b \quad (12)$$

where $k(x_i, x_j)$ denotes the inner product of two vectors in the feature space $\phi(x_i)$ and $\phi(x_j)$, and the Gaussian kernel function is the most commonly used kernel function.

When the length and quality of the training samples are fixed, three parameters are dominated the accuracy of the SVR model: C , which controls the empirical risk degree of the SVR, ε , which controls the width of the tube in the loss function; and σ , which controls the Gaussian function width of the kernel function [31–33].

2.4. Bayesian Model Averaging-Based Multimodel

The Bayesian model averaging (BMA)-based multimodel is the average of the considered models weighted by the likelihood that a considered model is correct given the observations, it has been used in various fields, such as statistics, geology, and hydrologic applications. The BMA model provides a more reliable description of the total predictive uncertainty, including the between-model-variance and the within-model-variance, than the original multimodels [34,35].

Consider that $f = [f_1, f_2, \dots, f_K]$ is an ensemble of K considered models and $M = [y_1, y_2, \dots, y_T]$ are observational datasets with data length T , the BMA predictive probability density functions (PDFs) are then:

$$p(y|M) = \sum_{k=1}^K p(f_k|M) * p_k(y|f_k, M) \quad (13)$$

where $p(f_k|M)$ is the posterior probability of model prediction f_k ; that is, the likelihood of the model prediction being correct given the observational data M . If we denote $w_k = p(f_k|M)$, we should obtain $\sum_{k=1}^K w_k = 1$. The posterior mean and variance of the BMA prediction can be expressed as:

$$E(y|M) = \sum_{k=1}^K p(f_k|M)w_k \left(f_k - \sum_{i=1}^K w_i f_i \right)^2 + \sum_{k=1}^K w_k \sigma_k^2 \quad (14)$$

$$\text{var}(y|M) = \sum_{k=1}^K w_k \left(f_k - \sum_{i=1}^K w_i f_i \right)^2 + \sum_{k=1}^K w_k \sigma_k^2 \quad (15)$$

where σ_k^2 is the variance associated with model prediction.

An expectation-maximization (EM) algorithm was used to estimate w_k and σ_k^2 . If $\theta = (w_k, \sigma_k^2, \dots, k = 1, 2, \dots, K)$, the log-likelihood function can be approximated as:

$$l(\theta) = \log \left(\sum_{k=1}^K w_k * p_k(y|f_k, M) \right) \quad (16)$$

For this study, a latent variable $z_{k,t}$ is introduced:

$$z_{k,t} = \begin{cases} 1 & \text{the } k\text{th model ensemble is the best prediction at time } t \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

At any time t , there is only one $z_{k,t}$ equal to 1, and the rest are equal to 0. The EM algorithm starts with an initial value for the parameter θ . In the expectation (E) step, $z_{k,t}$ is estimated given the current guess of θ . In the maximization (M) step, θ is estimated given the current values of $z_{k,t}$. The EM steps are repeated until certain convergence criteria are satisfied.

2.5. Model Performance Evaluation

Four standard statistical measures, root-mean-square error (RMSE), mean relative error (MAE), Nash-Sutcliffe (NS) efficiency coefficient, and determination coefficient (R^2), are employed to evaluate the performance of all the forecasting models.

(1) RMSE:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (q_i - Q_i)^2}{N}} \quad (18)$$

where q_i and Q_i are the observed and predicted data at the i th time period, respectively, and N is the number of considered data.

RMSE is frequently used to measure the residual between observed and predicted runoff.

(2) MAE:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |Q_i - q_i| \quad (19)$$

MAE indicates the mean deviation between observed and predicted runoff.

(3) NS efficiency coefficient:

$$\text{NS} = 1 - \frac{\sum_{i=1}^n (q_i - Q_i)^2}{\sum_{i=1}^n (q_i - \bar{q}_i)^2} \quad (20)$$

where \bar{q}_i is the mean observed runoff. The NS efficiency coefficient is used to measure the capability of the model to forecast runoff away from the mean, which is sensitive to extreme values.

(4) R^2 :

$$R^2 = \left[\frac{\sum_{i=1}^n (q_i - \bar{q}_i)(Q_i - \bar{Q}_i)}{\sqrt{\sum_{i=1}^n (q_i - \bar{q}_i)^2 \sum_{i=1}^n (Q_i - \bar{Q}_i)^2}} \right]^2 \quad (21)$$

where \bar{Q}_i is the mean predicted runoff.

R^2 evaluates the linear correlation by summarizing the discrepancy between the observed and predicted runoff. The values of NS and R^2 range between 0 and 1.0. Essentially, the smaller the RMSE and MAE is, the more accurate the runoff expectation is; and the closer NS and R^2 is to 1, the more accurate the runoff expectation is.

3. Study Area

The Tangnaihai hydrological station is the control station for the largest reservoir in the upstream part of the Yellow River, which also serves as the control outlet of the basin; and it is sensitive to climate changes. There are no large hydraulic engineering mechanisms in place, so runoff is only slightly influenced by human activities. Therefore, runoff forecasting has important significance for predicting ecological water demand, flood control operation, water allocation, and hydropower scheduling for the Yellow River.

The study area is located in the eastern part of Qinghai Province between latitudes 32°5' N–36°30' N and longitudes 95°30' E–103°30' E, including Maduo, Maqin, Jiuzhi, Dari, Hongyuan, Ruorgai, Xinghai, Tongde, Zeku, and Henan counties, among others, and covers an area approximately $12.20 \times 10^4 \text{ km}^2$ in size above Tangnaihai station (Figure 1). The average elevation is approximately 4217 m and ranges from 2568 to 6264 m. It has a typical plateau continental monsoon region climate; the annual temperature difference is small, but the daily temperature difference is significant. The annual average temperature in the study area ranges from -5.38°C to 4.14°C , the annual average evaporation ranges between 730 and 1700 mm, the annual average rainfall ranges between 262.2 and 772.8 mm, and the average annual runoff of the Yellow River at Tangnaihai station is $1.39 \times 10^{10} \text{ m}^3$.

Here, four climatic indices, sea surface temperatures in different areas (SST), precipitation, and temperature were used. The climatic indices included the strength of the East Asian trough (EAT), West Pacific Subtropical High (WPSH), Northern Hemisphere polar vortex (NH), Tibetan Plateau Index B (TPI-B). Monthly EAT, WPSH, NH, and the TPI-B data used were retrieved from the Climate Prediction Center of the US National Oceanic and Atmospheric Administration. Monthly temperature (T) and precipitation (P) from 1956 to 2014 were obtained from the China Meteorological Administration. Sea surface temperature data during 1956–2014 were obtained from Met Office Hadley Centre observation datasets. Three zones of monthly SST were used: zone 1 (90°E – 105°E , 30°S – 35°S), zone 2 (122.5°E – 177.5°E , 17.5°N – 27.5°N), and zone 3 (160°E – 170°E , 25°S – 35°S).

In this study, the monthly runoff data from Tangnaihai station were selected for model calibration and validation. The dataset for the period January 1956–December 2010 was used for model calibration and the dataset for January 2011–December 2014 for model validation.

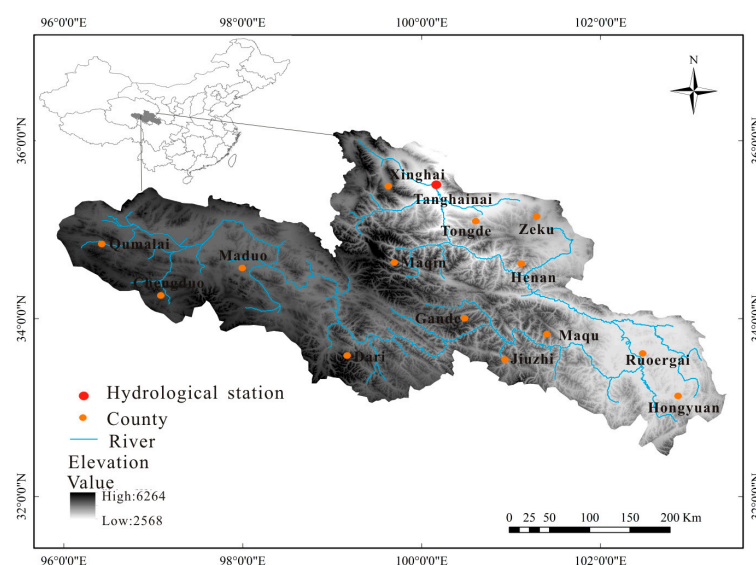


Figure 1. Location of the study area.

4. Results and Discussion

4.1. BMA-Based Multi Model Modeling Process

The BMA-based multi model was developed based on the MLR, RBFNN, and SVR models; hence, the SVR model, RBFNN model, and MLR model were first employed for forecasting medium- and long-term runoff. Then, an SVR model was selected as an example to illustrate the modeling process, and a detailed analysis of how the main parameters influenced the performance of the SVR model was performed.

The SVR model was built for monthly runoff forecasting, and the model-calibration process was carried out to obtain the optimal parameters and so that, finally, runoff forecasting could be conducted according to the training parameters. In the application, nine input variables including monthly EAT, WPSH, NH, TPI-B, three zones of monthly SST (SST1, SST2, SST3), evaporation, temperature, and precipitation, and the corresponding monthly runoff were used as the output, details of validation samples are presented in Table 1.

Table 1. Input-output pairs for model-validation process.

Sample	Input									Output
	WPSH	EAT	NH	TPI-B	SST1	SST2	SST3	P	T	Runoff
1	1	64	271	470	20	17	23	76	−12	179
2	0	113	360	492	21	16	24	8	−7	158
3	3	103	304	516	21	16	24	139	−6	216
4	12	190	220	548	19	17	23	234	1	384
5	36	226	143	574	18	18	22	693	5	460
6	62	226	115	578	17	19	20	1071	8	1170
...
42	29	274	116	580	17	18	20	1152	9	969
43	47	274	99	596	16	21	19	591	11	958
44	47	274	125	595	16	22	19	1788	10	1043
45	43	305	195	595	15	22	19	733	7	1454
46	18	236	253	576	16	21	19	223	1	1016
47	37	176	279	543	18	20	21	33	−4	455
48	20	89	301	521	19	18	22	83	−11	273

The important part in building a SVR model is training the parameters, including the regularization parameter, spread and tube width, which significantly influence forecasting accuracy. Hence, the parameters should be selected carefully.

The optimal parameters were determined by trial and error, and are as follows: the spread σ ranges from 0.1 to 10, the regularization parameter C from 0.1 to 10, and the tube width ϵ from 0.0001 to 0.1. The optimal values of the three parameters were selected when R^2 reached its maximum value as the values of the three parameters changed. Figure 2 shows the changing accuracy of the test datasets when the spread $\sigma = 0.6$, the tube width $\epsilon = 0.001$ and the regularization parameter C values were 0.5, 1.5, 2.5, and 3.5.

As shown in Figure 2, in the model-validation process, when $C = 0.5$, R^2 was approximately 0.807, and when $C = 2.5$, R^2 was approximately 0.890, when $C = 3.5$, R^2 was approximately 0.853. The results show that the best value for regularization parameter C was 1.5 with R^2 approximately 0.905. Hence, the SVR model using a regularization parameter C value of 1.5 was selected for monthly runoff forecasting in this study. In addition, regularization parameter $C = 1.5$, spread $\sigma = 0.6$, and regression tube width $\epsilon = 0.001$ were selected as the other optimal parameters for runoff forecasting at Tangnaihai station.

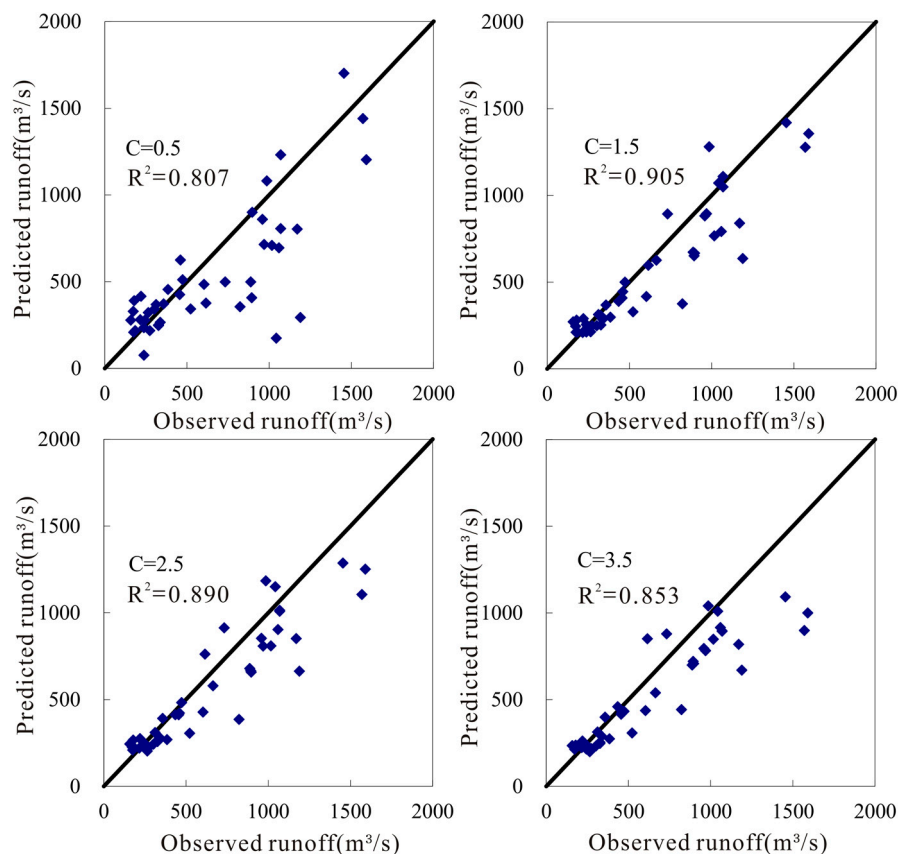


Figure 2. Performance of SVR model with different values of regularization parameter C .

RBFNN and MLR models were also applied for forecasting monthly runoff, and based on the MLR, RBFNN, and SVR models, a multimodel approach was developed to further improve the accuracy and reliability of medium- and long-term runoff forecasting. A BMA-based multimodel was developed for more reliable probabilistic runoff forecasting, which inferred consensus predictions by weighing individual predictions. The weights for the three models, MLR, RBF, and SVR were 0.005, 0.140, and 0.875, respectively. The results indicated the better an individual model performed, the higher the weight of that individual model is.

4.2. Comparative Analysis of MLR, RBFNN, SVR, and BMA-Based Multimodels

The results from the BMA-based multimodel were compared to those of the MLR, RBFNN, and SVR models, and the results of the comparison are shown in Figure 3. It can be seen that the SVR and RBFNN models performed better than the MLR model during the model-calibration process, and that the MLR model generally underestimated the monthly runoff compared to the observed values. During model calibration, the RMSEs for the SVR, RBFNN, and MLR models were 260, 280, and 304, respectively; the MAEs for the same three models were 0.174, 0.329, and 0.383, respectively; their R^2 values were 0.899, 0.834, and 0.799, respectively; and their NS values were 0.735, 0.696, and 0.638, respectively (Table 2). However, the BMA-based multimodel performed better than the SVR model during model calibration, giving the best RMSE, MAE, R^2 , and NS statistics of 180, 0.126, 0.945, and 0.879, respectively.

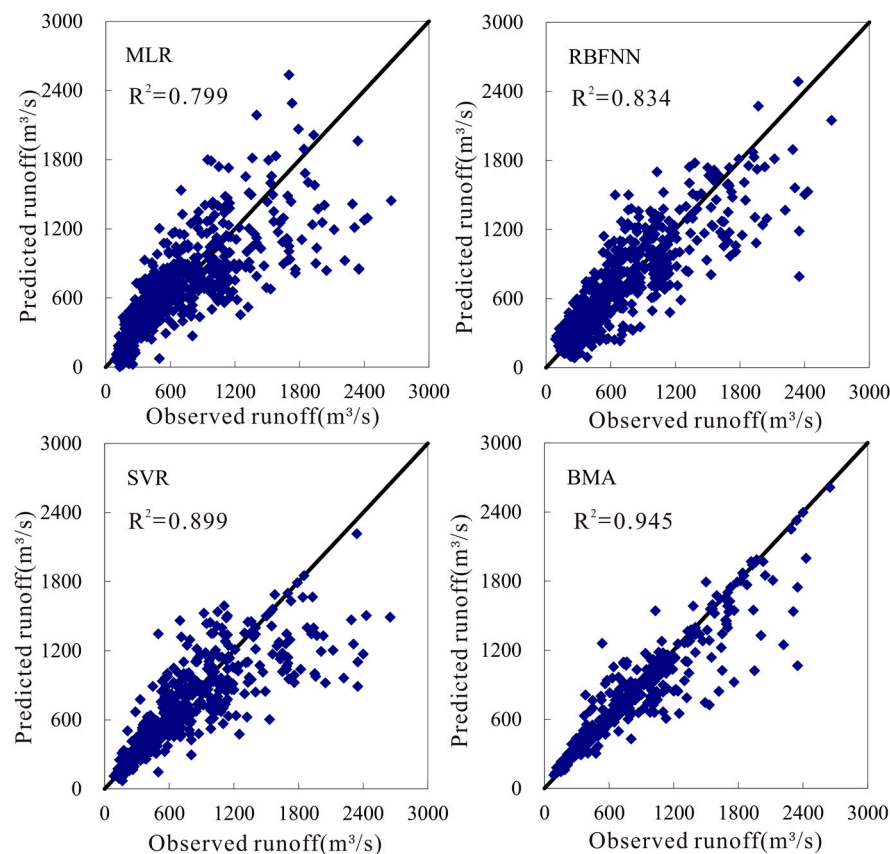


Figure 3. Predicted and observed runoff during model calibration using MLR, RBFNN, SVR, and the BMA-based multimodel.

Table 2. RMSE, MAE, R^2 , and NS values for the forecasting models.

		RMSE	MAE	R^2	NS
Calibration	MLR	304	0.383	0.799	0.638
	RBFNN	280	0.329	0.834	0.696
	SVR	260	0.174	0.899	0.735
	BMA	180	0.126	0.945	0.879
Validation	MLR	296	0.287	0.777	0.697
	RBFNN	327	0.307	0.798	0.725
	SVR	274	0.250	0.905	0.740
	BMA	261	0.196	0.937	0.785

Figure 4 displays the comparison between predicted and observed runoff during model validation using MLR, RBFNN, SVR, and the BMA-based multimodel for Tangnaihai station. It can be seen that the SVR model obviously outperformed the MLR and RBFNN models during model validation in terms of the standard statistical measures (Table 2), and both the SVR and RBFNN models were able to forecast runoff. Specifically, the RMSEs for the SVR and RBFNN models were 274 and 327; the MAEs for the same two models were 0.250 and 0.307, respectively; their R^2 values were 0.905 and 0.798, respectively; and their NS values were 0.740 and 0.725, respectively. However, all of the other models performed better than the MLR model, which had the worst RMSE, MAE, R^2 , and NS statistics, 296, 0.287, 0.777, and 0.697, respectively. Compared with all of the other models during model validation, the BMA-based multimodel obtained the best RMSE, MAE, R^2 , and NS statistics, of 261, 0.196, 0.937, and 0.785, respectively. The classic MLR model is relatively easy to

construct with the simplest type of parameters, and it can capture the global trend over an entire input space. Its accuracy, however, is not satisfactory, which may not meet the requirements of medium- and long-term runoff forecasting. The RBFNN model is capable of identifying complex nonlinear relationships between input and output data, and its accuracy is satisfactory for runoff forecasting, but there is a risk of over-fitting. The SVR model is also appropriate for reproducing the nonlinear problem, which can provide a suitable mapping between input and output data in a higher-dimensionality feature space to improve the forecasting accuracy. Its parameters need to be determined carefully due to the fact that they significantly influence the accuracy of the SVR model. Considering the overall characteristics of the above-mentioned models, the BMA-based multimodel significantly improves predictive performance.

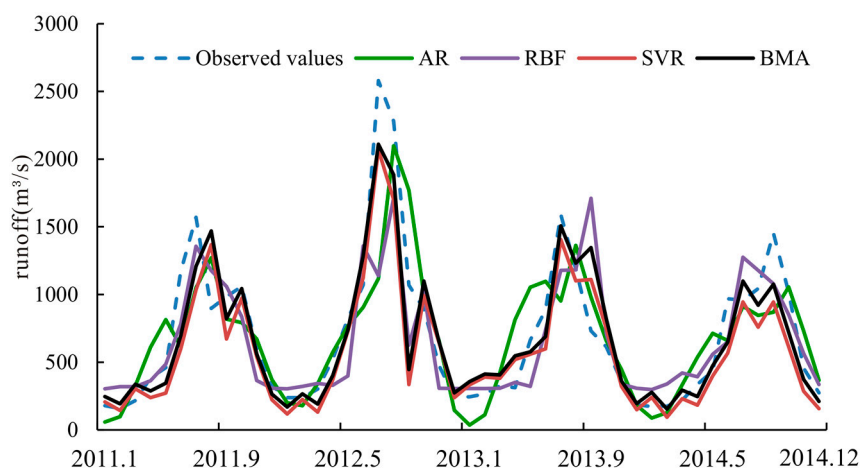


Figure 4. Predicted and observed runoff during model validation using MLR, RBFNN, and SVR.

4.3. High-Runoff Forecasting Analysis

Figure 5 displays the results of BMA-based multimodel forecasting along with the 90%, 95%, and 99% confidence intervals. The performance of BMA confidence interval is evaluated by coverage ratio and interval width. As the confidence level gets higher, the corresponding confidence interval will be wider. Coverage ratio is the ratio of the number of points, which ranges between the confidence interval, to the number of total points. The larger the coverage ratio and the smaller the interval width are, the better the BMA model performs. Figure 5 shows that most of the observed values are within a 90% confidence interval. The coverage ratio can reach 80%, and the interval width is narrow in general, but most of the high-runoff observations are outside of the confidence interval, and the interval width is wider, which indicates the uncertainty of forecasting at high values is great.

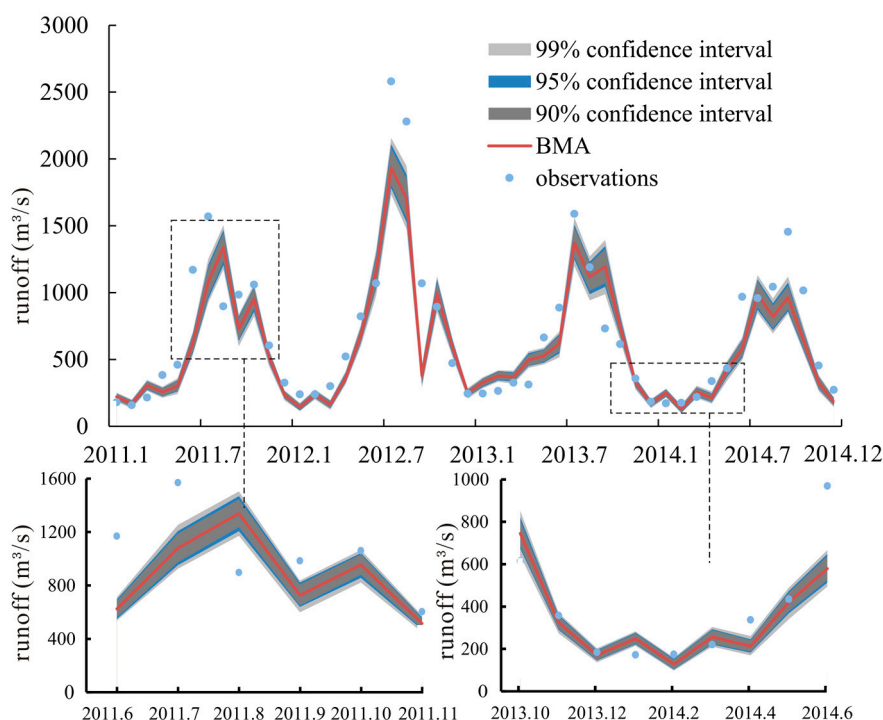


Figure 5. BMA-based multimodel runoff forecasting and 90% confidence interval compared to observations.

A runoff of over $1700 \text{ m}^3/\text{s}$ was selected to study the forecasting ability of high runoff for different models. The results are shown in Figure 6. The number of points used was 35, and the absolute differences between observed values and predicted values of all the points for different models were calculated. The minimal absolute differences for the BMA-based multimodel extended to 21 of 35 points, and the minimal absolute differences for the SVR, RBFNN, and MLR models extended to 5, 7, and 1 of 35 points, respectively. The BMA model has the best ability to forecast high runoff compared to all other models, the forecasting ability of the MLR model is the poorest, and the forecasting ability of the RBFNN and SVR models was better than that of the MLR model.

Generally, the BMA-based multimodel predicted values that are lower than the observed values when predicting high runoff. Since accurate prediction of extreme events would provide the most economic, environmental, and societal benefits, further improvements in high-runoff prediction are required.

There are significant differences between the observed values and the values predicted using all models for September 1981, and all of the models failed to capture extreme runoff events ($3500 \text{ m}^3/\text{s}$). The runoff in September 1981 was nearly two times greater than that for the same month in other years. Continuous rainfall occurred in the watershed above Tangnaihah station from August to September 1981, which resulted in serious flooding. As the major source of runoff, rainfall usually has the most significant impact on runoff, in addition, runoff also has a close relationship with evaporation, temperature, and other factors; although only the limited historical runoff time series available was applied in the paper, more runoff data needs to be considered for medium- and long-term runoff forecasting in the future.

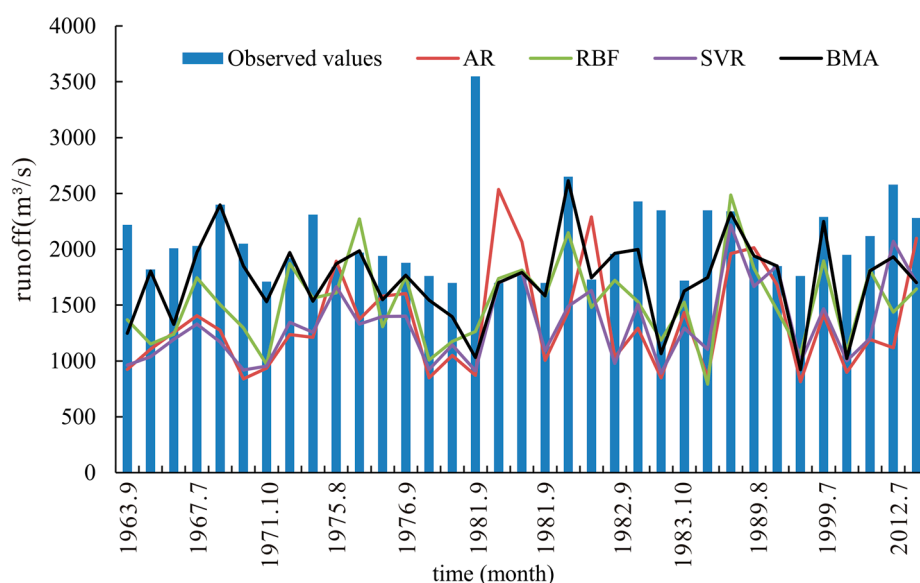


Figure 6. Predicted and observed peak runoff using MLR, RBFNN, and SVR models and the BMA-based multimodel.

Then another BMA-based multimodel was built for comparison based on only runoff time series. In the application, the first nine months of monthly runoff was used as the input, and the 10th monthly runoff as the output, details of validation samples were presented in Table 3.

Table 3. Input-output pairs of runoff (m^3/s) for model-validation process (without climatic factors).

Sample	Input (Runoff)									Output (Runoff)
1	317	373	1010	2120	945	689	617	422	203	179
2	373	1010	2120	945	689	617	422	203	179	158
3	1010	2120	945	689	617	422	203	179	158	216
4	2120	945	689	617	422	203	179	158	216	384
5	945	689	617	422	203	179	158	216	384	460
6	689	617	422	203	179	158	216	384	460	1170
...
42	29	274	116	580	17	18	20	1152	9	969
43	47	274	99	596	16	21	19	591	11	958
44	47	274	125	595	16	22	19	1788	10	1043
45	43	305	195	595	15	22	19	733	7	1454
46	18	236	253	576	16	21	19	223	1	1016
47	37	176	279	543	18	20	21	33	−4	455
48	20	89	301	521	19	18	22	83	−11	273

This displays the comparison between predicted and observed runoff using the BMA-based multimodel with and without climatic factors in Figure 7. Specifically, the R^2 for BMA-based multimodel, with climatic factors and without climatic factors, models were 0.945 and 0.884 during model calibration, and 0.937 and 0.846 during model validation, respectively. It can be seen that the BMA-based multimodel with climatic factors was generally superior to that without climatic factors during model calibration and validation. Large-scale oceanic–atmospheric conditions and local-scale climate indices have a direct or indirect influence on runoff variability in the source region of Yellow River, and it certainly improves medium- and long-term runoff forecasting with the linkage to different climatic factors.

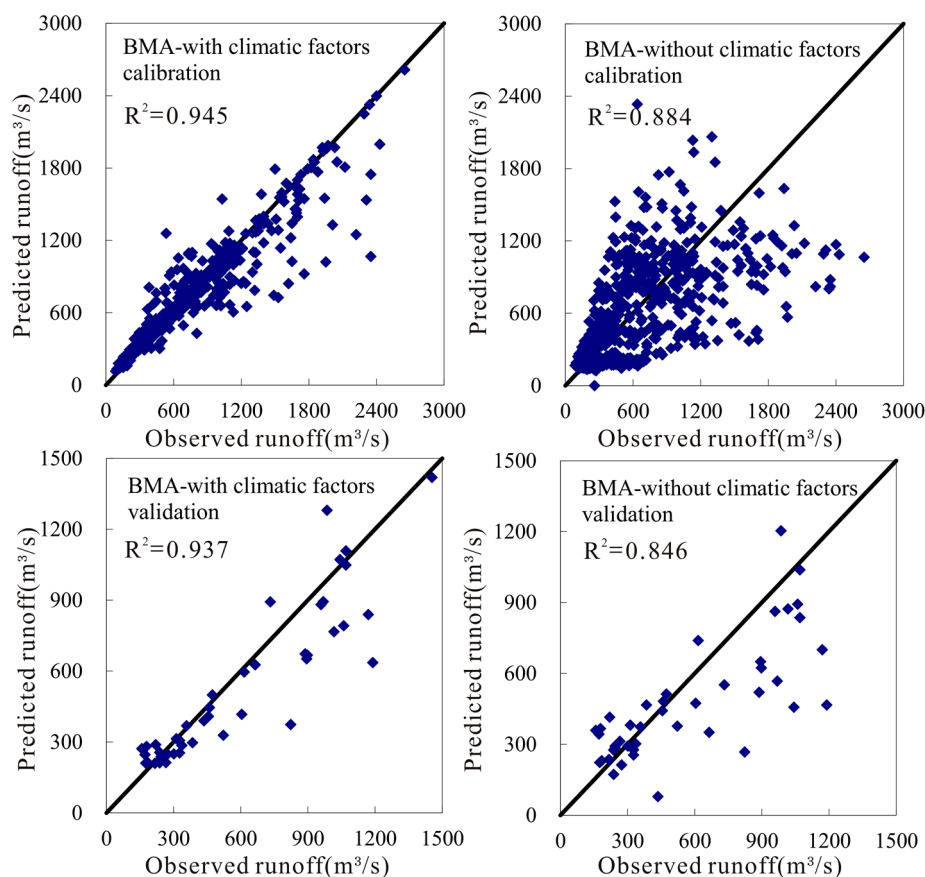


Figure 7. Predicted and observed runoff using the BMA-based multimodel, with and without climatic factors.

5. Conclusions

The Yellow River headwater region is a headstream of other major rivers in China, and water resources are important for the surrounding area, so accurate runoff forecasting is of great significance for rational development and utilization of water resources. The monthly runoff data from Tangnaihai station in the Yellow River headwater region of China were analyzed as the case study. In this paper, under the full consideration of climatic factors, MLR, RBFNN, and SVR models were built for medium- and long-term runoff forecasting. The performance of the SVR model has changed significantly with different parameters. Hence, the main parameters of the models should be selected carefully. To further improve medium- and long-term runoff forecasting, a BMA-based multimodel was developed by the weighted sum of these aforementioned three models, and the forecasting results obtained using the BMA-based multimodel were compared to those obtained using the MLR, RBFNN, and SVR models during model calibration and validation. The results showed that the BMA-based multimodel obviously outperformed the RBFNN and MLR models during model calibration and validation. The BMA-based multimodel improved forecasting accuracy in terms of the standard statistical measures, and it also performed better for high-runoff forecasting than the other models, and gave forecasting results under the 90%, 95%, and 99% confidence intervals, respectively. In addition, the multimodel, with and without climatic factors, were built for comparison. However, the values predicted by the BMA-based multimodel are generally lower than the observed values when forecasting high runoff, and all of the models failed to capture extreme runoff events ($3500 \text{ m}^3/\text{s}$). Since accurate prediction of extreme events would provide the most economic, environmental, and societal benefit, further improvements in high-runoff prediction are required. Future research aimed at improving the forecasting model will take into consideration about some new approaches, such as

deep learning model to build the complex nonlinear relationship between runoff and precipitation, temperature, and large-scale climate information.

Acknowledgments: This work has been sponsored in part by the National Key Research and Development Project from the Ministry of Science and Technology during the Thirteenth Five-year Plan Period (2017YFC0403600), National Natural Science Foundation of China (51459003), the Chinese Ministry of Water Resources special funds for scientific research on public causes (Grant No. 201501028), the Science and Technology Projects State Grid Corporation of China (Grant No. 52283014000T), and the China Postdoctoral Science Foundation (No. 41022207). Comments and suggestions from anonymous reviewers, the associate editor, and the editor are greatly appreciated.

Author Contributions: Haibo Chu conducted the data analysis work and wrote the paper; Jiahua Wei provided the method, guided the entire study, and contributed to the discussion part; and Jiaye Li, Zhen Qiao, and Jiongwei Cao helped in analyzing the data.

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

References

1. Zhang, X.; Srinivasan, R.; Debele, B.; Hao, F. Runoff simulation of the headwaters of the yellow river using the SWAT model with three snowmelt Algorithms1. *J. Am. Water Resour. Assoc.* **2008**, *44*, 48–61. [[CrossRef](#)]
2. Mao, T.; Wang, G.; Zhang, T. Impacts of Climatic Change on Hydrological Regime in the Three-River Headwaters Region, China, 1960–2009. *Water Resour. Manag.* **2016**, *30*, 115–131. [[CrossRef](#)]
3. Chua, L.H.; Wong, T.S. Runoff forecasting for an asphalt plane by Artificial Neural Networks and comparisons with kinematic wave and autoregressive moving average models. *J. Hydrol.* **2011**, *397*, 191–201. [[CrossRef](#)]
4. Li, H.; Xie, M.; Jiang, S. Recognition method for mid-to long-term runoff forecasting factors based on global sensitivity analysis in the Nenjiang River Basin. *Hydrol. Process.* **2012**, *26*, 2827–2837. [[CrossRef](#)]
5. Schneeberger, K.; Dobler, C.; Huttenlau, M.; Stötter, J. Assessing potential climate change impacts on the seasonality of runoff in an Alpine watershed. *J. Water Clim. Chang.* **2015**, *6*, 263–277. [[CrossRef](#)]
6. Shoaib, M.; Shamseldin, A.Y.; Melville, B.W.; Khan, M.M. Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. *J. Hydrol.* **2015**, *527*, 326–344. [[CrossRef](#)]
7. Shi, B.; Hu, C.H.; Yu, X.H.; Hu, X.X. New fuzzy neural network—Markov model and application in mid-to long-term runoff forecast. *Hydrol. Sci. J.* **2016**, *61*, 1157–1169. [[CrossRef](#)]
8. Abudu, S.; King, J.P.; Bawazir, A.S. Forecasting monthly streamflow of Spring-Summer runoff season in rio grande headwaters basin using stochastic hybrid modeling approach. *J. Hydrol. Eng.* **2010**, *16*, 384–390. [[CrossRef](#)]
9. Evsukoff, A.G.; de Lima, B.S.; Ebecken, N.F. Long-term runoff modeling using rainfall forecasts with application to the Iguaçu River Basin. *Water Resour. Manag.* **2011**, *25*, 963–985. [[CrossRef](#)]
10. Talei, A.; Chua, L.H.C.; Quek, C.; Jansson, P.E. Runoff forecasting using a Takagi–Sugeno neuro-fuzzy model with online learning. *J. Hydrol.* **2013**, *488*, 17–32. [[CrossRef](#)]
11. Elsanabary, M.H.; Gan, T.Y. Weekly streamflow forecasting using a statistical disaggregation model for the Upper Blue Nile basin, Ethiopia. *J. Hydrol. Eng.* **2014**, *20*, 04014064. [[CrossRef](#)]
12. Djibo, A.G.; Karambiri, H.; Seidou, O.; Sittichok, K.; Paturel, J.E.; Saley, H.M. Statistical seasonal streamflow forecasting using probabilistic approach over West African Sahel. *Nat. Hazards* **2015**, *79*, 699–722. [[CrossRef](#)]
13. Biondi, D.; De Luca, D.L. Rainfall-runoff model parameter conditioning on regional hydrological signatures: Application to ungauged basins in southern Italy. *Hydrol. Res.* **2017**, *48*, 714–725. [[CrossRef](#)]
14. De Luca, D.L. Analysis and modeling of rainfall fields at different resolutions in Southern Italy. *Hydrol. Sci. J.* **2014**, *59*, 1536–1558. [[CrossRef](#)]
15. Block, P.J.; Souza Filho, F.A.; Sun, L.; Kwon, H.H. A streamflow forecasting framework using multiple climate and hydrological models1. *J. Hydrol. Eng.* **2009**, *45*, 828–843.
16. Zhang, Q.; Wang, B.D.; He, B.; Peng, Y.; Ren, M.L. Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. *Water Resour. Manag.* **2011**, *25*, 2683–2703. [[CrossRef](#)]
17. Wang, W.C.; Chau, K.W.; Qiu, L.; Chen, Y.B. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environ. Res.* **2015**, *139*, 46–54. [[CrossRef](#)] [[PubMed](#)]

18. Zheng, H.; Zhang, L.; Liu, C.; Shao, Q.; Fukushima, Y. Changes in stream flow regime in headwater catchments of the Yellow River basin since the 1950s. *Hydrol. Process.* **2007**, *21*, 886–893. [[CrossRef](#)]
19. Lan, Y.; Zhao, G.; Zhang, Y.; Wen, J.; Hu, X.; Liu, J.; Gu, M.; Chang, J.; Ma, J. Response of runoff in the headwater region of the Yellow River to climate change and its sensitivity analysis. *J. Geogr. Sci.* **2010**, *20*, 848–860. [[CrossRef](#)]
20. Wang, W.; Van Gelder, P.H.; Vrijling, J.K.; Ma, J. Forecasting daily streamflow using hybrid ANN models. *J. Hydrol.* **2006**, *324*, 383–399. [[CrossRef](#)]
21. Zhang, F.; Dai, H.; Tang, D.; Sun, Y. Research on runoff predicting based on wavelet neural network conjunction model. In Proceedings of the IEEE 2013 International Conference on Information Science and Cloud Computing Companion (ISCC-C), Guangzhou, China, 7–8 December 2013; pp. 841–845.
22. Schilling, K.E.; Walter, C.F. Estimation of streamflow, base flow, and nitrate-nitrogen loads in Iowa using multiple linear regression model. *J. Am. Water Resour. Assoc.* **2005**, *41*, 1333–1346. [[CrossRef](#)]
23. Petković, D.; Gocic, M.; Shamshirband, S.; Qasem, S.N.; Trajkovic, S. Particle swarm optimization-based radial basis function network for estimation of reference evapotranspiration. *Theor. Appl. Climatol.* **2016**, *125*, 555–563. [[CrossRef](#)]
24. Phukoetphim, P.; Shamseldin, A.Y.; Adams, K. Multimodel Approach Using Neural Networks and Symbolic Regression to Combine the Estimated Discharges of Rainfall-Runoff Models. *J. Hydrol. Eng.* **2016**, *21*, 04016022. [[CrossRef](#)]
25. Kharroubi, O.; Blanpain, O.; Masson, E.; Lallahem, S. Application of artificial neural networks to predict hourly flows: Case study of the Eure basin, France. *Hydrol. Sci. J.* **2016**, *61*, 541–550. [[CrossRef](#)]
26. El Shafie, A.H.; El-Shafie, A.; Almukhtar, A.; Taha, M.R.; El Mazoghi, H.G.; Shehata, A. Radial basis function neural networks for reliably forecasting rainfall. *J. Water Clim. Chang.* **2012**, *3*, 125–138. [[CrossRef](#)]
27. Kalteh, A.M. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput. Geosci.* **2013**, *54*, 1–8. [[CrossRef](#)]
28. Jafarnejadsani, H.; Pieper, J.; Ehlers, J. Adaptive control of a variable-speed variable-pitch wind turbine using radial-basis function neural network. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 2264–2272. [[CrossRef](#)]
29. Yu, P.S.; Chen, S.T.; Chang, I.F. Support vector regression for real-time flood stage forecasting. *J. Hydrol.* **2006**, *328*, 704–716. [[CrossRef](#)]
30. Hong, W.C. Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model. *Energy Convers. Manag.* **2009**, *50*, 105–117. [[CrossRef](#)]
31. Clarke, S.M.; Griebisch, J.H.; Simpson, T.W. Analysis of support vector regression for approximation of complex engineering analyses. *J. Mech. Des.* **2005**, *127*, 1077–1087. [[CrossRef](#)]
32. Wu, C.H.; Tzeng, G.H.; Lin, R.H. A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. *Expert Syst. Appl.* **2009**, *36*, 4725–4735. [[CrossRef](#)]
33. Kang, P.; Kim, D.; Cho, S. Semi-supervised support vector regression based on self-training with label uncertainty: An application to virtual metrology in semiconductor manufacturing. *Expert Syst. Appl.* **2016**, *51*, 85–106. [[CrossRef](#)]
34. Duan, Q.; Ajami, N.K.; Gao, X.; Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Res.* **2007**, *30*, 1371–1386. [[CrossRef](#)]
35. McLean Sloughter, J.; Gneiting, T.; Raftery, A.E. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Weather Rev.* **2013**, *141*, 2107–2119. [[CrossRef](#)]

