

Research Article

A Comparative Study of VMD-Based Hybrid Forecasting Model for Nonstationary Daily Streamflow Time Series

Hui Hu, Jianfeng Zhang , and Tao Li 

State Key Laboratory Base of Eco-Hydraulic Engineering in Arid Area, School of Water Resources and Hydropower, Xi'an University of Technology, Xi'an 710048, China

Correspondence should be addressed to Tao Li; litao@xaut.edu.cn

Received 14 February 2020; Revised 21 June 2020; Accepted 27 June 2020; Published 30 July 2020

Academic Editor: Marcelo Messias

Copyright © 2020 Hui Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data-driven methods are very useful for streamflow forecasting when the underlying physical relationships are not entirely clear. However, obtaining an accurate data-driven model that is sufficiently performant for streamflow forecasting remains often challenging. This study proposes a new data-driven model that combined the variational mode decomposition (VMD) and the prediction models for daily streamflow forecasting. The prediction models include the autoregressive moving average (ARMA), the gradient boosting regression tree (GBRT), the support vector regression (SVR), and the backpropagation neural network (BPNN). The latest decomposition model, the VMD algorithm, was first applied to extract the multiscale features from the entire time series and to decompose them into several subseries, which were predicted after that using forecast models. The ensemble forecast was finally reconstructed by summing. Historical daily streamflow series recorded at the Wushan and Weijiabao hydrologic stations from 1 January 2001 to 31 December 2014 in China were investigated using the proposed VMD-based models. Three quantitative evaluation indexes, including the Nash–Sutcliffe efficiency coefficient (NSE), the root mean square error (RMSE), and the mean absolute error (MAE), were used to evaluate and compare the predicted results of the proposed VMD-based models with two other models such as nondecomposition method (BPNN) and BPNN based on ensemble empirical mode decomposition (EEMD-BPNN). Furthermore, a comparative analysis of the performance of the VMD-BPNN model under different forecast periods (1, 3, 5, and 7 days) was performed. The results evidenced that the proposed VMD-based models could always achieve good performance in the testing stage and had relatively good stability and representativeness. Specifically, the VMD-BPNN model considered both the prediction accuracy and computation efficiency. The results show that the reliability of the forecasting decreased as the foresight period increased. The model performed satisfactorily up to 7-d lead time. The VMD-BPNN model could be applied as a promising, reliable, and robust prediction tool for short-term streamflow forecasting modelling.

1. Introduction

Streamflow forecasting, especially daily streamflow forecasting, is an important task for optimizing the allocation of water resources and providing effective flood control measures [1]. For this reason, streamflow forecasting has received significant attention from the scientific community in the recent decades, and many models were proved to be instrumental for forecasting river flow to improve the prediction accuracy [2, 3].

Numerous classical black box time series models, which include the autoregression (AR) model, the autoregressive

moving average (ARMA) model, and the autoregressive integrated moving average (ARIMA) model [4, 5], have been used in streamflow forecasting since 1970. These models are linear and therefore miss the nonlinear and nonstationary characteristics that are hidden in the real streamflow series. Hence, researchers have focused on developing strong nonlinear mapping abilities (machine learning techniques) to overcome these drawbacks, including decision trees such as the gradient boosted regression tree (GBRT) [6, 7], the kernel methods such as the support vector machine (SVM) [8, 9], and the support vector regression (SVR) [10]. The SVM [11, 12] and the SVR [13, 14] have been used in the field

of streamflow forecasting research. It is important to remark that the artificial neural networks (ANN) represent the most widely applied artificial intelligence techniques for modelling [15–17], and it has been widely used in hydrology [18–20]. The backpropagation neural network (BPNN) is the improvement of the ANN learning representations by error backpropagation algorithm, which is the most popular neural network. Many extensions and modifications of the BPNN have gained development in different fields in the past few years [21–23], including for instance their application to rainfall-runoff modelling, which has been relatively successful [24–26]. However, some limitations still need to be addressed; the main one being the slow learning speeds and the overfitting [27, 28], which is caused by the insufficiency of generalization ability. Especially for the use of BPNN in the hydrology field, it is difficult to obtain satisfactory prediction accuracy due to the great heterogeneity of the rainfall-runoff process.

Although the abovementioned data-driven models have become appropriate alternatives to knowledge-driven models in hydrological forecasting and are both flexible and useful, they have limitations regarding the highly nonstationary characteristic of the hydrological series that vary over a range of scales (e.g., from daily to multidecadal) [29]. For this reason, recent developments in signal processing tools use data-driven models that can deal with the nonstationary datasets of hydrological signals and provide time-scale localization. There are many models for streamflow forecasting that are based on multidimensional feature extraction and feature learning. Wavelet Transforms (WT) [30], Empirical Mode Decomposition (EMD) [31], and Ensemble Empirical Mode Decomposition (EEMD) [32] are commonly used data-preprocessing techniques to feature the extraction of original data. WT has been applied in reservoir inflow modelling [33]. Although it possesses good time-frequency localization characteristics, the decomposition results mainly depend upon the mother wavelet and decomposition level, and its adaptability is relatively poor [34]. Several successful applications of EMD-based and EEMD-based modified model were ideally suitable for forecasting river flow [35–37]. However, one of the main drawbacks of the EMD is the frequent appearance of weak mode-mixing (i.e., the scale separation problem). EEMD is a significant improvement of the EMD since it adds white noise to compensate this drawback [32, 38]. Li and Liang [39] also showed that compared with the EMD and WT, EEMD has some special characteristics, including that it is self-adaptive, direct, intuitive, and empirical [40]. However, the EEMD involves a large number of calculations and the modal component is uncontrollable, which easily leads to the nonconvergence of the function and subsequently affects the accuracy of the algorithm [41].

To fill the aforementioned gaps, this paper designed a theoretically well-founded and robust decomposition method (VMD) and data-driven models referred to as VMD-based models to improve the accuracy and stability of the runoff forecasting. VMD was introduced into the prediction problem as a new adaptive time-frequency decomposition method and is substantially more sensitive to

sampling and noise than existing decomposition approaches, such as EMD and EEMD. Different VMD-based decomposed methods have been proposed and successfully applied for chatter detection in milling processes [42], vibroacoustic feature extraction [43], and container throughput forecasting [44]. However, they have been less applied for forecasting highly nonlinear and nonstationary streamflow series. In general, in streamflow forecasting, the model's performance (in terms of the model's accuracy) deteriorates as the lead time increases [45, 46]. Therefore, the single model based on VMD and BPNN, referred to as VMD-BPNN, is proposed to practically predict daily streamflow for different forecast periods (1, 3, 5, and 7 d) in this study.

In summary, the primary objective of this paper is to introduce a specific and novel decomposition method (VMD) and data-driven models referred to as VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN, which preserve the characteristics of hydrologic time series. To the best of the authors' knowledge, no existing studies have compared the different VMD and EEMD decomposition methods and the different GBRT, ARMA, SVR, and BPNN feature learning methods to forecast streamflows concerning both the model's accuracy and the precision in flow forecasting. The paper explores for the first time a new realm of hydrological modelling by combining VMD with prediction models and assessing both the model's accuracy and flow forecasting capability, which is a novel application of these decomposition algorithms in hydrology forecasting. Also, different lead-time forecasts were investigated to assess the accuracy of the models.

2. Methodology

2.1. VMD-Based Hybrid Model. To deal with the nonstationary problem of streamflows, a VMD-based hybrid model that incorporates the VMD algorithm with the ARMA, GBRT, SVR, and BPNN models was built for simulating the daily streamflow in the Wei River Basin. VMD was used to decompose the original time series into several subseries, and ARMA, GBRT, SVR, and BPNN were used to build the forecast model for each subseries. The subseries that were obtained by VMD were relatively stationary and could provide information about the original data structure and its periodicity. Therefore, the performance of the forecast models was expected to be improved by giving useful information on various resolution levels. The VMD-based models were based on the decomposition-ensemble framework, and it included the following three stages: (1) to decompose the runoff time series into a collection of IMFs using VMD, (2) to forecast each subseries using prediction models such as ARMA, GBRT, SVR, and BPNN, and (3) to obtain the final forecast by summing the outputs of the subseries. A series of experiments based on the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN model was conducted. A flowchart of the VMD-based models is illustrated in Figure 1. The framework of the decomposition-based data-driven model was represented as follows:

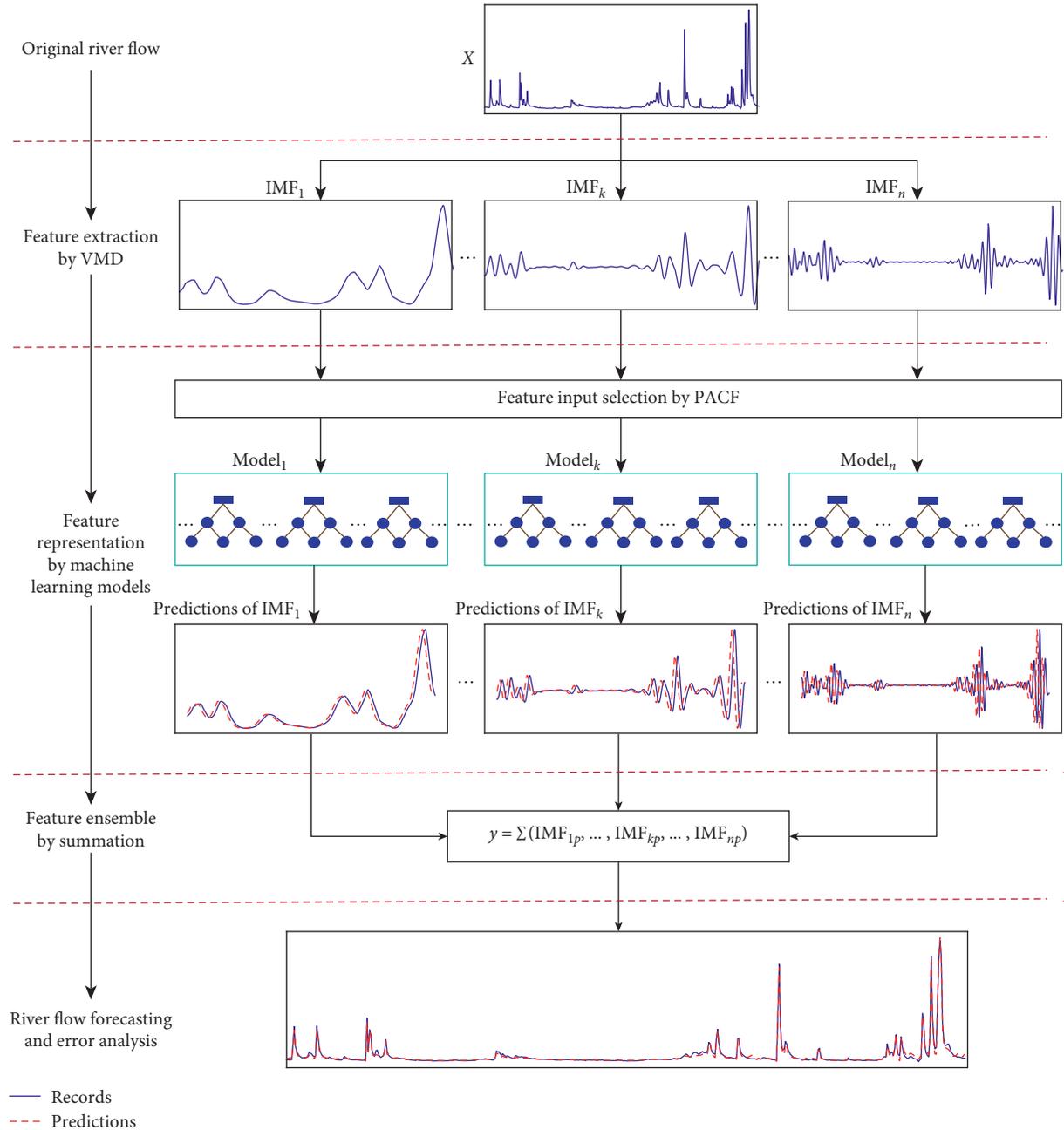


FIGURE 1: The technique of VMD-based multidimensional feature learning with hybrid models.

Step 1: decompose the entire original daily streamflow time series $Q = \{Q(1), Q(2), \dots, Q(N)\}$ into n IMFs using VMD and divide the entire period into the training and validation periods

Step 2: develop the ARMA, GBRT, SVR, and BPNN models for each subseries of the training period using the appropriate lags that are obtained by the PACF as inputs and then select the optimal model parameters using the error-index minimization criteria

Step 3: use the selected prediction models to forecast the IMFs

Step 4: obtain the final forecast of the training periods by summing the outputs of all the prediction models

Step 5: apply the ARMA, GBRT, SVR, and BPNN models to forecast the subseries in the test period and obtain the final forecast of the test period by adding the outputs of all prediction models

2.2. Multidimensional Feature Extraction with VMD. The VMD algorithm was used to decompose the original daily streamflow series into multiple components. It decomposed a complicated dataset into several intrinsic mode functions (IMFs) with different dominant frequencies and amplitudes.

Variational mode decomposition (VMD) first appeared in 2014 [47], and it was mainly used to decompose adaptively the input signal f into several subsignals (a modal

component function) $u_k(t)$. VMD decomposed the daily streamflow series X into $n+1$ IMFs, i.e., u_1, u_2, \dots, u_n . Each mode $u_k(t)$ had a finite bandwidth with different centre frequencies, and the sum of the estimated bandwidth of each mode was required to be minimal. For the sake of conciseness and convenience, all the IMFs and the original series are uniformly called subseries in the following section. The specific steps for the daily streamflow sequence decomposition are as follows:

Step 1: using the streamflow sequence as the original input signal f , the demodulated signal of the mode $u_k(t)$ is calculated using the Hilbert transform to obtain the corresponding unilateral spectrum. The spectrum of each mode is shifted to the baseband by mixing it with an exponential tuned to the predicted centre frequency. The squared L^2 -norm of the gradient of the demodulated signal is calculated to estimate the bandwidth of $u_k(t)$.

$$L(\{u_k\}, \{\omega_k\}, \lambda) := \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + f(t) - \sum_k u_k(t) \Big\|_2^2 + \lambda(t), f(t) - \sum_k u_k(t), \quad (2)$$

where the quadratic penalty α is a parameter to encourage reconstruction data fidelity and the Lagrangian multipliers $\lambda(t)$ are used to guarantee the strictness of the constraint.

$$u_k^{n+1} = \arg \min_{u_k \in X} \left\{ \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_i u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\}, \quad (3)$$

where ω_k is equivalent to ω_k^{n+1} and $\sum_i u_i(t)$ is equivalent to $\sum_{i \neq k} u_i(t)^{n+1}$.

Step 5: based on the Parseval/Plancherel Fourier transform, equation (3) is converted to equation (4) to update $u_k(t)$ in the frequency domain, and then the centre frequency ω_k is updated using equation (5). The update equation $\lambda(t)$ is shown as equation (6):

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^{n+1}(\omega) + (\hat{\lambda}(\omega)/2)}{1 + 2\alpha(\omega - \omega_k)^2}, \quad (4)$$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega}, \quad (5)$$

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega) \right), \quad (6)$$

Step 2: decompose the daily streamflow sequence (f) into the set of K modes and construct the corresponding constrained variational model [47] as follows:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (1)$$

$$\text{s.t.} \quad \sum_k u_k = f,$$

where $\{u_k\} := \{u_1, \dots, u_k\}$ and $\{\omega_k\} := \{\omega_1, \dots, \omega_k\}$ are shorthand notations of all the K modes and their centre frequencies, respectively, $\delta(t)$ represents the Dirac distribution, and t is the time. Equally, $\sum_{k=1}^K$ refers to the summation over the set of all modes.

Step 3: to render the constrained variational problem into an unconstrained one, the quadratic penalty term α and the Lagrangian multipliers $\lambda(t)$ are introduced. The augmented Lagrangian function is expressed as follows [47]:

Step 4: using the alternate direction method of the multipliers (ADMM) [48], u_k^{n+1} , ω_k^{n+1} , and λ_k^{n+1} are alternately updated to find the saddle point of the augmented Lagrange function of equation (2) such that u_k^{n+1} is expressed as follows:

where τ denotes the time step.

Step 6: in the process of solving the iterative solution of the variational model, repeat steps 2 to 5. The frequency centre and the bandwidth of each streamflow sequence exponential component are continuously updated to complete the adaptive segmentation of the signal band until the iteration stop condition is met:

$$\frac{\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < e. \quad (7)$$

Given the discriminative precision $e > 0$ for equation (7), the whole loop is ended. Finally, the mode $u_k(t)$ and the corresponding centre frequencies are obtained according to the frequency domain characteristics of the daily streamflow sequence.

2.3. Preliminary Processing-PACF. The development of the prediction model depends largely on the determination of the input variables. This requires some degree of a priori knowledge of the system to be modelled [49, 50]. However, it is difficult to select the initial input variables based on the underlying physical processes, especially in complex systems. Therefore, an analytical method, such as auto- or crosscorrelation, is often employed [51]. This paper used a statistical method (i.e., the partial autocorrelation function (PACF)) to address the limitation of neglecting the relationship between the number of parameters and the different antecedent values. Detailed descriptions of the PACF can be found in the previous literature [52]. The output variable was strongly correlated with $X(t)$ (current flow) and $X(t-i)$ (antecedent flow with i -day time lag $t-i$), which were thus selected as the input variables if the lag autocorrelation length i exceeded 95% confidence interval $[-(1.96/\sqrt{n}), (1.96/\sqrt{n})]$.

2.4. Models for Learning Prediction. Several techniques are employed for feature learning in this study, including the ARMA, GBRT, SVR, and BPNN models. A brief overview of these methods is presented here.

2.4.1. GBRT Model for Learning Prediction. The traditional GBRT model, according to Friedman [53], solves the following classification and the regression problem to combine the output of many weak prediction models or “learners” into a powerful “committee” [54]. At each stage $1 \leq m \leq M$ of the gradient boosting algorithm, we calculate the following:

$$F_m(x) = F_{m-1}(x) + h_m(x), \quad (8)$$

where $h_m(x)$ are the basic functions that are referred to as weak learners and small regression trees of fixed size in the

case of the GBRT. The GBRT model $F_m(x)$ is a summation of m small regression trees. For each boosting iteration from $m = 1$ to M , the gradient boosting algorithm improves the F_{m-1} that determines the new model F_m by adding a new regression tree h_m to improve the previous model. The procedure estimates the target value y based on a perfect h_m from the training set, such that h_m satisfies the following:

$$F_m(x) = F_{m-1}(x) + h_m(x) = y, \quad (9)$$

and also

$$h_m(x) = y - F_{m-1}(x). \quad (10)$$

Thus, h_m in equation (10) is the regression tree model fitting the residual $y - F_{m-1}(x)$ at the iteration m . Also, the current residuals for a given model are the negative gradients of the squared error loss function:

$$\frac{\partial (1/2)(y - F_{m-1}(x))^2}{\partial F_{m-1}(x)} = y - F_{m-1}(x). \quad (11)$$

This shows that h_m equals the negative gradient of the squared loss function. Therefore, equation (11) proves that the gradient boosting is a gradient descent algorithm minimizing the squared error loss function. This explanation can be generalized to other loss functions by replacing the squared error with a different loss function and its gradient. Further details of the gradient boosted regression trees are provided in Hastie et al. [54].

The objective of the model is to find an optimal estimate $\hat{F}(x)$ that minimizes the value of the loss function using the training set $\{(x_1, y_1), \dots, (x_n, y_n)\}$. A simplification is applied, and the steepest descent step is used to solve this minimization problem. The model is updated by the following equations [54, 55]:

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)), \gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L\left(y_i, F_{m-1}(x_i) - \gamma \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right), \quad (12)$$

where the derivatives are obtained with respect to the functions F_i for $i \in \{1, 2, \dots, m\}$.

In the m th step, a generic gradient boosting algorithm fits a regression tree $h_m(x)$ to the pseudoresiduals. Let J_m be the number of leaves. The regression tree model splits the input space into J_m separated regions $R_{1m}, \dots, R_{J_m m}$ and obtains a constant value for each region. For the input x , $h_m(x)$ is then written as the following sum [56]:

$$h_m(x) = \sum_{i=1}^{J_m} b_{jm} I(x \in R_{jm}), \quad (13)$$

where b_{jm} is the constant value that is calculated for the region R_{jm} . Each coefficient b_{jm} is multiplied by a different optimal γ_{jm} for each of the regions, which is carried out by a modified algorithm which was proposed by Friedman instead of using only one γ_m for the whole tree. After that the model is updated as follows [54, 55]:

$$F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}), \gamma_{jm} = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma). \quad (14)$$

To obtain the optimal model, understanding the effect of the different parameter values on the model's performance is critical. The well-known machine learning toolkit scikit-learn [57] was applied to train the GBRT model using training and developing dataset. To achieve lower prediction errors, 6-fold crossvalidation was used. The hyperparameters, i.e., the learning rate (LR), the maximum depth (MD), the maximum features (MF), the minimum sample split (MSS), and the minimum sample leaf (MSL), are searched for using Bayesian optimization based on the Gaussian processes.

2.4.2. ARMA Model for Learning Prediction. ARMA models [58] are linear stochastic models that are obtained by combining the AR and the MA models, and they are used to model the dependent stochastic components of a time series. ARMA models can be written as follows:

$$\varphi(B)Y_t = \theta(B)\varepsilon_t, \quad (15)$$

where $\varphi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ represents the autoregressive operator of the order p of the ARMA(p, q) model, B denotes the backward operator, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ is the moving average operator of the order q of the ARMA(p, q) model, and ε_t is the white noise sequence (residuals). The order of the model was identified in order to fit the ARMA(p, q) model to the data. Various types of ARMA(p, q) models should be validated to find the best choice for each series. For this purpose, the AIC is generally used such that the model with the lowest AIC value is accepted as the best among the alternative models and is written as follows:

$$\text{AIC} = N \times \ln(\sigma_\varepsilon^2) + 2 \times (p + q), \quad (16)$$

where N is the total number of observations and σ_ε^2 is the variance of the residual terms.

2.4.3. SVR Model for Learning Prediction. The support vector regression (SVR) is a powerful nonlinear regression model that was developed from the support vector machine (SVM) [59]. In the SVR, the aim is to solve the following minimization problem [60] to find the parameter C of the model for an appropriate function $f(x) = w\phi(x) + b$ with a small w (a weight vector) that has the least possible deviation (ε) for a set of N -training data, where $t = 1, 2, \dots, N$, $x_t \in \mathfrak{R}^d$, $y_t \in \mathfrak{R}$, and $\phi(x)$ is a nonlinear mapping function. Another parameter that defines the performance of an SVR model is the kernel parameter γ , which nonlinearly maps samples into a high-dimensional feature space. The loss function is defined as follows:

$$\min \phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^l \xi_i \right), \quad (17)$$

which is subjected to the following:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \dots, N, \quad (18)$$

where ξ_i is a slack variable, C is a positive regularized parameter, and b is the weight vector.

C , ε , and γ are determined through a trial-and-error process. The modelling process for an SVR model is presented as follows: (1) set all the possible combinations for $[C, \gamma, \varepsilon]$; (2) split the data into 5 folds with all possible combinations; (3) train and validate with all the possible C , ε , γ , and fold combinations; (4) keep the C , ε , and γ that yielded the lowest RMSE and the highest R^2 in the training and developing sets; (5) train the model using all the data with the selected C , ε , and γ ; and (6) test the SVR model using the testing data.

2.4.4. BPNN Model for Learning Prediction. BPNN is the error backpropagation algorithm. Developed by Rumelhart et al. [61], it is one of the most common and effective approaches amongst all neuron networks that consist of an input layer, a hidden layer, and an output layer. The number of hidden neurons depends on the complexity of the mathematical problem [62] and is often determined by trial and error. Each input I_j and output O_j of neuron j are calculated by

$$I_j = \sum_i w_{ij} O_i, \quad O_j = f(I_j + \theta_j), \quad (19)$$

where w_{ij} is the weight of the connection from the input to the hidden layer, or from the hidden to the output layer, $f(I_j + \theta_j)$ is an activation function, and θ_j is the bias input to the neuron.

For the training or learning process, where the weights are selected, the neural network uses the gradient descent learning method to modify the weights and to minimize the error between the actual output values and the expected values [63]. It stops when the errors are minimized or when another stopping criterion is met. The weights and biases are modified using the following formulas:

$$\begin{aligned} w_{ji}(k+1) &= w_{ji}(k) + \eta \delta_j O_i, \\ \theta_i(k+1) &= \theta_i(k) + \eta \delta_j, \end{aligned} \quad (20)$$

where k is the number of epochs, δ_j is the backpropagation errors in the hidden layer, and η is the learning rate.

2.5. Model Evaluation Criteria. Several error indices are commonly used to evaluate the prediction performance of the model. These include the root mean squared error (RMSE), the mean absolute error (MAE), and the Nash-Sutcliffe efficiency coefficient (NSE). The RMSE and MAE were used to characterize the overall precision and accuracy of the prediction results, respectively. The NSE was used to characterize the stability of the prediction results. If the forecasting errors exhibited large fluctuation, the model was deemed unreliable, even if the other error indices were high. The NSE was sensitive to the fluctuation of the data series and could describe the tracking ability of the prediction to the measurements. The RMSE, MAE, and NSE are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^N (Q_i - \hat{Q}_i)^2}, \text{MAE} = \frac{1}{n} \sum_{i=1}^N |Q_i - \hat{Q}_i|, \text{NSE} = 1 - \frac{\sum_{i=1}^N (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^N (Q_i - \bar{Q})^2}, \quad (21)$$

where n is the number of samples; Q_i and \hat{Q}_i are the observed and forecasted values, respectively; and \bar{Q} is the mean of the observed values. The best fit between the observed and forecasted values would yield $\text{RMSE} = 0$, $\text{NASH} = 1$, and $\text{MAE} = 0$.

2.6. Data Normalization. To achieve better performance and faster convergence, the data normalization was performed on the raw series according to the following equation:

$$Q_{i_normalized} = 2 * \frac{Q_i - Q_{min}}{Q_{max} - Q_{min}} - 1, \quad (22)$$

where $Q_{i_normalized}$ is the normalized data, Q_i is the raw data, and Q_{max} and Q_{min} are the maximum and minimum values of the sequence, respectively. The data series will be normalized in the range of -1 and 1 . Firstly, the original streamflow series were decomposed by the VMD method. The resulting decomposed data were normalized in equation (22), and the normalized data were subsequently used to learn feature by the different forecasting models.

3. Case Study and Catchment Description

The Wei River (Figure 2) was used as a case study in this research to demonstrate the effectiveness of the proposed hybrid model. The Wei River, which is the largest tributary of the Yellow River, originates from the Mountain Bird Mouse in Gansu Province, flows through the Gansu and Shaanxi provinces and converges into the Yellow River in Tongguan with a total length of 818 km. The basin covers an area of 135,000 km² (103.5°E–110.5°E and 33.5°N–37.5°N). Located in the continental monsoon climate zone, the Wei River Basin (WRB) is characterized by abundant precipitation in summer and rare precipitation in winter [27]. The basin has an annual precipitation of approximately 559 mm. Topographically, the altitude of the basin is high in the northwest mountainous areas and low in the Guanzhong Plain. It is worth mentioning that the Wei River is an important grain production area and a state key economic development zone in China's "one belt and one road" construction. Therefore, its economic development will directly affect the sustainable development of the economic society of the Gansu and Shaanxi provinces. Considering the development of the economy and the population growth, the WRB water resources demand increases noticeably. Therefore, it is of great significance to develop a hybrid forecasting model for streamflow in the WRB.

Because the observations of meteorological factors such as rainfall and temperature are incomplete in the study area, only the historical runoff data collected from the Wushan and the Weijiabao hydrological stations in the WRB were collected to assess the proposed model. The Wushan and the

Weijiabao hydrological stations correspond to the main upstream and midstream downstream stations of the Wei River, respectively. The daily streamflow data were collected from the published hydrologic manual of the WRB and the Hydrological and Water Resources Information Network of Shaanxi Province (<http://www.shxsw.com.cn/7/39/list.aspx>). The data quality was strictly controlled during its release. A consistency analysis has demonstrated that all the daily streamflow data used in this study were reliable. Based on the importance and representativeness of the hydrological series, the daily streamflow records (5113 total samples from 1 January 2001 to 31 December 2014) of the Wushan and Weijiabao hydrological station were used (Figure 3). To avoid local minima problems, the historical series of daily runoff data of the Wushan and the Weijiabao stations from 1 January 2001 to 31 December 2011, were selected for training the model structures in the model calibration, and the remainder (datasets from 1 January 2012 to 31 December 2014) for validation. The statistical properties of daily streamflow of two hydrological stations are shown in Table 1, indicating that the standard deviation values are higher than skewness coefficient in both stations.

4. Results and Discussion

4.1. Data Decomposition with VMD. According to the model established in Section 2.1, the initial daily streamflow series were decomposed into several IMFs using the VMD approach. The optimal value K could be determined by analyzing the centre frequency [64]. Noticeable aliasing of the centre frequency for the frequency spectrum was found to appear when the number of components increased to a certain value. To limit spurious components in the decomposition results to the minimum in this study, six multidimensional features (IMF_1 – IMF_6) and eleven decomposition components (IMF_1 – IMF_{11}) were decomposed by the VMD algorithm at the Wushan and the Weijiabao hydrological stations, respectively. Figure 4 shows the decomposition results of the different dominant frequencies and amplitudes for the Weijiabao station.

4.2. Selection of the Best Input Variables by PACF. Figure 5 depicts the PACF analysis results of the subseries for the Weijiabao station, where PACF_1 – PACF_{11} represent the respective PACFs of the normalized subseries. The input variables were determined by analyzing the resulting partial autocorrelation diagram, which corresponds to the plots of the PACFs with respect to the lag length (Figure 5). In the iterative process of the daily streamflow prediction, the input variables of eleven subseries and the number of inputs for the prediction models were determined, as illustrated in Table 2.

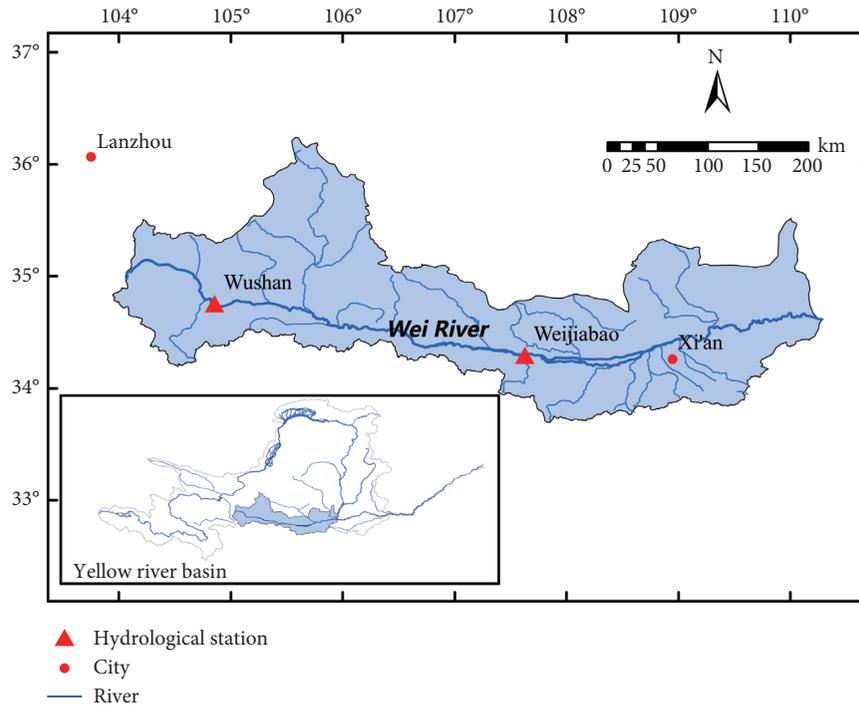


FIGURE 2: Study area and location of the two hydrological stations in the Wei River Basin, China.

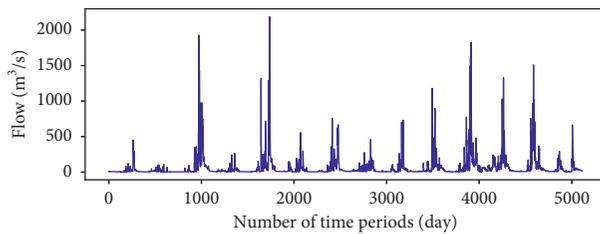


FIGURE 3: The observed daily runoff for the Weijiabao Hydrological Station from 2001 to 2014.

TABLE 1: The characteristic parameters of daily streamflow at the Wushan and Weijiabao hydrological station (m^3/s).

Station	Data range	Max	Min	Mean	S_x	C_{sx}
Wushan	2001/1/1~2014/12/31	643	0.03	11.17	17.87	11.48
Weijiabao	2001/1/1~2014/12/31	2190	1.98	57.09	137.38	6.45

Max, Min, Mean, S_x , and C_{sx} denote the minimum value, maximum value, average value, standard deviation value, and skewness coefficient of streamflow series, respectively.

4.3. Performance of the Different Forecasting Models and Decomposition in Preprocessing Methods

4.3.1. Performance of the GBRT Forecasting Model and Decomposition in Preprocessing Methods. The feasibility of the GBRT modelling was demonstrated by comparing the predicted streamflow with the observed results using the training set, which employed Bayesian optimization for the hyperparameter search. Table 3 presents the appropriate

values of five parameters, including the LR, MD, MF, MSS, and MSL for $\text{IMF}_1\text{--}\text{IMF}_{11}$ of the Weijiabao study sites. Table 3 shows that the RMSE is relatively small and tends towards zero, and the NSE values exceed 0.94, which indicates that the VMD-GBRT methods was very performant during both the training and developing periods. Therefore, this optimal model can be used to learn the features of the decomposed results in the test phase.

4.3.2. Performance of the ARMA Forecasting Model and Decomposition in Preprocessing Methods. According to the AIC model selection criterion, the values of the parameters p and q were assured during the training and developing phases. The best ARMA model has been determined as the one that yielded the lowest RMSE and the highest NSE values. Table 4 presents the values of the parameters and their errors for all the IMFs for the Weijiabao station. For all sequences, the RMSE tends towards zero, and the NSE values exceeded 0.995, which indicates that the VMD-ARMA methods have good performance. Thus, this optimal model could be used to learn the features of the decomposed results in the test phase.

4.3.3. Performance of SVR Forecasting Model and Decomposition in Preprocessing Methods. For the SVR model, the Bayesian optimization was employed to optimize the balance parameter C , the least possible deviation ϵ , and the kernel function parameter γ in the SVR models. Using the Weijiabao station as an example, eleven experiments, designated as $\text{IMF}_1\text{--}\text{IMF}_{11}$ series, were conducted for each training set of each mode. The parameter optimization results are listed

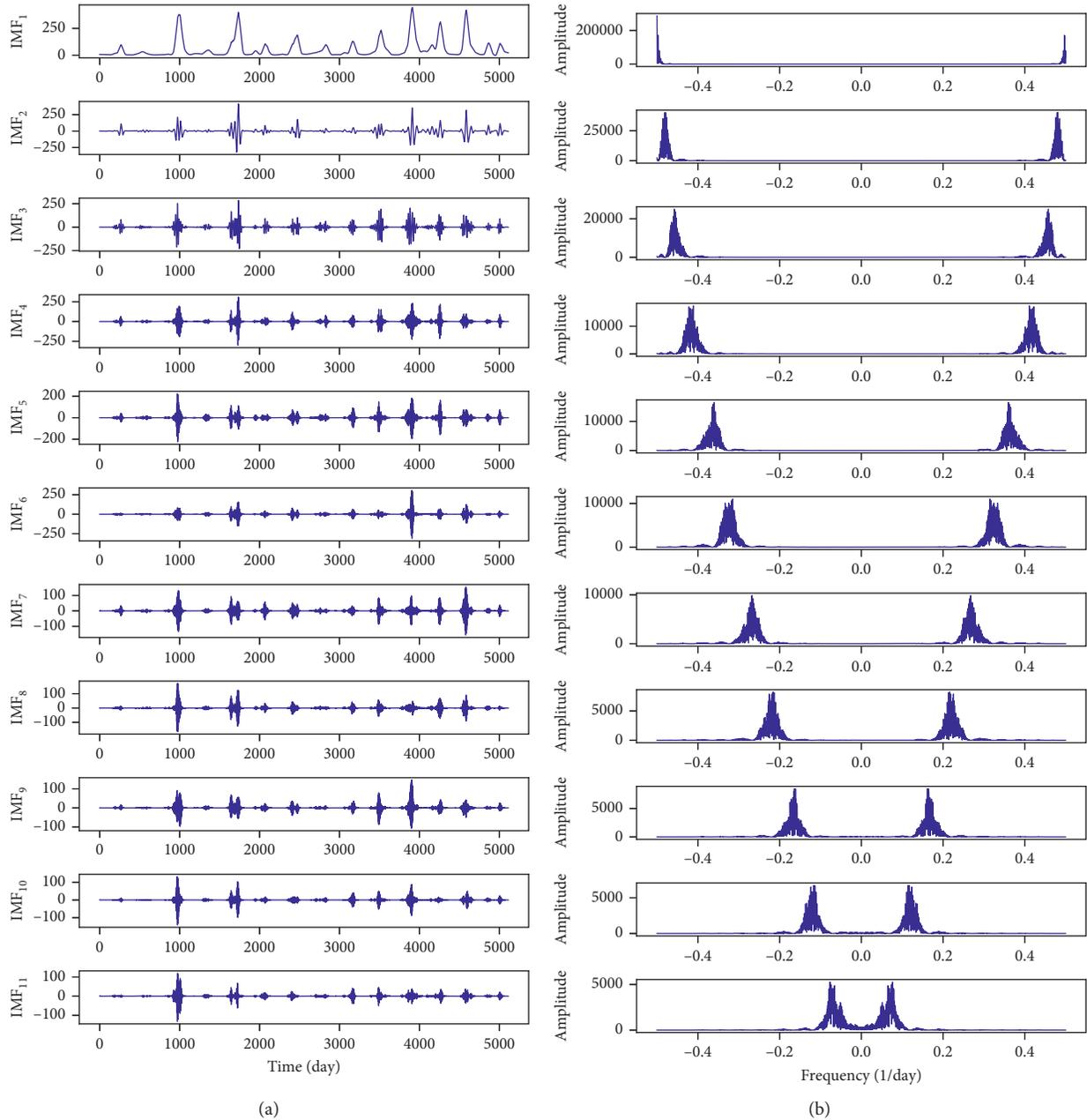


FIGURE 4: The decomposed subseries results of the daily streamflows using VMD at the Weijiabao station: (a) the decomposition sequence waveform and (b) the frequency spectrum representation.

in detail in Table 5. The results of the error show that these parameters were so appropriate that they could effectively eliminate the errors. Hence, the optimal model and parameters in Table 5 can be further applied to the learning of the decomposed IMFs in the test period.

4.3.4. Performance of the BPNN Forecasting Model and Decomposition in the Preprocessing Method. The BPNN model was developed for the feature learning of the decomposed IMFs. The optimum structure of the BPNN derived through trial and error is the one that decreases or increases the number of layers and neurons in each layer to

determine the optimal parameter settings. In this experiment, the BPNN training structures of IMF₁ for the Weijiabao station were initialized with 11 input neurons and 1 output neuron. The number of hidden layers was initially set to 1, the hidden neurons in a hidden layer were divided into 10 levels from 4 to 13 in training with a step size of 1, and the stopping condition was based on the BPNN reaching convergence. Table 6 shows the BPNN evaluation results with 11 input neurons and 1 hidden layer based on the two performance criteria. When the hidden neurons of one layer were 7, the NSE was 0.9999942, and the RMSE was 0.2182 m³/s, the model exhibited the best learning performance evaluation. The analysis of He et al.

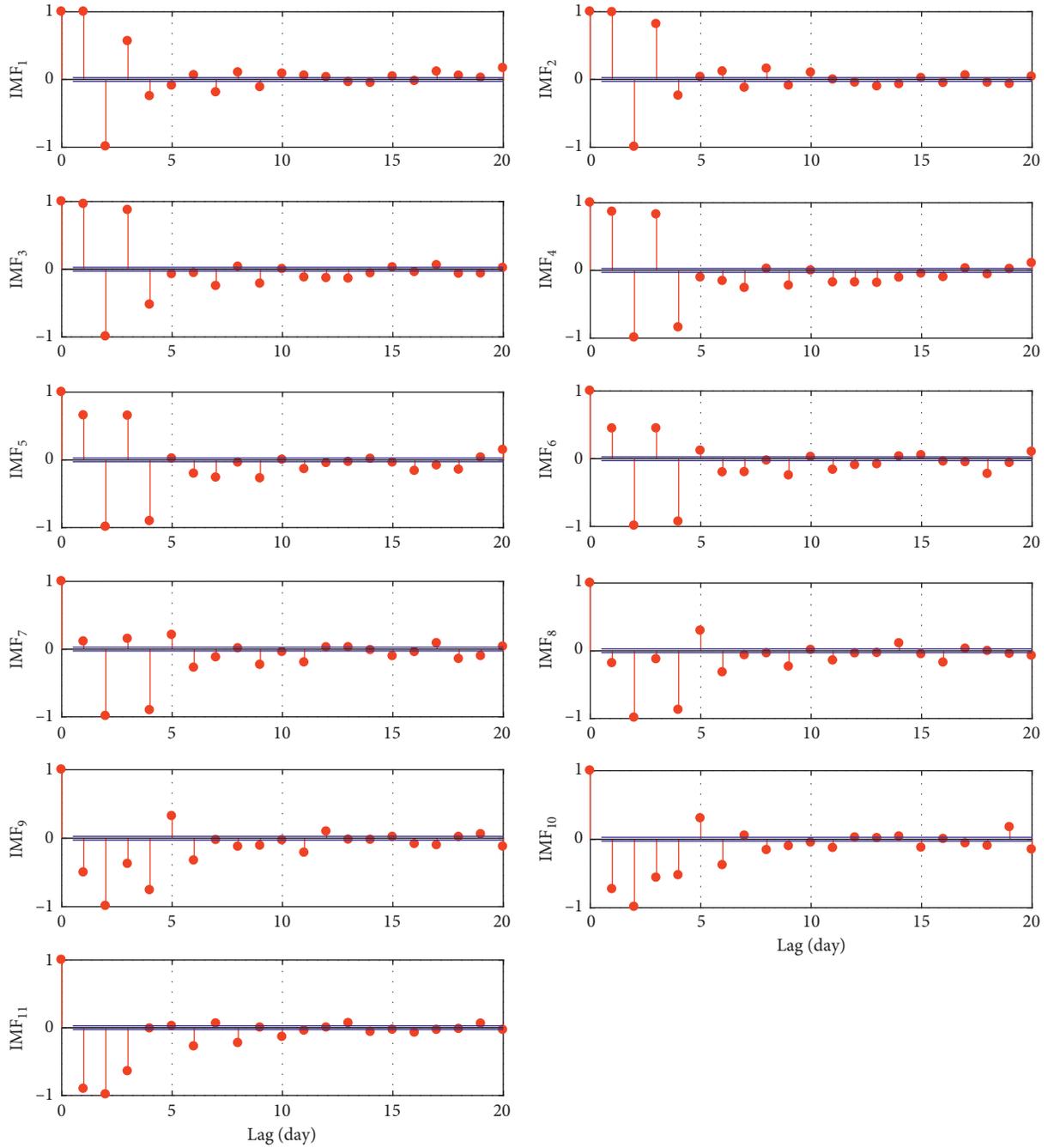


FIGURE 5: The PACFs of the raw data and IMF series that are obtained by the VMD at the Weijiabao station.

[65] revealed that the prediction performance of the neural network model with multiple hidden layers was not always better than the model with a single hidden layer, which therefore implies that the excess hidden layers of a neural network model might lead to an overfitting problem [66, 67]. The hidden layer of the BPNN can be one or multiple layers, but the BPNN with one hidden layer was able to complete the mapping of any continuous function with arbitrary accuracy. In order to avoid excessive slowdown of the training speed, one hidden layer was adopted for the BPNN model at the two selected hydrological stations.

Through the above method, the corresponding BPNN model structure of the other IMFs could be determined, and the prediction performance is also summarized in Table 7.

4.4. Forecasting Results of the Models and Model Comparison.

By summing all the components of the modelled IMFs of the daily streamflow during the testing period from 1 January 2012 to 31 December 2014, this section derived the resulting daily flow rates for the Wushan and Weijiabao hydrological stations. In order to verify the prediction accuracy and the stability of the different VMD-based models, six types of

TABLE 2: The input features and output targets for the VMD decomposition sequences of the raw data and its decomposition sequences at the Weijiabao station.

Feature term	Model inputs	Numbers of input	Outputs
Original	$X_0(t-i), i = 0, 1, 2$	3	$X_0(t+1)$
IMF ₁	$X_1(t-i), i = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$	11	$X_1(t+1)$
IMF ₂	$X_2(t-i), i = 0, 1, 2, 3$	4	$X_2(t+1)$
IMF ₃	$X_3(t-i), i = 0, 1, 2, 3, 4, 5, 6$	7	$X_3(t+1)$
IMF ₄	$X_4(t-i), i = 0, 1, 2, 3, 4, 5, 6$	7	$X_4(t+1)$
IMF ₅	$X_5(t-i), i = 0, 1, 2, 3$	4	$X_5(t+1)$
IMF ₆	$X_6(t-i), i = 0, 1, 2, 3, 4, 5, 6$	7	$X_6(t+1)$
IMF ₇	$X_7(t-i), i = 0, 1, 2, 3, 4, 5, 6$	7	$X_7(t+1)$
IMF ₈	$X_8(t-i), i = 0, 1, 2, 3, 4, 5, 6$	7	$X_8(t+1)$
IMF ₉	$X_9(t-i), i = 0, 1, 2, 3, 4, 5$	6	$X_9(t+1)$
IMF ₁₀	$X_{10}(t-i), i = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$	11	$X_{10}(t+1)$
IMF ₁₁	$X_{11}(t-i), i = 0, 1, 2$	3	$X_{11}(t+1)$

TABLE 3: Performance of the GBRT model during the training period at the Weijiabao station.

Sequence	GBRT parameters					Training	
	LR	MD	MF	MSS	MSL	RMSE (m ³ /s)	NSE
IMF ₁	0.17	9	10	170	63	3.311	0.99831
IMF ₂	0.37	15	4	200	1	0.452	0.99992
IMF ₃	0.27	15	7	200	1	0.624	0.99972
IMF ₄	0.15	6	3	130	37	5.748	0.96849
IMF ₅	0.25	15	1	34	5	1.070	0.99788
IMF ₆	1	15	7	200	1	0.137	0.99996
IMF ₇	0.20	15	7	2	1	0.042	0.99999
IMF ₈	1	1	7	200	1	2.049	0.97926
IMF ₉	0.13	15	6	2	1	0.040	0.99999
IMF ₁₀	0.15	6	4	130	37	2.912	0.94075
IMF ₁₁	1	15	3	200	1	0.101	0.99990

TABLE 4: Performance of the ARMA model during the training period at the Weijiabao station.

Sequence	ARMA parameters		Training	
	p	q	RMSE (m ³ /s)	NSE
IMF ₁	19	18	0.145	0.99997
IMF ₂	13	19	0.196	0.99987
IMF ₃	20	18	0.251	0.99929
IMF ₄	6	20	0.334	0.999874
IMF ₅	4	19	0.442	0.999604
IMF ₆	20	19	0.508	0.999132
IMF ₇	7	20	0.563	0.999325
IMF ₈	18	17	0.546	0.998371
IMF ₉	19	19	0.496	0.996429
IMF ₁₀	20	20	0.415	0.995971
IMF ₁₁	20	20	0.314	0.997730

daily-scale streamflow prediction models were established, including the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN. The comparison between the predicted results and the observed streamflow series is shown in Figures 6 and 7.

Figures 6 and 7 show that the test set forecasting results of various models at the Wushan and Weijiabao stations are in line with expectations, respectively. With regards to the

TABLE 5: Performance of the SVR model during the training period at the Weijiabao station.

Sequence	SVR parameters			Training	
	C	ϵ	γ	RMSE (m ³ /s)	NSE
IMF ₁	25	0.0000001	1	0.227	0.999992
IMF ₂	25	0.0000001	1	0.335	0.999954
IMF ₃	24.346	0.0000001	0.377	0.432	0.999867
IMF ₄	24.545	0.0000001	0.433	0.535	0.999728
IMF ₅	25	0.0000001	0.366	0.777	0.998884
IMF ₆	25	0.0000001	1	0.706	0.999041
IMF ₇	23.428	0.0002089	0.465	0.776	0.996972
IMF ₈	21.469	0.0005338	0.031	0.910	0.995910
IMF ₉	14.231	0.0000001	1	0.690	0.997577
IMF ₁₀	19.407	0.0000001	0.652	0.497	0.998274
IMF ₁₁	25	0.0000001	1	0.487	0.997650

TABLE 6: Evaluation of a BPNN with 1 hidden layer for the IMF1 of VMD at the Weijiabao station.

Predicted structure	NSE	RMSE (m ³ /s)
11-4-1	0.9999933	0.2333
11-5-1	0.9999938	0.2246
11-6-1	0.9999930	0.2389
11-7-1	0.9999942	0.2182
11-8-1	0.9999930	0.2382
11-9-1	0.9999930	0.2391
11-10-1	0.9999938	0.2242
11-11-1	0.9999935	0.2308
11-12-1	0.9999933	0.2337
11-13-1	0.9999938	0.2257

prediction accuracy, taking Weijiabao Station as an example, the models could be ranked from the highest to the lowest as follows: VMD-BPNN > VMD-SVR > VMD-ARMA > VMD-GBRT > EEMD-BPNN > BPNN. In the single prediction model, the BPNN model exhibited a poor-fitting ability and a high forecasting error, particularly for the days with relatively high flow rates. Compared with the single BPNN model forecasting results, the combined prediction model after being processed by EEMD and VMD modal decomposition was always better than single prediction model, which could, therefore, better fit the high flow and yield

TABLE 7: BPNN structure of the 1st–11th IMF after VMD decomposition at the Weijiabao station.

Decomposed IMFs	Predicted structure	NSE	RMSE (m ³ /s)
IMF ₁	11-7-1	0.9999942	0.2182
IMF ₂	4-12-1	0.9999694	0.3044
IMF ₃	7-5-1	0.9998331	0.3834
IMF ₄	7-6-1	0.9996981	0.5174
IMF ₅	4-5-1	0.9988959	0.7380
IMF ₆	7-12-1	0.9981018	0.7511
IMF ₇	7-5-1	0.9984073	0.8652
IMF ₈	7-3-1	0.9963585	0.8167
IMF ₉	6-10-1	0.9923446	0.7268
IMF ₁₀	11-10-1	0.9912652	0.6110
IMF ₁₁	3-11-1	0.9950110	0.4649

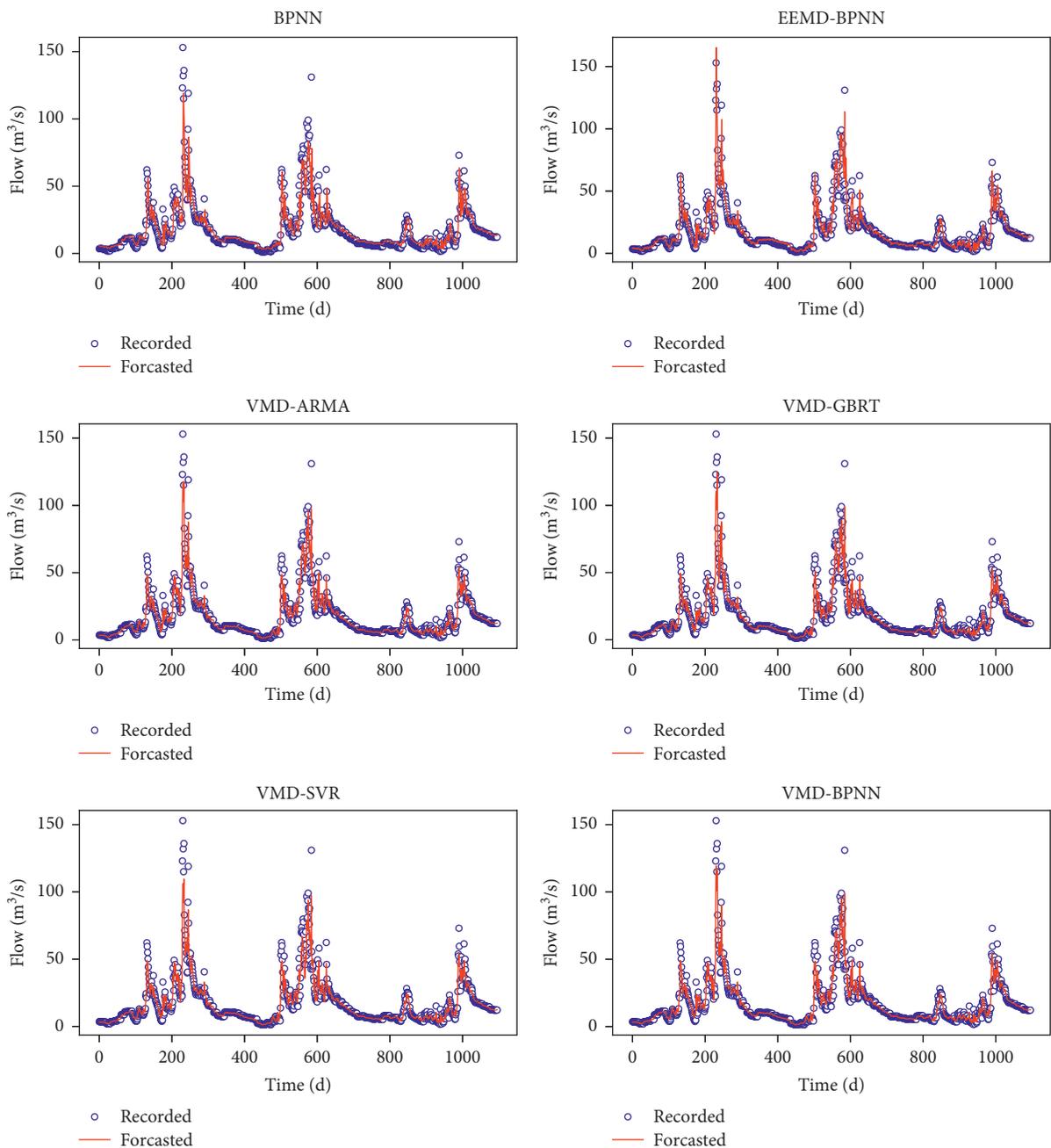


FIGURE 6: Comparison of the records and predicted streamflow results of the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models during the testing phase (01/01/2012–31/12/2014) at the Wushan station.

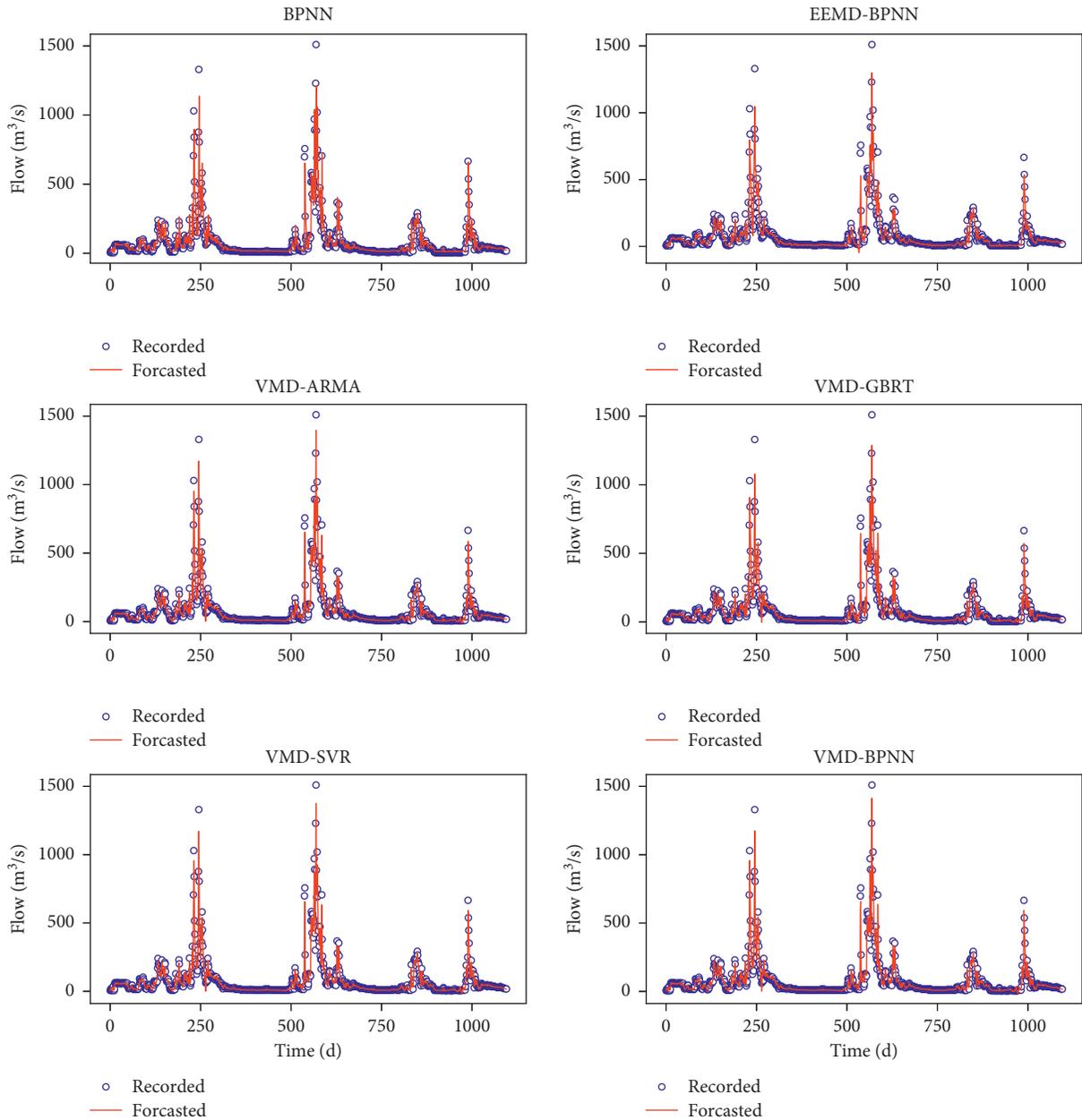


FIGURE 7: Comparison of the records and predicted streamflow results of the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models during the testing phase (01/01/2012–31/12/2014) at the Weijiabao station.

better prediction accuracy. Besides, the match between the recorded and the modelled daily runoff with the VMD-BPNN model was better than that with the EEMD-BPNN model, which is mainly because different decomposition algorithms had different control methods on the modal number, thereby affecting the prediction error. The centre frequency of the VMD method was controllable, which could effectively avoid modal aliasing in comparison with the EEMD method [68]. To further evidence the superiority of the VMD-based model's feature learning capability, four VMD-based models were also compared in terms of reproducing the statistical characteristics of observations, which included the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN (Figure 6). All the VMD-based models

closely represented the observed flows, and there was no major difference in the performances of the three models. For the two stations, the results show that the prediction results of the VMD-BPNN model were always better than all the other models and required the lowest computational effort [69].

4.5. Evaluating All Models' Forecasting Accuracy. The scatter plots of the observed and forecasted flows, as well as the fitting line of the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models in Figures 8 and 9 are closed to the ideal fitting line (the black fitting line in Figures 8 and 9), which indicates that the overall prediction effect was

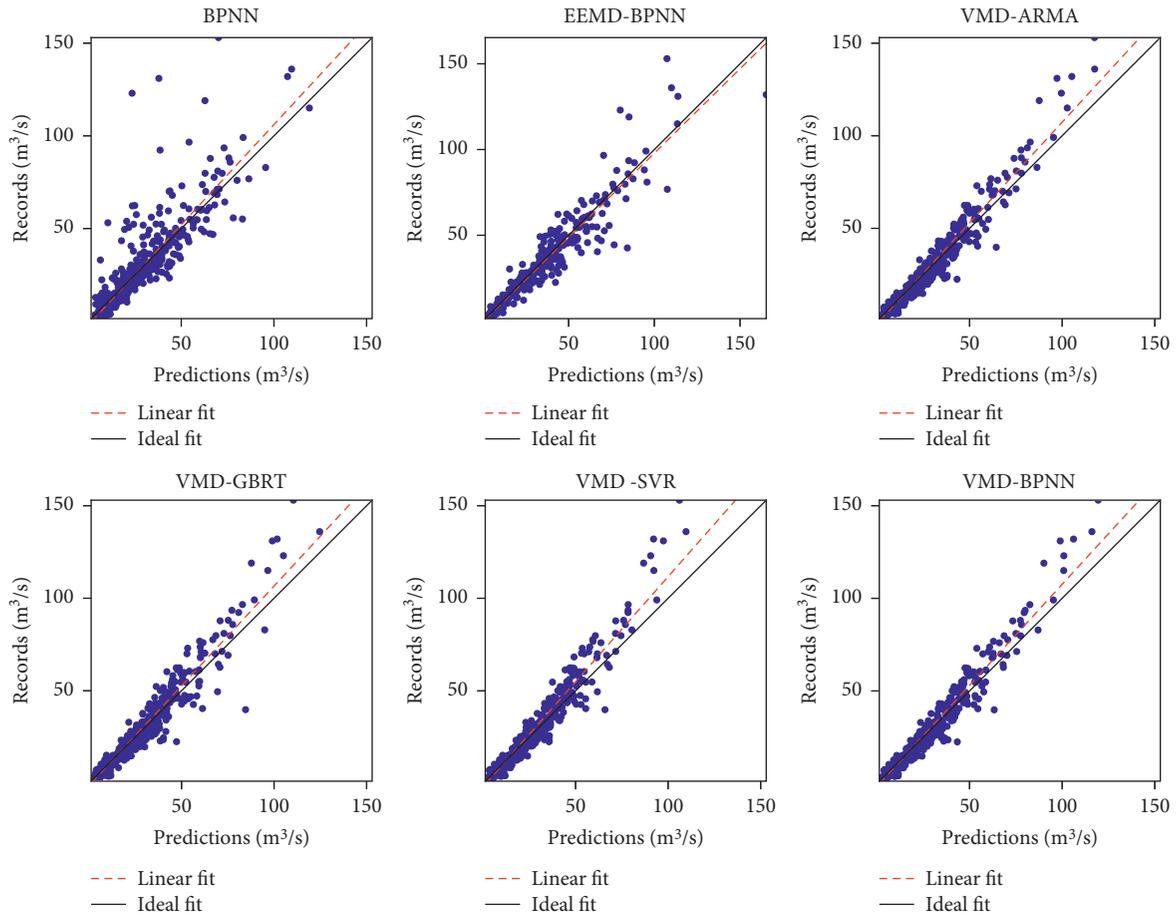


FIGURE 8: Comparison of scatter plots between the observed and simulated streamflows from the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models during the testing phase (01/01/2012–31/12/2014) at the Wushan station.

very well, and the high flow prediction error was the smallest, followed by BPNN, then EEMD-BPNN. The fitting effect of the VMD-BPNN model is better than the BPNN model, which exhibits the worst fitting effect. The scatter plots of the forecasted streamflow values were derived by the VMD-BPNN, and the observed flows follow the 45° lines, whereas the least square regression (LSR) lines of the BPNN deviate substantially from the 45° lines. These results confirm the underestimation of the flows, especially the high flows, and demonstrate that the BPNN model had relatively poor robustness and a weak forecasting ability if not integrated with any data processing techniques. This can be attributed to a separate forecasting model to learn feature for the original runoff series, which insufficiently represents the instinctual information of nonstationary and nonperiodicity data [70]. The points for the EEMD-BPNN model present obvious unorganized point-sets. Overall, the EEMD-BPNN model tends to overestimate the medium flows and underestimate the high flows. The superior performance of the VMD-BPNN methods becomes obvious when employing the VMD decomposition algorithm. The EEMD-BPNN results are close to the observed values in some days, but the EEMD-BPNN model always yielded less accurate forecasts than the VMD-BPNN model for different higher flows; the reason lies in the modes obtained

by VMD which are much smoother than components decomposed by EEMD [44].

As shown in the scatter plots in Figures 8 and 9, no significant difference terms of performance could be observed between the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models in the test phase. The results demonstrate that the forecasted flows of the VMD-based models had a good performance at forecasting the maximum and minimum streamflows. Compared with the VMD-ARMA, VMD-GBRT, and VMD-SVR models, the VMD-BPNN model has powerful approximation ability for solving complex nonlinear problems, which therefore caused the prediction accuracy to improve. Moreover, by using the VMD-based decomposed method to sufficiently represent the instinctual information of the streamflow series, the generalization ability of the BPNN model can be effectively enhanced.

The statistical indicators of performance can more accurately assess the predictive power of each model. Table 8 shows the model performance indices of the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models at the Wushan and Weijiabao stations during their respective testing period.

Taking Weijiabao Station as an example, the various performance statistics provided in Table 8 show that the

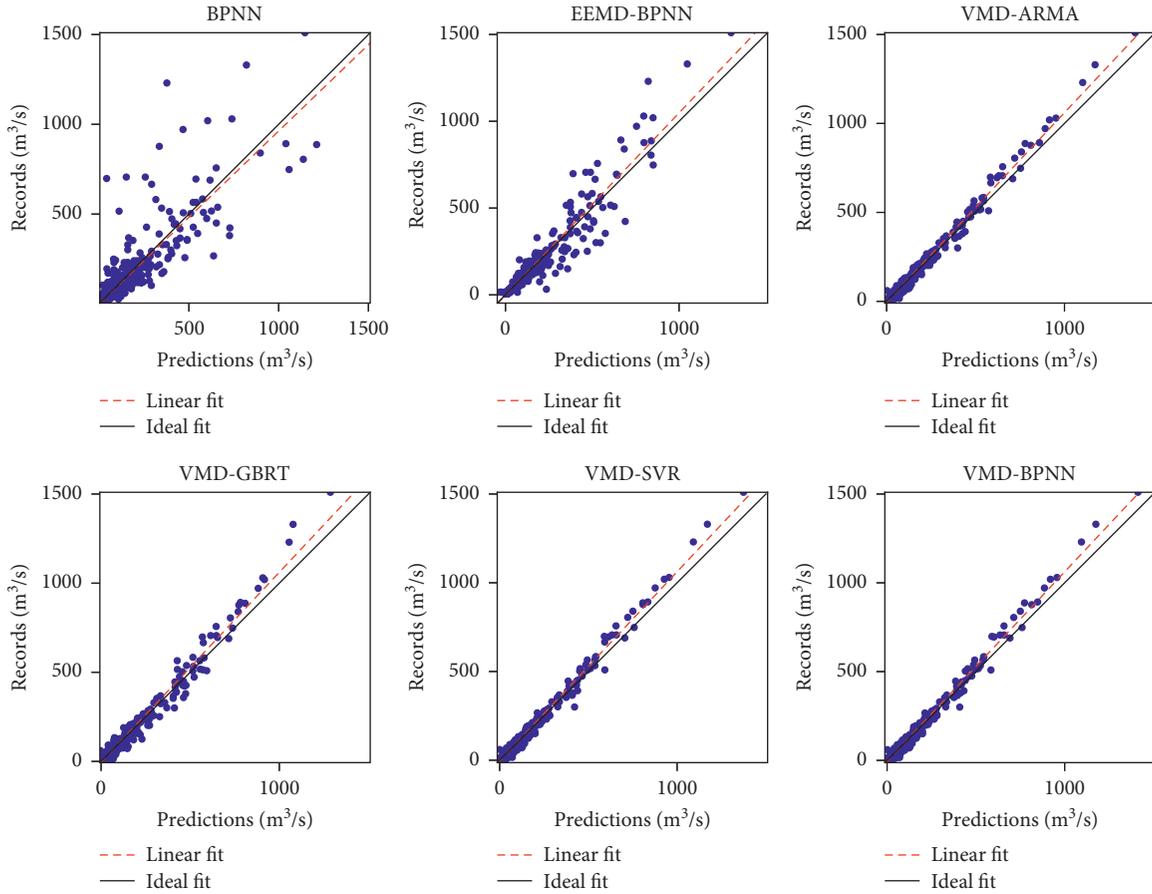


FIGURE 9: Comparison of scatter plots between the observed and simulated streamflows from the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models during the testing phase (01/01/2012–31/12/2014) at the Weijiabao station.

TABLE 8: The performance of the different prediction models during the testing period for the Wushan and Weijiabao hydrological stations in the Wei River Basin.

Stations	Models	NSE	RMSE (m ³ /s)	MAE (m ³ /s)	Time cost (s)
Wushan	BPNN	0.8132	7.87	3.57	5.33
	EEMD-BPNN	0.9346	4.65	2.00	176.23
	VMD-ARMA	0.9547	3.88	1.86	1941.42
	VMD-GBRT	0.9399	4.46	2.09	158.23
	VMD-SVR	0.9395	4.47	2.00	4330.57
	VMD-BPNN	0.9566	3.79	1.84	54.45
Weijiabao	BPNN	0.7698	68.86	24.79	4.37
	EEMD-BPNN	0.9236	39.66	13.96	198.88
	VMD-ARMA	0.9862	16.84	7.99	3211.78
	VMD-GBRT	0.9750	22.70	10.21	281.34
	VMD-SVR	0.9857	17.18	8.13	4673.64
	VMD-BPNN	0.9865	16.80	7.95	123.13

forecasted flows using the VMD-BPNN reproduced the statistical characteristics of the observations fairly well in overall, compared to the BPNN and EEMD-BPNN. The results concur with relatively high NSE values (0.9865 for the VMD-BPNN vs. 0.8132 for the BPNN, 0.9236 for the EEMD-BPNN, 0.9862 for the VMD-ARMA, 0.9750 for the VMD-GBRT, and 0.9857 for the VMD-SVR model); the NSE values of the VMD-BPNN model were increased by 28.16%, 6.81%, 0.03%, 1.18%, and 0.09%, respectively, compared to

other models. The RMSE values of each model are as follows: 16.80 m³/s (VMD-BPNN) vs. 68.86 m³/s (BPNN), 39.66 m³/s (EEMD-BPNN), 16.84 m³/s (VMD-ARMA), 22.70 m³/s (VMD-GBRT), and 17.18 m³/s (VMD-SVR); the RMSE values of the VMD-BPNN model was reduced by 75.60%, 57.63%, 0.24%, 25.96%, and 2.20%, respectively, compared to other models. The MAE values are, respectively, as follows: 7.95 m³/s (VMD-BPNN) vs. 24.79 m³/s (BPNN), 13.96 m³/s (EEMD-BPNN), 7.99 m³/s (VMD-ARMA),

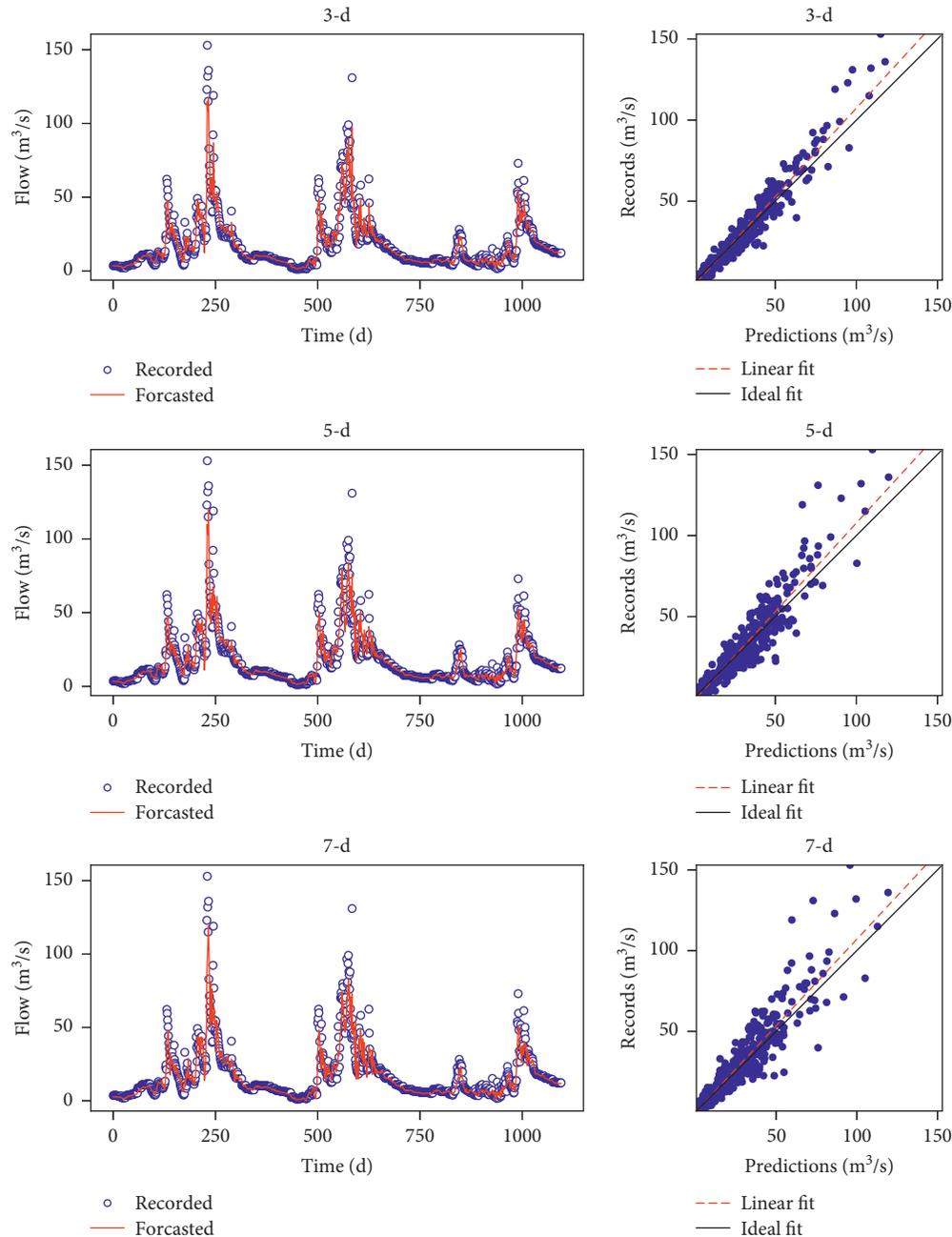


FIGURE 10: Final predicted results and the scatters of the VMD-BPNN for the 3-d, 5-d, and 7-d ahead forecasts at the Wushan station.

10.21 m^3/s (VMD-GBRT), and 8.13 m^3/s (VMD-SVR); the MAE values of the VMD-BPNN model was reduced by 67.95%, 43.08%, 0.60%, 22.19%, and 2.30%, respectively, compared to the other models. Through the calculation and comparison relative to error statistics (such as NSE, RMSE, and MAE) of testing set of each model, the results show that the VMD-BPNN model provided accurate and reliable streamflow forecasts for almost all models (Table 8), which demonstrates that the forecasting performance metric values of the BPNN were unsatisfactory, and the forecasted streamflow that was obtained by the EEMD-BPNN were rather accurate. These results imply that the non-decomposition method or EEMD-based decomposition

method may not be suitable for daily streamflow forecasts. The consistent forecast errors (Table 8) that were produced by the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models imply that the VMD-based model could be used for daily streamflow forecasts. The means of the predicted flows from the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models appear to be equivalent and closely match the measured data in the test phase, indicating that the amplitude and the shape of the subseries that are decomposed using VMD are well forecasted by the VMD-based model. There was no obvious difference in the performance amongst these four models in the test phase, but the performance errors predicted by the VMD-BPNN model

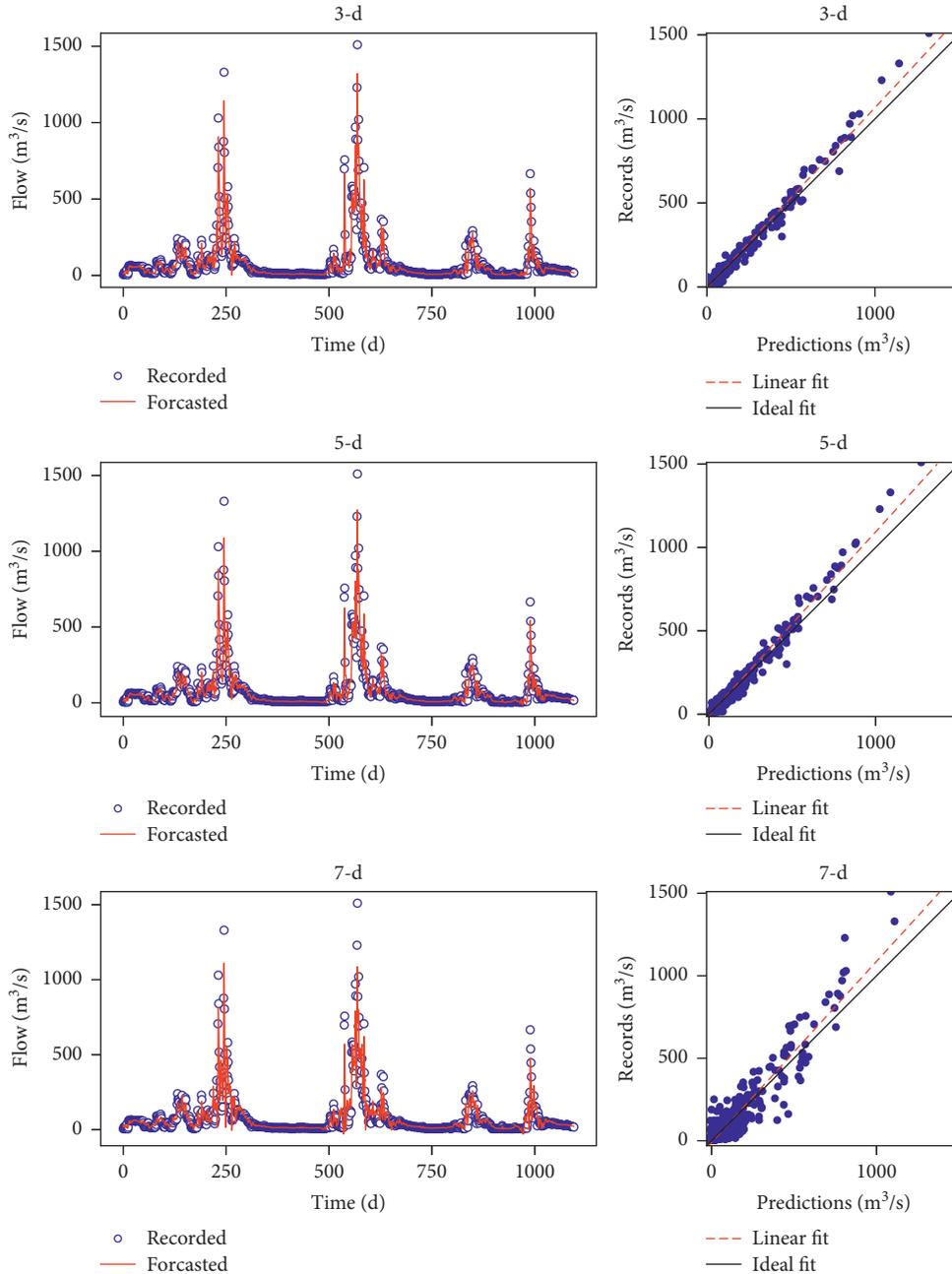


FIGURE 11: Final predicted results and the scatters of the VMD-BPNN for the 3-d, 5-d, and 7-d ahead forecasts at the Weijiabao station.

were slightly better than the other three models, and the computational run-time results of these four models also show that the computational efficiency was the highest by using the VMD-BPNN model (Table 8).

All the observations demonstrate the following: (1) hybrid models based on decomposition preprocessing methods such as VMD and EEMD surpass traditional models that perform no decomposition, (2) the VMD-based models are able to achieve almost perfect forecasting performance, and (3) EEMD cannot effectively extract useful information from the daily streamflow time series similar to its counterparts, such as VMD [71]. The performance of the models can be explained based on the mode-mixing

phenomena that are observed. For the BPNN, as discussed previously in Figure 4, the centred daily streamflow time series were heavily mode-mixed with various frequencies along the spectrum. The different frequencies were associated with different physical meanings. As a result, the forecasting performance of the BPNN was unsatisfactory, as observed in Table 8.

4.6. Forecasting Accuracy over Long Time Periods (>1 Day Ahead Time). The above studies show that all the BPNN, EEMD-BPNN, VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models yielded an NSE coefficient greater than 0.8 when the foreseen period was 1 day, both the RMSE

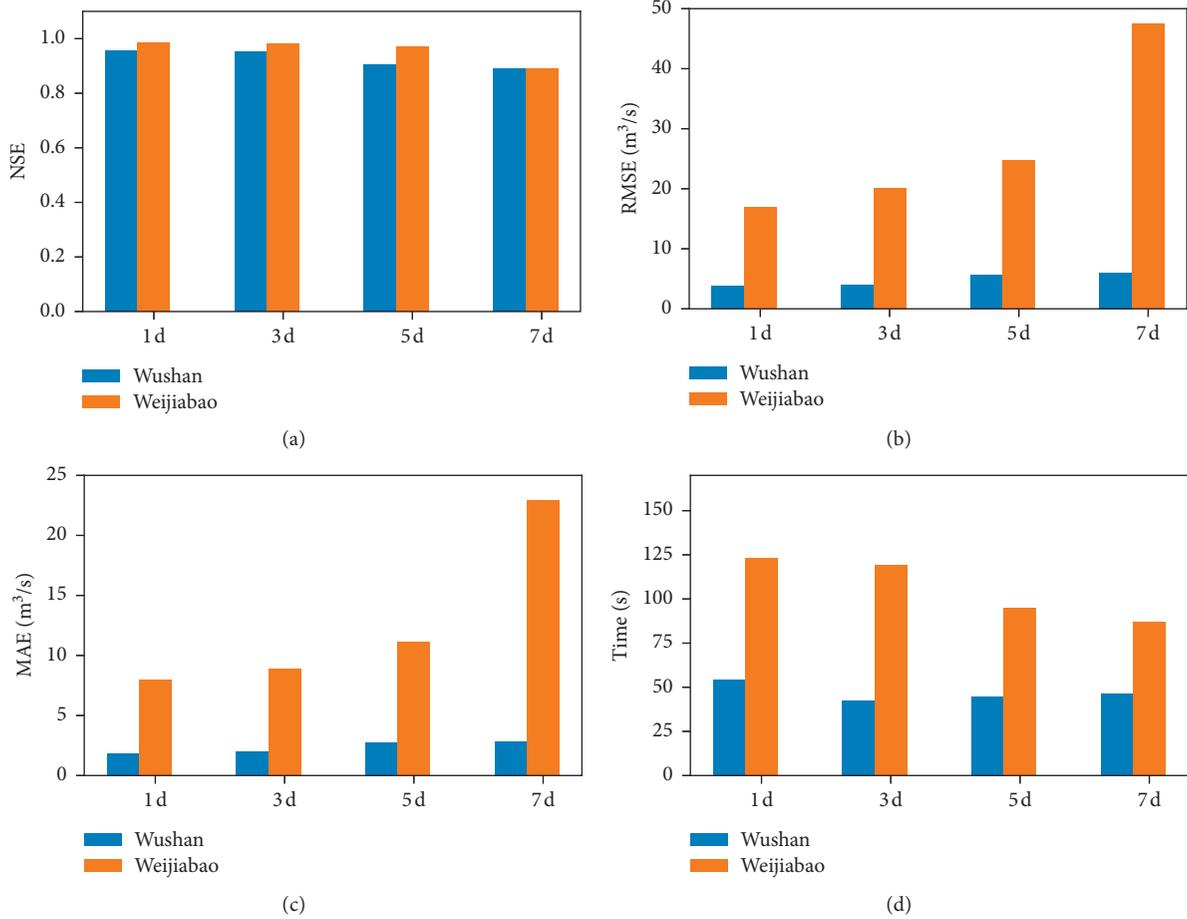


FIGURE 12: Streamflow forecast scores (computed over the verification period) for the VMD-BPNN model between the observations and predicted streamflows for the 1-d, 3-d, 5-d, and 7-d ahead forecasts, which were evaluated using the (a) Nash–Sutcliffe efficiency coefficient (NSE), (b) root mean squared error (RMSE), (c) mean absolute error (MAE), and (d) time cost.

and MAE were small, and the effect was satisfactory. Except that the hydrological forecasts focus on the prediction accuracy, the length of the foresight period is also of great significance for guiding the operation and dispatching of reservoirs. Now, taking the VMD-BPNN model as an example, the effects of the foresight period of 1, 3, 5, and 7 days on the prediction results were also examined. The forecasting results using the VMD-BPNN model for 1, 3, 5, and 7 days ahead in the Wushan and Weijiabao stations are displayed in Figures 10 and 11. The histograms of the four performance metrics of the NSE, RMSE, MAE, and the time cost along with the different lead-times of 1–7 days are presented in Figure 12, respectively.

From Figures 10 and 11, it is clear that the predicted data exhibits a stronger tracking ability for the recorded data for 3 and 5 days ahead, but has slightly poor performance for forecasting streamflow 7 day ahead, similar to that seen in scatter plot. If the NSE value was less than 0.8, the model forecasts are considered to be poor and are therefore not applied to predict streamflow [72]. Obviously, the VMD-BPNN model performed best in forecasting streamflow when the foresight period was 1 day, which was also better than the prediction accuracy yielded when the foresight

period was, respectively, 3 d, 5 d, and 7 d. It is worth noting that the worse performance occurred for longer foresight period (Figure 12), which was caused by continuous accumulation of errors. As the foresight period increases, the RMSE and MAE increase gradually, while the NSE coefficient decreases, but remains above 0.8 and can, therefore, satisfy the formulation of the daily operation plan of the reservoir. Figure 12 shows the acceptable reliability of the forecasting of the VMD-BPNN model under the foresight period of 1 d, 3 d, 5 d, and 7 d in the Wushan and Weijiabao stations.

5. Conclusions

This paper presented a comparative study of the VMD-based models for the forecasting daily streamflow of the Wushan and Weijiabao hydrological stations in China. To improve the forecasting accuracy, three aspects were considered in a decomposed and integrated fashion: (1) multiscale feature extraction, which consisted in extracting subseries (IMFs) by the VMD; (2) determination of the input variables by PACF, whereby the PACF was used to analyze the characteristics of each subseries for extracting the input variables; (3)

predicting each subseries based on the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models, and the final forecasting results were derived by summing; (4) using the NSE, RMSE, and MAE performance evaluation indices, the forecasting results of the VMD-ARMA, VMD-GBRT, VMD-SVR, and VMD-BPNN models were compared with that of two prediction models including BPNN and EEMD-BPNN; and (5) studying the effect of different foresight period on the streamflow forecast results for the VMD-BPNN model.

In general, the VMD-based models proposed in this paper could significantly improve the prediction accuracy of the daily time scale nonstationary streamflow series. Different from traditional daily streamflow forecasts, the key features of the proposed method lie in the following aspects: (1) the multiscale feature extraction using the VMD approach was much more robust to the sampling and noise, which could effectively distinguish the original nonstationary series; (2) the combination of the VMD decomposition method and the prediction models such as the ARMA, GBRT, SVR, and BPNN could improve the forecasting accuracy; (3) especially the VMD-BPNN method improved algorithm's stability with regards to the accuracy and limited computation time. The prediction results showed that among all the methods mentioned, the VMD-BPNN model was the best on prediction accuracy and computational efficiency. The model can effectively predict highly nonlinear and nonstationary hydrological streamflow series and is a powerful tool for daily streamflow forecasting.

Even though the hybrid model obtains better prediction accuracy than several traditional methods, they still have limitations that a black box model lacks explanation for its instinctive features. No matter how well the predicted results are, it is not incredibly reliable to use in practical engineering. Hence, in the future, the research of the interpretable machine learning is a point; in addition, the model can be retrained to take the complexity and representative of data into consideration to alleviate the potential influence of the limited data samples and has higher generalization ability.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported financially by the State Key Laboratory Base of Eco-Hydraulic Engineering in Arid Area, China (Grant no. 2017ZZKT-5), National Natural Science Foundation of China (Grant no. 51609197), CAS "Light of West China" Program (Grant no. XAB2016AW06), and Xi'an Science and Technology Program (Grant no. SF1335).

References

- [1] Z. M. Yaseen, I. Ebtehaj, H. Bonakdari et al., "Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model," *Journal of Hydrology*, vol. 554, pp. 263–276, 2017.
- [2] V. P. Singh and H. Cui, "Entropy theory for streamflow forecasting," *Environmental Processes*, vol. 2, no. 3, pp. 449–460, 2015.
- [3] N. F. Attar, Q. B. Pham, S. F. Nowbandegani et al., "Enhancing the prediction accuracy of data-driven models for monthly streamflow in Urmia lake basin based upon the autoregressive conditionally heteroskedastic time-series model," *Applied Sciences*, vol. 10, no. 2, pp. 571–590, 2020.
- [4] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," *Journal of Hydrology*, vol. 476, pp. 433–441, 2013.
- [5] H. Duan, G. R. Lei, and K. Shao, "Forecasting crude oil consumption in China using a grey prediction model with an optimal fractional-order accumulating operator," *Complexity*, vol. 2018, Article ID 3869619, 12 pages, 2018.
- [6] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [7] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061–1066, 2016.
- [8] G.-F. Lin, Y.-C. Chou, and M.-C. Wu, "Typhoon flood forecasting using integrated two-stage support vector machine approach," *Journal of Hydrology*, vol. 486, pp. 334–342, 2013.
- [9] Y. He, Y. Tian, J. Tang, and Y. Ma, "Unsupervised domain adaptation using exemplar-SVMs with adaptation regularization," *Complexity*, vol. 2018, Article ID 8425821, 13 pages, 2018.
- [10] M. Bray and D. Han, "Identification of support vector machines for runoff modelling," *Journal of Hydroinformatics*, vol. 6, no. 4, pp. 265–280, 2004.
- [11] Y. B. Dibike, S. Velickov, D. Solomatine, and M. B. Abbott, "Model induction with support vector machines: introduction and applications," *Journal of Computing in Civil Engineering*, vol. 15, no. 3, pp. 208–216, 2001.
- [12] Z. He, X. Wen, H. Liu, and J. Du, "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region," *Journal of Hydrology*, vol. 509, pp. 379–386, 2014.
- [13] P.-S. Yu, S.-T. Chen, and I.-F. Chang, "Support vector regression for real-time flood stage forecasting," *Journal of Hydrology*, vol. 328, no. 3-4, pp. 704–716, 2006.
- [14] R. Maity, P. P. Bhagwat, A. Bhatnagar et al., "Potential of support vector regression for prediction of monthly streamflow using endogenous property," *Hydrological Processes*, vol. 24, no. 7, pp. 917–923, 2010.
- [15] O. Abedinia and N. Amjadi, "Net demand prediction for power systems by a new neural network-based forecasting engine," *Complexity*, vol. 21, no. S2, pp. 296–308, 2016.
- [16] B. Gerardo, C. Fernando, R. E. Haber, Q. Ramón, and V. Alberto, "Coping with complexity when predicting surface roughness in milling processes: hybrid incremental model with optimal parametrization," *Complexity*, vol. 2017, Article ID 7317254, 11 pages, 2017.

- [17] S. Pan and K. Duraisamy, "Long-time predictive modeling of nonlinear dynamical systems using neural networks," *Complexity*, vol. 2018, Article ID 4801012, 26 pages, 2018.
- [18] E. Mutlu, I. Chaubey, H. Hexmoor, and S. G. Bajwa, "Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed," *Hydrological Processes*, vol. 22, no. 26, pp. 5097–5106, 2008.
- [19] C. Chiamsathit, A. J. Adeloye, and S. Bankaru-Swamy, "Inflow forecasting using artificial neural networks for reservoir operation," *Proceedings of the International Association of Hydrological Sciences*, vol. 373, pp. 209–214, 2016.
- [20] M. Rezaie-Balf, S. R. Naganna, O. Kisi, and A. El-Shafie, "Enhancing streamflow forecasting using the augmenting ensemble procedure coupled machine learning models: case study of Aswan high dam," *Hydrological Sciences Journal*, vol. 64, no. 13, pp. 1629–1646, 2019.
- [21] S.-G. Hong, S.-K. Oh, M.-S. Kim, and J.-J. Lee, "Nonlinear time series modelling and prediction using Gaussian RBF network with evolutionary structure optimisation," *Electronics Letters*, vol. 37, no. 10, pp. 639–640, 2001.
- [22] A. C. Marcelo, A. P. Braga, and B. R. de Menezes, "Improving neural networks generalization with new constructive and pruning methods," *Journal of Intelligent and Fuzzy Systems*, vol. 13, pp. 75–83, 2003.
- [23] X. G. Wang, Z. Tang, H. Tamura, and M. Ishii, "A modified error function for the backpropagation algorithm," *Neurocomputing*, vol. 57, pp. 477–484, 2004.
- [24] H. K. Cigizoglu and Ö. Kişi, "Flow prediction by three back propagation techniques using k -fold partitioning of neural network training data," *Hydrology Research*, vol. 36, no. 1, pp. 49–64, 2005.
- [25] C.-C. Yang and C.-S. Chen, "Application of integrated back-propagation network and self organizing map for flood forecasting," *Hydrological Processes*, vol. 23, no. 9, pp. 1313–1323, 2009.
- [26] X. Luo, Y. Xu, and J. Xu, "Regularized back-propagation neural network for rainfall-runoff modeling," in *Proceedings of the International Conference on Network Computing & Information Security*, May 2011.
- [27] S. Huang, J. Chang, Q. Huang, and Y. Chen, "Monthly streamflow prediction using modified EMD-based support vector machine," *Journal of Hydrology*, vol. 511, pp. 764–775, 2014.
- [28] Q. J. Liu, H. Y. Zhang, K. T. Gao, B. Xu, J. Z. Wu, and N. F. Fang, "Time-frequency analysis and simulation of the watershed suspended sediment concentration based on the Hilbert-Huang transform (HHT) and artificial neural network (ANN) methods: a case study in the Loess plateau of China," *Catena*, vol. 179, pp. 107–118, 2019.
- [29] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [30] O. Kisi and M. Cimen, "A wavelet-support vector machine conjunction model for monthly streamflow forecasting," *Journal of Hydrology*, vol. 399, no. 1-2, pp. 132–140, 2011.
- [31] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
- [33] U. Okkan and Z. Ali Serbes, "The combined use of wavelet transform and black box models in reservoir inflow modeling," *Journal of Hydrology and Hydromechanics*, vol. 61, no. 2, pp. 112–119, 2013.
- [34] S. J. Hadi and M. Tombul, "Streamflow forecasting using four wavelet transformation combinations approaches with data-driven models: a comparative study," *Water Resources Management*, vol. 32, no. 14, pp. 4661–4679, 2018.
- [35] O. Kisi, L. Latifoğlu, and F. Latifoğlu, "Investigation of empirical mode decomposition in forecasting of hydrological time series," *Water Resources Management*, vol. 28, no. 12, pp. 4045–4057, 2014.
- [36] J. Barge and H. Sharif, "An ensemble empirical mode decomposition, self-organizing map, and linear genetic programming approach for forecasting river streamflow," *Water*, vol. 8, no. 6, p. 247, 2016.
- [37] M. Rezaie-Balf, S. Kim, H. Fallah, and S. Alaghmand, "Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: application on the perennial rivers in Iran and South Korea," *Journal of Hydrology*, vol. 572, pp. 470–485, 2019.
- [38] Y. Yu, H. Zhang, and V. P. Singh, "Forward prediction of runoff data in data-scarce basins with an improved ensemble empirical mode decomposition (EEMD) model," *Water*, vol. 10, no. 4, 2018.
- [39] C. Li and M. Liang, "Extraction of oil debris signature using integral enhanced empirical mode decomposition and correlated reconstruction," *Measurement Science and Technology*, vol. 22, no. 8, pp. 85701–85710, 2011.
- [40] Q.-F. Tan, X.-H. Lei, X. Wang et al., "An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach," *Journal of Hydrology*, vol. 567, pp. 767–780, 2018.
- [41] X. Yu, X. Zhang, and H. Qin, "A data-driven model based on Fourier transform and support vector regression for monthly reservoir inflow forecasting," *Journal of Hydro-Environment Research*, vol. 18, pp. 12–24, 2018.
- [42] C. Liu, L. Zhu, and C. Ni, "Chatter detection in milling process based on VMD and energy entropy," *Mechanical Systems and Signal Processing*, vol. 105, pp. 169–182, 2018.
- [43] S. Mohanty, K. K. Gupta, and K. S. Raju, "Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD," *Measurement*, vol. 117, pp. 200–220, 2018.
- [44] M. Niu, Y. Hu, S. Sun, and Y. Liu, "A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting," *Applied Mathematical Modelling*, vol. 57, pp. 163–178, 2018.
- [45] A. K. Lohani, N. K. Goel, and K. K. S. Bhatia, "Improving real time flood forecasting using fuzzy inference system," *Journal of Hydrology*, vol. 509, pp. 25–41, 2014.
- [46] K. S. Kasiviswanathan, J. He, K. P. Sudheer, and J.-H. Tay, "Potential application of wavelet neural network ensemble to forecast streamflow for flood management," *Journal of Hydrology*, vol. 536, pp. 161–173, 2016.
- [47] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 531–544, 2014.
- [48] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

- [49] G. J. Bowden, G. C. Dandy, and H. R. Maier, "Input determination for neural network models in water resources applications. Part 1-background and methodology," *Journal of Hydrology*, vol. 301, no. 1–4, pp. 75–92, 2005a.
- [50] G. J. Bowden, H. R. Maier, and G. C. Dandy, "Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river," *Journal of Hydrology*, vol. 301, no. 1–4, pp. 93–107, 2005b.
- [51] J.-Y. Lin, C.-T. Cheng, and K.-W. Chau, "Using support vector machines for long-term discharge prediction," *Hydrological Sciences Journal*, vol. 51, no. 4, pp. 599–612, 2006.
- [52] W. C. Wang, K. W. Chau, C. T. Cheng, and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," *Journal of Hydrology*, vol. 374, no. 3-4, pp. 294–306, 2009.
- [53] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [54] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, New York, NY, USA, 2011.
- [55] S. B. Taieb and R. J. Hyndman, "A gradient boosting approach to the kaggle load forecasting competition," *International Journal of Forecasting*, vol. 30, no. 2, pp. 382–394, 2014.
- [56] J. Döpke, U. Fritsche, and C. Pierdzioch, "Predicting recessions with boosted regression trees," *International Journal of Forecasting*, vol. 33, no. 4, pp. 745–759, 2017.
- [57] A. Swami and R. Jain, "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 10, pp. 2825–2830, 2012.
- [58] Y. Hao, J. Wu, Q. Sun et al., "Simulating effect of anthropogenic activities and climate variation on Liulin springs discharge depletion by using the ARIMAX model," *Hydrological Processes*, vol. 27, no. 18, pp. 2605–2613, 2013.
- [59] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, NY, USA, 1995.
- [60] L. Iliadis, F. Maris, and S. Tachos, "Soft computing techniques toward modeling the water supplies of Cyprus," *Neural Networks*, vol. 24, no. 8, pp. 836–841, 2011.
- [61] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Nature*, vol. 323, no. 2, pp. 318–362, 1985.
- [62] Y. Tang, J. Ji, Y. Zhu, S. Gao, Z. Tang, and Y. Todo, "Eigen solution of neural networks and its application in prediction and analysis of controller parameters of grinding robot in complex environments," *Complexity*, vol. 2019, Article ID 5296123, 21 pages, 2019.
- [63] H. Naderpour, D. Rezazadeh Eidgahee, P. Fakharian, A. H. Rafiean, and S. M. Kalantari, "A new proposed approach for moment capacity estimation of ferrocement members using group method of data handling," *Engineering Science and Technology, an International Journal*, vol. 23, no. 2, pp. 382–391, 2020.
- [64] N. Huan, H. Chen, G. Cai, L. Fang, and Y. Wang, "Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier," *Sensors*, vol. 16, no. 11, p. 1887, 2016.
- [65] X. He, J. Luo, G. Zuo, and J. Xie, "Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks," *Water Resources Management*, vol. 33, no. 4, pp. 1571–1590, 2019.
- [66] D. R. Eidgahee, A. H. Rafiean, and A. Haddad, "A novel formulation for the compressive strength of IBP-based geopolymer stabilized clayey soils using ANN and GMDH-NN approaches," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 44, no. 1, pp. 219–229, 2020.
- [67] T. Xie, G. Zhang, J. Hou, J. Xie, M. Lv, and F. Liu, "Hybrid forecasting model for non-stationary daily runoff series: a case study in the Han river basin, China," *Journal of Hydrology*, vol. 577, Article ID 123915, 2019.
- [68] W. J. Niu, Z. K. Feng, Y. B. Chen et al., "Annual streamflow time series prediction using extreme learning machine based on gravitational search algorithm and variational mode decomposition," *Journal of Hydrologic Engineering*, vol. 25, no. 5, 2020.
- [69] M. Rezaie-Balf, S. Fani Nowbandegani, S. Z. Samadi, H. Fallah, and S. Alaghmand, "An ensemble decomposition-based artificial intelligence approach for daily streamflow prediction," *Water*, vol. 11, no. 4, p. 709, 2019.
- [70] N. E. Johnson, O. Ianiuk, D. Cazap et al., "Patterns of waste generation: a gradient boosting model for short-term waste prediction in New York city," *Waste Management*, vol. 62, pp. 3–11, 2017.
- [71] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on ceemd and gwo for short-term PM_{2.5} concentration forecasting," *Atmospheric Environment*, vol. 134, pp. 168–180, 2016.
- [72] G. Zuo, J. Luo, N. Wang, Y. Lian, and X. He, "Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting," *Journal of Hydrology*, vol. 585, Article ID 124776, 2020.