A hybrid model integrating Elman neural network with variational mode decomposition and Box-Cox transformation for monthly runoff time series prediction

Yan Kang (kangyan@nwsuaf.edu.cn)
Northwest A&F University: Northwest Agriculture and Forestry University
Fangqin Zhang
Northwest A&F University: Northwest Agriculture and Forestry University
Xiao Cheng
Northwest A&F University: Northwest Agriculture and Forestry University
Peiru Chen
Northwest A&F University: Northwest Agriculture and Forestry University
Songbai Song
Northwest A&F University: Northwest Agriculture and Forestry University

Research Article

Keywords: Monthly runoff prediction, non-stationary and skewed runoff time series, Variational mode decomposition, Box-Cox transformation, Elman neural network, Wei River Basin

Posted Date: March 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-807243/v1

A hybrid model integrating Elman neural network with variational mode decomposition and Box-Cox transformation for monthly runoff time series prediction

Fangqin Zhang[1,2], Yan Kang[1,2,*], Xiao Cheng[1], Peiru Chen[1], Songbai Song[1,2]

[1] College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling 712100, China.

[2] Key Laboratory of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling 712100, China.

Abstract: Precise and reliable monthly runoff prediction plays a vital role in optimal management of water resources but non-stationarity and skewness of monthly runoff time series can pose major challenges for developing appropriate prediction models. To address these issues, this paper proposes a novel hybrid prediction model based on Elman neural network (Elman), variational mode decomposition (VMD) and Box-Cox transformation (BC), named VMD-BC-Elman model. Firstly, the observed runoff is decomposed into sub-time series using VMD for the better frequency resolution. Secondly, the input datasets were transformed into normal distribution using Box-Cox, and as a result, skewedness in the data was removed and the correlation between the input and output variables enhanced. Finally, Elman is used to simulate the respective sub-time series. The proposed model is evaluated using monthly runoff time series at Zhangjiashan, Zhuangtou and Huaxian hydrological stations in Wei River Basin in China. The model performances are compared with those of single models (SVM, Elman), decomposition-based (VMD-SVM, VMD-Elman et.al) and BC-based models (BC-SVM and BC-Elman) by employing four metrics. The results show that the hybrid models outperform single models, and VMD-BC-Elman model performs best in all considered hybrid models with NSE greater than 0.95, R greater than 0.98, NMSE less than 4.73%, and PBIAS less than 0.39% in both training and testing periods. The study indicates that VMD-BC-Elman model is a satisfactory data-driven approach to predict the non-stationary and skewed monthly runoff time series, representing

an effective tool for predicting monthly runoff series.

Key words: Monthly runoff prediction; non-stationary and skewed runoff time series; Variational mode decomposition; Box-Cox transformation; Elman neural network; Wei River Basin

1. Introduction

Accurate and reliable runoff prediction is of great significance for the water resources planning and management, flood controlling and drought warning, which has attracted the extensive attention as a vital and difficult topic in hydrology modeling (Mohamad et al. 2017; Peng et al. 2011). However, runoff time series show complex non-linearity, non-stationarity and skewness (Barge and Sharif 2016), which are attributable to climate changes and frequent human activities, such as the variations of precipitation, conservancy project, urbanization, and so on. As a result, it is a great challenge to accurately simulate the intrinsic dynamic process of the inconsistent runoff series (Feng et al. 2020). In recent years, a multitude of runoff prediction models have been developed to model runoff time series, and representative models generally include two categories: process-driven models and data-driven models. Process-driven models, which are based on strict physical mechanisms, can successfully characterize the rainfall-runoff processes with complex mathematical formula, accurate meteorology and hydrological knowledge (Partington et al. 2012). However, there have been key modeling controversies about the issues such as adequacy of process parameterizations, data limitations and uncertainty, and computational constraints on model analysis (Demirel et al. 2009; Clark et al. 2017), which may lead up to poor performance of process-driven models. In contrast, data-driven approaches, which are based on the hydrological statistics and machine learning, can achieve

satisfactory performance without too much information input. The data-driven model can be categorized by three main methods: statistical methods, artificial intelligent methods (AI) and hybrid models (Myronidis et al. 2018; Kuremoto et al. 2014; Tiwari et al., 2013). Statistical models, such as autoregressive model (AR) (Sarlak 2008), autoregressive moving average model (ARMA) (Moeeni et al. 2017) and so on, have been widely applied to capture the stationary and linear relationship between variables and they, however, are not suitable for modeling non-stationary and non-linear runoff time series. AI models, i.e. neural network models (Sedki et al. 2008), support vector machines (Aggarwal et al. 2012), and random forest (Bojang et al. 2020), have been extensively applied in the hydrology prediction. Compared to the conventional regression models, AI models have greatly improved prediction performance for highly nonlinear runoff time series. As a typical recurrent neural network model, Elman neural network (Elman), equipped with a time delay operator, has short-term memory ability and is very suitable for the time series forecasting, and parameter selection is simple (Krishnan et al. 2019). Support vector machine (SVM), based on Vapnik-Chervonenkis dimension and structural risk minimization, can map data from input space to high-dimensional space in terms of a kernel function selected to suit the problem, and attain the optimal solution for regression issues (Karamouz et al. 2009). However, standalone AI models have a limited ability to identify the intrinsic non-stationary features of the input data, resulting in problems that the Elman model is inclined to fall into local optimum, over-fitting for, and the SVM model strongly dependents on the parameters selection (Wang et al. 2013). Therefore, a hybrid model which combines AI methods with data processing techniques has become the most popular option for predicting non-stationary runoff time series (Wen et al. 2019).

In recent years, signal decomposition algorithms, used as data processing techniques, are widely developed to effectively extract the multiple frequency information hidden from the complex runoff series, and generalizability of the data-driven model can be improved (Zhang et al. 2015). Empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD) are commonly employed as time-frequency decomposition algorithms. Among them, EMD has a strong adaptability due to a self-adaptive data-driven tool, but it lacks bases for rigorous mathematical theories and is prone to edge effect and ensuing modal aliasing (Huang et al. 2003; Sankaran and Reddy 2016). As an improved model of EMD, EEMD algorithm is proposed by adding white noise into a series of intrinsic mode functions (IMFs) which eliminates aliasing of modes effectively (Wu and Huang 2009). However, the amplitude of the added white noise in EEMD model is determined by expert experience, and the modal components cannot be controlled adaptively, which maybe lead to loss of information or incomplete decomposition (Yu et al. 2018). To address these issues, Variational Mode Decomposition (VMD), which has a sound mathematical basis, strong anti-aliasing ability and good decomposition potential, is introduced to recursively decompose original non-stationary series into a group of relatively stable sub-series (Drcgomiretskiy 2014). As a new signal decomposition method, VMD has been applied to speech recognition, fault analysis and hydrology forecasting in recent years, and the models have shown a satisfactory performance for signal decomposition (Deb et al. 2020; Mohanty et al. 2018; Li et al. 2020), and the hybrid model of VMD and support vector machine (SVM) has achieved good forecasting results (Feng et al. 2020).

The above hybrid models can process the non-stationary characteristics of runoff time series using decomposition, but the influence of runoff skewness on the forecasting performance has been rarely

86 explored by researchers. Normalization transformation techniques of data processing are worthy of
87 more attention. Box-Cox transformation can remove skewness of runoff data and transform the data
88 into more normal distribution which significantly stabilize variance of the runoff time series, and
89 improve the normality and linearity of data series, reduce the probability of pseudo regression, and
90 effectively improve the correlation between the input and output variables. The Box-Cox has achieved
91 good normality in hydrology frequency analysis and calculation (Vasiliades et al. 2009; Seong 2014),
92 but it has been rarely employed in runoff prediction models.

93 In summary, the AI models, especially the Elman neural network, has rarely integrated with VMD
94 and Box-Cox methods for runoff prediction. Therefore, to further improve the prediction accuracy, a
95 novel hybrid VMD-BC-Elman model is proposed for monthly runoff prediction in this paper The
96 objectives of this study are as followed: (1) VMD is used to decompose the original non-stationary
97 monthly runoff time series into several relatively stationary sub-time series; (2) the Box-Cox
98 transformation is employed to normalize the candidate input variables for each sub-time series; (3) a
99 hybrid model incorporating VMD, Box-Cox, and Elman is constructed, and tested using runoff data
100 collected from Wei River Basin; (4) the robustness and generalizability of the proposed hybrid model
101 is evaluated by comparing it to other combinations of models.

102 2. Methodologies

103 2.1 Variational Mode Decomposition

104 Variational mode decomposition (VMD), a novel non-recursive signal processing algorithm
105 proposed by Dragomiretskiy and Zosso, is utilized to adaptively decompose a complex non-stationary
106 signal into a set of discrete bandwidth-limited modes in a spectral domain by Wienner filering

(Dragomiretskiy and Zosso 2014). Given the original runoff time series $f(t)\left(t=1,2,..,n\right)$, regarded as a non-stationary signal, it is decomposed into K different modal functions $\mu_k(t)$（$k=1,2,..,K$）with central frequency ω_k. The bandwidth and central frequency of each mode can be estimated by seeking the optimums of constrained variational modes. The objective function is that the sum of the estimated bandwidths of all sub-series is minimized, while the constraint satisfies that the sum of each decomposed mode is equal to the original signal, which can be expressed as:

$$\begin{aligned}&\min_{\{\mu_k\},\{\omega_k\}}\left\{\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\otimes\mu_k(t)\right]e^{-j\omega_k t}\right\|_2^2\right\}\\&s.t.\sum_{k=1}^{K}\mu_k=f(t)\end{aligned}\tag{1}$$

where $\{\mu_k\}$ is the set of modes, $\{\omega_k\}$ is the set of central frequencies of modes, δ_t represents the Dirac distribution, K is the number of modes, and $\otimes$ denotes a convolution operation.

The above constrained problem can be transformed to an unconstrained variational problem by introducing Lagrange multiplier and quadratic penalty to construct the augmented Lagrange term L:

$$\begin{aligned}L\left(\{\mu_k\},\{\omega_k\},\lambda\right)=\alpha\sum_{k=1}^{K}\left\|\partial(t)\left[\left(\delta(t)+\frac{j}{\pi t}\right)\otimes\mu_k(t)\right]e^{-j\omega_k t}\right\|_2^2+\\\left\|f(t)-\sum_{k=1}^{K}\mu_k(t)\right\|_2^2+\left\langle\lambda(t)f(t)-\sum_{k=1}^{K}\mu_k(t)\right\rangle\end{aligned}\tag{2}$$

Where α and λ denote the penalty parameter and Lagrange multiplier, respectively.

The alternating direction method of multipliers (ADMM), an effective splitting method for separable optimization problems (Feng et al. 2020), is employed to seek out the saddle points of the augmented Lagrange term, and, therefore, the optimums of the decision variables μ_k and ω_k are presented as follows:

$$\begin{cases} \hat{\mu}_{\mathrm{k}}^{n+1}(\omega)=\dfrac{\hat{f}(\omega)-\sum\limits_{i\neq k}\hat{\mu}_i^n(\omega)+\dfrac{\hat{\lambda}^n(\omega)}{2}}{1+2\alpha\left(\omega-\omega_n^k\right)^2} \\ \omega_k^{n+1}=\dfrac{\int_0^\infty \omega\left|\hat{\mu}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{\mu}_k^{n+1}(\omega)\right|^2 d\omega} \\ \hat{\lambda}^{n+1}(\omega)=\hat{\lambda}^n(\omega)+\tau\left(\hat{f}(\omega)-\sum\limits_k \mu_k^{n+1}(\omega)\right) \end{cases} \quad (3)$$

where n is the number of iterations, τ is the update parameters of Lagrange multipliers, and $\hat{\mu}_{\mathrm{k}}^{n+1}(\omega)$, $\hat{f}(\omega)$, $\hat{\mu}_{\mathrm{i}}^n(\omega)$,and $\hat{\lambda}^{\mathrm{n+1}}(\omega)$ represent the fast Fourier transform of $\mu_k^{n+1}(t)$, $f(t)$, $\mu_i^n(t)$,and $\lambda^n(t)$, respectively.

2.2 Box-Cox transformation

The Box-Cox power transformation, proposed by Box and Cox in 1964, is applied to stabilize variance and remove skewness of time series, and convert them into more normal like distribution (Box and Cox 1964; Hamasaki and Kim 2008). Supposing y_i (i=1, 2, …n) is a positive random variable, the Box-Cox transformation with parameter λ is defined as:

$$y_i^{(\lambda)}=\begin{cases}\dfrac{y_i^\lambda-1}{\lambda}, \lambda\neq 0 \\ \log y_i, \lambda=0\end{cases}, i=1,2,...n \quad (4)$$

where y_i is the original signal (y_i >0), $y_i^{(\lambda)}$ is the transformed signal, and λ denotes the transform coefficient.

In order to estimate the parameters of the model, the transformed maximum likelihood function is adopted as the following (Hamasaki and Kim 2008),

$$\max(L)=-\frac{n}{2}\ln\sigma^2+(\lambda-1)\sum_{i=1}^{n}\ln y_i \quad (5)$$

where L denotes the likelihood function, σ^2 is the variance of $y_i^{(\lambda)}$.

Finally, the value of λ can be obtained by employing the optimization algorithm to solve the formula (5).

2.3 Elman Neural Network

The Elman neural network, first proposed by Elman in 1990, is a dynamic recurrent neural network with feedforward connections (Chandra 2015). The Elman network consists of four layers: input layer, hidden layer, context layer, and output layer.

Compared with feedforward neural networks, the additional context layer in the Elman neural network is employed to record the output information of the previous moment from the hidden layer and relay the former state in the current iteration, so the Elman neural network is sensitive to the historical input information and more suitable for time series forecasting (Mehrgini et al. 2019). Since the runoff time series have short-term dependency, the Elman neural network is suitable for runoff forecasting. The architecture is shown in Fig. 1.

Fig.1

The mathematical form can be expressed as (Li et al. 2019):

$$O_t = g\left(\omega_3 H_t + b_1\right) \tag{6}$$

$$H_t = f\left(\omega_1 C_t + \omega_2 I_t + b_2\right) \tag{7}$$

$$C_t = H_{t-1} \tag{8}$$

where I_t, H_t, C_t and $O(t)$ denote input, hidden, context and output layer vector, respectively. ω_1, ω_2 and ω_3 are weights connecting other two adjacent layers. g and f are the activation functions of the hidden layer and the output layer, respectively. b_1 and b_2 represent the bias vector.

2.4 Proposed VMD-BC-Elman Model

According to the above-mentioned methods, the hybrid model is proposed and diagrammed in Fig.2. The modeling procedures are described as follows:

Step1: The original runoff data is decomposed into k sub-time series that are bandwidth-limited with a center frequency using VMD.

Step2: Each sub-time series from the step 1 is reconstructed using phase space reconstruction (PSR) to generate candidate input variables series for predicting models, and each reconstructed candidate input variable series is normalized by Box-Cox.

Step3: For each candidate input variables series transformed by Box-Cox from the step 2, LASSO is employed to rank and select the optimal number of the input variables for Elman model.

Step4: Each sub-time series decomposed by VMD is divided into training dataset and testing dataset. The parameters can be calibrated by inputting the selected input variables into the Elman, and then the trained model is tested using the testing dataset.

Step5: The final estimates of runoff time series are obtained by summing up the sub-time series estimated from the predicting models, respectively.

Fig.2

2.5 Model performance evaluation

Nash-Suttcliffe efficiency coefficient (NSE) (Nash and Sutcliffe 1970), the coefficient of correlation (R), the normalized mean square error (NMSE) (Himanshu et al. 2017), and the percent of the bias error (PBIAS) (Wen et al. 2019), are adopted as performance metrics to evaluate the prediction

performance of different models. These metrics are defined as:

$$NSE = 1 - \frac{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - Q_{\mathrm{p}}(t)\right)^2}{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - \overline{Q_{\mathrm{o}}}\right)^2} \tag{9}$$

$$R = \frac{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - \overline{Q_{\mathrm{o}}}\right)\left(Q_{\mathrm{p}}(t) - \overline{Q_{\mathrm{p}}}\right)}{\sqrt{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - \overline{Q_{\mathrm{o}}}\right)^2\left(Q_{\mathrm{p}}(t) - \overline{Q_{\mathrm{p}}}\right)^2}} \tag{10}$$

$$NMSE = \frac{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - Q_{\mathrm{p}}(t)\right)^2}{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t)\right)^2} \times 100\% \tag{11}$$

$$PBIAS = \frac{\sum_{t=1}^{N}\left(Q_{\mathrm{o}}(t) - Q_{\mathrm{p}}(t)\right)}{\sum_{t=1}^{N} Q_{\mathrm{o}}(t)} \times 100\% \tag{12}$$

Where $Q_{\mathrm{o}}(t)$ and $Q_{\mathrm{p}}(t)$ is the observed and predicted runoff, respectively, $\overline{Q_{\mathrm{o}}}$ and $\overline{Q_{\mathrm{p}}}$ denotes the mean of observed and predicted monthly runoff time series, respectively. N is the number of $Q_{\mathrm{p}}(t)$.

3 Case study

3.1 study area and dataset

The Wei River originates from Weiyuan County in Gansu Province, extends more than 800 km and finally converges into Yellow River (Jiang et al. 2019). The Wei River Basin (WRB) covers an area of approximately $13.5\times10^4 km^2$ between 104°00′-110°20′E and 33°50′-37°18′N. Topographically, the elevation of the WRB ranges from 336 m to 3929 m (Wang et al. 2020), and a digital elevation model (DEM) of the WRB is presented in Fig.3. The basin is located in arid and semi-arid regions in China, affected by a continental monsoon, and the climate is warm and rainy in summer and cold and

dry in winter, and the precipitation and runoff, exhibiting extremely high intra-annual and inter-annual variability, mainly occur in the wet season from June to October accounting for almost 65% of the annual precipitation and over 50% of the annual runoff, respectively (Zou et al. 2018). Notably, the WRB has been an important agricultural and industrial region in Shaanxi Province, and the runoff has been disturbed by human activities. Thus, the WRB is a suitable candidate to verify the feasibility of proposed model for predicting the non-stationary runoff time series disturbed by climate change and human activities.

Fig.3

The monthly runoff time series employed in this paper were observed from the Zhangjiashan, Zhuangtou and Huaxian stations, which are the main control stations in Jing River, Beiluo River and Wei River, respectively, and data had been collected from 1/1933 to 12/2016, 1/1938 to 12/2016 and 1/1954 to 12/2016 in the published Hydrological Yearbook, respectively, as shown in Fig.4. It is worth mentioning that the data from Zhangjiashan station is first employed to develop the hybrid model and data from Zhuangtou and Huaxian stations are utilized to further validate the performance of the model.

Fig.4

Tab.1 presents the statistical characteristics of runoff series. It is shown that Cv values are over 0.90 in the three stations, and the runoff varies within a relatively large range, and Cs values are greater than zero, indicating that the runoff series have skewness.

Tab.1

3.2 Variability analysis of runoff time series

The monthly runoff hydrographs are plotted in Fig.4, which illustrates drastic intra- and inter-annual variability and high non-stationary. The cumulative departure curve in Fig.5 was used to identify variability points of the runoff time series. It can be clearly seen that the variability points of

the runoff time series occurred in 1996, 1994 and 1993 at Zhangjiashan, Zhuangtou and Huaxian stations, respectively. Considering that the training period of the model is expected to contain sufficient information on the runoff data after the variability points, the monthly runoff data set is partitioned into the training subset (80%, before December 2000) and the testing subsets (20%, from January 2001 to December 2016) as shown in Fig.4. In theory, the proposed model could capture variability of the runoff time series during the training period.

Fig.5

3.3 Decomposition Results with VMD

The K and α, subject to the constrains of the input runoff time series, are two key parameters of VMD algorithm. To achieve satisfactory performance, the center frequency iteration is run to determine the number of the sub-time series K, the results show that a convergence of the center frequency of the high frequency sequence is present when K=8-10, indicating the decomposition has been completed and K=8 is a determined value, which is consistent with earlier study (Huang et al. 2016). According to the multiple pre-experiment, α=1000 is determined. The default parameters will be adopted for other parameters of the VMD model as shown in Tab.2.

Tab.2

The sub-time series of the original runoff series in Zhangjiashan station are obtained by VMD decomposition, and are shown in Fig.6.

Fig.6

3.4 Reconstruction and transformation by PSR and Box-Cox

In order to explore the intrinsic characteristics of the runoff variation, each one-dimensional IMF (sub-time series) is reconstructed into a high-dimensional feature space using PSR. Considering the periodic nature of runoff in 12-month, when PSR dimension d=12, and delay time τ=1 (Packabd et al.1980), the t-th space vector in the reconstructed phase space is expressed as $X_t = \{x_{t-1}, x_{t-2}, \mathrm{L}\ , x_{t-12}\}$ $t = 13,14,\mathrm{L}\ ,n$, where, n is the length of the original monthly runoff series, and $x_{t-1}, x_{t-2}, ..., x_{t-12}$, respectively, denote 12 different dimensional series, namely 12 candidate input variables series.

To avoid the unsatisfactory prediction performance arising from the skewness of runoff data, each candidate input variables series is converted into normal distribution using Box-Cox. The optimal transformation coefficient λ is estimated by maximum likelihood method, and the estimated λ values of each candidate input variables series for runoff time series are shown as Tab.3.

Tab.3

Take the original runoff series as a case to illustrate the normalization of Box-Cox, the probability density of the untransformed and transformed candidate input variables are shown in Fig.7. It can be seen the candidate input variables series transformed by Box-Cox appear the normal distribution.

Fig.7

Fig.8 is used to analyze the correlation between the output variables and the 12 candidate input variables transformed by Box-Cox. It can be seen that correlation coefficients of the transformed data have been improved, indicating that Box-Cox strengthens the correlation of the data.

Fig.8

3.5 Input variables selection by LASSO regression

Least absolutely shrinkage and selection operator (LASSO) is a kind of variable selection method which performs well for handling multi-collinearity issue. The LASSO selects significant input variables by adjusting the value of tuning parameter ρ from large to small and controlling the number of non-zero coefficients of candidate input variables (Marami et al., 2016). The importance ranking and the optimal number of selected input variables are shown in the Tab.4. Take the original runoff series as an example, the optimal number of selected input variables is 4 using trial and error method, thus the input variables are $x_{t-1}, x_{t-12}, x_{t-11}, x_{t-2}$.

Tab.4

3. 6 Development and training of the forecasting models

In order to verify the performance of the VMD-BC-Elman model, single models and other combinations of models are used for comparison in Fig.9. The training of these models will be described in this section.

Fig.9

Group 1: Single models

The standard Elman and SVM are individually used as basic models for developing the hybrid models. The activation functions and parameters of Elman are set utilizing the training data sets with the maximum number of the training as 500, the target error 0.001, and the learning coefficient 0.01. The hidden layer activation function employs Tansig, the output layer activation function uses Purelin, and the training algorithm adopts L-M algorithm. The number of each layer is determined using the

277 trial-and-error method, and the trained Elman architecture is presented in Tab.5. In SVM model, the
278 radial basis kernel function is selected with insensitivity loss parameter at 0.01, penalty factor at 1 and
279 kernel function parameter at 2.

280 Group 2: Decomposition-based hybrid models

281 Decomposition-based hybrid models are developed by coupling EMD, EEMD and VMD with
282 Elman or SVM, respectively. The original runoff series are decomposed into 6, 11 and 8 sub-time
283 series by EMD, EEMD and VMD, respectively. Each sub-time series are modelled using the Elman or
284 SVM. In Elman training, the target error is set at 0.00001, and the remaining parameters and the
285 activation functions are constants as the single model have in Group 1. The three-layer architecture of
286 each sub-time series is determined by trial and error method, and the optimal number of the input, with
287 hidden and output neurons listed in Tab.5. For SVM training, the number of the input variables of each
288 sub-time series is selected in Tab.4 and the other parameters are the same with those of the standard
289 SVM.

290 Tab.5

291 Group 3: Box-Cox-based hybrid models

292 To reduce the negative effect of the skewed runoff data on the modeling, BC-based models are
293 developed by introducing Box-Cox into Elman or SVM, respectively. The input variables series are
294 transformed into normal distribution using Box-Cox with the parameters λ in Tab.3. The model
295 architecture and parameters are the same as the single models.

296 Group 4: Hybrid models–VMD-BC-Elman and VMD-BC-SVM

297 In order to further improve the models' performance and sufficiently overcome the problems
298 arising from the non-stationary and skewed runoff data, the VMD-BC-Elman and VMD-BC-SVM are

299 proposed by introducing Box-Cox and VMD into the Elman or SVM. In this section, the VMD-BC-
300 Elman architecture of each sub-time series is the same as those of VMD-Elman, and the VMD-BC-
301 SVM modeling parameters are the same as those of SVM.

302 3.7 Results and discussion for Zhangjiashan station

303 Based on the training models above, runoff time series are modelled using the different models,
304 respectively. And the performance statistical metrics are listed in the Tab.6. The hybrid models are
305 shown to be more accurate than their standalone counterparts; and the performance of the VMD-BC-
306 Elman is best in all models.

307 In the standalone models, Elman and SVM have unsatisfactory predicting results with NSE<0.6,
308 lower R and higher NMSE and PBIAS as shown in the Tab.6. Compared to SVM, the Elman provides
309 better general results in training period because its recurrent feedback networks with short-term
310 memory increase the ability to process dynamic information. However, the performance of the Elman
311 is unstable, which may lead to over-fitting. And the standalone SVM, which is based on the risk
312 minimization principle, has more robustness and generalizability for estimating non-stationary and
313 skewed time series than the Elman does. The Elman model provides better predicting results in training
314 period and SVM provides better predicting results in testing period, which is consistent with other
315 studies (Song et al. 2020; Feng et al. 2020).

316 Compared with the standalone models, the metric values of decomposition-based hybrid models
317 are satisfactory, as shown in the Tab.6. The six models yield R values in the range of 0.7878 to 0.9843,
318 and NSE values from 0.5754 to 0.9592, and NMSE values from 1.06% to 42.24%, and PBIAS values

319 from 1.26% to 19.26% in the testing period, thus the decomposition-based hybrid models are superior
320 to the counterpart single models (Napolitano et al. 2011; Mohammad et al. 2019). The metric values
321 of VMD-Elman are best in six models which indicates the VMD-Elman outperforms the other five
322 hybrid models and the VMD has a better ability to denoise the runoff data than EMD and EEMD do,
323 which agrees with the results of other researchers who have modelled runoff time series based on EMD
324 EEMD and VMD decomposition (Xie et al. 2019; Zuo et al. 2020).

325 In order to eliminate skewness of runoff time series, BC-based models are introduced. The Tab.6
326 shows that the statistical metrics of the BC-SVM and BC-Elman models are much better than those of
327 the single models. Taking NMSE in the testing period as an example, the NMSE for the BC-Elman,
328 BC-SVM, Elman and SVM models are 18.3%, 24.84%, 92.91% and 66.93%, respectively. This
329 illustrates that the performance of the BC-based hybrid models is superior to the single models, and
330 BC-Elman is better than BC-SVM. The excellent performance of BC-based models in this paper is
331 similar to the study on hydrology frequency analysis, which shows that Box-Cox is capable of
332 providing robustness in fitting models (Seong 2014).

333 The statistical metrics of the VMD-BC-Elman and VMD-BC-SVM are shown in the Tab.6. In the
334 testing period, the R of the VMD-BC-Elman are 0.04%, 6.05%, and 0.14% greater than that of VMD-
335 Elman, BC-Elman and VMD-BC-SVM models; its NSE is 0.08%, 17.65%, and 0.39% greater
336 respectively; its NMSE is 1.97%, 78.25%, and 8.50% smaller, and its PBIA is 68.29%, 96.54%, and
337 46.58% smaller, respectively. The statistical metrics values demonstrate the fact that, although VMD
338 or Box-Cox alone can improve the ability of a single Elman, integration of VMD and Box-Cox into
339 Elman can promote their predicting to a new level with the VMD-BC-Elman outperforming the VMD-

BC-SVM. Considering these results, it is evident that the proposed VMD-BC-Elman has more stable and consistent performance. It is affirmed that the AI models, combined the VMD and Box-Cox, represent a significant improvement in runoff prediction.

Tab.6

While the statistical metrics can evaluate the overall performance of models, hydrographs and scatter plots can further identify the temporal correspondence. Fig.10-11 display the hydrographs and scatter plots of the observed and predicted runoff from the different models in the testing periods. In Fig.10 (a) and Fig. 11(a)-(b), the single models are not able to accurately capture the observed runoff, especially for the peak flow, and the scatter plots of the observed and estimated runoff are more dispersed than the hybrid models. Notably, in Fig.10(b)-(c) and Fig.11(c)-(j), the estimated values generated by the decomposition-based and BC-based hybrid models are closer to the corresponding observed values and the scatter plots are clustered closer to the ideal fit. In Fig.10(d) and Fig.11(k)-(l), it is also obvious that VMD-BC-Elman and VMD-BC-SVM have a greater accuracy as the scatter plots are concentrated closer to the ideal fitting line, illustrating the VMD-BC-based hybrid models are more effective in estimating the peak values of data than other models. Particularly, the VMD-BC-Elman has the best capability to capture the information on the overall features of runoff time series. Therefore, it is also reaffirmed that the VMD and Box-Cox can effectively improve the stability and consistency of the models, and the VMD-BC-Elman exhibits the strongest performance.

Fig.10

Fig.11

In Fig.12, the boxplot diagram is designed to depict the errors distribution between estimated and

observed values in the entire tested dataset in order to further verify the prediction performance of all the models. It can be seen from Fig.12 that the predicted errors including the first, median and third quartiles, minimum and maximum non-outlier are smallest when the VMD-BC-Elman is employed compared with the compared models. It indicates that the VMD and Box-Cox enhance the robustness of the models, and the VMD-BC-Elman is superior to the other models, which is consistent with the above findings.

Fig.12

The Taylor diagram as shown in Fig.13, integrating multiple characteristics of models into a compact plot (Taylor 2001), can present the performance of the models by a comprehensive way. It shows how close or faraway the simulated values of each model to or from the observed runoff data in the testing period, quantified via the correlation coefficient, standard deviation and root mean square error in Taylor diagram. It is evident that VMD-BC-Elman is the closest to the observed reference point with the lowest RMSD, the highest correlation coefficient and the lower SD, following by the VMD-BC-based and the VMD-based models while the single models are positioned the furthest from the reference point. Obviously, the VMD-BC-Elman performs best for monthly runoff prediction, which can be explained that the provision of the VMD and Box-Cox contributes to the improved stability and consistency of the models, and avoid over-fitting.

Fig.13

3.8 Results and discussion for Zhuangtou and Huaxian station

In order to further verify the feasibility of the VMD-BC-Elman, the runoff at Zhuangtou and Huaxian stations are modelled. In Tab.7 and Tab.8, the hybrid models are better than the single models

382 in the testing period. The VMD-based models are generally more accurate and stable than the EMD-
383 based and the EEMD-based models. Meanwhile, the BC-based models outperform the single models.
384 Obviously, the VMD-BC-Elman and VMD-BC-SVM outperform the VMD-based and the BC-based
385 models, and the VMD-BC-Elman is superior to the VMD-BC-SVM which is similar to the results of
386 Zhangjiashan station, so it is safe to conclude that the VMD-BC-Elman has the best forecast
387 performance in this paper.

388 Tab.7

389 Tab.8

390 In Fig.14, the proposed hybrid VMD-BC-Elman exhibits the best fit for the observed runoff
391 compared with other models, and estimates by the VMD-BC-Elman are closest to the corresponding
392 observed values and follow the same trend in all plots. Their scatter plots are closest to the ideal line,
393 especially around the peak flow, which also indicates that the VMD-BC-Elman has best predictive
394 skills. In conclusion, the VMD-BC-Elman exhibits the best performance in terms of the accuracy,
395 stability and consistency, and the combination of VMD and Box-Cox has great potential in monthly
396 runoff prediction areas.

397 Fig.14

398 # 4 Conclusions

399 To estimate non-stationary and skewed monthly runoff series, a novel VMD-BC-Elman model
400 was developed in this paper and runoff data in Zhangjiashan, Zhuangtou and Huaxian stations in the
401 Wei River Basin were used. The steps of this model are as follows: (1) The original non-stationary
402 monthly runoff time series were decomposed into a set of relatively stationary sub-time series IMFs

by VMD. (2) Each IMF was reconstructed into a 12-dimensional phase space by PSR, and each candidate input variables series was normalized using Box-Cox. (3) LASSO was employed to determine the optimal number of the input variables. (4) Each sub-time series was simulated by the Elman. The final results are generated by summing up the estimates of all the sub-time series.

The proposed VMD-BC-Elman exhibited the best precision, stability and consistency for predicting the non-stationary and skewed monthly runoff series. The remarkable advantages were demonstrated in the following aspects: (1) The proposed model with a perfect denoising ability, can effectively distinguish the modalities of the complex non-stationary series, and adaptably decompose the original series into the relative stationary sub-time series using the VMD. (2) It can remove the skewness of the original runoff series using Box-Cox, and convert the skewed candidate input variables into the normal distribution, which can enhance the correlation between the input and output variables, and improve the mapping ability and self-learning ability of the Elman. (3) Integrating decomposition- normalization- simulation- reconstruction techniques together, VMD-BC-Elman can effectively identify and separate the different modal features, distinguish and construct the different distribution features of the runoff time series, and accurately sum up the optimal estimates of sub-time series to achieve the better prediction for the monthly runoff. Overall, the VMD-BC-Elman represents an effective tool for forecasting non-stationary and skewed monthly runoff series.

In spite of the remarkable performance of the VMD-BC-Elman over comparative models, the study has limitations that create opportunities for further research. Firstly, the prediction results of high IMF are unsatisfactory possibly due to too much noise. Secondly, the proposed model only considers one-step prediction, and multi-step prediction has not been discussed yet. Finally, the proposed model does not take in account the physical mechanism of runoff formation. The following aspects are worthy of further research. Firstly, the high- and low-IMFs are expected to develop different models to more

effectively predict the different frequency characteristics. Secondly, it would be useful to explore a model utilizing multi-step prediction approach. Thirdly, physically-based hydrology model can be introduced into the current VMD-BC-Elman, so this data-driven model can combine the strength of self-learning ability with enhanced physical mechanism.

Declarations

Acknowledgements This research was funded by Science-Technology Plan Program of Water Conservancy Fund of Shaanxi Province, grant number 2019slkj-14, and National Natural Science Foundation of China under grant 5149222, 52079110. Sincere gratitude is extended to the editor and anonymous reviewers for their professional comments and corrections.

Conflict of Interest The authors declared that they have no conflicts of interest to this work.

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

Code availability The code used to support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions Fangqin Zhang**:** Investigation, Modeling, Calculation, Writing-Original Draft. Yan Kang**:** Conceptualization, Methodology, Writing-Review & Editing, Supervision. Xiao Cheng: Investigation, Data curation. Peiru Chen**:** Investigation. Songbai Song: Writing-Review & Editing.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publication Not applicable.

References

Aggarwal SK, Goel A, Singh VP (2012). Stage and discharge forecasting by SVM and ANN techniques. Water Resource Management 26 (13): 3705–3724.

Barge JT, Sharif HO (2016). An ensemble empirical mode decomposition, self-organizing map, and linear genetic programming approach for forecasting river streamflow. Water 8(2):247.

Bojang PO, Yang TC, Pham QB (2020). Linking Singular Spectrum Analysis and Machine Learning for Monthly Rainfall Forecasting. Applied Sciences-Basel 10(09).

Box GEP, Cox, DR (1964). An analysis of transformations. Journal of the Royal Statistical Society B26: 211–252.

Chandra R (2015). Competition and Collaboration in Cooperative Coevolution of Elman Recurrent Neural Networks for Time-Series Prediction. IEEE Transactions on Neural Networks and Learning Systems 26(12):3123-3136.

Clark MP, Bierkens MFP, Samaniego L, et al (2017). The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. Hydrology and Earth System Sciences 21(07):3427-3440.

Deb S, Dandapat S, Krajewski (2020). Analysis and Classification of Cold Speech Using Variational Mode Decomposition. IEEE Transactions on Affective Computing 11(02):296-307.

Demirel MC, Venancio A, Kahya E, 2009. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. Advances in Engineering Software 40(07):467-473.

Dragomiretskiy K, Zosso D (2014). Variational mode decomposition. IEEE Trans Signal Process 62(3):531-544.

468 Feng ZK, Niu WJ, Tang ZY, Jiang ZQ, Xu Y, Liu Y, Zhang HR (2020). Monthly runoff time series
469 prediction by variational mode decomposition and support vector machine based on quantum-
470 behaved particle swarm optimization. Journal of Hydrology 583.
471 Hamasaki T, Kim SY (2017). Box and Cox power-transformation to confined and censored non-normal
472 responses in regression. Computational Statistics &Data Analysis 51(8):3788-3799.
473 Himanshu SK, Pandey A, Yadav B (2017). Ensemble Wavelet-Support Vector Machine Approach for
474 Prediction of Suspended Sediment Load Using Hydrometeorological Data. Journal of Hydrologic
475 Engineering 22(07).
476 Huang N, Chen H, Cai G, Fang L, Wang Y (2016). Mechanical fault diagnosis of high voltage circuit
477 breakers based on variational mode decomposition and multi-layer classifier. Sensors 16(11).
478 Huang NE, Wu MLC, Long SR, Shen SSP, Qu WD, Gloersen P, Fan KL (2003). A confidence limit
479 for the empirical mode decomposition and Hilbert spectral analysis. Proceedings of the royal
480 society a-mathematical physical and engineering sciences 459(08):2317-2345.
481 Jiang RG, Wang YP, Xie JC, Zhao Y, Li FW, Wang XJ (2019). Assessment of extreme precipitation
482 events and their teleconnections to El Nino southern oscillation, a case study in the Wei River
483 basin of China. Atmospheric Research 218:372–384.
484 Karamouz M, Ahmadi A, Moridi A (2009). Probabilistic reservoir operation using Bayesian stochastic
485 model and support vector machine. Advances in Water Resources 32(11):1588–1600.
486 Krishnan S, Lokesh S, Devi MR (2019). An efficient Elman neural network classifier with cloud
487 supported internet of things structure for health monitoring system. Computer Networks
488 14(5):201-210.

Kuremoto T, Kimura S, Kobayashi K, Obayashi M (2014). Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing 137: 47–56.

Li H, Fan BJ, Jia R, Zhai F, Bai L, Luo XQ (2020). Research on multi-domain fault diagnosis of gearbox of wind turbine based on adaptive variational mode decomposition and extreme learning machine algorithms. Energies 13(6):1375.

Li XY, Zhang L, Wang ZP, Dong P (2019). Remaining useful life prediction for lithium-ion batteries based on a hybrid model combining the long short-term memory and Elman neural networks. Journal of Energy Storage 21:510-518.

Mehrgini B, Izadi H, Memarian H (2019). Shear wave velocity prediction using Elman artificial neural network. Carbonates and Evaporites 34(04):1281-1291.

Moeeni H, Bonakdari H, Fatemi SE (2017). Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. Journal of Hydrology 547:348–364.

Mohamad JA, Mohamad RK, Ozgur K, Vahid N (2017). A new approach for simulating and forecasting the rainfall-runoff process within the next two months. Journal of Hydrology 548(4):588-597.

Mohanty S, Gupta KK, Raju KS (2018). Hurst based vibro-acoustic feature extraction of bearing using EMD and VMD. Measurement 117:200-220.

Myronidis D, Ioannou K, Fotakis D, Dorflinger G (2018). Streamflow and hydrological drought trend analysis and forecasting in Cyprus. Water Resources Management 32(5):1759–1776.

Napolitano G, Serinaldi F, See L (2011). Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: An empirical examination. Journal of Hydrology 406 (3):199–214.

510 Nash J, Sutcliffe J (1970). River flow forecasting through conceptual models part I—a discussion of
511 principles. Journal of Hydrology 10(3): 282–290.
512 Packard NH, Crutchfieid JP, Farmer JD, et al (1980). Geometry from a time series. Physical Review
513 Letters 45:712.
514 Partington D, Brunner P, Simmons CT, Werner AD, Therien R, Maier HP, Dandy GC (2012).
515 Evaluation of outputs from automated baseflow separation methods against simulated baseflow
516 from a physically based, surface water-groundwater flow model. Journal of Hydrology 458: 28–
517 39.
518 Peng Y, Wang GL, Tang GL, Zhou HC, Wang YJ, Jian DP (2011). Study on reservoir operation
519 optimization of Ertan hydropower Station considering GFS forecasted precipitation. Science
520 China-Technological Sciences 54: 76–82.
521 Sankaran A, Reddy MJ (2016). Analyzing the hydroclimatic teleconnections of summer monsoon
522 rainfall in Kerala, India, using multivariate empirical mode decomposition and time-dependent
523 intrinsic correlation. IEEE Geosci. Remote Sensing Lett 13(9):1221–1225.
524 Sarlak N (2008). Annual streamflow modelling with asymmetric distribution function. Hydrological
525 Processes 22(17):3403-3409.
526 Sedki A, Quazar D, EI Mazoudi E (2008). Evolving neural network using real coded genetic algorithm
527 for daily rainfall-runoff forecasting. Expert Systems with Applications 36:4523-4527.
528 Seong KW (2014). Deriving a practical form of IDF formula using transformed rainfall intensities.
529 Hydrological Processes 28(06):2881-2896.
530 Song PB, Liu WF, Sun JH, Wang C, Kong LZ, Nong ZX, Lei XH, Wang H (2020). Annual Runoff

Forecasting Based on Multi-Model Information Fusion and Residual Error Correction in the Ganjiang River Basin. Water 12(08).

Taylor KE (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research-Atmospheres 106 (D7):7183–7192.

Tiwari MK, Adamowaki J (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water resources research 49(10):6486-6507.

Vasiliades L, Loukas A (2009). Hydrological response to meteorological drought using the Palmer drought indices in Thessaly, Greece. Desalination 237(1-3):3-21.

Wang QY, Liu Y, Yue QM, Zheng YX, Yao XL, Yu JS (2020). Impact of Input Filtering and Architecture Selection Strategies on GRU Runoff Forecasting: A Case Study in the Wei River Basin, Shaanxi, China. Water 12(12).

Wang WC, Xu DM, Chau KW, Chen S (2013). Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. Journal of Hydroinform, 15(4):1377-1390.

Wen XH, Feng Q, Deo RC, et al (2019). Two-phase extreme learning machines integrated with the complete ensemble empirical mode decomposition with adaptive noise algorithm for multi-scale runoff prediction problems. Journal of Hydrology 570:167-184.

Wu Z, Huang NE (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances Adaptive Data Analysis 1(1):1–41.

Xie T, Zhang G, Hou JW, Xie JC, Lv M, Liu FC (2019). Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China. Journal of Hydrology 577(10).

Yu YH, Zhang HB, Singh VP (2018). Forward Prediction of Runoff Data in Data-Scarce Basins with

552 an Improved Ensemble Empirical Mode Decomposition (EEMD) Model. Water 10(04).
553 Zhang H, Singh VP, Wang B, Yu Y (2016). CEREF: a hybrid data-driven model for forecasting annual
554 streamflow from a socio-hydrological system. Journal of Hydrology 540:246–256.
555 Zhang X, Peng Y, Zhang C, Wang B (2015). Are hybrid models integrated with data preprocessing
556 techniques suitable for monthly streamflow forecasting? Some experiment evidences. Journal of
557 Hydrology 530: 137–152.
558 Zou L, Xia J, She DX (2018). Analysis of impacts of climate change and human activities on
559 hydrological drought: a case study in the Wei River basin, China. Water Resources Management
560 32(4):1421–1438.
561 Zuo GG, Luo JG, Wang N, et al (2020). Decomposition ensemble model based on variational mode
562 decomposition and long short-term memory for streamflow forecasting. Journal of Hydrology
563 585(06).
564

565 **Table Caption**

566 Tab.1 Profiles of hydrological stations and statistical characteristics of the runoff series at Zhangjiashan,

567 Zhuangtou and Huaxian stations.

568 Tab.2 Parameters of VMD determined in the iteration experiments.

569 Tab.3 The λ values of each candidate input variables series for the original runoff series and each

570 IMF at Zhangjiashan station.

571 Tab.4 Importance ranking of the selected input variables for original series and sub-time series

572 decomposed at Zhangjiashan station.

573 Tab.5 Elman predicted architecture of the original and each sub-time series based on different

574 decomposition methods at Zhangjiashan Station.

575 Tab.6 Statistical metrics for models' performance evaluation in the training and testing periods at

576 zhangjiashan station.

577 Tab.7 As in Tab.6, but for Zhuangtou station.

578 Tab.8 As in Tab.6, but for Huaxian station.

579 Tab.1 Profiles of hydrological stations and statistical characteristics of the runoff series at Zhangjiashan,

580 Zhuangtou and Huaxian stations.

Station	Longitude	Latitude	Control Area (km^2)	Data Sets	Data Length	Max (m^3 /s)	Min (m^3/s)	Mean (m^3/s)	Cv	Cs
Zhangjiashan	108°36′	34°38′	43,216	All	1933-2016	676.00	1.00	50.12	1.17	3.64
				Training	1933-2000	676.00	1.00	54.07	1.16	3.44
				Testing	2001-2016	213.00	7.00	33.34	0.97	3.42
Zhuangtou	109°50′	35°02′	25,154	All	1938-2016	199.00	2.00	24.86	0.90	3.19
				Training	1938-2000	199.00	2.00	26.41	0.88	3.07
				Testing	2001-2016	138.00	3.00	18.75	0.90	3.94
Huaxian	109°46′	34°35′	106,498	All	1954-2016	1690.00	2.70	206.65	1.15	2.69
				Training	1954-2000	1690.00	2.70	224.02	1.13	2.50
				Testing	2001-2016	1250.00	8.59	156.71	1.14	3.71

581 Max, Min, Mean, Cv and Cs represent maximum, minimum, mean value, coefficient of variation and coefficient of skewness of data

582 sets, respectively.

583

584 Tab.2 Parameters of VMD determined in the iteration experiments.

Parameters	α	τ	K	DC	Init	tol
Values	1000	0	8	0	1	10^{-7}

585

586 Tab.3 The λ values of each candidate input variables series for the original runoff series and each

587 IMF at Zhangjiashan station.

	x_{t-1}	x_{t-2}	x_{t-3}	x_{t-4}	x_{t-5}	x_{t-6}	x_{t-7}	x_{t-8}	x_{t-9}	x_{t-10}	x_{t-11}	x_{t-12}
Original	0.2104	0.2112	0.2112	0.2111	0.2109	0.2107	0.2108	0.2108	0.2114	0.2126	0.2126	0.2129
IMF1	0.2723	0.2796	0.2873	0.2953	0.3035	0.3118	0.3199	0.3276	0.3351	0.3420	0.3483	0.3536
IMF2	0.8304	0.8323	0.8345	0.8361	0.8359	0.8347	0.8343	0.8351	0.8352	0.8339	0.8333	0.8353
IMF3	1.0887	1.0871	1.0867	1.0869	1.0873	1.0881	1.0890	1.0891	1.0892	1.0899	1.0893	1.0909
IMF4	0.9419	0.9395	0.9384	0.9401	0.9434	0.9465	0.9473	0.9454	0.9458	0.9434	0.9460	0.9363
IMF5	1.1262	1.1273	1.1256	1.1247	1.1249	1.1259	1.1270	1.1276	1.1264	1.1267	1.1257	1.1268
IMF6	0.7677	0.7674	0.7678	0.7686	0.7678	0.7680	0.7690	0.7685	0.7679	0.7661	0.7653	0.7658
IMF7	1.1633	1.1634	1.1633	1.1631	1.1638	1.1630	1.1640	1.1635	1.1630	1.1636	1.1666	1.1659
IMF8	1.1784	1.1779	1.1789	1.1778	1.1792	1.1792	1.1788	1.1813	1.1801	1.1809	1.1822	1.1812

588

589 Tab.4 Importance ranking of the selected input variables for original series and sub-time series

590 decomposed at Zhangjiashan station.

No.	Series	Raking the selected input variables	Number of input variables
1	Original	$x_{t-1}, x_{t-12}, x_{t-11}, x_{t-2}, x_{t-10}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-9}, x_{t-6}, x_{t-8}, x_{t-7}$	4
2	IMF1	$x_{t-12}, x_{t-11}, x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$	3 (EEMD:6)
3	IMF2	$x_{t-9}, x_{t-3}, x_{t-10}, x_{t-8}, x_{t-4}, x_{t-2}, x_{t-11}, x_{t-12}, x_{t-7}, x_{t-5}, x_{t-6}, x_{t-1}$	7
4	IMF3	$x_{t-2}, x_{t-6}, x_{t-10}, x_{t-11}, x_{t-12}, x_{t-7}, x_{t-9}, x_{t-5}, x_{t-8}, x_{t-3}, x_{t-1}, x_{t-4}$	8
5	IMF4	$x_{t-11}, x_{t-10}, x_{t-8}, x_{t-7}, x_{t-4}, x_{t-5}, x_{t-1}, x_{t-2}, x_{t-12}, x_{t-9}, x_{t-6}, x_{t-3}$	8
6	IMF5	$x_{t-11}, x_{t-1}, x_{t-9}, x_{t-12}, x_{t-3}, x_{t-7}, x_{t-5}, x_{t-10}, x_{t-8}, x_{t-6}, x_{t-4}, x_{t-2}$	8
7	IMF6	$x_{t-12}, x_{t-10}, x_{t-11}, x_{t-7}, x_{t-9}, x_{t-4}, x_{t-8}, x_{t-1}, x_{t-5}, x_{t-6}, x_{t-2}, x_{t-3}$	8 (EMD:10)
8	IMF7	$x_{t-12}, x_{t-11}, x_{t-10}, x_{t-8}, x_{t-9}, x_{t-7}, x_{t-3}, x_{t-2}, x_{t-6}, x_{t-4}, x_{t-5}, x_{t-1}$	8
9	IMF8	$x_{t-7}, x_{t-8}, x_{t-6}, x_{t-9}, x_{t-12}, x_{t-10}, x_{t-11}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$	8

591

592 Tab.5 Elman predicted architecture of the original and each sub-time series based on different

593 decomposition methods at Zhangjiashan Station.

Series	Single model	Hybrid models with different decomposition methods		
	Elman	EMD-Elman	EEMD-Elman	VMD-Elman
Original	4-12-1	-	-	-
IMF1	-	3-15-1	6-15-1	3-10-1
IMF2	-	7-13-1	7-13-1	7-13-1
IMF3	-	8-13-1	8-13-1	8-13-1
IMF4	-	8-13-1	8-13-1	8-13-1
IMF5	-	8-13-1	8-13-1	8-13-1
IMF6	-	10-13-1	8-13-1	8-13-1
IMF7	-	-	8-13-1	8-13-1
IMF8	-	-	8-13-1	8-13-1
IMF9	-	-	8-13-1	-
IMF10	-	-	8-13-1	-
IMF11	-	-	8-13-1	-

594

595 Tab.6 Statistical metrics for models' performance evaluation in the training and testing periods at

596 zhangjiashan station.

Types	Models	Training Dataset				Testing Dataset			
		R	*NSE*	*NMSE*%	*PBIAS*%	*R*	*NSE*	*NMSE*%	*PBIAS*%
Original	SVM	0.6001	0.3229	67.63	18.39	0.5756	0.3272	66.93	5.83
	Elman	0.7395	0.5468	45.26	-0.53	0.5954	0.0660	92.91	-21.18
Decomposition	EMD-SVM	0.7735	0.5758	42.37	9.62	0.7878	0.5754	42.24	19.26
	EEMD-SVM	0.8772	0.7646	23.51	-3.25	0.9063	0.8028	19.62	-12.97
	VMD-SVM	0.9819	0.9524	4.75	0.43	0.9830	0.9552	4.45	1.81
	EMD-Elman	0.8983	0.8059	19.38	-0.94	0.8239	0.6371	36.10	2.65
	EEMD-Elman	0.9644	0.9301	6.98	0.04	0.9162	0.8200	17.91	-11.43
	VMD-Elman	0.9836	0.9580	4.19	0.13	0.9843	0.9592	4.06	1.23
Box-Cox transformation	BC-SVM	0.8674	0.7328	26.69	9.47	0.9059	0.7503	24.84	7.42
	BC-Elman	0.9409	0.8847	11.52	-2.81	0.9285	0.8160	18.30	11.28
Decomposition and Box-Cox	VMD-BC-SVM	0.9826	0.9537	4.63	0.19	0.9833	0.9563	4.35	0.73
	VMD-BC-Elman	0.9844	0.9596	4.03	0.10	0.9847	0.9600	3.98	0.39

597

598 Tab.7 As in Tab.6, but for Zhuangtou station.

Types	Model	Training dataset				Testing Dataset			
		R	*NSE*	*NMSE* %	*PBIAS*%	*R*	*NSE*	*NMSE*%	*PBIAS*%
Original	SVM	0.5693	0.2888	71.02	15.33	0.4928	0.2300	76.6	8.05
	Elman	0.7029	0.4940	50.53	0.16	0.3656	0.0588	93.63	-3.86
Decomposition	EMD-SVM	0.7756	0.5862	41.32	7.70	0.7665	0.5545	44.31	-15.74
	EEMD-SVM	0.8959	0.8007	19.90	2.60	0.9224	0.8424	15.68	-6.98
	VMD-SVM	0.9807	0.9495	5.04	0.03	0.9803	0.9445	5.52	-1.78
	EMD-Elman	0.9150	0.8372	16.26	-0.76	0.7889	0.4510	54.62	34.79
	EEMD-Elman	0.9580	0.9170	8.29	1.17	0.9202	0.8353	16.39	-3.12
	VMD-Elman	0.9829	0.9560	4.40	-0.09	0.9800	0.9483	5.15	0.18
Box-Cox transformation	BC-SVM	0.7999	0.6202	37.93	0.51	0.7840	0.5757	42.21	7.45
	BC-Elman	0.9304	0.8656	13.43	0.01	0.8029	0.6285	36.96	9.25
Decomposition and Box-Cox	VMD-BC-SVM	0.9815	0.9515	4.85	0.02	0.9793	0.9432	5.65	-1.82
	VMD-BC-Elman	0.9834	09565	4.35	-0.14	0.9827	0.9546	4.52	0.37

599

600 Tab.8 As in Tab.6, but for Huaxian station.

Types	Model	Training Dataset				Testing Dataset			
		R	*NSE*	*NMSE%*	*PBIAS%*	*R*	*NSE*	*NMSE%*	*PBIAS%*
Original	SVM	0.6478	0.3580	64.08	22.47	0.5335	0.2551	74.10	19.28
	Elman	0.7255	0.5260	47.31	-1.35	0.3596	-0.3594	135.23	-57.76
Decomposition	EMD-SVM	0.8463	0.6957	30.37	8.13	0.7819	0.5498	40.31	8.32
	EEMD-SVM	0.9164	0.8352	16.45	4.26	0.8969	0.8034	19.56	3.13
	VMD-SVM	0.9823	0.9555	4.44	0.45	0.9775	0.9436	5.61	0.62
	EMD-Elman	0.9489	0.8990	10.08	-3.46	0.6860	0.4184	57.86	10.15
	EEMD-Elman	0.9694	0.9393	6.06	2.47	0.8336	0.6814	31.69	10.17
	VMD-Elman	0.9853	0.9629	3.70	0.01	0.9788	0.9485	5.12	0.19
Box-Cox transformation	BC-SVM	0.8407	0.6845	31.49	13.26	0.8013	0.5994	39.85	10.12
	BC-Elman	0.9648	0.9309	6.90	0.01	0.7273	0.5202	47.73	10.17
Decomposition and Box-Cox	VMD-BC-SVM	0.9835	0.9579	4.20	0.26	0.9796	0.9473	5.24	0.24
	VMD-BC-Elman	0.9853	0.9631	3.68	0.01	0.9810	0.9525	4.73	-0.15

601

602 **Figure caption**

603 Fig.1 Architectural graph of Elman neural network.

604 Fig.2. Architecture of the proposed VMD-BC-Elman model (CIV= candidate input variables; IV=

605 input variables).

606 Fig.3 The study area showing the runoff observations sites in the Wei River Basin, China.

607 Fig.4 The observed monthly runoff series, the training and testing data sets at three hydrological

608 stations. (a)Zhangjiashan station; (b)Zhuangtou station; (c) Huaxian station.

609 Fig.5 The variability points of monthly runoff time series at three stations. (a)Zhangjiashan station;

610 (b)Zhuangtou station; (c)Huaxian station.

611 Fig.6 Decomposition results by VMD for monthly runoff time series at Zhangjiashan station. (a) IMF1-

612 IMF4; (b) IMF5-IMF8.

613 Fig.7 Probability density of the candidate input variables of the original runoff time series before and

614 after Box-Cox transformation at Zhangjiashan Station.

615 Fig.8 The correlation coefficients between the output and input variables before and after Box-Cox

616 transformation at Zhangjiashan Station.

617 Fig.9 The flowchart of modeling comparison.

618 Fig.10 The observed *vs.* predicted runoff in the testing period generated at Zhangjiashan station. (a)

619 Single models; (b) The decomposition-based hybrid models; (c) The Box-Cox transformation-based

620 hybrid models; (d) VMD-BC-SVM and VMD-BC-Elman.

621 Fig.11 Scatter plots of the predicted *vs.* observed runoff in testing period at Zhangjiashan station.

622 Fig.12 Boxplots of the predicted errors generated by 12 models in the testing period at Zhangjiashan

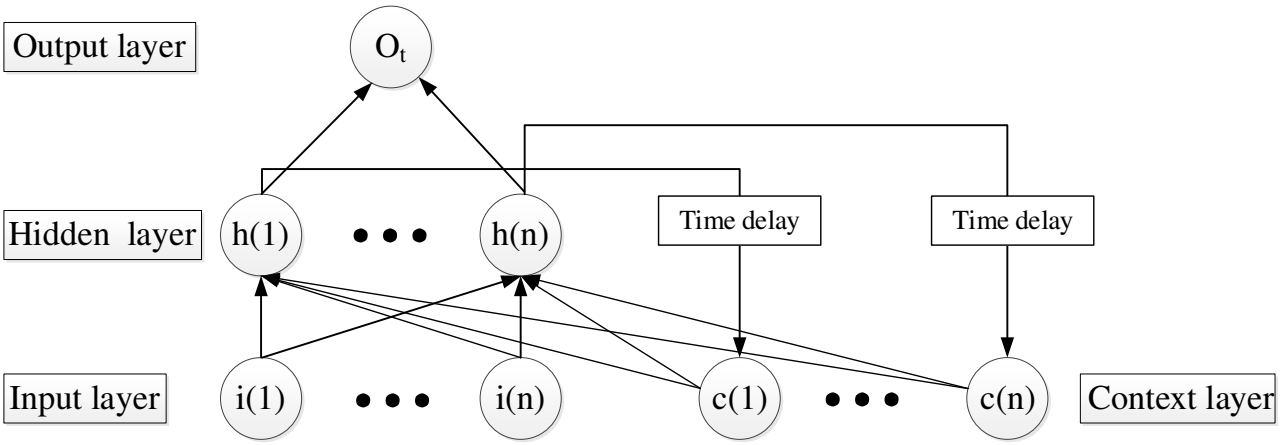
623 station (VMD-BC-Elman, VMD-BC-SVM, BC-Elman, BC-SVM, VMD-Elman, VMD-SVM, VMD-

624 Elman, EEMD-Elman, EEMD-SVM, EMD-Elman, EMD-SVM are abbreviated as VBE, VBS, BE,

625 BS, VE, VS, EEE, EES, EE, ES).

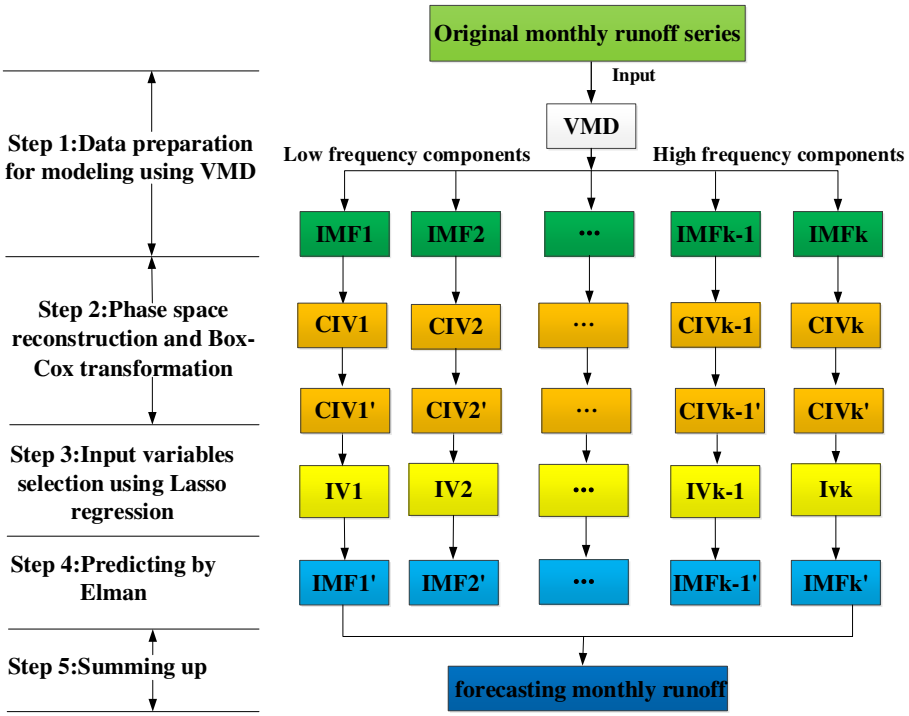
626 Fig.13 Taylor diagram depicting the predictive ability of 12 models in the testing period at
627 Zhangjiashan station.

628 Fig.14 The observed *vs.* predicted runoff in the testing period. (a) Zhuangtou station; (b) Huaxian
629 station.



630

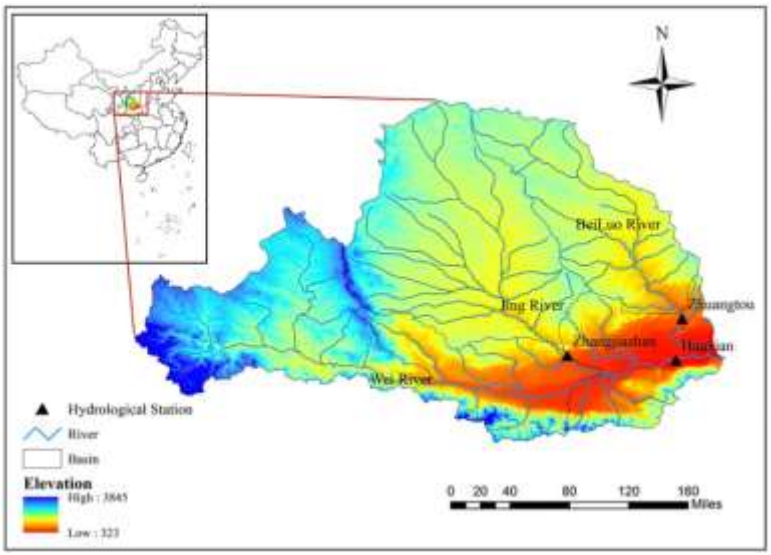
631 Fig.1 Architectural graph of Elman neural network.



632

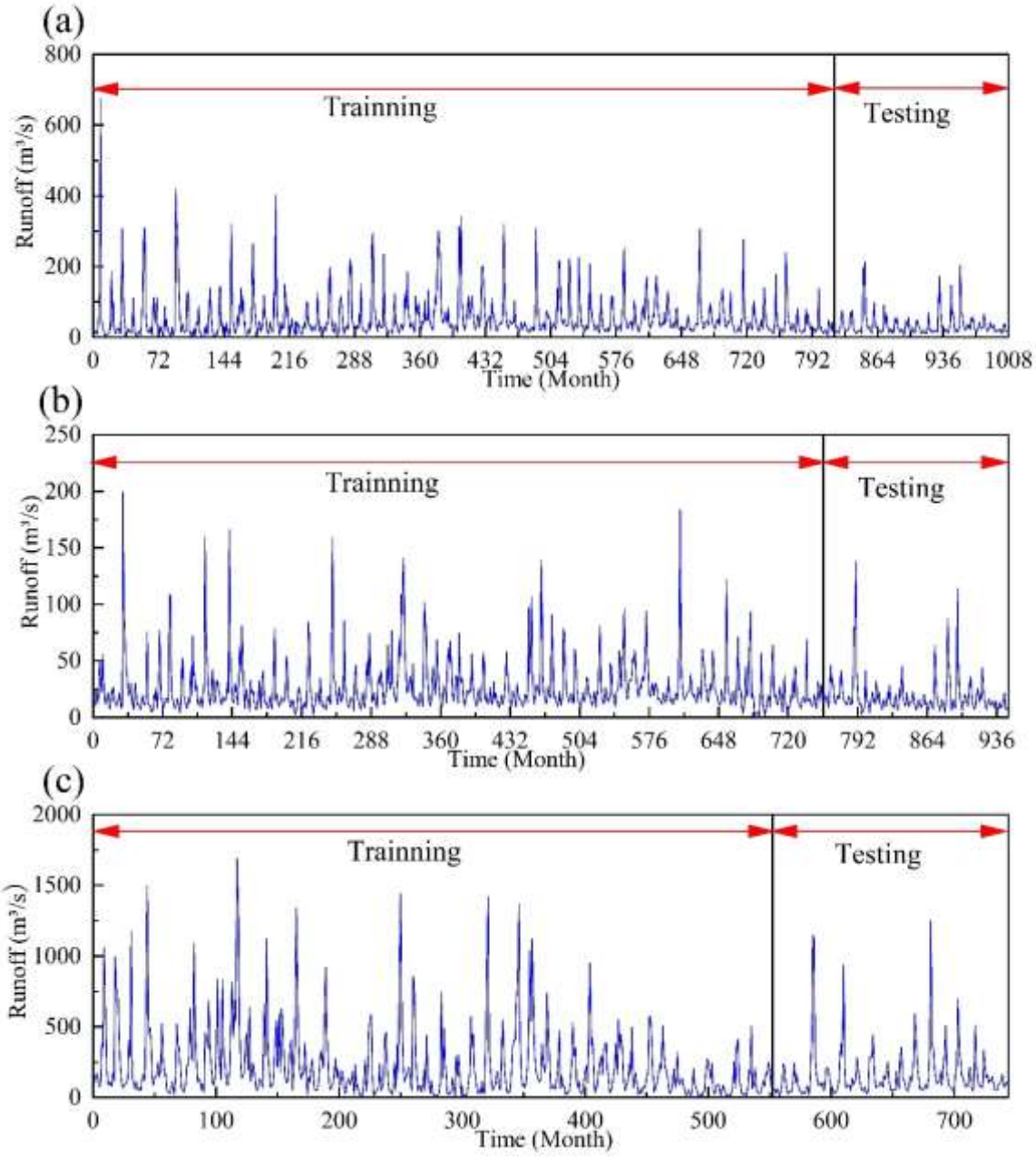
633 Fig.2 Architecture of the proposed VMD-BC-Elman model (CIV= candidate input variables; IV= input

634 variables).



635

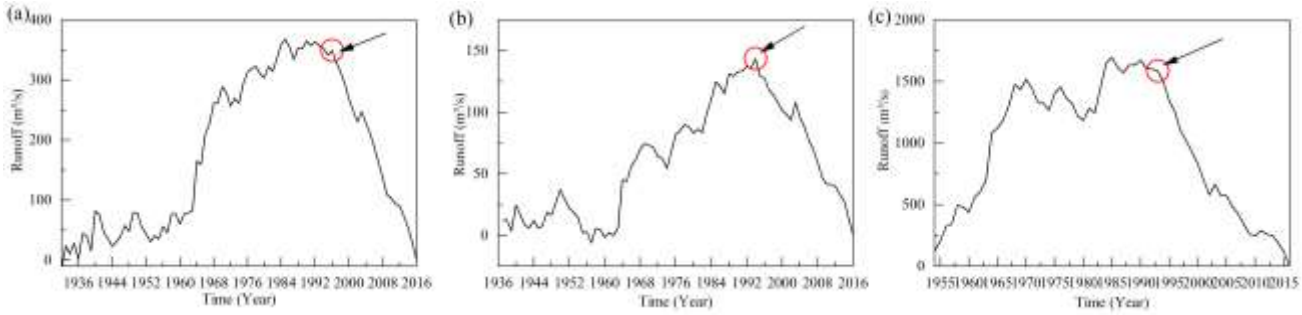
636 Fig.3 The study area showing the runoff observations sites in the Wei River Basin, China.



637

638 Fig.4 The observed monthly runoff series, the training and testing data sets at three hydrological

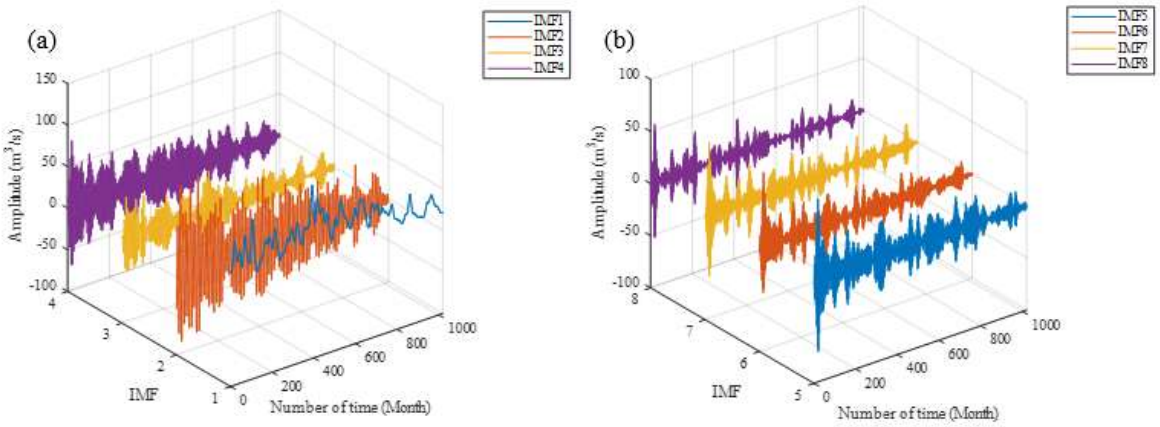
639 stations. (a)Zhangjiashan station; (b)Zhuangtou station; (c) Huaxian station.



640

641 Fig.5 The variability points of monthly runoff time series at three stations. (a)Zhangjiashan station;

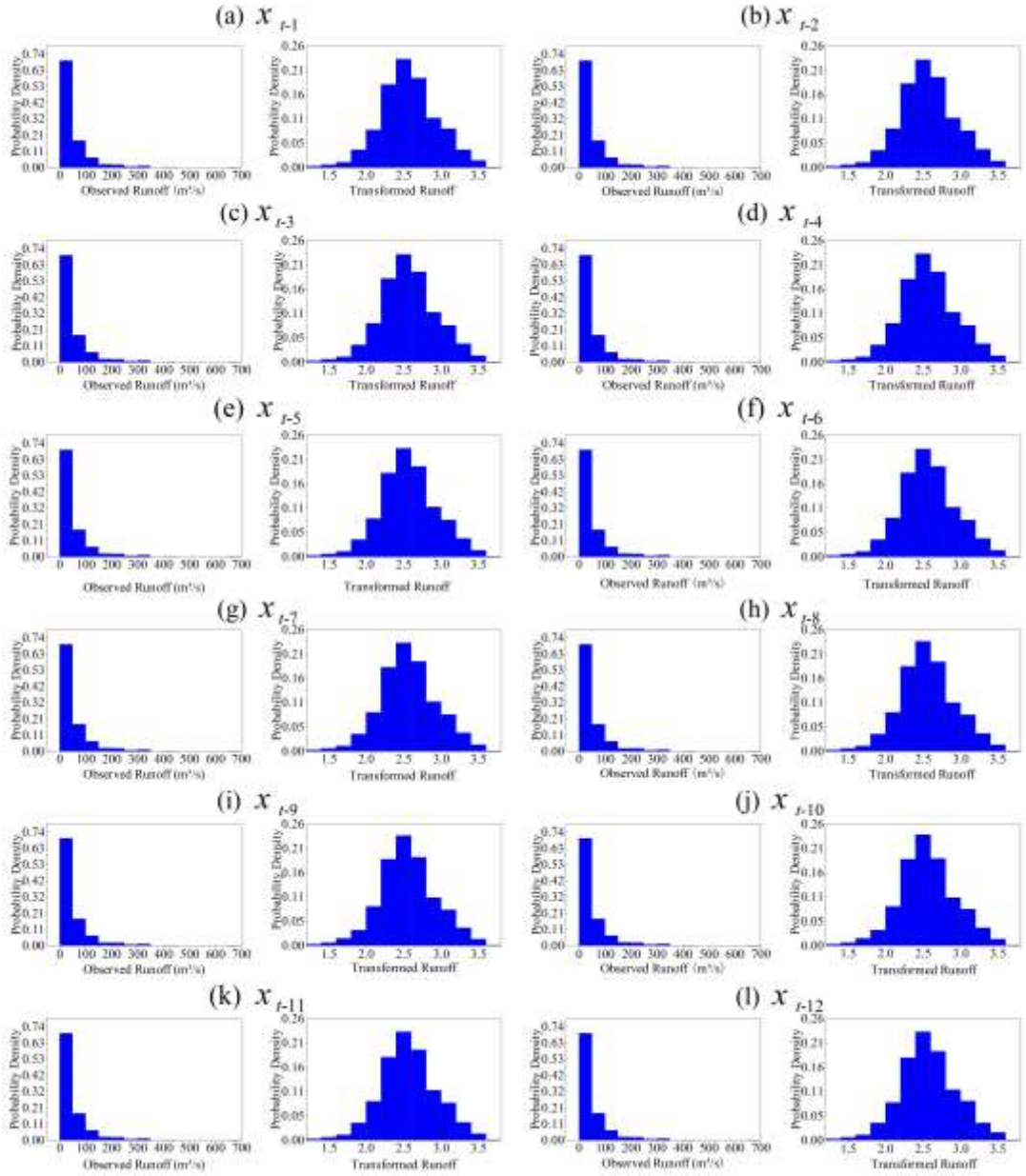
642 (b)Zhuangtou station; (c)Huaxian station.



643

644 Fig.6 Decomposition results by VMD for monthly runoff time series at Zhangjiashan station. (a) IMF1-

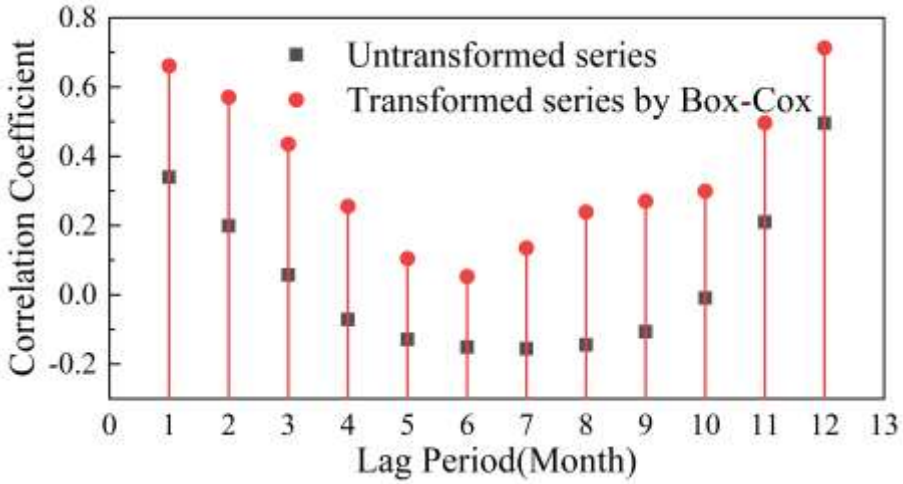
645 IMF4; (b) IMF5-IMF8.



646

647 Fig.7 Probability density of the candidate input variables of the original runoff time series before and

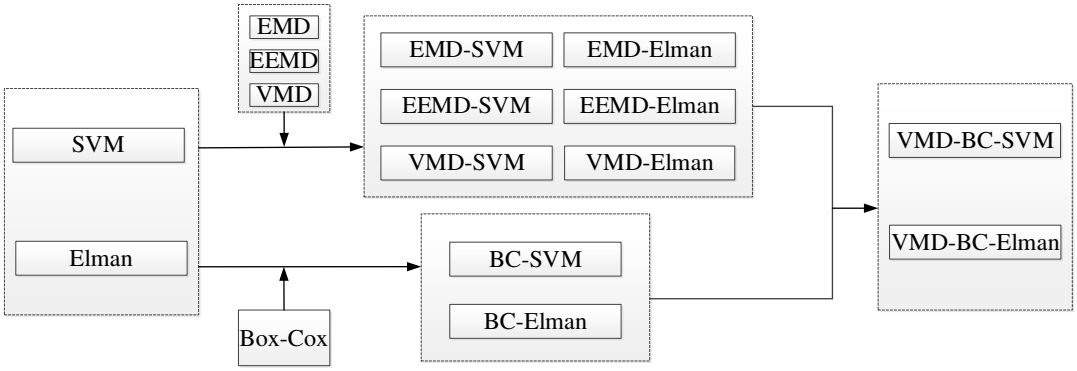
648 after Box-Cox transformation at Zhangjiashan Station.



649

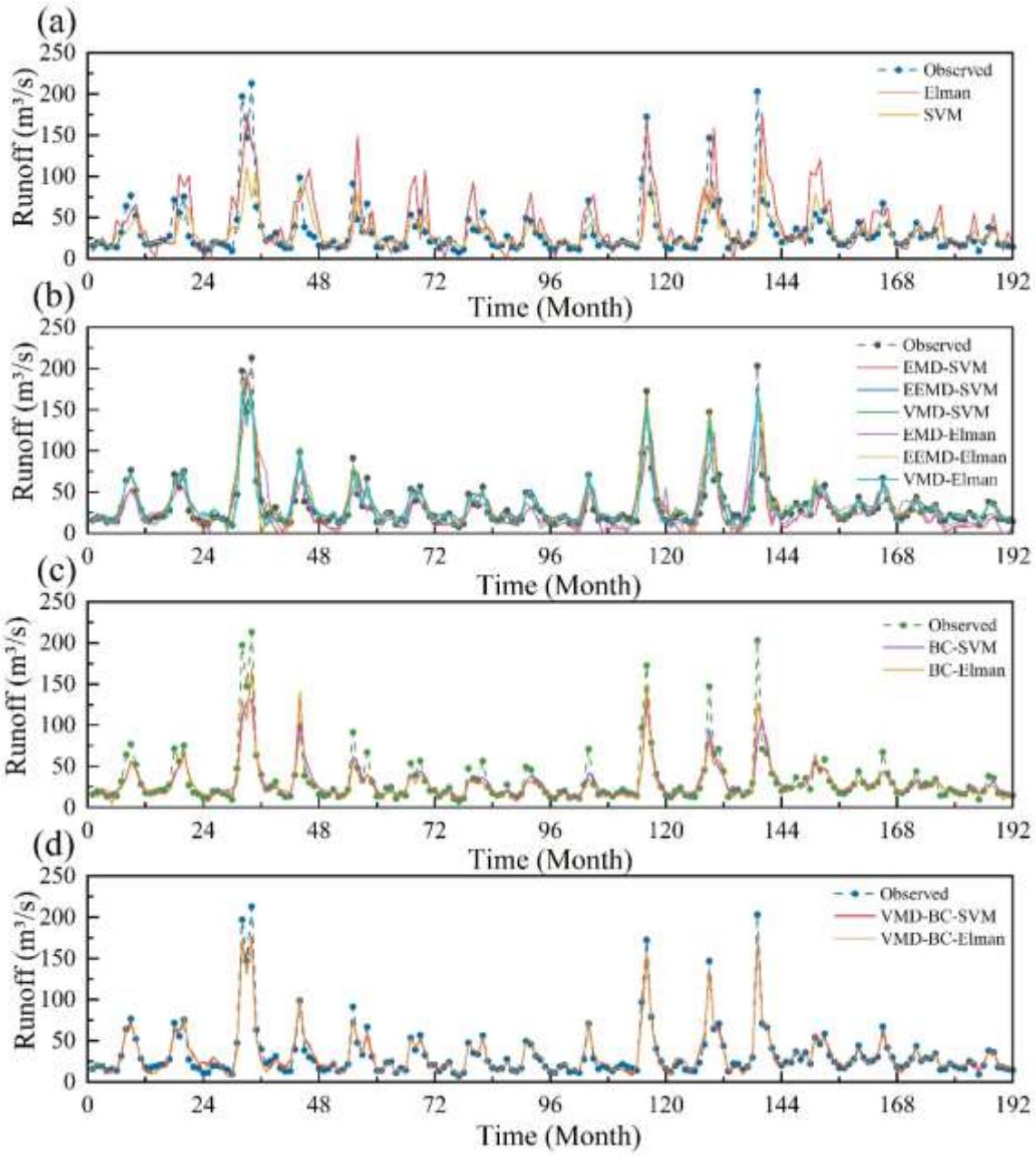
650 Fig.8 The correlation coefficients between the output and input variables before and after Box-Cox

651 transformation at Zhangjiashan Station.



652

653 Fig.9 The flowchart of modeling comparison.

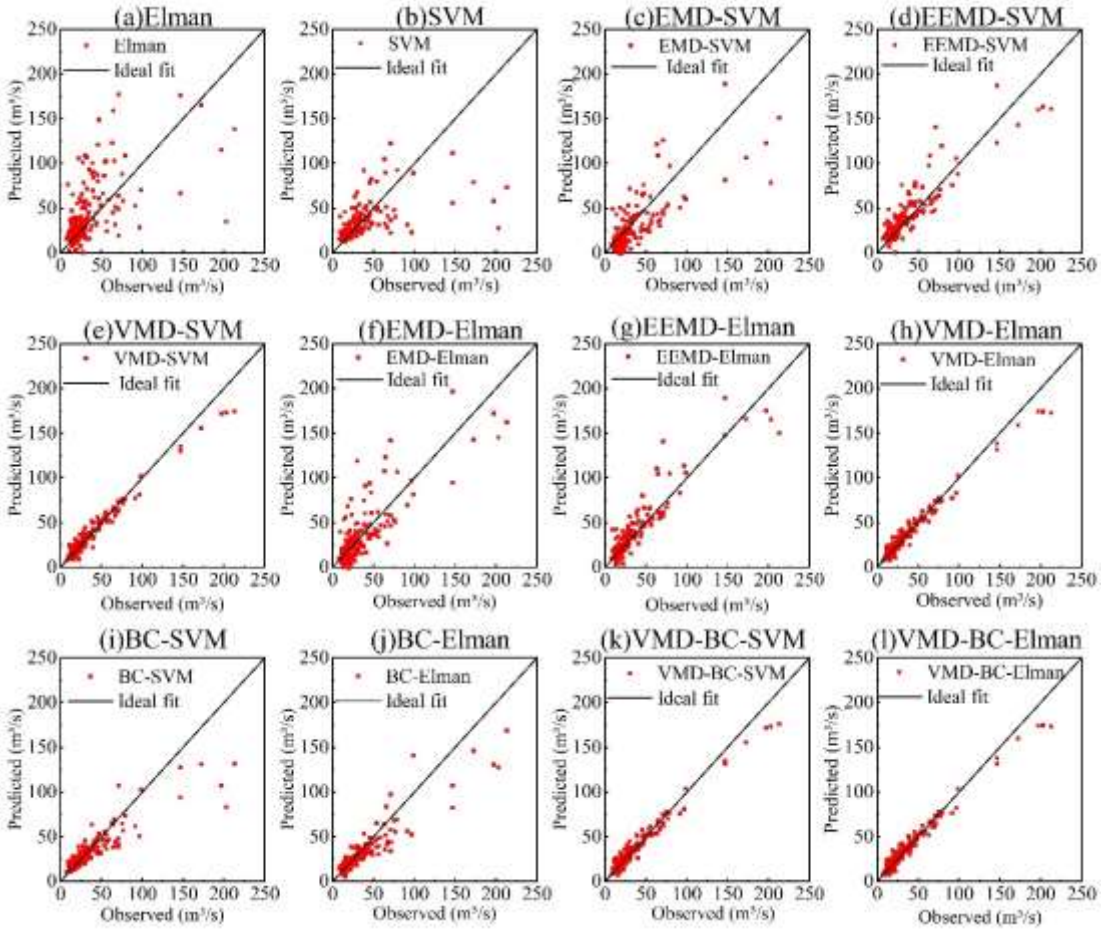


654

655 Fig.10 The observed *vs.* predicted runoff in the testing period generated at Zhangjiashan station. (a)

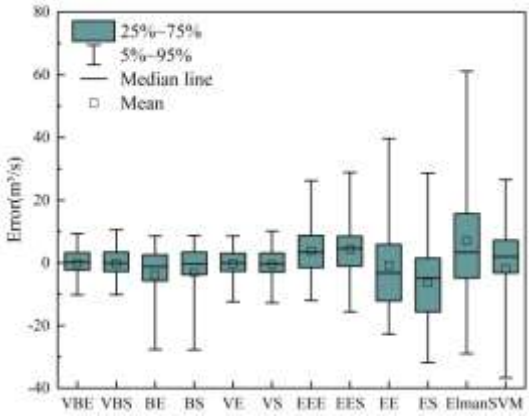
656 Single models; (b) The decomposition-based hybrid models; (c) The Box-Cox transformation-based

657 hybrid models; (d) VMD-BC-SVM and VMD-BC-Elman.



658

659 Fig.11 Scatter plots of the predicted *vs.* observed runoff in testing period at Zhangjiashan station.



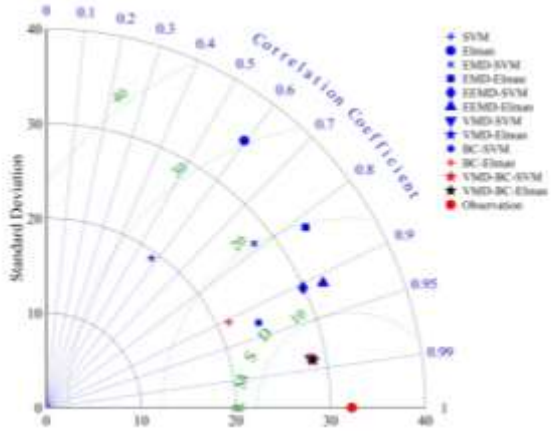
660

661 Fig.12 Boxplots of the predicted errors in the testing period at Zhangjiashan station (VMD-BC-

662 Elman, VMD-BC-SVM, BC-Elman, BC-SVM, VMD-Elman, VMD-SVM, VMD-Elman, EEMD-

663 Elman, EEMD-SVM, EMD-Elman, EMD-SVM are abbreviated as VBE, VBS, BE, BS, VE, VS,

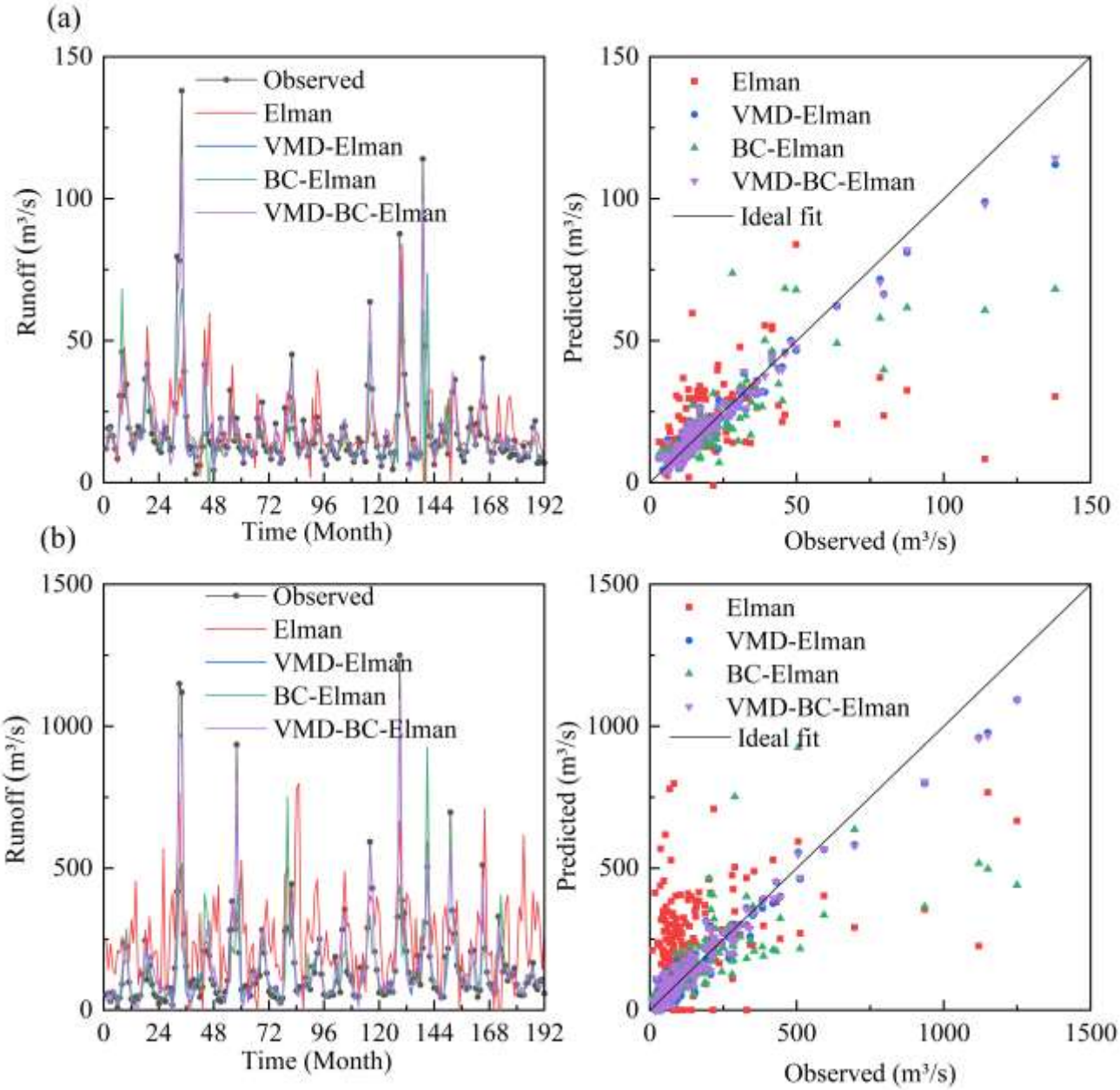
664 EEE, EES, EE, ES).



665

666 Fig.13 Taylor diagram depicting the predictive ability of 12 models in the testing period at

667 Zhangjiashan station.



668

669 Fig.14 The observed *vs.* predicted runoff in the testing period. (a) Zhuangtou station; (b) Huaxian

670 station.