A Hybrid Model Integrating Elman Neural Network with Variational Mode Decomposition and Box–Cox Transformation for Monthly Runoff Time Series Prediction

Fangqin Zhang1,2 · Yan Kang1,2 · Xiao Cheng1 · Peiru Chen1 · Songbai Song1,2

Received: 24 August 2021 / Accepted: 7 June 2022 / Published online: 23 June 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
Precise and reliable monthly runoff prediction plays a vital role in the optimal management of water resources, but the nonstationarity 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 by introducing variational mode decomposition (VMD) and Box–Cox transformation (BC) into the Elman neural network (Elman), named the VMD-BC-Elman model. First, the observed runoff is decomposed into sub-time series using VMD for better frequency resolution. Second, the input datasets are transformed into a normal distribution using Box–Cox, and as a result, skewedness in the data is removed, and the correlation between the input and output variables is enhanced. The proposed VMD-BC preprocessing technology is expected to overcome the problems arising from nonstationary and skewed runoff data. 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 the 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 the VMD-BC-Elman model performs best in all considered hybrid models with an NSE greater than 0.95, R greater than 0.98, NMSE less than 4.7%, and PBIAS less than 0.4% in both the training and testing periods. The study indicates that the VMD-BC-Elman model is a satisfactory data-driven approach to predict nonstationary and skewed monthly runoff time series, representing an effective tool for predicting monthly runoff series.

Keywords Monthly runoff prediction · Nonstationary and skewed runoff time series · Variational mode decomposition · Box–Cox transformation · Elman neural network · Wei River Basin

* Yan Kang kangyan@nwsuaf.edu.cn

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
1 Introduction

Accurate and reliable runoff prediction is of great significance for water resources planning and management, flood control and drought warning, which has attracted extensive attention as a vital and difficult topic in hydrology modeling (Peng et al. 2011). However, runoff time series show complex nonlinearity, nonstationarity and skewness (Barge and Sharif 2016), which are attributable to climate changes and frequent human activities, such as the variations in precipitation, conservancy projects, and urbanization. As a result, it is a great challenge to accurately simulate the intrinsic dynamic process of 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 rainfall-runoff processes with complex mathematical formulas, accurate meteorology and hydrological knowledge (Troin and Caya 2014). However, there have been key modeling controversies about issues such as the 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 to poor performance of process-driven models. In contrast, data-driven approaches, which are based on hydrological statistics and machine learning, can achieve satisfactory performance without too much information input. Data-driven models can be categorized into three main methods: statistical methods, artificial intelligence methods (AI) and hybrid models (Myronidis et al. 2018; Kuremoto et al. 2014; Tiwari and Adamowaki 2013). Statistical models, such as autoregressive models (AR) (Sarlak 2008), autoregressive moving average models (ARMA) (Pan et al. 2021) and so on, have been widely applied to capture the stationary and linear relationships between variables; however, they are not suitable for modeling nonstationary and nonlinear 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 hydrology prediction. Compared to conventional regression models, AI models have greatly improved prediction performance for highly nonlinear runoff time series. As a typical recurrent neural network model, the Elman neural network (Elman), equipped with a time delay operator, has short-term memory ability and is very suitable for time series forecasting, and parameter selection is simple (Krishnan et al. 2019). Support vector machine (SVM), based on the 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 nonstationary features of the input data, resulting in problems that the Elman model is inclined to fall into a local optimum, overfitting for, and the SVM model strongly depends on the parameter selection (Wang et al. 2013). Therefore, a hybrid model that combines AI methods with data processing techniques has become the most popular option for predicting nonstationary runoff time series (Wen et al. 2019).

In recent years, signal decomposition algorithms, used as data processing techniques, have been widely developed to effectively extract the multiple frequency information hidden from complex runoff series, and the 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 strong adaptability due to a self-adaptive
A Hybrid Model Integrating Elman Neural Network with Variational…

data-driven tool, but it lacks bases for rigorous mathematical theories and is prone to edge effects and ensuing modal aliasing (Huang et al. 2003; Sankaran and Reddy 2016). As an improved EMD model, the EEMD algorithm is proposed by adding white noise into a series of intrinsic mode functions (IMFs), which eliminates effective mode aliasing (Wu and Huang 2009). However, the amplitude of the added white noise in the EEMD model is determined by expert experience, and the modal components cannot be controlled adaptively, which may lead to information loss 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 nonstationary series into a group of relatively stable sub-series (Dragomiretskiy and Zosso 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 satisfactory performance for signal decomposition (Deb et al. 2020; Mohanty et al. 2018; Li et al. 2020).

The above hybrid models can process the nonstationary characteristics of runoff time series using decomposition, but the influence of runoff skewness on forecasting performance has rarely been explored by researchers. Normalization transformation techniques of data processing are worthy of more attention. Box–Cox transformation can remove the skewness of runoff data and transform the data into a more normal distribution, which significantly stabilizes the variance in the runoff time series, improves the normality and linearity of the data series, reduces the probability of pseudoregression, and effectively improves the correlation between the input and output variables. Box–Cox has achieved good normality in hydrology frequency analysis and calculation (Vasiliades and Loukas 2009; Seong 2014), but it has rarely been employed in runoff prediction models.

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

2 Methodologies

2.1 Variational Mode Decomposition

Variational mode decomposition (VMD), a novel nonrecursive signal processing algorithm proposed by Dragomiretskiy and Zosso, is utilized to adaptively decompose a complex nonstationary signal into a set of discrete bandwidth-limited modes in a spectral domain by Wiener filtering (Dragomiretskiy and Zosso 2014). The original runoff time series \( f(t) (t = 1, 2, \ldots, n) \), regarded as a nonstationary signal, is decomposed into \( K \) different modal functions \( \mu_k(t) \) \( (k = 1, 2, \ldots, K) \) with central frequency \( \omega_k \). The bandwidth and central frequency of each mode can be estimated by seeking the optimums of the constrained variational modes. The objective function is that the sum of the estimated...
bandwidths of all subseries is minimized, while the constraint satisfies that the sum of each decomposed mode is equal to the original signal, which can be expressed as:

$$\min_{\{\mu_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes \mu_k(t) \right] e^{-j\omega_k t} \right\|^2 \right\}$$

$$\text{s.t. } \sum_{k=1}^{K} \mu_k = f(t)$$

where \( \{\mu_k\} \) is the set of modes, \( \{\omega_k\} \) is the set of central frequencies of modes, \( \delta \) 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 a Lagrange multiplier and quadratic penalty to construct the augmented Lagrange term \( L \):

$$L(\{\mu_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) \otimes \mu_k(t) \right] e^{-j\omega_k t} \right\|^2$$

$$+ \left\| f(t) - \sum_{k=1}^{K} \mu_k(t) \right\|^2 + \left\langle \lambda(t)f(t) - \sum_{k=1}^{K} \mu_k(t) \right\rangle$$

where \( \alpha \) and \( \lambda \) 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 \( \mu_k \) and \( \omega_k \) are presented as follows:

$$\begin{align*}
\hat{\mu}_{k}^{n+1}(\omega) &= \frac{\hat{f}(\omega) - \sum \hat{\nu}_{i}(\omega) + \frac{2\nu(\omega)}{1+2\alpha(\omega-\omega_{k})}}{1+2\alpha(\omega-\omega_{k})} \\
\omega_{k}^{n+1} &= \frac{\int_{0}^{\pi} \hat{\mu}_{k}^{n+1}(\omega) d\omega}{\int_{0}^{\pi} \left| \hat{\mu}_{k}^{n+1}(\omega) \right|^2 d\omega} \\
\hat{\lambda}^{n+1}(\omega) &= \hat{\lambda}^{n}(\omega) + \tau \left( \hat{f}(\omega) - \sum \hat{\mu}_{k}^{n+1}(\omega) \right)
\end{align*}$$

where \( n \) is the number of iterations, \( \tau \) is the update parameters of Lagrange multipliers, and \( \hat{\mu}_{k}^{n+1}(\omega) \), \( \hat{f}(\omega) \), \( \hat{\nu}_{i}(\omega) \), and \( \hat{\lambda}^{n+1}(\omega) \) represent the fast Fourier transforms of \( \mu_{k}^{n+1}(t) \), \( f(t) \), \( \nu_{i}(t) \), 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 in time series and convert them into a more normal-like distribution (Box and Cox 1964; Hamasaki and Kim 2006). Assuming \( y_i \) \((i = 1, 2, \ldots, n)\) is a positive random variable, the Box–Cox transformation with parameter \( \lambda \) is defined as:
where $y_i$ is the original signal ($y_i > 0$), $y_i^{(\alpha)}$ is the transformed signal, and $\lambda$ denotes the transform coefficient.

To estimate the parameters of the model, the transformed maximum likelihood function is adopted as follows (Hamasaki and Kim 2006):

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

(5)

where $L$ denotes the likelihood function and $\sigma^2$ is the variance in $y_i^{(\alpha)}$.

Finally, the value of $\lambda$ can be obtained by employing the optimization algorithm to solve 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: an input layer, a hidden layer, a context layer, and an 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.

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

$$O_t = g(\omega_3 H_t + b_1)$$

(6)

$$H_t = f(\omega_1 C_t + \omega_2 I_t + b_2)$$

(7)

$$C_t = H_{t-1}$$

(8)
where $I_{t}$, $H_{t}$, $C_{t}$ and $O(t)$ denote the input, hidden, context and output layer vectors, respectively. $\omega_{1}$, $\omega_{2}$ and $\omega_{3}$ are weights connecting the 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 abovementioned methods, the hybrid model is proposed and diagrammed in Fig. 2. The modeling procedures are described as follows:

**Step 1:** The original runoff data are decomposed into $k$ sub-time series that are bandwidth-limited with a center frequency using VMD.

**Step 2:** Each sub-time series from step 1 is reconstructed using phase space reconstruction (PSR) to generate candidate input variable series for predicting models, and each reconstructed candidate input variable series is normalized by Box–Cox.

**Step 3:** For each candidate input variable series transformed by Box–Cox from step 2, LASSO is employed to rank and select the optimal number of input variables for the Elman model.

**Step 4:** Predicting by Elman

**Step 5:** Summing up

![Fig. 2 Architecture of the proposed VMD-BC-Elman model (CIV = candidate input variables; IV = input variables)](image-url)
Step 4: Each sub-time series decomposed by VMD is divided into a training dataset and a testing dataset. The parameters can be calibrated by inputting the selected input variables into the Elman model, and then the trained model is tested using the testing dataset.

Step 5: The final estimates of runoff time series are obtained by summing up the sub-time series estimated from the predicting models.

2.5 Model Performance Evaluation

The Nash–Sutcliffe 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 follows:

\[
NSE = 1 - \frac{\sum_{i=1}^{N} (Q_o(t) - Q_p(t))^2}{\sum_{i=1}^{N} (Q_o(t) - \overline{Q_o})^2}
\]

\[
R = \frac{\sum_{i=1}^{N} (Q_o(t) - \overline{Q_o})(Q_p(t) - \overline{Q_p})}{\sqrt{\sum_{i=1}^{N} (Q_o(t) - \overline{Q_o})^2 \left(Q_p(t) - \overline{Q_p}\right)^2}}
\]

\[
NMSE = \frac{\sum_{i=1}^{N} (Q_o(t) - Q_p(t))^2}{\sum_{i=1}^{N} (Q_o(t))^2} \times 100\%
\]

\[
PBIAS = \frac{\sum_{i=1}^{N} (Q_o(t) - Q_p(t))}{\sum_{i=1}^{N} Q_o(t)} \times 100\%
\]

where \(Q_o(t)\) and \(Q_p(t)\) are the observed and predicted runoff, respectively, and \(\overline{Q_o}\) and \(\overline{Q_p}\) denote the mean of the observed and predicted monthly runoff time series, respectively. \(N\) is the number of \(Q_o(t)\).

3 Study Area and Data Description

The Wei River originates from Weiyuan County in Gansu Province, extends more than 800 km and finally converges into the Yellow River (Jiang et al. 2019). The Wei River Basin (WRB) covers an area of approximately \(13.5 \times 10^4\) 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 3,929 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 semiarid regions in China, affected by a continental monsoon, and the climate is warm and rainy in summer and cold and dry in winter. Precipitation and runoff, exhibiting extremely high intra-annual and interannual 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 runoff has been disturbed by human activities. Thus, the WRB is a suitable candidate to verify the feasibility of the proposed model for predicting nonstationary runoff time series disturbed by climate change and human activities.

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 the Jing River, Beiluo River and Wei River, respectively, and data were 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 are first employed to develop the hybrid model, and data from Zhuangtou and Huaxian stations are utilized to further validate the performance of the model.

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

The monthly runoff hydrographs are plotted in Fig. 4, which illustrates drastic intra- and interannual variability and high nonstationary. The cumulative departure curve in Fig. 5 was used to identify variability points of the runoff time series. 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 dataset 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 the variability of the runoff time series during the training period.
Table 1 Profiles of hydrological stations and statistical characteristics of the runoff series at Zhangjiashan, Zhuangtou and Huaxian stations

| Station  | Longitude | Latitude | Control Area (km²) | Data Sets | Data Length | Max (m³/s) | Min (m³/s) | Mean (m³/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|

Max, Min, Mean, Cv and Cs represent maximum, minimum, mean value, coefficient of variation and coefficient of skewness of data sets, respectively.
4 Results and Discussion

4.1 Decomposition Results with VMD

\( K \) and \( \alpha \), subject to the constraints of the input runoff time series, are two key parameters of the VMD algorithm. To achieve satisfactory performance, the center frequency iteration is run to determine the number of 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 that the decomposition has been completed and \( K = 8 \) is a determined value, which is consistent with an earlier study (Huang et al. 2016). According to the multiple preexperiment, \( \alpha = 1,000 \) is determined. The default parameters will be adopted for other parameters of the VMD model, as shown in Table 2.

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

4.2 Reconstruction and Transformation by PSR and Box–Cox

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 over 12 months, when PSR dimension \( d = 12 \) and delay time \( \tau = 1 \) (Packard et al. 1980), the \( t \)-th space vector in the reconstructed phase space is expressed as

\[
X_t = \{x_{t-1}, x_{t-2}, \ldots, x_{t-12}\}, \quad t = 13, 14, \ldots, n,
\]

where \( n \) is the length of the original monthly runoff series, and \( x_{t-1}, x_{t-2}, \ldots, x_{t-12} \), respectively, denote 12 different dimensional series, namely, 12 candidate input variable series.

To avoid the unsatisfactory prediction performance arising from the skewness of runoff data, each candidate input variable series is converted into a normal distribution using Box–Cox. The optimal transformation coefficient \( \lambda \) is estimated by the maximum likelihood method, and the estimated \( \lambda \) values of each candidate input variable series for runoff time series are shown in Table 3.

Taking 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 that the candidate input variable series transformed by Box–Cox exhibits a normal distribution.

| Table 2 Parameters of VMD determined in the iteration experiments |
|---------------------------------------------------------------|
| Parameters | \( \alpha \) | \( \tau \) | \( K \) | DC | Init | tol |
| Values     | 1000 | 0 | 8 | 0 | 1 | \( 10^{-7} \) |
Figure 8 is used to analyze the correlation between the output variables and the 12 candidate input variables transformed by Box–Cox. The correlation coefficients of the transformed data have been improved, indicating that Box–Cox strengthens the correlation of the data.

4.3 Input Variable Selection

Least absolute shrinkage and selection operator (LASSO) is a kind of variable selection method that performs well for handling multicollinearity issues. LASSO selects significant input variables by adjusting the value of tuning parameter $\rho$ from large to small and controlling the number of nonzero coefficients of candidate input variables (Milani et al. 2016). The importance ranking and the optimal number of selected input variables are shown in Table 4. Taking the original runoff series as an example, the optimal number of selected input variables is 4 using the trial-and-error method; thus, the input variables are $x_{t-1}$, $x_{t-12}$, $x_{t-11}$, $x_{t-2}$.

4.4 Development and Training of the Forecasting Models

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 is described in this section.

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 datasets with a maximum number of training iterations of 500, a target error of 0.001, and a learning coefficient of 0.01. The hidden layer activation function employs Tansig, the output layer activation function uses Purelin, and the training algorithm adopts the L-M algorithm. The number of each layer is determined using the trial-and-error method, and the trained Elman architecture is presented in Table 5.
Table 3  The $\lambda$ values of each candidate input variables series for the original runoff series and each 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    | 0.2111    | 0.2109    | 0.2107    | 0.2108    | 0.2114    | 0.2126    | 0.2126    | 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    |
Group 2: Decomposition-based hybrid models

Decomposition-based hybrid models are developed by coupling EMD, EEMD and VMD with Elman or SVM. The original runoff series are decomposed into 6, 11 and 8 sub-time series by EMD, EEMD and VMD, respectively. Each sub-time series is modeled using the Elman or SVM. In Elman training, the target error is set at 0.00001, and the remaining parameters and the activation functions are constants as the single model has in Group 1. The three-layer architecture of each sub-time series is determined by

the SVM model, the radial basis kernel function is selected with an insensitivity loss parameter of 0.01, a penalty factor of 1 and a kernel function parameter of 2.

Fig. 7 Probability density of the candidate input variables of the original runoff time series before and after Box-Cox transformation at Zhangjiashan Station
the trial-and-error method, and the optimal number of inputs, with hidden and output neurons listed in Table 5. For SVM training, the number of input variables of each sub-time series is selected in Table 4, and the other parameters are the same as those of the standard SVM.

Group 3: Box–Cox-based hybrid models

To reduce the negative effect of the skewed runoff data on the modeling, BC-based models are developed by introducing Box–Cox into Elman or SVM. The input variable series are transformed into a normal distribution using Box–Cox with the parameters $\lambda$ in Table 3. The model architecture and parameters are the same as those of the single models.

Table 4 Importance ranking of the selected input variables for original series and sub-time series decomposed at Zhangjiashan Station

| No. | Series | Raking the selected input variables | Number of input variables |
|-----|--------|-------------------------------------|---------------------------|
| 1   | Original | $x_{t-1}, x_{t-2}, x_{t-11}, x_{t-2}, x_{t-10}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-9}, x_{t-6}, x_{t-8}, x_{t-7}$ | 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   | $x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 7 |
| 4   | IMF3   | $x_{t-22}, x_{t-12}, x_{t-11}, x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 |
| 5   | IMF4   | $x_{t-11}, x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 |
| 6   | IMF5   | $x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 |
| 7   | IMF6   | $x_{t-12}, x_{t-11}, x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 (EMD:10) |
| 8   | IMF7   | $x_{t-12}, x_{t-11}, x_{t-10}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-6}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 |
| 9   | IMF8   | $x_{t-7}, x_{t-8}, x_{t-9}, x_{t-12}, x_{t-10}, x_{t-11}, x_{t-5}, x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}$ | 8 |
Group 4: Hybrid models—VMD-BC-Elman and VMD-BC-SVM

To further improve the models’ performance and sufficiently overcome the problems arising from the nonstationary and skewed runoff data, the VMD-BC-Elman and VMD-BC-SVM are proposed by introducing Box–Cox and VMD into the Elman or SVM. In this section, the VMD-BC-Elman architecture of each sub-time series is the same as those of VMD-Elman, and the VMD-BC-SVM modeling parameters are the same as those of SVM.

4.5 Forecasting Results and Discussion for Zhangjiashan Station

Based on the training models above, runoff time series are modeled using the different models. The performance statistical metrics are listed in Table 6. The hybrid models are shown to be more accurate than their standalone counterparts, and the performance of VMD-BC-Elman is the best among all models.

| Series | Single model architecture of the original and each sub-time series based on different decomposition methods at Zhangjiashan Station |
|--------|------------------------------------------------------------------------------------------------------------|
| 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 8–13-1 |
| IMF8   | - 8–13-1 8–13-1 8–13-1 |
| IMF9   | - 8–13-1 8–13-1 8–13-1 |
| IMF10  | - 8–13-1 8–13-1 8–13-1 |
| IMF11  | - 8–13-1 8–13-1 8–13-1 |
In the standalone models, Elman and SVM have unsatisfactory prediction results with NSE < 0.6, lower R and higher NMSE and PBIAS, as shown in Table 6. Compared to SVM, Elman provides better general results in the training period because its recurrent feedback networks with short-term memory increase the ability to process dynamic information. However, the performance of Elman is unstable, which may lead to overfitting. The standalone SVM, which is based on the risk minimization principle, has more robustness and generalizability for estimating nonstationary and skewed time series than the Elman does. The Elman model provides better prediction results in the training period, and the SVM provides better prediction results in the testing period, which is consistent with other studies (Song et al. 2020; Feng et al. 2020).

Compared with the standalone models, the metric values of decomposition-based hybrid models are satisfactory, as shown in Table 6. The six models yield R values in the range of 0.7878 to 0.9843, NSE values from 0.5754 to 0.9592, NMSE values from 1.06% to 42.24%, and PBIAS values from 1.26% to 19.26% in the testing period; thus, the decomposition-based hybrid models are superior to the counterpart single models. The metric values of VMD-Elman are the best among the six models, which indicates that VMD-Elman outperforms the other five hybrid models and that VMD has a better ability to denoise runoff data than EMD and EEMD do, which agrees with the results of other researchers who have modeled runoff time series based on EMD, EEMD, and VMD decomposition (He et al. 2019).

To eliminate the skewness of runoff time series, BC-based models are introduced. Table 6 shows that the statistical metrics of the BC-SVM and BC-Elman models are much better than those of the single models. Taking the NMSE in the testing period as an example, the NMSEs for the BC-Elman, BC-SVM, Elman and SVM models are 18.3%, 24.84%, 92.91% and 66.93%, respectively. This illustrates that the performance of the BC-based hybrid models is superior to that of the single models, and BC-Elman is better than BC-SVM. The excellent performance of BC-based models in this paper is similar to the study on hydrology frequency analysis, which shows that Box–Cox can provide robustness in fitting models (Seong 2014).

| Types                        | Models          | Training Dataset | Testing Dataset |                |                |                |                |
|------------------------------|-----------------|------------------|-----------------|----------------|----------------|----------------|----------------|
|                              |                 |                  |                 | R   | NSE  | NMSE  | PBIAS | R   | NSE  | NMSE  | PBIAS |
| Original                     | SVM             | 0.600            | 0.323           | 0.676 | 0.184 | 0.576  | 0.327  | 0.670 | 0.058 |
|                              | Elman           | 0.740            | 0.547           | 0.453 | -0.005 | 0.595  | 0.066  | 0.929 | -0.212 |
| Decomposition                | EMD-SVM         | 0.774            | 0.576           | 0.424 | 0.096 | 0.788  | 0.575  | 0.422 | 0.193 |
|                              | EEMD-SVM        | 0.877            | 0.765           | 0.235 | -0.033 | 0.906  | 0.803  | 0.196 | -0.130 |
|                              | VMD-SVM         | 0.982            | 0.952           | 0.048 | 0.004 | 0.983  | 0.955  | 0.045 | 0.018 |
|                              | EMD-Elman       | 0.898            | 0.806           | 0.194 | -0.009 | 0.824  | 0.637  | 0.361 | 0.027 |
|                              | EEMD-Elman      | 0.964            | 0.930           | 0.070 | 0.000 | 0.916  | 0.820  | 0.179 | -0.114 |
|                              | VMD-Elman       | 0.984            | 0.958           | 0.041 | 0.001 | 0.984  | 0.959  | 0.041 | 0.012 |
| Box-Cox transformation       | BC-SVM          | 0.867            | 0.733           | 0.267 | 0.095 | 0.906  | 0.750  | 0.248 | 0.074 |
|                              | BC-Elman        | 0.941            | 0.885           | 0.115 | -0.028 | 0.929  | 0.816  | 0.183 | 0.113 |
| Decomposition and Box-Cox    | VMD-BC-SVM      | 0.983            | 0.954           | 0.046 | 0.002 | 0.983  | 0.956  | 0.044 | 0.007 |
|                              | VMD-BC-Elman    | 0.984            | 0.960           | 0.040 | 0.001 | 0.985  | 0.960  | 0.040 | 0.004 |
The statistical metrics of VMD-BC-Elman and VMD-BC-SVM are shown in Table 6. In the testing period, the R values of the VMD-BC-Elman model are 0.04%, 6.05%, and 0.14% greater than those of the VMD-Elman, BC-Elman and VMD-BC-SVM models; its NSE is 0.08%, 17.65%, and 0.39% greater, respectively; its NMSE is 1.97%, 78.25%, and 8.50% smaller, and its PBIA is 68.29%, 96.54%, and 46.58% smaller, respectively. The statistical metric values demonstrate that although VMD or Box–Cox alone can improve the ability of a single Elman, the integration of VMD and Box–Cox into Elman can promote their prediction to a new level, with VMD-BC-Elman outperforming VMD-BC-SVM. Considering these results, it is evident that the proposed VMD-BC-Elman has a more stable and consistent performance. It is affirmed that the AI models, combining VMD and Box–Cox, represent a significant improvement in runoff prediction.

While the statistical metrics can evaluate the overall performance of models, hydrographs and scatter plots can further identify the temporal correspondence. Figures 10 and 11 display the hydrographs and scatter plots of the observed and predicted runoff from the different models in the testing periods. In Figs. 10a and 11a–b, the single models cannot 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 Figs. 10b–c and 11c–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 Figs. 10d and 11k–l, it is also obvious that VMD-BC-Elman and VMD-BC-SVM have greater accuracy as the scatter plots are concentrated closer to the ideal fitting line, illustrating that the VMD-BC-based hybrid models are more effective in estimating the peak values of data than other models. In particular, VMD-BC-Elman has the best capability to capture information on the overall features of runoff time series. Therefore, it is also reaffirmed that VMD and Box–Cox can effectively improve the stability and consistency of the models, and VMD-BC-Elman exhibits the strongest performance.

In Fig. 12, the boxplot diagram is designed to depict the error distribution between the estimated and observed values in the entire tested dataset 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 and minimum and maximum nonoutliers, are smallest when VMD-BC-Elman is employed compared with the compared models. This indicates that VMD and Box–Cox enhance the robustness of the models, and VMD-BC-Elman is superior to the other models, which is consistent with the above findings.

The Taylor diagram shown in Fig. 13, integrating multiple characteristics of models into a compact plot (Taylor 2001), can present the performance of the models comprehensively. It shows how close or far away the simulated values of each model are to or from the observed runoff data in the testing period, quantified via the correlation coefficient, standard deviation and root mean square error in the 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 lowest SD, followed by the VMD-BC-based and the VMD-based models, while the single models are positioned the furthest from the reference point. Obviously, VMD-BC-Elman performs best for monthly runoff prediction, which can be explained by the fact that the provision of VMD and Box–Cox contributes to the improved stability and consistency of the models and avoids overfitting.
Fig. 10  The observed vs. predicted runoff in the testing period generated at Zhangjiashan station. (a) Single models; (b) The decomposition-based hybrid models; (c) The Box-Cox transformation-based hybrid models; (d) VMD-BC-SVM and VMD-BC-Elman
4.6 Forecasting Results and Discussion for Zhuangtou and Huaxian Stations

To further verify the feasibility of the VMD-BC-Elman, the runoff at the Zhuangtou and Huaxian stations is modeled. In Tables 7 and 8, the hybrid models are better than the single models in the testing period. VMD-based models are generally more accurate and stable.
than EMD-based and EEMD-based models. Additionally, the BC-based models outperform the single models. Obviously, VMD-BC-Elman and VMD-BC-SVM outperform the VMD-based and BC-based models, and VMD-BC-Elman is superior to VMD-BC-SVM, which is similar to the results of Zhangjiashan station, so it is safe to conclude that VMD-BC-Elman has the best forecast performance in this paper.

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

**Table 7** As in Table 6, but for Zhuangtou station

| Types                      | Model     | Training dataset |                | Testing Dataset |        |
|----------------------------|-----------|-----------------|----------------|-----------------|-------|
|                            |           |                 | $R$  | NSE  | NMSE | PBIAS | $R$  | NSE  | NMSE | PBIAS |
| Original                   | SVM       |                 | 0.569 | 0.289 | 0.710 | 0.153 | 0.493 | 0.230 | 0.766 | 0.081 |
|                            | Elman     |                 | 0.703 | 0.494 | 0.505 | 0.002 | 0.366 | 0.059 | 0.936 | -0.039 |
| Decomposition              | EMD-SVM   |                 | 0.776 | 0.586 | 0.413 | 0.077 | 0.767 | 0.555 | 0.443 | -0.157 |
|                            | EEMD-SVM  |                 | 0.896 | 0.801 | 0.199 | 0.026 | 0.922 | 0.842 | 0.157 | -0.070 |
|                            | VMD-SVM   |                 | 0.981 | 0.950 | 0.050 | 0.000 | 0.980 | 0.945 | 0.055 | -0.018 |
|                            | EMD-Elman |                 | 0.915 | 0.837 | 0.163 | -0.008 | 0.789 | 0.451 | 0.546 | 0.348 |
|                            | EEMD-Elman|                 | 0.958 | 0.917 | 0.083 | 0.012 | 0.920 | 0.835 | 0.164 | -0.031 |
|                            | VMD-Elman |                 | 0.983 | 0.956 | 0.044 | -0.001 | 0.980 | 0.948 | 0.052 | 0.002 |
| Box-Cox transformation     | BC-SVM    |                 | 0.800 | 0.620 | 0.379 | 0.005 | 0.784 | 0.576 | 0.422 | 0.075 |
|                            | BC-Elman  |                 | 0.930 | 0.866 | 0.134 | 0.000 | 0.803 | 0.629 | 0.370 | 0.093 |
| Decomposition and Box-Cox | VMD-BC-SVM|                 | 0.982 | 0.952 | 0.049 | 0.000 | 0.979 | 0.943 | 0.057 | -0.018 |
|                            | VMD-BC-Elman|               | 0.983 | 0.957 | 0.044 | -0.001 | 0.983 | 0.955 | 0.045 | 0.004 |
Table 8  As in Table 6, but for Huaxian station

| Types                      | Model       | Training Dataset | Testing Dataset |
|----------------------------|-------------|------------------|-----------------|
|                            |             | $R$    | NSE   | NMSE  | PBIAS | $R$    | NSE   | NMSE  | PBIAS |
| Original                   | SVM         | 0.648 | 0.358 | 0.641 | 0.225 | 0.534 | 0.255 | 0.741 | 0.193 |
|                            | Elman       | 0.726 | 0.526 | 0.473 | -0.012| 0.360 | -0.359| 1.352 | -0.578|
| Decomposition              | EMD-SVM     | 0.846 | 0.696 | 0.304 | 0.081 | 0.782 | 0.550 | 0.403 | 0.083 |
|                            | EEMD-SVM    | 0.916 | 0.835 | 0.165 | 0.043 | 0.897 | 0.803 | 0.196 | 0.031 |
|                            | VMD-SVM     | 0.982 | 0.956 | 0.044 | 0.005 | 0.978 | 0.944 | 0.056 | 0.006 |
|                            | EMD-Elman   | 0.949 | 0.899 | 0.101 | -0.035| 0.686 | 0.418 | 0.579 | 0.102 |
|                            | EEMD-Elman  | 0.969 | 0.939 | 0.061 | 0.025 | 0.834 | 0.681 | 0.317 | 0.102 |
|                            | VMD-Elman   | 0.985 | 0.963 | 0.037 | 0.000 | 0.979 | 0.949 | 0.051 | 0.002 |
| Box-Cox transformation     | BC-SVM      | 0.841 | 0.685 | 0.315 | 0.133 | 0.801 | 0.599 | 0.399 | 0.101 |
|                            | BC-Elman    | 0.965 | 0.931 | 0.069 | 0.000 | 0.727 | 0.520 | 0.477 | 0.102 |
| Decomposition and Box-Cox  | VMD-BC-SVM  | 0.984 | 0.958 | 0.042 | 0.003 | 0.980 | 0.947 | 0.052 | 0.002 |
|                            | VMD-BC-Elman| 0.985 | 0.963 | 0.037 | 0.000 | 0.981 | 0.953 | 0.047 | -0.002|

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

To estimate nonstationary and skewed monthly runoff series, a novel VMD-BC-Elman model was developed in this paper, and runoff data from Zhangjiashan, Zhuangtou and Huaxian stations in the Wei River Basin were used. The steps of this model are as follows. (1) The original nonstationary 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 variable series was normalized using Box–Cox. (3) LASSO was employed to determine the optimal number of input variables. (4) Each sub-time series was simulated by Elman. The final results are generated by summing the estimates of all the sub-time series.

The proposed VMD-BC-Elman exhibited the best precision, stability and consistency for predicting the nonstationary 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 nonstationary series and adaptably decompose the original series into the relative stationary sub-time series using VMD. (2) It can remove the skewness of the original runoff series using Box–Cox and convert the skewed candidate input variables into a normal distribution, which can enhance the correlation between the input and output variables and improve the mapping ability and self-learning ability of Elman. (3) Integrating decomposition-normalization-simulation-reconstruction techniques, 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 summarize the optimal estimates of sub-time series to achieve better prediction for the monthly runoff. Overall, VMD-BC-Elman represents an effective tool for forecasting nonstationary and skewed monthly runoff series.

Despite the remarkable performance of VMD-BC-Elman over comparative models, the study has limitations that create opportunities for further research. First, the prediction results of high IMF are unsatisfactory, possibly due to too much noise. Second, the proposed model only considers one-step prediction, and multistep prediction has not yet been discussed. Finally, the proposed model does not consider the physical mechanism of runoff formation. The following aspects are worthy of further research. First, the high- and low-IMFs are expected to develop different models to more effectively predict the different frequency characteristics. Second, it will be useful to explore a model utilizing a multistep prediction approach. Third, a physically-based hydrology model can be introduced into the current VMD-BC-Elman, so this data-driven model can combine the strength of the self-learning ability with an enhanced physical mechanism.

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.

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.

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.
Declarations

Ethics Approval  Not applicable.

Consent to Participate  Not applicable.

Consent to Publication  Not applicable.

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

References

Aggarwal SK, Goel A, Singh VP (2012) Stage and discharge forecasting by SVM and ANN techniques. Water Resour Manag 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

Bojiang PO, Yang TC, Pham QB (2020) Linking singular spectrum analysis and machine learning for monthly rainfall forecasting. Appl Sci Basel 10(09)

Box GEP, Cox DR (1964) An analysis of transformations. J Roy Stat Soc B26:211–252

Chandra R (2015) Competition and collaboration in cooperative coevolution of elman recurrent neural networks for time-series prediction. IEEE Trans Neural Netw Learn Syst 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. Hydrol Earth Syst Sci 21(07):3427–3440

Deb S, Dandapat S, Krajewski J (2020) Analysis and classification of cold speech using variational mode decomposition. IEEE Trans Affect Comput 11(02):296–307

Demirel MC, Venancio A, Kahya E (2009) Flow forecast by SWAT model and ANN in Pracana basin, Portugal. Adv Eng Softw 40(07):467–473

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

Feng ZK, Niu WJ, Tang ZY, Jiang ZQ, Xu Y, Liu Y, Zhang HR (2020) Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. J Hydrol 583

Hamasaki T, Kim SY (2006) Box and Cox power-transformation to confined and censored non-normal responses in regression. Comput Stat Data Anal 51(8):3788–3799

He XX, Luo JG, Zuo GG, Xie JC (2019) Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks. Water Resour Manag 33(04):1571–1590

Himanshu SK, Pandey A, Yadav B (2017) Ensemble wavelet-support vector machine approach for prediction of suspended sediment load using hydrometeorological data. J Hydrol Eng 22(07)

Huang N, Chen H, Cai G, Fang L, Wang Y (2016) Mechanical fault diagnosis of high voltage circuit breakers based on variational mode decomposition and multi-layer classifier. Sensors 16(11)

Huang NE, Wu MLC, Long SR, Shen SSP, Qu WD, Gloersen P, Fan KL (2003) A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proc R Soc Math Phys Eng Sci 459(08):2317–2345

Jiang RG, Wang YP, Xie JC, Zhao Y, Li FW, Wang XJ (2019) Assessment of extreme precipitation events and their teleconnections to El Nino southern oscillation, a case study in the Wei River basin of China. Atmos Res 218:372–384

Karamouz M, Ahmadi A, Moridi A (2009) Probabilistic reservoir operation using Bayesian stochastic model and support vector machine. Adv Water Resour 32(11):1588–1600

Krishnan S, Lokesh S, Devi MR (2019) An efficient Elman neural network classifier with cloud supported internet of things structure for health monitoring system. Comput Netw 14(5):201–210

Kuremoto T, Kimura S, Kobayashi K, Obayashi 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
A Hybrid Model Integrating Elman Neural Network with Variational...

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. J Energy Stor 21:510–518

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

Milani MRM, Hense A, Rahmani E, Ploeger A (2016) Applying least absolute shrinkage selection operator and akaike information criterion analysis to find the best multiple linear regression models between climate indices and components of Cow’s Milk. Foods 5(4):52. https://doi.org/10.3390/foods5030052

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 Resour Manag 32(5):1759–1776

Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 10(3):282–290

Packard NH, Crutchfield JP, Farmer JD et al (1980) Geometry from a time series. Phys Rev Lett 45:712

Pan B, Han M, Wei F, Tian LX, Liu YT, Li YL, Wang M (2021) Analysis of the variation characteristics of runoff and sediment in the yellow river within 70 years. Water Resour 48(05):676–689

Peng Y, Wang GL, Tang GL, Zhou HC, Wang YJ, Jian DP (2011) Study on reservoir operation optimization of Ertan hydropower Station considering GFS forecasted precipitation. Sci China-Technol Sci 54:76–82

Sankaran A, Reddy MJ (2016) Analyzing the hydroclimatic teleconnections of summer monsoon rainfall in Kerala, India, using multivariate empirical mode decomposition and time-dependent intrinsic correlation. IEEE Geosci Remote Sens Lett 13(9):1221–1225

Saridak N (2008) Annual streamflow modelling with asymmetric distribution function. Hydrol Process 22(17):3403–3409

Sedki A, Quazar D, EI Mazoudi E (2008) Evolving neural network using real coded genetic algorithm for daily rainfall-runoff forecasting. Expert Syst Appl 36:4523–4527

Seong KW (2014) Deriving a practical form of IDF formula using transformed rainfall intensities. Hydro Process 28(06):2881–2896

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. J Gerontol Ser A Biol Med Sci 106(D7):7183–7192

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

Troin M, Caya D (2014) Evaluating the SWAT’s snow hydrology over a Northern Quebec watershed. Hydro Process 28(04):1858–1873

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. J Hydroinf 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. J Hydrol 570:167–184

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

Yu YH, Zhang HB, Singh VP (2018) Forward prediction of runoff data in data-scarce basins with an improved ensemble empirical mode decomposition (EEMD) Model. Water 10(04)

Zhang X, Peng Y, Zhang C, Wang B (2015) Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J Hydrol 530:137–152

Zou L, Xia J, She DX (2018) Analysis of impacts of climate change and human activities on hydrological drought: a case study in the Wei River basin, China. Water Resour Manag 32(4):1421–1438

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.