Streamflow Forecasting Using Empirical Wavelet Transform and Artificial Neural Networks

Tian Peng [1,2], Jianzhong Zhou [1,2,*], Chu Zhang [1,2] and Wenlong Fu [3]

[1] School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; husthydtopt@126.com (T.P.); zhangchuhust@foxmail.com (C.Z.)
[2] Hubei Key Laboratory of Digital Valley Science and Technology, Wuhan 430074, China
[3] School of Hydraulic and Environmental Engineering, Three Gorges University, Yichang 443002, China; husthydrofwl@126.com
[*] Correspondence: jz.zhou@mail.hust.edu.cn; Tel.: +86-027-8754-3127

Abstract: Accurate and reliable streamflow forecasting plays an important role in various aspects of water resources management such as reservoir scheduling and water supply. This paper shows the development of a novel hybrid model for streamflow forecasting and demonstrates its efficiency. In the proposed hybrid model for streamflow forecasting, the empirical wavelet transform (EWT) is firstly employed to eliminate the redundant noises from the original streamflow series. Secondly, the partial autocorrelation function (PACF) values are explored to identify the inputs for the artificial neural network (ANN) models. Thirdly, the weights and biases of the ANN architecture are tuned and optimized by the multi-verse optimizer (MVO) algorithm. Finally, the simulated streamflow is obtained using the well-trained MVO-ANN model. The proposed hybrid model has been applied to annual streamflow observations from four hydrological stations in the upper reaches of the Yangtze River, China. Parallel experiments using non-denoising models, the back propagation neural network (BPNN) and the ANN optimized by the particle swarm optimization algorithm (PSO-ANN) have been designed and conducted to compare with the proposed model. Results obtained from this study indicate that the proposed hybrid model can capture the nonlinear characteristics of the streamflow time series and thus provides more accurate forecasting results.

Keywords: empirical wavelet transform; multi-verse optimizer; artificial neural networks; annual streamflow forecasting; data-driven modeling

1. Introduction

Assessing and estimating water available in a basin over long periods (from months to years) is an important precondition for efficient reservoir scheduling and water supply. For this reason, constructing a forecasting model for long-term streamflow time series is essential [1]. At present, methods for long-term streamflow forecasting can be mainly grouped into two categories: physically-based models and data-driven models. Technically, physical-based models usually consider multi-factors including rainfall, runoff, evaporation, soil moisture, infiltration, snow cover and melt, etc., while data-driven models usually take advantage of the implicit information contained in the historical streamflow time series to forecast the future streamflow [2]. Physically-based models can suffer from the drawbacks that physical relationships between different hydrological features are complex, and sometimes the lack of hydrological data makes it difficult to complete physical modeling [3].

In the last few decades, numerous data-driven models for hydrological time series forecasting have been proposed to increase forecasting accuracy. The autoregressive moving average (ARMA) approach, since first being proposed by Box and Jenkins [4] and then popularized by Carlson et al. [5], has been

one of the most widely-used methods for hydrological forecasting [6–8]. The ARMA technique assumes the time series to be stationary and to follow the normal distribution [4]. However, streamflow time series is usually characterized by features of both nonlinearity and unstableness; thus, linear-related time series forecasting techniques are not sufficient to capture the characteristics of hydrological time series [9]. In this case, artificial neural networks (ANNs) have been put forward and widely exploited for hydrological forecasting to deal with the nonlinearity and instability of hydrological time series. Hornik et al. [10] have demonstrated that ANN can approximate any measurable function to a certain degree. One of the most obvious advantages of the ANN technique is that one does not need to have a well-defined process for algorithmically converting inputs into outputs [11]. The advantages and disadvantages of ANN in the application of hydrologic research have been discussed in a comprehensive review by Govindaraju [12,13].

Although there have been many achievements in the application of the ANN technique using single ANN, there is still much room to improve. One traditional method is to optimize the weights and biases of ANN. In the training process of ANN, the target is to find the optimal weights and biases to minimize the learning error. Since the behavior of ANN is highly dependent on the initial values of weights and biases, employing different heuristic searching algorithms becomes popular in the training process [14–18]. Mirjalili et al. [14] employed the gravitational search algorithm (GSA) and PSOGSA (particle swarm optimization, gravitational search algorithm) as training methods to train the neural networks. Khan and Sahai [15] demonstrated the superiority of the new metaheuristic bat algorithm (BA) over the genetic algorithm (GA), particle swarm optimization (PSO), the "standard" back propagation (BP) algorithm and the Levenberg-Marquardt algorithm. In Ref. [16], the GA was exploited to search for the optimal weights and biases of the back propagation neural network (BPNN). Cheng et al. [17] proposed an ANN model based on quantum-behaved PSO (ANN-QPSO) for daily runoff forecasting. In Ref. [18], a gravitational search algorithm (GSA)-based radial basis function (RBF) approach (GSA-RBF) was exploited for predicting debris flow velocity.

Another new trend to improve the performance of the single ANN model is to take advantage of data pre-analysis methodologies such as eliminating stochastic volatilities from the original time series [19, 20], using decomposed subseries as inputs [21,22] and employing a decompose-predict-ensemble framework [23,24]. Wu et al. [19] combined moving average (MA), singular spectrum analysis (SSA), wavelet multi-resolution analysis (WMRA) and ANNs to improve the forecasting accuracy of daily flows. In Ref. [20], the MA method, principle component analysis (PCA) and SSA were coupled with ANNs to predict rainfall time series. Besides these aforementioned data preprocessing-based techniques, wavelet and empirical mode decomposition (EMD) have been the most commonly-used data pre-analysis techniques in recent years. Kisi et al. [21] developed a wavelet-support vector regression (WSVR) model in which the discrete wavelet transform (DWT) was employed as a data preprocessing tool. The discrete wavelet transform has been proven to have improved the forecasting accuracy. Budu et al. [22] decomposed the observed reservoir inflow time series into subseries using DWT with different mother wavelet functions. Then appropriate subseries were used as inputs for constructing ANN models. Fu et al. (2014) developed a hybrid EMD-RBFNN model for predicting regional groundwater depth time series [23]. Zhao et al. [24] combined chaotic least squares support vector machine (CLSSVM) with EMD for annual runoff forecasting. The EMD-CLSSVM model has been proven to have attained significant improvement compared with the single CLSSVM model in long-term runoff time series forecasting. Nourani et al. [25] provided a comprehensive introduction to the application of the wavelet artificial intelligence (AI) conjunction model in simulating important processes at the hydrologic cycle.

Although wavelet transform (WT) and EMD have been used widely in recent years, the behavior of different WT-based forecasting models depends highly on the selection of mother wavelets, and EMD suffers from the drawbacks of a lack of mathematical theory and sensitivity to both noise and sampling. In this case, Gilles [26] proposed a new adaptive data analysis method, empirical wavelet transform (EWT), to overcome the drawbacks of these decomposition techniques. The efficiency of

EWT has been demonstrated by a few benchmarks in Ref. [26]. Additionally, EWT has been applied to several forecasting models successfully since being proposed in 2013. Hu and Wang [27] combined the EWT with Gaussian process regression (GPR) for forecasting short-term wind speed. Wang and Hu [28] proposed a GPR model combined with autoregressive integrated moving average (ARIMA), extreme learning machine (ELM), SVM and LSSVM for short-term wind speed forecasting. In both Ref. [27] and Ref. [28], EWT was employed as a data pre-analysis technique to filter the noises of the original time series.

Multi-verse optimizer (MVO) is a novel nature-inspired optimization algorithm, which was originally introduced by Mirjalili et al. [29]. In Ref. [29], the efficiency of MVO has been demonstrated by 19 benchmark functions and five real engineering optimization problems. Results have proven that the MVO algorithm outperforms the best algorithms proposed in the literature on the majority of the test beds. Additionally, the MVO algorithm has exhibited good performance in optimizing the weights and biases of a feedforward multilayer perceptron neural network [30], as well as the parameters of SVM [31].

In consideration of the good performance of EWT and MVO, a hybrid EWT-MVO-ANN model is proposed for more accurate annual streamflow forecasting results in which the EWT is employed to decompose the streamflow time series into subseries and eliminate the noises; the MVO algorithm is adopted to optimize the parameters of ANN; and ANN trained with MVO is employed for time series forecasting. Before constructing MVO-ANN models, the input variables are selected according to the PACF (partial autocorrelation function) values. Parallel experiments as BPNN, PSO-ANN and MVO-ANN with and without the EWT denoising process have been designed and conducted to apply to annual streamflow data of four hydrological stations in China to highlight the superiority of the proposed EWT-MVO-ANN model. The results of the experiments of this study have demonstrated that EWT is an effective way to filter the noises of annual streamflow series and thus improving prediction accuracy, and also, the MVO algorithm exhibits better performance compared with the Levenberg–Marquart (LM) and PSO algorithms in the training of ANN models for annual streamflow forecasting. It has been demonstrated by the results of this study that the proposed hybrid model can provide an effective modeling approach to capture the nonlinear characteristics of annual streamflow series, thus providing more satisfactory forecasting results.

2. Methodology

2.1. Empirical Wavelet Transform

The empirical wavelet transform (EWT) was first introduced by Jerome Gilles [26] in 2013. The main idea of this method is to extract the amplitude modulated-frequency modulated (AM-FM) signals of the given signal by designing an appropriate wavelet filter bank adaptively. EWT produces intrinsic mode functions (IMFs) according to the information contained in the spectrum of the signal. The decomposition process of signal $x(t)$ using EWT can be described as follows:

Step 1　Compute the Fourier spectrum $F(\omega)$ of the original time series $x(t)$ using the fast Fourier transform algorithm (FFT).

Step 2　Extract different modes by proper segmentation of the Fourier spectrum and determine the boundaries.

Step 3　After the boundaries are determined by appropriate segmentation, calculate the approximation and detail coefficients by applying the scaling function and empirical wavelets that correspond to each detected segment.

Equipped with the obtained set of local maxima plus 0 and π, the boundaries ω_i of each segment are defined as the center between two consecutive maxima. Let $\{f_0, f_1, f_2, ..., f_N\}$ be the

set of frequencies corresponding to the set of local maxima and $f_0 = 0, f_N = \pi$. The boundaries ω_i can be given as:

$$\omega_i = \frac{f_i + f_{i+1}}{2} \qquad for \ 1 \le i \le N - 1 \tag{1}$$

then the Fourier segments will be $[0, \omega_1], [\omega_1, \omega_2], ..., [\omega_{N-1}, \pi]$. A transition phase T_n of width $2\tau_n$ is defined centered around each ω_n [26].

After determining the set of boundaries, a bank of N wavelet filters, comprised of one low-pass filter and $N - 1$ band-pass filters, is defined for each segment based on the detected boundaries. The expressions for empirical wavelets $\psi_i(\omega)$ and scaling function $\varphi_i(\omega)$ are given by Gilles [26], as shown in Equations (3) and (4), respectively:

$$\psi_i(\omega) = \begin{cases} 1 & , if (1+\gamma)\omega_i \le |\omega| \le (1-\gamma)\omega_{i+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_{i+1}}(|\omega| - (1-\gamma)\omega_{i+1})\right)\right] & , if (1-\gamma)\omega_{i+1} \le |\omega| \le (1+\gamma)\omega_{i+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_i}(|\omega| - (1-\gamma)\omega_i)\right)\right] & , if (1-\gamma)\omega_i \le |\omega| \le (1+\gamma)\omega_i \\ 0 & , otherwise \end{cases} \tag{2}$$

$$\varphi_1(\omega) = \begin{cases} 1 & , \quad if \ |\omega| \le (1-\gamma)\omega_1 \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_1}(|\omega| - (1-\gamma)\omega_1)\right)\right] & , \quad if (1-\gamma)\omega_1 \le |\omega| \le (1+\gamma)\omega_1 \\ 0 & , \quad otherwise \end{cases} \tag{3}$$

where $\beta(x)$ is an arbitrary function such that:

$$\beta(x) = \begin{cases} 0 & , if \ x \le 0 \\ & and \quad \beta(x) + \beta(1-x) = 1 \quad \forall x \in [0,1] \\ 1 & , if \ x \ge 1 \end{cases} \tag{4}$$

The widely-used function $\beta(x)$ according to Ref. [32] is:

$$\beta(x) = x^4\left(35 - 84x + 70x^2 - 20x^3\right) \tag{5}$$

Additionally, $\tau_n = \gamma\omega_n, 0 < \gamma < 1, \gamma$ can be selected using the below expression:

$$\gamma < \min_i \left(\frac{\omega_{i+1} - \omega_i}{\omega_{i+1} + \omega_i}\right) \tag{6}$$

The detail coefficients are given by the inner products with the empirical wavelets:

$$W_x(i,t) = \langle x(t), \psi_i(t) \rangle = \int x(\tau)\psi_i(\tau - t)d\tau = F^{-1}[x(\omega)\psi(\omega)] \tag{7}$$

while the approximation coefficient by the inner products with the scaling function:

$$W_x(1,t) = \langle x(t), \varphi_1(t) \rangle = \int x(\tau)\varphi_1(\tau - t)d\tau = F^{-1}[x(\omega)\varphi_1(\omega)] \tag{8}$$

where F^{-1} denotes the inverse Fourier transform.

Finally, the reconstruction of the original signal is obtained according to:

$$x(t) = W_x(1,t) * \varphi_1(t) + \sum_{i=2}^{N} W_x(i,t) * \psi_i(t) \tag{9}$$

where $*$ means convolution.

2.2. The Proposed MVO-ANN Model

2.2.1. Multi-Verse Optimizer

The multi-verse optimizer (MVO) algorithm, proposed by Mirjalili et al. [29], is inspired by the theory of the multi-verse in physics. Multi-verse theory believes that there is more than one big bang, and each big bang causes the birth of a universe [29]. It builds a mathematical model of the white/black holes (wormholes) tunnels that connect two universes and exchange the objects of universes using the wheel mechanism. At every iteration, the universes are sorted according to their inflation rates, and the one with the highest inflation rate is chosen by the roulette wheel to have a white hole. MVO is mathematically modeled as follows:

Assume that:

$$U = \begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^d \\ x_2^1 & x_2^2 & \cdots & x_2^d \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^d \end{bmatrix} \tag{10}$$

where d is the number of parameters (variables) and n is the number of universes (candidate solutions):

$$x_k^j = \begin{cases} x_k^j & r1 < NI(Ui) \\ x_i^j & r1 \geq NI(Ui) \end{cases} \tag{11}$$

where x_i^j denotes the j-th parameter of the i-th universe, Ui denotes the i-th universe, $NI(Ui)$ is the normalized inflation rate of the i-th universe and $r1$ ranges from 0 to 1.

Assuming that there is a wormhole tunnel between one universe and the best universe, the transportation can be described as follows:

$$x_i^j = \begin{cases} \begin{cases} X_j + \text{TDR} \times ((ub_j - lb_j) \times r4 + lb_j) & r3 < 0.5 \\ X_j - \text{TDR} \times ((ub_j - lb_j) \times r4 + lb_j) & r3 \geq 0.5 \end{cases} & r2 < \text{WEP} \\ x_i^j & r2 \geq \text{WEP} \end{cases} \tag{12}$$

where X_j denotes the j-th parameter of the best universe, WEP and TDR are worm existence probability and travelling distance rate, respectively, lb_j and ub_j denote the lower and upper bound of the j-th variable, respectively, and $r2$, $r3$, $r4$ are random numbers in [0, 1]. The formulae for WEP and TDR are defined as follows:

$$\text{WEP} = \min + l \times \left(\frac{\max - \min}{L} \right) \tag{13}$$

$$\text{TDR} = 1 - \frac{l^{1/p}}{L^{1/p}} \tag{14}$$

where min and max are the minimum and maximum, which are defined as 0.2 and 1, respectively, l denotes the current iteration, L denotes the maximum iterations and p is the exploitation accuracy over the iterations and is defined as 6. The number of universes is defined as 30, and the maximum number of iterations is 500 in this study.

2.2.2. Artificial Neural Network

As a typical intelligent learning paradigm, the artificial neural network has been widely used in many areas of science and engineering such as daily water level forecasting [33], rainfall-runoff simulation [34] and wind speed forecasting [35]. As can be seen in Figure 1, a typical three-layer feed-forward ANN usually consists of three layers including input, hidden and output layers [36]. Each layer contains nodes that are connected to nodes of the other layer(s). Additionally, the connection

is associated with a weight. Assume a simplified neural network with three layers. The output of the *j*-th hidden node can be obtained as given in Equation (15):

$$H_j = f\left(\sum_{i=1}^{n} w_{ij}x_i - a_j\right) \qquad j = 1, 2, ..., l \tag{15}$$

where $f(x) = 1/(1 + \exp(-x))$ and is the transfer function of the hidden layer; n denotes the number of nodes in the input layer; l denotes the number of nodes in the hidden layer; w_{ij} represents the connection weight from the *i*-th input node to the *j*-th hidden node; and a_j represents the bias of the *j*-th hidden node.
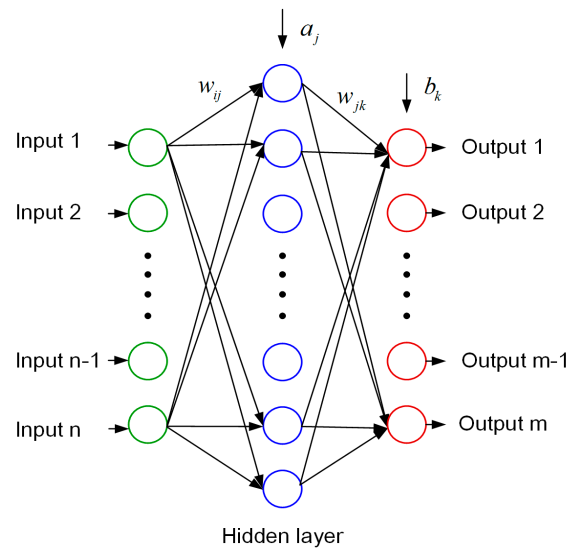

Figure 1. Architecture of a typical three-layer ANN.

After calculating the outputs of the hidden layer, the final output can be given as follows:

$$O_k = \sum_{j=1}^{l} H_j w_{jk} - b_k \qquad k = 1, 2, ..., m \tag{16}$$

where m denotes the number of nodes in the output layer, w_{jk} denotes the connection weight from the *j*-th hidden node to the *k*-th output node and b_k is the bias of the *k*-th output node.

2.2.3. ANN Trained with MVO

Training of neural networks is a form of supervised learning in which the weights and thresholds of the network are updated for the purpose of reducing the function error whenever the error of the training outputs does not match the expected accuracy. Unlike traditional local optimization training algorithms, such as the gradient descent method and Levenberg–Marquart (LM) algorithm, the global optimization algorithms can overcome the drawbacks of local minima and obtain more efficient solutions. The MVO algorithm is exploited to find an optimal combination of connection weights and biases to minimize the fitness error. In the MVO algorithm, the iteration begins with creating a set of random universes. When training the neural network, the weights and thresholds of the network are represented by the parameters of the MVO algorithm. In each iteration of MVO, the weights and thresholds are optimized.

In the MVO-ANN model, the ANN model is trained with MVO. In order to evaluate the performance of the MVO-ANN model, root mean square error (*RMSE*) is adopted as the fitness

function. Define the number of training samples as q, then the fitness function can be described as follows:

$$E_i = \sum_{k=1}^{m} \left(Y_k^i - O_k^i \right)^2 \tag{17}$$

$$fitness = \sqrt{\frac{1}{q} \sum_{i=1}^{q} E_i} \tag{18}$$

where Y_k^i and O_k^i denote the predicted and observed values of the k-th output node in the i-th training sample, respectively.

2.3. Partial Autocorrelation Function

Selecting appropriate input variables for ANN models is an essential precondition for ANN modeling. The relationship between inputs and outputs has a significant impact on the forecasting accuracy of ANN models. To address this matter, PACF values of the streamflow time series are adopted to identify the input variables. Assume $x_t (t = 1, 2, ..., n)$ to be the target variable. If PACF values at lags greater than k are inside the 95% confidence interval, $[-1.96/\sqrt{n}, 1.96/\sqrt{n}]$, $x_{t-1}, x_{t-2}, ..., x_{t-k}$ are employed as input variables. If all PACF values are in the 95% confidence interval, the input variables are chosen according to a trial and error method by systematically increasing the number of sequential data from the latest year to the sixth year. The calculation steps of the PACF values of the target variable $x_t (t = 1, 2, ..., n)$ can be briefly described as follows [37]:

Firstly, let γ_k denote the covariance at lag k (γ_0 represents the variance), γ_k can be calculated as follows:

$$\gamma_k = \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x}), k = 0, 1, ..., M \tag{19}$$

Secondly, the autocorrelation coefficient at lag k, denoted by ρ_k, is calculated:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{20}$$

Finally, the partial autocorrelation coefficient at lag k, denoted as α_{kk}, is obtained as follows:

$$\left. \begin{array}{l} \alpha_{11} = \rho_1 \\ \alpha_{k+1,k+1} = \dfrac{\rho_{k+1} - \sum_{k=1}^{k} \rho_{k+1-j} \alpha_{kj}}{1 - \sum_{j=1}^{k} \rho_j \alpha_{kj}} \\ \alpha_{k+1,j} = \alpha_{kj} - \alpha_{k+1,k+1} \cdot \alpha_{k,k-j+1} \ (j = 1, 2, ..., k) \end{array} \right\} \tag{21}$$

where $k = 1, 2, ..., M$.

2.4. The Hybrid EWT-MVO-ANN Model

The overall structure of the proposed hybrid EWT-MVO-ANN model is shown in Figure 2. As can be seen in Figure 2, the detailed procedure of the proposed hybrid EWT-MVO-ANN model can be described in four steps:

Firstly, the EWT is exploited to decompose the original streamflow series into several detailed components and an approximated component. The approximated component corresponds to the general trend of the original series while the detailed components are high-frequency IMFs that belong to different frequency bands of the streamflow series. Secondly, the residual series is thrown away to eliminate the noises and smooth the original streamflow series. The sum of the remaining modes produces the reconstructed series. Thirdly, the PACF values are calculated to determine the best combination of input variables. Fourthly, the MVO algorithm is exploited to find the optimal weights

and biases of the ANN architecture to minimize the fitness error. Finally, the forecasting results of the streamflow series are obtained according to the well-trained MVO-ANN model.
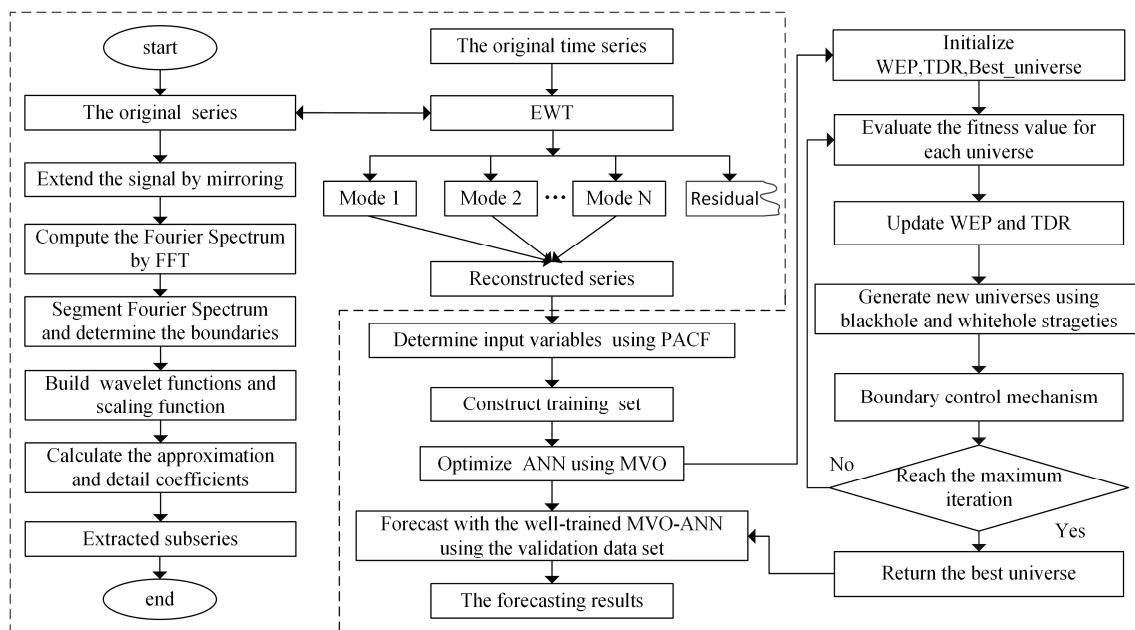

Figure 2. The overall framework of the proposed empirical wavelet transform (EWT)-multi-verse optimizer (MVO)-ANN model. PACF, partial autocorrelation function; TDR, travelling distance rate; WEP, worm existence probability.

2.5. Model Performance Evaluation

It is important to apply multiple error measures when evaluating the forecasting ability of the developed models. This paper considers four statistical measures including *RMSE*, mean absolute error (*MAE*), mean absolute percent error (*MAPE*) and the coefficient of correlation (*R*). Among the four statistical measures, *RMSE* is sensitive to extremely large or small values of a time series and reflects the degree of variation; *MAE* reflects the actual forecasting error in a more balanced perspective; and *MAPE* is a measure of accuracy with no units. The *RMSE*, *MAE* and *MAPE* are defined as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(q_p(i) - q_o(i)\right)^2} \tag{22}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|q_p(i) - q_o(i)\right| \tag{23}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|q_p(i) - q_o(i)\right|}{q_o(i)} \times 100\% \tag{24}$$

where $q_p(i)$ and $q_o(i)$ denote the predicted and observed annual streamflow series, respectively, and N denotes the length of data.

R shows the degree to which two datasets are related to each other and ranges from -1 to 1. The larger the absolute value of *R*, the more the predicted and observed data are related. The coefficient of correlation (*R*) is defined as:

$$R = \frac{\sum_{i=1}^{N}\left(q_p(i) - \bar{q}_p\right)\left(q_o(i) - \bar{q}_o\right)}{\sqrt{\sum_{i=1}^{N}\left(q_p(i) - \bar{q}_p\right)^2} \times \sqrt{\sum_{i=1}^{N}\left(q_o(i) - \bar{q}_o\right)^2}} \tag{25}$$

where $\bar{q}_p$ and $\bar{q}_o$ denote the average values of the predicted and observed runoffs, respectively.

3. Model Construction and Development

3.1. Study Area and Data Collection

Annual streamflow time series of four hydraulic gauging stations located in two river basins in the upper reach of the Yangtze River in China, i.e., Xiangjiaba, Panzhihua, Luoxidu and Sanleiba, are researched in this study.

The first river basin is referred to as the Jinsha River Basin. It flows through the provinces of Qinghai, Sichuan and Yunnan in western China. Originating in the uppermost of the Yangtze River, the Jinsha River has a length of 3486 km, accounting for 77% of the total length of the upper Yangtze River, and has a drainage area of 480,000 km^2, accounting for 50% of the total area of the upper Yangtze River. The observed annual streamflow data of Xiangjiaba and Panzhihua stations, which lie on the lower reaches of the Jinsha River basin, are selected as the first two study areas. For Xiangjiaba station, annual streamflow data from 1959 to 2012 (54 observations) are studied. The training dataset is from 1959 to 2000 (42 observations), whilst the validation data set is from 2001 to 2012 (12 observations). For Panzhihua station, streamflow data from 1953 to 2012 (60 observations) are studied. The training dataset is from 1953 to 1998 (46 observations), whilst the validation dataset is from 1999 to 2012 (14 observations).

The second river basin is referred to as the Jialing River Basin. Jialing River is a major tributary of the Yangtze River in the Sichuan Basin. The Jialing River Basin has a main stream length of 1120 km and a drainage area of 160,000 km^2, accounting for about 17% of the drainage area of the Yangtze River Basin. Two hydraulic gauging stations, which are called Luoduxi and Sanleiba, are selected in this study. Annual streamflow data from 1954 to 2008 (55 observations) for both Luoduxi and Sanleiba are studied. Data from 1954 to 1996 (43 observations) are for training, whilst those from 1997 to 2008 (12 observations) are for validation.

Figure 3 provides a schematic of the Jinsha River Basin and the Jialing River Basin, as well as the four hydrological gauging stations researched in this study. Annual streamflow data of the four stations are presented in Figure 4. Basic statistical information, such as the mean value, standard deviation (SD), minimum (Min.), maximum (Max.), skewness (Skew.) and kurtosis (Kurt.), of these datasets is illustrated in Table 1.

Table 1. The statistical information of the streamflow data of the four stations. Skew., skewness; Kurt., kurtosis.

| Station | Catchment Area (km^2) | Statistics Information | | | | | | Time Range | |
		Mean (m^3/s)	SD (m^3/s)	Min. (m^3/s)	Max. (m^3/s)	Skew.	Kurt.	Training	Validation
Xiangjiaba	458,800	4593	730	3245	6286	0.39	2.31	1959–2000	2001–2012
Panzhihua	259,177	1830	298	1211	2423	0.13	2.03	1953–1998	1999–2012
Luoduxi	38,064	686	229	300	1400	0.40	2.94	1954–1996	1997–2008
Sanleiba	28,896	308	73	188	496	0.58	2.72	1954–1996	1997–2008
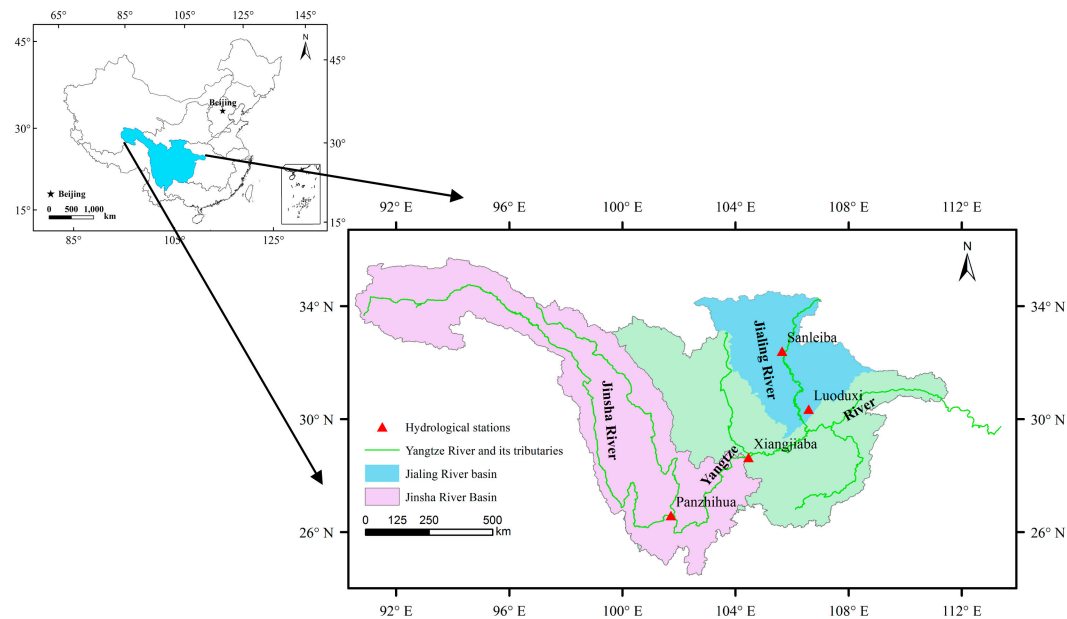
Figure 3. Locations of upper Yangtze River, as well as the two study areas and the four gauging stations.
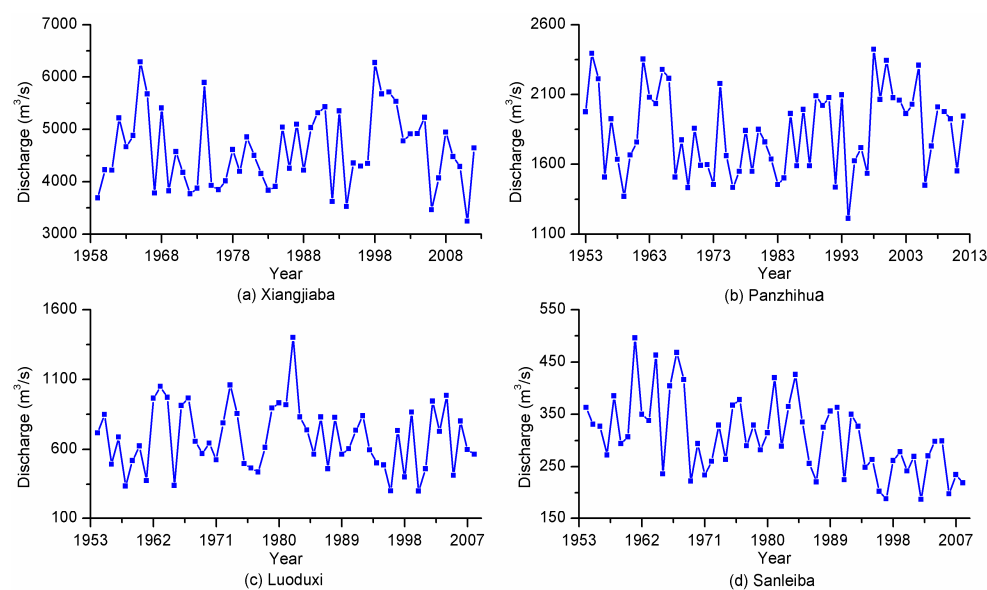

Figure 4. Annual streamflow series for the four stations.

3.2. Decomposing and Reconstruction Using EWT

Since EWT borrows the framework of classic WT in the decomposition and reconstruction processes [27], and three-level decompositions have been considered as the most appropriate in various studies, such as wind speed forecasting [27,28], drought index prediction [38] and daily river flow forecasting [39]; the decomposition level is determined as three in this study. Graphical representations of the decomposed subseries using EWT for four study stations are illustrated in Figure 5. As can be seen in Figure 5, the original annual streamflow series is decomposed into three independent modes and a residual for the four stations, respectively.
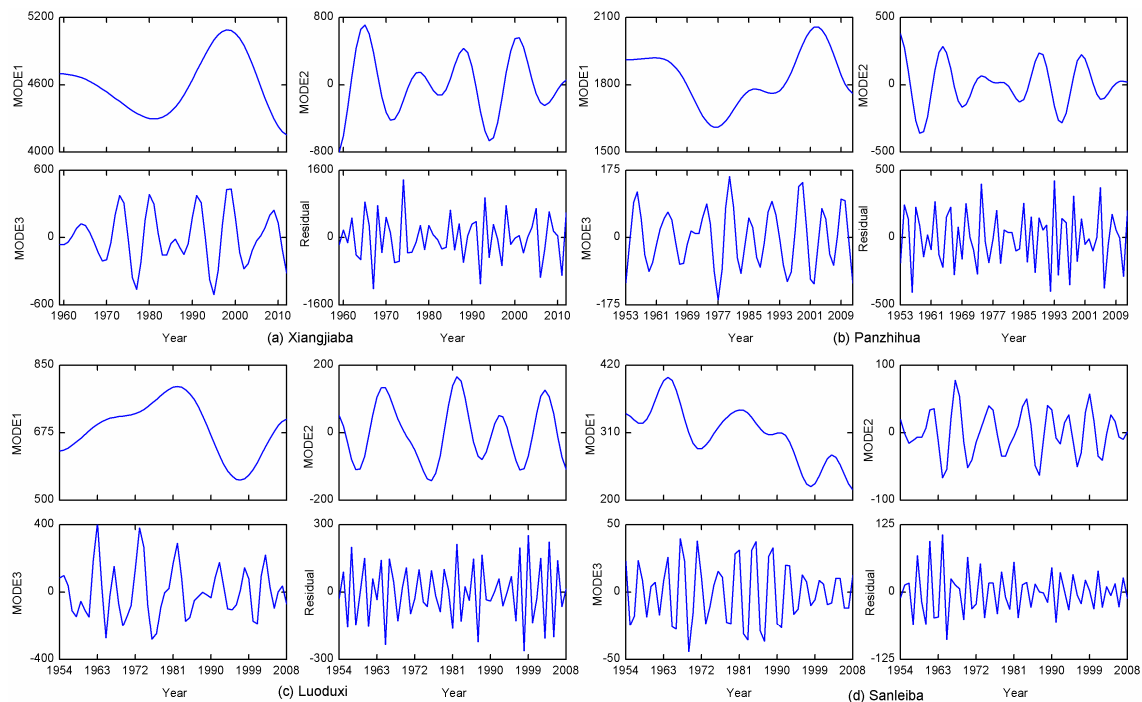
Figure 5. Decomposed subseries of annual streamflow series for the four stations.

Among these modes, Mode 1 is a trend term that describes the general tendency of the original annual streamflow series. The modes present changing frequencies, amplitudes and wavelengths. For all stations, Mode 1 has the lowest frequency, the longest wavelength and contributes the most to the original time series. For the following subseries, the frequency is increased, while the wavelength and contribution are decreased. After decomposing the original streamflow series, the residue is thrown away because it mainly contains noisy information, and the remaining modes are reconstructed in a new series. The graphical representations of the reconstructed series are shown in Figure 6. As can be seen in Figure 6, the reconstructed series tends to be smoother after the denoising processes.
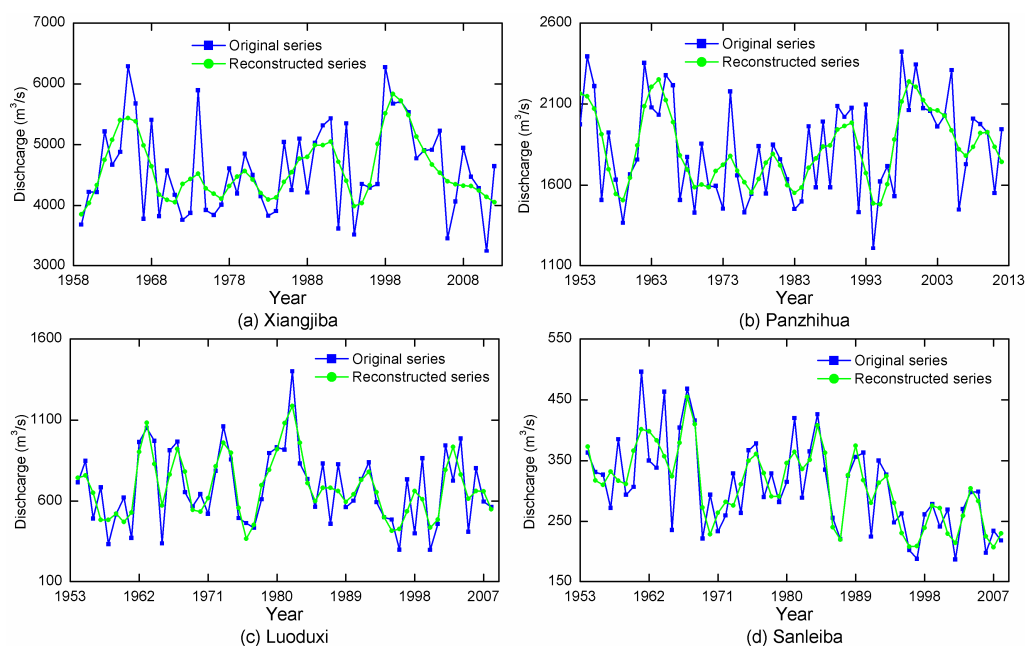

Figure 6. Comparison between the original and reconstructed streamflow series for the four stations.

3.3. Input Determination

To obtain the input variables for each model, the PACF values between each streamflow series and their antecedent values at lag *k* are calculated using the method described in Section 2.3, as well as the PACF values between the reconstructed streamflow series and their antecedent values. Graphical representations of the PACF values are shown in Figure 7 for the four stations.
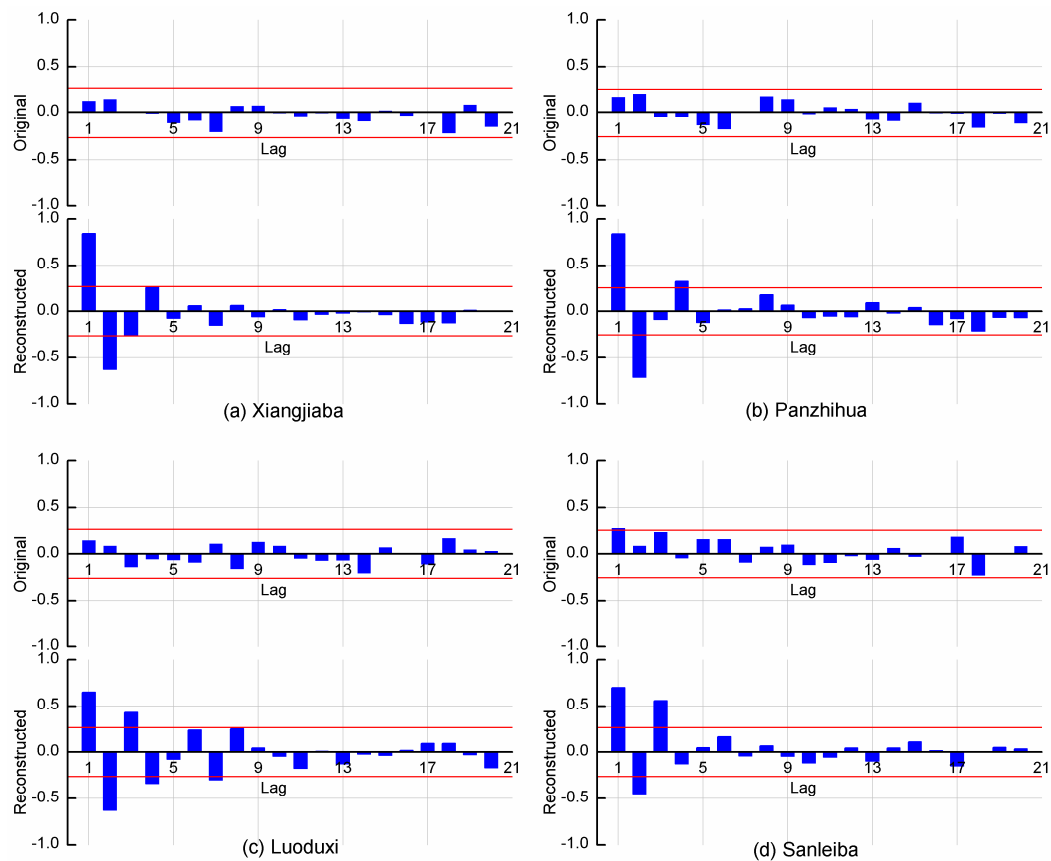

Figure 7. PACF plot with 95% confidence bands for the original and reconstructed series.

According to the input variable selection method described in Section 2.3, the input variables of the models for the original and the reconstructed series of the four stations are finally determined as described in Table 2.

Table 2. Input determination for the original and reconstructed series.

Stations	Inputs Combination	
	Original	**Reconstructed**
Xiangjiaba	x_{t-1}, x_{t-2}	x_{t-1}, x_{t-2}
Panzhihua	x_{t-1}, x_{t-2}	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$
Luoduxi	x_{t-1}, x_{t-2}	$x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}$
Sanleiba	x_{t-1}, x_{t-2}	$x_{t-1}, x_{t-2}, x_{t-3}$

3.4. Parameter Settings

To verify the efficiency of the proposed MVO-ANN model, the back propagation neural network (BPNN) trained with the LM algorithm and ANN trained with the PSO algorithm (PSO-ANN) have been constructed for annual streamflow forecasting.

After determining the optimal combination of input variables using the method described in Section 2.3, the streamflow data are normalized to [0, 1]. In most cases, the sigmoid function is used as the transfer function in the hidden layer, while the linear activation function is used in the output layer for ANNs [40]. In this study, all of the ANN models exploit the tan-sigmoid function and purelin function as the activation function for the hidden layer and output layer, respectively. For the single BPNN models, the LM training function (TrainLM) is exploited to speed up the training process. For the MVO-ANN models, the number of universes is defined as 30, and the maximum number of iterations is 500 according to Ref. [29]. A value of 1000 for the maximum iteration number has also been tried for the MVO-ANN model in the experimental process. Results have shown that there is no significant improvement in accuracy, but the experiment time is much longer. The WEP is increased linearly from 0.2 to 1, while TDR is decreased from 0.6 down to zero according to Equations (13) and (14), respectively. For the PSO-ANN models, the population size is set as 30 and the maximum number of iteration is set as 500 to compare with the MVO-ANN model. The optimal numbers of hidden nodes for all ANN models including BPNN, PSO-ANN and MVO-ANN with and without EWT are determined using grid search (GS) algorithm. In the GS algorithm, the searching range is set as Ref. [2,29] and the grid step is set as one [35]. Each experiment is repeated 10 times, and results with the best accuracy are presented.

4. Results and Analysis

To evaluate the performance of these models, *RMSE*, *MAE*, *MAPE* and *R* of the BPNN, PSO-ANN and MVO-ANN models with and without the denoising process are analyzed and compared. The computations related to all of the models including BPNN, PSO-ANN, MVO-ANN, EWT-BPNN, EWT-PSO-ANN and EWT-MVO-ANN are implemented in the MATLAB environment in a computer with Intel core i5, 2.6-GHz CPU and 4 GB of RAM. The performance indices of the four stations in the training and validation periods are given in Tables 3–6, respectively.

Table 3. Forecasting performance of the models for Xiangjiaba Station.

Models	Training				Validation			
	RMSE (m^3/s)	*MAE* (m^3/s)	*MAPE* (%)	*R*	*RMSE* (m^3/s)	*MAE* (m^3/s)	*MAPE* (%)	*R*
BPNN	660.37	534.91	11.5	0.525	721.17	602.00	14.7	0.223
PSO-ANN	603.42	480.81	10.4	0.629	725.10	590.04	14.0	0.227
MVO-ANN	576.73	447.18	9.5	0.672	774.46	655.93	15.3	0.514
EWT-BPNN	578.23	387.91	8.9	0.708	577.72	500.92	11.2	0.649
EWT-PSO-ANN	482.79	363.97	8.1	0.786	543.31	465.48	10.6	0.665
EWT-MVO-ANN	431.27	316.94	7.2	0.838	513.98	448.51	10.3	0.724

Table 4. Forecasting performance of the models for Panzhihua Station.

Models	Training				Validation			
	RMSE (m^3/s)	*MAE* (m^3/s)	*MAPE* (%)	*R*	*RMSE* (m^3/s)	*MAE* (m^3/s)	*MAPE* (%)	*R*
BPNN	285.79	224.09	12.3	0.443	226.40	191.85	10.0	0.494
PSO-ANN	264.85	214.35	12.2	0.453	212.22	176.76	9.6	0.496
MVO-ANN	251.34	202.00	11.4	0.537	217.47	177.89	9.7	0.538
EWT-BPNN	193.35	140.15	8.1	0.766	208.66	187.40	9.7	0.632
EWT-PSO-ANN	180.35	144.71	8.2	0.800	209.13	169.76	9.3	0.612
EWT-MVO-ANN	159.39	124.89	7.2	0.844	181.19	158.20	8.0	0.727

Table 5. Forecasting performance of the models for Luoduxi Station.

Models	Training				Validation			
	RMSE (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*	*RMSE* (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*
BPNN	209.72	164.64	24.5	0.526	398.13	356.80	57.2	0.406
PSO-ANN	196.87	161.81	25.5	0.540	327.48	286.82	50.2	0.380
MVO-ANN	177.94	136.75	21.8	0.648	284.38	257.35	44.4	0.532
EWT-BPNN	184.41	137.98	20.6	0.790	271.71	216.21	40.6	0.571
EWT-PSO-ANN	122.40	85.17	13.6	0.854	241.08	208.13	33.6	0.623
EWT-MVO-ANN	118.85	91.11	13.8	0.863	184.66	157.95	27.3	0.611

Table 6. Forecasting performance of the models for Sanleiba Station.

Models	Training				Validation			
	RMSE (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*	*RMSE* (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*
BPNN	66.88	45.62	16.2	0.614	53.33	41.28	17.1	0.496
PSO-ANN	65.88	50.16	16.8	0.601	38.62	33.59	13.2	0.489
MVO-ANN	60.32	49.81	16.3	0.618	36.01	30.07	12.0	0.521
EWT-BPNN	49.12	37.56	12.2	0.766	35.98	31.13	11.8	0.697
EWT-PSO-ANN	45.02	36.05	11.6	0.813	31.37	25.06	9.6	0.720
EWT-MVO-ANN	38.87	26.86	8.4	0.861	35.24	29.68	11.9	0.752

It is clear from Tables 3–6 that the proposed EWT-MVO-ANN model obtains the smallest *RMSE*, *MAE* and *MAPE* and the biggest *R* compared with the other five models. Comparison between the models with and without the EWT algorithm reveals that combining EWT with the ANN model is an effective way to increase the forecasting accuracy. Though the proposed EWT-MVO-ANN model exhibits much better performance than the other models, the configuration is more complex on account of data pre-processing and parameter optimization. Thus, the computation time required is much more than the single BPNN model without the denoising and optimization processes. Detailed analyses are submitted to further evaluate the performance of these models.

Table 3 displays the forecasting results of the six models in terms of *RMSE*, *MAE*, *MAPE* and *R* for Xiangjiaba Station. For models without denoising process, the MVO-ANN model outperforms the other models in the training period in terms of the four indices. However, the *RMSE*, *MAE* and *MAPE* are a little bigger than the other models in the validation period. The comparison between BPNN, PSO-ANN and MVO-ANN demonstrates that the simple LM algorithm may be more suitable for forecasting in some occasions. For models with EWT, the EWT-MVO-ANN model outperforms the other models both in the training and validation periods in terms of the four indices. Overall, the EWT-MVO-ANN method outperforms the other models. For example, the *R* of the EWT-MVO-ANN model is 0.724 while those of the BPNN, PSO-ANN, MVO-ANN, EWT-ANN and EWT-PSO-ANN are 0.223, 0.227, 0.514, 0.649 and 0.665 in the validation period, respectively.

Table 4 displays the forecasting performance of the six models in terms of *RMSE*, *MAE*, *MAPE* and *R* for Panzhihua Station. For the four indices, the MVO-ANN model improved the BPNN model by 3.94% decreasing in *RMSE*, 7.28% decreasing in *MAE*, 3% decreasing in *MAPE* and 8.91% increasing in *R* in the validation period while the improvement of the EWT-MVO-ANN model to the BPNN model is 19.97%, 17.54% and 20% decreasing in *RMSE*, *MAE* and *MAPE* and 47.17% increasing in *R*, respectively.

Table 5 displays the forecasting performance of the six models in terms of *RMSE*, *MAE*, *MAPE* and *R* for Luoduxi Station. For the four indices, the MVO-ANN improved the BPNN model by 28.57% decreasing in *RMSE*, 27.87% decreasing in *MAE*, 22.38% decreasing in *MAPE* and 31.03% increasing in *R* in the validation period, while the improvement of the EWT-MVO-ANN model to the BPNN model is 53.61%, 55.73% and 52.27% decreasing in *RMSE*, *MAE* and *MAPE* and 50.49% increasing in *R*, respectively. These results demonstrate that the MVO algorithm outperforms the simple LM algorithm in the training of ANN models, and the use of EWT contributes greatly to the forecasting accuracy of streamflow time series.

Similarly, Table 6 displays the forecasting performance of the six models in terms of *RMSE*, *MAE*, *MAPE* and *R* for Sanleiba Station. For the four indices, the MVO-ANN improved the BPNN model by 32.48% decreasing in *RMSE*, 27.16% decreasing in *MAE*, 29.82% decreasing in *MAPE* and 0.05% increasing in *R* in the validation period while the improvement of the EWT-MVO-ANN model to the BPNN model is 33.92%, 28.10% and 30.41% decreasing in *RMSE*, *MAE* and *MAPE* and 51.61% increasing in *R*, respectively.

Table 7 illustrates the comparisons of the forecasting performance between EWT-PSO-ANN and EWT-MVO-ANN models. It can be concluded from Table 7 that the EWT-MVO-ANN models outperform the EWT-PSO-ANN models in *RMSE*, *MAE*, *MAPE* and *R* for the four stations in both training and validation periods, which demonstrate the superiority of MVO algorithm in training ANNs.

Table 7. The comparison between the EWT-PSO-ANN and EWT-MVO-ANN models for the four stations.

Stations	Models	Training				Validation			
		RMSE (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*	*RMSE* (m³/s)	*MAE* (m³/s)	*MAPE* (%)	*R*
Xiangjiaba	EWT-PSO-ANN	482.79	363.97	8.1	0.786	543.31	465.48	10.6	0.665
	EWT-MVO-ANN	431.27	316.94	7.2	0.838	513.98	448.51	10.3	0.724
Panzhihua	EWT-PSO-ANN	181.40	143.40	8.3	0.793	223.91	198.04	10.0	0.612
	EWT-MVO-ANN	159.39	124.89	7.2	0.844	181.19	158.2	8.0	0.727
Luoduxi	EWT-PSO-ANN	147.36	117.73	18.5	0.783	211.60	173.32	29.3	0.582
	EWT-MVO-ANN	118.85	91.11	13.8	0.863	184.66	157.95	27.3	0.611
Sanleiba	EWT-PSO-ANN	48.10	37.42	12.4	0.788	34.59	29.13	11.2	0.705
	EWT-MVO-ANN	38.87	26.86	8.4	0.861	35.24	29.68	11.9	0.752

To further describe the effectiveness of EWT algorithm in enhancing the accuracy of ANN models in streamflow forecasting, forecasting results of the MVO-ANN and EWT-MVO-ANN models for the four stations are illustrated in Figure 8, where line charts with markers on the left side represent the forecasted and observed values of streamflow, and scatter plots with a linear fitting line of the predicted and observed values are exhibited on the right side of Figure 8. Subgraphs on the right side of Figure 8 are drawn using the "Fit linear" tool in the OriginPro environment. The intercepts of the linear fitting lines of all subgraphs have been fixed as zero to better display the differences between two different models. From the left side of Figure 8, it is obvious that the prediction lines of the EWT-MVO-ANN models approximate the observed values better than the MVO-ANN models. From the right side of Figure 8, it can be seen that the scatters of the EWT-MVO-ANN models distribute a little more tightly around the linear fitting line compared with the MVO-ANN models. From the description above, the effectiveness of EWT in the forecasting of annual streamflow time series is demonstrated.
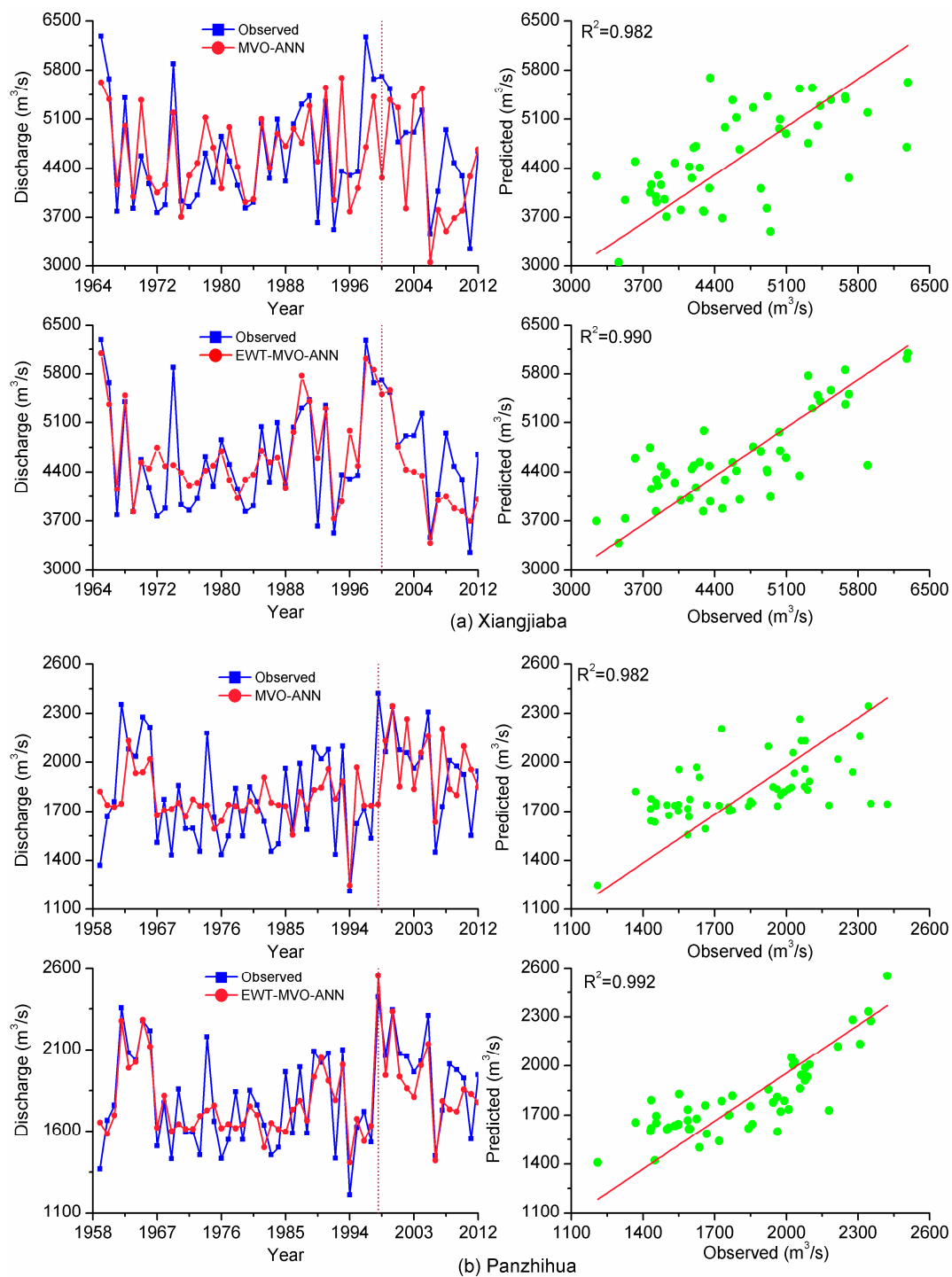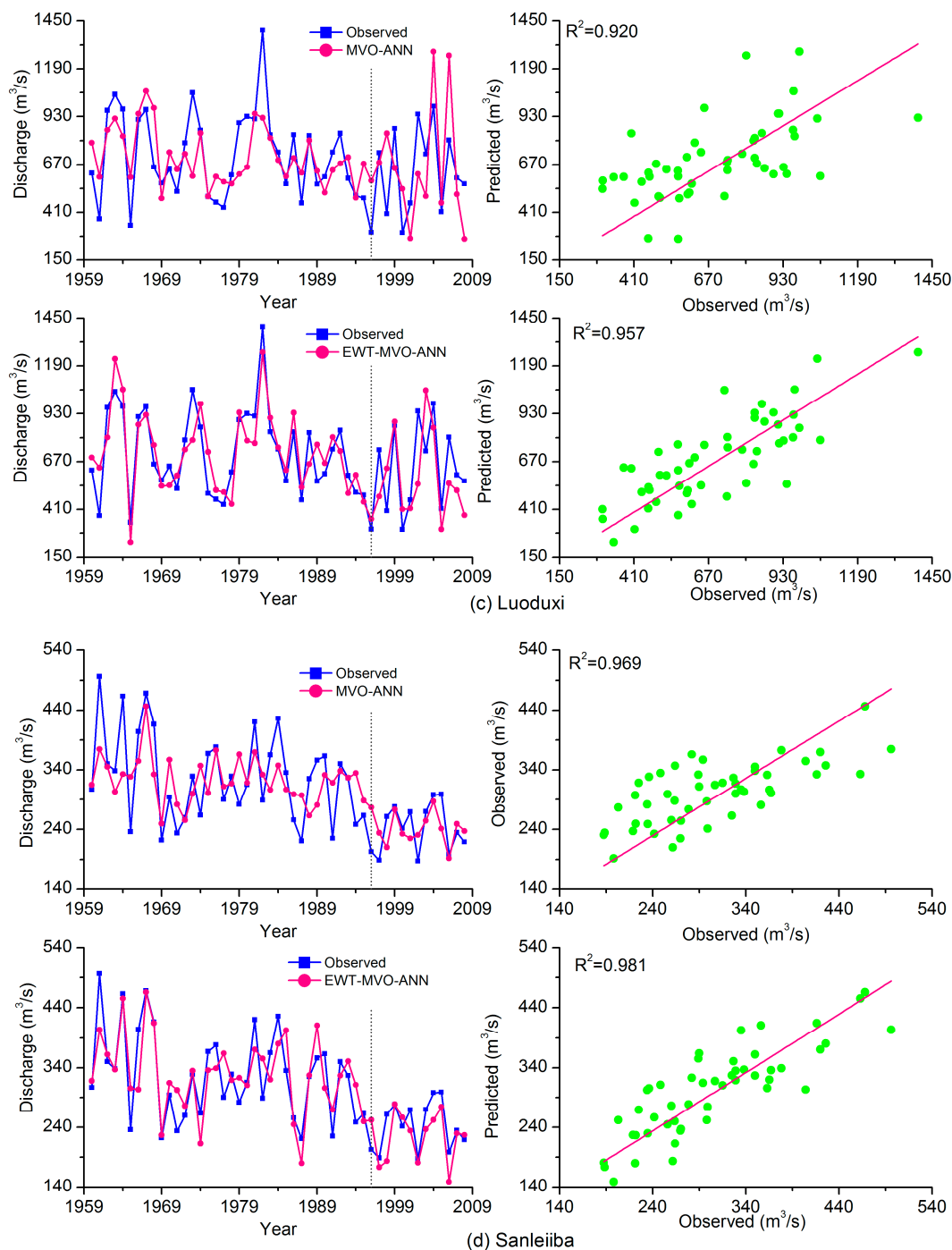
Figure 8. *Cont.*

Figure 8. Forecasting results of the MVO-ANN and the EWT-MVO-ANN models for the four stations (**left**, forecasted streamflow and observed streamflow with a dotted line dividing the training and validation stages; **right**, scatter plot with a linear fitting line).

Generally, it can be concluded from the forecasting results of the experiments of this study that the ANN model trained by LM cannot give a satisfactory performance in the forecasting of annual streamflow series. Additionally, the MVO method can achieve better performance in the training of ANN models compared with the LM and PSO algorithms. Compared with the MVO-ANN model, the EWT-MVO-ANN model leads to a much better performance in terms of the four evaluation indices, which demonstrates that EWT is an effective alternative decomposition technique in improving the forecasting accuracy of annual streamflow time series.

5. Conclusions

This paper develops a hybrid EWT-MVO-ANN model for streamflow time series forecasting. EWT is exploited to extract useful information and eliminate stochastic volatility from the original streamflow series. The ANN architecture trained with MVO method is employed as a predictor. To fully demonstrate the effectiveness of the proposed model, data from four hydrological stations with different levels of discharge and different volatilities in the upper reaches of the Yangtze River, China, are studied. Among the six models, including BPNN, PSO-ANN, MVO-ANN, EWT-BPNN, EWT-PSO-ANN and EWT-MVO-ANN, the EWT-MVO-ANN model performs the best according to four statistical indices (*RMSE*, *MAE*, *MAPE* and *R*). The following results have been concluded:

1 Compared with the BPNN model, the ANN models trained by intelligent algorithms obtain better forecasting results. The performance of the MVO algorithm is better than the PSO algorithm.
2 By discarding the residue of the subseries, the forecasting ability of the EWT-MVO-ANN model has been improved, which demonstrated that EWT is suitable for data pre-analysis for streamflow time series.
3 The proposed model is appropriate for annual streamflow series forecasting for streamflow data with different magnitudes and fluctuations in the four hydraulic gauging stations in China.

This study proposes a novel evolutionary algorithm to optimize the weights and biases of a traditional ANN, but the performances of ANNs with different network structures, such as the radial basis function network, the general regression neural network and the ELM, have not been studied. More attention will be paid to the performances of different ANNs for forecasting in the future study.

Author Contributions: Tian Peng and Chu Zhang designed and conducted the experiments. Tian Peng wrote the draft of the paper. Chu Zhang prepared the figures for this paper. Jianzhong Zhou proposed the main structure of this study. Jianzhong Zhou and Wenlong Fu provided useful advice and made some corrections. All authors read and approved the final manuscript.

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

References

1. Yang, T.C.; Yu, P.S.; Chen, C.C. Long-term runoff forecasting by combining hydrological models and meteorological records. *Hydrol. Process.* **2005**, *19*, 1967–1981. [CrossRef]
2. Barge, J.T.; Sharif, H.O. An Ensemble Empirical Mode Decomposition, Self-Organizing Map, and Linear Genetic Programming Approach for Forecasting River Streamflow. *Water* **2016**, *8*, 247. [CrossRef]
3. Napolitano, G.; Serinaldi, F.; See, L. Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: An empirical examination. *J. Hydrol.* **2011**, *406*, 199–214. [CrossRef]
4. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Wiley: Tokyo, Japan, 1970; ISBN 0-8162-1104-3.
5. Carlson, R.F.; MacCormick, A.J.A.; Watts, D.G. Application of linear random models to four annual streamflow series. *Water Resour. Res.* **1970**, *6*, 1070–1078. [CrossRef]
6. Khashei, M.; Bijari, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Appl. Soft Comput.* **2011**, *11*, 2664–2675. [CrossRef]
7. Zhao, X.H.; Chen, X. Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. *Water Resour. Manag.* **2015**, *29*, 2913–2926. [CrossRef]
8. Zounemat-Kermani, M. Investigating Chaos and Nonlinear Forecasting in Short Term and Mid-term River Discharge. *Water Resour. Manag.* **2016**, *30*, 1851–1865. [CrossRef]

9. Wei, S.; Zuo, D.; Song, J. Improving prediction accuracy of river discharge time series using a Wavelet-NAR artificial neural network. *J. Hydroinform.* **2012**, *14*, 974–991. [CrossRef]
10. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
11. Sudheer, K.P.; Gosain, A.K.; Mohana Rangan, D.; Saheb, S.M. Modelling evaporation using an artificial neural network algorithm. *Hydrol. Process.* **2002**, *16*, 3189–3202. [CrossRef]
12. Govindaraju, R.S. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123.
13. Govindaraju, R.S. Artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Eng.* **2000**, *5*, 124–137.
14. Mirjalili, S.; Hashim, S.Z.M.; Sardroudi, H.M. Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. *Appl. Math. Comput.* **2012**, *218*, 11125–11137. [CrossRef]
15. Khan, K.; Sahai, A. A comparison of BA, GA, PSO, BP and LM for training feed forward neural networks in e-learning context. *Int. J. Intell. Syst. Appl.* **2012**, *4*, 23. [CrossRef]
16. Liu, W.C.; Chung, C.E. Enhancing the predicting accuracy of the water stage using a physical-based model and an artificial neural network-genetic algorithm in a river system. *Water* **2014**, *6*, 1642–1661. [CrossRef]
17. Cheng, C.T.; Niu, W.J.; Feng, Z.K.; Shen, J.J.; Chau, K.W. Daily reservoir runoff forecasting method using artificial neural network based on quantum-behaved particle swarm optimization. *Water* **2015**, *7*, 4232–4246. [CrossRef]
18. Cao, C.; Song, S.; Chen, J.; Zheng, L.; Kong, Y. An Approach to Predict Debris Flow Average Velocity. *Water* **2017**, *9*, 205. [CrossRef]
19. Wu, C.L.; Chau, K.W.; Li, Y.S. Methods to improve neural network performance in daily flows prediction. *J. Hydrol.* **2009**, *372*, 80–93. [CrossRef]
20. Wu, C.L.; Chau, K.W.; Fan, C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *J. Hydrol.* **2010**, *389*, 146–167. [CrossRef]
21. Kisi, O.; Cimen, M. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J. Hydrol.* **2011**, *399*, 132–140. [CrossRef]
22. Budu, K. Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. *J. Hydrol. Eng.* **2013**, *19*, 1385–1400. [CrossRef]
23. Fu, Q.; Liu, D.; Li, T.; Cui, S.; Hu, Y. EMD-RBFNN Coupling Prediction Model of Complex Regional Groundwater Depth Series: A Case Study of the Jiansanjiang Administration of Heilongjiang Land Reclamation in China. *Water* **2016**, *8*, 340. [CrossRef]
24. Zhao, X.; Chen, X.; Xu, Y.; Xi, D.; Zhang, Y.; Zheng, X. An EMD-Based Chaotic Least Squares Support Vector Machine Hybrid Model for Annual Runoff Forecasting. *Water* **2017**, *9*, 153. [CrossRef]
25. Nourani, V.; Baghanam, A.H.; Adamowski, J.; Kisi, O. Applications of hybrid wavelet–Artificial Intelligence models in hydrology: A review. *J. Hydrol.* **2014**, *514*, 358–377. [CrossRef]
26. Gilles, J. Empirical wavelet transform. *IEEE Trans. Signal Proc.* **2013**, *61*, 3999–4010. [CrossRef]
27. Hu, J.; Wang, J. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy* **2015**, *93*, 1456–1466. [CrossRef]
28. Wang, J.; Hu, J. A robust combination approach for short-term wind speed forecasting and analysis—Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy* **2015**, *93*, 41–56.
29. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A. Multi-verse optimizer: A nature-inspired algorithm for global optimization. *Neural Comput. Appl.* **2016**, *27*, 495–513. [CrossRef]
30. Faris, H.; Aljarah, I.; Mirjalili, S. Training feedforward neural networks using multi-verse optimizer for binary classification problems. *Appl. Intell.* **2016**, *45*, 322–332. [CrossRef]
31. Faris, H.; Hassonah, M.A.; Ala'M, A.Z.; Mirjalili, S.; Aljarah, I. A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Comput. Appl.* **2017**, *28*, 1–15. [CrossRef]
32. Daubechies, I. *Ten Lectures on Wavelets*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992.

33. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **2015**, *520*, 224–243. [CrossRef]

34. Chen, L.; Singh, V.P.; Guo, S. Copula entropy coupled with artificial neural network for rainfall–runoff simulation. *Stoch. Environ. Res. Risk Assess.* **2014**, *28*, 1755–1767. [CrossRef]

35. Zhang, C.; Zhou, J.; Li, C.; Fu, W.; Peng, T. A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *143*, 360–376. [CrossRef]

36. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P.; Parasuraman, K. River stage forecasting using wavelet packet decomposition and machine learning models. *Water Resour. Manag.* **2016**, *30*, 4011–4035. [CrossRef]

37. Wang, H.; Zhao, W. Arima model estimated by particle swarm optimization algorithm for consumer price index forecasting. In *International Conference on Artificial Intelligence and Computational Intelligence*; Springer: Berlin, Germany, 2009.

38. Deo, R.C.; Tiwari, M.K.; Adamowski, J.F.; Quilty, J.M. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1–30. [CrossRef]

39. Shafaei, M.; Kisi, O. Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models. *Neural Comput. Appl.* **2016**, *27*, 1–14. [CrossRef]

40. Ahila, R.; Sadasivam, V.; Manimala, K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. *Appl. Soft Comput.* **2015**, *32*, 23–37. [CrossRef]