An Ensemble Decomposition-Based Artificial Intelligence Approach for Daily Streamflow Prediction

Mohammad Rezaie-Balf [1], **Sajad Fani Nowbandegani** [1,*], **S. Zahra Samadi** [2], **Hossein Fallah** [1] **and Sina Alaghmand** [3,*]

1 Department of Civil Engineering, Graduate University of Advanced Technology-Kerman, Kerman P.O. Box 76315-116, Iran; moe.rezaie69@gmail.com (M.R.-B.); hosein.fallah.0511@gmail.com (H.F.)
2 Department of Civil and Environmental Engineering, University of South Carolina, Columbia, SC 29208, USA; samadi@cec.sc.edu
3 Department of Civil Engineering, Monash University, 23 College Walk, Clayton, VIC 3800, Australia
* Correspondence: sajad.fani68@gmail.com (S.F.N.); Sina.Alaghmand@monash.edu (S.A.); Tel.: +61-3-9905-4679 (S.A.)

check for updates

Abstract: Accurate prediction of daily streamflow plays an essential role in various applications of water resources engineering, such as flood mitigation and urban and agricultural planning. This study investigated a hybrid ensemble decomposition technique based on ensemble empirical mode decomposition (EEMD) and variational mode decomposition (VMD) with gene expression programming (GEP) and random forest regression (RFR) algorithms for daily streamflow simulation across three mountainous stations, Siira, Bilghan, and Gachsar, in Karaj, Iran. To determine the appropriate corresponding input variables with optimal lag time the partial auto-correlation function (PACF) and auto-correlation function (ACF) were used for streamflow prediction purpose. Calibration and validation datasets were separately decomposed by EEMD that eventually improved standalone predictive models. Further, the component of highest pass (IMF1) was decomposed by the VMD approach to breakdown the distinctive characteristic of the variables. Results suggested that the EEMD-VMD algorithm significantly enhanced model calibration. Moreover, the EEMD-VMD-RFR algorithm as a hybrid ensemble model outperformed better than other techniques (EEMD-VMD-GEP, RFR and GEP) for daily streamflow prediction of the selected gauging stations. Overall, the proposed methodology indicated the superiority of hybrid ensemble models compare to standalone in predicting streamflow time series particularly in case of high fluctuations and different patterns in datasets.

Keywords: streamflow; ensemble empirical mode decomposition (EEMD); variational mode decomposition (VMD); artificial intelligence (AI) approach; mountainous watershed

1. Introduction

Developing an efficient predictive technique for both long- and short-term streamflow is a challenge in hydrology which is crucial for resource planning and management. This is because streamflow is influenced by various dynamic nonlinear processes, such as rainfall, runoff yield and confluence, evaporation, topography, and anthropic activities (e.g., [1]). In addition, streamflow forecasting has attracted more attention because of reservoir operations and irrigation management decisions. Over the past decades, researchers have carried out different attempts to forecast monthly streamflow. Streamflow forecasting models consist of two general categories: process-driven and data-driven. Process-driven models are based on empirical formulas that focus on principal physical processes of the hydrological cycle [2].

However, AI models which are fundamentally "black-box" avoid analyzing the physical processes. In other words, these models are more user-friendly and can be used without accurate knowledge of the physical mechanisms of the watershed system [3]. Thus, AI modeling has been widely used for streamflow forecasting in recent years because of the availability of long-term gauging data, the ever-increasing computational power. For example, Chitsaz et al. [4] proposed a pre-processing based model to forecast monthly streamflow. They considered three data-driven models, generalized regression neural network (GRNN), multi-layer perceptron (MLP), and adaptive neuro fuzzy inference system (ANFIS) and pre-processing techniques, such as principal component analysis (PCA), singular value decomposition (SVD) and average values of inputs for their aims. The results revealed that SVD highly improved forecasting values and was effective as a pre-processing method in hydrological simulation. A hybrid technique was presented by Humphrey et al. [5] which integrated hydrological model outputs (GR4J conceptual rainfall–runoff model) into a Bayesian artificial neural network to forecast streamflow. The results indicated that hybrid models outperform the standalone models, particularly in term of high flows. Another attempt was conducted by Abdollahi et al. [6] to predict daily mean streamflow of perennial and non-perennial rivers in semi-arid region of Zagros Mountains in Iran using standalone and hybrid models. The applicability of proposed models was evaluated applying several performance metrics. They concluded that gene expression programming (GEP) integrating wavelet transform (WT) was the most accurate in predicting streamflow. Yuan et al. [1] investigated the accuracy of hybrid long short-term memory (LSTM) for which its parameters were optimized by the ant lion optimizer (ALO) algorithm in forecasting the monthly streamflow. The results proved that the historical monthly streamflow was predicted more accurately when using LSTM-ALO compared to other models. Mehr [7] presented an integrated predictive model for streamflow forecasting in intermittent streams. He implemented the GEP model combined genetic algorithm (GA) in which sub-expression trees of the best evolved GEP model were rescaled by appropriate weighting coefficients through the use of GA optimizer. The optimum input variables were identified by auto-correlation and partial auto-correlation functions of the streamflow records. The outcomes showed that the GEP based genetic algorithm outperforms genetic programming, GEP, multiple linear regression and GEP-linear regression as benchmarks models.

In computer science, AI is demonstrated by machines, in contrast to the natural intelligence represent by animals and humans. In recent years, AI techniques have been increasingly used to solve a large number of environmental and water engineering problems. These include evolutionary polynomial regression (EPR) [8,9], ANFIS [10,11], gene expression programming (GEP) [12,13], model tree (MT) [14,15], support vector machine (SVM) [16–18], and extreme learning machine (ELM) [8,19]. Various researches have also used AI approaches particularly for river flow forecasting [20–24]. However, developing a sophisticated model to forecast flow is a challenge since many (nonlinear) factors control rainfall–runoff processes in the catchment. Thereby, using the raw data for modeling directly may not provide permissible results, but applying a pre-processing method may improve model performance [25].

Recently, a vast number of data pre-processing methods have been developed, such as continuous wavelet transform (CWT) [8,26], maximum entropy spectral analysis (MESA) [27], wavelet multi-resolution analysis (WMRA) [28], PCA [29,30], singular spectrum analysis (SSA) [31,32], and moving average (MA) [33], which enhanced model predictability. Moreover, noise-assisted data analysis method namely ensemble empirical mode decomposition (EEMD) have been pioneered by Wu and Huang [31]. It is an advanced form of the traditional decomposition approach (e.g., wavelet and Fourier decompositions), and is an empirical, intuitive, and self-adaptive data processing method developed for non-stationary, nonlinear, and signal sequences [34]. The VMD is another non-stationary signal decomposition technique which is presented to overcome the lack of exact mathematical model [35–37].

Many EEMD and variational mode decomposition (VMD) based algorithms have been applied in hydrology to predict rainfall and runoff records [38–41]. The work of Napolitano et al. [42] was the first

attempt in using artificial neural networks (ANNs) based on EEMD for daily streamflow prediction. In the sequel, Wang et al. [43] coupled EEMD-SVM approach as an adaptive data analysis method with particle swarm optimization to decompose annual rainfall time series records for annual streamflow prediction. Soon after, Li et al. [44] proposed a hybrid prediction model based on VMD-ELM to improve the capability of the monthly precipitation prediction model.

Appling the hybrid approaches for daily streamflow forecasting has reached enormous popularity since it takes the advantages of different methods. Consequently, the purpose of this study was to employ two-phase decomposition based random forest algorithm and gene expression programming for daily streamflow prediction at three stations, Siira, Bilghan, and Gachsar, in Karaj basin, Iran. This work is one of the first attempts known to the authors that used the ensemble empirical mode decomposition as the primary decomposition technique, to enhance the prediction of daily streamflow records with random patterns and fluctuations. So, to diminish noise, trend, and stochastic behaviors of the data, an additional decomposition approach (i.e., VMD) is proposed and incorporated. This additional decomposition procedure is an iterative process. The proposed integrated models are benchmarked and evaluated with standalone models (GEP and random forest regression (RFR)) using several well-established performance's criteria. Finally, the diagnostic analyses are adopted to show how daily streamflow forecasting of the ensemble decomposition based models can improve model fidelity more intuitively.

The rest of the present research has been organized as follows: Section 2.1 to Section 2.4, the methodology of EEMD and original forecasting models (i.e., GEP and RFR) are introduced. Sections 2.5 and 2.6 include a description of the applied statistical performance benchmarks and uncertainty analysis for evaluating model performance. Section 2.7 describes the data and study areas. Section 3 compares the results of the proposed models, and finally in Section 4, a summary and conclusion is provided.

2. Methodology and Data Description

2.1. Gene Expression Programming

One of the most common AI models is GEP, which is inspired by Darwin's theory and developed form of genetic programming technique. GEP evolves computer programing in different forms consisting decision trees and logical and mathematical expressions [45–47]. Moreover, the GEP approach has high performance in solving a large number of problems in different fields especially water engineering [48–51]. Here, mathematical equations based on the GEP model have been applied to predict daily streamflow records in Karaj basin, Iran. Gene expression programming is coded as linear chromosomes, which is represented in form of expression trees (ETs). In point of fact, ETs are complicated computer programs which are evolved for solving practical problems and are determined based on their fitness at solving those problems. The corresponding mathematical expressions can also be extracted from these tree structures. An ETs population can discover traits. Hence, it will adapt to the specific problem which they are used to solve [46]. GEP development involves five steps. Determine the fitness function (*fi*) for each individual program (*i*) is the first step, which is computed as:

$$f_i = \sum_{j=1}^{C_i} \left(M - \left| C_{(i,j)} - T_j \right| \right) \tag{1}$$

where M, T_j, and $C_{(i,j)}$ are, respectively, the range of selection range, the maximum value for fitness case j, and provided value using the individual chromosome i in fitness case j. Second, the function and terminals set are determined for chromosomes generating. It should be noted that peer reviewing the prior investigations is necessary to find a suitable function set. Hence, in this study, several basic operators $(+, -, \times, /)$ and mathematical functions ($\sqrt{}$, power, exp) have been employed to model daily streamflow data. Third, the chromosomal architecture is configured. The liking function and the

genetic operator's sets are selected in the fourth step and case variation and their rate are determined in the final step. Figure 1a illustrates a simple tree representation of classical GEP. More details about GEP and its architecture have been provided in the various literatures, such as Ferreira [46] and Shahrara et al. [13].
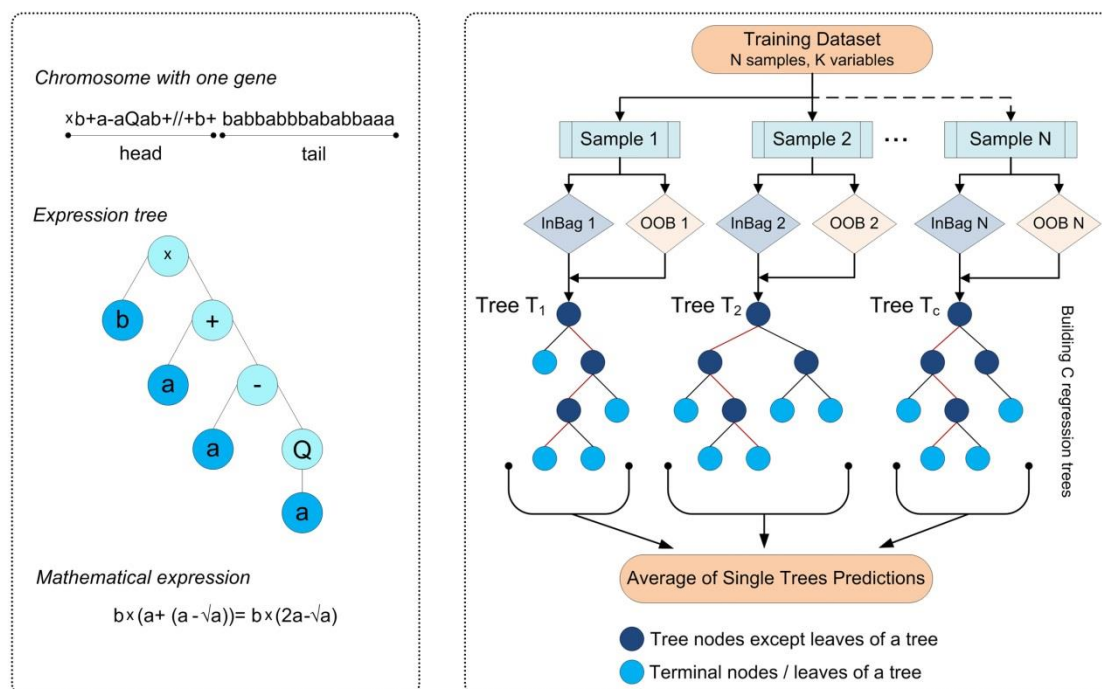

Figure 1. The topological structure of (**a**) gene expression programming (GEP) and (**b**) random forest regression (RFR) models.

2.2. Random Forest Regression

RFR is a state-of-the-art hierarchical algorithm proposed by Breiman [52] that computes the appropriate relationship between input and output parameters. It is a learning algorithm of tree-based ensemble, which during the model building process, raises various decision trees and each of them in the ensemble model is trained by bootstrapped sample of the main input dataset. The output estimation values are then computed using the average of these estimation trees. The proposed RFR approach is illustrated in Figure 1b with a specific algorithmic procedure introduced by Svetnik et al. [53] and Cootes et al. [54] as follows:

(1) Use the original training dataset X (N samples) to draw k samples randomly by the bootstrap resampling technique to construct k regression trees. In this process, the probability related to the samples that would not be drawn can be computed by $p = (1 - 1/N)^N$. If N achieved to infinity, $p \approx 0.37$ which expresses that roughly 37% of the samples of original training dataset X are not drawn and these data are known as out-of-bag (OOB) data. Likewise, the training dataset, these OOB data can be applied for testing samples.

(2) Moreover, unpruned regression trees corresponding to k bootstrap samples are created. During the growing process of trees, in each node, a attribute is randomly considered from all A attributes (input parameters) as internal nodes ($a < A$). Then, based on the principle of minimum Gini index (a measure of how each variable contributes to the homogeneity of the nodes and leaves), an optimum attribute is determined from an attribute as a split variable to build the branches hierarchy.

(3)   The final random forest regression model is constituted by generated k regression trees. To evaluate the model estimation performance, two indices namely coefficients of determination (R^2_{RFR}) and mean square error of OOB (MSE_{OOB}) are employed.

$$MSE_{OOB} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{2}$$

$$R^2_{RFR} = 1 - \frac{MSE_{OOB}}{\hat{\sigma}^2_y} \tag{3}$$

where n is the total OOB samples, y_i and $\hat{y}_i$ are the observed and predicted output values, respectively, and $\hat{\sigma}^2_y$ is the OOB variance of predicted output.

2.3. Ensemble Empirical Mode Decomposition

EEMD as an improvement of EMD is a type of mathematical function and data pre-processing technique which represents a non-stationary and nonlinear signal from original dataset, expressed by Wu and Huang [31]. One of the EEMD features is decomposing the signals of the original dataset to a small and the finite length of oscillatory modes according to the local properties' time scales [55]. The oscillatory modes are described by components of intrinsic mode functions (IMFs) where embedded in the data. EMD is a self-adaptive time-frequency approach, without elementary information about the nature that its IMF indicates a signal as a sum of mean zero well-behaved both slow and fast oscillation modes related to IMFs given by two principles [31,55]. These include: (i) the number of extrema and zero-crossings must either be equal or at least different from one throughout the whole length and (ii) for every point, the mean value of the local minimum and maximum envelope is equal to zero. Generally, an IMF indicates a simple oscillatory mode in comparison with the simple harmonic function. According to the definition, a shifting process of original daily streamflow time series dataset can be briefly described throughout six steps as stressed by Huang et al. [55].

Step 1: Determine all local minima and maxima (extrema) points of a given time series $y(t)$;

Step 2: Connect local extrema points for forming the bottom and upper envelopes ($e_{max}(t)$ and $e_{min}(t)$) with spline interpolation, respectively;

Step 3: Determine the mean $m(t)$ between two maxima and minima envelopes as follow:

$$m(t) = (e_{max}(t) + e_{min}(t))/2 \tag{4}$$

Step 4: Diminish the average value $m(t)$ from the data to determine an IMF candidate as follow: $h(t) = y(t) - m(t)$.

Step 5: If $h(t)$ meets two IMF properties based on the predefined stopping benchmarks, $h(t)$ is defined as the first IMF (written as $c_1(t)$ which 1 is its index); else, $y(t)$ is replaced with $h(t)$ and repeat steps 1 to 4 until $h(t)$ meet the two IMF properties.

Step 6: The residual $r_1(t) = y(t) - c_1(t)$ is defined as new dataset subjected to the same shifting proceeding as explained for other IMFs from $r_1(t)$. In this step, the shifting processing can be ceased, if the $r(t)$ has a steady trend or has a local extrema point from no IMF may be provided [56]. Finally, in this method, the y is created from the aggregation of intrinsic mode functions and residual as:

$$y(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{5}$$

where n is the IMF number, $r_n(t)$ is the final $r(t)$, and $c_i(t)$ are closely orthogonal together, and all of them have mean zero. More informative about the EMD technique and the ceasing benchmarks have been provided by Huang et al. [55,56]. Hence, evaluations show EMD is still unstable because of the mode mixing [31]. Mode mixing is defined as either a single IMF including a wide range of disparate scales

components or a similar scale component residing in various IMFs [57]. For overcoming the mode mixing problem in EMD, the EEMD is modified as an improved technique. Regarding this argument, the whole space of time-frequency becomes uniformly filled by adding the white noise which can simplify a normal divorce of the frequency scales and decay the mode mixing event.

2.4. Variational Mode Decomposition

The main aim of this study is developing and investigating a two-layer integrated predictive approach. Therefore, variational mode decomposition (VMD) as a pre-processing technique for signal decomposition was applied [58]. The VMD algorithm has an ability in decomposing a signal $x(t)$ to K discrete sub-signals number. The VMD is employed as a constrained optimization issue, expressed by Dragomiretskiy and Zosso [58]:

$$f = \min_{\{U_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times U_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \; subject \; to \; \sum_k U_k \tag{6}$$

where $\{U_k\} : \{U_1, U_2, \ldots, U_K\}$ and $\{\omega_k\} : \{\omega_1, \omega_2, \ldots, \omega_K\}$ indicate stenography, impression related to all modes set and their central frequency, respectively. Moreover, $\delta(t)$ and $\times$ denotes the Dirac distribution and convolution, respectively. So, the Lagrangian multipliers and term of quadratic penalty are defined to convert this constrained optimization issue to an unconstrained problem [38]:

$$L(\{U_k\}, \{\omega_k\}, \lambda) = \alpha \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) \times U_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} + \left\| f(t) - \sum_k U_k(t) \right\|_2^2$$
$$+ \left\langle \lambda(t), f(t) - \sum_k U_k(t) \right\rangle \tag{7}$$

Equation (7) can be resolved with alternative approaches. In this way, Equation (7) has been shown as two stages,

(1) U_k Minimization:

$$\hat{U}_k^{n+1} = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{U}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{8}$$

(2) ω_k Minimization:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{U}_k(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{U}_k(\omega) \right|^2 d\omega} \tag{9}$$

where $\hat{U}_i(\omega)$, $\hat{U}_k^{n+1}$, $\hat{f}(\omega)$, $\hat{\lambda}(\omega)$ indicate the Fourier transform of $U_i(\omega)$, U_k^{n+1}, f(ω), λ(ω), respectively, and n is the iterations number. The use of VMD diminishes remaining noises in the modes and avoids the redundant modes compared to EEMD approach. Moreover, VMD decomposes a multi-elements signal to some mode functions of the quasi-orthogonal intrinsic [59]. During VMD and EEMD combination, resolving the patterns of low frequency which exist in the IMF1 signal is the main goal of applying different AI methods. This helps the model to be more responsive to the patterns of the fine-scale dataset that is important for simulating inconsistencies in the dataset.

2.5. Model Assessment Criteria

Several standard statistical criteria are used in the present research for evaluating the performance of the modeled and observed streamflow. In addition to common criterion, such as root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE), and mean absolute error (MAE), the indices below were applied to assess the fidelity of predicted model.

1. Ratio of RMSE to standard deviation (RSD): RSD, proposed by Singh et al. [60] is a model evaluation metric to assess the differences between a model's prediction and the observed data

in hydrological simulation. This metric is calculated based on RMSE and standard deviation (STDEV) of the observed data points. The lower the value of the RSD the higher the performance of the model.

$$RSD = \frac{RMSE}{STDEV_{obs}} = \frac{\left[\sqrt{\sum_{i=1}^{N}\left(Q_{obs} - Q_{for}\right)^2}\right]}{\left[\sqrt{\sum_{i=1}^{N}\left(Q_{obs} - \overline{Q_{obs}}\right)}\right]} \tag{10}$$

2. Uncertainty at 95% (U95): U95 is defined as a 95% uncertainty confidence.

$$U_{95} = 1.96\sqrt{\left(STDEV^2 + RMSE^2\right)} \tag{11}$$

3. Reliability of model (%): This statistic shows the satisfactory state of the model's forecast by the probability [19].

$$Reliability = \frac{\sum_{i=1}^{N} K_i}{N} \times 100\% \tag{12}$$

$$K_i = \begin{cases} 1, & if\,(RAE_i \le \delta) \\ 0, & else \end{cases} \tag{13}$$

$$RAE_i = \frac{\left|Q_{for}(i) - Q_{obs}(i)\right|}{Q_{obs}(i)} \times 100\%, \quad RAE_i \ge 0 \tag{14}$$

4. Resilience of model (%): This indicator defines how quickly the model forecast is likely to recover once an unqualified forecast has occurred [61].

$$Resilience = \begin{cases} 100\%, & if\,(Reliability = 100\%) \\ \frac{\sum_{i=1}^{N-1} R_i}{N - \sum_{i=1}^{N} K_i} \times 100\%, & else \end{cases} \tag{15}$$

$$R_i = \begin{cases} 1, & if\,(RAE_i > \delta\ and\ RAE_{i+1} \le \delta) \\ 0, & else \end{cases} \tag{16}$$

where Q_{obs} and Q_{for} denote the measured and estimated values, respectively; $\overline{Q_{obs}}$ is the average of observed values, and N is the number of datasets. RAE_i is the value for the *ith* data, K_i is the number of times that the threshold value (δ) of the qualified forecast is greater than or equal to the RAE value. δ is set to 20% according to the Chinese standard (GB/T 22482-2008). R_i is the number of times that model forecast is likely to transfer from unqualified into qualified forecast in the *i*th data.

2.6. Uncertainty Analysis

A quantitative evaluation of the uncertainty in streamflow prediction is carried out using standalone and hybrid models. The uncertainty analysis is used to validate the dataset for each gauging stations. It should be noted that daily streamflow time series may contain uncertainty. For example, errors in streamflow data resulting from uncertainty in the stage-flow rating curve may affect the simulation results. This happens especially during repeated rainfall events when the measured values are extrapolated outside the range of the established rating curve [62]. In this study, a quantitative uncertainty evaluation of streamflow predicting rate E is computed by using EEMD-VMD based GEP and RFR models. The uncertainty analysis is performed for lag values of daily streamflow. The individual error of prediction Equation (17) is computed for all datasets to compute mean ($\overline{e}$) and standard deviation (S_e) of estimating error using Equations (18) and (19) [63].

$$e_i = \log_{10}(Q^{pre}_i) - \log_{10}\left(Q^{obs}_i\right) \tag{17}$$

$$\bar{e} = \sum_{i=1}^{n} e_i \tag{18}$$

$$S_e = \sqrt{\sum_{i=1}^{n}\left(\frac{(e_i - \bar{e})^2}{n-1}\right)} \tag{19}$$

where n is the number of datasets, and positive and negative mean forecasting error values represent, respectively, overestimation and underestimation of measured flow. To this end, a band of confidence is expressed around forecasted error values by the Wilson score method [64]. In addition, using $\pm 1.96\ S_e$ gives about 95% confidence band surrounding predicted P_i as follow:

$$\left\{P_i \times 10^{-\bar{e}-1.96S_e}, P_i \times 10^{-\bar{e}+1.96S_e}\right\} \tag{20}$$

2.7. Case Study and Data Analysis

Iran is mostly an arid region with 242 mm annual precipitation which is equal to one third of the annual average precipitation of the world [65]. Thus, the surface water resources are not sufficient to meet all demand, and, therefore, consumers have been tapping into groundwater resources. Tehran is the capital and the most populated city in Iran where about 340×10^6 m^3 of its water demands are supplied from the Karaj reservoir located on Karaj River in Varian Strait 23 km from the city of Karaj (Figure 2) [66]. Karaj reservoir with an area of approximately 764 km^2 was constructed in the period from 1957 to 1961 to supply drinking water for Tehran residents, hydroelectric power production, and irrigation of the Karaj plain [67].
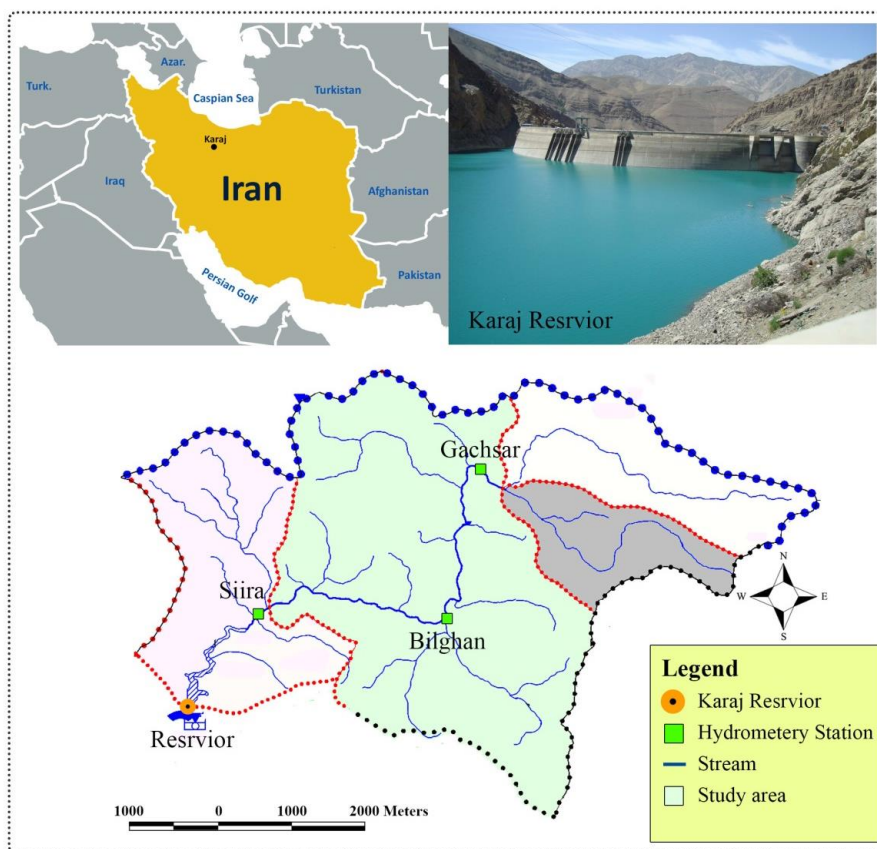

Figure 2. Location map of Karaj reservoir within the Karaj watershed.

Karaj drainage system is located between 52°2' to 51°32' E and 35°52' to 36°11' N with an area of 850 km^2 and a circumference of 146 km on the southern portion of Alborz mountain range. Average

annual flow for Karaj Dam is 472 million cubic meters. The maximum elevation of the watershed is about 4200 m, and the lowest is 1600 m. The average daily discharge is estimated at 154.54 m³/s. In this study, three hydrometric stations were selected, namely Gachsar, Bilghan, and Siira. Details of datasets and the parameters used for the proposed models are given in Table 1.

Table 1. Descriptive statistics of river flow data for the study site located at the Karaj basin.

Station	Crd	El (m)	X_{min} (m³/s)	X_{max} (m³/s)	X_{mean} (m³/s)	S_x	C_v	C_{sx}
Siira	36°01′ N 51°09′ E	1898	0	160	11.8	12.54	157.35	2.64
Bilghan	36°00′ N 51°17′ E	2138	0	194	15.81	14.04	197.17	3.21
Gachsar	36°06′ N 51°19′ E	2258	0	61	3.93	3.99	15.96	2.34

Crd, El, X_{min}, X_{max}, X_{mean}, S_x, C_v, and C_{sx} denote the coordinates, elevation, minimum flow, maximum flow, mean flow, standard deviation, variation coefficient, and skewness coefficient, respectively.

In addition, Figure 3a–c display the frequency of observed daily streamflow records using normal fit distribution to show the distribution of data by forming bins along the range of the data and then drawing bars to show the number of observations that fall in each bin. The histograms of the streamflow for all three stations follow a bell shape and indicate that the time series records follow a normal distribution. Time series data for Siira, Bilghan, and Gachsar stations were available, respectively, since 1954–1998, 1947–1993, and 1982–2001. These datasets were used to calibrate the proposed models. 1999–2008, 1994–2008, and 2001–2008 were used for validation.
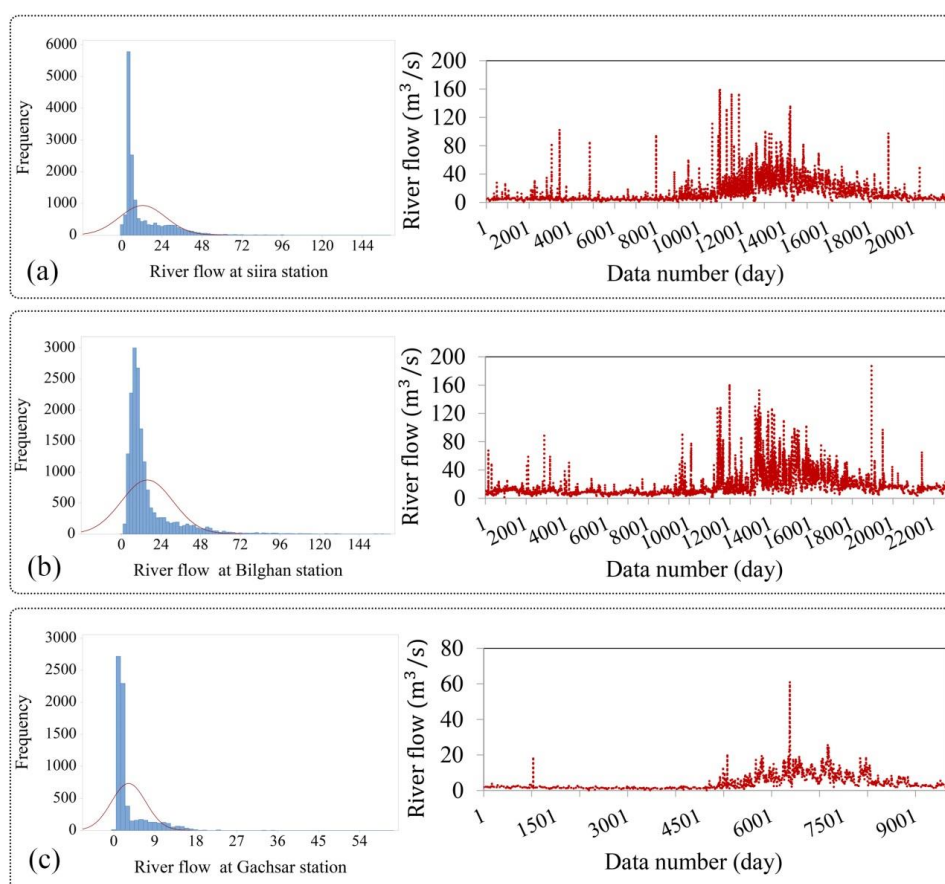

Figure 3. (**Left**) histogram with normal distribution overlay (red), and (**right**) time variation for the observed river flows; (**a**) Siira station, (**b**) Bilghan station, and (**c**) Gachsar station.

3. Results and Discussion

A hybrid ensemble model (e.g., EEMD-VMD-GEP and EEMD-VMD-RFR) is first coded and implemented to forecast daily streamflow with a long time dataset series at the three selected gauging stations. To achieve more reliable and accurate model, outliers which may be recorded by a faulty device, typos, etc., are identified and replaced with the average of previous and subsequent reliable records. The results of each model were then computed and evaluated by several model performance metrics in the calibration and validation periods. For better understanding, the results of the three stations have been evaluated separately which have been discussed below.

3.1. Input Variables Selection and Model Development

To simulate daily streamflow records in the selected stations, a lag time was applied to each station to define the appropriate dataset that correspond to the streamflow records at the outlet. We applied a partial auto-correlation function (PACF) and an auto-correlation function (ACF) value to the time series to determine the corresponding datasets. Figure 4 represents the ACF and PACF analyses of the daily streamflow data for each application. Based on this analysis, three lags of 1 to 7, 6, and 4 days are considered, respectively, for Siira, Bilghan, and Gachsar gauging stations. That is, the values of lag (red lines in Figure 4) which are upper or lower than $\pm 1.96/\sqrt{n}$ should be considered as the effective input variables for streamflow forecasting.
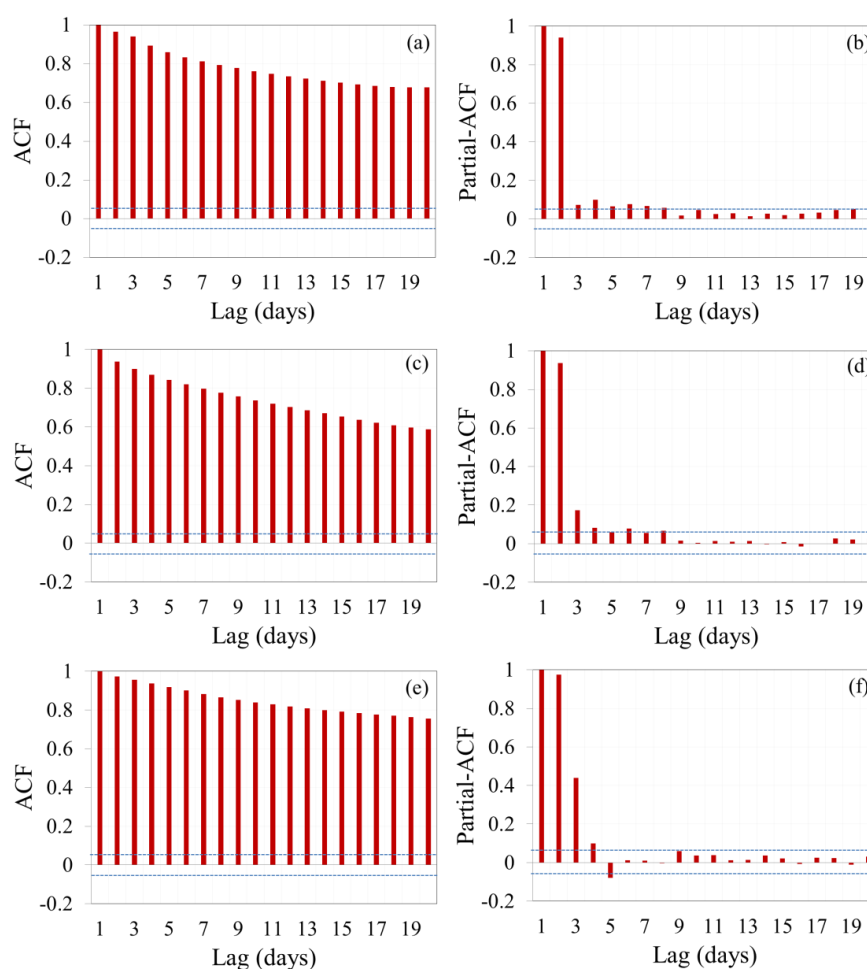

Figure 4. The auto-correlation function (ACF) and partial auto-correlation function (PACF) graphs of the daily streamflow for Siira (**a**,**b**), Bilghan (**c**,**d**), and Gachsar (**e**,**f**) stations used to develop predictive models.

The monotonic trends in daily streamflow are analyzed by Spearman's rank correlation test which is one of the nonparametric techniques and quantitatively assessed how strongly the two data sets are related to each other. This approach with uniform power for linear and non-linear trends is widely applied to verify the absence of trends. In the present study, the null hypothesis of the analysis is that two variables are independent and identically distributed (*iid*) and the alternative hypothesis is that increasing or decreasing trends may exist. As demonstrated in Table 2, correlations between the output (present streamflow) and input (antecedent values of streamflow) variables for all three stations are evaluated by the Spearman's rank correlation test which is performed using SpSS-18. Furthermore, P values <0.05 are selected as the confidence level for the correlation test.

Table 2. Values of the correlation coefficient between the daily streamflow and the individual input variables.

Station	Parameter Value	Input Variable						
		Qs(t−1)	Qs(t−2)	Qs(t−3)	Qs(t−4)	Qs(t−5)	Qs(t−6)	Qs(t−7)
Siira	R	0.97**	0.95**	0.94**	0.92**	0.89**	0.86**	0.82**
	P-value	0.00	0.00	0.00	0.00	0.002	0.005	0.008
Bilghan	R	0.97**	0.95**	0.93**	0.91**	0.88**	0.84**	-
	P-value	0.00	0.00	0.00	0.001	0.002	0.004	-
Gachsar	R	0.95**	0.92**	0.88**	0.82**	-	-	-
	P-value	0.00	0.00	0.002	0.006	-	-	-

** Correlation is significant at the 0.05 level.

Spearman's rank correlation coefficients for the input variables at three stations are more than 0.8, and the P-values are less than 0.05 (confidence level). Therefore, the null hypothesis that the two populations are independent is rejected at a level of significance of 5 % and the output variable (daily streamflow) is judged to be dependent on input variables.

Furthermore, one of the important analyses that confirms that the hydrological time series are not affected by non-climatic factors, such as station environment, station location, observation practices, and instruments, is the homogeneity test. In hydro–climatological studies, this non-parametric test is considered as a critical quality control method to detect the variability of climatology dataset [68]. Generally speaking, having completely homogeneous data due to non-climatic causes that occur by the changes in the time-series and the obvious nonstationarity in the study area (changes in metadata, etc.) surrounding a meteorological station can be disregarded as neglectable.

Therefore, the temporal homogeneity of time-series is required in hydrological modeling. In this regard, the homogeneity of the streamflow data in three hydrometric stations is analyzed using two absolute homogeneity tests, the Neumann ratio test and Pettitt test, that are mostly used in the hydro–climate nonstationary test [68–71]. The annual mean as the testing variable is considered in Siira, Bilghan, and Gachsar flow stations because of the presence of the missing values on daily streamflows. The streamflow series of each station were computed for a significance level of 0.05. The annual quantities of the testing variables under the null hypothesis are independent, and the series is assumed as homogeneous. On the other hand, Pettitt test assumes the series consisted of a break in the mean and are considered as inhomogeneous, while the Von Neumann ratio test assumes that the series is not randomly distributed under the alternative hypothesis.

As defined in Table 3, the null hypothesis of the homogeneity analyses for annual mean streamflow at three stations is not rejected at 5% level of significance for the Pettitt and Van Neumann tests. Based on the criteria presented by Wijngaard et al. [71], the results are placed into the three groups (class A: (useful) zero or one rejection; class B: (doubtful) two rejections; and class C: (suspect) three or more rejections). From Table 3, all three stations fall under class A (useful) and can be used for further analysis.

Table 3. Results of homogeneity tests for annual mean streamflow records.

Station	Pettitt Test			Van Neumann Test		
	P-value	**Statistics**	**Class**	**P-value**	**Statistics**	**Class**
Siira	0.462	252	A	0.006	2.38	A
Bilghan	0.69	240	A	0.026	1.74	A
Gachsar	0.133	92	A	0.02	1.52	A

Critical values for Pettitt test are <306, <318, and <95, and Von Neumann test are <1.57, <1.59, and <137, respectively, for the Siira, Bilghan, and Gachsar stations.

For developing a two-phase hybrid, EEMD-VMD-based GEP and RFR models, streamflow records are divided into calibration (approximately 75% of the whole dataset) and validation datasets (reminder of whole dataset). Both calibration and validation datasets were, respectively, decomposed into different components of IMFs. To pre-process the data, a residual value which computes the details of high and low flow frequencies in the data was then included as the model input data series. This helped to pre-process the data and eliminate trend and pattern in time series. The decomposition of the dataset into nine IMFs and one residual component is illustrated in Figure 5.

Due to high fluctuations of oscillatory in the first IMF (IMF1), a secondary VMD-IMF1 decomposition was carried out. It should be noted that the number of intrinsic modes determination plays a vital role in the VMD process because it can define appropriately resolved time-series data for a permissible prediction model. If this process does not proceed, the specification of the hidden original time series dataset cannot be provided fully which may result in a lower component. Conversely, a large number of intrinsic modes can cause a poor performance leading to the influences of error accumulation from each prediction module in the aggregation stage as seen elsewhere [72].

Notwithstanding the above necessity, the intrinsic modes given by the VMD algorithm may be smoother compared to decomposed components by other traditional techniques, such as wavelet decomposition and the ensemble empirical mode decomposition methods as stressed by Niu et al. [73]. This mitigates the effects of error accumulation over time. To create the VMD decomposition, several parameters must be selected which are set with the trial and error procedure. By means of the VMD process, the IMF1 time-series was decomposed into eleven various components for IMF1. The decomposed time-series of the IMF1 dataset using the VMD approach is presented in Figure 5.

Finally, the predicted values community of these IMFs was applied to provide the estimated IMF1 value and the VMD-EEMD. Figure 6 presents a diagram of the developed ensemble decomposition-based proposed models. Furthermore, design parameters for both GEP and RFR models in calibration and validation periods were considered constant which are shown in Table 4.

Table 4. Design parameters of gene expression programming (GEP) and random forest regression (RFR) models for calibration and validation stages at the three gauging stations.

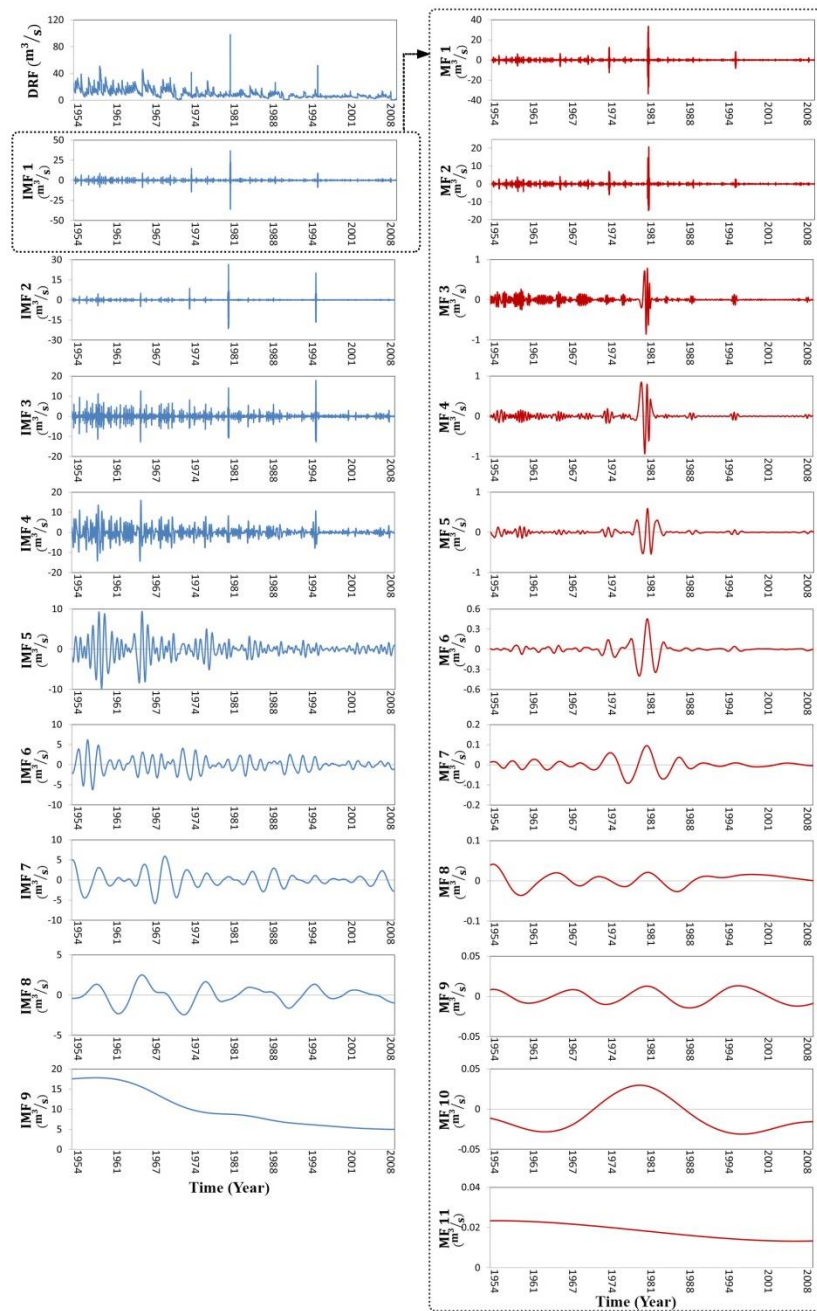| Model	Design Parameters							
GEP	Chromosomes	Gene size	HEAD SIZE	Linking Function	Mutation Rate	Crossover Rate	One and two point recombination rate	IS and RIS transposition rate
	40	3	8	Addition	0.01	0.8	0.3	0.1
RFR	Bag size percent	Leaf	Batch size	Surrogate	Delta Criterion	Number of depth	Seed	Out-of-Bag error
	200	8	100	True	0.1007	0	1	False

Figure 5. Ensemble empirical mode decomposition (EEMD) (Blue line) and variational mode decomposition (VMD) (Red line) of daily streamflow time series at the Siira station.
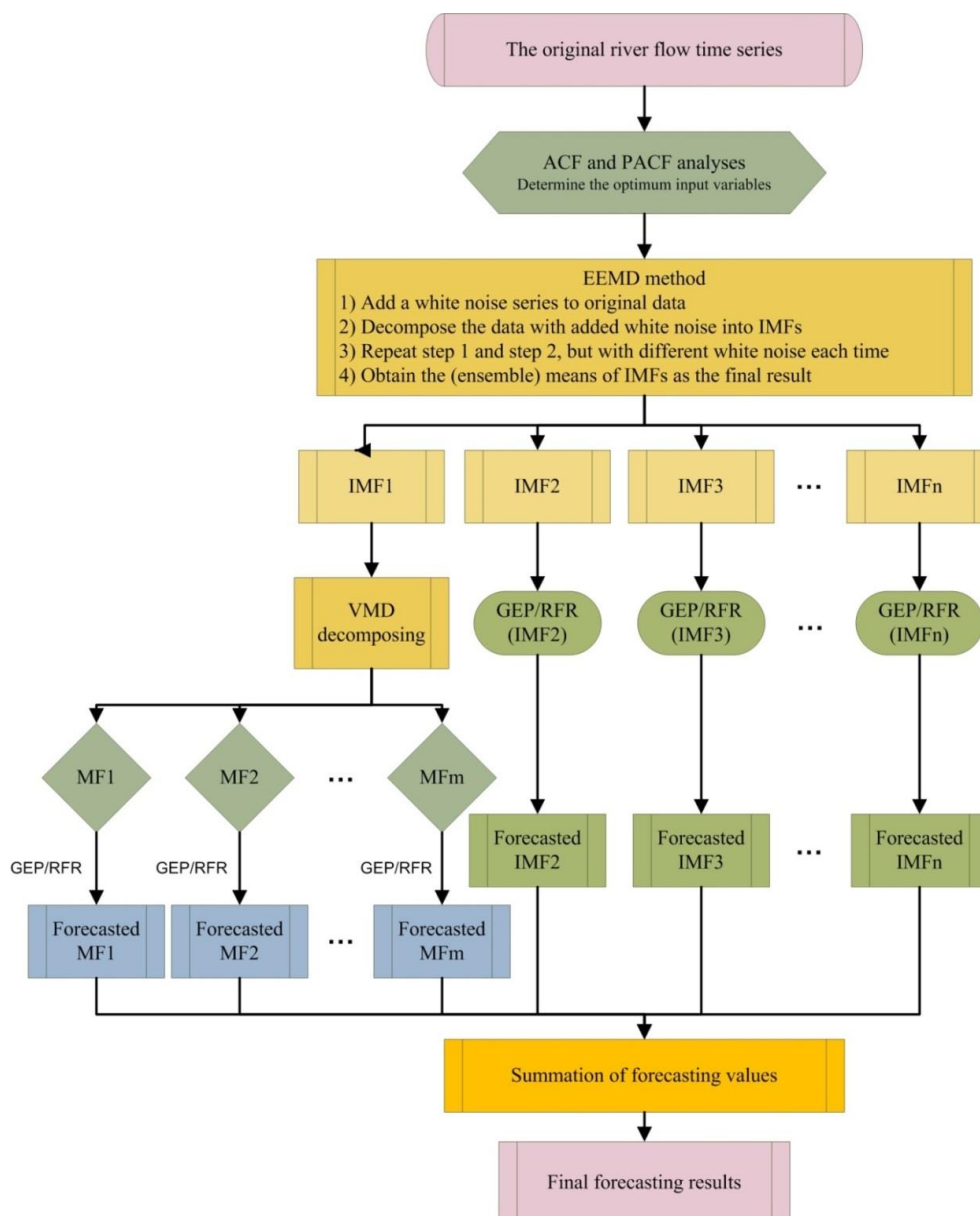
Figure 6. A flow chart of the ensemble GEP and RFR integrated with the EEMD and VMD approaches.

3.2. Application

3.2.1. The Siira Gauging Station

Results from standalone and hybrid models for forecasting daily streamflow at Siira are presented in Table 5 which the statistical metrics are applied for evaluation of the performance of GEP, RFR, EEMD-VMD-GEP, and EEMD-VMD-RFR models in calibration (1954–1994) and validation periods (1994–2008).

Performance assessment in Table 5 with respect to the statistical indices, NSE, RMSE, MAE, and RSD, indicated remarkable outcomes. In this station, the most accurate model in term of NSE was recorded from EEMD-VMD-RFR (NSE = 0.97) compared with GEP (0.92), RFR (0.93), and EEMD-VMD-GEP (0.94), while the largest error in term of RMSE, and RSD was calculated for GEP, 3.731 m^3/s and 0.269 m^3/s, respectively. In terms of MAE, the RFR revealed the least performance with 1.358, while EEMD-VMD-RFR has the best performance (MAE = 1.056 m^3/s) for which the predicted values of streamflow is well-correlated with the calibration period.

Table 5. Evaluation metrics of the proposed models in the calibration and validation periods at the Siira station.

Models	Statistical Error Indices			
	GEP	RFR	EEMD-VMD-GEP	EEMD-VMD-RFR
Total Available Data in Calibration Stage				
NSE	0.927	0.93	0.944	0.972
RMSE (m³/s)	3.731	3.482	3.26	2.3
MAE (m³/s)	1.313	1.358	1.301	1.056
RSD (m³/s)	0.269	0.251	0.235	0.166
U95	28.116	27.99	27.82	27.32
Reliability (%)	85.94	86.47	89.08	90.96
Resilience (%)	67.24	54.79	69.44	71.21
Total Available Data in Validation Stage				
NSE	0.892	0.75	0.933	0.944
RMSE (m³/s)	2.286	3.491	1.806	1.645
MAE (m³/s)	0.906	1.282	0.722	0.683
RSD (m³/s)	0.327	0.499	0.258	0.235
U95	14.403	15.303	14.13	14.06
Reliability (%)	89.49	85.68	92.69	95.88
Resilience (%)	68.44	61.674	78.47	82.43

To provide more detail in the results, three additional evaluation indices based on U95, reliability and resilience of models are considered to examine the proposed predictive models. To determine the model's deviation in term of predicted daily streamflow, U95 is calculated and that 1.96 is a coverage factor corresponding to the 95% confidence level of predicted values of the models. Therefore, the hybrid EEMD-VMD-RFR model registers a lower value of U95, of approximately 27.32% during the calibration period, whereas the reliability (90.96%) and resilience (71.21%) values of this model showed relatively larger values compared with the other three approaches in the calibration period.

The modeling results during the validation period revealed that RFR model poorly predicted streamflow records over time. RMSE is approximately 1.84 (m³/s) lower when daily streamflow is predicted using EEMD-VMD-RFR compared to the standalone RFR, and approximately 0.641(m³/s) and 0.161(m³/s) lower compared with the GEP and EEMD-VMD-GEP model, respectively (see Table 3). The Nash-Sutcliffe coefficient for EEMD-VMD-RFR showed very good performance compared with original GEP and RFR and hybrid EEMD-VMD-GEP models. Moreover, the RFR model produced the highest values of RSD (0.499 (m³/s)), and U95 (15.303%) indicating poor performance of the model for the Siira streamflow prediction. Conversely, the integration of RFR with EEMD-VMD led to improving the simulation remarkably (e.g., Resilience = 82.43% and Reliability = 95.88%).

The uncertainty analysis was also undertaken in the validation period to evaluate the predictive ability of the models. As shown in Table 6, mean predictive error for EEMD-VMD-RFR (0.172) is lower compared with the rest of AI-based models, though all proposed models overestimated daily streamflow by presenting positive values for mean predictive error. Moreover, the uncertainty band for the RFR is larger than those of the available predictive approaches, which is ±6.831.

Table 6. Uncertainty analyses for streamflow prediction using predictive models at the Siira station.

Models	Mean Prediction Error	Width of Uncertainty Band	Median	95% Predictive Error Interval
Siira Station				
GEP	0.236	±4.456	7.719	−4.22 to 4.69
RFR	0.189	±6.831	7.788	−6.64 to 7.02
EEMD-VMD-GEP	0.208	±3.511	7.695	−3.28 to 3.74
EEMD-VMD-RFR	0.172	±3.192	7.401	−2.95 to 3.42

The degree of agreement between observed and predicted streamflow at calibration and validation periods was visually evaluated using scatterplots (Figure 7). Analysis suggested that EEMD-VMD to RFR combination made a stronger correlation between predicted and observed streamflow compared to other predictive models. Further, with a comparison between modeled and observed daily streamflow indicated that the standalone RFR model performed weaker than GEP in daily streamflow prediction and is not able to capture high and low flow values as well as the GEP model (see Figure 8). However, the two-layer decomposition-based approaches (especially EEMD-VMD-RFR) revealed a very good simulation performance and were most effective in high flow prediction during validation period.
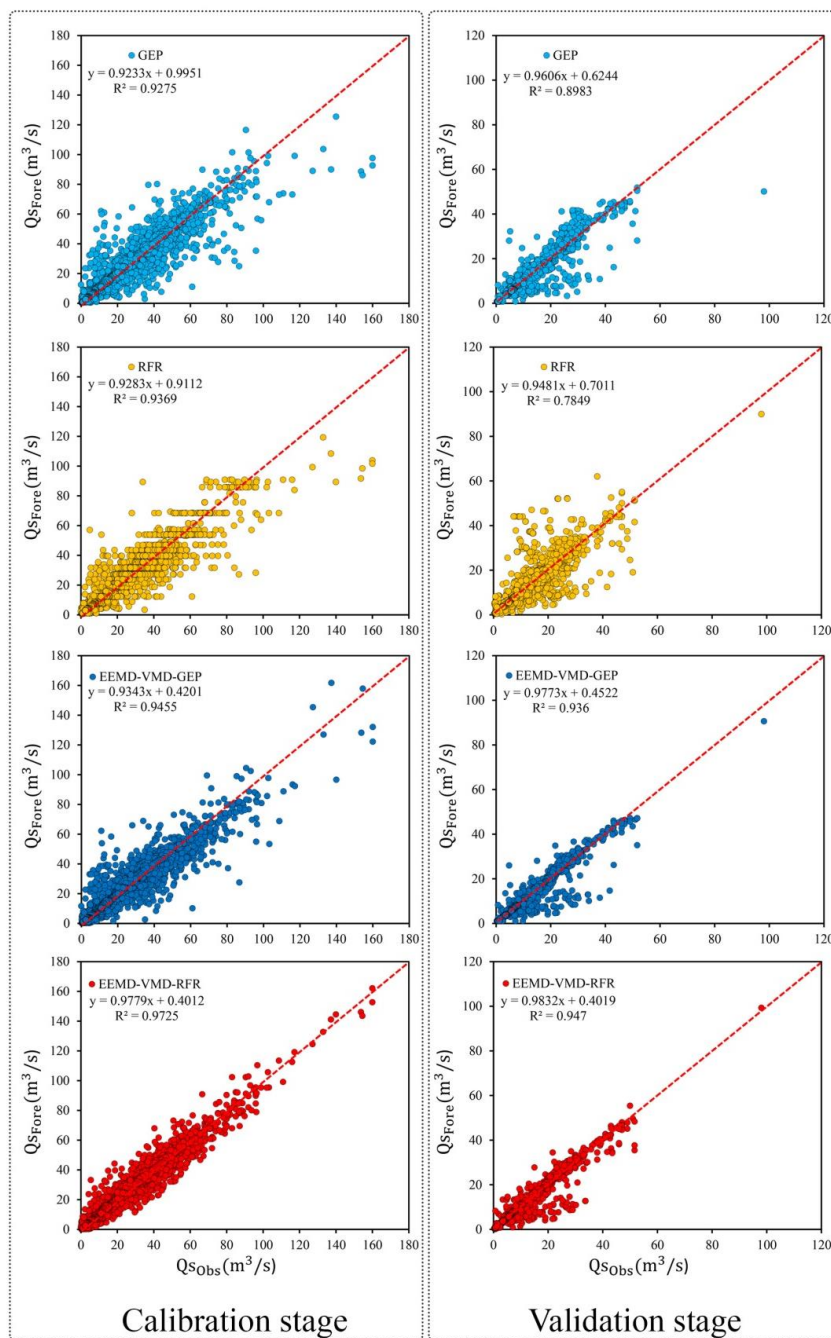
Figure 7. Scatterplot of forecasted (Qs) and observed (Qs) river flow using the standalone GEP, RFR, and ensemble EEMD-VMD-GEP, EEMD-VMD-RFR models for the calibration and validation stages at the Siira station.
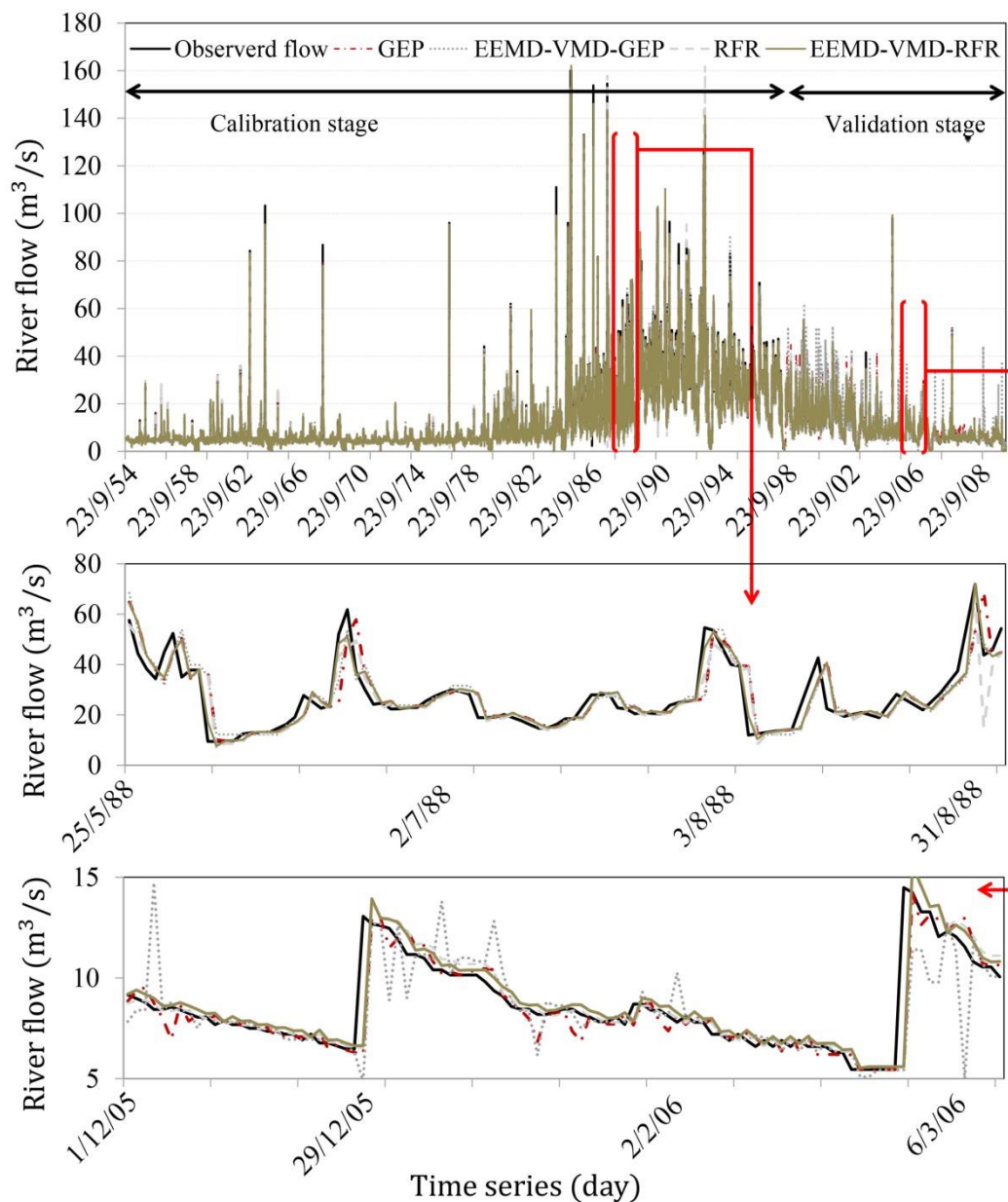
Figure 8. A time-series of observed and forecasted river daily flow of ensemble EEMD-VMD-GEP and EEMD-VMD-RFR models compared with standalone GEP and RFR models for the calibration and validation stages at the Siira station.

3.2.2. The Bilghan Gauging Station

Similar to prior application, an integration of EEMD, VMD decomposition and RFR models outperformed the best prediction in terms of performance criteria. In the calibration period, the calculated RMSE for two-phase decomposition GEP and RFR is 3.73 (m^3/s) and 1.768 (m^3/s), respectively (see Table 7). This reveals the proposed approach is excellent for daily streamflow simulation across the mountainous region. Results of standalone GEP and RFR results showed poor performances compared with hybrid decomposition based models in terms of RMSE. EEMD-VMD based GEP model ranked as the second-best AI model for daily streamflow prediction at the Bilghan station.

Table 7. Evaluation metrics of the proposed models in the calibration and validation periods at the Bilghan station.

Models	Statistical Error Indices			
	GEP	RFR	EEMD-VMD-GEP	EEMD-VMD-RFR
Total Available Data in Calibration Stage				
NSE	0.915	0.932	0.94	0.98
RMSE (m³/s)	4.55	4.063	3.73	1.768
MAE (m³/s)	1.655	1.508	2.11	0.747
RSD (m³/s)	0.291	0.261	0.24	0.11
U95	31.84	31.61	31.46	30.79
Reliability (%)	88.36	90.27	93.52	96.49
Resilience (%)	59.66	64.93	71.67	79.15
Total Available Data in Validation Period				
NSE	0.85	0.87	0.90	0.92
RMSE (m³/s)	2.704	2.61	2.358	2.061
MAE (m³/s)	1.442	1.29	1.08	1.02
RSD (m³/s)	0.357	0.33	0.311	0.272
U95	15.75	15.83	15.24	14.66
Reliability (%)	88.507	93.89	95.64	96.82
Resilience (%)	55.39	81.97	82.02	86.85

According to Table 7, this model has an acceptable prediction in term of accuracy (NSE = 0.90) and has a lower error values of RMSE, MAE, and RSD, with 2.358 (m³/s), 1.08 (m³/s), and 0.311 (m³/s), correspondingly, when compared with GEP (RMSE = 2.704 (m³/s), MAE = 1.442 (m³/s), and 0.357(m³/s)) and RFR (RMSE = 2.61 (m³/s), MAE = 1.29 (m³/s), and 0.33(m³/s)) in the validation stage. Moreover, by comparing the results of models, the reliability and resilience of EEMD-VMD-RFR show an increase by approximately 10% and 21% in comparison with RFR, where these criteria are around 92% and 78% for EEMD-VMD coupled with GEP in the validation period. Although, the uncertainty analysis resulted that all four AI models overpredicted daily streamflow (i.e., mean prediction error for all models is positive), it is interesting to note that EEMD-VMD-RFR produced a smaller range of the uncertainty bands with ±3.922 compared with GEP (±5.058), RFR (±5.514), and EEMD-VMD-GEP (±4.563) (Table 8). In addition, the lowest 95% predictive error interval is calculated for the EEMD-VMD-RFR.

Table 8. Uncertainty analyses for streamflow prediction using predictive models at Bilghan station.

Models	Mean Prediction Error	Width of Uncertainty Band	Median	95% Predictive Error Interval
Bilghan Station				
GEP	1.102	±5.058	16.52	−4.25 to 5.86
RFR	0.811	±5.514	15.684	−5.53 to 5.49
EEMD-VMD-GEP	0.483	±4.563	15.804	−4.18 to 4.94
EEMD-VMD-RFR	0.301	±3.922	15.78	−3.62 to 4.22

It is evident from the scatterplots (see Figure 9) that the EEMD-VMD-RFR model's line of linear agreement between forecasted (y) and observed (x) daily streamflow in validation stage is closer to the ideal line (1:1) compared with the GEP, RFR, and EEMD-VMD-GEP models. It is clear that the EEMD-VMD-RFR model can provide an accurate value for streamflow at Bilghan compared with the other three forecasting models.

Figure 10 illustrates the predicted hydrographs using the four proposed models. As illustrated, the EEMD-VMD-FRF showed a much better agreement between predicted and observed times series which demonstrates the suitability of EEMD and VMD in pre-processing inputs/output data and predicting daily flow.
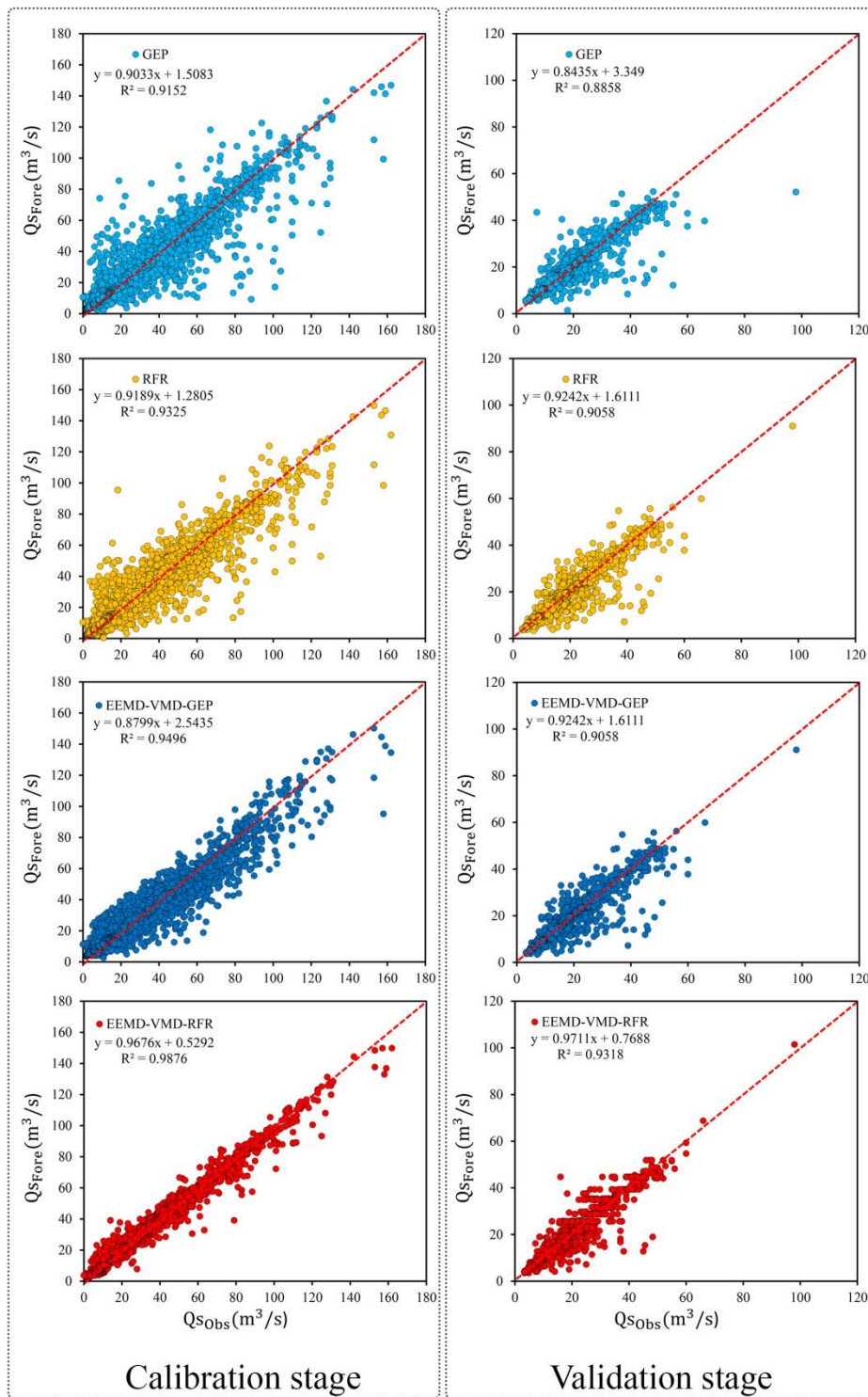
Figure 9. Scatterplot of forecasted (Qs) and observed (Qs) daily streamflow using the standalone GEP, RFR, and ensemble EEMD-VMD-GEP, EEMD-VMD-RFR models for the calibration and validation stages at the Bilghan station.
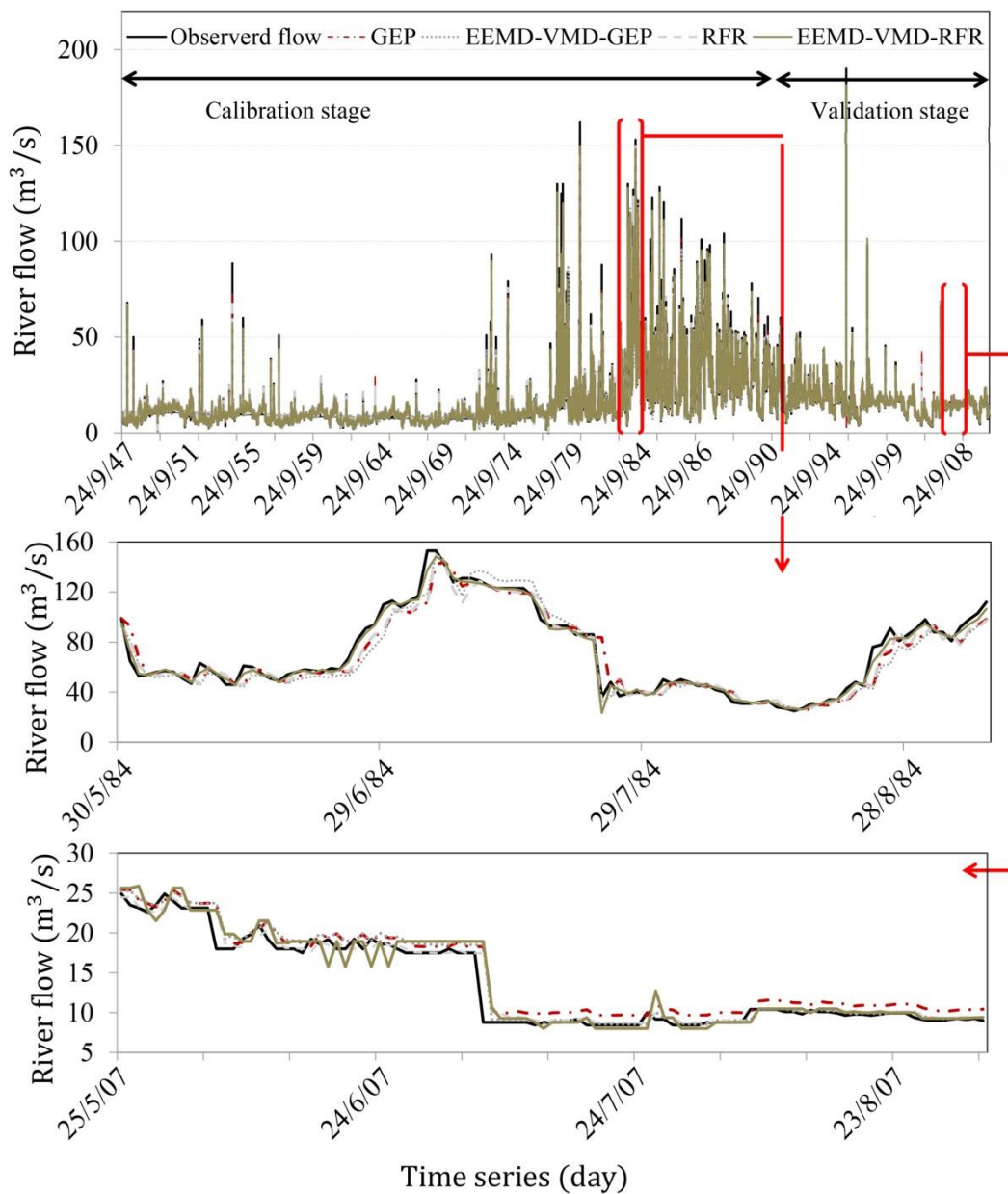
Figure 10. A time-series of observed and forecasted river daily streamflow of ensemble EEMD-VMD-GEP and EEMD-VMD-RFR models compared with standalone GEP and RFR models for the calibration and validation stages at the Bilghan station.

3.2.3. The Gachsar Gauging Station

Concurrent with prior stations (Siira and Bilghan stations), EEMD-VMD based GEP and RFR models outperformed in the calibration phase. The values of NSE, RMSE, MAE, RSD, U95, reliability, and resilience for each model at the Gachsar gauging station are presented in Table 6. Compare with the original GEP and RFR models in which no pre-processing procedure of model inputs/output were carried out, the accuracy (NSE = 0.98) of the EEMD-VMD-RFR is far better than RFR (NSE = 0.93) particularly in the calibration period. The EEMD-VMD approach, in addition, improved GEP accuracy by about 7%, whereas standalone GEP generated an NSE value of 0.89. The results in the validation stage show that EEMD-VMD-GEP had a lower percentage error for streamflow prediction in terms of RSD (37.43%) and U95 (3.60%) than standalone GEP model (RSD = 0.366 and U95 = 8.38). Regarding the two-phase decomposition approach (EEMD-VMD), the outcomes present that the RFR model yielded reliability of 97.56% and resilience of 88.29% for daily streamflow prediction, while these values

are 96.27% and 92.15% for the GEP, indicating much better performance of an evolutionary algorithm (Table 9).

Table 9. Evaluation metrics of the proposed models in the calibration and validation periods at the Gachsar station.

Models	Statistical Error Indices			
	GEP	RFR	EEMD-VMD-GEP	EEMD-VMD-RFR
Total available data in calibration stage				
NSE	0.89	0.934	0.95	0.98
RMSE (m^3/s)	1.879	0.964	0.84	0.53
MAE (m^3/s)	0.519	0.256	0.237	0.137
RSD (m^3/s)	0.497	0.255	0.222	0.141
U95	5.274	7.64	7.59	7.482
Reliability (%)	91.52	93.98	94.06	96.16
Resilience (%)	54.49	78.65	77.18	85.33
Total available data in validation period				
NSE	0.865	0.93	0.95	0.97
RMSE (m^3/s)	1.47	1.013	0.62	0.516
MAE(m^3/s)	0.647	0.357	0.302	0.29
RSD (m^3/s)	0.366	0.252	0.229	0.178
U95	8.38	8.12	8.078	7.95
Reliability (%)	86.7	95.91	96.27	97.56
Resilience (%)	69.97	84.82	92.15	88.29

Table 10 summarizes the outcomes of the uncertainty analysis. Analysis revealed that 95% predictive error interval for ensemble GEP ranges from −1.78 to +1.83 and for RFR varies from −1.54 to +1.66 are likely similar, while the mean predictive error for EEMD-RFR (0.006) is far lower than the EEMD-VMD-GEP (0.021). Moreover, the uncertainty bands for all four models vary significantly, in which the band for EEMD-VMD-RFR (±1.402) is closer than EEMD-VMD-GEP (±1.808), standalone GEP (±2.876) and RFR (±1.986).

Table 10. Uncertainty analyses for streamflow prediction using predictive models at Gachsar station.

Models	Mean Prediction Error	Width of Uncertainty Band	Median	95% predictive Error Interval
Gachsar station				
GEP	0.103	±2.876	4.11	−2.97 to 2.77
RFR	0.084	±1.986	4.236	−1.99 to 1.98
EEMD-VMD-GEP	0.021	±1.808	4.186	−1.78 to 1.83
EEMD-VMD-RFR	0.006	±1.402	4.15	−1.54 to 1.66

Figure 11 represents the scatter plots of predicted and observed streamflows for the Gachsar station. Although RFR prediction is mostly underestimated, pre-processing technique remarkably improved the performance criteria throughout the simulation period. It is clearly observed that EEMD-VMD process highly influenced the model performances and the slopes of the EEMD-VMD-GEP and EEMD-VMD-RFR are very close to the perfect line compared to the GEP and RFR models. Additionally, the correlations between observed and predicted streamflow values for EEMD-VMD-RFR (for calibration R^2 = 0.98 and validation R^2 = 0.96) are higher than the EEMD-VMD-GEP (for calibration R^2 = 0.95 and validation R^2 = 0.94) technique.

Furthermore, calibration and validation results (Figure 12) indicate that predicted streamflow using EEMD-VMD technique is in well agreement with observational data. However, over- and underestimations of the time series values, particularly for the GEP simulation, can be visually observed from the time variation graphs. To this end, the two hybrid ensemble models provided skillful simulation in terms of predicting high and low flow variability.
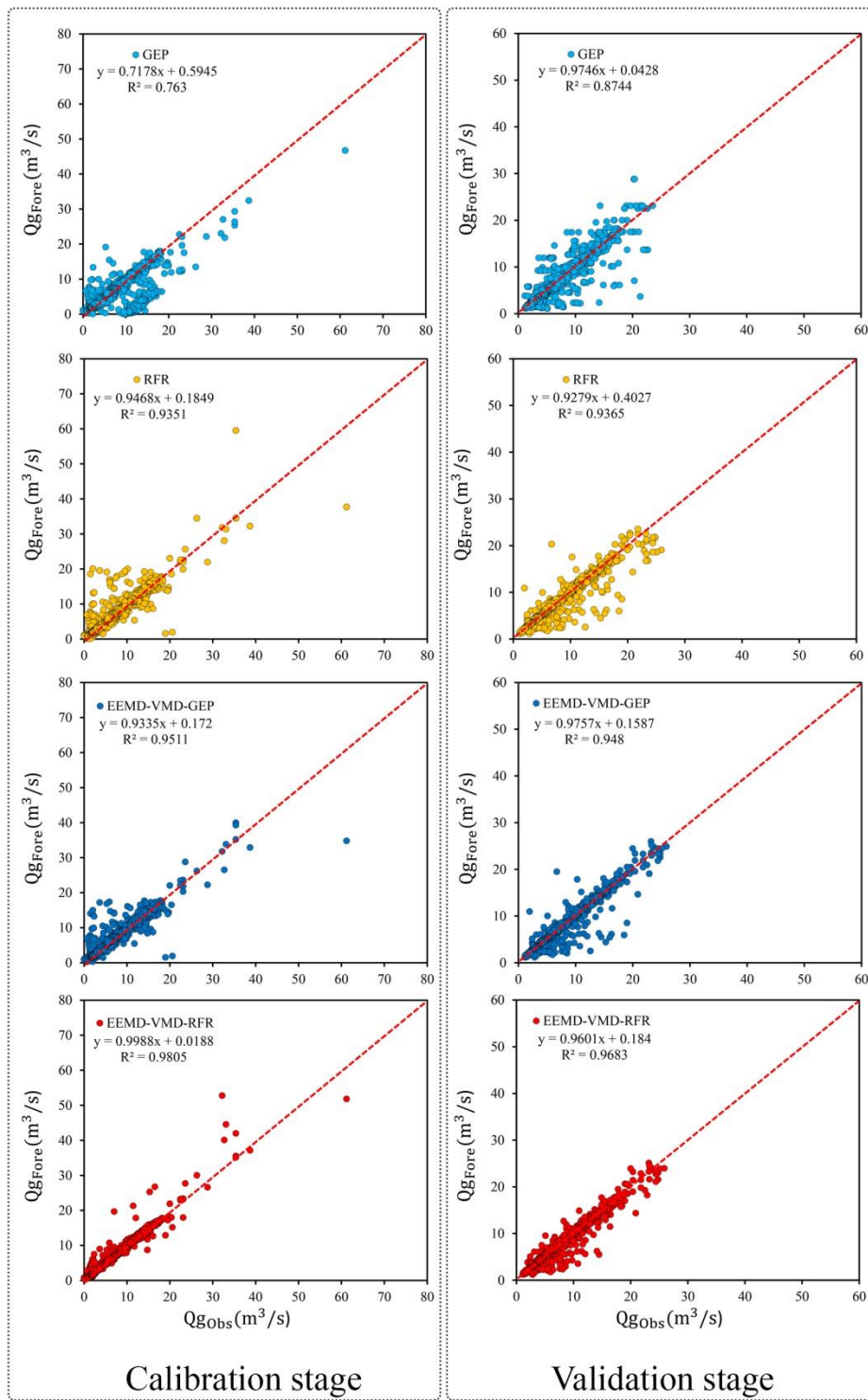
Figure 11. Scatterplot of forecasted (Q_{Obs}) and observed (Q_{Fore}) river flow using the standalone GEP, RFR, and ensemble EEMD-VMD-GEP, EEMD-VMD-RFR models for the calibration and validation periods at the Gachsar station.
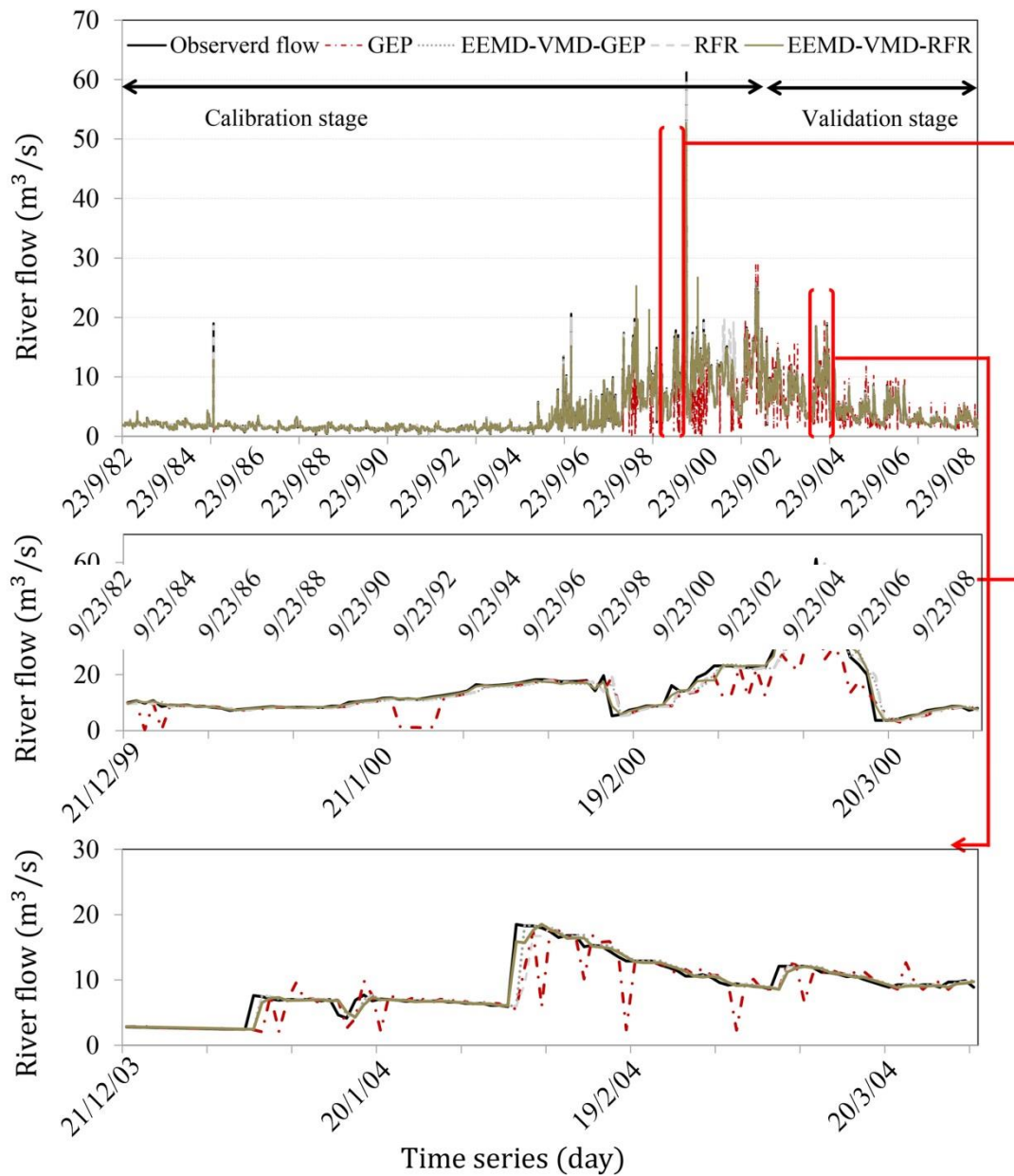
Figure 12. A time-series of observed and forecasted daily streamflow of ensemble EEMD-VMD-GEP and EEMD-VMD-RFR models compared with standalone GEP and RFR models for the calibration and validation periods at the Gachsar station.

3.2.4. Further Comparison Among Proposed Models

To visually evaluate the performance of models in the validation period, box plot and empirical cumulative distribution (ECDF) are illustrated in Figures 13 and 14. Box plots indicated the spread of predicted and observed streamflow based on the quartiles, that the whiskers showed the variability outside of the 25th and 75th percentiles.
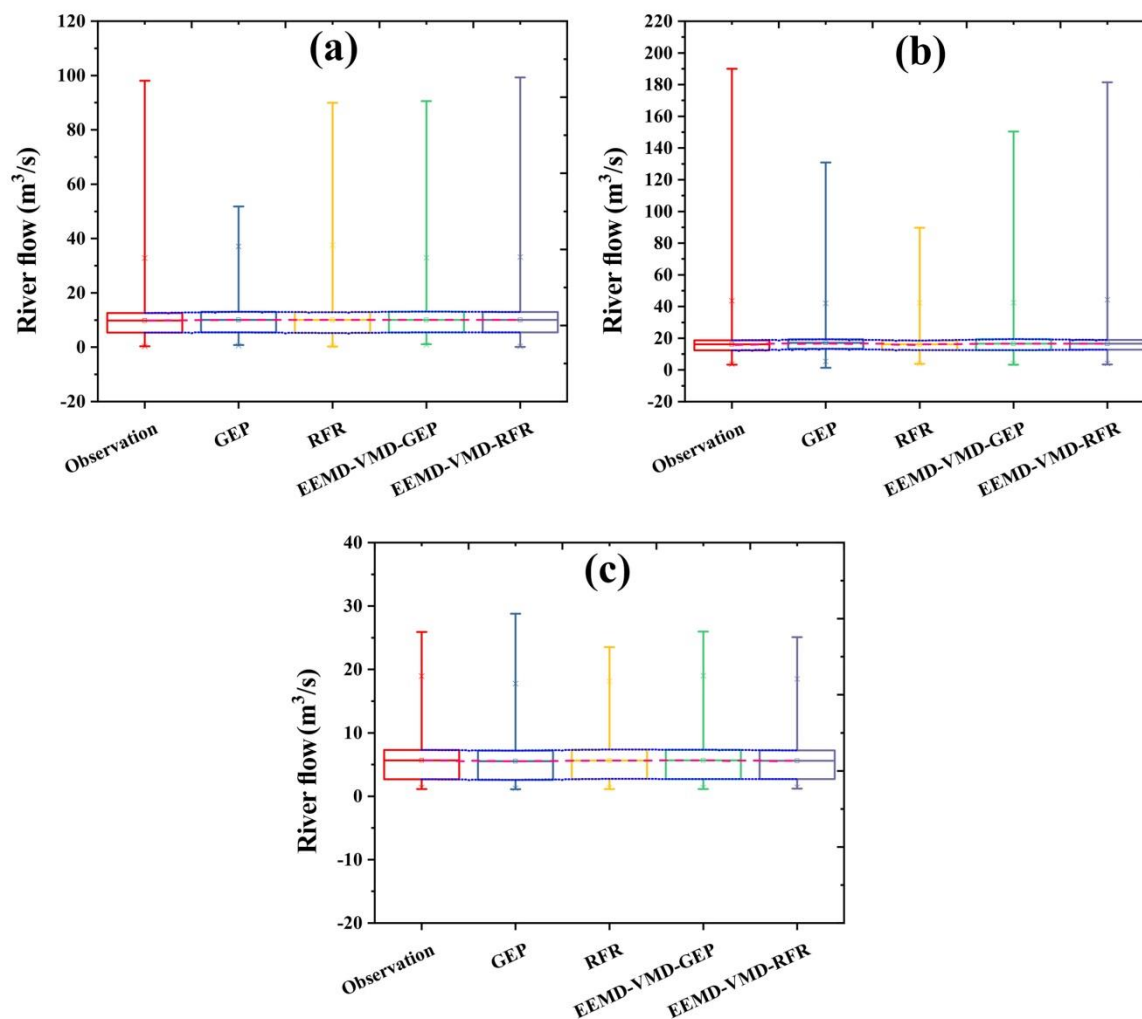
Figure 13. Boxplots of the measured and predicted daily streamflow by the standalone (i.e., GEP and RFR) and hybrid ensemble (i.e., EEMD-VMD-GEP and EEMD-VMD-RFR) models in the validation phase. (**a**) Siira, (**b**) Bilghan, and (**c**) Gachsar stations.

Moving from Siira, to Bilghan, and Gachsar stations, more dispersion and skewness is noted in the Gachsar prediction. Further, the prediction of EEMD-VMD-RFP is more skillful for Siira and Bilghan rather than the Gachsar station. In all cases, the GEP standalone prediction seems to poorly predict daily streamflow records underestimating the majority of high flow events.

Further, the percentage of the absolute forecasted error statistics value (|FE|) through the empirical cumulative distribution function (ECDF) were computed and analyzed for at three stations (see Figure 14). Regarding the percentage of errors illustrated in the minimum error bracket (i.e., from -5 to ±5 m^3/s), the ECDF indicated the superiority of EEMD-VMD-RFR and EEMD-VMD-GEP models in predicting streamflow fluctuations. According to this error bracket, the EEMD-VMD-RFR performed slightly better than the rest of the models.
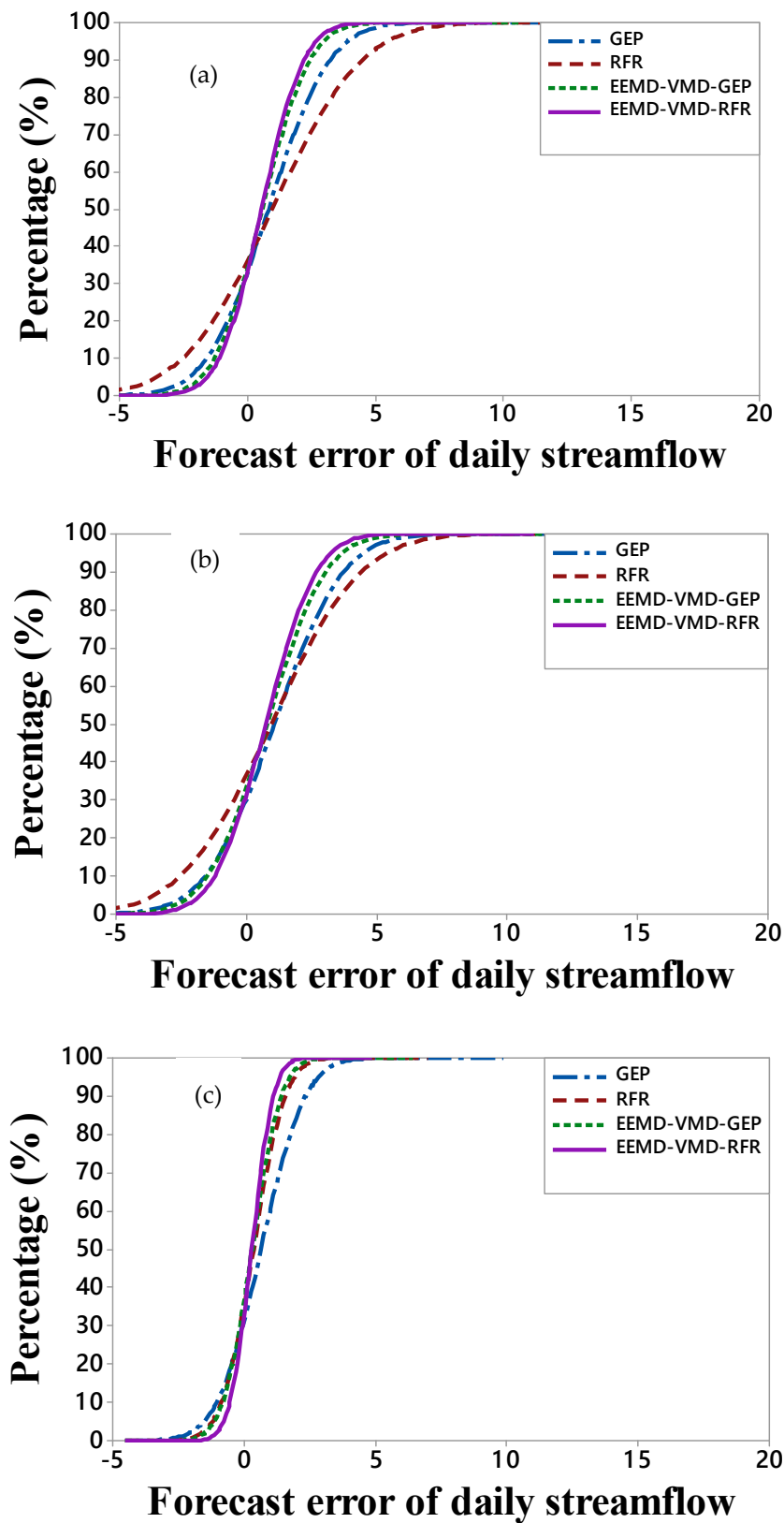
Figure 14. Empirical cumulative distribution (ECDF) of the absolute forecasted error |FE| (m^3/s) for the EEMD-VMD-RFR model compared to the other models at (**a**) Siira, (**b**) Bilghan, and (**c**) Gachsar stations in the validation phase.

There are various limitations should be addressed to illustrate future avenues for further research, despite the outstanding accuracy of the ensemble decomposition-based GEP and RFR models for daily streamflow forecasting. It is a need to design an expert system for a real-time short-term forecasting tool that a technique should be performed at much shorter-term periods, such as hourly step-ahead flow forecasting. This requires adequate and accurate time-series data that was not available for the present study. One of the drawbacks of the EEMD-VMD-MARS/RFR is that its implementation is that it is time-consuming because it produces several numbers of IMFs. In this respect, this study suggests applying other techniques or modifying the proposed methods to speed up the simulation process.

4. Conclusions

This paper examined the performance of a hybrid ensemble decomposition method embedded with the EEMD combined with the VMD as a preprocessing approach. Two kinds of AI methods namely RFR and GEP were implemented for predicting daily streamflow across three different gauging stations, Siira, Bilghan, and Gachsar, at a mountainous area in Karaj basin, Iran

Compare to the results of standalone (GEP and RFR) and ensemble-based (EEMD-VMD-GEP/RFR) AI models, it is apparent that of the performance of all four developed algorithms are relatively high. Among all models, EEMD-VMD approach showed promising results in term of predicting high and low values of daily streamflow. The predictive errors and uncertainties associated with the developed EEMD-VMD- RFR model were smaller than those associated with other forecast models. The analysis suggested that data pre-processing approaches (e.g., EEMD and VMD) are key approaches in accurate prediction of daily streamflow dynamics over time. This result is vital for reliable management of study region where the available dataset may involve noises and higher order trends.

Overall, the EEMD-VMD-RFR algorithm outperformed the other models. The results obtained by the best performing model are satisfactory as evidence by highest NSE and lowest RMSE values. This reflects that a hybrid-based AI approach is a robust and useful tool for simulating daily streamflow over the mountainous region. However, most of the models could not successfully predict the extreme (high and low) flow events. The reason is the developed models are not predominantly trained on extreme events. This can be overcome in the future by the flexibility to incorporate auxiliary information and soft data, such as expert knowledge, into the algorithmic framework, which could provide more flexibility in simulation and assist water managers and dam operators. In summary, this research presents a novel study of testing various AI-based algorithms for streamflow prediction and gives a comprehensive comparison among popular AI methods in hydrologic simulation. The authors believe this unique feature of the AI methods, especially the EEMD-VMD-RFR algorithm, is to be able to be further employed in the study region and provide more flexibility by adding desired decision variables for reservoir management. However, the proposed methodologies in this study are universally applicable to other mountainous regions, and are flexible to incorporate and test for other hydrologic time series data, such as flood records.

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

References

1. Yuan, X.; Chen, C.; Lei, X.; Yuan, Y.; Adnan, R.M. Monthly runoff forecasting based on LSTM–ALO model. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 2199–2212. [CrossRef]

2. Yu, Y.; Zhang, H.; Singh, V. Forward prediction of runoff data in data-scarce basins with an improved ensemble empirical mode decomposition (EEMD) model. *Water* **2018**, *10*, 388. [CrossRef]

3. Adamowski, J.; Sun, K. Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* **2010**, *390*, 85–91. [CrossRef]

4. Chitsaz, N.; Azarnivand, A.; Araghinejad, S. Pre-processing of data-driven river flow forecasting models by singular value decomposition (SVD) technique. *Hydrol. Sci. J.* **2016**, *61*, 2164–2178. [CrossRef]

5. Humphrey, G.B.; Gibbs, M.S.; Dandy, G.C.; Maier, H.R. A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. *J. Hydrol.* **2016**, *540*, 623–640. [CrossRef]

6. Abdollahi, S.; Raeisi, J.; Khalilianpour, M.; Ahmadi, F.; Kisi, O. Daily mean streamflow prediction in perennial and non-perennial rivers using four data driven techniques. *Water Resour. Manag.* **2017**, *31*, 4855–4874. [CrossRef]

7. Mehr, A.D. An improved gene expression programming model for streamflow forecasting in intermittent streams. *J. Hydrol.* **2018**, *563*, 669–678. [CrossRef]

8. Rezaie-Balf, M.; Kisi, O. New formulation for forecasting streamflow: Evolutionary polynomial regression vs. extreme learning machine. *Hydrol. Res.* **2018**, *49*, 939–953. [CrossRef]

9. Rezaie-Balf, M.; Noori, R.; Berndtsson, R.; Ghaemi, A.; Ghiasi, B. Evolutionary polynomial regression approach to predict longitudinal dispersion coefficient in rivers. *J. Water Supply Res. Technol. AQUA* **2018**, *67*, 447–457.

10. Chang, T.K.; Talei, A.; Alaghmand, S.; Ooi, M.P.L. Choice of rainfall inputs for event-based rainfall-runoff modeling in a catchment with multiple rainfall stations using data-driven techniques. *J. Hydrol.* **2017**, *545*, 100–108. [CrossRef]

11. Zare, M.; Koch, M. Groundwater level fluctuations simulation and prediction by ANFIS-and hybrid Wavelet-ANFIS/Fuzzy C-Means (FCM) clustering models: Application to the Miandarband plain. *J. Hydro. Environ. Res.* **2018**, *18*, 63–76. [CrossRef]

12. Najafzadeh, M.; Rezaie Balf, M.; Rashedi, E. Prediction of maximum scour depth around piers with debris accumulation using EPR, MT, and GEP models. *J. Hydroinform.* **2016**, *18*, 867–884. [CrossRef]

13. Shahrara, N.; Çelik, T.; Gandomi, A.H. Gene expression programming approach to cost estimation formulation for utility projects. *J. Civ. Eng. Manag.* **2017**, *23*, 85–95. [CrossRef]

14. Wang, L.; Kisi, O.; Zounemat-Kermani, M.; Zhu, Z.; Gong, W.; Niu, Z.; Liu, H.; Liu, Z. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. *Int. J. Climatol.* **2017**, *37*, 1141–1155. [CrossRef]

15. Sattar, A.A.; Elhakeem, M.; Rezaie-Balf, M.; Gharabaghi, B.; Bonakdari, H. Artificial intelligence models for prediction of the aeration efficiency of the stepped weir. *Flow Meas. Instrum.* **2019**, *65*, 78–89. [CrossRef]

16. Li, X.; Sha, J.; Wang, Z.-L. A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. *Hydrol. Res.* **2017**, *48*, 1214–1225. [CrossRef]

17. Kisi, O.; Parmar, K.S.; Soni, K.; Demir, V. Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline, and M5 model tree models. *Air Qual. Atmos. Health* **2017**, *10*, 873–883. [CrossRef]

18. Zhang, D.; Lin, J.; Peng, Q.; Wang, D.; Yang, T.; Sorooshian, S.; Liu, X.; Zhuang, J. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *J. Hydrol.* **2018**, *565*, 720–736. [CrossRef]

19. Chen, L.; Sun, N.; Zhou, C.; Zhou, J.; Zhou, Y.; Zhang, J.; Zhou, Q. Flood Forecasting Based on an Improved Extreme Learning Machine Model Combined with the Backtracking Search Optimization Algorithm. *Water* **2018**, *10*, 1362. [CrossRef]

20. Chen, X.Y.; Chau, K.W.; Busari, A.O. A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. *Eng. Appl. Artif. Intell.* **2015**, *46*, 258–268. [CrossRef]

21. Ahani, A.; Shourian, M.; Rad, P.R. Performance assessment of the linear, nonlinear and nonparametric data driven models in river flow forecasting. *Water Resour. Manag.* **2018**, *32*, 383–399. [CrossRef]

22. Baydaroğlu, Ö.; Koçak, K.; Duran, K. River flow prediction using hybrid models of support vector regression with the wavelet transform, singular spectrum analysis and chaotic approach. *Meteorol. Atmos. Phys.* **2018**, *130*, 349–359. [CrossRef]

23. Bou-Fakhreddine, B.; Mougharbel, I.; Faye, A.; Chakra, S.A.; Pollet, Y. Daily river flow prediction based on Two-Phase Constructive Fuzzy Systems Modeling: A case of hydrological–meteorological measurements asymmetry. *J. Hydrol.* **2018**, *558*, 255–265. [CrossRef]

24. Yin, Z.; Feng, Q.; Wen, X.; Deo, R.C.; Yang, L.; Si, J.; He, Z. Design and evaluation of SVR, MARS and M5Tree models for 1, 2 and 3-day lead time forecasting of river flow data in a semiarid mountainous catchment. *Stoch. Environ. Res. Risk A* **2018**, *32*, 2457–2476. [CrossRef]

25. Wu, C.L.; Chau, K.W. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J. Hydrol.* **2011**, *399*, 394–409. [CrossRef]

26. Kim, S.; Seo, Y.; Rezaie-Balf, M.; Kisi, O.; Ghorbani, M.A.; Singh, V.P. Evaluation of daily solar radiation flux using soft computing approaches based on different meteorological information: Peninsula vs. continent. *Theor. Appl. Climatol.* **2018**. [CrossRef]

27. Benedetto, F.; Giunta, G.; Mastroeni, L. A maximum entropy method to assess the predictability of financial and commodity prices. *Digit. Signal Process.* **2015**, *46*, 19–31. [CrossRef]

28. Zakhrouf, M.; Bouchelkia, H.; Stamboul, M.; Kim, S.; Heddam, S. Time series forecasting of river flow using an integrated approach of wavelet multi-resolution analysis and evolutionary data-driven models. A case study: Sebaou River (Algeria). *Phys. Geogr.* **2018**, *39*, 506–522. [CrossRef]

29. Hu, T.; Wu, F.; Zhang, X. Rainfall–runoff modeling using principal component analysis and neural network. *Hydrol. Res.* **2007**, *38*, 235–248. [CrossRef]

30. Ravikumar, P.; Somashekar, R.K. Principal component analysis and hydrochemical facies characterization to evaluate groundwater quality in Varahi river basin, Karnataka state, India. *Appl. Water Sci.* **2017**, *7*, 745–755. [CrossRef]

31. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [CrossRef]

32. Baykasoğlu, A.; Güllü, H.; Çanakçı, H.; Özbakır, L. Prediction of compressive and tensile strength of limestone via genetic programming. *Expert Syst. Appl.* **2008**, *35*, 111–123. [CrossRef]

33. Yuan, X.; Tan, Q.; Lei, X.; Yuan, Y.; Wu, X. Wind power prediction using hybrid autoregressive fractionally integrated moving average and least square support vector machine. *Energy* **2017**, *129*, 122–137. [CrossRef]

34. Huang, N.E.; Wu, Z. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Rev. Geophys.* **2008**, *46*. [CrossRef]

35. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [CrossRef]

36. Liu, H.; Mi, X.; Li, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Convers. Manag.* **2018**, *159*, 54–64. [CrossRef]

37. Naik, J.; Dash, S.; Dash, P.K.; Bisoi, R. Short term wind power forecasting using hybrid variational mode decomposition and multi-kernel regularized pseudo inverse neural network. *Renew. Energy* **2018**, *118*, 180–212. [CrossRef]

38. Wang, Y.; Markert, R.; Xiang, J.; Zheng, W. Research on variational mode decomposition and its application in detecting rub-impact fault of the rotor system. *Mech. Syst. Signal Process.* **2015**, *60*, 243–251. [CrossRef]

39. Seo, Y.; Kim, S.; Singh, V.P. Comparison of different heuristic and decomposition techniques for river stage modeling. *Environ. Monit. Assess.* **2018**, *190*, 392. [CrossRef]

40. Seo, Y.; Kim, S.; Singh, V. Machine learning models coupled with variational mode decomposition: A new approach for modeling daily rainfall-runoff. *Atmosphere* **2018**, *9*, 251. [CrossRef]

41. Rezaie-Balf, M.; Kisi, O.; Chua, L.H. Application of ensemble empirical mode decomposition based on machine learning methodologies in forecasting monthly pan evaporation. *Hydrol. Res.* **2018**. [CrossRef]

42. Napolitano, G.; Serinaldi, F.; See, L. Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily stream flow series: An empirical examination. *J. Hydrol.* **2011**, *406*, 199–214. [CrossRef]

43. Wang, W.C.; Xu, D.M.; Chau, K.W.; Chen, S. Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *J. Hydroinform.* **2013**, *15*, 1377–1390. [CrossRef]

44. Li, G.; Ma, X.; Yang, H. A hybrid model for monthly precipitation time series forecasting based on variational mode decomposition with extreme learning machine. *Information* **2018**, *9*, 177. [CrossRef]

45. Ferreira, C. Gene expression programming and the evolution of computer programs. In *Recent Developments in Biologically Inspired Computing*; Igi Global: Hershey, PA, USA, 2005; pp. 82–103.

46. Ferreira, C. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 21.

47. Gandomi, A.H.; Babanajad, S.K.; Alavi, A.H.; Farnam, Y. Novel approach to strength modeling of concrete under triaxial compression. *J. Mater. Civ. Eng.* **2012**, *24*, 1132–1143. [CrossRef]

48. Gandomi, A.H.; Alavi, A.H.; Kazemi, S.; Gandomi, M. Formulation of shear strength of slender RC beams using gene expression programming, part I: Without shear reinforcement. *Autom. Constr.* **2014**, *42*, 112–121. [CrossRef]

49. Dey, P.; Das, A.K. A utilization of GEP (gene expression programming) metamodel and PSO (particle swarm optimization) tool to predict and optimize the forced convection around a cylinder. *Energy* **2016**, *95*, 447–458. [CrossRef]

50. Kaboli, S.H.A.; Fallahpour, A.; Selvaraj, J.; Rahim, N.A. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. *Energy* **2017**, *126*, 144–164. [CrossRef]

51. Samadianfard, S.; Asadi, E.; Jarhan, S.; Kazemi, H.; Kheshtgar, S.; Kisi, O.; Sajjadi, S.; Manaf, A.A. Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. *Soil Tillage Res.* **2018**, *175*, 37–50. [CrossRef]

52. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.

53. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958. [CrossRef]

54. Cootes, T.F.; Ionita, M.C.; Lindner, C.; Sauer, P. Robust and accurate shape model fitting using random forest regression voting. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 278–291.

55. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Chi, C.T.; Liu, H.H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]

56. Huang, N.E.; Wu, M.L.C.; Long, S.R.; Shen, S.S.; Qu, W.; Gloersen, P.; Fan, K.L. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. Math. Phys. Eng. Sci.* **2003**, *459*, 2317–2345. [CrossRef]

57. Lei, Y.; He, Z.; Zi, Y. Application of the EEMD method to rotor fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **2009**, *23*, 1327–1338. [CrossRef]

58. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]

59. Liu, Y.; Yang, G.; Li, M.; Yin, H. Variational mode decomposition denoising combined the detrended fluctuation analysis. *Signal Process.* **2016**, *125*, 349–364. [CrossRef]

60. Singh, J.; Knapp, H.V.; Arnold, J.G.; Demissie, M. Hydrological modeling of the iroquois river watershed using HSPF and SWAT 1. *J. Am. Water Resour. Assoc.* **2005**, *41*, 343–360. [CrossRef]

61. Zhou, Y.; Chang, F.J.; Guo, S.; Ba, H.; He, S. A robust recurrent anfis for modeling multi-step-ahead flood forecast of three gorges reservoir in the Yangtze River. *Hydrol. Earth Syst. Sci. Discuss.* **2017**. [CrossRef]

62. Samadi, S.; Tufford, D.; Carbone, G. Assessing prediction uncertainty of a semi-distributed hydrology model for a shallow aquifer dominated environmental system. *J. Am. Water Resour. Assoc.* **2017**, *53*, 1368–1389. [CrossRef]

63. Najafzadeh, M.; Rezaie-Balf, M.; Tafarojnoruz, A. Prediction of riprap stone size under overtopping flow using data-driven models. *Int. J. River Basin Manag.* **2018**, *16*, 505–512. [CrossRef]

64. Newcombe, R.G. Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Stat. Med.* **1998**, *17*, 857–872. [CrossRef]

65. Tabari, H.; Aghajanloo, M.B. Temporal pattern of aridity index in Iran with considering precipitation and evapotranspiration trends. *Int. J. Climatol.* **2013**, *33*, 396–409. [CrossRef]

66. Fallah-Mehdipour, E.; Bozorg Haddad, O.; Mariño, M.A. Extraction of optimal operation rules in an aquifer-dam system: Genetic programming approach. *J. Irrig. Drain. Eng.* **2013**, *139*, 872–879. [CrossRef]

67. Heidarnejad, M.; Golmaee, S.H.; Mosaedi, A.; Ahmadi, M.Z. Estimation of sediment volume in Karaj Dam Reservoir (Iran) by hydrometry method and a comparison with hydrography method. *Lake Reserv. Manag.* **2006**, *22*, 233–239. [CrossRef]

68. Yozgatligil, C.; Yazici, C. Comparison of homogeneity tests for temperature using a simulation study. *Int. J. Climatol.* **2016**, *36*, 62–81. [CrossRef]

69. Nkiaka, E.; Nawaz, N.R.; Lovett, J.C. Analysis of rainfall variability in the Logone catchment, Lake Chad basin. *Int. J. Climatol.* **2017**, *37*, 3553–3564. [CrossRef]

70. Kazemzadeh, M.; Malekian, A. Homogeneity analysis of streamflow records in arid and semi-arid regions of northwestern Iran. *J. Arid Land* **2018**, *10*, 493–506. [CrossRef]

71. Wijngaard, J.B.; Tank, A.K.; Können, G.P. Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol.* **2003**, *23*, 679–692. [CrossRef]

72. Zhang, X.; Zhang, Q.; Zhang, G.; Nie, Z.; Gui, Z.; Que, H. A novel hybrid data-driven model for daily land surface temperature forecasting using long short-term memory neural network based on ensemble empirical mode decomposition. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1032. [CrossRef] [PubMed]

73. Niu, M.; Hu, Y.; Sun, S.; Liu, Y. A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting. *Appl. Math. Model.* **2018**, *57*, 163–178. [CrossRef]