

Article

Deterministic Analysis and Uncertainty Analysis of Ensemble Forecasting Model Based on Variational Mode Decomposition for Estimation of Monthly Groundwater Level

Min Wu ^{1,2,3}, Qi Feng ^{1,3,*} , Xiaohu Wen ^{1,3,*}, Zhenliang Yin ^{1,3} , Linshan Yang ^{1,3}  and Danrui Sheng ^{1,2,3}

¹ Key Laboratory of Ecohydrology of Inland River Basin, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; wumin@nieer.ac.cn (M.W.); yinzhenliang@lzb.ac.cn (Z.Y.); yanglsh08@lzb.ac.cn (L.Y.); shengdanrui@lzb.ac.cn (D.S.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Qilian Mountains Eco-Environment Research Center in Gansu Province, Lanzhou 730000, China

* Correspondence: qifeng@lzb.ac.cn (Q.F.); xhwen@lzb.ac.cn (X.W.)

Abstract: Precise multi-time scales prediction of groundwater level is essential for water resources planning and management. However, credible and reliable predicting results are hard to achieve even to extensively applied artificial intelligence (AI) models considering the uncontrollable error, indefinite inputs and unneglectable uncertainty during the modelling process. The AI model ensemble with the data pretreatment technique, the input selection method, or uncertainty analysis has been successfully used to tackle this issue, whereas studies about the comprehensive deterministic and uncertainty analysis of hybrid models in groundwater level forecast are rarely reported. In this study, a novel hybrid predictive model combining the variational mode decomposition (VMD) data pretreatment technique, Boruta input selection method, bootstrap based uncertainty analysis, and the extreme learning machine (ELM) model named VBELM was developed for 1-, 2- and 3-month ahead groundwater level prediction in a typical arid oasis area of northwestern China. The historical observed monthly groundwater level, precipitation and temperature data were used as inputs to train and test the model. Specifically, the VMD was used to decompose all the input-outputs into a set of intrinsic mode functions (IMFs), the Boruta method was applied to determine input variables, and the ELM was employed to forecast the value of each IMF. In order to ascertain the efficiency of the proposed VBELM model, the performance of the coupled model (VELM) hybridizing VMD with ELM algorithm and the single ELM model were estimated in comparison. The results indicate that the VBELM performed best, while the single ELM model performed the worst among the three models. Furthermore, the VBELM model presented lower uncertainty than the VELM model with more observed groundwater level values falling inside the confidence interval. In summary, the VBELM model demonstrated an excellent performance for both certainty and uncertainty analyses, and can serve as an effective tool for multi-scale groundwater level forecasting.

Keywords: uncertainty analysis; groundwater level prediction; hybrid predictive model; variational mode decomposition; Boruta technique; extreme learning machine



Citation: Wu, M.; Feng, Q.; Wen, X.; Yin, Z.; Yang, L.; Sheng, D. Deterministic Analysis and Uncertainty Analysis of Ensemble Forecasting Model Based on Variational Mode Decomposition for Estimation of Monthly Groundwater Level. *Water* **2021**, *13*, 139. <https://doi.org/10.3390/w13020139>

Received: 12 November 2020

Accepted: 6 January 2021

Published: 9 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Groundwater, a significant water supply for agricultural, drinking, domestic and industrial purposes, plays a vital role in the sustainable development of society and ecology in arid regions where the rainfall is extremely scarce [1,2]. As the most critical factor, the groundwater level is fit to quantify the groundwater resource. However, the dwindling of groundwater level caused by external pressure of population increase, economic development, climate change and pollution over-exploitation is threatening the sustainability of water resources in arid areas [3,4]. Therefore, deriving accurate and specific groundwater level estimation is essential for policy makers and planners to evaluate and manage the groundwater resources as well as prevent over-abstraction more effectively.

Generally, the groundwater level can be assessed by physically-based and data-based models. In the physically based models, a detailed interaction of various physical processes that control the hydrologic behavior of the groundwater system is involved using a simplified control equation [5]. To achieve this, proper initial and boundary conditions calculated by numerical methods are indispensable [6]. Nevertheless, a large quantity of hydrogeological data and the physical properties of groundwater such as hydraulic conductivity, volumetric water content and matric potential are hard to access even with expensive site investigations. Moreover, the computational costs derived from partitioning of the physical domain required for the numerical solution is extremely high [7,8]. Besides, limited insight of the researchers about the physical process of groundwater flow restricts application of the models from a practical perspective. Thus, satisfactory results from the numerical groundwater models may be constrained considering these facts. As a result, an alternative artificial intelligence (AI) model, which formulates groundwater level nonlinearity merely by relying on the historical data or broader exogenous data as inputs, is necessary and significant under such conditions.

In recent years, a wide variety of AI models, such as the Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), Extreme Learning Machine (ELM) and genetic programming (GP), has been extensively recommended and applied to investigate hotspots in hydrologic research, mainly focusing on rainfall, runoff, reference evapotranspiration, flood, water quality and groundwater level forecast [9–15]. In terms of groundwater level prediction, the applicability and potential of these AI methods have been confirmed [16]. Among these models, the ELM, as a typical representative of AI models, has received increasing attention in hydrological modeling due to the fast learning speed and strong generalization capability. Compared with the traditional ANN methodology, the ELM algorithm does not need prior tuning of meta-parameters like input weights and hidden layer biases [10]. Thus, a global solution and more accurate prediction are able to be attained.

However, deficiencies still exist in groundwater modelling despite the effectiveness and applicability of the conventional AI methods. Firstly, the complexity of the groundwater dynamic system itself makes it difficult to distinguish and recognize groundwater features accurately [17]. In addition, uncertainties from non-stationary inputs caused by trends or seasonal variation of the data may largely affect the model's performance if no preprocessing procedure is applied [18,19]. Under such circumstances, linking data pretreatment process with the conventional AI models can be considered as an effective attempt at careful analysis of the dynamic characteristics and precise accuracy of these models.

In recent years, a number of studies have verified and reported the superiority and availability of AI models coupled with data preprocessing techniques [4,18,20–26]. In review of this, the wavelet transformation (WT) algorithm, empirical mode decomposition (EMD) and ensemble empirical mode decomposition (EEMD) are three commonly used techniques that can decompose original time series into various sub-series components and extract valuable information concealed in the datasets [27–29]. Several successful applications of the WT, EMD and EEMD have also been conducted in hydrological forecasting, as described in Napolitano et al. [30], Kisi et al. [31], Feng et al. [32], Rezaie-balf et al. [33], Hadi et al. [34], Yu et al. [35], and Rezaie-Balf et al. [36]. These data pre-processing methods reliably obviate the shortcoming of AI models in dealing with the nonstationary and nonlinear signals by decomposing these time series into a set of simpler components to attain a deeper insight into the features [22,37]. Therefore, the prediction precision can be improved to a greater extent. Despite this, the weaknesses of these data pre-processing techniques, for instance, the high dependence of wavelet decomposition on the mother wavelet functions [38], the disadvantages in mode mixing and the lack of the strict mathematical theory of the EMD method [39], the large amount of computation cost, the uncontrollable modal components, and the unremovable noise of the EEMD [17] may affect the accuracy of predicting results.

The variational mode decomposition, newly proposed by Dragomiretskiy and Zosso. [40], is a completely non-recursive variation model for signal decomposition. This method has attracted much attention due to its solid theoretical foundation, strong noise robustness and precise component separation [41]. The hybrids AI and VMD models have successfully been employed in power quality events recognition [42], short-term load forecasting [43], time frequency analysis of Mirnov coil [44], stock price and movement prediction [45], short-term wind power generation forecasting [46], wind speed forecasting [47], and solar radiation forecasting [48]. In the hydrological domain, runoff and rainfall-runoff predictions were mainly focused upon. Seo et al. [49] proposed VMD-based ELM (VMD-ELM) and VMD-based least squares support vector regression (VMD-LSSVR) models for daily rainfall-runoff modeling in the Geumho River Watershed, South Korea. The results showed that the VMD-ELM and the VMD-LSSVR models presented the best performance when compared with the VMD-based ANN (VMD-ANN), discrete wavelet transform (DWT)-based single models (DWT-ELM, DWT-LSSVR, and DWT-ANN) and single models (ELM, LSSVR, and ANN). He et al. [41] simulated daily runoff by using the deep neural networks (DNN) coupled with VMD (VMD-DNN) in the Zhangjiashan hydrological station of Jing River, China. The results confirmed the superiority and novelty of the proposed hybrid model in daily runoff forecasting. Furthermore, the good stability and representativeness of the deep belief network (DBN) coupled with VMD in short-term runoff prediction were also reported by Xie et al. [17]. Despite the demonstrated significant potential and advantages of the AI coupled VMD methods, the use of these models in groundwater level prediction is rarely recorded, especially in the arid environment where groundwater has high irregularity, complex nonlinearity and multi-scale variability. It is, thus, important for groundwater level prediction to use the VMD based hybrid models to provide reliable scientific references in such conditions.

Especially, the uncertainty residing in the stochastic AI algorithms remains a problem that cannot be ignored when conducting these models. Therefore, uncertainty analysis is an indispensable procedure to get reliable simulation results since these models are susceptible to input data, and incapable of reproducing the same results even in identical situations. However, the uncertainty of hybrid models in the estimation process is often neglected in most cases, regarding the fact that the deterministic analysis of hybrid models is the main focus. The bootstrap method coupled with a resampling technique can be an effective method for uncertainty analysis due to the advantages of the more convenient computation process relative to derivatives and the Hessian-matrix involved in the delta method, or the Monte Carlo simulations involved in the Bayesian approach, and has been successfully used in a wide range of problems in hydrological modeling [50–53]. The applications of combining a machine learning method based on data decomposition technology and a bootstrap resamples method are described in detail in [51,52,54]. These researchers developed a hybrid wavelet-bootstrap-neural network (WBNN) model for forecasting hourly flood, daily discharge, and medium term urban water demand, respectively, and confirmed the superiority of the bootstrap method in assessing uncertainty. However, the uncertainty of the AI coupled decomposition ensemble models is usually ignored especially in groundwater level prediction. Therefore, the uncertainty analysis based on the bootstrap method was conducted to assess the precision of the hybrid models.

According to the above considerations, the main objective of this study is to propose a hybrid model combining the signal decomposition (VMD) with feature extraction (Boruta) and ELM (VMD-Boruta-ELM, briefed as VBELM) for the accurate simulation of 1, 2 and 3 month-ahead groundwater level. Specifically, the VMD was first used to decompose the time series of groundwater level, rainfall and temperature into various multiple intrinsic narrow-band sub-components called IMFs; then the feature selection was carried out by using the Boruta method to reduce input dimensions and to extract input variables of each IMF; next, the obtained IMFs of input parameters were imported into the ELM model to derive a corresponding forecast; last, all the predicted IMFs were summed as final output result. In addition, the single ELM model and the coupled VELM model without

Boruta method were also developed for comparison. Moreover, uncertainty analysis was performed for the VBELM model by the bootstrap sampling technique for the purpose of more accurate and reliable groundwater level prediction results.

2. Methodologies

2.1. Data Decomposition (VMD)

The VMD is established based on the Wiener filtering theory, Hilbert transform theory, and frequency shifting theory by adaptively decomposing a sophisticated signal into several band-limited IMFs (BLIMFs) with varying amplitude and frequency [44]. According to the VMD, the original signal $f(t)$ is decomposed into amplitude-modulated-frequency-modulated (AM-FM) signals in band-limited modes u_k ; the equation can be described as:

$$u_k(t) = A_k(t) \cos(\varphi_k(t)) \quad (1)$$

where $u_k(t)$ denotes the k th mode IMFs, and $A_k(t)$ is the instantaneous amplitude and phase. Besides, the derivative of $\varphi_k(t)$ is $\omega_k(t)$ which is called the instantaneous frequency.

There are three steps to estimate the bandwidth of $u_k(t)$,

- Step 1: Compute the related analytic signal of $u_k(t)$ by using the Hilbert transform to obtain a unilateral frequency spectrum.
- Step 2: Shift frequency spectrum of $u_k(t)$ to the baseband by mixing with an exponent tuned to the respective computed central frequency.
- Step 3: Estimate the bandwidth through H^1 Gaussian smoothness of the demodulated signal.

The objective of the optimization is to minimize the sum value of the estimated bandwidth. The constrained variational problem can be expressed as:

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \text{ s.t. } \sum_{k=1}^K u_k(t) = f(t) \quad (2)$$

where $\{u_k\}$ represents the decomposed modes, and can be expressed as $\{u_1, u_2, \dots, u_K\}$; $\{\omega_k\}$ is the corresponding center frequencies, namely $\{\omega_1, \omega_2, \dots, \omega_K\}$; k and t denote the number of modes and time, restively; and $\delta(t)$ denotes the Dirac function. The expression $\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t)$ can transform $u_k(t)$ into the analytical signal to form a unilateral frequency spectrum. Using $e^{-j\omega_k t}$ the spectrum of each mode can be shifted into the baseband, and ω_k is the center frequency of u_k .

To transform the above constrained variational problem into an unconstrained optimization problem, the quadratic penalty term and Lagrangian multiplier are considered; the augmented Lagrange function is given as follows:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) &= \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ &+ \|f(t) - \sum_{k=1}^K u_k(t)\|_2^2 + \langle \lambda(t) f(t) - \sum_{k=1}^K u_k(t) \rangle \end{aligned} \quad (3)$$

where $\|f(t) - \sum_{k=1}^K u_k(t)\|_2^2$ is a quadratic penalty term used to guarantee the good convergence velocity, and α and λ denote the balancing parameter of data fidelity constraint and Lagrange multiplier, respectively.

In the VMD algorithm, the optimization technique called alternate direction method of multipliers (ADMM) is adopted to seek out the saddle point of the augmented Lagrange function [55]. Based on the ADMM optimization approach, the u_k and ω_k can be updated in two directions to complete the analysis process of VMD, and the solution for u_k and ω_k is shown as [56]:

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

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

where $\hat{y}(\omega)$, $\hat{u}_i(\omega)$, $\hat{\lambda}(\omega)$ and $\hat{u}_k^{n+1}(\omega)$ denote the Fourier transform of $y(t)$, $u_i(t)$, $\lambda(t)$ and $u_k^{n+1}(t)$, respectively.

2.2. Feature Selection of Input Variables (Boruta)

Input variable selection is crucial to the development of AI models and is particularly relevant in water resources modeling [57]. In the current study, the Boruta method is employed to search for important predictors. Developed by Kursa et al. [58], the algorithm is a wrapper around the Random Forest classification algorithm implemented by the Boruta package in the R environment and is a promising feature selection algorithm that incorporates the interactions between the features [59]. The classification performed in this method is fulfilled by voting of multiple unbiased weak classifiers—decision trees [60]. The method has been successfully used and is strongly recommended [61].

The algorithm steps can be summarized as follows [60]:

Firstly, create shuffled copies of all features which are called Shadow Features to add randomness to the given data set and remove correlations between shadow variables and the response.

Then, train a random forest classifier in the extended information system and evaluate the importance of each feature using Mean Decrease Accuracy.

Next, determine the maximum Z score (computed by step 2) among shadow attributes (MZSA), and compare it with the corresponding shadows. Mark the attributes as “important” (tagged “Confirmed”) when its importance significantly higher than MZSA, otherwise regard as “unimportant” (tagged “Rejected”) and remove from the information system.

Finally, the Boruta algorithm ends when all attributes are confirmed or rejected or it reaches a specified limit of random forest runs. Noted that unassigned inputs after reaching limitation are classed as “Tentative” and obtain final decision (confirmed or rejected) by comparing the respective median Z scores with the median Z-scores of the best shadow variable [61].

2.3. Prediction Model (ELM)

The ELM proposed by Huang et al. [62] is a single-hidden layer feedforward network (SLFN) with a number of nodes that can be randomly generated in the hidden layer. The ELM method is easy to use and no parameters need to be tuned except for determining network architecture, for instance, the inputs and number of hidden layer neurons, therefore avoiding many complications faced by the gradient-based algorithms such as learning rate, learning epochs, and local minima [19,63]. Apart from the faster learning speed and better generalization performance, the ELM algorithm does not have problems such as local minima and over-fitting [64].

SLFNs functions with n hidden nodes and activation function $g(x)$ can be represented as:

$$f_n(x_j) = \sum_{i=1}^n \beta_i g_i(x_j) = \sum_{i=1}^n \beta_i g(\omega_i x_j + b_i) = t_i, \quad j = 1, 2, \dots, N \quad (6)$$

where β_i is the weight vector connecting the i th hidden node and the output node, ω_i is the weight vector linking the input layer to the i th hidden node, b_i is the threshold of the i th hidden node, and $\omega_i x_j$ denotes the inner product of vector ω_i and x_j . In order to comprehend further the computation process, Equation (6) can be rewritten as follow:

$$H\beta = T \quad (7)$$

where H is the hidden layer output matrix, and β represents the output weight matrix.

In total, the steps of ELM algorithm can be summarized as: (a) randomly generate input weights ω_i and threshold b_i before training; (b) calculate the hidden layer output matrix H ; (c) calculate the output weight matrix $\beta = H^+T$, where H^+ is called the Moore–Penrose generalized inverse of the hidden layer matrix H [65].

2.4. Uncertainty Analysis (Bootstrap)

The bootstrap is a resampling technique for statistical inference and also a computational, data-driven simulation method that generates multiple realizations from one dataset of a distribution or process [66]. It is commonly used to estimate confidence intervals, bias and variance of an estimator or to calibrate hypothesis tests. In terms of assessment of confidence intervals, the bootstrap is a widely accepted and reliable method relative to others [54]. Moreover, unlike Monte Carlo solutions relying on definite probability distribution, the bootstrap can estimate unknown sampling distribution using the data and computer power. The essential idea of the bootstrap approach is based on resampling with replacement of the available dataset and training of an individual network in each resampled instance of the original dataset [52]. The precision of a bootstrap estimate depends on the number of bootstrap replicates randomly resampled from the original dataset; the number of bootstrap samples should be at least 1000 samples as suggested by Efron and Tibshirani [67]. In this study, we randomly resample each IMF with replacement 1000 times, keeping the ratio between the training and testing data sets constant; training and testing the model obtains 1000 outputs of each IMF, and all output results are correspondingly summed as final result.

Given a set of independently and identically distributed random sample $T_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ consisting of a predictor vector x_i and the corresponding output variable y_i , n is a random dataset sample drawn from an unknown probability distribution F . Each (x_i, y_i) is randomly sampled with replacement from the original datasets. The set of S bootstrap samples generated and which can be denoted as $T^1, T^2, \dots, T^s, \dots, T^S$, S represents the total number of bootstrap samples and the value is 1000 in the present study. Each T^s is employed to train the VBELM model, and the output obtained is represented as $g_{VBELM}(x_i, \omega_b/T^s)$ which is evaluated using the observation pairs that are not included in a bootstrap sample T^s . Afterwards, the performance of the VBELM model in these validation datasets is averaged as an estimate of the generalization error of the VBELM model developed on T_n . This generalization error can be estimated by E_0 [68]:

$$E_0 = \frac{\sum_{s=1}^S \sum_{i \in A_s} [y_i - g_{VBELM}(x_i, \frac{\omega_s}{T^s})]^2}{\sum_{s=1}^S \#(A_s)} \quad (8)$$

where x_i denotes the input vector, ω_s denotes the weight vector, A_s is the set of indices for the observation pairs not included in the bootstrap sample T^s and $\#A_s$ is the number of observation pair indices in A_s . The VBELM estimate $y(x)$ is given by the mean of the S bootstrapped estimates [52]:

$$y(x) = \frac{1}{S} \sum_{s=1}^S g_{VBELM}(x_i, \omega_s/T^s) \quad (9)$$

The variance of the outputs of each VBELM is expressed as:

$$\sigma^2(x) = \frac{\sum_{s=1}^S \sum_{i \in A_s} [y_i - g_{VBELM}(x_i, \frac{\omega_s}{T^s})]^2}{S - 1} \quad (10)$$

The 95% confidence interval is used in this study for intuitive evaluation of uncertainty. It is determined by finding the 2.5th (X_L) and 97.5th (X_U) percentiles of 1000 outputs. The ratio of observed values within the 95% confidence interval is calculated to judge the

uncertainty of the final model; the higher the ratio is, the smaller the uncertainty is, and vice versa. This is represented as:

$$\text{Bracketed by 95PPU} = \frac{1}{n} \text{Count}(N | X_L \leq N \leq X_U) \times 100 \quad (11)$$

where n indicates the number of the observed data. N increases with observations falling between the corresponding X_L and X_U increase; the “Bracketed by 95PPU” denotes the number of observed data bracked by 95% confidence interval, this value is equal to 100 when all of the observed data are within the range of $X_L \leq N \leq X_U$.

Besides, the d -factor is applied for computing the average width of the confidence interval, which can be evaluated as [69]:

$$d - \text{factor} = \bar{d}_x / \bar{\sigma}_x \quad (12)$$

$$\bar{d}_x = \frac{1}{n} \sum_{i=1}^n (X_U - X_L) \quad (13)$$

where $\bar{d}_x$ is the average distance between the upper (97.5th) and lower (2.5th) bands, and $\bar{\sigma}_x$ is the standard deviation of observations. The better results indicate that the width of confidence interval is narrow, namely a d -factor value which is close to 0.

3. Model Development

In light of the foregoing discussion, a hybrid VBELM model integrating modal decomposition, feature selection and AI methods was proposed for monthly groundwater level prediction. As for the input data, the antecedent groundwater level and weather data were employed in most studies [4,8,33]. In the present study, the past total precipitation (P), average temperature (T) and groundwater level (GW) were used as inputs to predict future groundwater level. The maximum time lag was identified as 3, namely that the time series of P , T and GW for the past three months (1 time step represents 1 month) were employed to forecast the GW at the $t + 1$, $t + 2$ and $t + 3$ timescales Equation (14).

$$GW_{t+\Delta t} = f(P_{t2}, P_{t1}, P_t, T_{t2}, T_{t1}, T_t, GW_{t2}, GW_{t1}, GW_t) \quad (14)$$

where t denotes the current time, Δt is the lead-time period, and 1-, 2- and 3- time steps were selected in this study. t_2 and t_1 denote the prior 2- month time step and the prior 1- month time step, respectively.

For model construction, firstly, the VMD technique was adopted to decompose the original precipitation, temperature and groundwater series (for both training and testing sets) into the same number of IMFs with corresponding low–high frequency components to obtain stationary subseries. Secondly, the Boruta feature selection algorithm was used to select appropriate input components for each IMF during the training phase in order to build concise and efficient models. Thirdly, the ELM model was applied to simulate all IMFs. Finally, the predicted IMFs were obtained by aggregating the results of all obtained IMF components. This process was carried out for deterministic analysis. In terms of uncertainty analysis, each selected IMF by Boruta was randomly resampled 1000 times (see Section 2.4; running the ELM model obtains 1000 outputs for each IMF series. The prediction results of all IMFs are aggregated 1000 times to obtain the final 1000 outputs used as uncertainty analysis. The schematics of the process are illustrated in Figure 1.

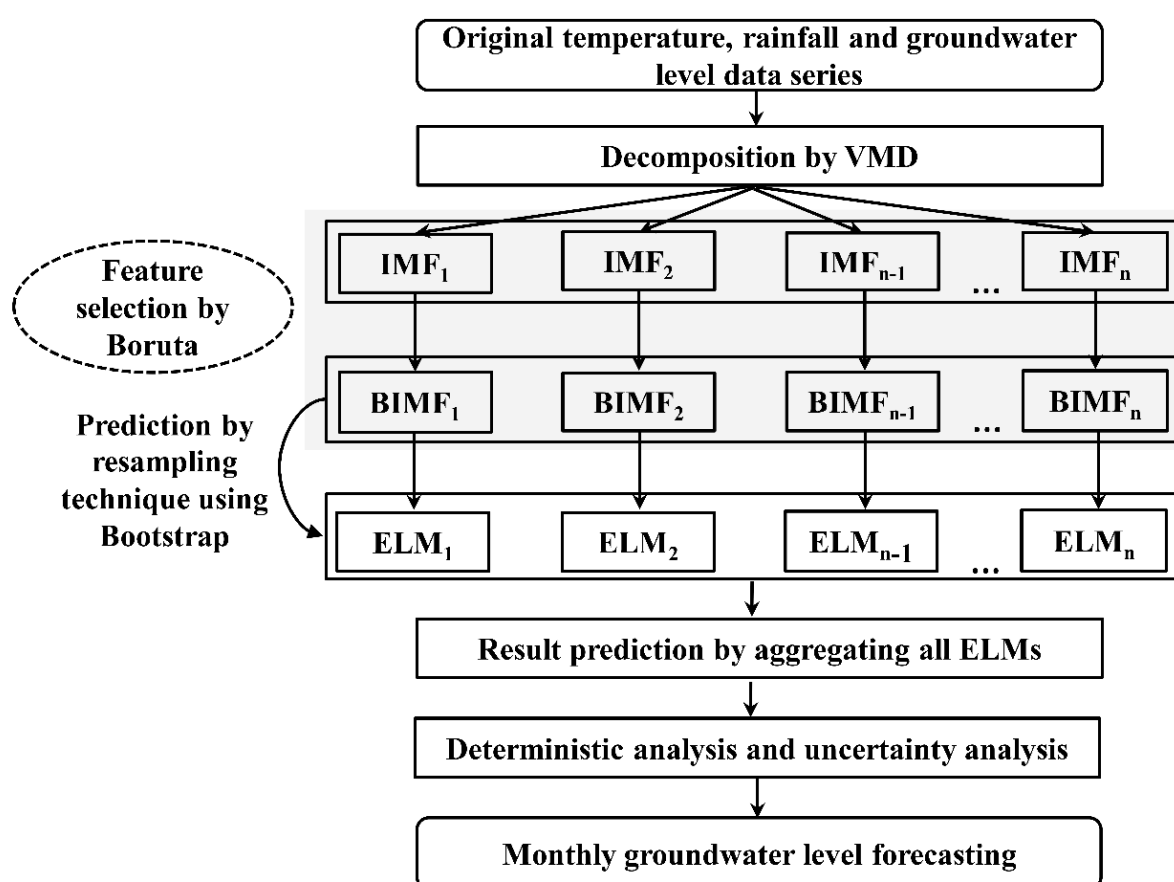


Figure 1. The schematic structure of the proposed forecasting modeling.

Note that the input and output data were normalized with a mean of 0 and a variance of 1 before running the model in order to eliminate influence of the dimensions. The equation used is as follows:

$$x_{new} = (x - \mu) / \sigma \quad (15)$$

where x_{new} is the normalized dimensionless data, and μ and σ are the average and standard deviation respectively.

4. Case Study

4.1. Study Area and Hydrological Data

The present study was carried out in the Zhangye basin, which is in the middle reaches of the Heihe River, northwest China (Figure 2), and holds an area of about $1.08 \times 10^4 \text{ km}^2$. It is a corridor plain between the Qilian and the Arkin mountains in the south, and the Mazong, the Heli, and the Longshou mountains in the north. The region belongs to a typical continental arid climate characterized by scarce but concentrated precipitation, intense evaporation and strong solar radiation, with the average annual air temperature changing from 3 to 7 °C year⁻¹, the mean precipitation ranging from about 50–150 mm·year⁻¹, and the mean potential evaporation rate varying from 2000 to 2200 mm·year⁻¹. The terrain of the study area gradually drops from the southeast to the northwest with a slope varying from 25% to 4%. The landform is divided into two types: the piedmont alluvial–proluvial Gobi plain in the south and the alluvial–proluvial fine soil plain in the middle of the region. An independent hydrogeological unit with recharge, runoff, and drainage processes is formed in the basin. Groundwater in this area is mainly the Quaternary pore water, and the depth of groundwater decreases gradually from southeast to northwest.

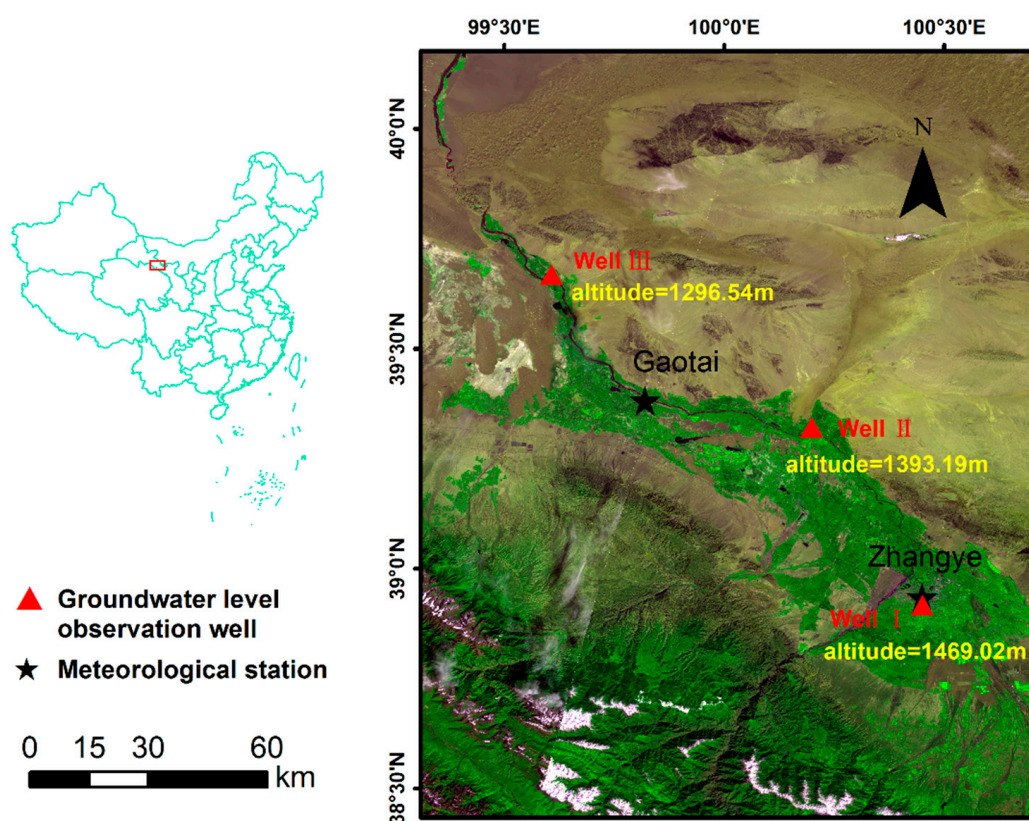


Figure 2. Location study area with the groundwater level observation well sites and meteorological sites.

The Zhangye basin, which can be regarded as the main commodity grain production base with the typical irrigated agriculture and dense population of arid regions in China, occupies a very important position along the Silk Road Economic Belt. Groundwater is the major source of water supply for crop irrigation, production and domestic use. However, the rapid population growth, oasis expansion and urbanization have recently increased the amount of groundwater use resulting in an over-extraction problem in this area [70]. The 2016 government report shows that the total area of groundwater overexploitation was 2418.30 km² and the amount of over-exploitation reached 192.64 million m³ in this area. The over-extraction of groundwater further decreases groundwater levels, threatening the sustainable development of the regional ecosystem. Therefore, accurate determination of groundwater level is important for the rational planning of water resources. In this study, three typical groundwater observation wells, namely well I, well II and well III, were selected according to altitude gradients, ranked from high to low as well I > well II > well III.

The monthly average groundwater level observations for the three wells were collected from 1 January 2000 to 31 December 2017. Accordingly, the meteorological data (P and T) of the Zhangye and Gaotai stations, which are located at this study area and closest to the observation wells, were obtained from the National Climatic Centre of the China Meteorological Administration at the same time scale. The groundwater level data of the three observation wells along with P and T records were selected as the model inputs. It should be noted that for well I the data from the Zhangye station were used while data observed from the Gaotai station were applied for well II and well III, respectively. Particularly, the whole 216 monthly data records were divided into training and testing parts. In the training phase, 168 months data from 1 January 2000 to 31 December 2013 accounting for about 77.8% of the total data set was used, while the remaining 48 months data ranging from 1 January 2014 to 31 December 2017 was used at the testing part. The statistical characteristics of monthly meteorological data and groundwater level data for the three wells are showed in Table 1. It can be seen that the groundwater level with

average values ranging from 1460.60 m to 1298.69 m during the observation period became shallower as the altitude decreased from well I to well III. Additionally, no significant differences between the training and the testing data were found, indicating the reasonable division of training and testing datasets. Thus, the training data records can be regarded as representatives of the total datasets, as sufficient information on groundwater system behavior was captured.

Table 1. Statistical parameters of groundwater level and climate data at three sites.

		Well I	Well II	Well III	Zhangye		Gaotai	
		GWL (m)	GWL (m)	GWL (m)	P (mm)	T (°C)	P (mm)	T (°C)
Max	All	1464.24	1384.72	1299.88	60.4	25.72	65.7	25.62
	Training	1464.24	1384.72	1299.99	55.7	25.72	65.7	25.62
	Testing	1463.51	1384.62	1299.75	56.7	24.33	61.4	24.89
Min	All	1455.02	1382.39	1295.23	0	−14.91	0	−14.4
	Training	1455.02	1382.39	1297.64	0	−14.91	0	−14.4
	Testing	1459.67	1382.43	1295.23	0	−11.29	0	−10.36
Mean	All	1460.60	1383.25	1298.69	11.433	8.472	9.688	8.795
	Training	1460.38	1383.25	1298.82	11.29	8.372	9.718	8.609
	Testing	1461.47	1383.25	1298.23	11.974	8.863	9.583	9.444
Std	All	1.951	0.467	0.864	14.316	11.374	12.100	11.332
	Training	2.074	0.444	0.414	14.432	11.401	11.976	11.421
	Testing	0.992	0.543	1.591	14.008	11.394	12.654	11.109
SK	All	−0.48	0.629	−2.604	1.515	−0.234	1.999	−0.237
	Training	−0.273	0.674	−0.313	1.554	−0.229	2.022	−0.241
	Testing	0.247	0.528	−1.087	1.410	−0.263	1.989	−0.215

Max is the maximum, Min is the minimum, Std is the standard deviation, SK is the skewness.

4.2. Decomposition Results

The VMD was used to decompose the raw data into high-frequency and low-frequency signals with the number of K . It should be noted that the parameter K , which must be pre-defined, is an important parameter in VMD decomposition. The number of K has a significant influence on the decomposed results because too few K may result in the loss of the signal, while too many K may yield abundant frequency or extra noise [46,55]. To find the appropriate number of K , the central frequency method, which is based on the change of the center frequency (ω_k) of the last IMF, was adopted in this paper [71]. There are 9 input variables (P_{t2} , P_{t1} , P_t , P_{t2} , T_{t1} , T_{t2} , T_{t1} , GW_{t2} , GW_{t1} , GW_t) and 3 outputs (GW_{t+1} , GW_{t+2} , GW_{t+3}) for each site. The K for the three wells was determined starting from $K = 1$, and it was found that the number of K changed from 9 to 14, and the ω_k of IMF gradually became smaller as the number of K increased; moreover, slight variation of the ω_k for all the input-outputs could be found when K is more than 10. In this context, $K = 10$ was chosen as the optimal value for the three observation wells. It is worth noting that the determination of the number of K aforementioned was practiced by training data series; for the testing data series, the same number of modes is applied meaning that 10 IMFs (denoted as IMF_1 , IMF_2 , ..., IMF_{10}) were decomposed. The parameters of the VMD algorithms are presented in Table 2.

Table 2. Statistical parameters of groundwater level and climate data at three sites.

Parameters	Alpha	Tau	K	DC	Init	Tol
Value	2000	0	10	0	1	1×10^{-7}

Taking the raw groundwater level data during the training periods of the three observation wells as examples, Figure 3 shows the decomposition results of VMD. It can be

seen that the frequencies of subcomponents raised gradually as the K values increased. Each component played a distinct role in the raw data series with the lower frequency subseries representing the periodicity and trend (the IMF_1 contains low frequency components), and the higher ones reflecting the local volatility tendency and the small-scale features of the original groundwater level series. Besides, the difference between the subcomponents in the three observation wells vary substantially, mainly caused by different groundwater behavior of the three sites. The finding shows the feasibility and flexibility of the VMD method in extracting the intrinsic information from original groundwater level data.

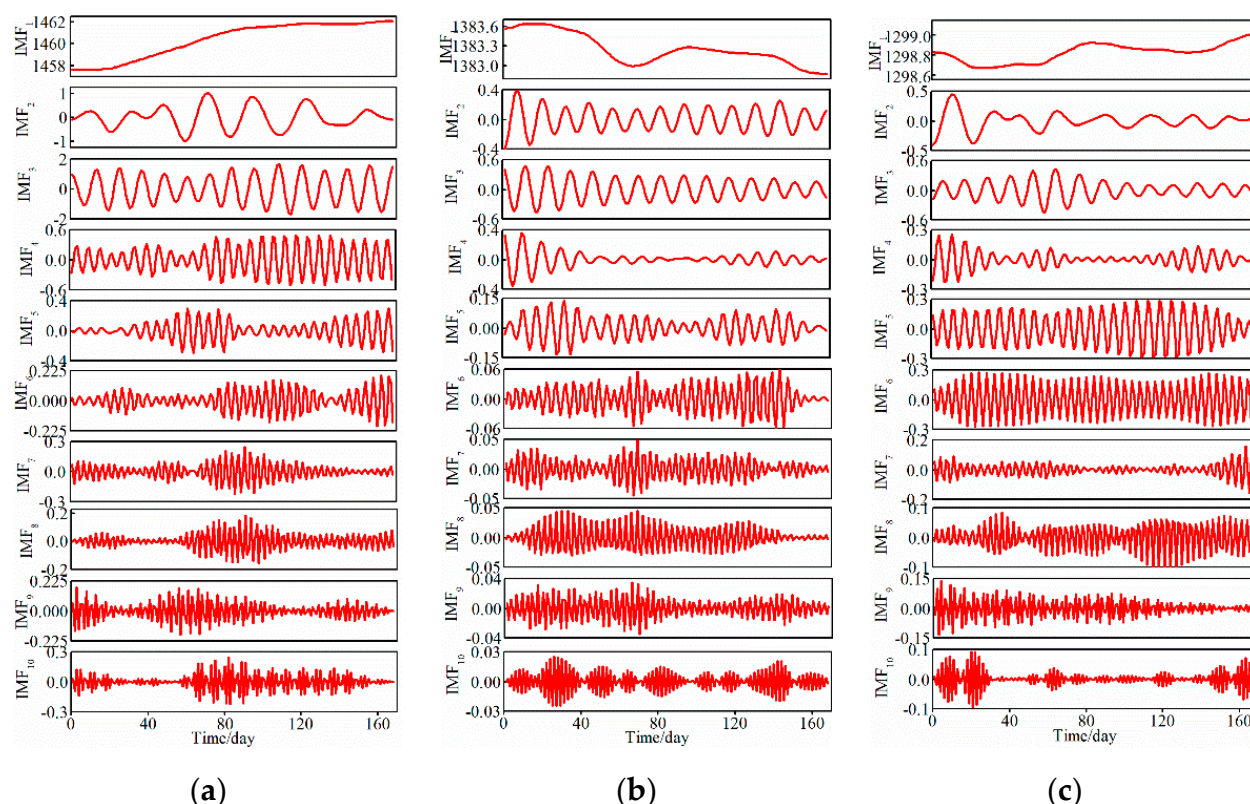


Figure 3. Decomposed results of training periods by variational mode decomposition (VMD) for monthly groundwater level at Well I (a), Well II (b) and Well III (c).

4.3. Determination of Input Variables

The determination of important input features has direct impacts on the model's performance. However, no definite statements about the choice of the methods is reported so far. It is certain that the feature selection process used in this study, which is intended to reduce the input dimensions and to provide an insight into underlying physical processes without altering data, is beneficial for increasing efficiency and optimizing the model's performances [61]. It is relevant to note that, prior to feature selection, all input and response variables should be normalized between 0 and 1.

For detailed explanation, the IMF_5 of well I was used as an example to explain the process, and to show the important input variables of IMF_5 (GW_{t+1}), as displayed in Figure 4.

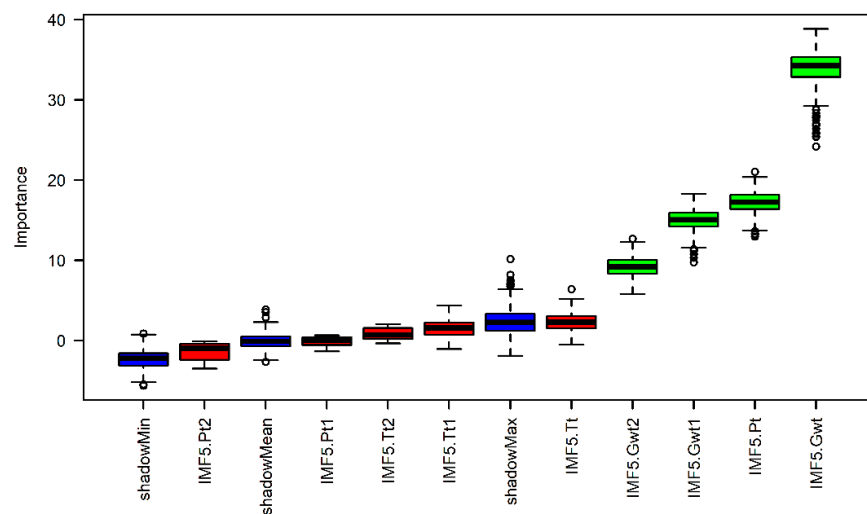


Figure 4. Box result plot by Boruta input selection algorithm.

In terms of the Boruta result plot, the blue boxplots represent the minimal, average and maximum Z scores of the shadow attribute. The red and green boxplots correspond to the Z scores of the rejected and confirmed attributes, respectively. It can be seen from this figure that five attributes were identified as unimportant variables (the fifth mode of P_{t2} , P_{t1} , T_{t2} , T_{t1} and T_t), and four attributes were tagged as important variables including IMF_5 (P_t), IMF_5 (GW_{t2}), IMF_5 (GW_{t1}), IMF_5 (GW_t).

4.4. Performance Criteria

The statistical indices including the coefficient of correlation (R), root mean square error (RMSE), mean absolute error (MAE) and Nash-Sutcliffe efficiency coefficient (NS) were applied in this study. The R measures the correlation between the estimated and observed values, and the NS is used to assess efficiency of forecasting model. The smaller the differences between the R or NS values and 1, the higher the accuracy and stability of the prediction models. RMSE and MAE provide distinctive qualifications for the prediction capability of the models. The RMSE demonstrates the goodness-of-fit relevant to the high values whereas the MAE provides a more balanced perspective, especially for the moderate values [72]. The small RMSE and MAE values mean a tiny error between the estimated and observed values, as well as satisfactory performance of the models. The R , RMSE, MAE and NS can be calculated by the following equations:

$$R = \frac{\sum_{i=1}^n (EP(i) - \overline{EP(i)}) (EO(i) - \overline{EO(i)})}{\sqrt{\sum_{i=1}^n (EP(i) - \overline{EP(i)})^2 \sum_{i=1}^n (EO(i) - \overline{EO(i)})^2}} \quad (16)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (EP(i) - EO(i))^2}{n}} \quad (17)$$

$$MAE = \frac{\sum_{i=1}^n |(EP(i) - EO(i)) / EO(i)|}{n} \quad (18)$$

$$NS = 1 - \frac{\sum_{i=1}^n (EO(i) - EP(i))^2}{\sum_{i=1}^n (EO(i) - \overline{EO(i)})^2} \quad (19)$$

where $EP(i)$ and $EO(i)$ represent the predicted and observed groundwater level for the i -th sample, respectively, and $\overline{EP(i)}$ and $\overline{EO(i)}$ are the average of the $EP(i)$ and $EO(i)$, respectively.

n means the sample number. The combination of $R = 1$, $RMSE$ and $MAE = 0$, and $NS = 1$ denotes the perfect fit.

5. Results and Discussion

In this section, the performance of the proposed hybrid VBELM model was discussed to forecast the 1-, 2- and 3-month ahead groundwater level for three wells. The VELM model without feature selection technique and the single ELM model were also assessed for comparison. It should be noted that it is during the testing phase that the performance in the prediction of groundwater level was demonstrated and compared of these models. In addition, the uncertainty assessment was further adopted to verify the simulation results of the proposed VBELM model.

5.1. Model Performance

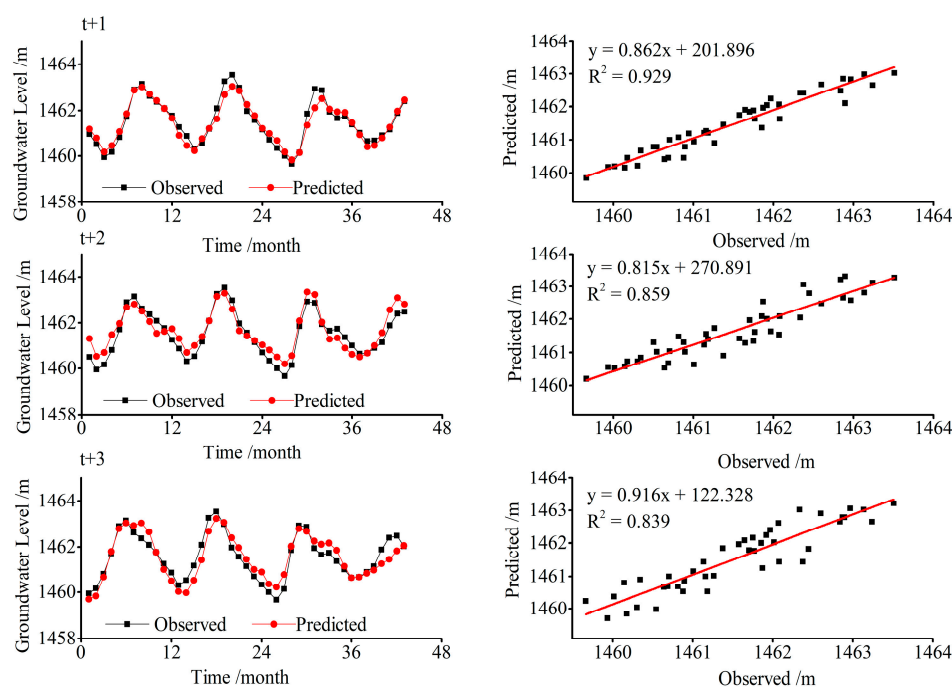
The precision of the proposed VBELM models for groundwater level forecasting at the period of 1-, 2- and 3-months is shown in Table 3 in terms of R , $RMSE$, MAE and NS . It can be seen that the VBELM models exhibited good prediction accuracy at the three sites, with high R and NS values as well as low $RMSE$ and MAE values. In particular, all of the R and NS values surpassed 0.9 and 0.8, respectively, while the $RMSE$ and MAE values were less than 0.6 and 0.5, respectively, indicating the satisfactory performances of this model in groundwater level predicting at multiple timescales. In addition, the accuracy of the VBELM models decreased with the increasing lead time of forecast, as displayed unambiguously and consistently in the three tables. This may be due to the reduction of data patterns. Concretely, from 1-month ahead to 3-month ahead prediction, the R , $RMSE$, MAE and NS values deteriorated by 5.2%, 32.6%, 35.6% and 11.2% for well I, respectively; 4.1%, 10.8%, 14.1% and 6.2% for well II, respectively; and 3.8%, 37.5%, 40.6% and 7.1% for well III, respectively. Although the precision of the hybrid model gradually deteriorated from the 1 monthly time-step to the 3 monthly time-step, the four statistical evaluation values remained within reasonable accurate range, providing satisfactory simulating results at all lead timescales for the three sites.

Table 3. Performance analysis of the VBELM models during testing periods for three sites.

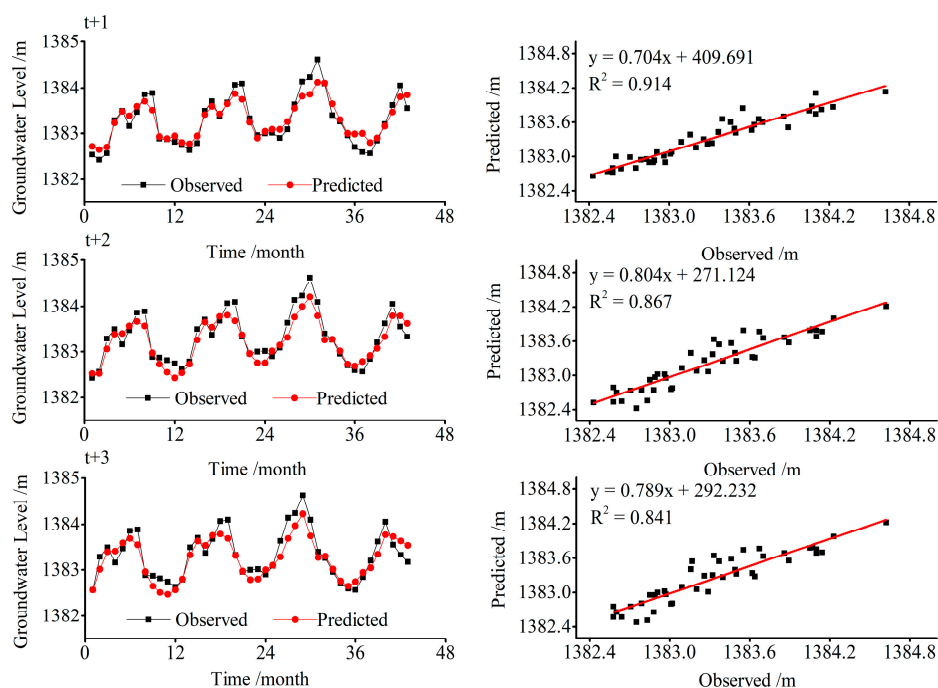
Observation Sites	Lead Month	R	Root Mean Square Error (RMSE) (m)	Mean Absolute Error (MAE) (m)	NS
Well I	1-month ahead	0.964	0.273	0.217	0.924
	2-month ahead	0.927	0.399	0.352	0.840
	3-month ahead	0.916	0.405	0.337	0.831
Well II	1-month ahead	0.956	0.199	0.159	0.865
	2-month ahead	0.932	0.214	0.180	0.838
	3-month ahead	0.917	0.223	0.185	0.811
Well III	1-month ahead	0.979	0.354	0.280	0.948
	2-month ahead	0.976	0.420	0.337	0.931
	3-month ahead	0.942	0.566	0.471	0.881

The hydrographs and scatter plots of groundwater level estimated by the VBELM model during the testing period for the three wells at 1, 2 and 3 months ahead forecasting are displayed in Figure 5. The left hydrographs clearly illustrate that the prediction results of the three wells were in line with the expectation values. That is to say, the evaluated groundwater levels were closer to the observations and followed the same trend. Meanwhile, the longer the lead times, the worse the simulation performances for groundwater level forecasting with the results of the 1-month ahead fitting the best, while those of the 3-months ahead fitted the worst. Besides, the superior simulation results of well I can also be distinctly discovered when comparing with that of well II and well III as the predicted values effectively fitted the fluctuations of the real values at both peaks and valleys for well I. Besides, it can also be observed that the scatter plots for well I

are less dispersive and closer to the fitted lines than for well II and well III, which again proves the above result. Moreover, the same gradually decreasing tendency can also be discovered from the R^2 values with the prediction time getting longer. This finding further certifies the deteriorating fitting performance of the hybrid model with an increase in the forecast time. This is in accordance with many other studies in groundwater level estimation [2,23,73].



(a)



(b)

Figure 5. Cont.

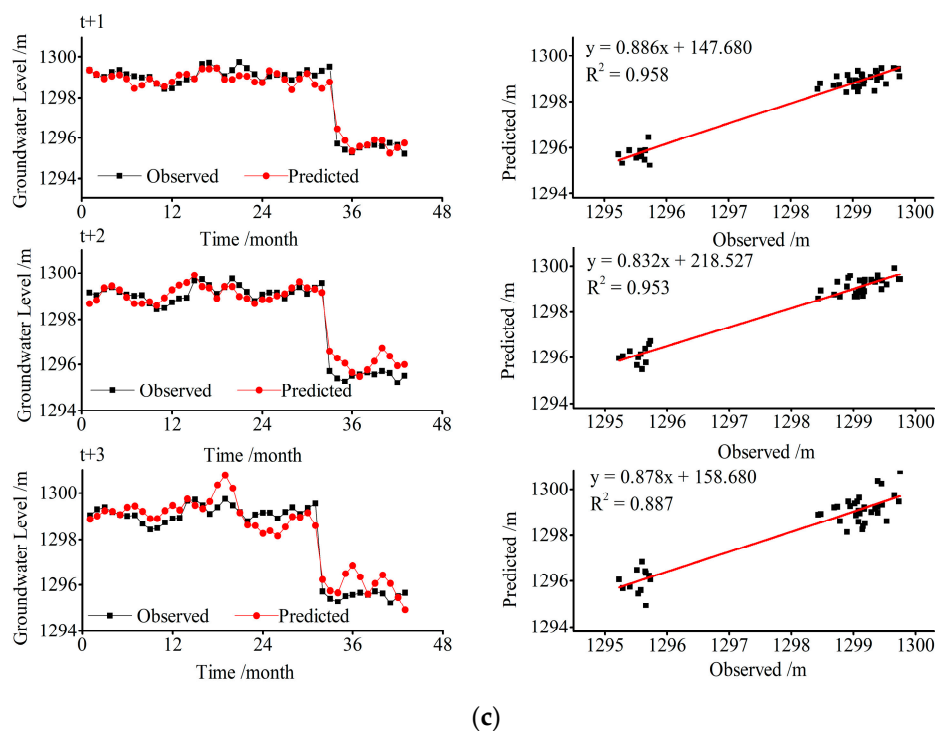


Figure 5. Observed groundwater level and forecasted by VBELM model for 1, 2 and 3-month ahead during the testing period at Well I (a), Well II (b) and Well III (c).

5.2. Model Comparison

In this study, the capability of the VELM model with the absence of a feature extraction method, and the single ELM model taking the original data series as sole inputs, were discussed in comparison in order to evaluate prediction accuracy of the VBELM model for 1, 2 and 3 monthly ahead groundwater level forecasting. Tables 4 and 5 list the four statistical indicators of the VELM and ELM, respectively, for well I, well II and well III. By contrast, the better performance that the VBELM models derived than the VELM and ELM models for 1-, 2- and 3-month ahead groundwater level forecasting can be observed. The simulation precision of the three models can be ranked as: VBELM > VELM > ELM. Concretely speaking, the VBELM achieved higher R and NS values as well as lower RMSE and MAE values for the three observation wells, and the standalone ELM model was found to be the worst. In terms of the four evaluation indexes, the performances of the two hybrid models exceeded the standalone ELM model too far, since all the R and NS values were more than 0.9 and 0.8 respectively, and the RMSE and NS values were less than 0.7 and 0.6 of the two hybrid models, respectively. Nevertheless, the ELM model yielded significant reduction of the R and NS values, with evident elevation of the RMSE and MAE values. This is especially true for the 2- and the 3-monthly ahead groundwater level forecasts since the NS value of the 3 months ahead forecast was only up to 0.706 for well I, the NS values were less than 0.7 for the 2 and 3 months ahead forecast for well II and well III, and the R value was less than 0.8 for 3 months ahead forecast for well III. This is mainly due to the fact that the VMD method remarkably improved the performance of the combined models when embedded into the ELM model. In addition, the VBELM model performed slightly better than the VELM model in terms of the statistics indicators, indicating that the Boruta technique contributed to an improvement in the performance of the VBELM model. Besides, the outcomes reaffirm the fact that the performance of the VELM and ELM models deteriorated when the lead time of forecast increased. The finding is consistent with the results obtained from the metrics values. When it comes to the stability of the models, it can be seen that the ELM led to the most dramatic deteriorations in light of the metrics values, followed by the VELM model, while the VBELM model exhibited the most

stable deteriorating rate. Thus, it can be concluded that the proposed VBELM method is an effective tool for groundwater level forecast, especially for the longer lead time prediction.

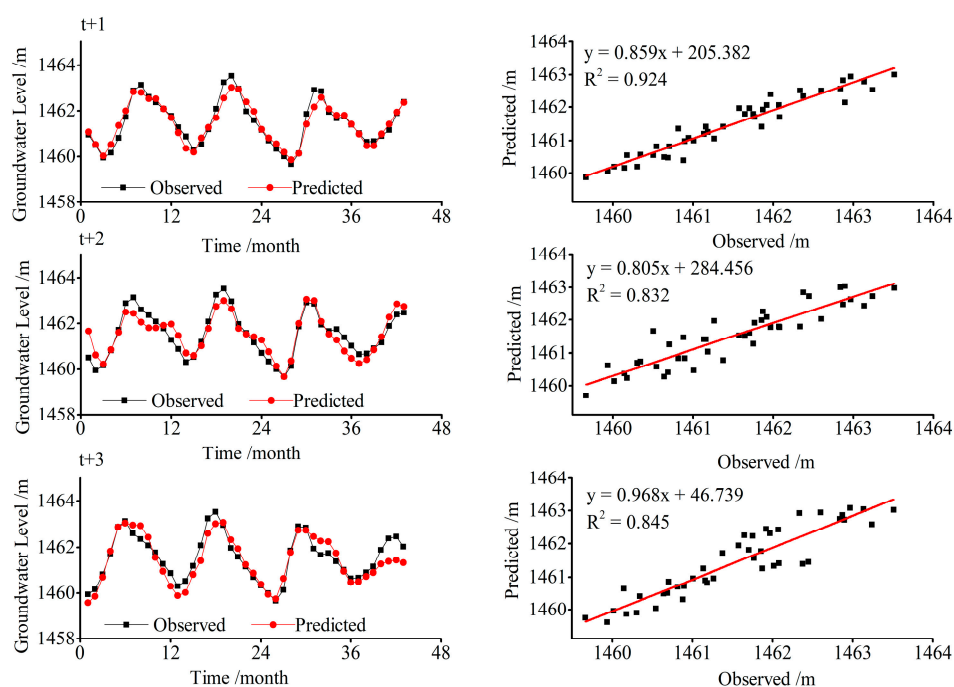
Table 4. Performance analysis of the VELM models during testing periods for three sites.

Observation Sites	Lead Month	R	RMSE (m)	MAE (m)	NS
Well I	1-month ahead	0.961	0.282	0.210	0.919
	2-month ahead	0.912	0.410	0.335	0.830
	3-month ahead	0.919	0.423	0.344	0.817
Well II	1-month ahead	0.924	0.209	0.167	0.852
	2-month ahead	0.927	0.220	0.183	0.827
	3-month ahead	0.914	0.229	0.198	0.801
Well III	1-month ahead	0.977	0.362	0.285	0.946
	2-month ahead	0.972	0.450	0.344	0.921
	3-month ahead	0.949	0.641	0.504	0.847

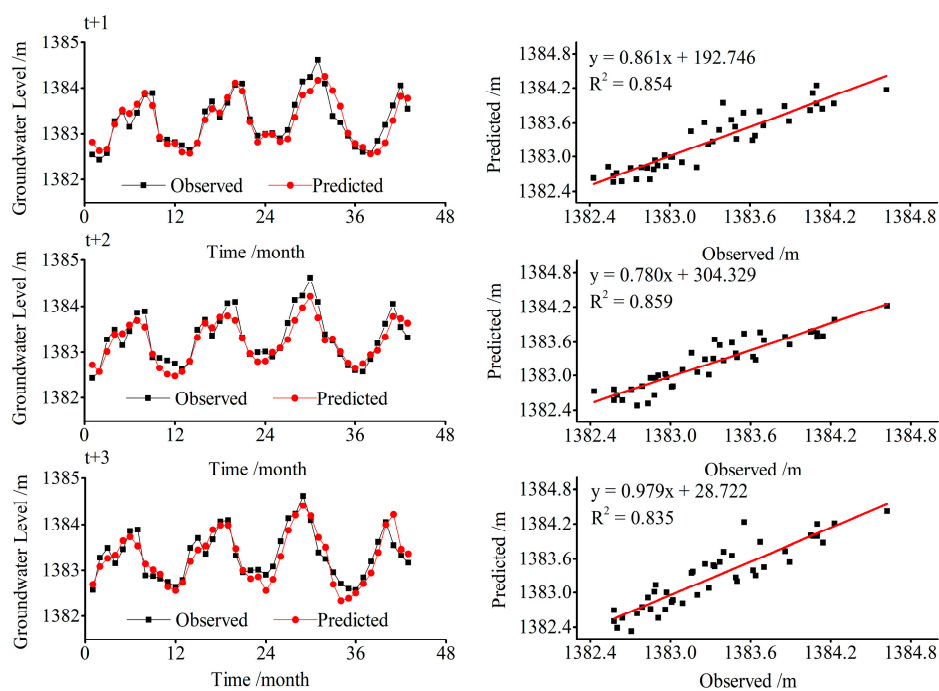
Table 5. Performance analysis of the ELM models during testing periods for three sites.

Observation Sites	Lead Month	R	RMSE (m)	MAE (m)	NS
Well I	1-month ahead	0.928	0.374	0.288	0.857
	2-month ahead	0.915	0.424	0.342	0.819
	3-month ahead	0.853	0.535	0.417	0.706
Well II	1-month ahead	0.903	0.238	0.191	0.808
	2-month ahead	0.835	0.307	0.264	0.666
	3-month ahead	0.815	0.314	0.261	0.625
Well III	1-month ahead	0.913	0.645	0.368	0.828
	2-month ahead	0.843	0.890	0.574	0.691
	3-month ahead	0.786	1.016	0.629	0.616

Figures 6 and 7 show the hydrographs and scatter plots of the VELM and ELM models for the three wells. It can be observed that the predicted values fitted the trend of the real values for both the two methods satisfactorily. Combining with Figure 4, it can be found that the predicted values obtained by the VBELM model better traced the fluctuation of the real values relative to the VELM and ELM models. Besides, it is worthwhile to note that the VBELM model achieved the best performance at both peaks and valleys among the three models, while the poorest performance can be found in the ELM model. Likewise, the best fitting outcomes from the short to long scenarios were still derived by the VBELM regarding the fact that the fitting performance of the three models descended with the increase in lead time. The results reasserted the superior performance of the proposed VBELM method for monthly groundwater level forecasting.



(a)



(b)

Figure 6. Cont.

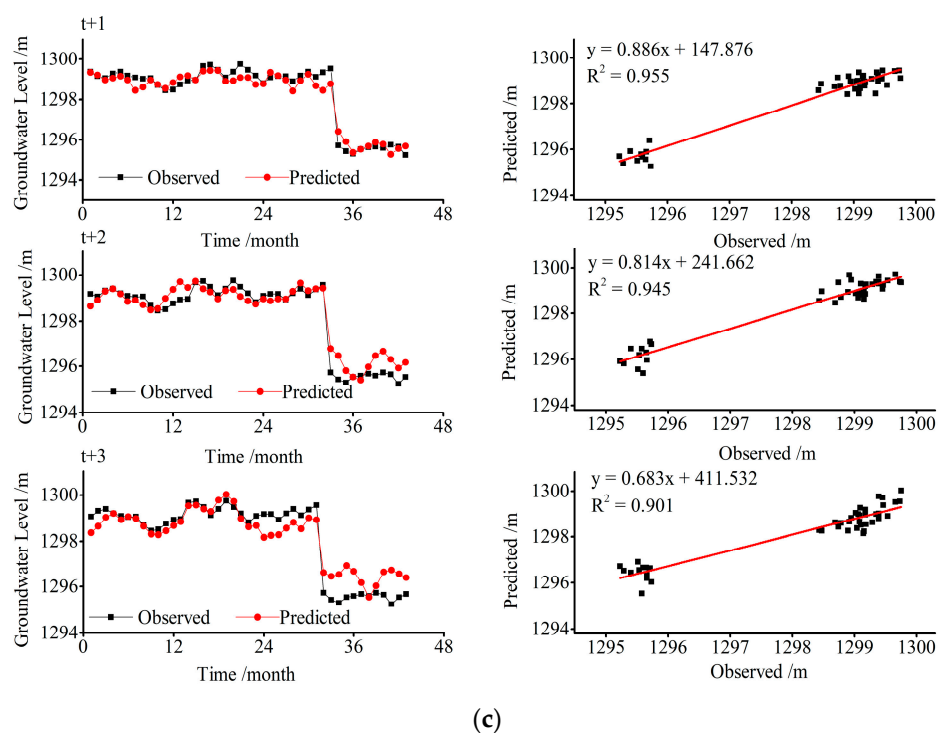


Figure 6. Observed groundwater level and forecasted by VELM model for 1, 2 and 3-months ahead during the testing period at Well I (a), Well II (b) and Well III (c).

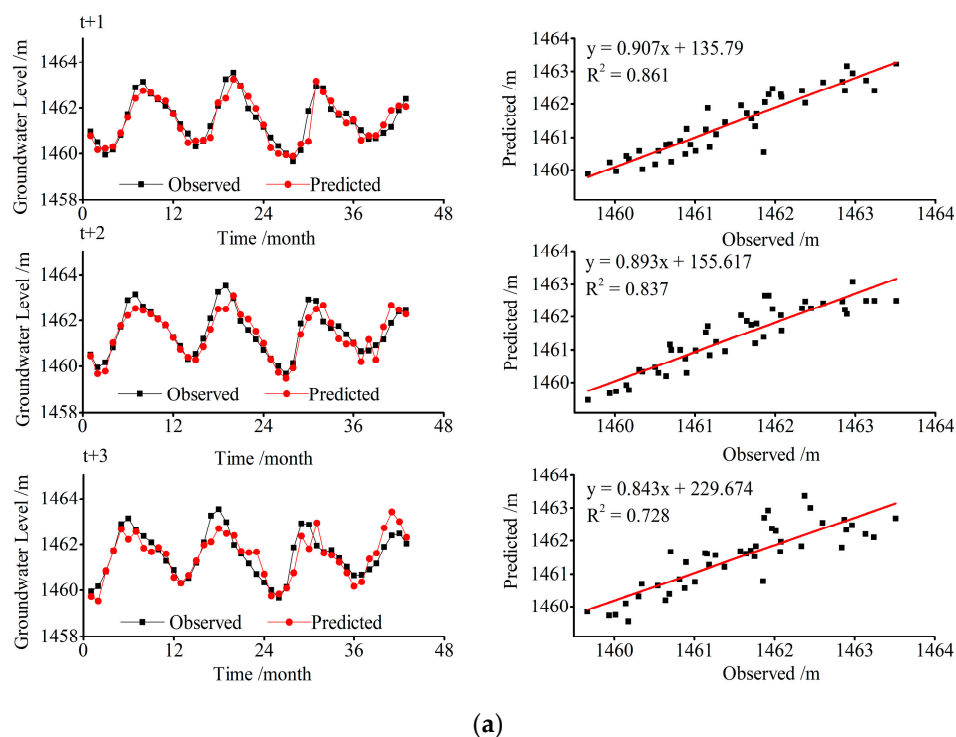


Figure 7. Cont.

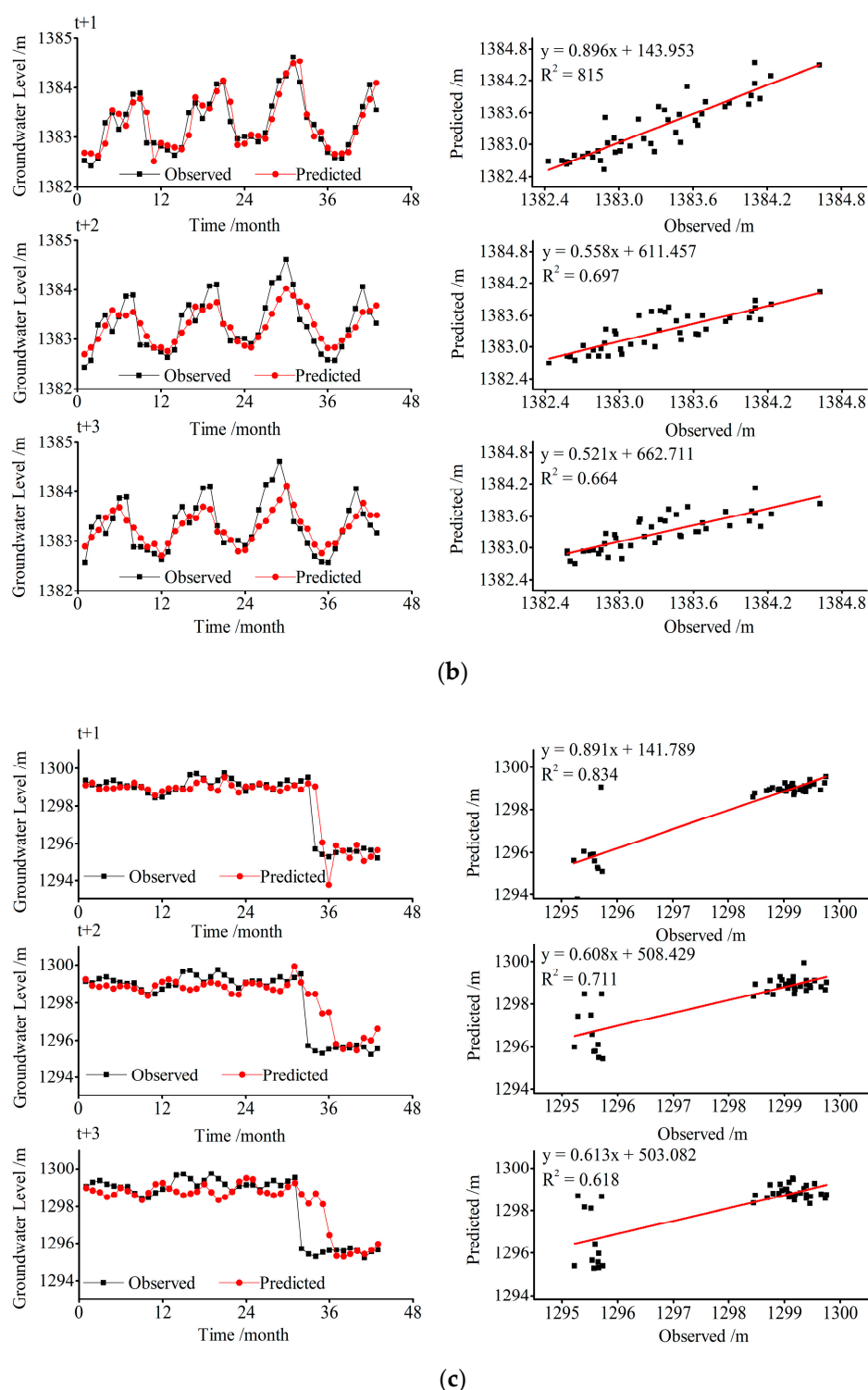


Figure 7. Observed groundwater level and forecasted by ELM model for 1, 2 and 3-months ahead during the testing period at Well I (a), Well II (b) and Well III (c).

It is important to evaluate the distribution of estimation errors as well as the average estimation error when assessing the performance of AI models in groundwater level modeling. Figure 8 shows the boxplots of error distribution of the VBELM, VELM and ELM methods for the three wells by comparing the degree of dispersion and skewness of the error. Thereto, the lower and the upper end of the boxplot denote the 25th and the 75th percentiles, respectively, the line and the small square inside the box denote the

median and mean, respectively, the outliers outside the box represent the values >1.5 interquartile (the cross line). What can be derived from Figure 8 is that the VBELM and VELM models produced less error distribution range than the ELM at 1, 2 and 3 months ahead groundwater level prediction for the three wells. Alternatively, the error of the two coupled models were more closely arranged with fewer outliers as compared with the standalone ELM model. For the two proposed hybrid models, the VELM performed slightly worse than the VBELM with respect to distribution and arrangement of the error. The availability and efficiency of the hybrid VBELM algorithm was then further verified.

According to the above analysis, the superiority of the hybrid VBELM and VELM models in groundwater level prediction over the standalone ELM model for the current study area was verified. Thanks to the advantage of the distinguished VMD pre-processing technology, frequency changes were effectively identified, and the complexity and non-stationary nature of the original time series were commendably reduced by decomposing the original monthly groundwater level, rainfall and temperature data into several sub-components with natural frequencies. Besides, the importance of feature selection in improving predicting accuracy was proved by the outperformed VBELM method. Concretely, by coupling with the Boruta method, the dimension, error and complexity of the VBELM model were conspicuously reduced from the correlated variables, biases and noises. In a word, the proposed hybrid VBELM model can be treated as a promising alternative for both short-term and long-term monthly groundwater level forecast in view of the great improvement in prediction accuracy.

5.3. Uncertainty Analysis

Despite the excellent outcome yielded by the VBELM in groundwater level forecasting, uncertainty may exist. To solve this, uncertainty assessment can be an indispensable procedure for reliable simulation results. In this section, the bootstrap method was carried out to evaluate the uncertainty of the VBELM by predicting 95 PPU of the results, and to corroborate the applicability of this model in modeling groundwater level. Note that it is the uncertainty of the two hybrid VBELM and VELM models that was compared in this section, in consideration of their better performance.

Figures 9 and 10 illustrate the 95% confidence interval of the 1-, 2- and 3-monthly ahead groundwater level estimates for the three wells during the testing period. The number of the observed data lying within the 95 PPU and the *d-factor* were applied to quantify the uncertainty of simulation results. Both the figures demonstrated a good match between the 95 PPU and the obtained results, as most of the observed values fell within the confidence intervals. It is also observable that, by and large, the number of the actual values bracketed by the 95 PPU decreased as the lead time increased. In terms of the width of the confidence intervals, all the *d-factor* values of the three wells were less than 1 (*d-factor* < 1 is considered as appropriate), indicating that the predicted confidence band was narrow and reached a lower level. Besides, it should be noticed that the width of the 95% confidence interval for a short lead time was less than that for long lead times as a whole. Although the *d-factors* for the 2-months ahead forecasts were smaller compared with the 1-month ahead results for well I and well II, the number of the observed values of the two wells falling within the confidence interval for the 1-month ahead forecast was higher than those of the 2-months ahead forecast. In this context, the uncertainty of the 1-month ahead forecast was considered as low for the two wells.

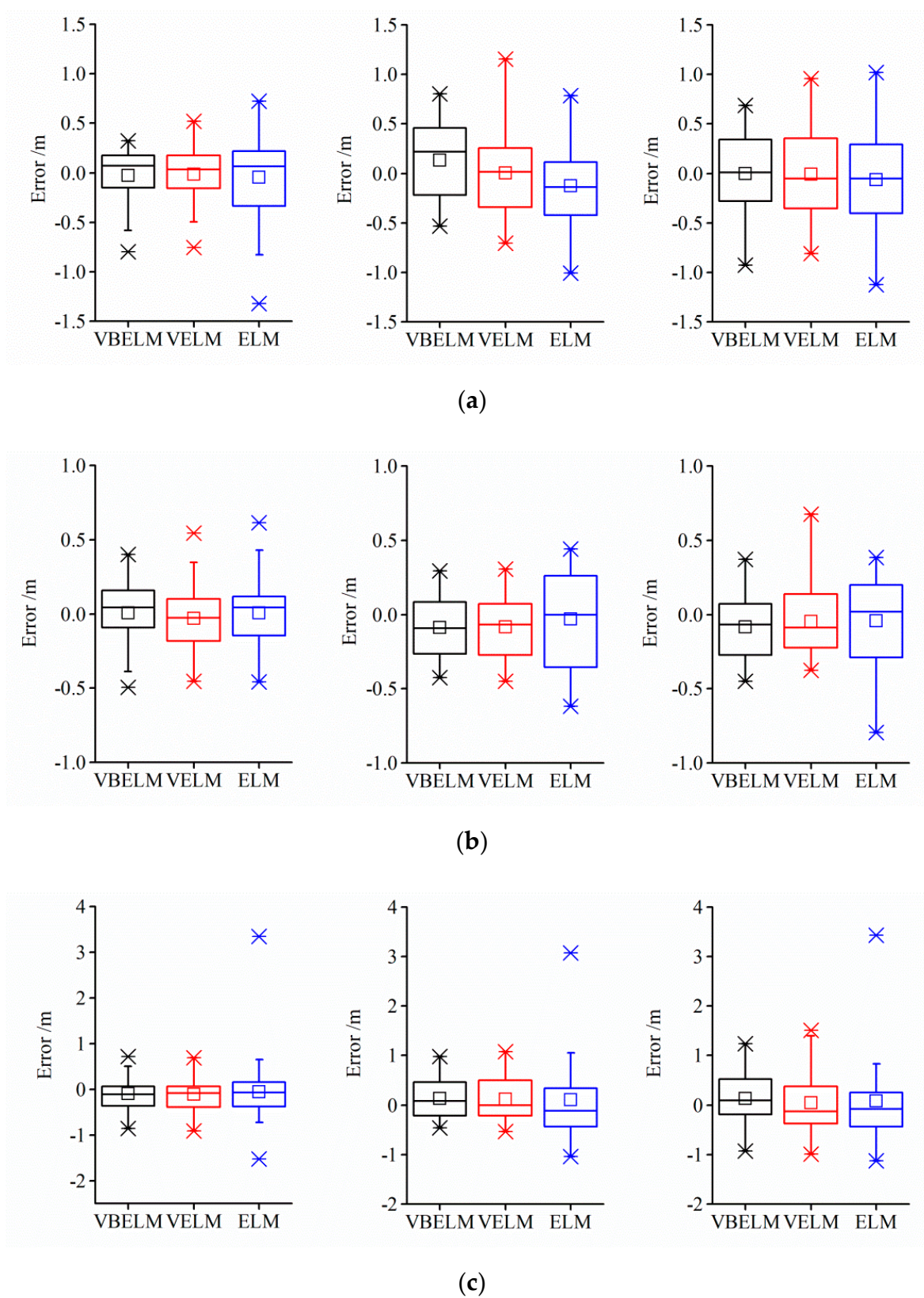
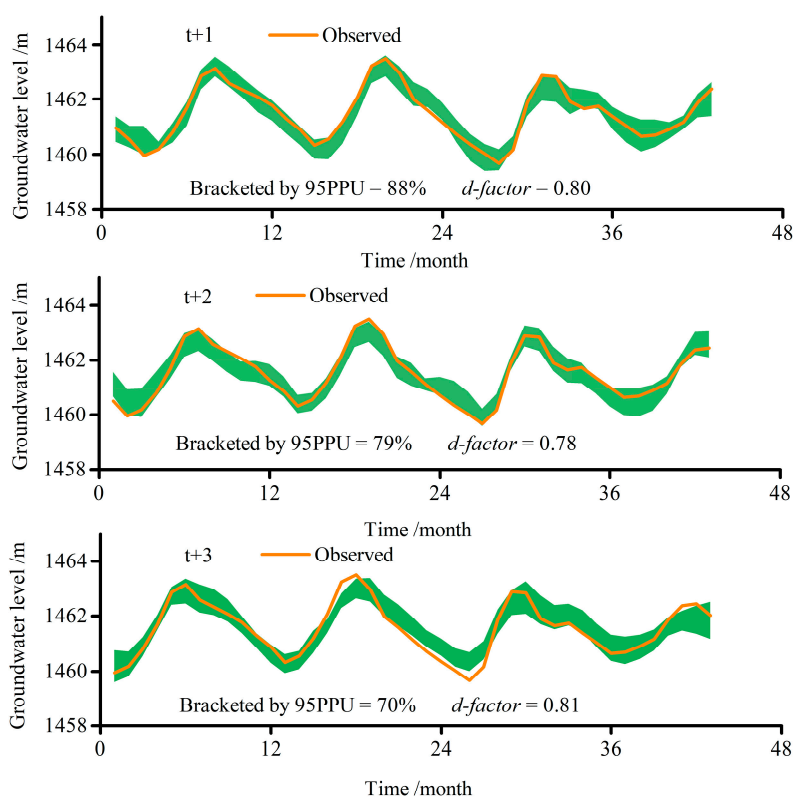
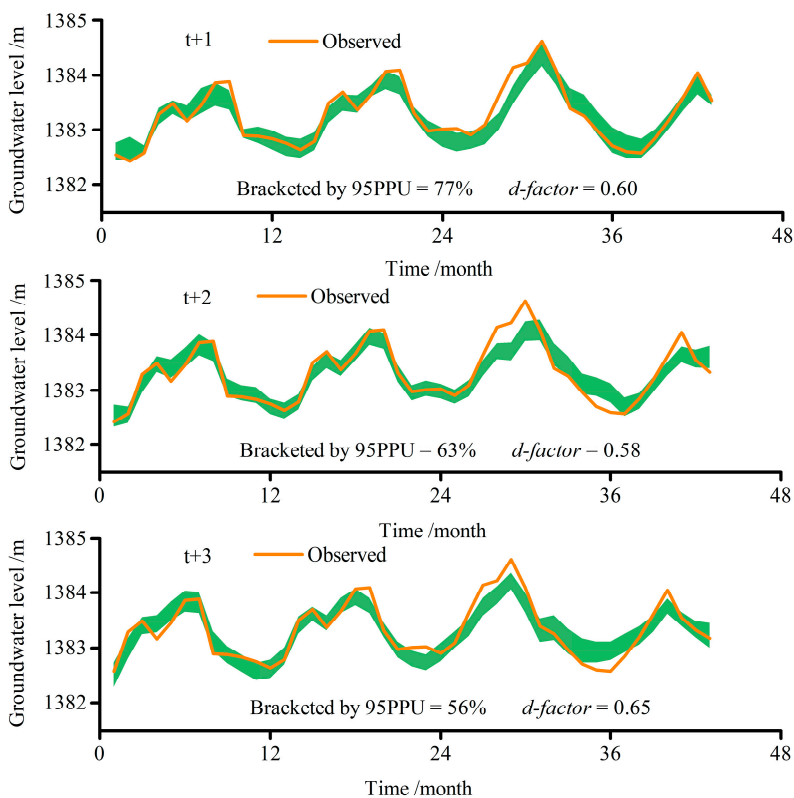


Figure 8. Boxplot of the error of 1-, 2-, and 3-months ahead groundwater level forecasted by VBELM, VELM and ELM models during the testing period at Well I (a), Well II (b) and Well III (c).



(a)



(b)

Figure 9. Cont.

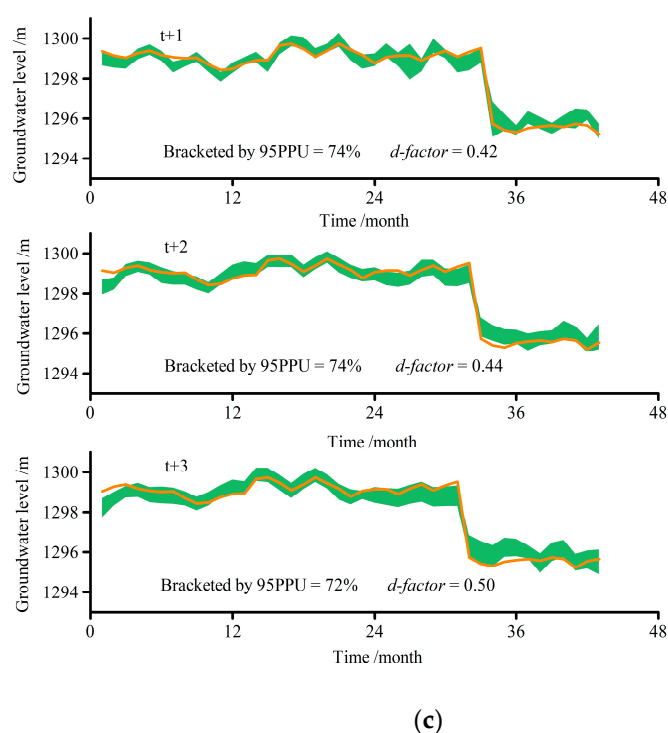


Figure 9. Observed groundwater level and the 95% confidence intervals forecasted by VBELM model with randomly sampled input vectors for 1, 2 and 3-months ahead forest during the testing period at Well I (a), Well II (b) and Well III (c).

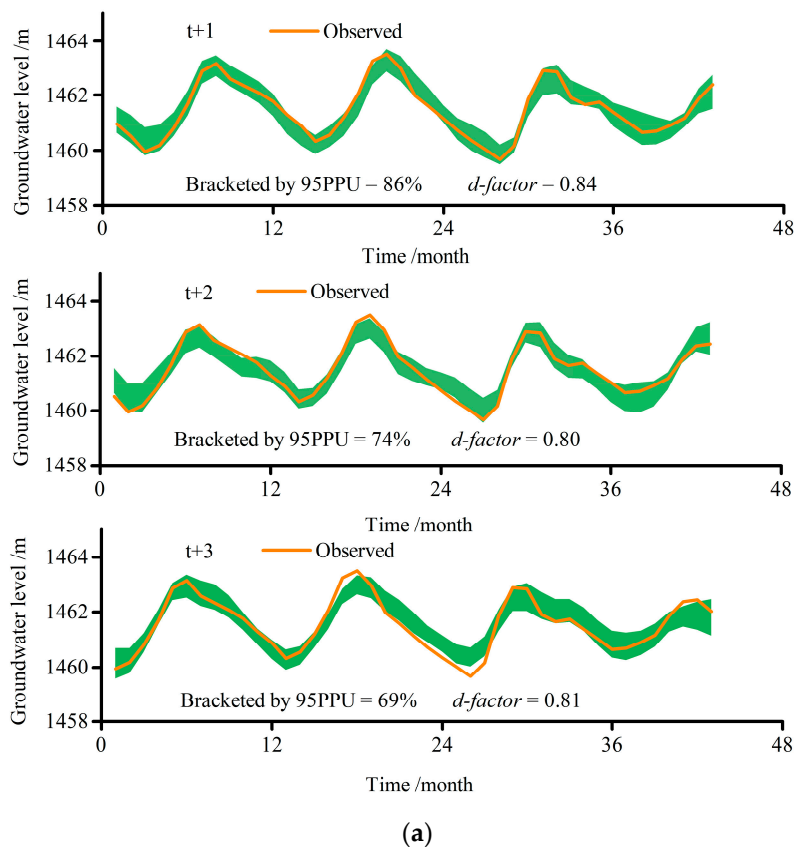
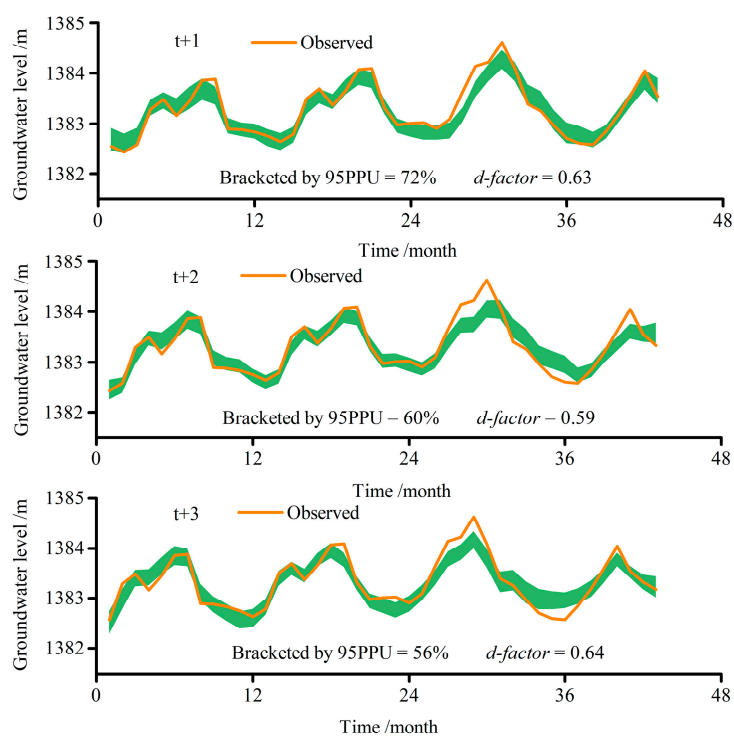
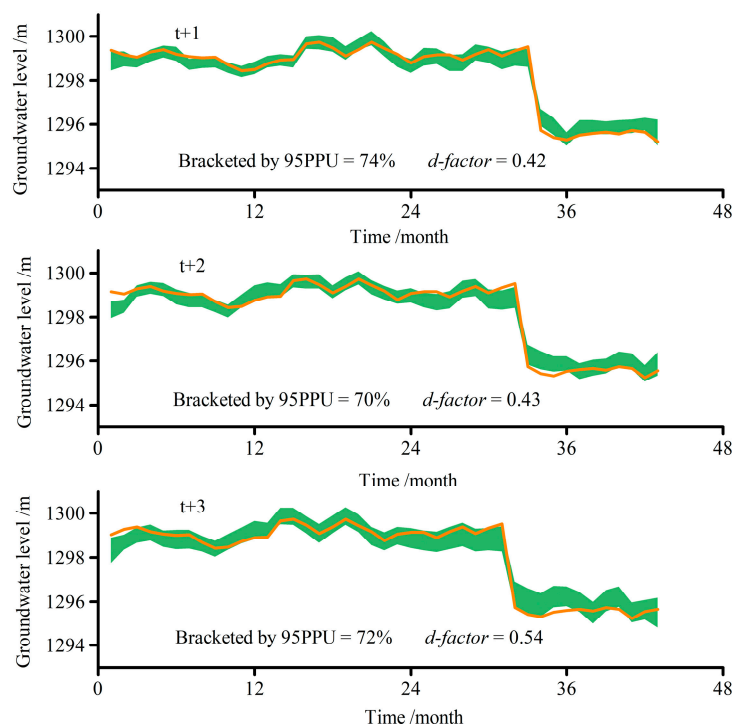


Figure 10. Cont.



(b)



(c)

Figure 10. Observed groundwater level and the 95% confidence intervals forecasted by VELM model with randomly sampled input vectors for 1, 2 and 3-months ahead forest during the testing period at Well I (a), Well II (b) and Well III (c).

Comparing the uncertainty of the two models, it can be discovered that the better performance, with more observed values of groundwater level falling within the 95 PPU and smaller *d-factor* values, can be obtained by the VBELM model. Despite the fact that the uncertainty of the VBELM model was slightly higher than that of the VELM model for 3-months ahead forecast of well II, in most cases the VBELM presented a lower uncertainty relative to the VELM model. In a word, the uncertainty analysis results of the proposed VBELM model were satisfactory with more observed data falling inside the confidence interval, and lower and acceptable *d-factor* values. This further illustrates that the proposed VBELM model can be regarded as a reliable technique to simulate monthly groundwater for all the lead times explored in this study.

6. Conclusions

Accurate and reliable groundwater level prediction can provide government with valuable information about water resources planning and management. In this study, a hybrid VBELM model coupled with the VMD data decomposition method, the Boruta feature selection technique and the ELM model for 1-, 2-, and 3-months ahead groundwater level forecasting at well I, well II and well III in an arid region, northwest China was developed. The performance of the VBELM model was compared with that of the VELM and the ELM models in light of R, RMSE, MAE and NS. Rather than the frequently used deterministic analysis, uncertainty analysis was carried out to evaluate the proposed model by the bootstrap technique.

The results demonstrate that the VBELM model could provide accurate forecast of 1-, 2-, and 3-month ahead groundwater level for all three wells. The best outcome can be derived from the 1-month ahead forecast, while the performance deteriorated gradually with an extension of the prediction time. The comparison results show that the VBELM model performed better than the VELM and ELM models for all the prediction period. In terms of the evaluation indices, the VBELM model obtained higher R and NS values as well as lower RMSE and MAE values compared with those of the other two models. Besides, the VBELM model is capable of tracing the fluctuations in the peak and valley of groundwater level. In terms of the three models, the VBELM performed slightly better than the VELM model, while the single ELM model demonstrated the worst prediction outcome. The result implies that there is a significant improvement in the accuracy of the hybrid models when incorporating data decomposition and inputs selection technologies in monthly groundwater level prediction. In addition, the results of the uncertainty analysis indicate that the performance of the proposed VBELM model was satisfactory, with more observed data falling inside the confidence interval, and lower and acceptable *d-factor* values. Though the uncertainty of the VBELM model decreased along with forecast time increases, the results remained acceptable. In summary, the VBELM model is able to provide an appropriate way for groundwater level forecast and can be regarded as an alternative tool to explore a complex groundwater system in an arid environment.

Although the VBELM method was found to be a promising method for groundwater level modeling, improvement is still needed. First, the performance of the developed models for the three observation wells varies substantially. This is mainly caused by the difference in the internal mechanism of the groundwater systems, which was neglected in this study. Second, the performance of the hybrid VBELM model in groundwater level forecasting was explored without considering the possibility of any other hybrid models. Third, only the uncertainty caused by input data in forecasting results was discussed, while uncertainty resulting from other sources were barely involved. Fourth, extreme climate change has an important impact on groundwater level, such as extreme precipitation and flood events, and groundwater level change due to these factors have not been discussed. To get more reliable results in groundwater level predictions, further studies should concentrate on exploring more hybrid models, introducing a physical mechanism into the predicting models, investigating uncertainty in the model structure

and parameters, and combining extreme climate change to further clarify the reason for groundwater level fluctuation.

Author Contributions: Conceptualization, M.W. and Q.F. Methodology: M.W., Q.F. and X.W. Writing—Original Draft Preparation, M.W. Writing—Review & Editing, Z.Y., L.Y. and D.S. Supervision, X.W. Funding Acquisition, X.W. and Q.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program, China (2017YFC0404305), the Key Research Program of Frontier Sciences, CAS (QYZDJ-SSW-DQC031), Natural Science Foundation of Gansu province, China (18JR2RA026, 18JR3RA393), CAS “Light of West China” Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: 3rd Party Data.

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

Abbreviations

VMD	Variational mode decomposition
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
ELM	Extreme learning machines
VELM	Variational mode decomposition with extreme learning machine
VBELM	Variational mode decomposition with Boruta and extreme learning machine
ANN	Artificial Neural Network
SVM	Support Vector Machine
WT	Wavelet Transform
IMF	Intrinsic mode function
R	Coefficient of correlation
RMSE	Root mean square error
MAE	Mean absolute error
NS	Nash-Sutcliffe efficiency coefficient
Max	Maximum
Min	Minimum
Std	Standard deviation
SK	Skewness

References

1. Cui, Y.; Shao, J. The Role of Ground Water in Arid/Semiarid Ecosystems, Northwest China. *Ground Water* **2005**, *43*, 471–477. [[CrossRef](#)] [[PubMed](#)]
2. Yadav, B.; Gupta, P.K.; Patidar, N.; Himanshu, S.K. Ensemble modelling framework for groundwater level prediction in urban areas of India. *Sci. Total. Environ.* **2020**, *712*, 135539. [[CrossRef](#)] [[PubMed](#)]
3. Mohanty, S.; Jha, M.K.; Kumar, A.; Sudheer, K.P. Artificial Neural Network Modeling for Groundwater Level Forecasting in a River Island of Eastern India. *Water Resour. Manag.* **2010**, *24*, 1845–1865. [[CrossRef](#)]
4. Wen, X.; Feng, Q.; Deo, R.C.; Wu, M.; Si, J. Wavelet analysis-artificial neural network conjunction models for multi-scale monthly groundwater level predicting in an arid inland river basin, northwestern China. *Hydrol. Res.* **2016**, *48*, 1710–1729. [[CrossRef](#)]
5. Wang, W.-C.; Xu, D.-M.; Chau, K.-W.; Chen, S. Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *J. Hydroinformatics* **2013**, *15*, 1377–1390. [[CrossRef](#)]
6. Wen, X.; Feng, Q.; Yu, H.; Wu, J.; Si, J.; Chang, Z.; Xi, H. Wavelet and adaptive neuro-fuzzy inference system conjunction model for groundwater level predicting in a coastal aquifer. *Neural Comput. Appl.* **2015**, *26*, 1203–1215. [[CrossRef](#)]
7. Taormina, R.; Chau, K.-W.; Sethi, R. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* **2012**, *25*, 1670–1676. [[CrossRef](#)]
8. Shiri, J.; Kisi, O.; Yoon, H.; Lee, K.-K.; Nazemi, A.H. Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. *Comput. Geosci.* **2013**, *56*, 32–44. [[CrossRef](#)]
9. Fallah-Mehdipour, E.; Haddad, O.B.; Marini, M.A. Prediction and simulation of monthly groundwater levels by genetic programming. *J. Hydro-Environ. Res.* **2013**, *7*, 253–260. [[CrossRef](#)]

10. Patil, A.P.; Deka, P.C. An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs. *Comput. Electron. Agric.* **2016**, *121*, 385–392. [\[CrossRef\]](#)
11. Chen, X.; Li, F.-W.; Feng, P. A new hybrid model for nonlinear and non-stationary runoff prediction at annual and monthly time scales. *HydroResearch* **2018**, *20*, 77–92. [\[CrossRef\]](#)
12. Zhou, Y.; Guo, S.; Chang, F.J. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *J. Hydrol.* **2019**, *570*, 343–355. [\[CrossRef\]](#)
13. Chen, K.; Chen, H.; Zhou, C.; Huang, Y.; Qi, X.; Shen, R.; Liu, F.; Zuo, M.; Zou, X.; Wang, J.; et al. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* **2020**, *171*, 115454. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Diez-Sierra, J.; Del Jesus, M. Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. *J. Hydrol.* **2020**, *586*, 124789. [\[CrossRef\]](#)
15. Tikhamarine, Y.; Souag-Gamane, D.; Ahmed, A.N.; Kisi, O.; El-Shafie, A. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. *J. Hydrol.* **2020**, *582*, 124435. [\[CrossRef\]](#)
16. Rajaei, T.; Ebrahimi, H.; Nourani, V. A review of the artificial intelligence methods in groundwater level modeling. *J. Hydrol.* **2019**, *572*, 336–351. [\[CrossRef\]](#)
17. Xie, T.; Zhang, G.; Hou, J.; Xie, J.; Lv, M.; Liu, F. Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China. *J. Hydrol.* **2019**, *577*, 123915. [\[CrossRef\]](#)
18. Adamowski, J.; Chan, H.F. A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **2011**, *407*, 28–40. [\[CrossRef\]](#)
19. 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.* **2017**, *31*, 1211–1240. [\[CrossRef\]](#)
20. Huang, S.; Chang, J.; Huang, Q.; Chen, Y. Monthly streamflow prediction using modified EMD-based support vector machine. *J. Hydrol.* **2014**, *511*, 764–775. [\[CrossRef\]](#)
21. Wang, W.-C.; Ahmadi, M.H.; Qiu, L.; Chen, Y. Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Environ. Res.* **2015**, *139*, 46–54. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Zhang, X.; Peng, Y.; Zhang, C.; Wang, B. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *J. Hydrol.* **2015**, *530*, 137–152. [\[CrossRef\]](#)
23. Barzegar, R.; Fijani, E.; Moghaddam, A.A.; Tziritis, E.P. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. *Sci. Total. Environ.* **2017**, *20*–31. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Zhou, T.; Wang, F.; Yang, Z. Comparative Analysis of ANN and SVM Models Combined with Wavelet Preprocess for Groundwater Depth Prediction. *Water* **2017**, *9*, 781. [\[CrossRef\]](#)
25. Wang, Z.-Y.; Qiu, J.; Li, F. Hybrid Models Combining EMD/EEMD and ARIMA for Long-Term Streamflow Forecasting. *Water* **2018**, *10*, 853. [\[CrossRef\]](#)
26. Niu, W.-J.; Feng, Z.-K.; Zeng, M.; Feng, B.-F.; Min, Y.-W.; Cheng, C.-T.; Zhou, J.-Z. Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. *Appl. Soft Comput.* **2019**, *82*, 105589. [\[CrossRef\]](#)
27. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory* **1990**, *36*, 961–1005. [\[CrossRef\]](#)
28. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [\[CrossRef\]](#)
29. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. Kisi, O.; Latifoğlu, L.; Latifoğlu, F. Investigation of Empirical Mode Decomposition in Forecasting of Hydrological Time Series. *Water Resour. Manag.* **2014**, *28*, 4045–4057. [\[CrossRef\]](#)
32. Feng, Q.; Wen, X.; Li, J. Wavelet Analysis-Support Vector Machine Coupled Models for Monthly Rainfall Forecasting in Arid Regions. *Water Resour. Manag.* **2015**, *29*, 1049–1065. [\[CrossRef\]](#)
33. Rezaie-Balf, M.; Naganna, S.R.; Ghaemi, A.; Deka, P.C. Wavelet coupled MARS and M5 Model Tree approaches for groundwater level forecasting. *J. Hydrol.* **2017**, *553*, 356–373. [\[CrossRef\]](#)
34. Hadi, S.J.; Tombul, M. Streamflow Forecasting Using Four Wavelet Transformation Combinations Approaches with Data-Driven Models: A Comparative Study. *Water Resour. Manag.* **2018**, *32*, 4661–4679. [\[CrossRef\]](#)
35. Yu, Y.; Zhang, H.; Singh, V.P. Forward Prediction of Runoff Data in Data-Scarce Basins with an Improved Ensemble Empirical Mode Decomposition (EEMD) Model. *Water* **2018**, *10*, 388. [\[CrossRef\]](#)
36. Rezaie-Balf, M.; Kim, S.; Fallah, H.; Alaghmand, S. Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in Iran and South Korea. *J. Hydrol.* **2019**, *572*, 470–485. [\[CrossRef\]](#)
37. 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. [\[CrossRef\]](#)

38. Liu, Y.; Yang, C.; Huang, K.; Gui, W. Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network. *Knowledge-Based Syst.* **2020**, *188*, 105006. [\[CrossRef\]](#)
39. Zuo, G.; Luo, J.; Wang, N.; Lian, Y.; He, X. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. *J. Hydrol.* **2020**, *585*, 124776. [\[CrossRef\]](#)
40. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
41. He, X.; Luo, J.; Zuo, G.; Xie, J. Daily Runoff Forecasting Using a Hybrid Model Based on Variational Mode Decomposition and Deep Neural Networks. *Water Resour. Manag.* **2019**, *33*, 1571–1590. [\[CrossRef\]](#)
42. Sahani, M.; Dash, P.K.; Samal, D. A real-time power quality events recognition using variational mode decomposition and online-sequential extreme learning machine. *Measurement* **2020**, *157*, 107597. [\[CrossRef\]](#)
43. He, F.; Zhou, J.Z.; Feng, Z.-K.; Liu, G.; Yang, Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Appl. Energy* **2019**, *237*, 103–116. [\[CrossRef\]](#)
44. Jayakumar, C.; Sangeetha, J. Kernellized support vector regressive machine based variational mode decomposition for time frequency analysis of Mirnov coil. *Microprocess. Microsyst.* **2020**, *75*, 103036. [\[CrossRef\]](#)
45. Bisoi, R.; Dash, P.K.; Parida, A.K. Hybrid Variational Mode Decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis. *Appl. Soft Comput.* **2019**, *74*, 652–678. [\[CrossRef\]](#)
46. Zhang, Y.; Liu, K.; Qin, L.; An, X. Deterministic and probabilistic interval prediction for short-term wind power generation based on variational mode decomposition and machine learning methods. *Energy Convers. Manag.* **2016**, *112*, 208–219. [\[CrossRef\]](#)
47. Wang, D.; Luo, H.; Grunder, O.; Lin, Y. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction. *Renew. Energy* **2017**, *113*, 1345–1358. [\[CrossRef\]](#)
48. Majumder, I.; Dash, P.; Bisoi, R. Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting. *Energy Convers. Manag.* **2018**, *171*, 787–806. [\[CrossRef\]](#)
49. Seo, Y.; Kim, S.; Singh, V.P. Machine Learning Models Coupled with Variational Mode Decomposition: A New Approach for Modeling Daily Rainfall-Runoff. *Atmosphere* **2018**, *9*, 251. [\[CrossRef\]](#)
50. Srivastav, R.K.; Sudheer, K.; Chaubey, I. A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resour. Res.* **2007**, *43*, 43. [\[CrossRef\]](#)
51. Tiwari, M.K.; Chatterjee, C. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J. Hydrol.* **2010**, *382*, 20–33. [\[CrossRef\]](#)
52. Tiwari, M.K.; Chatterjee, C. A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *J. Hydroinformatics* **2010**, *13*, 500–519. [\[CrossRef\]](#)
53. Gopalan, S.P.; Kawamura, A.; Amaguchi, H.; Takasaki, T.; Azhikodan, G. A bootstrap approach for the parameter uncertainty of an urban-specific rainfall-runoff model. *J. Hydrol.* **2019**, *579*, 124195. [\[CrossRef\]](#)
54. Tiwari, M.K.; Adamowski, J.F. Medium-Term Urban Water Demand Forecasting with Limited Data Using an Ensemble Wavelet-Bootstrap Machine-Learning Approach. *J. Water Resour. Plan. Manag.* **2015**, *141*, 04014053. [\[CrossRef\]](#)
55. Feng, Z.-K.; Niu, W.-J.; Tang, Z.-Y.; Jiang, Z.-Q.; Xu, Y.; Liu, Y.; Zhang, H.-R. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *J. Hydrol.* **2020**, *583*, 124627. [\[CrossRef\]](#)
56. 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. [\[CrossRef\]](#)
57. Sahoo, S.; Russo, T.; Elliott, J.; Foster, I. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S. *Water Resour. Res.* **2017**, *53*, 3878–3895. [\[CrossRef\]](#)
58. Kursa, M.B.; Jankowski, A.; Rudnicki, W.R. Boruta—A system for feature selection. *Fundam. Inform.* **2010**, *101*, 271–285. [\[CrossRef\]](#)
59. Christa, M.; Kempa-Liehrb, A.W.; Feindta, M. Distributed and parallel time series feature extraction for industrial big data applications. *arXiv* **2016**, arXiv:1610.07717.
60. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [\[CrossRef\]](#)
61. Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *Catena* **2019**, *177*, 149–166. [\[CrossRef\]](#)
62. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [\[CrossRef\]](#)
63. Deo, R.C.; Şahin, M. Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. *Atmos. Res.* **2015**, *153*, 512–525. [\[CrossRef\]](#)
64. Zhu, B.; Feng, Y.; Gong, D.; Jiang, S.; Zhao, L.; Cui, N. Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data. *Comput. Electron. Agric.* **2020**, *173*, 105430. [\[CrossRef\]](#)
65. Huang, G.; Zhu, Q.; Siew, C. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. In Proceedings of the 2004 IEEE international joint conference on neural networks, Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 985–990.
66. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [\[CrossRef\]](#)
67. Efron, B.J.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall: London, UK, 1993.

-
68. Twomey, J.M.; Smith, A.E. Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **1998**, *28*, 417–430. [[CrossRef](#)]
 69. Ghorbani, M.A.; Zadeh, H.A.; Isazadeh, M.; Terzi, O. A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environ. Earth Sci.* **2016**, *75*, 1–14. [[CrossRef](#)]
 70. Wang, S.; Liu, H.; Yu, Y.; Zhao, W.; Yang, Q.; Liu, J. Evaluation of groundwater sustainability in the arid Hexi Corridor of Northwestern China, using GRACE, GLDAS and measured groundwater data products. *Sci. Total. Environ.* **2020**, *705*, 135829. [[CrossRef](#)]
 71. Huang, N.; Chen, H.; Cai, G.; Fang, L.; Wang, Y. Mechanical Fault Diagnosis of High Voltage Circuit Breakers Based on Variational Mode Decomposition and Multi-Layer Classifier. *Sensors* **2016**, *16*, 1887. [[CrossRef](#)]
 72. Citakoglu, H.; Cobaner, M.; Haktanir, T.; Kisi, O. Estimation of Monthly Mean Reference Evapotranspiration in Turkey. *Water Resour. Manag.* **2014**, *28*, 99–113. [[CrossRef](#)]
 73. Nayak, P.C.; Rao, Y.R.S.; Sudheer, K.P. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. *Water Resour. Manag.* **2006**, *20*, 77–90. [[CrossRef](#)]