

Article

# A Comparative Study of Groundwater Level Forecasting Using Data-Driven Models Based on Ensemble Empirical Mode Decomposition

Yicheng Gong <sup>1,2,\*</sup>, Zhongjing Wang <sup>1,2,3,\*</sup>, Guoyin Xu <sup>1</sup> and Zixiong Zhang <sup>1</sup>

<sup>1</sup> Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China; xgy13@mails.tsinghua.edu.cn (G.X.); zx-zhang15@mails.tsinghua.edu.cn (Z.Z.)

<sup>2</sup> State Key Laboratory of Hydro-Science and Engineering, Tsinghua University, Beijing 100084, China

<sup>3</sup> State Key Lab of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

\* Correspondence: gongyc@tsinghua.edu.cn (Y.G.); zj.wang@tsinghua.edu.cn (Z.W.); Tel.: +86-010-6278-2021 (Z.W.)

Received: 15 March 2018; Accepted: 29 May 2018; Published: 4 June 2018



**Abstract:** The reliable and accurate prediction of groundwater levels is important to improve water-use efficiency in the development and management of water resources. Three nonlinear time-series intelligence hybrid models were proposed to predict groundwater level fluctuations through a combination of ensemble empirical mode decomposition (EEMD) and data-driven models (i.e., artificial neural networks (ANN), support vector machines (SVM) and adaptive neuro fuzzy inference systems (ANFIS)), respectively. The prediction capability of EEMD-ANN, EEMD-SVM, and EEMD-ANFIS hybrid models was investigated using a monthly groundwater level time series collected from two observation wells near Lake Okeechobee in Florida. The statistical parameters correlation coefficient (R), normalized mean square error (NMSE), root mean square error (RMSE), Nash–Sutcliffe efficiency coefficient (NS), and Akaike information criteria (AIC) were used to assess the performance of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models. The results achieved from the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models were compared with those from the ANN, SVM and ANFIS models. The three hybrid models (i.e., EEMD-ANN, EEMD-SVM, and EEMD-ANFIS) proved to be applicable to forecast the groundwater level fluctuations. The values of the statistical parameters indicated that the EEMD-ANFIS and EEMD-SVM models achieved better prediction results than the EEMD-ANN model. Meanwhile, the three models coupled with EEMD were found have better prediction results than the models that were not. The findings from this study indicate that the proposed nonlinear time-series intelligence hybrid models could improve the prediction capability in forecasting groundwater level fluctuations, and serve as useful and helpful guidelines for the management of sustainable water resources.

**Keywords:** groundwater level; ensemble empirical mode decomposition; artificial neural network; support vector machine; adaptive neuro fuzzy inference system

## 1. Introduction

Groundwater is an increasingly important water resource for irrigation, and domestic and industrial activities in many countries. More reliable and accurate estimation of groundwater levels can help prevent groundwater overexploitation and improve water-use efficiency for water resource management. However, in some regions, the groundwater has been pumped out much faster than it can be replenished, which eventually reduces the groundwater level. In addition, groundwater level time series are highly non-linear and non-stationary in nature, and prediction depends on many complex environmental factors, such as groundwater aquifers, precipitation, etc. Groundwater aquifers

are intrinsically heterogeneous systems that are affected by complex hydrogeological conditions with groundwater-surface water interactions at various temporal-spatial scales [1,2]. Therefore, it is essential to develop more effective models for groundwater level prediction. Many groundwater modeling approaches and data-driven models have been implemented to forecast groundwater levels [2–4]. A groundwater numerical model is firstly developed from a conceptual model, which often includes only the main and fundamental principles but neglects the complexities of groundwater systems [1]. It is hard to establish groundwater flow equations and set the values of hydrogeological parameters for conceptual models. Also, it is difficult to obtain long time series data for groundwater numerical modeling; there are still challenges and uncertainties in modelling processes [1,5].

Recently, data-driven models such as artificial neural networks (ANN), support vector machines (SVM), adaptive neuro-fuzzy inference systems (ANFIS) and genetic programming (GP), and time series methods such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) have been proved efficient in forecasting hydrologic time series (e.g., groundwater level, water demand and inflow) [5–18]. Yoon et al. [5] developed ANN and SVM models to forecast groundwater level fluctuations in a coastal aquifer. The performance of the SVM was similar to or even better than that of the ANN model, according to the results from the validation. Shirmohammadi et al. [10] applied time series, system identification and ANFIS models to predict groundwater levels. The obtained results showed that ANFIS is superior to time series and system identification. Moosavi [11] proposed a model to forecast groundwater levels for different prediction periods. The achieved results showed that predictive ability of the wavelet–ANFIS model was more accurate than those of the ANN, ANFIS and wavelet–ANN models. Shiri and Kişi [12] investigated the ability of adaptive neuro-fuzzy inference systems (ANFIS) and genetic programming (GP) to predict water table depth fluctuations. The results showed that the ANFIS and GP models can be applied successfully in groundwater depth prediction. Fallah-Mehdipour et al. [6] investigated the capability of ANFIS and GP to forecast and simulate groundwater levels in the Karaj plain of Iran. These results indicated that GP is an effective method for predicting groundwater levels. Valipour et al. [13] compared autoregressive moving average (ARMA) with autoregressive integrated moving average (ARIMA) models in forecasting the inflow of the Dez dam reservoir. The ARIMA model was proved have better predictive ability compared to the ARMA model. Xu et al. [14] built the instance-based weighting and support vector regression data-driven models to reduce the predictive error of groundwater models. The results of two real-world case studies showed that data-driven models can be applied effectively to reduce the root mean square error of the groundwater models. Asefa et al. [15] used support vector machines (SVMs) to monitor network design. The obtained results showed that SVMs can select the best configurations of well networks by reproducing the behavior of Monte Carlo flow.

Ensemble empirical mode decomposition (EEMD) is a scale-adaptive method proposed by Wu and Huang [19], which improved on empirical mode decomposition (EMD) to avoid the drawback of mode mixing [20]. EEMD has a high ability to decompose the original signal into intrinsic mode function (IMF) components and residual components for nonstationary and nonlinear signal sequences. In the hydrological and environmental research field, EEMD has been successfully applied to predict nonlinear problems such as runoff [21–23], wind speed [24], wave height [25], particulate matter 2.5 (PM<sub>2.5</sub>) [26], streamflow [27,28], vegetation dynamics [29], etc. Wang et al. [21] proposed an EEMD-ARIMA model for the forecasting of annual runoff time series. Quantitative standard statistical performance values proved that EEMD-ARIMA is a superior model to ARIMA for the forecasting of annual runoff. Liu et al. [24] applied the fast ensemble empirical mode decomposition–multilayer perceptron (FEEMD-MLP), Fast Ensemble Empirical Mode Decomposition-Adaptive neuro fuzzy inference system (FEEMD-ANFIS), wavelet packet–multilayer perceptron (WP-MLP) and WP-ANFIS models to forecast wind speed. The results suggested that all of the proposed hybrid algorithms are suitable for wind speed prediction. Duan et al. [25] developed the improved empirical model decomposition–support vector regression (EMD-SVR) model for the short-term prediction of wave height and found that

the EMD-SVR model performs better than the wavelet-decomposition-based SVR (WD-SVR) model. Ausati and Amanollahi [26] used ensemble empirical mode decomposition–general regression neural network (EEMD-GRNN), principal component regression (PCR), adaptive neuro-fuzzy inference system (ANFIS) and multiple linear regression (MLR) models to predict concentrations of particulate matter 2.5 (PM<sub>2.5</sub>) in the city of Sanandaj. It was concluded that the EEMD-GRNN hybrid model had higher accuracy than the linear model in forecasting the particulate matter 2.5 (PM<sub>2.5</sub>) concentration. Karthikeyan and Nagesh [28] tested the predictability of monthly streamflow time series using wavelet- and EMD-based ARMA models. The result showed that the wavelet-based ARMA method outperformed the EMD-based ARMA method. Hawinkel et al. [29] used EEMD to decompose the time series of the normalized difference vegetation index (NDVI) and extracted climate-driven interannual vegetation dynamics. The result indicated that the EEMD was a feasible method for the assessment of climate-driven interannual vegetation dynamics.

ANN, SVN and ANFIS data-driven models and EEMD were applied in formal hydrology studies. All of these models were used alone or in combination to predict nonlinear and nonstationary time series. In this study, the ANN, SVN and ANFIS nonlinear time-series model coupled with EEMD were applied to groundwater level prediction, for validity and accuracy testing. The main advantage of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS hybrid models is that exogenous factors affecting the groundwater level do not need to be considered in the prediction. In order to compare the forecasting performance of the three hybrid models, single ANN, SVM and ANFIS models were applied to predict the groundwater level by considering exogenous factors, such as precipitation, temperature and surface water level, etc. The prediction results obtained from models with and without coupling were compared based on standard statistical analysis.

This paper is organized as follows: Section 2 briefly describes the basic theory of EEMD, ANN, SVM and ANFIS, and presents the framework of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS hybrid models; Section 3 introduced the study area, available data and performance criteria. Section 4 explained the groundwater level decomposition by the EEMD for modeling and application; Section 5 analyzed the predicting results of the proposed EEMD-ANN, EEMD-SVM and EEMD-ANFIS hybrid models; Section 6 concluded this study.

## 2. Methods

### 2.1. Ensemble Empirical Mode Decomposition

Ensemble empirical mode decomposition (EEMD) is a new empirical mode decomposition (EMD)-based algorithm, which is a fully data-adaptive method to analyze nonlinear and nonstationary signals [19,20]. EMD can decompose the original signal into the number of intrinsic mode functions (IMFs) when the signal meets two conditions: (1) the functions are symmetric and the mean value of the upper and lower envelopes should be zero; (2) the number of extrema and the number of zero crossings must be equal or differ at most by one. However, EMD still has some drawbacks, such as mode mixing. Ensemble empirical mode decomposition (EEMD) was proposed to combine the white-noise-assisted system based on EMD to solve the problem of mode mixing [19,30]. If  $x(t)$  is the signal or time series to be decomposed, the EMD algorithm can be briefly summarized as follows:

- Step 1. Identify all local maxima and minima points of the time series  $x(t)$ ;
- Step 2. Interpolate between all local maxima and minima points of  $x(t)$  to form the upper envelope  $e_{\max}(t)$  and lower envelope  $e_{\min}(t)$ ;
- Step 3. Compute the mean envelope  $m(t)$  between the upper envelope  $e_{\max}(t)$  and the lower envelope  $e_{\min}(t)$ ;

$$m(t) = (e_{\max}(t) + e_{\min}(t))/2 \quad (1)$$

Step 4. Calculate the IMF candidate;

$$h(t+1) = x(t) - m(t) \quad (2)$$

Step 5. Determine whether or not  $h(t+1)$  satisfies the two conditions of IMF. Is  $h(t+1)$  an IMF?

If yes, save  $h(t+1)$ , calculate the residue  $r(t+1) = x(t) - \sum_{i=1}^{t+1} h(i)$ ,  $t = t + 1$ , and treat  $r(t)$  as input data in step 2. If no, treat  $h(t+1)$  and as input data in step 2 until it satisfies the two conditions of IMF.

Step 6. Continue until the final residue meets some predefined stopping criteria.

At the end of the shifting process, the final residual term  $r_n(t)$  has less than two local extrema, and the original signal or time series can be decomposed into a set of IMFs and a residue as follows:

$$x(t) = \sum_{i=1}^n h_i(t) + r(t) \quad (3)$$

where  $n$  is the number of IMFs,  $r_n(t)$  is the final residue and  $h_i(t)$  as IMFs are nearly orthogonal to each other.

Ensemble empirical mode decomposition (EEMD) was proposed by adding a finite amount of Gaussian white noise to series with EMD mixture, and the mode mixing is mostly eliminated. Based on the EMD method, the EEMD can be briefly summarized as follows [19,31]:

Step 1. Initialize the ensemble number and the amplitude of the added white noise.

Step 2. Add random white noise to produce the noise-added data.

Step 3. Identify the local maxima and minima and obtain the upper and lower envelopes.

Step 4. Compute the mean of the upper and lower envelopes.

Step 5. Decompose the data with added random white noise into IMFs.

Step 6. Repeat step 3 to step 5 until the stopping criteria. After the shift processing, the IMFs and the residue are obtained.

## 2.2. Artificial Neural Network

An artificial neural network (ANN) is a black box tool that has certain performance characteristics that resemble the biological neural networks in the human brain [32]. Feed-forward neural networks (FFNN) are the most commonly used artificial neural networks, and have been broadly employed for modeling and forecasting in hydrogeology [33,34]. The architecture of the FFNN model consists of an input layer, hidden layer, output layer and artificial neurons (Figure 1). In a feed forward network, the neurons in each layer are connected to those in the next layer. Therefore, the output of a node in a layer only depends on the input it received from previous layer, determined weight, and type of transform function. The present research used the three-layer FFNN model, which was trained with a Bayesian regularization (BR) algorithm. The tan-sigmoid transfer function was selected by previous studies [35]. The mathematical expression can be expressed as follows:

$$y_j = f\left(\sum_{i=1}^N w_{ji}x_i + b_j\right) \quad (4)$$

where  $x_i$  is the input vector,  $y_j$  is the output,  $b_i$  is the bias,  $w_{ji}$  is a weight connecting  $x_i$  and  $y_j$ ,  $N$  is the number of nodes, and  $f$  is the activation function.

### 2.3. Support Vector Machine

A support vector machine (SVM) is a novel machine learning technique based on statistical learning theory [36,37]. In a regression SVM model, a regression hyperplane with a  $\epsilon$ -insensitivity loss function is a convex dual optimization problem. The solution can be obtained from the optimization algorithm. The deterministic function of the SVM can be expressed as follows:

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \tag{5}$$

where  $\mathbf{w}_i$  is a weight vector, and  $b$  is a bias.  $\mathbf{x}$  is mapped to a high-dimensional feature space by nonlinear transfer function  $\phi$ . In this study, the sequential minimal optimization (SMO) algorithm was used to solve the dual optimization problem of SVM [38,39]. The SMO algorithm can be directly acquired from an analytical solution without invoking a quadratic optimizer. The programming codes of the library for support vector machines (LIBSVM) was applied to predict a non-linear time series in the validation and calibration [40]. Figure 2 shows the schematic diagram of the SVM model.

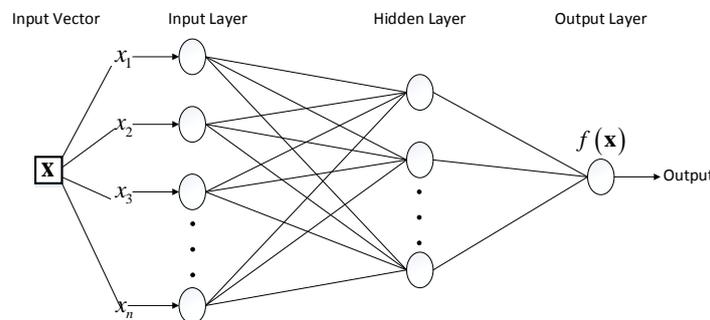


Figure 1. Schematic diagram of an ANN.

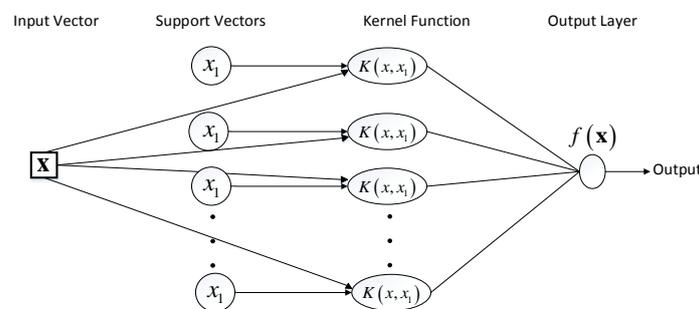


Figure 2. Schematic diagram of a SVM.

### 2.4. Adaptive Neuro Fuzzy Inference System

An adaptive neuro fuzzy inference system (ANFIS) as a hybrid algorithm is an adaptive neural network based on a fuzzy inference system [41]. ANFIS is capable of approximating any real continuous function in a compact set to any degree of accuracy. This study used the first order Sugeno–fuzzy model of ANFIS [41,42]. To simplify, it was assumed that the fuzzy inference system has two inputs  $x$  and  $y$  and one output  $f$ . For the first-order Sugeno inference system, the typical rules can be expressed as:

$$\text{Rule 1: If } x \text{ is } A_1 \text{ and } y \text{ is } B_1; \text{ then } f_1 = p_1x + q_1y + r_1 \tag{6}$$

$$\text{Rule 2: If } x \text{ is } A_2 \text{ and } y \text{ is } B_2; \text{ then } f_2 = p_2x + q_2y + r_2 \tag{7}$$

where  $x$  and  $y$  are the crisp inputs to the node  $i$ ,  $A_i$  and  $B_i$  are the linguistic labels (low, medium, high, etc.,) characterized by the convenient membership functions.  $p_i, q_i,$  and  $r_i (i = 1, 2)$  are the parameters

in the then part (consequent part) of the first-order Sugeno fuzzy model. The Sugeno–fuzzy structure of ANFIS comprised a five-layer network (Figure 3). More comprehensive information about ANFIS can be found in the literature [11,43–45].

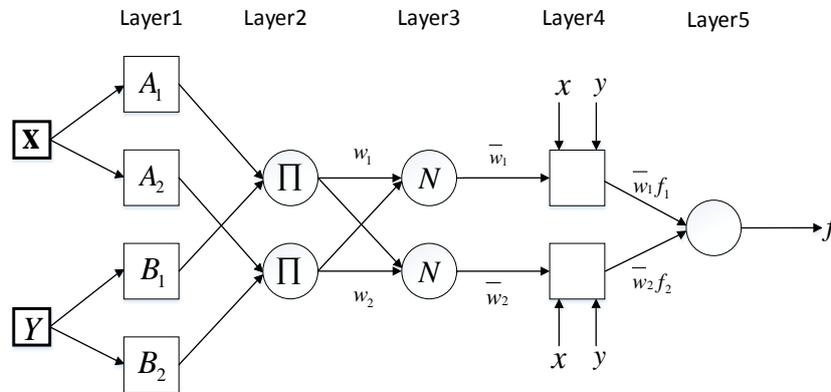


Figure 3. Schematic diagram of ANFIS.

2.5. The Hybrid EEMD-ANN, EEMD-SVM and EEMD-ANFIS Forecasting Models

The goal of this study was to improve the forecasting accuracy of groundwater levels by coupling the EEMD and data-driven models. A schematic diagram of the hybrid EEMD-ANN, EEMD-SVM and EEMD-ANFIS models is illustrated in Figure 4. As shown in Figure 4, the process of running the three hybrid models consists of three main steps:

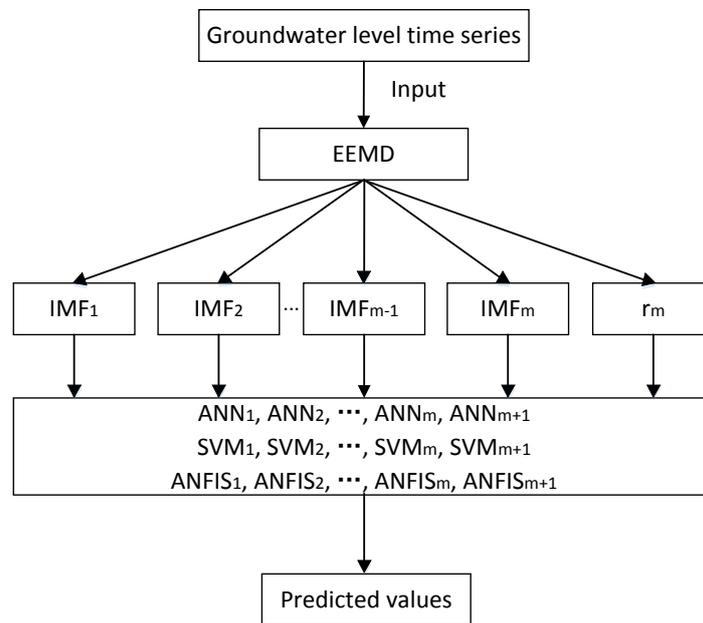


Figure 4. Schematic diagram of EEMD-ANN, EEMD-SVM and EEMD-ANFIS.

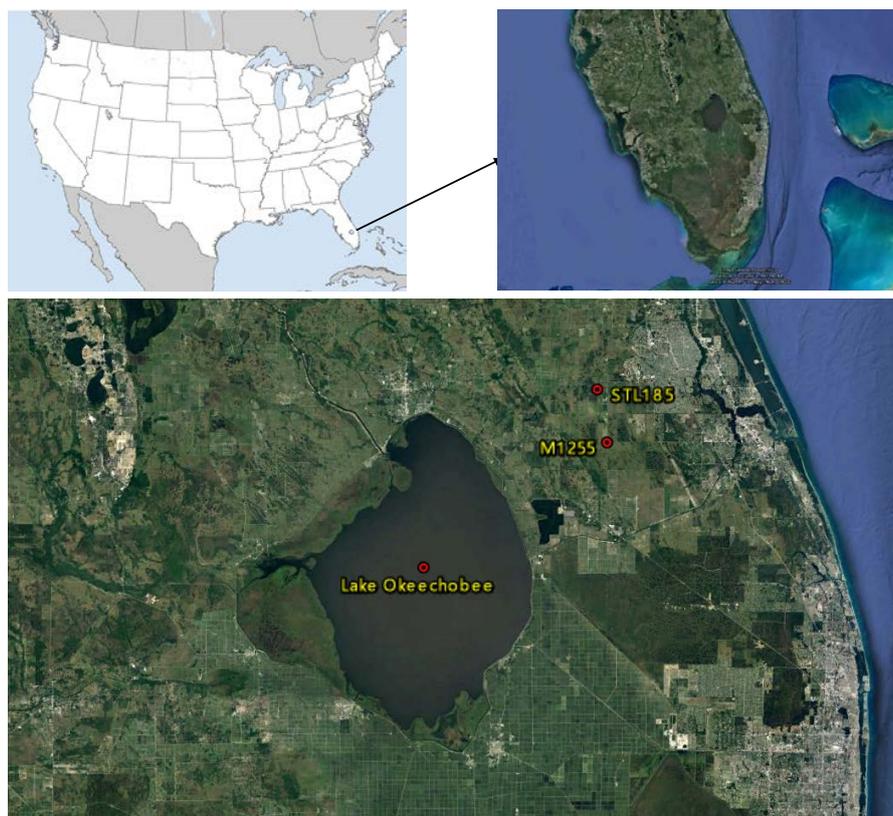
- (1) Firstly, use the EEMD technique to decompose the original groundwater level fluctuation time series  $x(t) (t = 1, 2, \dots, n; i = 1, 2, \dots, m)$  into an IMF component  $c_i(t)$  and one residual component  $r_m(t)$ .
- (2) Secondly, the three data driven models ANN, SVM and ANFIS are developed to make the corresponding prediction using extracted IMF component and the residual component, respectively.

- (3) Finally, all of the predicting results from the ANN, SVM and ANFIS models are combined to obtain the new output values, which are the final forecasting result for the groundwater level prediction.

### 3. Study Area and Available Data

#### 3.1. Study Site and Data Preprocessing

The study area was located in Florida, United States. The annual mean temperature in the study area was 23.36 °C and the average annual mean precipitation was 980.61 mm in the past 16 years. All the observed data (including the monthly mean precipitation, monthly mean temperature, monthly maximum temperature, monthly minimum temperature, monthly mean lake level and monthly mean groundwater level) were collected to forecast the future groundwater fluctuations at well M1255 and STL185. For the obtained data from the past 16 years (from 1997 to 2012), data from the first 14 years was used for training and the last two years of data was used for validation. Meanwhile, the IMF components and one residual component of the groundwater level time-series were used to predict future groundwater fluctuations based on the EEMD. The results of the groundwater level prediction using the EEMD were analyzed and compared with the results obtained not using the EEMD. The locations of the observed wells in the study area are shown in Figure 5. Figure 6a–d illustrates the monthly mean precipitation, monthly mean temperature, monthly mean lake level, and monthly maximum temperature. Figure 7a,b illustrates the monthly groundwater levels at well M1255 and STL185.



**Figure 5.** Location map of the United States and observation wells (Source: Imagery © Data SIO, NOAA, U.S. Navy, NGA, GEBCO, Image Landsat/Copernicus).

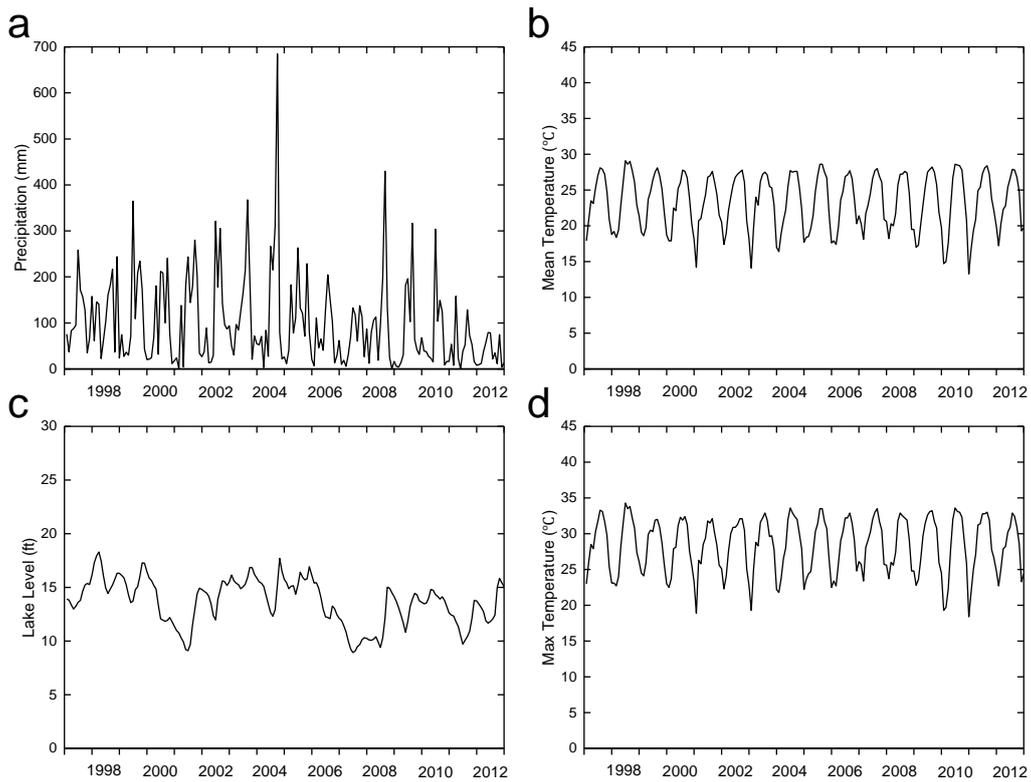


Figure 6. Plots of (a) precipitation (b) mean temperature (c) lake level (d) max. temperature.

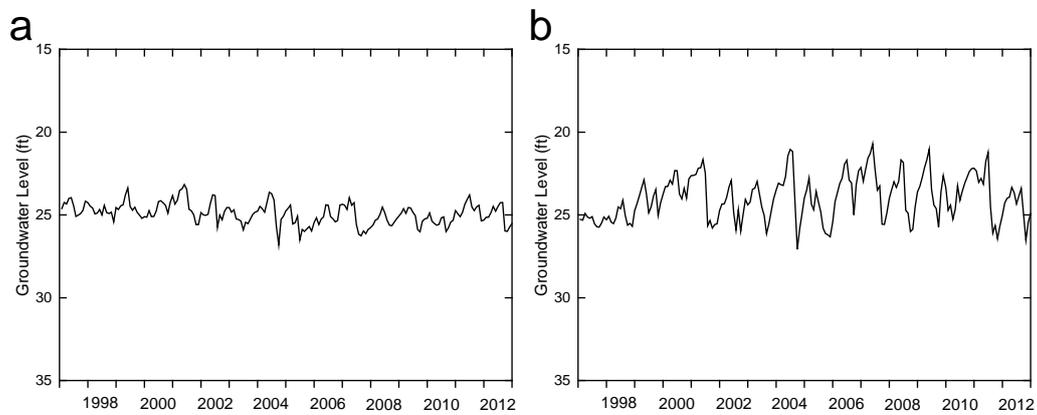


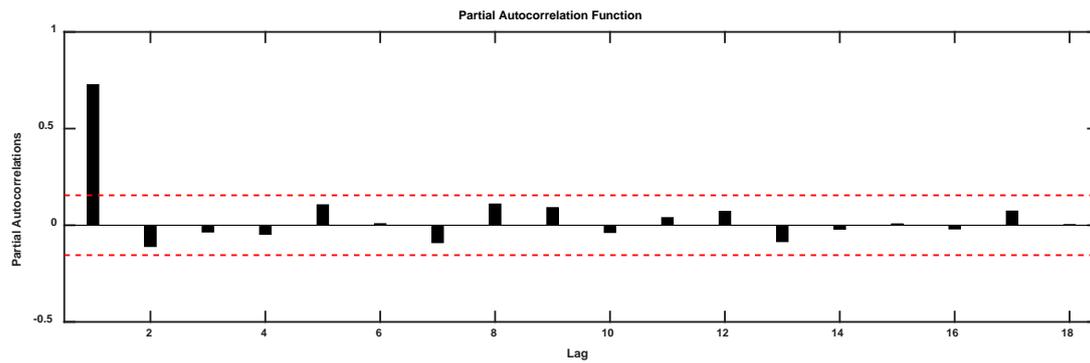
Figure 7. Plots of (a) groundwater level of M1255 (b) groundwater level of STL185.

To eliminate the dimensional differences, the time-series data was normalized using the following equation before the training stage.

$$x_i = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{8}$$

where  $x_i$  is the normalized data,  $x$  is the time-series data,  $x_{\min}$  is the minimum value and  $x_{\max}$  is the maximum value.

The partial autocorrelation function (PACF), as one of the statistical methods, was frequently used to select suitable models [10]. The PACF of well site M1255 from lag-0 to lag-18 is shown in Figure 8. Figure 8 suggests a significant correlation up to the lag-1 month for this time series within the confidence interval. The partial autocorrelation coefficients indicated that there was a one-month time lag for the monthly groundwater level time series as the input vector of the ANN, SVM and ANFIS models.



**Figure 8.** The partial autocorrelation function (PACF) of the groundwater level series at site M1255.

### 3.2. Performance Criteria

Five statistical performance evaluation parameters—correlation coefficient (R), normalized mean square error (NMSE), root mean squared error (RMSE), Nash–Sutcliffe efficiency coefficient (NS) and Akaike information criteria (AIC)—were used to estimate the performance of the models proposed in this study.

Correlation coefficient (R):

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \quad (9)$$

Normalized mean square error (NMSE):

$$NMSE = \frac{n-1}{n} \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (10)$$

Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (11)$$

Nash–Sutcliffe efficiency coefficient (NS):

$$NS = 1 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (12)$$

Akaike information criteria (AIC):

$$AIC(k) = N \cdot \ln \left( \frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2 \right) + 2k \quad (13)$$

where  $O_i$  is the observed groundwater level value,  $P_i$  is the predicted groundwater level value,  $\bar{O}$  is the average of the observed groundwater level value,  $\bar{P}$  is the average of the predicted groundwater level value,  $N$  is the number of input samples, and  $k$  is the number of free parameters used in those models. Suppose  $k$  is 0 in this paper. The Akaike information criteria (AIC) was employed to select the best time series mode [46]. The best fit between the predicted value and observed value would have  $R = 1$ ,  $NMSE = 0$ ,  $RMSE = 0$ ,  $NS = 1$ ,  $AIC = -\infty$ , respectively.

#### 4. Decomposing, Modeling and Application

The time series of the original groundwater levels can be decomposed into several independent IMFs and one residue using the EEMD technique. The results are shown in Figures 9 and 10. It is noticeable that the two time series of the groundwater levels were resolved into six independent IMF and one residue component. Six independent IMFs were successively displayed from the highest to the lowest frequency, respectively. The extracted IMF components and the residual component as input variables were used to forecast the groundwater level. EEMD, as a new noise-assisted data analysis method, overcame the scale separation problem, avoided the mode mixing problem and conserved the physical uniqueness of the decomposition [19]. Therefore, the decomposition could be helpful to transform non-linear and non-stationary time series to independent IMF components from high to low frequency and could be useful to improve the forecast accuracy.

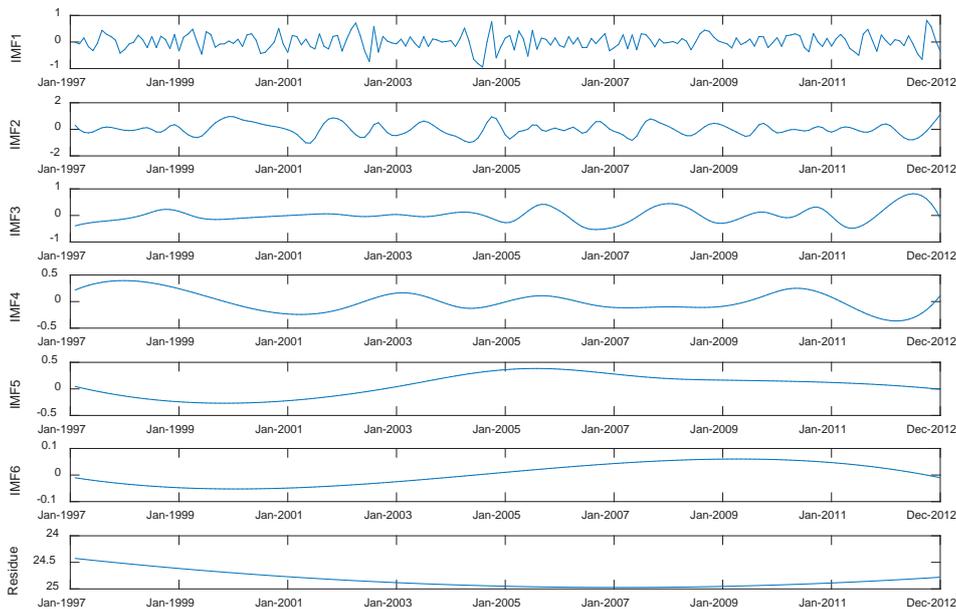


Figure 9. Decomposition of groundwater level time-series at the M1255 observation well.

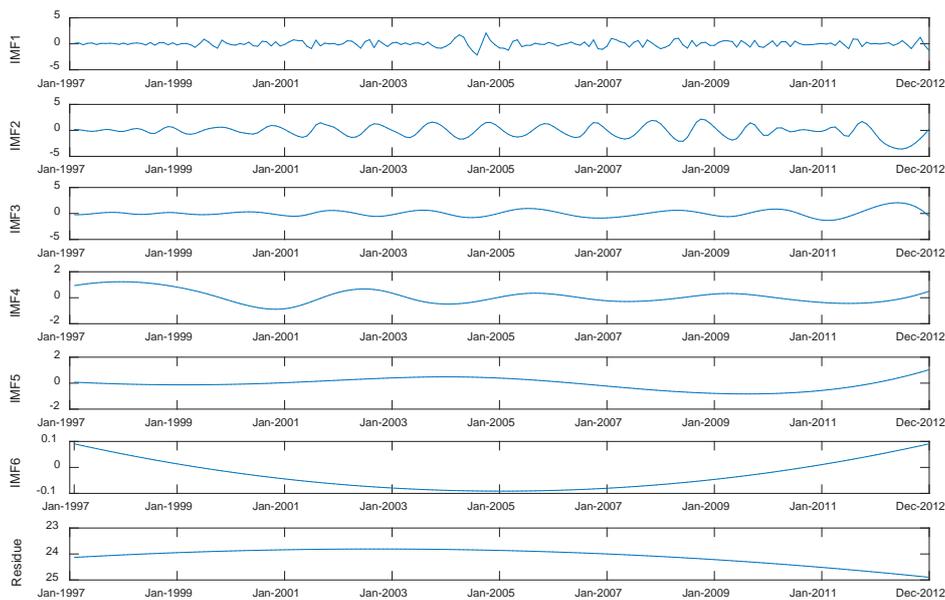


Figure 10. Decomposition of groundwater level time-series at the STL185 observation well.

The selection of input variables depended on the availability of the inputs and historically observed records. Several combinations of the exogenous factors were applied in the ANN, SVM and ANFIS models for groundwater level prediction. The input variables in the ANN model were the monthly time-series data of the precipitation, temperature (maximum, mean and minimum), and groundwater level. ANN+L representing the lake level was added to input variables in the ANN model. EEMD-ANN represents that the input variables were IMF components and one residual component. The input variables of the SVM and ANFIS models were similar to that of the ANN models (see Tables 1–3).

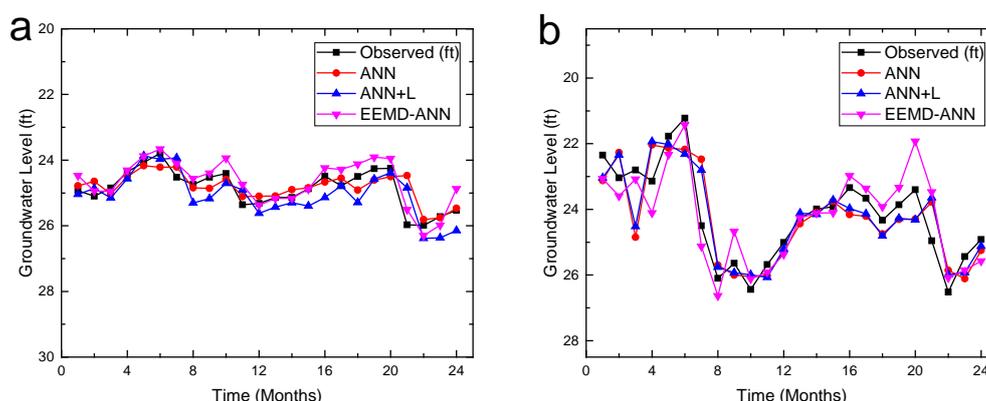
### 5. Results and Discussion

#### 5.1. EEMD-ANN and ANN Models

For the investigation of the effects of the input structure and input data on the performance of the ANN model, three input structures were considered, as shown in Table 1. The results showed that the input variables of the EEMD-ANN at two sites has the minimum RMSE and AIC values in the validation. Meanwhile, the input variables of EEMD-ANN had the maximum R value in the validation. At site M1255, the minimum RMSE and AIC of EEMD-ANN were 0.329 and  $-53.350$ , respectively; at site STL185, the minimum RMSE and AIC of EEMD-ANN were 0.646 and  $-20.979$ , respectively. The maximum R of EEMD-ANN in the validation was 0.926 at site M1255 and 0.891 at site STL185. Table 1 suggested that R values of the four input variables of ANN+L at two sites were better than that of three input variables of ANN in the validation stage. The R of ANN+L at site M1255 was 0.754. The R value for ANN+L applied at site STL185 was 0.839. Figure 11 shows the observed and predicted groundwater levels using ANN, ANN+L and EEMD-ANN in the validation stage.

**Table 1.** Result of modeling from the ANN model at site M1255 and STL185.

		Training					Validation				
		R	NMSE	RMSE	NS	AIC	R	NMSE	RMSE	NS	AIC
M1255	EEMD-ANN	0.932	0.131	0.234	0.869	$-488.040$	0.926	0.317	0.329	0.669	$-53.350$
	ANN	0.808	0.345	0.380	0.653	$-325.009$	0.737	0.442	0.389	0.538	$-45.377$
	ANN+L	0.816	0.333	0.374	0.665	$-330.668$	0.754	0.675	0.480	0.296	$-35.246$
STL185	EEMD-ANN	0.977	0.044	0.281	0.955	$-426.784$	0.891	0.208	0.646	0.783	$-20.979$
	ANN	0.869	0.244	0.659	0.754	$-140.343$	0.813	0.353	0.842	0.632	$-8.239$
	ANN+L	0.872	0.239	0.651	0.760	$-144.026$	0.839	0.298	0.774	0.689	$-12.323$



**Figure 11.** Observed and predicted groundwater level using ANN, ANN+L and EEMD-ANN (a) M1255 (b) STL185.

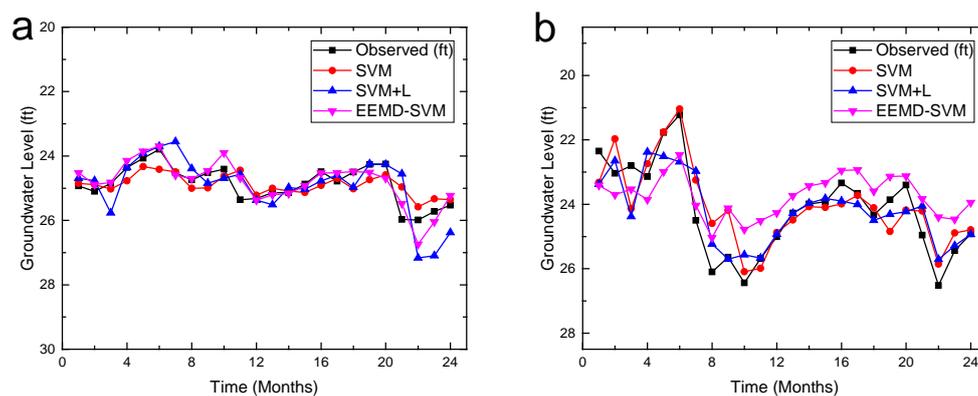
#### 5.2. EEMD-SVM and SVM Models

The same input structures were introduced to the SVM model, as shown in Table 2. The results showed that the input variables of EEMD-SVM at site M1255 and the three input variables of SVM

at site STL185 had the minimum RMSE and AIC values in the validation. The minimum RMSE and AIC of EEMD-SVM at site M1255 were 0.315 and 221255.373, respectively. The minimum RMSE and AIC of SVM at site STL185 were 0.755 and  $-13.512$ , respectively. Table 2 suggests that the maximum  $R$  value of EEMD-SVM was 0.885 at site M1255 and 0.860 at site STL185. The  $R$  values of the four input variables of SVM+L at the two sites were higher than or equal to those of the three input variables of SVM in the validation stage. The  $R$  values of SVM+L at site M1255 were 0.758 and 0.730 in the training and validation stage, respectively. The  $R$  values of SVM+L at site STL185 were 0.858 and 0.845 in the training and validation stage, respectively. Figure 12 shows the observed and predicted groundwater levels using SVM, SVM+L and EEMD-SVM in the validation stage.

**Table 2.** Results of modeling from the SVM model at site M1255 and STL185.

		Training					Validation				
		R	NMSE	RMSE	NS	AIC	R	NMSE	RMSE	NS	AIC
M1255	EEMD-SVM	0.948	0.102	0.206	0.898	-530.463	0.885	0.292	0.315	0.696	-55.373
	SVM	0.786	0.450	0.434	0.548	-280.434	0.7089	0.508	0.416	0.470	-42.067
	SVM+L	0.758	0.424	0.421	0.574	-290.376	0.730	1.149	0.626	-0.199	-22.467
STL185	EEMD-SVM	0.972	0.073	0.360	0.927	-343.467	0.860	0.507	1.009	0.471	0.447
	SVM	0.931	0.134	0.488	0.865	-241.064	0.845	0.283	0.755	0.704	-13.512
	SVM+L	0.858	0.273	0.696	0.725	-121.615	0.845	0.284	0.756	0.703	-13.441



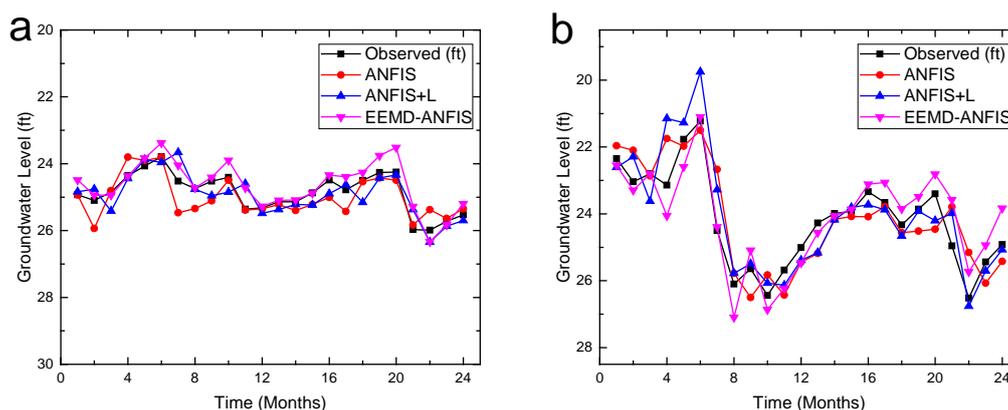
**Figure 12.** Observed and predicted groundwater levels using SVM, SVM+L and EEMD-SVM (a) M1255 (b) STL185.

### 5.3. EEMD-ANFIS and ANFIS Models

In the ANFIS model, the grid partition algorithm was used to generate a fuzzy inference system (FIS) structure from the training data [47]. The same input structures were introduced to the ANFIS model, the results are shown in Table 3. Table 3 suggested that the input variables of EEMD-ANFIS at two sites had the minimum RMSE and AIC values in the validation. The minimum RMSE and AIC of EEMD-ANFIS at site M1255 were 0.360 and  $-49.100$ , respectively. The minimum RMSE and AIC of EEMD-ANFIS at site STL185 were 0.606 and  $-24.035$ , respectively. The result showed that the maximum  $R$  value of EEMD-ANFIS in the validation was 0.926 at site M1255 and 0.909 at site STL185. The  $R$  values of four input variables of ANFIS+L at two sites were greater than that of the three input variables of ANFIS in the validation stage. The  $R$  value of ANFIS+L at site M1255 was 0.799 in the validation. The  $R$  value of ANFIS+L at site STL185 was 0.910 in the validation. Figure 13 shows the observed and predicted groundwater levels using ANFIS, ANFIS+L and EEMD-ANFIS in the validation stage.

**Table 3.** Result of modeling from the ANFIS model at site M1255 and STL185.

		Training					Validation				
		R	NMSE	RMSE	NS	AIC	R	NMSE	RMSE	NS	AIC
M1255	EEMD-ANFIS	0.968	0.063	0.162	0.937	-611.840	0.926	0.379	0.360	0.605	-49.100
	ANFIS	0.918	0.157	0.257	0.842	-457.099	0.785	0.496	0.412	0.482	-42.602
	ANFIS+L	0.971	0.058	0.156	0.942	-624.287	0.799	0.443	0.389	0.538	-45.341
STL185	EEMD-ANFIS	0.982	0.035	0.249	0.965	-466.512	0.909	0.183	0.606	0.809	-24.035
	ANFIS	0.940	0.117	0.455	0.883	-264.408	0.855	0.318	0.799	0.668	-10.756
	ANFIS+L	0.980	0.039	0.264	0.961	-447.574	0.910	0.262	0.726	0.726	-15.360

**Figure 13.** Observed and predicted groundwater levels using ANFIS, ANFIS+L and EEMD-ANFIS (a) M1255 (b) STL185.

#### 5.4. Comparison of EEMD-ANN, EEMD-SVM and EEMD-ANFIS

The best statistical parameter values of validation are shown in Tables 1–3, when the IMF components and one residual component were considered as the input variables. Also, based on the results of the statistical analysis, it was found that taking the lake level into account for the input variables led to better prediction in terms of the accuracy. Since the two well sites for data collection were close to the lake, groundwater-lake interaction could affect the groundwater level fluctuations. The results showed that the lake level should be considered as an input variable when exogenous factors were used to forecast the groundwater level for those well sites.

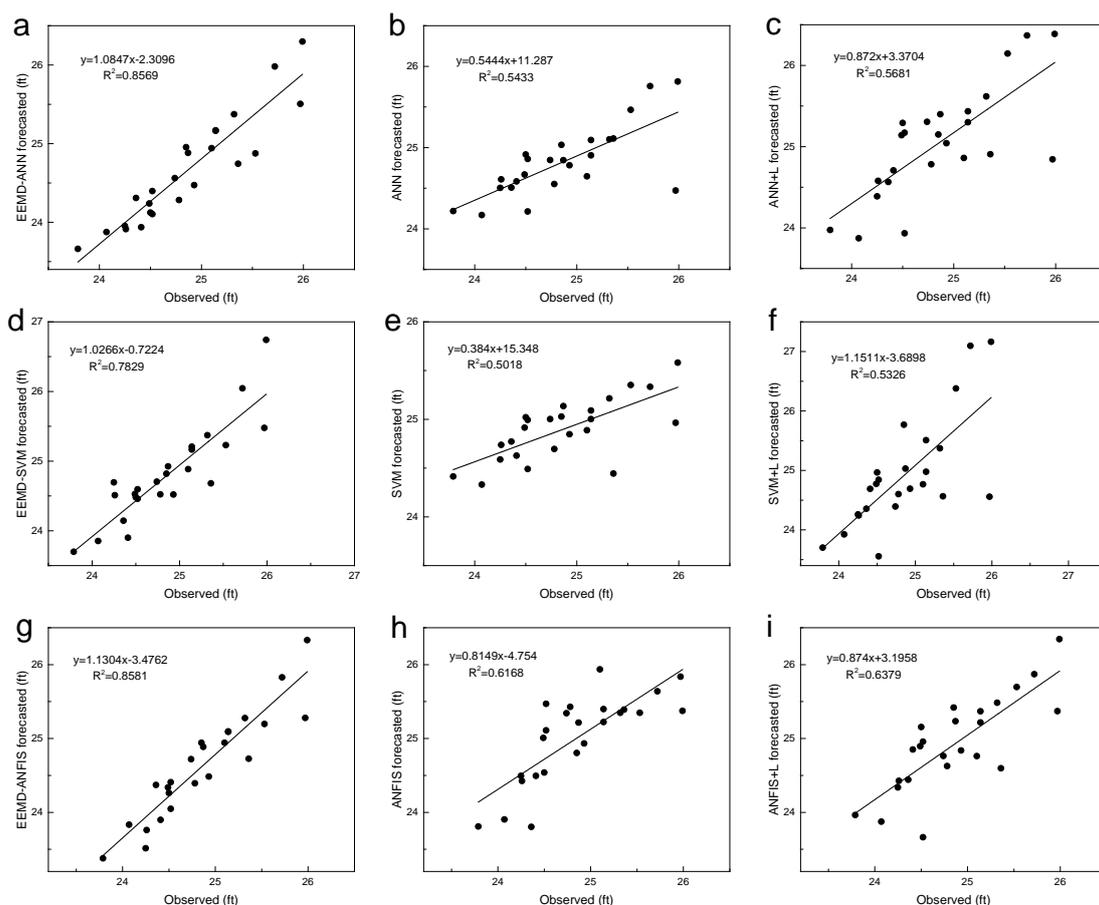
The analysis results from the model training and validation stages for two well sites are listed in Tables 1–3. At the training stage, the RMSE values in the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models for well M1255 were 0.234, 0.206 and 0.162 respectively; the RMSE values in these models for well STL185 were 0.281, 0.360 and 0.249 respectively. The RMSE value of the EEMD-ANFIS model was smaller than those of other two models in the training stage, which implied that the prediction ability of the EEMD-ANFIS model was better than that of the other two models for the given data. In the validation stage, the RMSE values for the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models for well M1255 were 0.329, 0.315 and 0.360 respectively; the RMSE of these models for well STL185 were 0.646, 1.009 and 0.606 respectively. For well M1255, the prediction result based on EEMD-ANFIS was close to that obtained from the other two models; the prediction result of EEMD-ANFIS was more accurate than that of the other two models for well STL185.

Generally, the modeling result is regarded as a perfect estimation when the NS criterion is equal to 1. If the NS criterion is higher than 0.8, the model can be recognized as effective and accurate [48]. The NS values for the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models in the training stage were all greater than 0.8. This indicated that the results from all these hybrid models are acceptable for forecasting groundwater levels. In the validation stage, the NS values for the EEMD-SVM model at site M1255 were greater than those for the EEMD-ANN and EEMD-ANFIS models. The NS values for the EEMD-ANFIS model at site STL185 were higher than those for the EEMD-ANN and EEMD-SVM

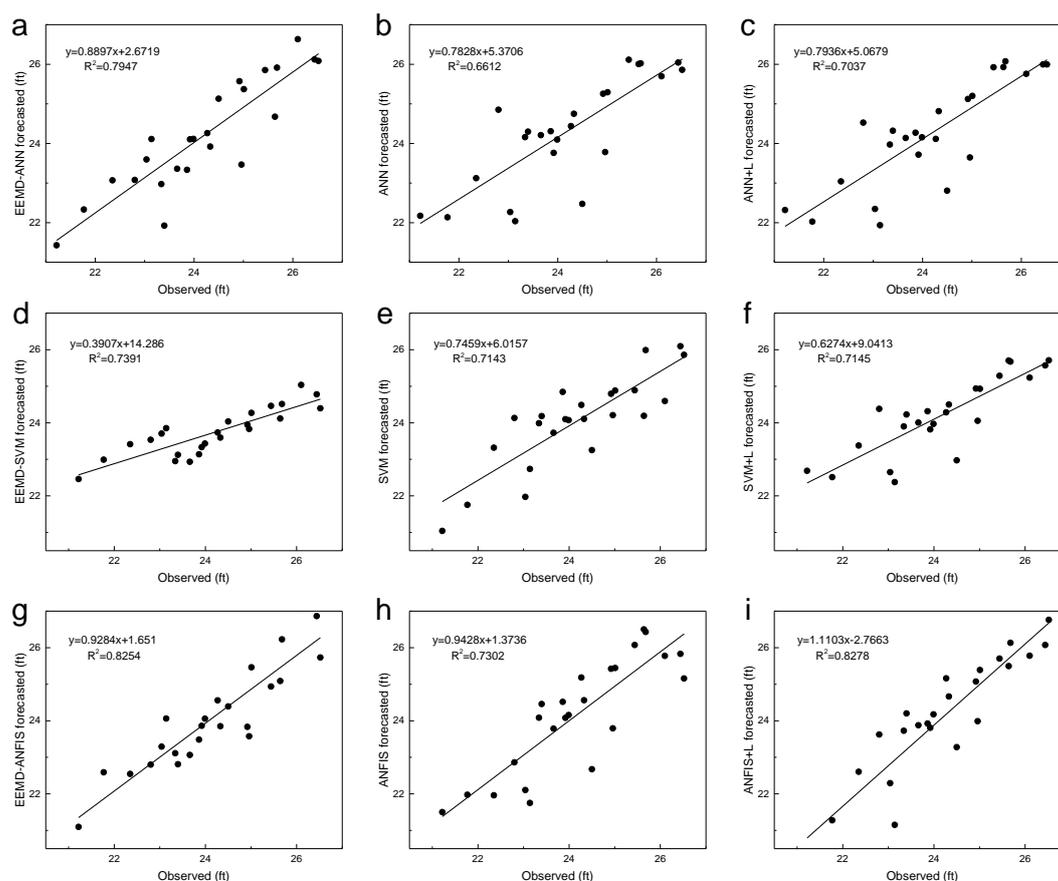
models in the validation stage. This indicated that EEMD-ANFIS and EEMD-SVM had an overall better estimation quality in comparison with the EEMD-ANN model. (See Tables 1–3).

When comparing the RMSE and  $R$  values of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models in the validation, the RMSE value of the EEMD-SVM model at site M1255 was less than that in both the EEMD-ANN and the EEMD-ANFIS model. The RMSE value of the EEMD-ANFIS model at site STL185 was higher than both the EEMD-ANN and the EEMD-SVM model. For the results for the two well sites, EEMD-ANFIS had higher  $R$  values compared to the other models. Obviously, the  $R$  and  $NS$  values of the EEMD-ANFIS and the EEMD-SVM model in the validation stage were greater than that of the EEMD-ANN model. Therefore, the EEMD-ANFIS and EEMD-SVM model could be good data-driven model in the validation stage. (See Tables 1–3).

Figures 14 and 15 illustrate the coefficient of determination ( $R^2$ ) values, corresponding to the predicted values in the scatter plots at the M1255 and STL185 observation wells, using the EEMD-ANN, EEMD-SVM, EEMD-ANFIS, ANN, SVM, ANFIS, ANN+L, SVM+L and ANFIS+L models. The scatter plots revealed the relationships between the predicted and observed groundwater levels for two observation wells. It can be seen clearly from the scatter plots that the EEMD-ANFIS model forecast the groundwater levels with less scatter for the two observed wells. Figures 14 and 15 show that EEMD-ANFIS had the best fit line compared to the other models. Figure 16a,b show the forecast groundwater levels versus observed groundwater levels using all of the models in the training stage and the validation stage.



**Figure 14.** Observed and predicted groundwater levels at the M1255 observation well using (a) EEMD-ANN; (b) ANN; (c) ANN+L; (d) EEMD-SVM; (e) SVM; (f) SVM+L; (g) EEMD-ANFIS; (h) ANFIS; (i) ANFIS+L. ( $x$  is observed groundwater level,  $y$  is predicted groundwater level).



**Figure 15.** Observed and predicted groundwater levels at the STL185 observation well using (a) EEMD-ANN; (b) ANN; (c) ANN+L; (d) EEMD-SVM; (e) SVM; (f) SVM+L; (g) EEMD-ANFIS; (h) ANFIS; (i) ANFIS+L. ( $x$  is observed groundwater level,  $y$  is predicted groundwater level).

According to the analysis, the  $R^2$  value for the data at site M1255 indicated that EEMD-ANFIS performed better than the other models, although the RMSE value for the EEMD-SVM was less than that for EEMD-ANFIS. The  $R^2$  value of the three hybrid models (i.e., EEMD coupled) were better than that of the other models. Therefore, both the EEMD-ANFIS and EEMD-SVM models can be considered good data-driven models at site M1255. Figure 15 shows that the  $R^2$  value of EEMD-ANFIS at site STL185 was nearly equal to that of ANFIS+L and was a bit higher than that of EEMD-ANN. The RMSE value for EEMD-ANFIS was less than that for EEMD-ANN and ANFIS+L. Thus, the EEMD-ANFIS model can be considered the best estimation model at site STL185. For the two observation sites, the forecast results obtained from the three hybrid models were suggested to have better quality compared to those not coupled with EEMD.

Boxplot is importantly used to check whether the data-driven models are able to forecast these variations and corresponding prediction errors. It intuitively depicts the quartile values for the prediction error of groundwater data. Figures 17 and 18 display that the median value of the training errors is close to zero, which indicates the good performance of the data-driven models in the training stage in terms of the efficiency. Figures 17 and 18 show the comparison of errors between the results obtained by the EEMD-ANN, EEMD-SVM, EEMD-ANFIS, ANN, SVM, ANFIS, ANN+L, SVM+L and ANFIS+L models in the training period and the validation period. The results of the boxplots indicated that EEMD-ANFIS was the most accurate model in the training stage. In the validation stage, the result of the EEMD-ANFIS for well M1255 was close to that of the EEMD-SVM and EEMD-ANN; while in the prediction of the groundwater level for well STL185, the EEMD-ANFIS performed more accurately than the EEMD-SVM and EEMD-ANN.

Overall, the performance of the EEMD-ANFIS, EEMD-SVM and EEMD-ANN models was superior to the other models. The results of the ANFIS+L, SVM+L and ANN+L model were better than that of ANFIS, SVM and ANN in terms of R. It indicated that lake level fluctuations as an input variable is important in the prediction of the groundwater level in the near-lake area. Comparing the results of the EEMD-ANFIS, EEMD-ANN and EEMD-SVM models, the three models had their own advantages and disadvantages. The selection of the prediction model should balance the effects and benefits of the statistical parameters (e.g., R and RMSE) in both the training period and the validation period. The prediction results of the EEMD-ANFIS were close to that of the EEMD-SVM and EEMD-ANN models at site M1255; the prediction results of the EEMD-ANFIS were a bit more accurate than that of the EEMD-SVM and EEMD-ANN models at site STL185. The results of the application suggested that the EEMD-ANFIS, EEMD-SVM and EEMD-ANN models were feasible and effective. Also, EEMD can be used to improve the accuracy of predicting nonlinear and nonstationary time series. The results in this study are consistent with those acquired by [22,26,29].

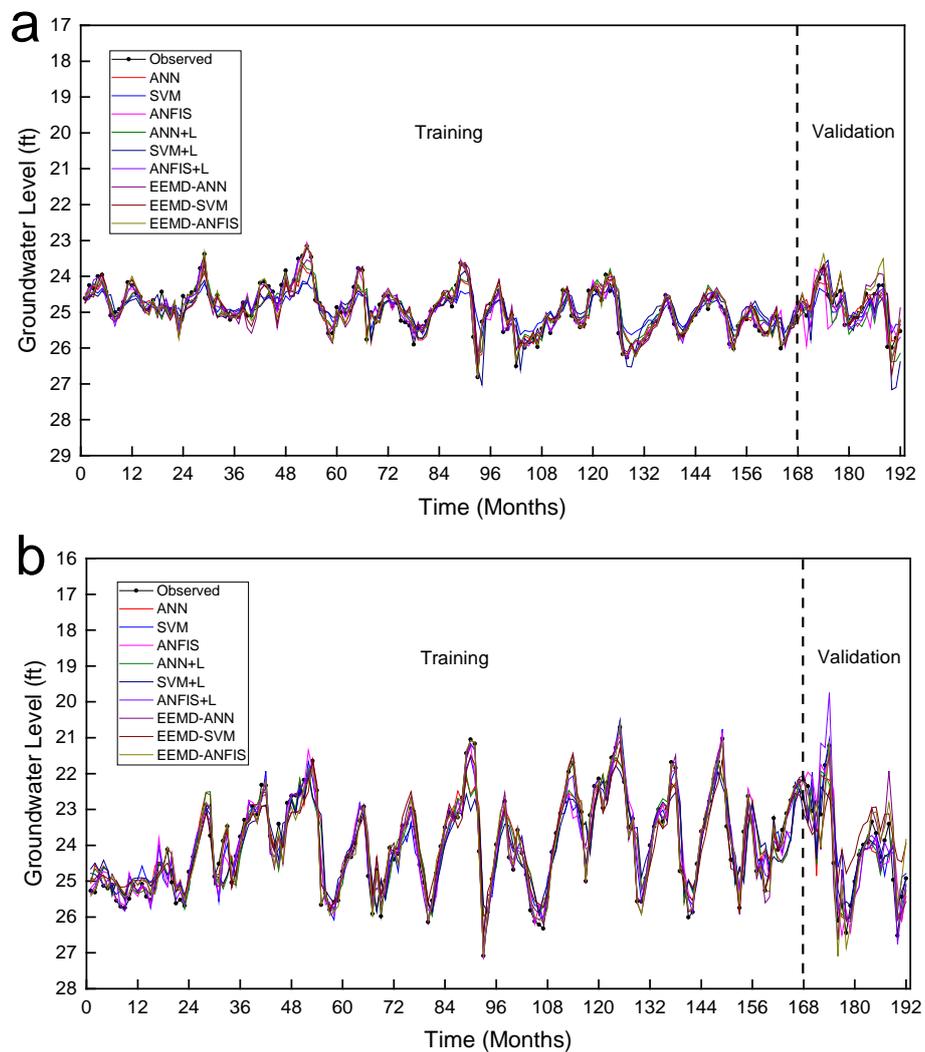


Figure 16. Model predictions versus observed data (a) M1255 and (b) STL185.

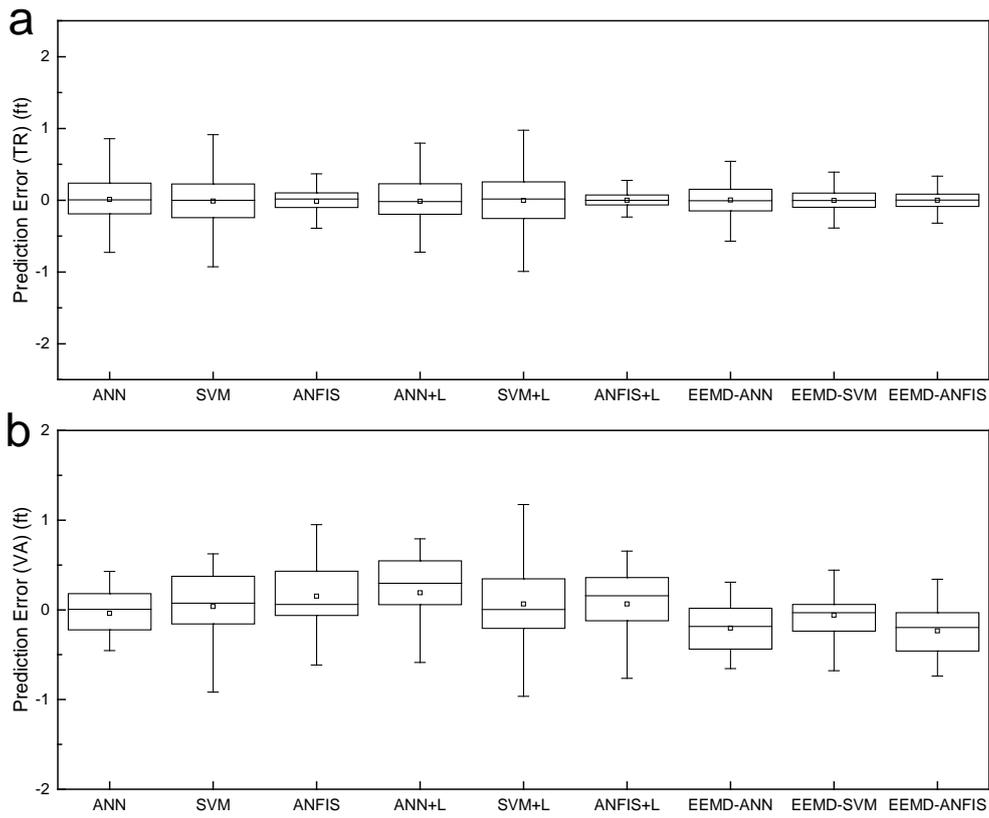


Figure 17. Box plots of the prediction error at site M1255 (a) Tr—Training; (b) Va—Validation.

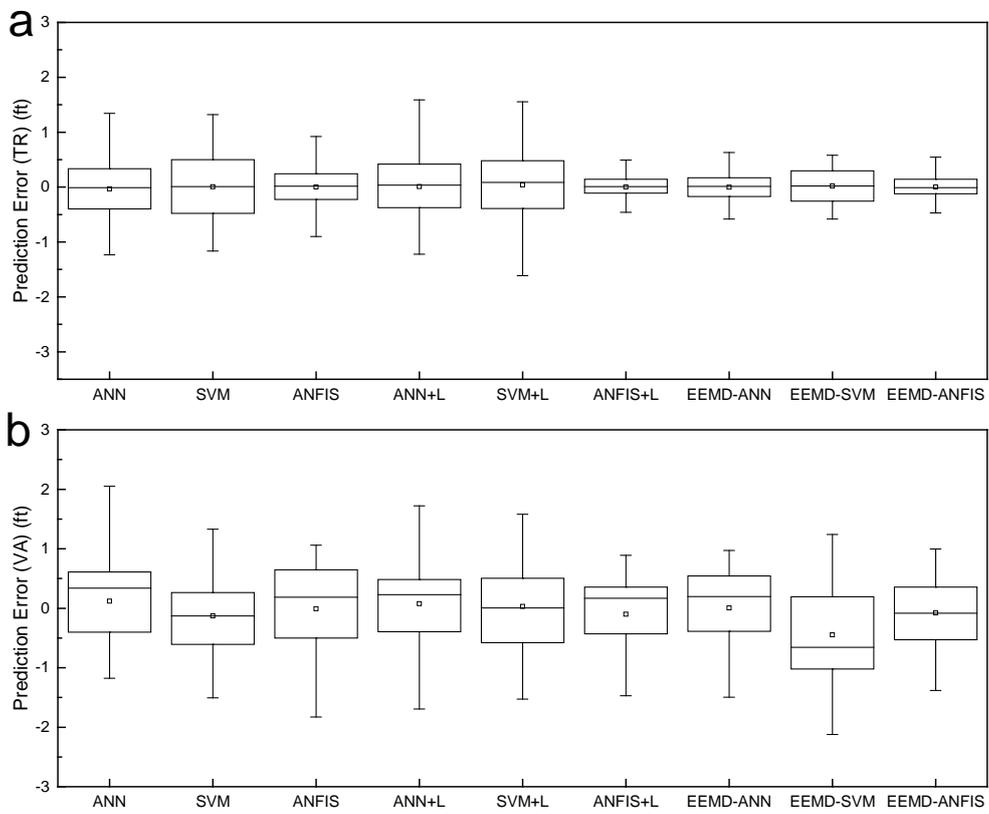


Figure 18. Box plots of prediction error at site STL185 (a) Tr—Training; (b) Va—Validation.

## 6. Conclusions

The reliable and accurate estimation of groundwater level fluctuation is essential in order to manage water resources and improve water-use efficiency. In this study, the prediction capability of the ANN, SVM and ANFIS models based on ensemble empirical mode decomposition (EEMD) were investigated using monthly groundwater level data collected at the M1255 and STL185 observation wells. The statistical parameters  $R$ , NMSE, RMSE, NS and AIC were used to assess the performance of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models. The results from the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models were analyzed and compared with the results from the ANN, SVM and ANFIS models. The values of the statistical parameters indicated that the EEMD-ANN, EEMD-SVM, and EEMD-ANFIS models achieved better prediction results than the ANN, SVM and ANFIS. The average  $R$  value of three hybrid models was higher than that of the ANN, SVM and ANFIS models, and the average RMSE value of these hybrid models was less than that of the ANN, SVM and ANFIS models.

The results in this study suggested that EEMD can effectively enhance predicting accuracy. The proposed EEMD could significantly improve the performance of the ANN, SVM and ANFIS data-driven models in groundwater level forecasting. The proposed three hybrid models based on EEMD had several obvious advantages: (a) it was convenient and effective to combine the EEMD with the ANN, SVM and ANFIS to forecast the nonstationary and nonlinear groundwater level fluctuations; (b) time series data on the groundwater level was only required in the hybrid models and exogenous factors affecting the groundwater level do not need to be considered in the research area; and (c) the prediction results of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models were more accurate when using the groundwater level time series decomposition. Therefore, this study supported the validity and applicability of the EEMD-ANN, EEMD-SVM and EEMD-ANFIS models in the prediction of groundwater levels. The results from this research would be beneficial for sustainable water resource management.

**Author Contributions:** Y.G. wrote the manuscript; Z.W., G.X. and Z.Z. revised and proofread the manuscript.

**Acknowledgments:** This study was supported by the National Key Research and Development Program (No. 2016YFC0402900, 2017YFC0403501), Key Research and Development Program of Qinghai Province (No. 2017-SF-116), National Natural Science Foundation of China (No. 41671020) and China Postdoctoral Science Foundation (No. 2017M610909). The authors gratefully acknowledge the helpful and insightful comments of the editor and anonymous reviewers.

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

## References

1. Chang, F.J.; Chang, L.C.; Huang, C.W.; Kao, I.F. Prediction of monthly regional groundwater levels through hybrid soft-computing techniques. *J. Hydrol.* **2016**, *541*, 965–976. [[CrossRef](#)]
2. Chang, J.; Wang, G.; Mao, T. Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. *J. Hydrol.* **2015**, *529*, 1211–1220. [[CrossRef](#)]
3. Emamgholizadeh, S.; Moslemi, K.; Karami, G. Prediction the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). *Water Resour. Manag.* **2014**, *28*, 5433–5446. [[CrossRef](#)]
4. Yoon, H.; Hyun, Y.; Ha, K.; Lee, K.-K.; Kim, G.-B. A method to improve the stability and accuracy of ANN- and SVM-based time series models for long-term groundwater level predictions. *Comput. Geosci.* **2016**, *90*, 144–155. [[CrossRef](#)]
5. Yoon, H.; Jun, S.C.; Hyun, Y.; Bae, G.O.; Lee, K.K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [[CrossRef](#)]
6. Fallah-Mehdipour, E.; Bozorg Haddad, O.; Mariño, M.A. Prediction and simulation of monthly groundwater levels by genetic programming. *J. Hydro-Environ. Res.* **2013**, *7*, 253–260. [[CrossRef](#)]

7. Nourani, V.; Alami, M.T.; Vousoughi, F.D. Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *J. Hydrol.* **2015**, *524*, 255–269. [[CrossRef](#)]
8. Suryanarayana, C.; Sudheer, C.; Mahmood, V.; Panigrahi, B.K. An integrated wavelet-support vector machine for groundwater level prediction in visakhapatnam, india. *Neurocomputing* **2014**, *145*, 324–335. [[CrossRef](#)]
9. Jiang, Q.; Tang, Y. A general approximate method for the groundwater response problem caused by water level variation. *J. Hydrol.* **2015**, *529*, 398–409. [[CrossRef](#)]
10. Shirmohammadi, B.; Vafakhah, M.; Moosavi, V.; Moghaddamnia, A. Application of several data-driven techniques for predicting groundwater level. *Water Resour. Manag.* **2013**, *27*, 419–432. [[CrossRef](#)]
11. Moosavi, V.; Vafakhah, M.; Shirmohammadi, B.; Behnia, N. A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. *Water Resour. Manag.* **2013**, *27*, 1301–1321. [[CrossRef](#)]
12. Shiri, J.; Kişi, Ö. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Comput. Geosci.* **2011**, *37*, 1692–1701. [[CrossRef](#)]
13. Valipour, M.; Banihabib, M.E.; Behbahani, S.M.R. Comparison of the arma, arima, and the autoregressive artificial neural network models in forecasting the monthly inflow of dez dam reservoir. *J. Hydrol.* **2013**, *476*, 433–441. [[CrossRef](#)]
14. Xu, T.; Valocchi, A.J.; Choi, J.; Amir, E. Use of machine learning methods to reduce predictive error of groundwater models. *Ground Water* **2014**, *52*, 448–460. [[CrossRef](#)] [[PubMed](#)]
15. Asefa, T.; Kemblowski, M.; Urroz, G.; Mckee, M. Support vector machines (SVMs) for monitoring network design. *Ground Water* **2005**, *43*, 413–422. [[CrossRef](#)] [[PubMed](#)]
16. White, S.; Robinson, J.; Cordell, D.; Jha, M.; Milne, G. *Urban Water Demand Forecasting and Demand Management: Research Needs Review and Recommendations*; Water Services Association of Australia: Melbourne, Australia, 2003.
17. Wang, Y.-H.; Yeh, C.H.; Young, H.W.V.; Hu, K.; Lo, M.T. On the computational complexity of the empirical mode decomposition algorithm. *Phys. A Stat. Mech. Appl.* **2014**, *400*, 159–167. [[CrossRef](#)]
18. Ghalekhondabi, I.; Ardjmand, E.; Young, W.A., II; Weckman, G.R. Water demand forecasting: Review of soft computing methods. *Environ. Monit. Assess.* **2017**, *189*, 313. [[CrossRef](#)] [[PubMed](#)]
19. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
20. 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](#)]
21. Wang, W.C.; Chau, K.W.; Xu, D.M.; Chen, X.Y. Improving forecasting accuracy of annual runoff time series using arima based on EEMD decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675. [[CrossRef](#)]
22. Wang, W.C.; Chau, K.W.; Qiu, L.; Chen, Y.B. 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](#)]
23. 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](#)]
24. Liu, H.; Tian, H.-Q.; Li, Y.-F. Comparison of new hybrid feemd-mlp, feemd-ANFIS, wavelet packet-mlp and wavelet packet-ANFIS for wind speed predictions. *Energy Convers. Manag.* **2015**, *89*, 1–11. [[CrossRef](#)]
25. Duan, W.Y.; Han, Y.; Huang, L.M.; Zhao, B.B.; Wang, M.H. A hybrid emd-svr model for the short-term prediction of significant wave height. *Ocean Eng.* **2016**, *124*, 54–73. [[CrossRef](#)]
26. Ausati, S.; Amanollahi, J. Assessing the accuracy of ANFIS, EEMD-GRNN, PCR, and MLR models in predicting PM 2.5. *Atmos. Environ.* **2016**, *142*, 465–474. [[CrossRef](#)]
27. 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](#)]
28. Karthikeyan, L.; Nagesh Kumar, D. Predictability of nonstationary time series using wavelet and emd based arma models. *J. Hydrol.* **2013**, *502*, 103–119. [[CrossRef](#)]
29. Hawinkel, P.; Swinnen, E.; Lhermitte, S.; Verbist, B.; Van Orshoven, J.; Muys, B. A time series processing tool to extract climate-driven interannual vegetation dynamics using ensemble empirical mode decomposition (EEMD). *Remote Sens. Environ.* **2015**, *169*, 375–389. [[CrossRef](#)]

30. Humeau-Heurtier, A.; Abraham, P.; Mahe, G. Analysis of laser speckle contrast images variability using a novel empirical mode decomposition: Comparison of results with laser doppler flowmetry signals variability. *IEEE Trans. Med. Imaging* **2015**, *34*, 618–627. [[CrossRef](#)] [[PubMed](#)]
31. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. An unconstrained optimization approach to empirical mode decomposition. *Digit. Signal Process.* **2015**, *40*, 164–175. [[CrossRef](#)]
32. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
33. ASCE. Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng.* **2000**, *5*, 115–123.
34. ASCE. Artificial neural networks in hydrology. II: Hydrologic applications. *J. Hydrol. Eng.* **2000**, *5*, 124–137.
35. Gong, Y.; Zhang, Y.; Lan, S.; Wang, H. A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near lake okeechobee, florida. *Water Resour. Manag.* **2016**, *30*, 375–391. [[CrossRef](#)]
36. Vapnik, V.N. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
37. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
38. Platt, J.C. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods*; MIT Press: Cambridge, MA, USA, 1999.
39. Scholkopf, B.; Smola, A.J. *Learning with Kernels: SUPPORT Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2001.
40. Chang, C.C.; Lin, C.J. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]
41. Jang, J.S.R. ANFIS: Adaptive ne twork based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
42. Jang, J.S.R.; Sun, C.T.; Mizutani, E. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*; Prentice Hall: Upper Saddle River, NJ, USA, 1996.
43. Valizadeh, N.; El-Shafie, A. Forecasting the level of reservoirs using multiple input fuzzification in ANFIS. *Water Resour. Manag.* **2013**, *27*, 3319–3331. [[CrossRef](#)]
44. Akrami, S.A.; El-Shafie, A.; Jaafar, O. Improving rainfall forecasting efficiency using modified adaptive neuro-fuzzy inference system (MANFIS). *Water Resour. Manag.* **2013**, *27*, 3507–3523. [[CrossRef](#)]
45. Nourani, V.; Komasi, M. A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *J. Hydrol.* **2013**, *490*, 41–55. [[CrossRef](#)]
46. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *AC-19*, 716–723. [[CrossRef](#)]
47. *The Math Works*; Matlab; Version 9.0.0; The Math Works Inc.: Natick, MA, USA. Available online: <http://www.mathworks.com> (accessed on 3 March 2016).
48. Shu, C.; Ouarda, T.B.M.J. Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* **2008**, *349*, 31–43. [[CrossRef](#)]

