

Article

# Simulation of Titicaca Lake Water Level Fluctuations Using Hybrid Machine Learning Technique Integrated with Grey Wolf Optimizer Algorithm

Babak Mohammadi <sup>1</sup>, Yiqing Guan <sup>1,\*</sup>, Pouya Aghelpour <sup>2</sup>, Samad Emamgholizadeh <sup>3</sup>,  
Ramiro Pillco Zolá <sup>4</sup> and Danrong Zhang <sup>1,\*</sup>

<sup>1</sup> College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; Babakmohammadi@aol.com

<sup>2</sup> MSc graduated of Agricultural Meteorology, Department of Water Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan 65178-38695, Iran; aghelpour\_p68@yahoo.com

<sup>3</sup> Department of Water and Environmental Engineering, Faculty of Civil Engineering, Shahrood University of Technology, Shahrood 36199-95161, Iran; s\_gholizaeh517@yahoo.com

<sup>4</sup> Instituto de Hidráulica e Hidrología, Universidad Mayor de San Andrés, La Paz 699, Bolivia; rami\_lund99@hotmail.com

\* Correspondence: yiqingguan@hhu.edu.cn (Y.G.); danrong\_zhang@hhu.edu.cn (D.Z.)

Received: 27 July 2020; Accepted: 23 October 2020; Published: 27 October 2020



**Abstract:** Lakes have an important role in storing water for drinking, producing hydroelectric power, and environmental, agricultural, and industrial uses. In order to optimize the use of lakes, precise prediction of the lake water level (LWL) is a main issue in water resources management. Due to the existence of nonlinear relations, uncertainty, and characteristics of the time series variables, the exact prediction of the lake water level is difficult. In this study the hybrid support vector regression (SVR) and the grey wolf algorithm (GWO) are used to predict lake water level fluctuations. Also, three types of data preprocessing methods, namely Principal component analysis, Random forest, and Relief algorithm were used for finding the best input variables for prediction LWL by the SVR and SVR-GWO models. Before the LWL simulation on monthly time step using the hybrid model, an evolutionary approach based on different monthly lags was conducted for determining the best mask of the input variables. Results showed that based on the random forest method, the best scenario of the inputs was  $X_{t-1}$ ,  $X_{t-2}$ ,  $X_{t-3}$ ,  $X_{t-4}$  for the SVR-GWO model. Also, the performance of the SVR-GWO model indicated that it could simulate the LWL with acceptable accuracy (with RMSE = 0.08 m, MAE = 0.06 m, and  $R^2 = 0.96$ ).

**Keywords:** lake water level; prediction; data-driven techniques; hybrid model; support vector regression; Titicaca Lake

## 1. Introduction

Lake water level fluctuations have important impacts on water resource management planning, such as water supply management, irrigation and drinking, recharge of groundwater aquifers, ecological and environmental changes, fishing, tourism, and many other forms of activities [1,2]. Changes in lake levels are affected by many factors such as precipitation, direct and indirect runoff from adjacent basins, evaporation from the free surface of the lake, and interactions between the lake and the groundwater table. The study of the lake water fluctuations was always of interest to hydrologists and environmental experts [3]. The accurate prediction of a lake water level intervals, has always been one of the challenges faced by hydrologists. Prediction of the lake water level (LWL) fluctuations at various time intervals is a challenging task in water resources management, and forecasting it using

the records of past time series provides useful data for planning, designing, and constructing water resource projects. Hence, researchers are trying to predict the LWL through accurate and cost-effective ways. Lake water level forecasting can be done using some complex models that include many relevant parameters, but the sensitivity of the LWL to these factors may vary from one area to another [4].

Over recent decades, several theory-driven (e.g., conceptual hydrological models) and data-driven techniques have been adopted and extended to the use for lake water fluctuations simulation [5]. The process-driven models, e.g., IHACRES [6], HSPF [7], SWAT [8], stochastic dynamic models [9], are useful for understanding the physically-based hydrological process in a watershed. However, the main constraint for the use of these models is their extensive data requirements which restrict their implementation with intensively monitored watersheds. Generally, hydrologists are concerned about the accurate prediction at specific locations with affordable time and cost. On the other hand, the data-driven techniques are estimated not from the physically-based hydrological process, but from the concurrent hydro-meteorological input-output data time series [10]. In comparison with theory-driven techniques, fewer parameters are needed for data-driven models. Consequently, the data requirement of these models is significantly less than those of theory-driven models. Various types of data-driven techniques such as Support Vector Regression (SVR) [11], artificial neural networks (ANNs) [12], adaptive neuro-fuzzy inference system [13], genetic expression programming and M5 model trees [14], have been extensively applied in hydro-environmental studies including prediction of the hydrological variables.

In recent decades, artificial intelligence models have become a proven tool in modeling nonlinear parameters and have been utilized by researchers in different hydrological studies [15–20]. Support vector regression is one of the advanced models of machine learning that operates based on minimizing error [21]. Although many studies show the ability of support vector regression (SVR) models to predict hydrological variables [22], SVR performance can be enhanced by hybridization with the Grey Wolf Optimizer algorithm (GWO) [23]. The GWO algorithm was first introduced by Mirjalili et al. [24] and was used in hydrology and other associated areas [25].

Providing water needed for agriculture, industry, and beverage requires precise and long-term planning. Hence, the prediction of hydrological variables in order to manage the utilization of water resources is necessary, in which the level of lakes as a natural heritage is of special importance. To this end, it is very important to use methods that achieve the most accurate result with the least time and cost. Pillco and Bengtsson [26] used the water balance method to predict the water level of Poopo Lake at a monthly scale. The main inflow water source of this lake is from the Desaguadero River, which is an outflow of Titicaca Lake. They discussed the dynamic of water levels of Poopo Lake related to climate variability, as well as climate change scenarios. Their results indicated that lowering the water depth decreased one meter in a wet season; the Poopo lake will be dry in the dry season.

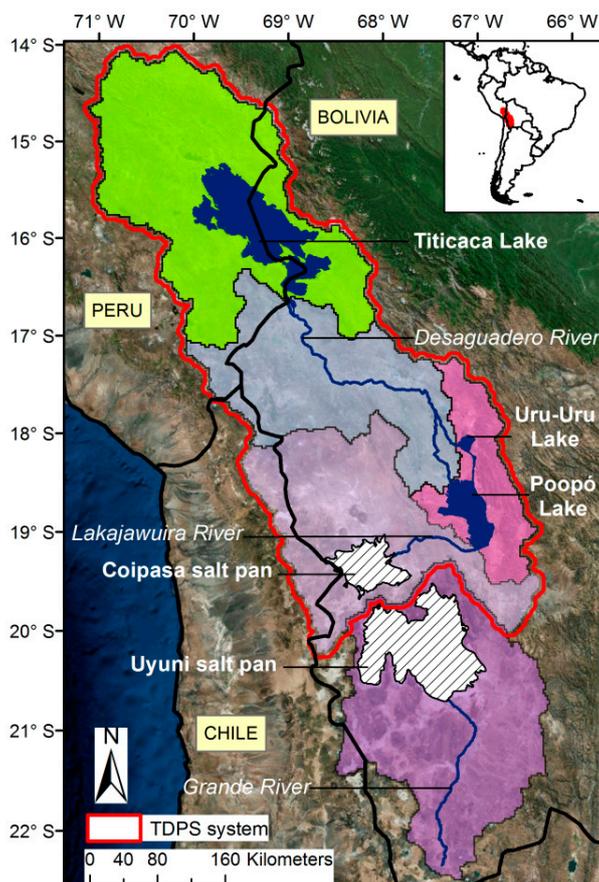
The objective of this study is to examine the ability of new hybrid support vector regression (SVR) model and the Grey Wolf Optimizer algorithm (GWO) for short-term prediction of lake water level. To do that, the data of the Titicaca Lake was used as a case study for the training and testing model. Finally, the performance of the hybrid SVR-GWO model was compared to the SVR model. Therefore, in this study, the hybrid support vector regression model and a Grey Wolf Optimizer model is used to improve the prediction of Titicaca lake water level at a monthly scale dataset during the 1973–2017 period. Also, in this research, we used three different pre-processing methods (principal component analysis, random forest, and Relief algorithm) to find the best mask of inputs for the proposed model.

## 2. Materials and Methods

### 2.1. Study Area

Titicaca Lake (TL) is one of the most important water body resources on the South American continent, plays an important role in the agriculture, habitat for animals, human being, and plant species; as well as in ecology of the region, economic and industrial activities, and the environment [27].

This lake is located at  $68^{\circ}33'–70^{\circ}1' W$  and  $15^{\circ}6'–16^{\circ}50' S$  between Peru and Bolivia (Figure 1). Titicaca Lake is an international wetland, the largest, deepest, and the highest navigable water body in the world; also, it is important from the point of view tourism for South America. Furthermore, this lake feeds significantly to the second largest and shallow Poopó Lake in Bolivia, placed at 300 km downstream from Titicaca [26]. The total lake surface is  $8560 \text{ km}^2$  at an elevation of 3810 m a.m.s.l, with a mean depth of 105 m. The outlet is still at 3807 m a.m.s.l. [28].



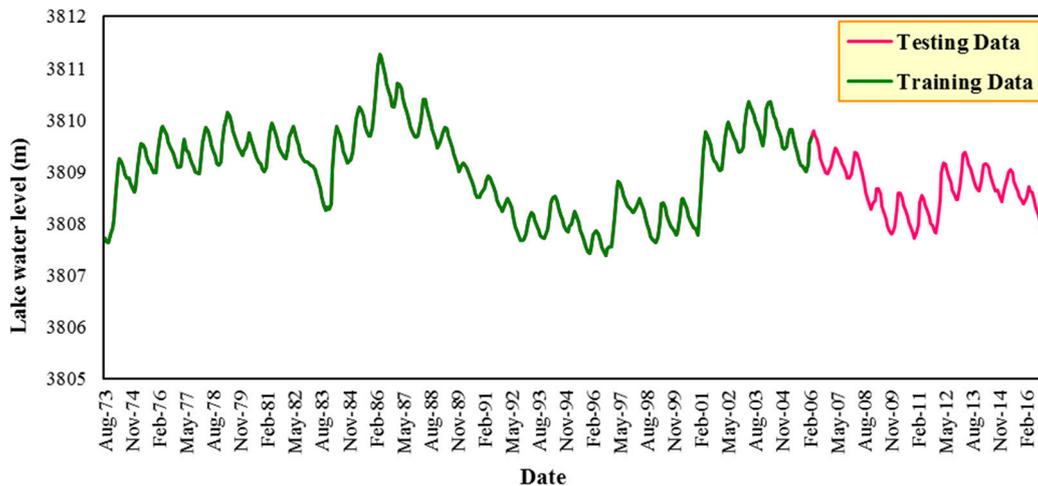
**Figure 1.** Location of Titicaca Lake in the study area.

## 2.2. Data Used

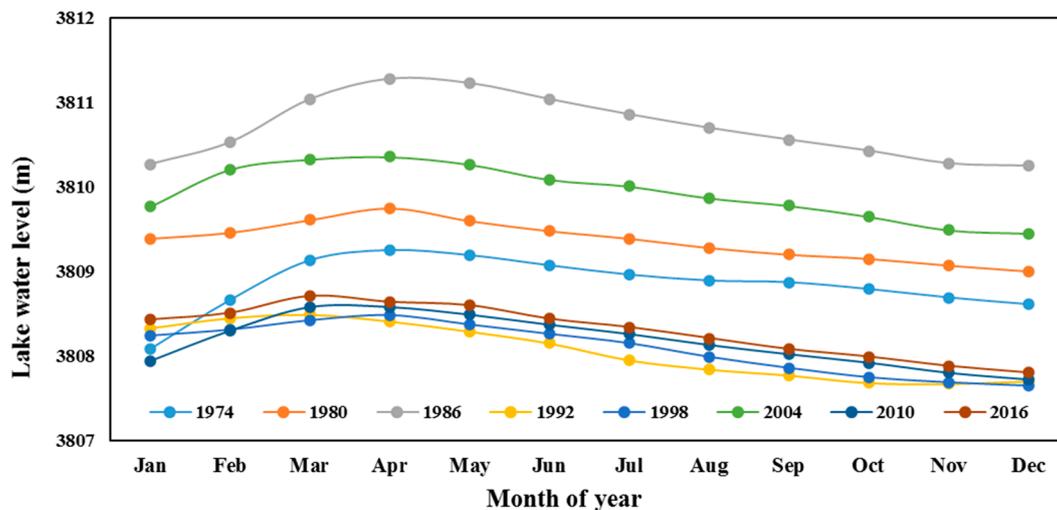
To conduct this research, the time series data of the water level which was measured in the hydrometric station of Titicaca Lake (located in southeast Peru) were used. As shown in Figure 1, this station is the closest representative station which present water level variations within the lake. The monthly time series data was used for tuning the parameters of the investigated models. For this aim, the selection dataset is from August 1973 to January 2017. 75% of the dataset (i.e., 392 data points) used for the training phase from August 1973 to March 2006 and 25% of dataset (i.e., 130 data points) used for the testing phase from April 2006 to January 2017. The average yearly precipitation of the Titicaca Lake basin is 609.5 mm/year, the average yearly minimum air temperature is  $-0.08 \text{ }^{\circ}\text{C}$ , and the average yearly maximum air temperature is  $17.2 \text{ }^{\circ}\text{C}$ . Table 1 shows the statistical characteristics of data used including length, minimum, maximum, mean, standard deviation, Skewness, Kurtosis, and confidence level. Additionally, Figures 2 and 3 indicate the time series of the water level of Titicaca for the duration of the study period (1973–2017) and also, monthly patterns of some years, respectively.

**Table 1.** Water level data used for the duration of the study period (1973–2017).

Dataset	Min (m)	Max (m)	Mean (m)	SD (m)	Skewness (m)	Kurtosis (m)	Confidence Level (95%)
Total	3807.387	3811.277	3808.959	0.781	0.11	−0.584	0.067
Training	3807.387	3811.277	3809.059	0.834	−0.101	−0.743	0.082
Testing	3807.717	3809.797	3808.658	0.483	−0.054	−0.701	0.083



**Figure 2.** Time series of Titicaca water level for the duration of the study period (1973–2017).



**Figure 3.** Monthly time series of water level for some years of the study period.

### 2.3. Preprocessing Methods

Model input combination and proper training and testing data lengths are among the main limiting factors on the predictive accuracy of any non-linear model [29]. In order to determine the best mask of input variables, the Principal component analysis (PCA), Random forest (RF) and Relief (RL) are used to abstract the most correlated variables.

#### 2.3.1. Principal Component Analysis (PCA)

The data dimension is the number of variables that are measured per view. Many of the data sets may have a high-dimension space, and therefore their interpretation will be difficult. Therefore, we tried to reduce the dimensions of data. Principal component analysis (PCA) is one of the most common methods of reducing the dimensions with the minimum error between the original data set

and the new obtained dimensions. PCA is an unsupervised learning method, is similar to clustering methods [30], and can be used for extracting important variables (in the form of the component) of the large set of variables in a dataset. In this method, the variables in a multimode correlated area are considered to be a collection of non-correlated components. The PCA method reduces the number of variables and finds the relationship structure between variables that are the same classification of variables. By applying the PCA variables, the main input is converted to new variables that are non-correlation. The PCA changes the input variables to the main components that are independent and linear combinations of input variables. In this method, data will be provided with a minimum of losses in the main components [10]. In this study, the first components of data sets, which has at least 80% of the total variance, are used as the input data for the predictive models.

### 2.3.2. Random Forest (RF)

Random Forests (RF) is a popular and powerful tool in order to select proper input variables to build predictive models. This method is used extensively across a multitude of engineering fields. It is also known as random decision forests. RF is one of the latest techniques which can be used for classification and regression-based analyses [31]. In this technique, the number of trees with different verity was used for forecasting or estimation. Tree predictors were used as random numerical values to class labels in the random forest classifier. Random forest regression used in this study contains an assembly of input parameters or arbitrarily chosen parameters at each node to grow a tree. RF technique requires only two user-defined parameters, such as the number of parameters used at each node and the number of trees [32].

### 2.3.3. Relief (RL)

Relief (RL) is an algorithm introduced by Kira and Rendell [33] that takes a filter-method approach to feature selection that is notably sensitive to feature interactions. The RL method is a random search method based on the filter model and no feedback on the learning algorithm is applied to assess the subset of the selected features. The stages of the RL, as the algorithm for reducing input data dimensions, are as follows: In the first step, each ranking correlation feature is given to the ultimate goal. In the second step, the ranks are updated using randomly selected samples. Then these ratings are sorted and the features with the lowest rank are removed, and the threshold is used to ranks the feature. The remaining features are retained as the superior features are kept in the input of a batch algorithm. The RL algorithm works fine for the noise characteristics or features of a good correlation, and the complexity of the time is linear and functional of the number of features and the number of exemplars the sample collection. It works well for both continuous and discrete datasets. The RL algorithm uses an evaluation function that calculates the sum of a statistically significant criterion and a criterion of complexity makes it a minimum. This algorithm finds the first feature to be able to make classes even better. It then finds properties, which in combination with the selected features, increase the separation of classes. This process is stopped when it gets to the least expected criteria for the pending representation [33].

## 2.4. Predictor Methods

### 2.4.1. Support Vector Regression

Support vector regression (SVR) is a set of supervised learning methods and can be used for classification and regression tasks. This method is based on binary classification in the space of arbitrary properties and is, therefore, a suitable method for predictive problems [21]. SVR can learn and model the nonlinear relation between the input and output data in the higher dimension, which minimizes the observed training error and the distribution error extent to attain generalized regression efficiency [10]. In fact, support vector regression is an efficient learning system based on effective optimization theory.

This technique implements the principle of inductive minimization of structural error in order to attain a general optimal solution. The structure of the SVR is presented in Figure 4.

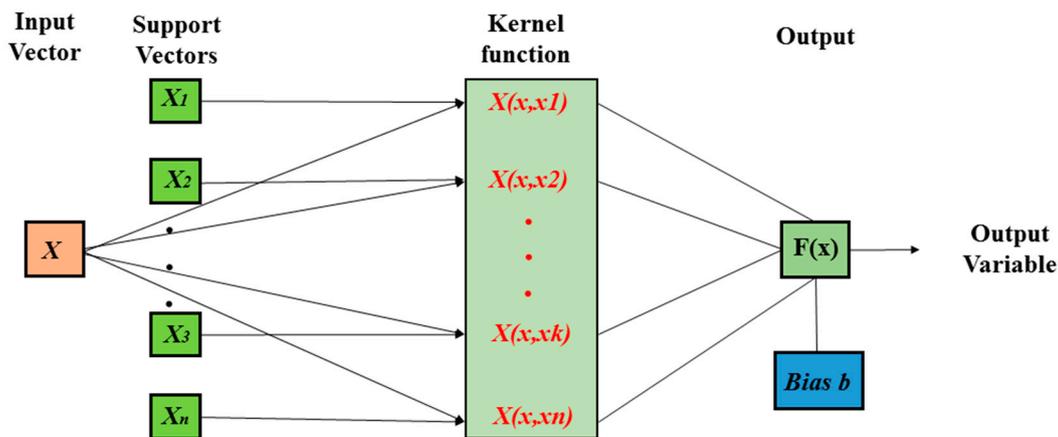


Figure 4. Support vector regression structure including input variables and output variable.

This study used radial basis function (RBF) as the learning algorithm for SVR. Let the training data set  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  where  $x_i (i = 1 - n)$  and  $y_i (i = 1 - n)$  indicate the  $i$ -th input and output vector, respectively, and  $n$  is regarded as the number of training data pairs. The SVR function attempts to detect a suitable function  $f(x)$  by minimizing the distribution error limits to attain general regression efficiency as follows:

$$f(x) = w^T \times \varphi(x) + b \tag{1}$$

where  $\varphi(x)$  represents the non-linear projection function which maps the input vector  $x$  into a higher-dimensional feature space where the trained data show linearity. However,  $w$  and  $b$  are the weight and constant coefficients, respectively, which are estimated by minimizing the arranged risk function.

#### 2.4.2. Grey Wolf Optimizer

The grey wolf optimizer (GWO) algorithm is a nature-inspired swarm-based algorithm that emulates the social hierarchy of wolves and their behavior in approaching, encircling, and attacking the prey [24–44]. Each member in a grey wolf herd categorized as  $\alpha, \beta, \delta$  and  $\omega$  based on the effectiveness and decision-making power in the group. The alpha wolf usually corresponds to the strongest and dominant wolf who leads the herd, and his/her order should be enforced by the rest of the group. The  $\beta$  wolves are the second group that plays the role of advisors for alphas. The beta wolves assist the alpha wolves by reinforcing their commands through the rest of the wolves [24]. The  $\delta$  wolves are ranked below the  $\alpha$  and  $\beta$  wolves and above the  $\omega$  wolves in terms of leadership hierarchy. They are guards, sentinels, hunters, and caretakers of the group. The  $\omega$  wolves are placed in the lowest order of decision-making and have to submit to all other wolves. To address an optimization problem using the GWO algorithm, the social behavior of the grey wolves in the hunting process is modeled in a statistical framework. The hunting process of the grey wolves is followed in three stages: tracking, encircling and attacking the prey [34]. The hunting process is organized by  $\alpha, \beta$  and  $\delta$  wolves. The  $\omega$  wolves follow  $\alpha, \beta$  and  $\delta$  wolves [35]. The best solution is considered as  $\alpha$ . The  $\beta, \delta$ , and  $\omega$  are the next subsequent solutions in terms of priority. In GWO iterations, the wolves evaluate the possible hunt situation, and update their status according to that. The mathematical formulation of the encircling process is as follows:

$$\vec{X}(t + 1) = \vec{X}_p(t) - \vec{A} \times \vec{C} \times \vec{X}_p(t) - \vec{X}(t) \tag{2}$$

In the above equation,  $t$  denotes the current iteration,  $\vec{X}_P$  and  $\vec{X}$  respectively implicates the position vectors of the hunt and hunter (candidate solutions),  $\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$  and  $\vec{C} = 2\vec{r}_2$  indicate coefficient vectors with  $\vec{r}_1$  and  $\vec{r}_2$  being as random vectors (between 0, 1) that allow the wolves update their status in the hunt space. Values of  $\vec{a}$  are linearly reduced from 2 to zero during the process as follows:

$$a(t) = 2 - \frac{2t}{\sum t} \quad (3)$$

where  $t$  signifies the running iteration and  $\sum t$  signify the total number of iterations. During the search process,  $\omega$  wolves update their status according to the best exploration factors i.e.,  $\alpha$ ,  $\beta$  and  $\delta$ , which represented by the following equations:

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \times \left| \vec{C}_1 \times \vec{X}_\alpha - \vec{X} \right| \quad (4)$$

$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \times \left| \vec{C}_2 \times \vec{X}_\beta - \vec{X} \right| \quad (5)$$

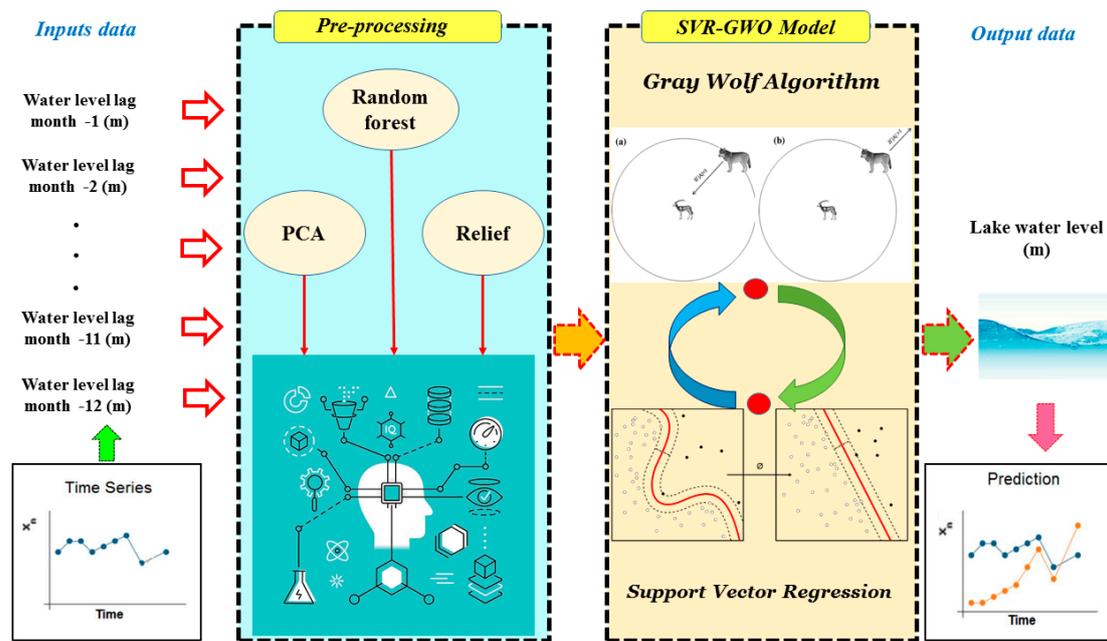
$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \times \left| \vec{C}_3 \times \vec{X}_\delta - \vec{X} \right| \quad (6)$$

$$\vec{X}(t + 1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (7)$$

#### 2.4.3. Hybrid SVR-GWO Model

In the proposed method, the Grey Wolf optimization algorithm is coupled with support vector regression (SVR-GWO). Indeed, the GWO used to optimize the SVR parameters. Figure 3 illustrates adjusting the search between local and global optima. In the weight optimization process, the weights are randomly assigned initially. The initial weights are changed and modified in each repetition as long as the difference of support vector regression output and the output related to the actual inputs is less than a specified limit.

In order to use the GWO for SVR weight updates, training SVR as a function should be highlighted and the problem aims to optimize it in an n-dimensional space. In this regard, the SVR is considered a grey wolf optimizer in d-dimensional space, d equals to the total number of weights and SVR bias. In other words, the location of each grey wolf has a dimension, which is equal to the total number of weights and SVR bias. This algorithm modifies each grey wolf location and returns the best wolf as the answer. Then, the weight of SVR is updated based on the values obtained from the best position of wolves. The best weight is acquired if one of two final conditions is established: first when the network minimum squared errors is less than a certain threshold, and second the number of iterations which is already known. Figure 3 displays a schematic view of the SVR model integrated with GWO optimizer applied for predicting lake water lake by using data-driven techniques method for data assimilation (Figure 5). It is an estimator hybrid procedure that utilizes the ANFIS capability together with optimization algorithms robustness. It is already confirmed that predictor tools can predict more successful results by coupling with optimization approaches [36–41].



**Figure 5.** Schematic view of predicting lake water lake by SVR and SVR-GWO by using data-driven techniques.

### 2.5. Performance Indexes

Various performance evaluation criteria such as root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R-squared) were utilized to evaluate the simulated results of each model, as illustrated in Equations (8)–(11), respectively:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{y})^2}{N}} \tag{8}$$

$$MAE = \frac{\sum (|x_i - y_i|)}{n} \tag{9}$$

$$R^2 = \left( \frac{1}{n} \times \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(\sigma_x)(\sigma_y)} \right)^2 \tag{10}$$

where  $x_i$  is the observed value,  $\bar{x}$  is the mean of observed values,  $y_i$  is the estimated value,  $\bar{y}$  is the mean of estimated values, and  $\sigma_x$  and  $\sigma_y$  are the standard deviation of the observed and estimated data, and  $n$  is the number of observations.

## 3. Results and Discussion

### 3.1. Implementation of the Preprocessing Methods

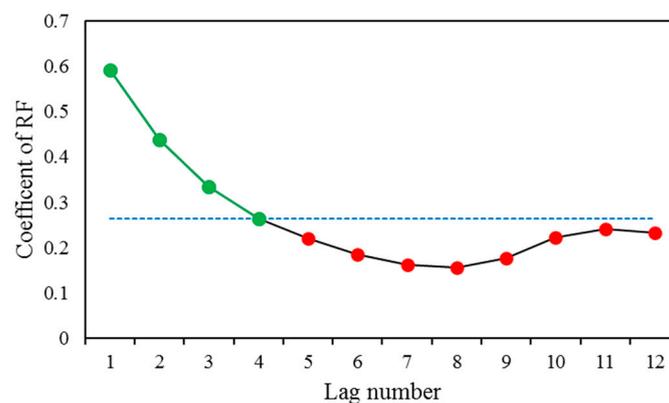
The predictive accuracy of the machine learning approaches relies significantly on the proper input combination and efficient size of training data. In this regard, three methods of the PCA, RF and RL, were implemented to find the best mask of the input variables. The predictor inputs in this study are selected among the time lags of the output series, as a time series prediction problem. This approach is applicable for forecasting the variable for one step (in this study one month) ahead. Table 2 lists the result of the PCA method to investigate the effective input variables. For this aim, the PCA method was employed to analyze the monthly time lags of water level. This method was implemented for 12 months of time lags and results show that PCA1 (component 1) covered 84.98% of all variances so

that PCA1 is selected as the best mask of the input variable for the predictor models. Also, PCA1 as the most effective input by PCA method introduced ( $PCA1 = 0.1117 \times L1 + 0.1138 \times L2 + 0.1155 \times L3 + 0.1169 \times L4 + 0.1179 \times L5 + 0.1183 \times L6 + 0.1183 \times L7 + 0.1179 \times L8 + 0.1169 \times L9 + 0.1155 \times L10 + 0.1138 \times L11 + 0.1117 \times L12 - 5287.7$ ).

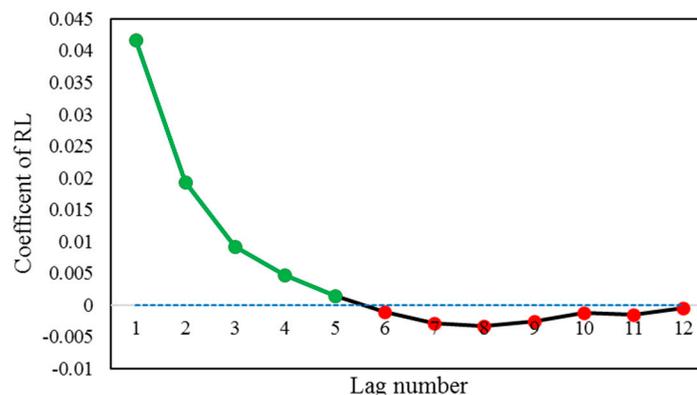
**Table 2.** Result of the PCA method for the pre-processing of Titicaca lake water level data.

Component	% of Variance	Cumulative %
1	84.98	84.98
2	8.05	93.03
3	5.24	98.27
4	1.05	99.32
5	0.38	99.7
6	0.18	99.88
7	0.05	99.93
8	0.03	99.96
9	0.01	99.97
10	0.01	99.98
11	0.01	99.99
12	0.01	100

Figures 6 and 7 show the results of the Random Forest (RF) (Figure 6) and Relief (RL) (Figure 7) method for pre-processing of the lake water lake dataset, respectively. Based on the results of these figures, RF have selected 4-months' time lag (i.e.,  $L(t - 1)$ ,  $L(t - 2)$ ,  $L(t - 3)$ ,  $L(t - 4)$ ), and RL have selected 5-months' time lag (i.e.,  $L(t - 1)$ ,  $L(t - 2)$ ,  $L(t - 3)$ ,  $L(t - 4)$ ,  $L(t - 5)$ ) for the best mask of the input variable which can be used for models. In addition, upper the line (green points) are effective lag numbers and below of line (red points) are not effective lag times.



**Figure 6.** Result of the RF method for the pre-processing of the lake water level data.



**Figure 7.** Result of the RL method for the pre-processing of the lake water level data.

After implementation of the pre-processing methods on historical lake water level data, three scenarios were defined for the predictor models (SVR and SVR-GWO) to be implemented. Table 3 shows three scenarios defined according to the selected input variables.

**Table 3.** Scenarios defined according to the selected input variables by PCA, RF, and RL pre-processing methods.

No.	Preprocessing Method	Input Combinations	Model Designation	
			SVR	SVR-GWO
1	Principal component analysis	PCA1	SVR1	SVR-GWO1
2	Random forest	$L(t-1), L(t-2), L(t-3), L(t-4)$	SVR2	SVR-GWO2
3	Relief	$L(t-1), L(t-2), L(t-3), L(t-4), L(t-5)$	SVR3	SVR-GWO3

### 3.2. Performances of the SVR and SVR-GWO Models

The SVR-GWO, meta-optimized hybrid version of the base SVR with the GWO algorithm, was utilized for the simulation of Titicaca lake water level in monthly scale. The training and testing of the SVR-GWO model were done with that measured data. The parameter settings of the SVR-GWO model have selected by trial and error method; whereas, the population size set to 20 and the max number of iterations is set to 500. The statistical results of the performance evaluation of the investigated models are also provided in Table 4. As shown in Table 4, when used, the SVR model and the GWO algorithm improved the SVR model efficiency in terms of RMSE, MAE, and  $R^2$ . The SVR-GWO model in the testing period and in the best scenario (with input of  $L(t-1), L(t-2), L(t-3), L(t-4)$ ) with the RMSE = 0.087 m, MAE = 0.066 m, and  $R^2 = 0.967$  outperformed the ordinary SVR model (with input of  $L(t-1), L(t-2), L(t-3), L(t-4), L(t-5)$ ) with the RMSE = 0.1 m, MAE = 0.082 m, and  $R^2 = 0.961$ .

**Table 4.** Performance evaluation of the SVR and SVRGWO models by different pre-processing methods in train and test stages.

Partition	Models Name	RMSE (m)	MAE (m)	$R^2$
Train phase	SVR1	0.395	0.312	0.774
	SVR2	0.11	0.082	0.982
	SVR3	0.108	0.08	0.983
	SVR-GWO1	0.304	0.208	0.877
	SVR-GWO2	0.084	0.06	0.989
	SVR-GWO3	0.079	0.052	0.99
Test phase	SVR1	0.329	0.276	0.558
	SVR2	0.101	0.084	0.96
	SVR3	0.1	0.082	0.961
	SVR-GWO1	0.27	0.223	0.727
	SVR-GWO2	0.087	0.066	0.967
	SVR-GWO3	0.089	0.064	0.966

Among the pre-processing used methods (i.e., PCA, RF and RL), the RL method has been the most successful method compared to the PCA and RF methods for determining inputs. Based on statistical indices, the third combination of data (SVR-GWO3) was selected as the best input data for prediction of the lake water level. The main reason for that was that this method introduced fewer inputs; furthermore, the performance of SVR-GWO3 was better than the other scenarios, while the RF method in SVR2 model selected more inputs and had weaker results. In general, the new hybrid model (SVR-GWO) had better result with less input than the ordinary SVR models. Also, it can be perceived that the GWO algorithm attained the best fitness values compared to ordinary SVR trainers' optimizers. So, in Table 4, the MAE's index for the training part of SVR model is 0.312 m (for SVR1), 0.082 m (for SVR2), and 0.08 m (for SVR3), respectively. The same magnitude for the training part of the

SVR-GWO model is 0.208 m (for SVR-GWO1), 0.06 m (for SVR-GWO2), and 0.052 m (for SVR-GWO3). This suggests that combining the grey wolf optimization algorithm with the support vector regression model can create a new hybrid model with high learning ability, which is a powerful tool (SVR-GWO) for estimating the level of lake water. The optimal parameters of the SVR ranged between  $\gamma$  (Radial basis function parameter) = 0.45 - 20.2 and C (Trade-off parameter) = 1.68 - 19.14. Also, for running the GWO used 500 as the maximum number of iterations, 50 as the number of agents.

As the comparison between results of Table 4 shows that using data with four lag times in model SVR-GWO2 and with five lag times in SVR-GWO3 model would not further increase the prediction accuracy. In other word, increasing lag times would not improve the accuracy of the model and the results have been converged. Also, this topic shows lag number 5 is not very effective in lake water level for Titicaca Lake. Then, by ignoring some minor errors, this study suggests using less inputs for models. Figure 8 reported the scatter plots of the predictions by the SVR and SVR-GWO models versus the actual values in the training and testing stages. Based on the presented graphical variation between the observed lake water level and predicted values, the SVR-GWO model attained the distinguished correlation with the highest value at all scenarios compared to the ordinary SVR model. This is evidenced by the potential of the SVR-GWO predictive model to capture the variance observations of the water level of Titicaca Lake with a couple of scattered observations. Indeed, using the GWO optimizer improved the performance of hybrid SVR-GWO compared to the original SVR model in all scenarios. Also, according to Figure 7, the PCA pre-processing method has a weak performance in determining the input variables for prediction of the lake water level, while pre-processing methods RF and RL have been successful in determining inputs.

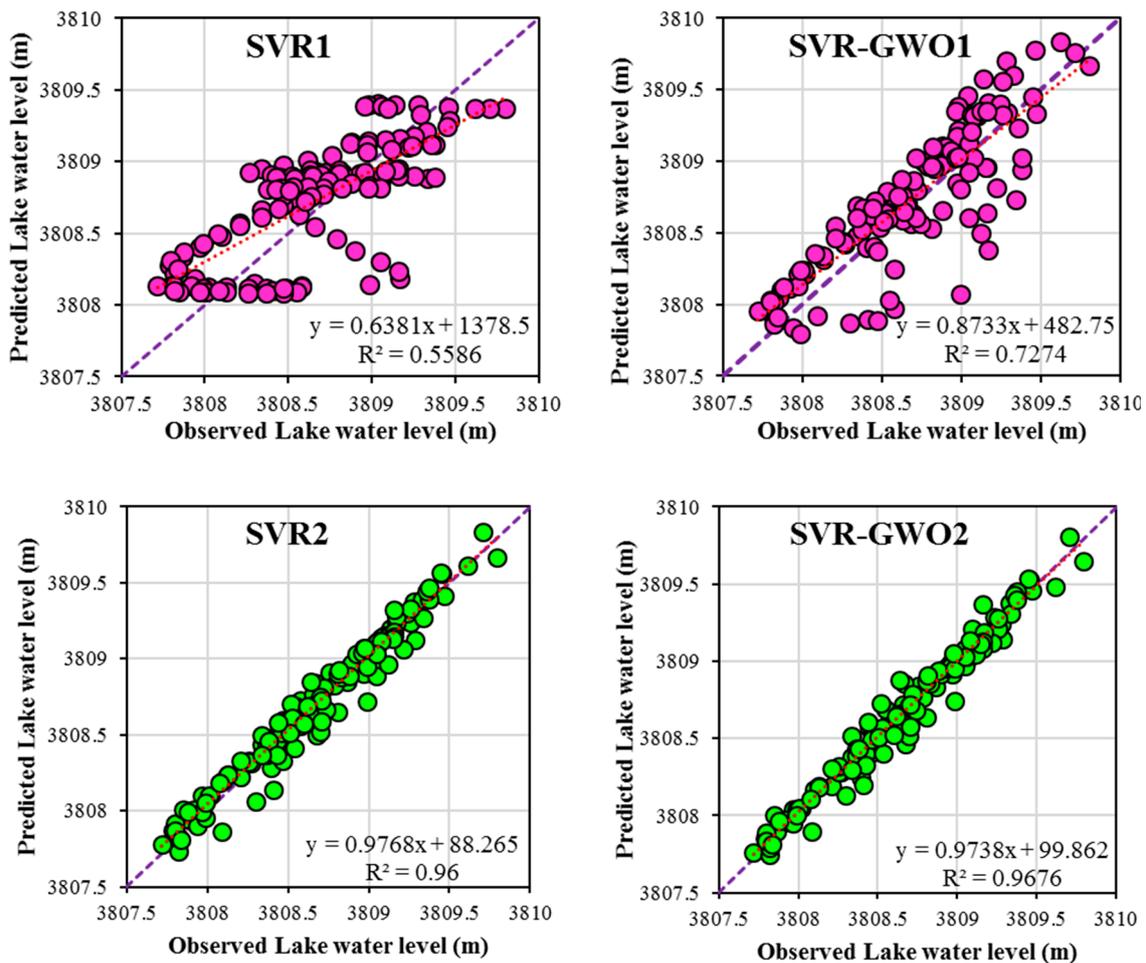
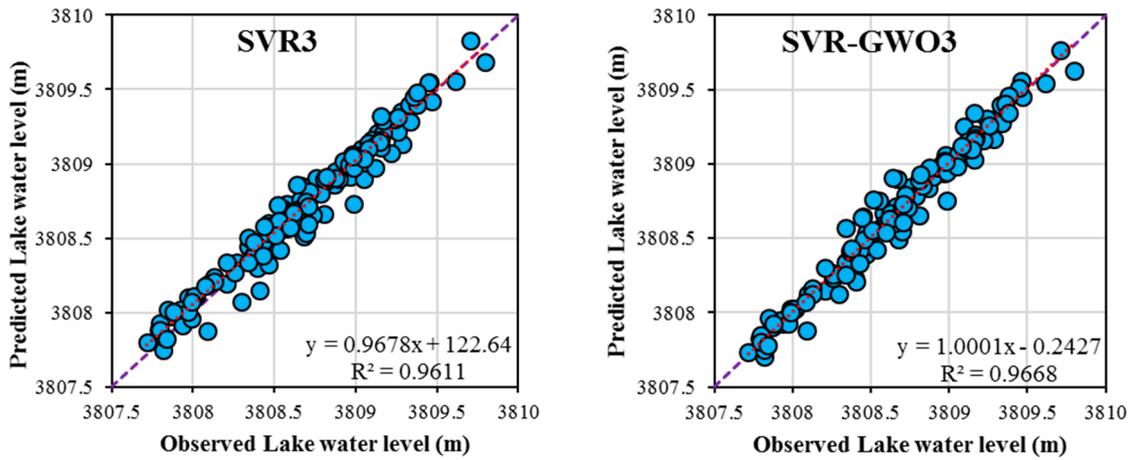
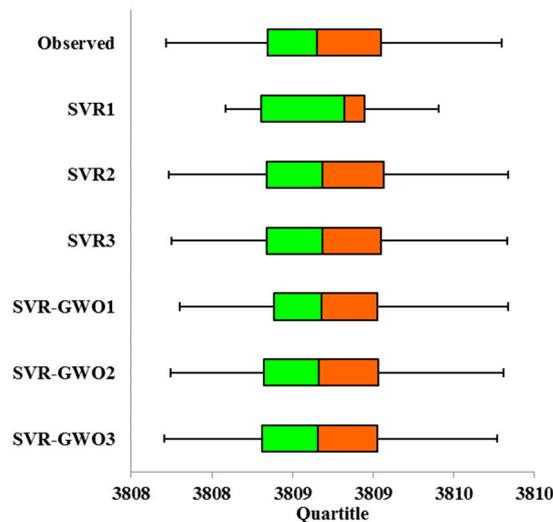


Figure 8. Cont.



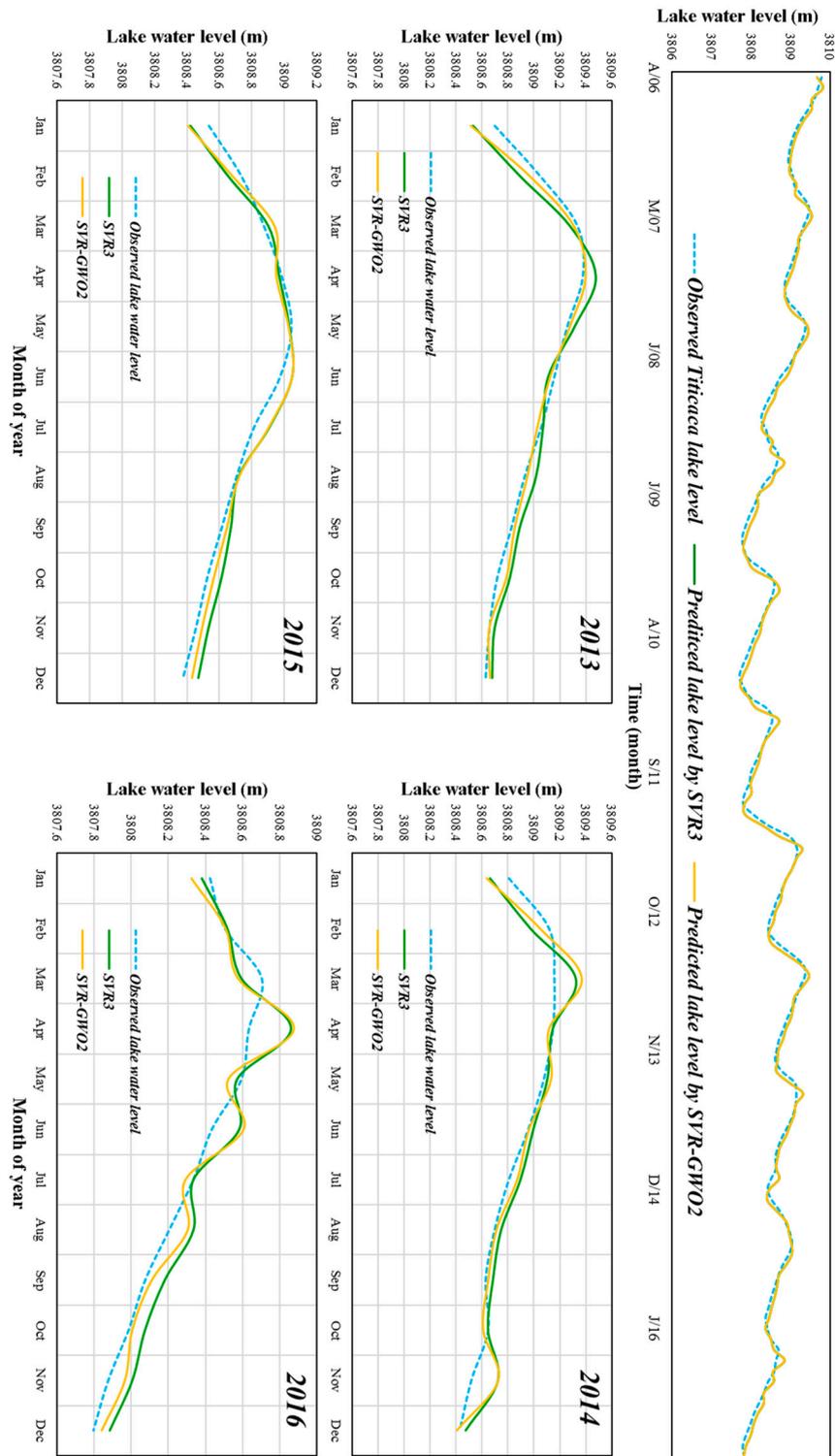
**Figure 8.** Scatter plot of the predicted lake water level from the SVR and SVR-GWO models versus the corresponding measurements for the testing stage.

Figure 9 shows the box plot of the measured and the predicted values by the SVR and SVR-GWO models for the testing stage. Based on these figure results obtained by the RL and RF with SVR-GWO model, it shows the highest similarities with observation so that the SVR-GWO2 and SVR-GWO3 can accurately simulate the minimum, maximum, 25% range and 75% range intervals. Based on the results of this study, the SVR-GWO model has been performing better and more accurately in predicting lake water level. Finally, SVR1 is the weakest model and SVR-GWO2 is a strong model for the prediction of lake water level.



**Figure 9.** Box plot of observed and predicted lake water level for the testing period of the SVR and SVR-GWO models in all scenarios.

In Figure 10, the best output of each model has plotted versus measured values for four years (2013, 2014, 2015, and 2016). This figure is drawn up to better understand the ability of each model to predict LWL for the testing period. The hybrid SVR-GWO2 model has better overlap with observational values and its disagreements with observational values are significantly less than the ordinary SVR model. Also, based on the results of this figure, most of the forecast errors in four years were from March to July in both models.



**Figure 10.** Time series plot of observed and predicted LWL for the best SVR and SVR-GWO models for four years of 2013, 2014, 2015 and 2016.

Figure 11 shows the prediction error of the SVR and SVR-GWO models for the testing period. According to this figure, the SVR-GWO2 and SVR-GWO3 models predict the LWL with low error, and subsequently they reduce the average error by 27% and 28% compared to the SVR1, SVR2, SVR3, and SVR-GWO1 models for testing stages, respectively. Accordingly, two models of the SVR-GWO2 and SVR-GWO3 models offer much better performance in terms of prediction of the high and the low

values of the Titicaca Lake water level dataset. Also, the combination of different models indicates that using the integrated of the SVR model and GWO algorithm increased the accuracy of predicting the lake water level. In addition, Figure 11 shows that two methods of the Random forest and Relief have a good ability for LWL data pre-processing and these methods can find the best mask of input variables.

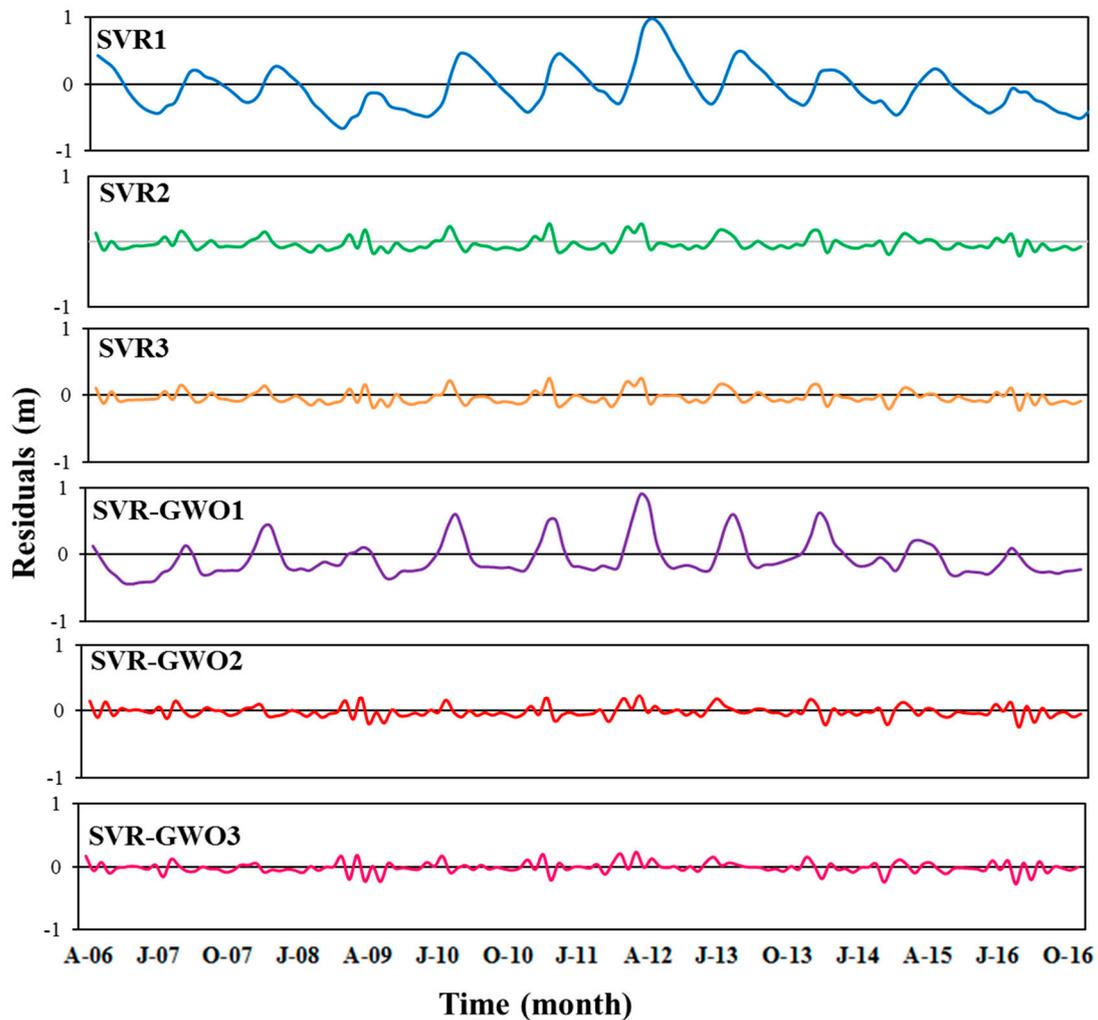


Figure 11. Time series of prediction error for the SVR and SVR-GWO models for the testing period.

Histogram plot is another method of evaluation ability of models. Figure 12 shows the histogram diagram for the testing period of all scenarios. Based on the result of this figure, the PCA method is a weak method for identifying effective variables for the prediction lake water level, so that both the SVR1 and SVR-GWO1 models had a poor performance, while RL and RF methods had better performance in identifying effective variables for lake water level prediction, because the SVR2, SVR3, SVR-GWO2 and SVR-GWO3 had better performance. Overall, the pre-processing of the RF method was more successful in identifying input variable to predict Titicaca lake water level compared the RL and PCA methods, and using this method improved the efficiency of the SVR-GWO model.

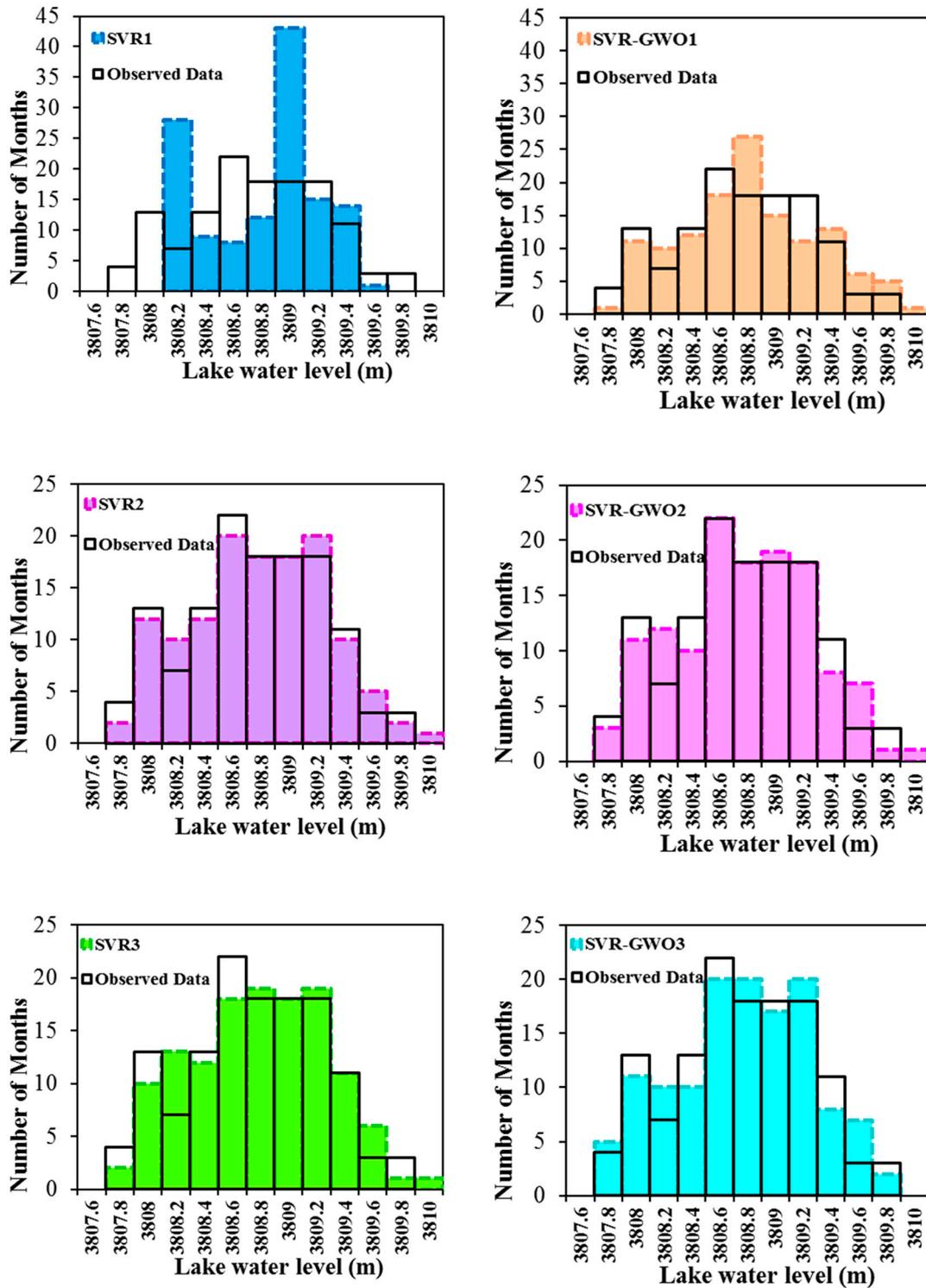


Figure 12. The histogram of the predicted and the observed ratio of lake level in the testing period, the data comparison intervals were selected from 3807.6 m to 3810 m.

There are some similar studies to discuss regarding the current study’s results. An SVR model was used to predict the water level of Poyang Lake in China [42], which showed similar results. In Li et al. [42] the  $R^2$  value of SVR predictions were about 0.95 & 0.96 and approved the high capability of this model in lake water level prediction; which is in line with the results of the current research, despite the lakes being located in two completely different areas. To evaluate the positive effect of merging SVR with GWO, there are no studies about lake water level prediction. However, in other

areas such as monthly streamflow forecasting, this was a successful and efficient combination; so that in hybrid forms of SVR, the GWO was superior in competition with the algorithms such as Particle Swarm Optimization (PSO), Multi-Verse Optimization (MVO) and Shuffled Complex Evolution (SCE) [43].

#### 4. Conclusions

In the present study, support vector regression (SVR) integrated with the grey wolf optimizer algorithm (GWO) has been adopted for the forecasting of the lake water level (LWL). The historical water level data of Titicaca Lake for the period of 1973–2017 were used as a case study. Three types of data preprocessing methods (PCA, RF, and RL) were used for finding the best input variables for prediction the water level of Titicaca Lake by the SVR and hybrid SVR-GWO models. Based on the results of three preprocessing methods, six scenarios are considered for developing these models. The results have been evaluated with several statistical score metrics (i.e., RMSE, MAE, and  $R^2$ ) and visual displays (i.e., scatter plot, box plot, etc.). Comparing the results of six scenarios from the implementation of the SVR and SVR-GWO models showed that the performance of the RF pre-processing method was better than PCA and RF methods for finding the best input for predictor models. The results demonstrated that the meta-optimized hybrid model (SVR-GWO) enhanced the capability of the original SVR model for the reproduction of the monthly lake water level. The SVR-GWO model with inputs of  $L(t-1)$ ,  $L(t-2)$ ,  $L(t-3)$ ,  $L(t-4)$  was found to be the most suitable model for prediction LWL. The results of this study suggest that the Grey Wolf Optimizer algorithm is a useful add-on tool for enhancing the accuracy of forecasting SVR model to predict LWL. Also, this research provided evidence for the effectiveness of the hybrid model, which can be utilized and investigated in hydrology for forecasting time series data such as lake water level (LWL). The results show that the models can be used to represent a 1-month (ahead) prediction of the lake water level. This can be applicable for the agricultural, industrial, environmental and urban sectors and systems related to Titicaca Lake and their managers and planners, to inform about the lake's water level status in the coming month.

**Author Contributions:** Conceptualization, B.M., P.A. and S.E.; methodology, B.M. and P.A.; software, B.M. and P.A.; validation, B.M., P.A., S.E. and R.P.Z.; data curation, B.M. and R.P.Z.; writing—original draft preparation, B.M., S.E. and R.P.Z.; writing—review and editing, Y.G. and D.Z.; visualization, B.M. and P.A.; supervision, Y.G. and D.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (NSFC51209064 and NSFC51579067), and the Qing Lan Project and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX17\_0416).

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

#### References

1. Ebtehaj, I.; Bonakdari, H.; Gharabaghi, B. A reliable linear method for modeling lake level fluctuations. *J. Hydrol.* **2019**, *570*. [[CrossRef](#)]
2. Seo, Y.; Kim, S.; Kisi, O.; Singh, V.P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J. Hydrol.* **2015**, *520*. [[CrossRef](#)]
3. Hu, T.; Mao, J.; Pan, S.; Dai, L.; Zhang, P.; Xu, D.; Dai, H. Water level management of lakes connected to regulated rivers: An integrated modeling and analytical methodology. *J. Hydrol.* **2018**, *562*. [[CrossRef](#)]
4. Ghorbani, M.A.; Deo, R.C.; Karimi, V.; Yaseen, Z.M.; Terzi, O. Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*. [[CrossRef](#)]
5. Rajurkar, M.P.; Kothiyari, U.C.; Chaube, U.C. Modeling of the daily rainfall-runoff relationship with artificial neural network. *J. Hydrol.* **2004**, *285*. [[CrossRef](#)]
6. Croke, B.F.W.; Andrews, F.; Jakeman, A.J.; Cuddy, S.M.; Luddy, A. IHACRES Classic Plus: A redesign of the IHACRES rainfall-runoff model. *Environ. Model. Softw.* **2006**, *21*. [[CrossRef](#)]
7. Bicknell, B.R.; Imhoff, J.C.; Kittle, J.L., Jr.; Donigan, A.S., Jr.; Johanson, R.C.; Barnwell, T.O. *Hydrological Simulation Program-Fortran User's Manual for Release 11*; U.S. Environ. Prot. Agency: Washington, DC, USA, 1996.

8. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large area hydrologic modeling and assessment part I: Model development. *J. Am. Water Resour. Assoc.* **1998**, *34*. [[CrossRef](#)]
9. Privalsky, V.E. Modeling long term lake variations by physically based stochastic dynamic models. *Stoch. Hydrol. Hydraul.* **1988**, *2*. [[CrossRef](#)]
10. Mohammadi, B.; Mehdizadeh, S. Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm. *Agric. Water Manag.* **2020**, *237*. [[CrossRef](#)]
11. Misra, D.; Oommen, T.; Agarwal, A.; Mishra, S.K.; Thompson, A.M. Application and analysis of support vector machine based simulation for runoff and sediment yield. *Biosyst. Eng.* **2009**, *103*. [[CrossRef](#)]
12. Emamgholizadeh, S.; Demneh, R.K. A comparison of artificial intelligence models for the estimation of daily suspended sediment load: A case study on the telar and kasilian rivers in Iran. *Water Sci. Technol. Water Supply* **2019**, *19*. [[CrossRef](#)]
13. 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*. [[CrossRef](#)]
14. Solomatine, D.P.; Xue, Y. M5 model trees and neural networks: Application to flood forecasting in the upper reach of the Huai River in China. *J. Hydrol. Eng.* **2004**, *9*. [[CrossRef](#)]
15. Emamgholizadeh, S.; Bahman, K.; Bateni, S.M.; Ghorbani, H.; Marofpoor, I.; Nielson, J.R. Estimation of soil dispersivity using soft computing approaches. *Neural Comput. Appl.* **2017**, *28*. [[CrossRef](#)]
16. Aghelpour, P.; Bahrami-Pichaghchi, H.; Kisi, O. Comparison of three different bio-inspired algorithms to improve ability of neuro fuzzy approach in prediction of agricultural drought, based on three different indexes. *Comput. Electron. Agric.* **2020**, *170*. [[CrossRef](#)]
17. Aghelpour, P.; Varshavian, V. Evaluation of stochastic and artificial intelligence models in modeling and predicting of river daily flow time series. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*. [[CrossRef](#)]
18. Guan, Y.; Mohammadi, B.; Pham, Q.B.; Adarsh, S.; Balkhair, K.S.; Rahman, K.U.; Linh, N.T.T.; Tri, D.Q. A novel approach for predicting daily pan evaporation in the coastal regions of Iran using support vector regression coupled with krill herd algorithm model. *Theor. Appl. Climatol.* **2020**, *142*. [[CrossRef](#)]
19. Aghelpour, P.; Mohammadi, B.; Biazar, S.M. Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. *Theor. Appl. Climatol.* **2019**, *138*. [[CrossRef](#)]
20. Buyukyildiz, M.; Tezel, G.; Yilmaz, V. Estimation of the change in lake water level by artificial intelligence methods. *Water Resour. Manag.* **2014**, *28*. [[CrossRef](#)]
21. Moazenzadeh, R.; Mohammadi, B.; Shamshirband, S.; Chau, K. Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Eng. Appl. Comput. Fluid Mech.* **2018**, *12*, 584–597. [[CrossRef](#)]
22. Liu, Z.; Zhou, P.; Zhang, Y. A probabilistic wavelet-support vector regression model for streamflow forecasting with rainfall and climate information input. *J. Hydrometeorol.* **2015**, *16*. [[CrossRef](#)]
23. Niu, M.; Wang, Y.; Sun, S.; Li, Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM2.5 concentration forecasting. *Atmos. Environ.* **2016**, *134*. [[CrossRef](#)]
24. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*. [[CrossRef](#)]
25. Dehghani, M.; Seifi, A.; Riahi-Madvar, H. Novel forecasting models for immediate-short-term to long-term influent flow prediction by combining ANFIS and grey wolf optimization. *J. Hydrol.* **2019**, *576*. [[CrossRef](#)]
26. Zolá, R.P.; Bengtsson, L. Long-term and extreme water level variations of the shallow Lake Poopó, Bolivia. *Hydrol. Sci. J.* **2006**, *51*. [[CrossRef](#)]
27. Hastenrath, S.; Kutzbach, J. Late Pleistocene climate and water budget of the south American Altiplano. *Quat. Res.* **1985**, *24*. [[CrossRef](#)]
28. Zolá, R.P.; Bengtsson, L.; Berndtsson, R.; Martí-Cardona, B.; Satgé, F.; Timouk, F.; Bonnet, M.P.; Mollericon, L.; Gamarra, C.; Pasapera, J. Modelling Lake Titicaca's daily and monthly evaporation. *Hydrol. Earth Syst. Sci.* **2019**, *23*. [[CrossRef](#)]
29. Choubin, B.; Malekian, A. Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. *Environ. Earth Sci.* **2017**, *76*. [[CrossRef](#)]
30. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*. [[CrossRef](#)]

31. Cutler, D.R.; Edwards, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for classification in ecology. *Ecology* **2007**, *88*. [[CrossRef](#)]
32. Díaz-Uriarte, R.; de Andrés, S.A. Gene selection and classification of microarray data using random forest. *BMC Bioinform.* **2006**, *7*. [[CrossRef](#)] [[PubMed](#)]
33. Kira, K.; Rendell, L.A. Feature selection problem: Traditional methods and a new algorithm. In Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992.
34. Faris, H.; Aljarah, I.; Al-Betar, M.A.; Mirjalili, S. Grey wolf optimizer: A review of recent variants and applications. *Neural Comput. Appl.* **2018**, *30*. [[CrossRef](#)]
35. Mirjalili, S.; Saremi, S.; Mirjalili, S.M.; Coelho, L.D.S. Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization. *Expert Syst. Appl.* **2016**, *47*. [[CrossRef](#)]
36. Mohammadi, B.; Ahmadi, F.; Mehdizadeh, S.; Guan, Y.; Pham, Q.B.; Linh, N.T.T.; Tri, D.Q. Developing novel robust models to improve the accuracy of daily streamflow modeling. *Water Resour. Manag.* **2020**, *34*. [[CrossRef](#)]
37. Mehdizadeh, S.; Mohammadi, B.; Pham, Q.B.; Khoi, D.N.; Linh, N.T.T. Implementing novel hybrid models to improve indirect measurement of the daily soil temperature: Elman neural network coupled with gravitational search algorithm and ant colony optimization. *Meas. J. Int. Meas. Confed.* **2020**, *165*. [[CrossRef](#)]
38. Mohammadi, B.; Aghashariatmadari, Z. Estimation of solar radiation using neighboring stations through hybrid support vector regression boosted by Krill Herd algorithm. *Arab. J. Geosci.* **2020**, *13*. [[CrossRef](#)]
39. Mohammadi, B.; Linh, N.T.T.; Pham, Q.B.; Ahmed, A.N.; Vojteková, J.; Guan, Y.; Abba, S.I.; El-Shafie, A. Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series. *Hydrol. Sci. J.* **2020**, *65*. [[CrossRef](#)]
40. Vaheddoost, B.; Guan, Y.; Mohammadi, B. Application of hybrid ANN-whale optimization model in evaluation of the field capacity and the permanent wilting point of the soils. *Environ. Sci. Pollut. Res.* **2020**, *27*. [[CrossRef](#)]
41. Moazenzadeh, R.; Mohammadi, B. Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. *Geoderma* **2019**, *353*. [[CrossRef](#)]
42. Li, B.; Yang, G.; Wan, R.; Dai, X.; Zhang, Y. Comparison of random forests and other statistical methods for the prediction of lake water level: A case study of the Poyang Lake in China. *Hydrol. Res.* **2016**, *47*. [[CrossRef](#)]
43. Tikhmarine, Y.; Souag-Gamane, D.; Kisi, O. A new intelligent method for monthly streamflow prediction: Hybrid wavelet support vector regression based on grey wolf optimizer (WSVR–GWO). *Arab. J. Geosci.* **2019**, *12*. [[CrossRef](#)]
44. Aghelpour, P.; Guan, Y.; Bahrami-Pichaghchi, H.; Mohammadi, B.; Kisi, O.; Zhang, D. Using the MODIS Sensor for Snow Cover Modeling and the Assessment of Drought Effects on Snow Cover in a Mountainous Area. *Remote Sens.* **2020**, *12*, 3437. [[CrossRef](#)]

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).