

Article

# Establishment of Landslide Groundwater Level Prediction Model Based on GA-SVM and Influencing Factor Analysis

Ying Cao <sup>1</sup>, Kunlong Yin <sup>1,\*</sup>, Chao Zhou <sup>2,\*</sup> and Bayes Ahmed <sup>3</sup>

<sup>1</sup> Faculty of Engineering, China University of Geosciences, Wuhan, 430074, China; yingcao@cug.edu.cn

<sup>2</sup> School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430078, China

<sup>3</sup> Institute for Risk and Disaster Reduction, University College London (UCL), Gower Street, London, WC1E 6BT, UK; bayesahmed@gmail.com

\* Correspondence: yinkl@cug.edu.cn(K.Y.); zhouchao@cug.edu.cn (C.Z.)

Received: 20 December 2019; Accepted: 03 February 2020; Published: 5 February 2020

**Abstract:** The monitoring and prediction of the landslide groundwater level is a crucial part of landslide early warning systems. In this study, Tangjiao landslide in the Three Gorges Reservoir area (TGRA) in China was taken as a case study. Three groundwater level monitoring sensors were installed in different locations of the landslide. The monitoring data indicated that the fluctuation of groundwater level is significantly consistent with rainfall and reservoir level in time, but there is a lag. In addition, there is a spatial difference in the impact of reservoir levels on the landslide groundwater level. The data of two monitoring locations were selected for establishing the prediction model of groundwater. Combined with the qualitative and quantitative analysis, the influencing factors were selected, respectively, to establish the hybrid Genetic Algorithm-Support Vector Machine (GA-SVM) prediction model. The single-factor GA-SVM without considering influencing factors and the backpropagation neural network (BPNN) model were adopted to make comparisons. The results showed that the multi-factor GA-SVM performed the best, followed by multi-factor BPNN and single-factor GA-SVM. We found that the prediction accuracy can be improved by considering the influencing factor. The proposed GA-SVM model combines the advantages of each algorithm; it can effectively construct the response relationship between groundwater level fluctuations and influencing factors. Above all, the multi-factor GA-SVM is an effective method for the prediction of landslides groundwater in the TGRA.

**Keywords:** landslides monitoring; groundwater level prediction; Support Vector Machine; influencing factors; Three Gorges Reservoir area

---

## 1. Introduction

The groundwater, with storage and migration behaviors, is one of the main natural factors affecting landslide stability. It impacts the various stages of landslide development [1,2]. According to the statistics, 30–40% of dam failures are caused by the damage of groundwater seepage [3], and more than 55% of soil landslides are caused by the effect of groundwater in China. The monitoring and prediction of the landslide groundwater level is a crucial part of landslide early warning systems [4].

The groundwater level in a landslide is mainly affected by external input factors (rainfall, reservoir level, irrigation, etc.) and the permeability of the sliding body (looseness of the soil, development of earth cracks, etc.). For example, for landslides with loose soil and well-developed cracks, the groundwater level clearly rises after rainfall infiltration, which becomes an unfavorable factor for landslide stability [5]. There are mainly three aspects of groundwater that affect the stability of landslides [6,7], including physical and chemical functions of rock and soil, the mechanical

function of pore water pressure, and hydrodynamic pressure conditions of landslides. Trigo et al. proposed that landslides of variable movement types are triggered by groundwater level rising and shear strength reduction [8]. Ray and Jacobs and Tommasi et al. also considered that most of the slope failures are caused by soil moisture or groundwater because they increase pore water pressure and decrease shear strength [9–11]. It can be seen that the effect of groundwater on landslides is reflected by the change of landslide stability, which is caused by the change of the water level [12]. Hence, real-time process control of groundwater levels can provide an effective judgment on the stability of landslides.

For the landslides in the Three Gorges Reservoir area (TGRA) in China, under the conditions of water-impoundment of the reservoir, the slopes are recharged directly by groundwater or surface water chronically with strong alternating and cycling actions [13]. The dynamic variation process of the landslide groundwater level is a complex power system. It is controlled by the hydrogeological conditions and the combined action of internal and external factors, such as precipitation and the fluctuation of reservoir water level [14]. However, it is difficult to comprehensively analyze the overall stability of landslides through an accurate analytical method of the measured pore water pressure [15,16]. In general, the change of groundwater level is composed of the deterministic and random components, and the components are influenced by both deterministic and non-deterministic factors. The action rules of these factors will be reflected in the measured data as long as the observation period of the groundwater level is long enough. It is of great significance for landslide forecasting to study the variation characteristics of groundwater under variable conditions of rainfall and reservoir water level, and make a real-time prediction of groundwater level in a potential susceptible landslide.

In recent years, numerical simulation techniques have been utilized to study the groundwater level change with external factors, including rainfall [17], reservoir impoundment [18], artificial infiltration, drainage, and so on. However, the simulation technique is always complicated, and it requires accurate soil parameters, which are difficult to obtain. Recently, with the development of machine learning (ML) technique, numerous ML models, including hybrid, ensemble, deep learning, etc., have been widely utilized in geology and environment studies [19–21]. Choubin et al. applied simulated annealing feature selection to identify key features, and five ML models were used to predict the earth's fissure hazard [22]. Shamshirband et al. proposed the ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll-a concentration in coastal waters [23]. Lian et al. established a landslide displacement model based on a modified ensemble empirical mode decomposition and extreme learning machine (ELM) [20]. Nguyen et al. presented three novel hybrid ML models for landslide susceptibility modeling, all of which achieved good performance in the area of Vietnam [21]. The advanced ML models have been applied in land subsidence hazard prediction, landslide susceptibility mapping, landslide failure time prediction, and the other fields [22, 24–26]. However, far too little attention has been paid to the application of ML for the prediction of landslide groundwater level fluctuation. Support Vector Machine (SVM) model was established based on statistical learning theory, which replaced the empirical risk minimization principle of traditional ML methods with the structural risk minimization principle [27]. The SVM model shows a perfect generalization ability to overcome the deficiencies of the traditional artificial neural networks [28,29]. The developed studies present that SVM has achieved excellent performance in both accuracy and stability [25, 30, 31].

A hybrid model can integrate the advantages of each model. Nowadays, using hybrid models has become popular in modeling and estimating studies. These models combine more than one classifier or use one for estimating and the other for parameter optimization. Based on the previous literature, the hybrid model performs better than the single model [32]. SVM advances a lot in dealing with the problem of nonlinear regression prediction. However, its generalization performance is sensitive to the selection of the parameters [33–36]. The difficulties in capturing the critical modeling variables are noted as the major drawbacks to SVM [22]. Hence, it is urgent to apply optimization algorithms to search for the optimal parameters of SVM. The Genetic Algorithms (GA) is an optimization algorithm; it simulates the mechanism of genetic variation and the theory of biological

evolution. GA has the advantages of parallelism and global optimization. It has achieved excellent optimization results in plenty of studies [37–39]. In the current study, GA was utilized to select and optimize the parameters of SVM.

Three sensors were installed in Tangjiao landslide to monitor the fluctuation of the groundwater level. Based on the in-suit monitoring data, this study analyzed the variation characteristics of the groundwater level and its relationship with reservoir level and rainfall. Two monitoring positions were selected for establishing a prediction model. The proposed multi-factor GA-SVM, backpropagation neural network (BPNN) model, and the single-factor GA-SVM without considering the factors were utilized to predict the groundwater level of Tangjiao landslide. A comprehensive comparison and assessment of these models will be presented in this study.

## 2. The Forecast Model of Groundwater Level

### 2.1. Proposed Prediction Model

In the reservoir area, the change of groundwater level in a landslide is associated with geological conditions of the landslide itself, the scheduling of reservoir level, and the rainfall. According to the data of reservoir scheduling, rainfall, and the monitoring data of groundwater level in the Tangjiao landslide, the time-series expression is established as follows:

$$g_i = f(x_i, x_{i+1}, \dots, x_{i+n}, y_i, y_{i+1}, \dots, y_{i+n}) (i = 1, 2, \dots, n) \quad (1)$$

Where  $f$  is the expression of the SVM,  $g_i$  is groundwater level,  $x_i$  is reservoir level, and  $y_i$  is rainfall. Reservoir level and rainfall are factors independent of each other.

### 2.2. Introduction of the GA-SVM Model

The SVM model is a nonlinear regression forecast method proposed by Vapnik et al. (1995). The input variables are mapped to a high-dimensional linear feature space by the nonlinear transformation. The optimal decision function is constructed. Then the dot product operation of high-dimensional feature space is replaced by the kernel function of the original space. The optimal global solution is obtained by learning and training of finite samples [40]. The regression function is as follows:

$$f(x) = \langle W \cdot \Phi(x) \rangle + b \quad (2)$$

The estimation function is transformed into a function minimization problem by the insensitive loss function  $\varepsilon$ :

$$R_{\min} = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) \quad (3)$$

The constraints are as follows:

$$\begin{cases} W^T \phi(x_i) + b_i - y_i \leq \varepsilon + \xi_i; \\ y_i - W^T \phi(x_i) - b_i \leq \varepsilon + \xi_i^*; \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l. \end{cases} \quad (4)$$

Where,  $C$  is the penalty factor,  $\xi$  and  $\xi_i^*$  are relaxation factors,  $b$  is the offset value. Finally, by introducing Lagrange multipliers and the application of Wolf duality theory, it is translated into an equivalent dual problem as follows:

$$\min \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \quad (5)$$

The constraints are as follows:

$$\begin{cases} \sum_{i=1}^l (\alpha_i + \alpha_i^*) = 0; \\ 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, l. \end{cases} \quad (6)$$

Where,  $Q_{ij} = K(x_i, x_j) = f(x_i)^T f(x_j)$

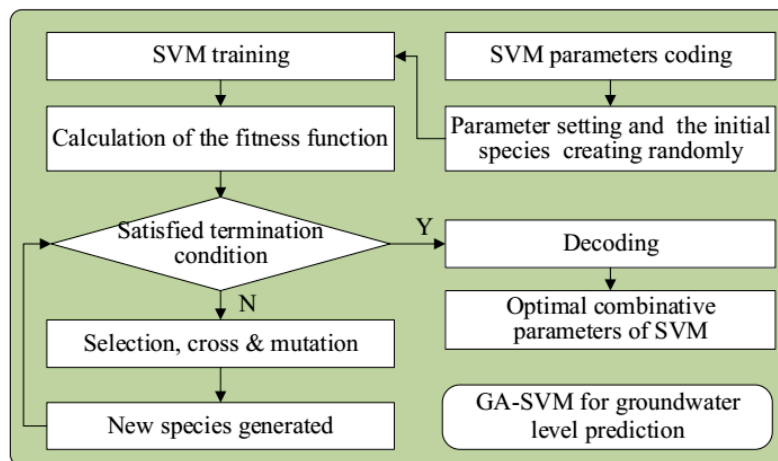
By quadratic programming, the regression forecast model is obtained as follows:

$$f(x, \alpha_i^*, \alpha_i) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (7)$$

Where,  $K(x_i, x)$  is the kernel function of the SVM. At present, there are four commonly used kernel functions—linear kernel function, polynomial kernel function, radial basis kernel function (RBF), and sigmoid function.

The genetic algorithm is an artificial intelligence algorithm. It searches the optimal solution by simulating the process of natural evolution. This algorithm simulates phenomena that occurred in natural selections and genetics, such as reproduction, crossover, and mutation. Starting from an initial population, through the operation of selecting, crossing and mutating, a group of better-adapted individuals is produced. These individuals make the group evolve into a better area in the search space. Then through the constant multiplications and evolutions, the optimal solution can be obtained by converging to the individuals, which can best adapt to the environment.

Considering that the SVM is sensitive to the model parameters, GA was adopted to optimize the parameters, and the GA-SVM coupling model was established. The flowchart is depicted in Figure 1.



**Figure 1.** Analysis flowchart of the GA-SVM prediction model.

### 2.3. Introduction of the BPNN Model

BPNN is a multilayer feedforward neural network. It has the characteristics of error reverse transmission and signal forward transmission. BPNN is learned by example, which consists of some input examples and the known correct output for each case. In the first step of the BPNN process [41], the learning samples are added to the input network. Then, the back-propagation algorithm is used to train the weights and deviations in the network. It more closely matches the output vectors and expectation vectors. If there is a big error between the output result and the expected result, the error propagates backward, adjusts the network threshold and weight, and is repeated to eventually make the predicted output result approach the expected output result. The topological structure of BPNN is shown in Figure 2. Previous research has shown that the basic architecture of a three-layered neural

network consists of an input layer, a hidden layer, and an output layer. It can accurately fit nonlinear mapping relationships [42]. Additionally, the trial-and-error approach is used to choose an appropriate number of hidden neurons in the BPNN model.

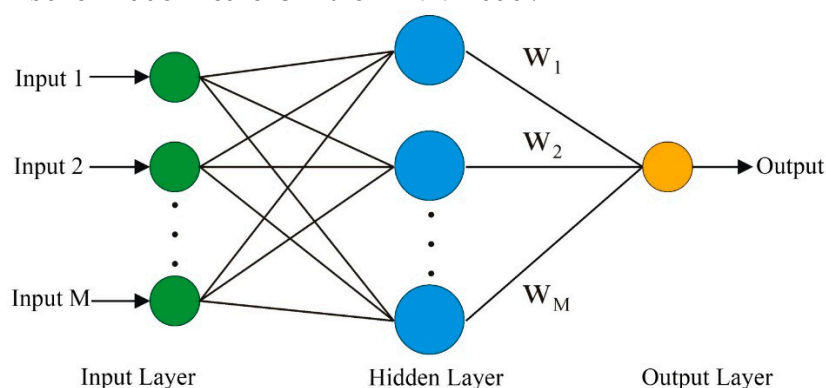


Figure 2. The basic structure of BPNN.

#### 2.4. Evaluation Model of the Prediction Accuracy

Root mean square error (*RMSE*, Equation (8)) and the mean absolute percentage error (*MAPE*, Equation (9)) are commonly used indexes for accuracy testing of prediction models. The smaller the values of *RMSE* and *MAPE* are, the better the prediction result is. However, the values of *RMSE* and *MAPE* are associated with the value of the motoring data. The correlation coefficient (Equation (10)) is not related to the monitoring data; it can reflect the correlation between the variables. This paper has adopted the correlation coefficient for the forecast accuracy analysis [43].

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (8)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad (9)$$

$$R = \frac{\sum_{i=1}^N (x_i - \bar{x})(\hat{x}_i - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (\hat{x}_i - \bar{\hat{x}})^2}} \quad (10)$$

Where,  $x_i$  is the measured value;  $\hat{x}_i$  is the predicted value;  $N$  is the number of predicted values;  $\bar{x}$  is the mean of the measured values;  $\bar{\hat{x}}$  is the mean of predicted values.

### 3. Sensors Installation and Data Acquisition

#### 3.1. Geologic Conditions of the Monitored Landslide—the Tangjiao Landslide

The Tangjiao landslide is located in the Wanzhou district of Chongqing City, the east edge of the Sichuan basin and the south side of the Yangtze River (Figure 3). Once the landslide fails, it will endanger the lives and property of 709 people in the landslide area. It will also threaten the safety of the road and shipping of the Yangtze River.

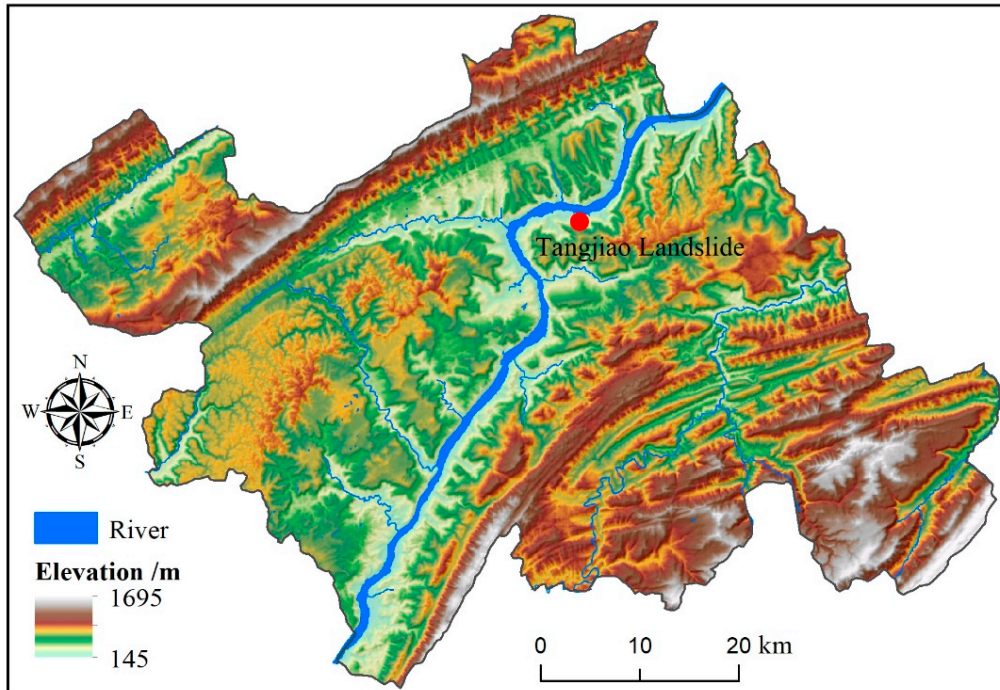


Figure 3. Location map of the Tangjiao landslide in the Wanzhou district.

The Tangjiao landslide location belongs to the bank accumulation landforms of the Yangtze River valley. In the front of the landslide, the terraces and flood plains develop well with gentle step surfaces. They bulge out like the tongue and show arc shapes in contours. The slope inclines with radial shape. The middle and rear part of the landslide is gentler than the front. The back edge of the landslide is concave southwards with cirque shape. the slope gradient is 25°–45° with many bedrocks exposed. The landslide is an oversized soil deformation. It is fan-shaped on the whole platform, with the main sliding direction of 359°. It has an estimated volume of  $2672.4 \times 10^4 \text{ m}^3$ , with an average thickness of 20 m. The entire landslide covers an area of  $133.62 \times 10^4 \text{ m}^2$ , with a maximum longitudinal dimension of 1020 m and an average width of 1310 m (Figure 4).

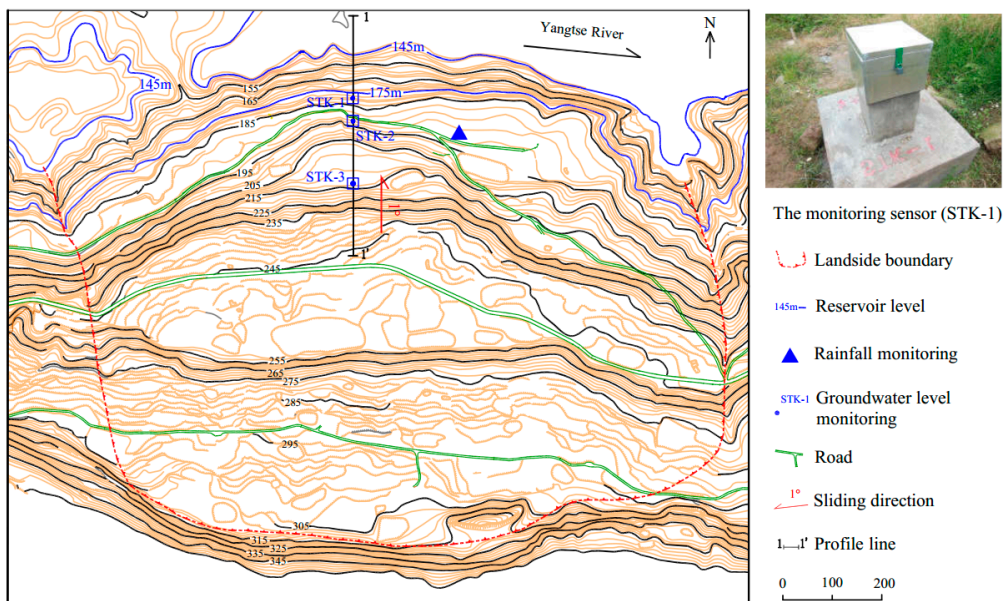


Figure 4. Topographical map of the Tangjiao landslide, with monitoring points.

The landslide is located in the core area of Wanxian syncline. Its axis trend is NE30°-60°. The attitude of the core rock layer is 153°∠4° with two asymmetrical wings, the steep northwest one (40°–80°) and the gentle southeast one (15°–40°). The landslide is locally covered by thick gravelly soil accumulated from quaternary collapse-slide and eluvial-colluvial deposits. The underlying bedrock is interbedded with sandstone and mudstone of the Shaximiao Formation of Middle Jurassic, the strata attitude of which is 170°∠5° (Figure 5, 6).

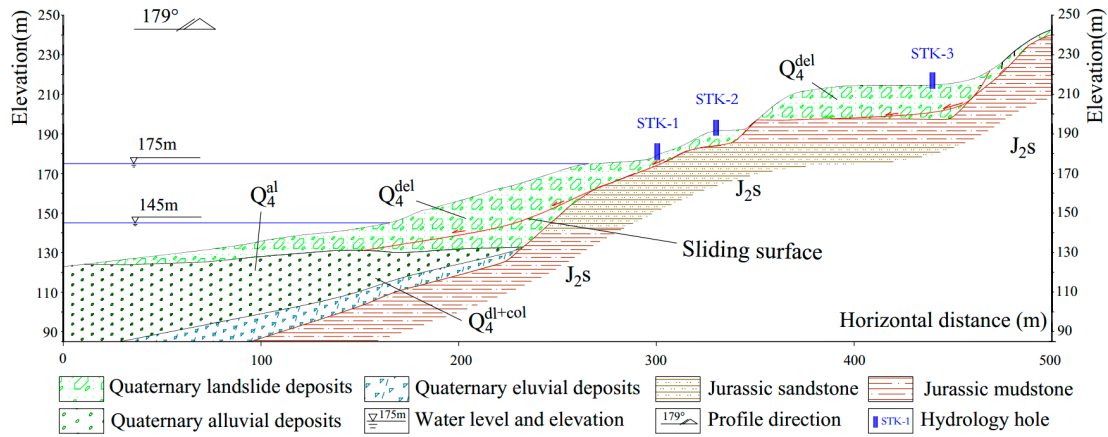


Figure 5. Schematic I-I' geological cross-section of Tangjiao landslide.

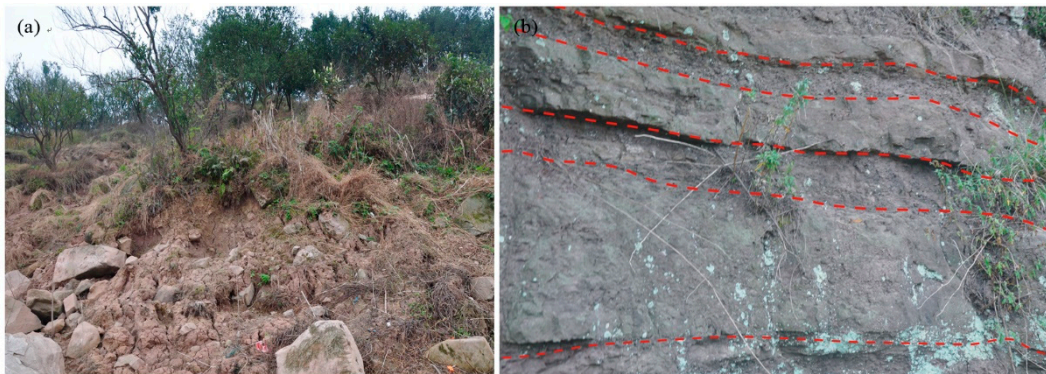


Figure 6. (a) Deposits with rubble and block stones; (b) Bedrock interbedded with sandstones and mudstones.

### 3.2. In-suit Monitoring and Data Acquisition

The principle and in-suit installation of groundwater level monitoring device are shown in Figure 7.

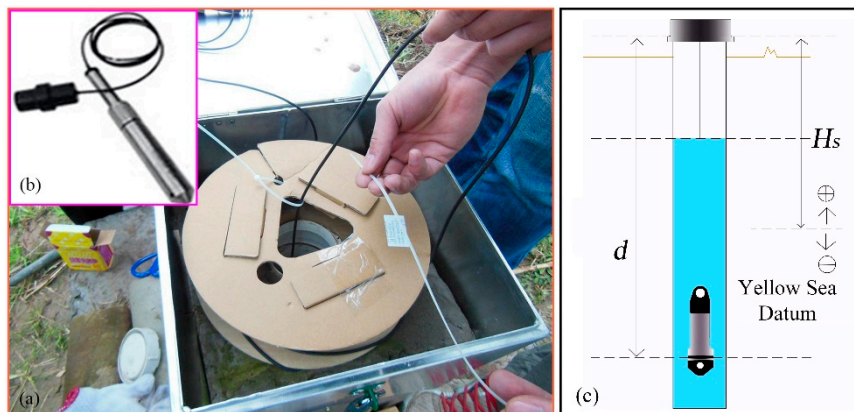


Figure 7. The principle and installation of groundwater level monitoring device (a) Sensor installation (b) a sensor for monitoring and (c) the principle of monitoring.

The formula of the groundwater level is shown below:

$$H = h/100 - h_s + H_s - d \tag{11}$$

Where,  $H$  is the groundwater level;  $h$  is the device records(cm);  $h_s$  is the height of water column at standard atmospheric pressure,  $h_s = 10.336$  m;  $H_s$  the elevation of the hydrologic hole(m);  $d$  is the distance between monitoring probe and hydrologic hole, in this paper,  $d = 5$  m.

Three groundwater monitoring points (STK) were installed on the Tangjiao landslide. They were respectively labelled as STK-1, STK-2, and STK-3, and their positions are shown on Figure 4, 5, and 8.

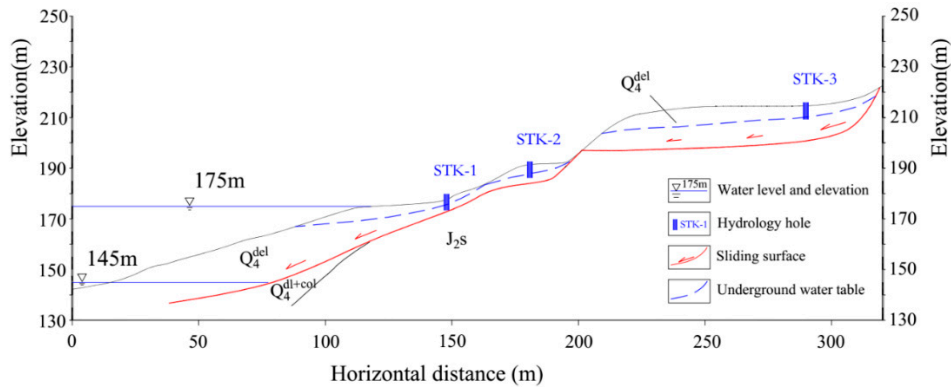


Figure 8. Generalized schematic section of groundwater in the toe of the landslide

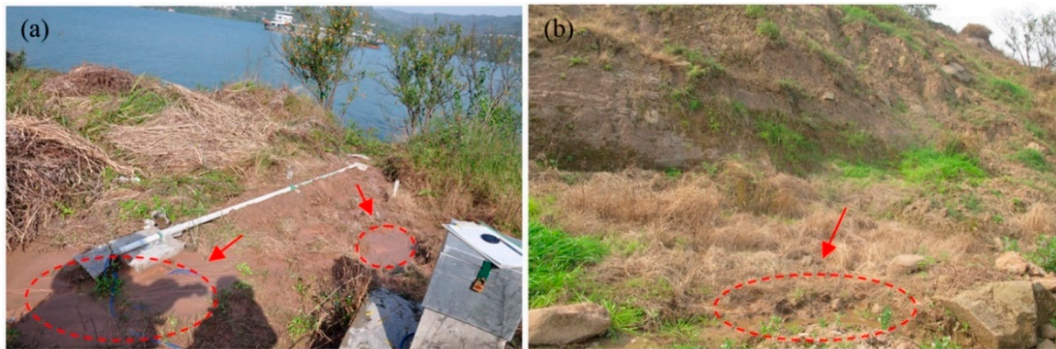


Figure 9. (a) Water exuded on the ground near STK-1; (b) Water exuded on the ground near STK-2

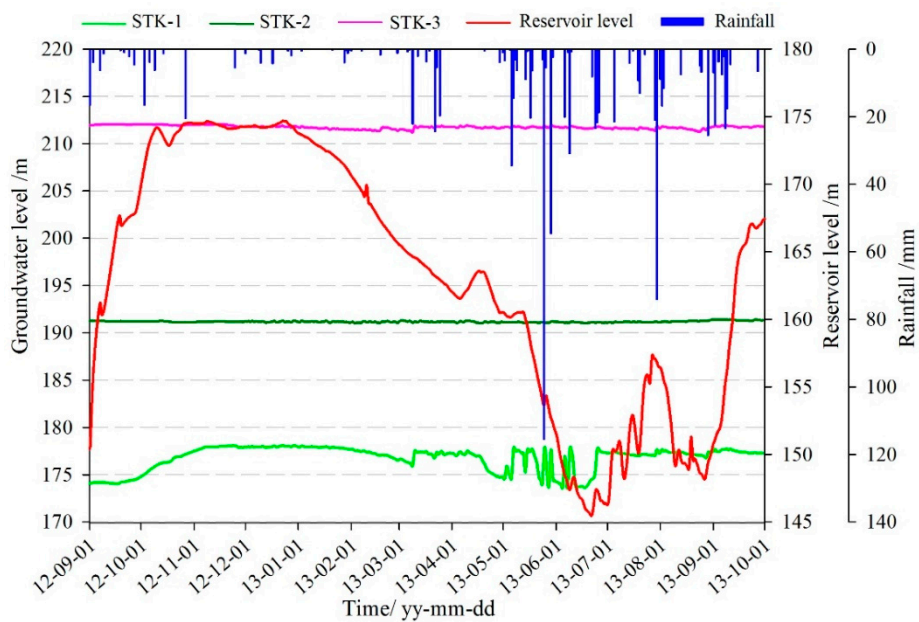


Figure 10. Monitoring curves of groundwater level of the Tangjiao landslide.



The three monitoring points of groundwater level were all installed on the terrace III. The groundwater level of STK-1 varied obviously along with the fluctuation of the reservoir level, while the STK-2 and the STK-3 almost remained unchanged (Figure 8). STK-1 is located on the toe of the front part with an elevation of 179 m, and it is close to the river; therefore, the groundwater can be directly recharged or discharged by the reservoir water. The elevations of the STK-2 and STK-3 are 192 m and 213 m, respectively. The STK-3 is located on the second platform of the junction of terrace III and IV. The maximum height of the reservoir level is about 175 m, which shows less impact on the groundwater of the two positions. From field investigations, it was found that during the high reservoir level season, groundwater is often exposed to small scarps (Figure 9a); and similar results were found for the STK-2 in rainy seasons (Figure 9b). Consequently, it is reasonable to infer that the groundwater in the toe of the Tangjiao landslide consists of different hydrologic systems. The groundwater of STK-1 and STK-2 are influenced both by the reservoir level and rainfall. The reservoir level is the most obvious factor for STK-1, while STK-3 is mainly influenced by rainfall. The groundwater level of STK-1 and STK-3 were used to establish the prediction model in this paper (Figure 10).

#### 4. The Groundwater Level Prediction of STK-1

##### 4.1. Monitoring Data of Groundwater Level

The water-impoundment of the Three Gorges Reservoir has reached full capacity since 2007. The fluctuation of the reservoir level ranges from 145 m to 177 m with a variation of 30 m. The scheduling process is divided into four stages (Figure 11). It makes the hydrogeological conditions of the reservoir area in Wanzhou represent different characteristics in different periods during the year [44].

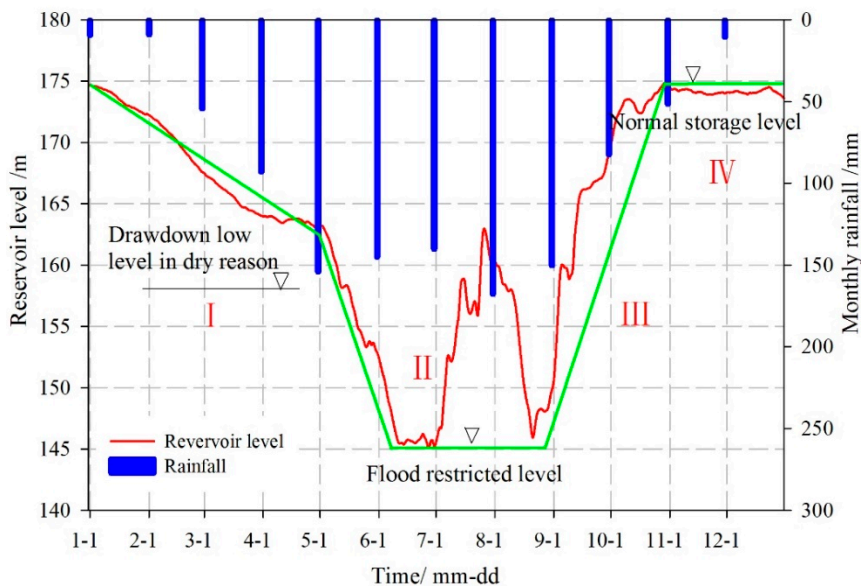
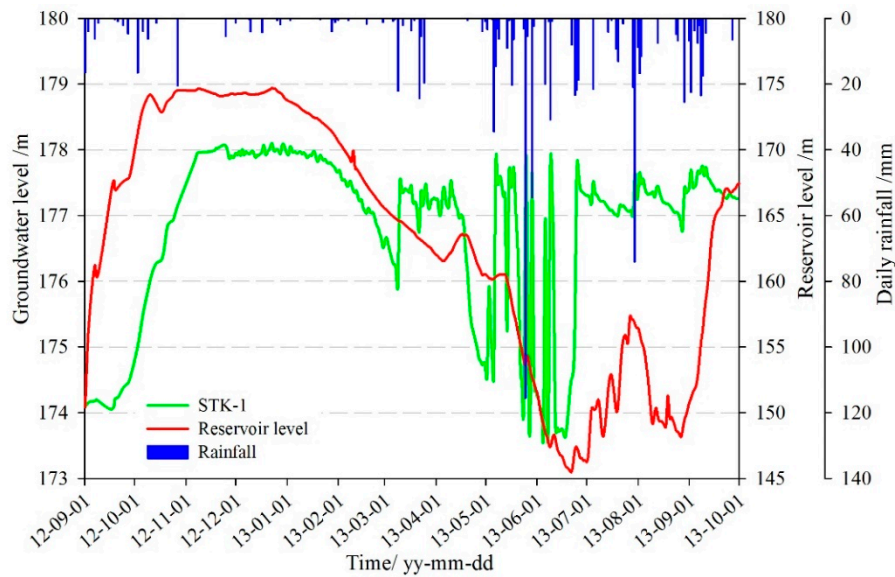


Figure 11. Water level regulation of the Three Gorges Reservoir in 2012 [36].



**Figure 12.** Monitoring curves of groundwater level in STK-1.

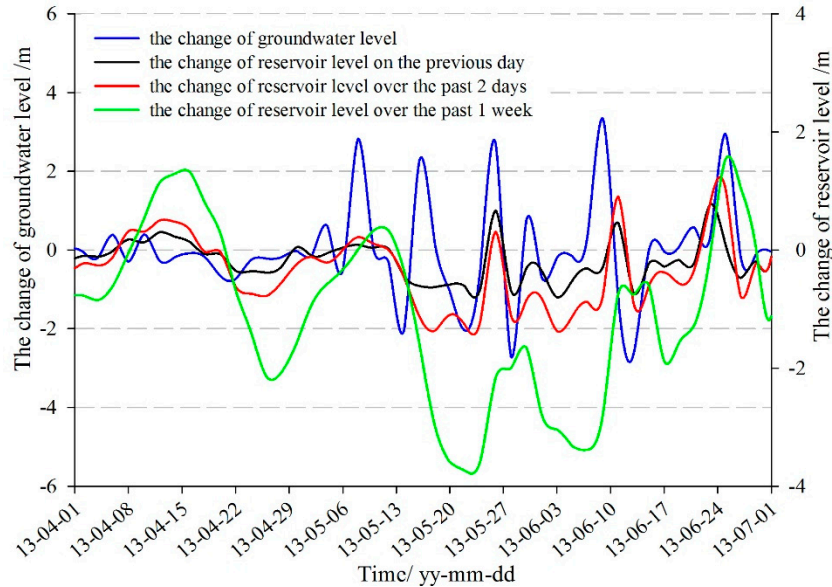
The groundwater level with time was analyzed and summarized below (Figure 12):

- (1) Between September 2012 and May 2013, the fluctuation of the reservoir level maintained a high degree of consistency with the groundwater level of hydrological hole STK-1. From May to July 2013, the reservoir level was continuously decreasing from 160 m to 145 m, while the groundwater level clearly fluctuated under the influence of rainfall. The reason is that the front part of the landslide is covered by loose deposits, the high permeability of the soil, and the penetrative pressure induced by short-time intensive rainfall make the groundwater recharge quickly by the rainfall. So, the continuous rise of the groundwater level was affected by the rainfall more than the reservoir level during this period.
- (2) On May 25<sup>th</sup>, 2013, the rainfall reached 115 mm/d, and the groundwater level rose from 173.5 m to 177.5 m which increased 4 m within 24 hours. Meanwhile, the reservoir level fell at a speed of 1.1 m/d. Three days after the rain stopped, the groundwater level dropped to 174 m again, and then rose rapidly when the rainfall reached 54.6 mm/d on May 29<sup>th</sup>.
- (3) Between September and November 2013, the reservoir level increased into the normal storage level of 175 m, but the groundwater level showed a slight downward trend.

The characteristics of the groundwater level in STK-1 indicated that the groundwater maintained close hydraulic connection and significant response relationship with the reservoir level and rainfall. They were the main factors of the changes in the groundwater level. Due to the geological conditions of the landslide, there was a hysteresis effect when the groundwater level was changing with the variation of rainfall and reservoir level, but the trends were in agreement substantially. When the reservoir level rose, and the rainfall increased, the groundwater would be recharged to a corresponding uplift. On the contrary, when the reservoir level dropped, and the rainfall decreased, the groundwater level would decline.

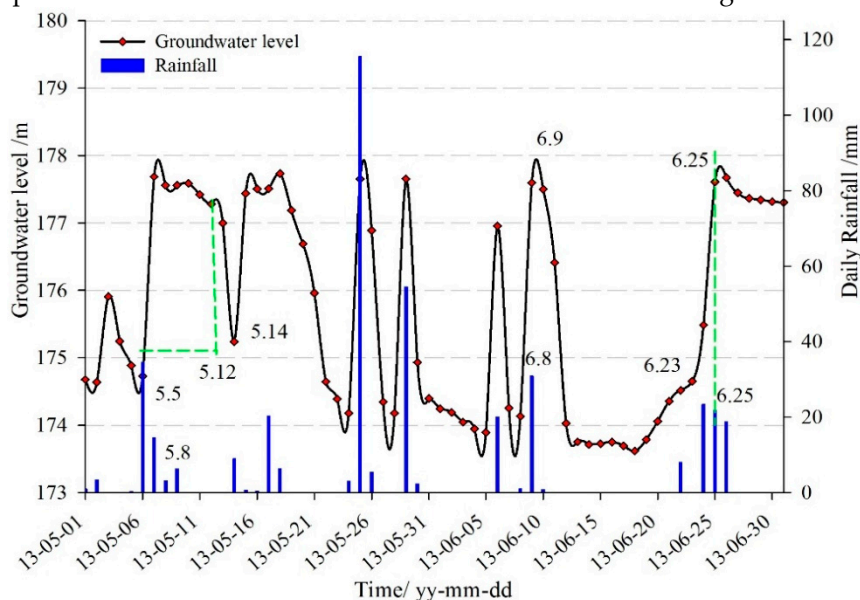
#### 4.2. Determination of Influencing Factors

The variable speed of the groundwater level is related to the fluctuation speed of reservoir level and the permeability of the soil in front of the landslide. The greater the speed of groundwater level declines, the faster the decrease of the seepage line in a landslide will be. When the reservoir level drops at a certain speed and the rains last in a certain duration, the position of the seepage line would depend on the rainfall intensity. With the changes in the external conditions and the soil permeability, the response of the groundwater level shows hysteresis effects in various degrees.



**Figure 13.** The relation between the groundwater level and the reservoir level.

The hydrology hole STK-1 is located on the front part of the Tangjiao landslide at a lower altitude. The soil is loose with large permeability coefficient, so the lag-time, the groundwater response to the reservoir water, is shorter than other parts of the landslide. The monitoring data in Figure 12 and Figure 13 showed that the shape of the change rate curves of the groundwater level in the Tangjiao landslide was similar to that of the change of the reservoir level on the previous day, the change of the reservoir level over the past two days, and the change of the reservoir level over the past one week. Hence, the reservoir level on the current day, the change of the reservoir level on the previous day, the change of the reservoir level over the past two days, and the change of the reservoir level over the past one week were selected as the reservoir level factors of groundwater level.



**Figure 14.** The relationship between rainfall and groundwater level from May to July

Figure 14 showed that it was at a lower level of the reservoir from May to July, and the rainfall period as well. So, besides the fluctuation influence of the reservoir level, the groundwater level was also influenced by the rainfall. The groundwater increased within 1–2 days when heavy rainfall occurred, which was seen from June 8<sup>th</sup>–9<sup>th</sup>. The relationship between the groundwater level and the rainfall during May 5<sup>th</sup>–8<sup>th</sup> and June 23<sup>rd</sup>–25<sup>th</sup> showed that, when the continuous heavy rainfall occurred, the groundwater rose rapidly and was maintained at a high level, about one week after the

rainfall stopped. Then because of the geotechnical characteristics of the landslide, surface evaporation, and flow discharge, the groundwater returned to the initial state rapidly within a short period. The continuous heavy rainfall was the main reason for the rise of the groundwater level.

The degree of grey correlation is applied to make a dynamic analysis on the degree of similarity or dissimilarity of the development trend of multi-variable nonlinear time series. It is a quantitative index for measuring the correlation. As the relationship between the groundwater level and its factors was dynamically changing with time, the calculation of the absolute difference and the degree of grey correlation were performed through the normalized processing. The values of grey correlation are shown in Table 1.

**Table 1.** Calculation results of grey correlation between groundwater level and factors of STK-1

Influencing factor	Degree of grey correlation
The reservoir level on the current day	0.988
The change of the reservoir level on the previous day	0.827
The change of the reservoir level over the past two days	0.827
The change of the reservoir level over the past one week	0.828
The rainfall on the current day	0.854
The cumulative rainfall on the previous day	0.854
The cumulative rainfall over the past two days	0.859
The cumulative rainfall over the past one week	0.879

The results showed that the degree of grey correlation between the influencing factors and the groundwater level were all greater than 0.800. It indicated a great consistency of the changing tendency. The influence of the reservoir level was stronger, which was in a valid agreement with the monitoring data.

This paper selected the reservoir level on current day, the change of the reservoir level on the previous day, the change of the reservoir level over the past two days, the change of the reservoir level over the past one week, the rainfall on current day, the cumulative rainfall on the previous day, the cumulative rainfall over the past two days and the cumulative rainfall over the past one week as the influencing factors to establish the response model between the factors and the groundwater level and to predict the groundwater level in the STK-1.

#### 4.3. The Result Analysis of the Prediction Model

The monitoring data from February 15<sup>th</sup> to June 7<sup>th</sup>, 2013 were selected as the training sample of the time-series for the GA-SVM prediction model. The data from June 9<sup>th</sup> to 27<sup>th</sup>, 2013 were selected as the test sample, which was within the period when the groundwater level fluctuated the most.

In order to eliminate the influence of the dimension, a normalization processing was applied to all the data, and the normalization formula is shown as follows:

$$x = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (12)$$

Where,  $x$  are the normalized values,  $X$  are the original values,  $X_{max}$  is the maximum value of a sequence, and  $X_{min}$  is the minimum value of a sequence.

The population of the GA was set as  $pop = 30$  and the maximum iterative steps number was 100. The optimal parameters of the penalty factor  $C$  and the radial basis kernel function (RBF)  $\gamma$  were searched by the GA (Figure 15), which was  $C=6.9709$  and  $\gamma=0.58537$ , respectively. Using the optimal parameters, the multi-factor SVM model was established to find the response relationship between the groundwater level and the influencing factors. The model was trained by the training sample, then the trained model was used to make a prediction of the groundwater level.

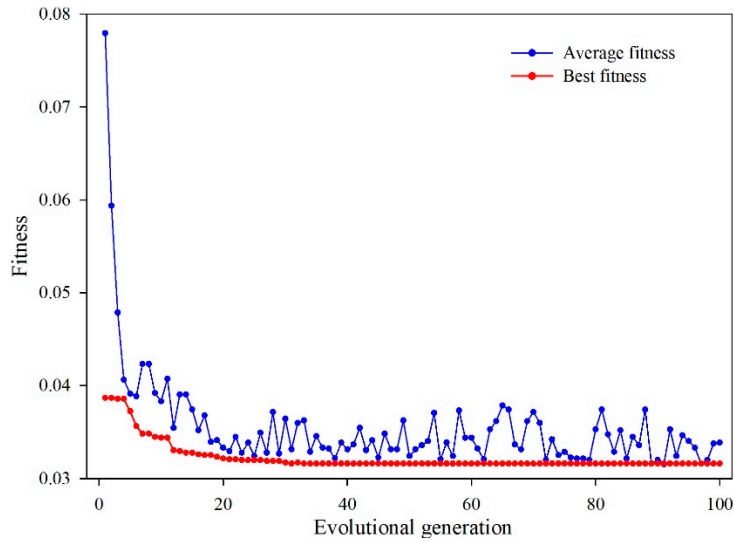


Figure 15. The parameters searching process of GA.

In addition, the single-factor SVM prediction model without considering the factors and the BPNN model were adopted to make a comparative analysis with the proposed multi-factor model. The comparison curves between the predicted and measured groundwater levels are shown in Figure 16, 17. The evaluation results of the prediction accuracy are shown in Table 2.

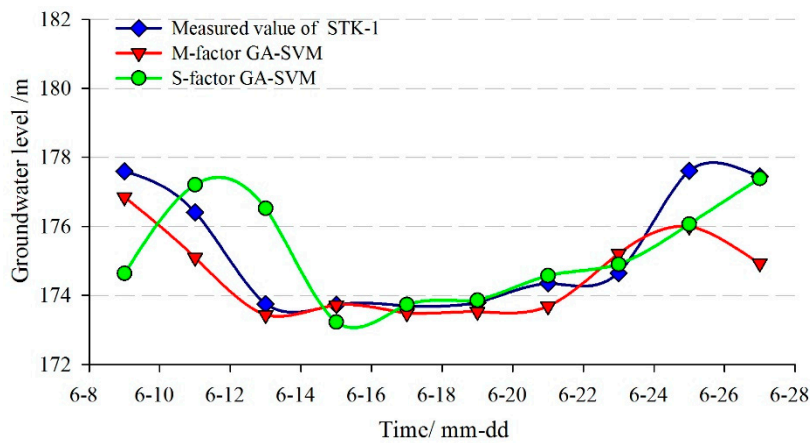


Figure 16. The predicted result of groundwater level in STK-1 by the GA-SVM model.

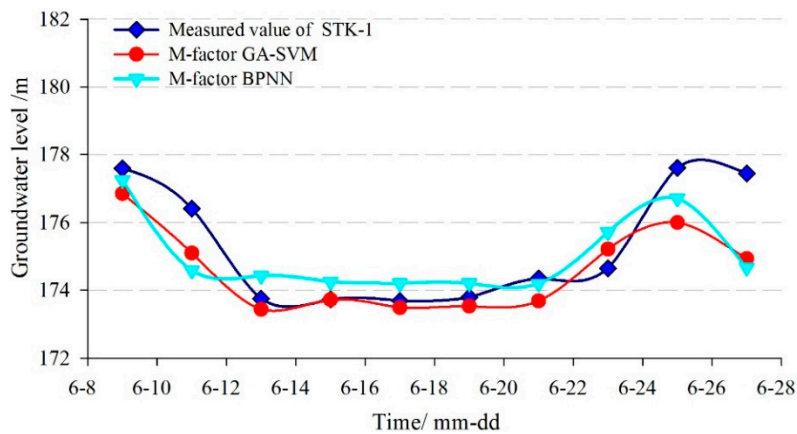


Figure 17. The predicted result of groundwater level in STK-1 by the BPNN model.

**Table 2.** Evaluation results of the prediction accuracy of STK-1

Prediction model		Multi-factor GA-SVM	Single-factor GA-SVM	Multi-factor BPNN
Accuracy	<i>RMSE/m</i>	1.104	1.409	1.195
	<i>MAPE/%</i>	0.465	0.525	0.522
	<i>R</i>	0.881	0.591	0.718

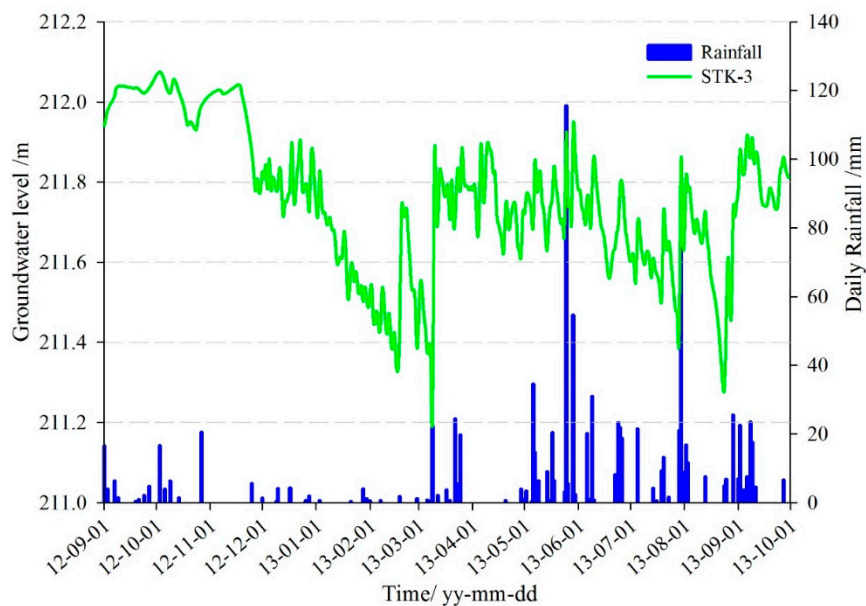
Figure 16 and Figure 17 showed that the predicted results of the three models were all consistent with the trend of measured values, but the prediction error at the turning points was relatively larger using the single-factor model.

Table 2 showed that the *RMSE* and the *MAPE* of the multi-factor GA-SVM model were 1.104 m and 0.465%, while the values of the multi-factor BPNN model were 1.195 m and 0.522%, respectively. The correlation coefficient (*R*) of the former was 0.881, which was greater than the latter (0.718). So, the multi-factor GA-SVM model outperforms the multi-factor BPNN model. The *RMSE*, *MAPE*, and *R* of the single-factor GA-SVM model were 1.409 m and 0.525%, 0.591, respectively. The prediction accuracies of the multi-factor models were much greater than the single-factor model, so the consideration of the influencing factors in the multi-factor model can make the results superior to that of the single-factor model. The main reasons are, on the one hand, the groundwater level of the landslide is under influencing factors such as reservoir level and rainfall. On the other hand, because of the complex geological conditions of the Tangjiao landslide, there is a complicated nonlinear relationship between the groundwater level and its influencing factors. In conclusion, the multi-factor prediction model considering the factors that showed a higher prediction accuracy.

## 5. The Groundwater Level Prediction of STK-3

### 5.1. Monitoring Data of Groundwater Level

As can be seen from the monitoring curves (Figure 18), the rainfall reached 115 mm/d on May 25<sup>th</sup>, 2013, the groundwater level rose from 211.6 m to 211.9 m, which increased 0.3 m within 24 hours. Three days after the rain stopped, the groundwater dropped to 211.7 m again. Then it rose rapidly when the rainfall reached 54.6 mm/d on May 29<sup>th</sup>. Although the groundwater level of STK-3 fluctuating within a narrow range, it was because of high sensitivity to rainfall.

**Figure 18.** Monitoring curves of groundwater level in STK-3

### 5.2. Determination of Influencing Factors

The changes of groundwater level of STK-3 was mainly caused by rainfall. It was hardly influenced by the reservoir level. So, the rainfall was selected as the factor for the groundwater level prediction of STK-3. The degree of grey correlation between the groundwater level and the factors are shown in Table 3.

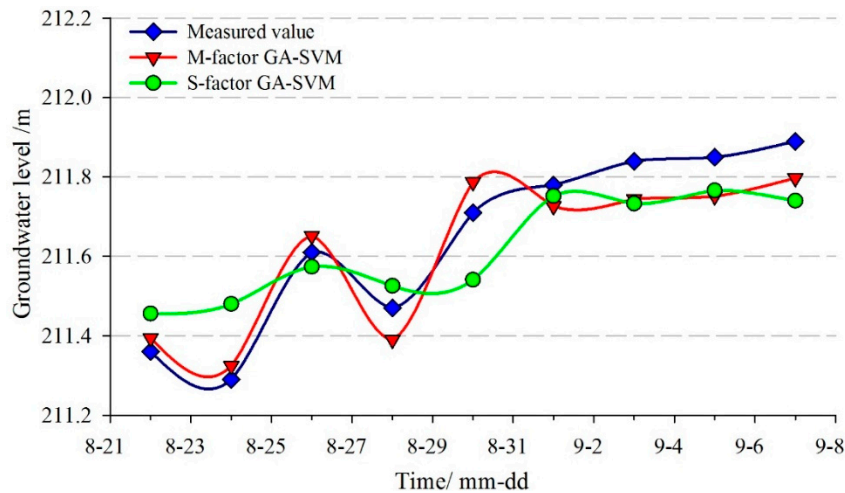
**Table 3.** Calculation results of grey correlation between groundwater level and factors of STK-3

Influencing factor	Degree of grey correlation
The rainfall on the current day	0.856
the cumulative rainfall on the previous day	0.856
the cumulative rainfall over the past two days	0.860
the cumulative rainfall over the past one week	0.880

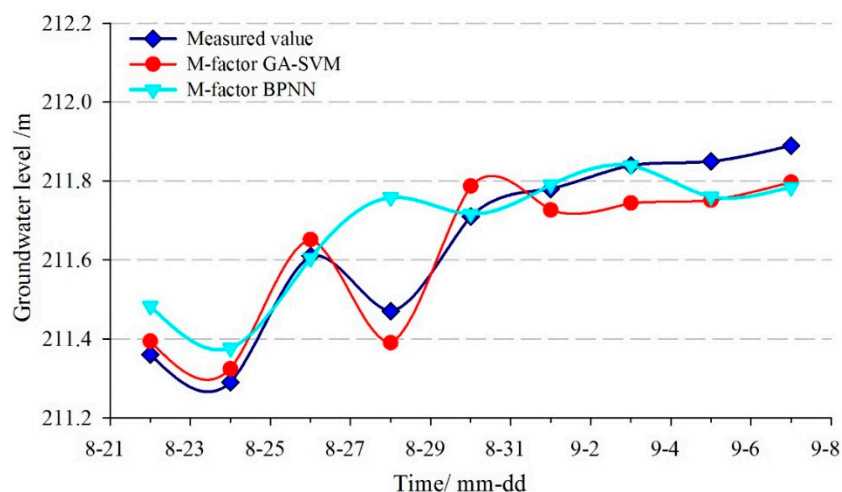
The results show that the groundwater level kept a great consistency of changing tendency with the influencing factors, which was in a valid agreement with the monitoring data. This paper selected the rainfall on the current day, the cumulative rainfall on the previous day, the cumulative rainfall over the past two days, and the cumulative rainfall over the past one week as the influencing factors to predict the groundwater level of STK-3.

### 5.3. The Result Analysis of the Prediction Model

The monitoring data from January 20<sup>th</sup> to August 20<sup>th</sup>, 2013, were selected as the training sample of the time-series prediction model for STK-3. The data from August 22<sup>nd</sup> to September 7<sup>th</sup>, 2013, were selected as the test sample. The population of the Genetic Algorithm was set as  $pop = 35$  and the maximum iterative steps number was 35. The optimal parameters of the penalty factor  $C$  and the radial basis kernel function  $\gamma$  were searched by Genetic Algorithm, which was  $C = 4.8183$  and  $\gamma = 30.0069$ , respectively. The comparison curves between the predicted and measured groundwater levels are shown in Figure 19 and Figure 20, and the evaluation results of the prediction accuracy are shown in Table 4.



**Figure 19.** The predicted result of groundwater level in STK-3 by GA-SVM model.



**Figure 20.** The predicted result of groundwater level in STK-3 by BPNN model.

**Table 4.** Evaluation results of the prediction accuracy of STK-3

Prediction model		Multi-factor GA-SVM	Single-factor GA-SVM	Multi-factor BPNN
Accuracy	RMSE/m	0.072	0.116	0.117
	MAPE/%	0.032	0.048	0.0376
	R	0.953	0.914	0.860

Figure 19 and Figure 20 showed that the goodness-of-fit of the multi-factor GA-SVM model was greater than the single-factor GA-SVM model and the multi-factor BPNN model, especially when the strong rainfall occurred and the groundwater level was fluctuating (24<sup>th</sup>–31<sup>st</sup>, August). The prediction results of the multi-factor GA-SVM model were closer to the measured values. Table 4 shows that the RMSE and the MAPE of the multi-factor GA-SVM model were 0.072 m and 0.032%, while the values of the BPNN model were 0.117 m and 0.0376%, respectively. The correlation coefficient of the former was 0.953; it was greater than the latter (0.860). Compared with the multi-factor models, the RMSE, MAPE, and R of the single-factor GA-SVM model were 0.116 and 0.048%, 0.914, respectively. The prediction results of the multi-factor model were superior to that of the single-factor model. These suggest that based on deep analysis of the geological conditions of landslide and selecting the appropriate factors, an accurate prediction of the groundwater level can be achieved.

## 6. Discussion

### 6.1. Performance Comparison of GA-SVM and BPNN

In order to compare the prediction stability of the GA-SVM and BPNN, three sets of prediction experiments were obtained with the same inputs and model parameters (Figure 21, 22). The REMS of prediction results are shown in Table 5. Although the inputs and parameter of the BPNN are the same, three sets of predicted results are different, while the prediction results of GA-SVM is constant. Overall, the proposed multi-factor GA-SVM method can achieve accurate and stable predictions in the groundwater level of landslides.



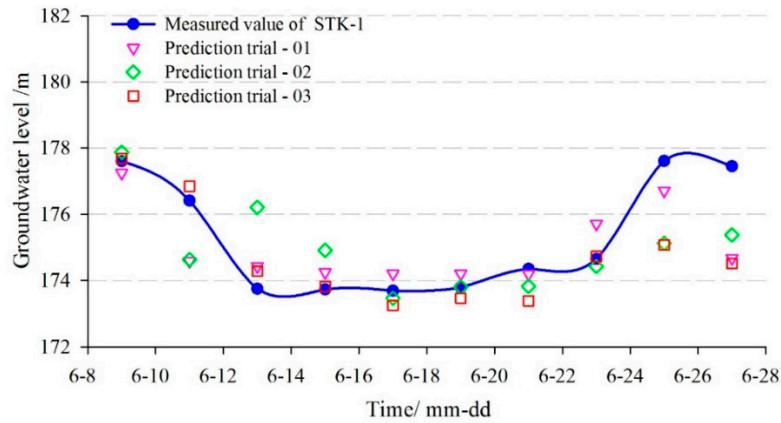


Figure 21. The prediction accuracy comparisons between different trials of BPNN(STK-1).

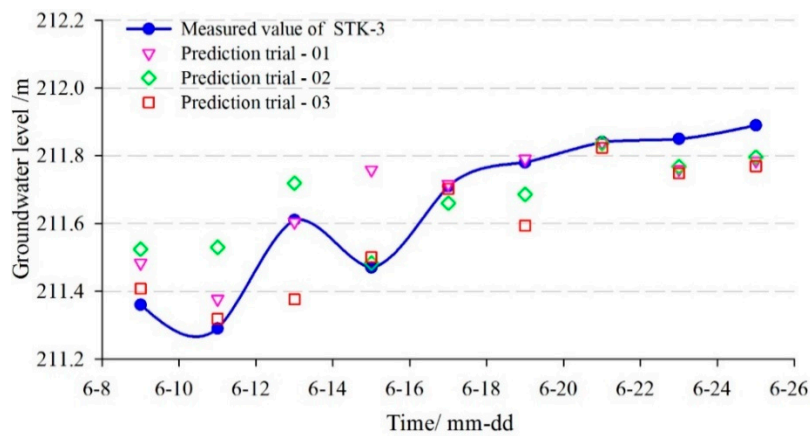


Figure 22. The prediction accuracy comparisons between different trials of BPNN(STK-3).

Table 5. The prediction accuracy comparisons between different trials of BPNN

Model	STK-1					STK-3				
	1st	2nd	3rd	Mean	Variance	1st	2nd	3rd	Mean	Variance
BPNN	1.195	1.466	1.295	1.319	0.112	0.118	0.117	0.115	0.117	0.001
GA-SVM	1.104	1.104	1.104	1.104	0.000	0.072	0.072	0.072	0.072	0.000

6.2. Future Developments of ML Methods

The groundwater level is a key parameter in the assessment of landslide stability. Accurate monitoring and prediction of groundwater level are essential for landslide early warning. Machine learning is a very effective prediction technology. It has been applied to landslide prediction and has achieved excellent results, such as landslide susceptibility mapping. In order to reduce the dependence on tedious numerical simulations, how to make full use of ML technology for landslide groundwater level prediction needs to be studied further.

There have been few reports on the application of ML techniques to groundwater level prediction up to now. We compared the performance of ML models in the application, mainly the accuracy and stability of prediction. In this study, through an optimum parameter searching by the GA optimization algorithm, a hybrid model of the coupled SVM and GA algorithm was established for prediction. Although GA-SVM performed well, in future studies, it is necessary to research the prediction performance of more advanced ML algorithms, such as ELM, Adaptive Network-based Fuzzy Inference System, optimization algorithms (Particle Swarm Optimization, Artificial Bee

Colony, Grey Wolf Optimizer, etc.), and ensemble learning (Bagging, Boosting, Stacking, etc.), Etc.. In order to achieve an accurate and stable prediction of landslide groundwater level, its urgent to study the combination of various algorithms and integrate their advantages.

### 6.3. Geological Factors of Groundwater Level

The groundwater level is not only affected by external inputs (rainfall, reservoir level, irrigation, etc.), but also closely related to the appearance condition of the slope, including the landslide structure and the macroscopic deformation characteristics (earth cracks, etc.). During the development process of a landslide, earth cracks are important companion products [45]. The surface cracks destroy the integrity of the soil, and significantly change the infiltration path and the distribution of the seepage field as well. They provide a preferential flow path for the rapid migration of water to deep soil. The study found that a slope would maintain a high saturation under long-term continuous rainfall, and the top-down development of the wetting front would eventually lead to direct precipitation of groundwater [46]. For landslides with loose mass and well-developed cracks, the groundwater will fill the cracks rapidly in a short-term concentrated rainfall. The anisotropy of crack parameters has different influence rules and degrees on groundwater level. The larger the permeability coefficient corresponding to the parameters, including the location, width, depth, density, length, fracture size, and suction characteristics of earth cracks, the faster the groundwater responses. Hence, how to predict the groundwater level by combining the characteristics of the influencing factors, the sliding body and the macroscopic deformation of a landslide needs further study.

## 7. Conclusions

The monitoring and prediction of the groundwater level are significant for landslide early warning. In this study, the Tangjiao landslide in the TGRA was taken as an example. Three groundwater level monitoring sensors were installed in different locations of the landslide. Based on the monitoring data, the response relationship between landslide groundwater level fluctuations and factors is analyzed, multi-factor GA-SVM, multi-factor BPNN, and single-factor GA-SVM prediction models were established. The main conclusions are as follows: 1) The fluctuation of groundwater level is significantly consistent with rainfall and reservoir level in time, and there is also a lag. There is a spatial difference in the impact of reservoir level on the groundwater level of the reservoir landslide. The closer to the Yangtze River, the stronger the reservoir level effect and the shorter the lag time; 2) On the basis of qualitative and quantitative analysis, the model inputs of STK-1 and STK-3 were established, respectively. Comparing the prediction results of multi-factor GA-SVM and single-factor GA-SVM; it can be found that multi-factor GA-SVM has higher accuracy at both locations. In the establishment of landslide groundwater level prediction model, it can significantly improve the prediction accuracy to consider the response of factors; 3) Comparing the prediction results of the multi-factor GA-SVM and the multi-factor BPNN models, it demonstrated that the prediction accuracy and stability of the GA-SVM model are both superior to the traditional BPNN model. The proposed GA-SVM model integrates the advantages of GA and SVM, it shows excellent modeling efficiency and accuracy. In general, by combining the advantages of different artificial intelligence algorithms, the multi-factor GA-SVM model proposed in this paper can effectively construct the response relationship between groundwater level fluctuations and influencing factors. It is a reliable prediction model, which can be recommended in predicting landslide groundwater levels.

**Author Contributions:** Writing—original draft: Y.C. and C.Z.; Writing—review & editing: Y.C., C.Z., K.Y., and B.A.; Supervision: K.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (No. 41702330 and No. 41907253).

**Acknowledgments:** We greatly appreciate the careful reviews and thoughtful suggestions by reviewers. We also thank the assistance of the Research Center of Geohazard Monitoring and Warning in the Three Gorges Reservoir.

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

## References

1. Asch, T.W.J.V.; Buma, J.; Beek, L.P.H.V. A view on some hydrological triggering systems in landslides. *Geomorphology* **1999**, *30*, 25-32.
2. Zhi, M.M.; Shang, Y.Q.; Zhao, Y.; Lü, Q.; Sun, H. Investigation and monitoring on a rainfall-induced deep-seated landslide. *Arab. J. Geosci.* **2016**, *9*, 182.
3. Chang, H.; Wang, X.S. Stability variation of a slope associated with transient groundwater flow: a case study on Shiliushubao landslide in Huanglashi landslides, Three Gorges Project Area, China. *Geol. Sci. Technol. Inf.* **2004**, *23*, 94-98.
4. Baharuddin, M.F.T.; Hazreek, Z.A.M.; Azman, M.A.A.; Madun, A. Prediction of Groundwater Level at Slope Areas using Electrical Resistivity Method. *J. Phys. Conf. Ser.* **2018**, *955*, 1742-6596.
5. Mathew, J.; Babu, D.G.; Kundu, S.; Kumar, K.V.; Pant, C.C. Integrating intensity-duration-based rainfall threshold and antecedent rainfall-based probability estimate towards generating early warning for rainfall-induced landslides in parts of the Garhwal Himalaya, India. *Landslides* **2014**, *11*(4), 575-588.
6. Shrestha, H.K.; Yatabe, R.; Bhandary, N.P. Groundwater flow modeling for effective implementation of landslide stability enhancement measures. *Landslides* **2008**, *5*, 281-290.
7. He, K.Q.; Wang, Z.L.; Ma, X.Y.; Lee, Z.T. Research on the displacement response ratio of groundwater dynamic augment and its application in evaluation of the slope stability. *Environ. Earth Sci.* **2015**, *74*, 5773-5791.
8. Trigo, R.M.; Zezere, J.L.; Rodrigues, M.L.; Trigo, I.F. The influence of the North Atlantic Oscillation on rainfall triggering of landslides near Lisbon. *Nat. Hazards* **2005**, *36*, 331-354.
9. Ray, R.L.; Jacobs, J.M. Relationships among remotely sensed soil moisture, precipitation and landslide events. *Nat. Hazards* **2007**, *43*, 211-222.
10. Tommasi, P.; Pellegrini, P.; Boldini, D.; Ribacchi, R. Influence of rainfall regime on hydraulic conditions and movement rates in the overconsolidated clayey slope of the Orvieto hill (central Italy). *Can. Geotech. J.* **2006**, *43*, 70-86.
11. Zhang, Z.; Wang, C.; Zhang, H.; Tang, Y.; Liu, X. Analysis of permafrost region coherence variation in the Qinghai-Tibet plateau with a high-resolution TerraSAR-X image. *Remote Sens.* **2018**, *10*, 298.
12. Cascini, L.; Calvello, M.; Grimaldi, G.M. Modelling the transient groundwater regime for the displacement analysis of slow-moving active landslides. *Tenth Int. Symp. Landslides Eng. Slopes* **2008**, *2*, 607-613.
13. Zhang, M.S.; Dong, Y.; Sun, P.P. Impact of reservoir impoundment-caused groundwater level changes on regional slope stability: a case study in the Loess Plateau of Western China. *Environ. Earth Sci.* **2012**, *66*, 1715-1725.
14. Zhao, Y.L. Exploration and research of the influence of the rise of groundwater on slopes. *J. Sichuan Univ. Sci. Eng. (Nat. Sci. Ed.)* **2008**, *21*(6), 110-112.
15. Kiawabe, H.; Shibano, H.; Nishio, K.; Yamaguchi, I. On the influence of Pore Water Pressure on the Land Deformation of the Landslide. *Landslides* **1991**, *27*, 33-36.
16. Tsao, T.M.; Wang, M.K.; Chen, M.C. Takeuchi Y, Matsuura S, Ochiai H. A case study of the pore water pressure fluctuation on the slip surface using horizontal borehole works on drainage well. *Eng. Geo.* **2004**, *78*, 105-118.

17. Martinović, K.; Gavin, K.; Reale, C.; Mangan, C. Rainfall thresholds as a landslide indicator for engineered slopes on the Irish Rail network. *Geomorphology* **2018**, *306*, 40-50.
18. Li, H.J.; Wang, X. Landslide groundwater flow field simulation analysis based on the finite element method. *Western Explor. Eng.* **2009**, *21*, 27-29.
19. Zhou, C.; Yin, K.L.; Cao, Y., Ahmed, B.; Fu, X.L. A novel method for landslide displacement prediction by integrating advanced computational intelligence algorithms. *Sci. Rep.* **2018**, *8*, 1-12.
20. Lian, C.; Zeng, Z.G.; Yao, W.; Tang, H.M. Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine. *Nat. Hazards* **2013**, *66*, 759-771.
21. Nguyen, V.V.; Pham, B.T.; Vu, B.T.; Prakash, I.; Jha, S.; Shahabi, H.; Shirzadi, A.; Ba, D.N.; Kumar, R.; Chatterjee, J.M.; Bui, D.T. Hybrid machine learning approaches for landslide susceptibility modeling. *Forests* **2019**, *10*, 157.
22. Choubin, B.; Mosavi, A.; Alamdarloo, E.H.; Hosseini, F.S.; Shamshirband, S.; Dashtekian, K.; Ghamisi, P. Earth fissure hazard prediction using machine learning models. *Environ. Res.* **2019**, *179*, 108770.
23. Shamshirband, S.; Nodoushan, E.J.; Adolf, J.E.; Manaf, A.A.; Mosavi, A.; Chau, K.W. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. *Eng. Appl. Comp. Fluid* **2019**, *13*, 91-101.
24. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361-378.
25. Zhou, C.; Yin, K. Landslide displacement prediction of WA-SVM coupling model based on chaotic sequence. *Electr. J. Geol. Eng.* **2014**, *19*, 2973-2987.
26. Qiang, L.; Yu, G.; Jing, J. Classification of multiple chinese liquors by means of a qcm-based e-nose and mds-svm classifier. *Sensors* **2017**, *17*, 272-286.
27. Yuan, Y.B. Forecasting the movement direction of exchange rate with polynomial smooth support vector machine. *Math. Comput. Model.* **2013**, *57*, 932-944.
28. Pham, B.T.; Jaafari, A.; Prakash, I., Bui, D.T. A Novel Hybrid Intelligent Model of Support Vector Machines and MultiBoost Ensemble for Landslide Susceptibility Modeling. *B. Eng. Geol. Environ.* **2018**, *78*, 2865-2886.
29. Karballaezadeh, N.; Mohammadzadeh S, D.; Shamshirband, S.; Hajikhodaverdikhan, P.; Mosavi, A.; Chau, K.W. Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan-Firuzkuh road). *Eng. Appl. Comp. Fluid* **2019**, *13*, 188-198.
30. Zhou, C.; Yin, K.L.; Cao, Y., Ahmed, B.; Li, Y.Y.; Catani, F.; Pourghasemie, H.R. Landslide susceptibility modeling applying machine learning methods: a case study from longju in the three gorges reservoir area, China. *Comput. Geosci.* **2018**, *112*, 23-37.
31. Pham, B.T.; Pradhan, B.; Tien, B.D.; Prakash, I.; Dholakia, M.B. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environ. Modell. Softw.* **2016**, *84*, 240-250.
32. Faizollahzadeh Ardabili, S.; Najafi, B.; Alizamir, M.; Mosavi, A.; Shamshirband, S.; Rabczuk, T. Using SVM-RSM and ELM-RSM approaches for optimizing the production process of methyl and ethyl esters. *Energies* **2018**, *11*, 2889.
33. Bui, D.T.; Pham, B.T.; Nguyen, Q.P.; Hoang, N.D. Spatial prediction of rainfall-induced shallow landslides using hybrid integration approach of least-squares support vector machines and differential evolution optimization: a case study in central vietnam. *Int. J. Digit. Earth.* **2016**, *9*, 1077-1097.

34. Chou, J.S.; Cheng, M.Y.; Wu, Y.W.; Pham, A.D. Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Syst. Appl.* **2014**, *41*, 3955-3964.
35. Zhou, C.; Yin, K.L.; Cao, Y.; Ahmed, B. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the three gorges reservoir, China. *Eng. Geol.* **2016**, *204*, 108-120.
36. Zhou, C.; Yin, K.L.; Cao, Y.; Intrieri, E.; Ahmed, B.; Catani, F. Displacement prediction of a step-like landslide by applying a novel kernel extreme learning machine method. *Landslides* **2018**, *15*, 2211-2225.
37. Zhang, W.; Gao, L.; Jiao, X.; Yu, J.; Su, X.; Du, S. Occurrence assessment of earth fissure based on genetic algorithms and artificial neural networks in Su-Xi-Chang land subsidence area, China. *Geosci. J.* **2014**, *18*, 485-493.
38. Betiku, E.; Okunsolawo, S.S.; Ajala, S.O.; Odedele, O.S. Performance evaluation of artificial neural network coupled with generic algorithm and response surface methodology in modeling and optimization of biodiesel production process parameters from shea tree (*Vitellaria paradoxa*) nut butter. *Renew. Energy* **2015**, *76*, 408-417.
39. Cai, Z.; Xu, W.; Meng, Y.; Shi, C.; Wang, R. Prediction of landslide displacement based on GA-LSSVM with multiple factors. *B. Eng. Geol. Environ.* **2016**, *75*, 637-646.
40. Wang, X.C.; Shi, F.; Yu, L. MATLAB neural network analysis of 43 cases. Beijing University Press: Bei Jing, China, **2013**.
41. Goh, A.T.C. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.* **1995**, *9*, 143-151.
42. Yang, Z.P.; Lu, W.X.; Long, Y.Q.; Li, P. Application and comparison of two prediction models for groundwater levels: A case study in Western Jilin Province, China. *J. Arid. Environ.* **2009**, *73*, 487-492.
43. Samadianfard, S.; Jarhan, S.; Salwana, E.; Mosavi, A.; Shamshirband, S.; Akib, S. Support Vector Regression Integrated with Fruit Fly Optimization Algorithm for River Flow Forecasting in Lake Urmia Basin. *Water* **2019**, *11*, 1934.
44. Cao, Y.; Yin, K.L.; Alexander, D.E.; Zhou, C. Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors. *Landslides* **2016**, *13*, 725-736.
45. Jin, D.S. Identification and analysis of landslide fissures. *Chin. J. Geol. Hazard Control* **2002**, *2*, 52-54.
46. Xu, Q.; Fan, X.M.; Li, Y.; Zhang S. Formation condition, genetic mechanism and treatment measures of plate-shaped landslide. *Chin. J. Rock Mech. Eng.* **2010**, *29*, 242-250.

