J. Mt. Sci. (2018) 15(1): 156-166 e-mail: jms@imde.ac.cn http://jms.imde.ac.cn
https://doi.org/10.1007/s11629-016-4245-3

A WD-GA-LSSVM model for rainfall-triggered landslide displacement prediction

ZHU Xing[1,2] http://orcid.org/0000-0002-0602-7832; e-mail: zhuxing330@163.com

MA Shu-qi[2*] http://orcid.org/0000-0002-5820-4394; e-mail: sqma@ntu.edu.sg

XU Qiang[1*] http://orcid.org/0000-0002-5192-6116; e-mail: xuqiang_68@126.com

LIU Wen-de[1] http://orcid.org/0000-0002.9312-0401; e-mail: 619518901@qq.com

** Corresponding author*

1 State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China

2 School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 637787, Singapore

Citation: Zhu X, Ma SQ, Xu Q, et al. (2018) A WD-GA-LSSVM model for rainfall-triggered landslide displacement prediction. Journal of Mountain Science 15(1). https://doi.org/10.1007/s11629-016-4245-3

Abstract: This paper proposes a WD-GA-LSSVM model for predicting the displacement of a deep-seated landslide triggered by seasonal rainfall, in which wavelet denoising (WD) is used in displacement time series of landslide to eliminate the GPS observation noise in the original data, and genetic algorithm (GA) is applied to obtain optimal parameters of least squares support vector machines (LSSVM) model. The model is first trained and then evaluated by using data from a gentle dipping (~2°-5°) landslide triggered by seasonal rainfall in the southwest of China. Performance comparisons of WD-GA-LSSVM model with Back Propagation Neural Network (BPNN) model and LSSVM are presented, individually. The results indicate that the adoption of WD-GA-LSSVM model significantly improves the robustness and accuracy of the displacement prediction and it provides a powerful technique for predicting the displacement of a rainfall-triggered landslide.

Keywords: WD-GA-LSSVM; Landslide; Rainfall; Displacement prediction; Wavelet denoising

Received: 12 October 2016
Revised: 15 February 2017
Accepted: 1 August 2017

Introduction

Landslide is one of the major widespread geological hazards throughout the world, causing thousands of fatalities, and extensive property damage every year. Especially over recent years, the rate of landslides occurrence has increased in the mountain areas of South-West China due to the post-effects of huge earthquakes (e.g. the 5·12 Wenchuan earthquake in 2008). Rainfall becomes one of the main triggers for land sliding in those areas. It is significantly important to have a profound knowledge of the dynamic characteristics of rain-triggered landslides in order to prevent or reduce property damage and loss to the largest extent the risk. For this purpose, monitoring and predicting landslide-prone slope deformation is very useful. It is well known that the evolution of a landslide from a stable state to an unstable state is a highly non-linear process, which is caused by the complicated geological settings, complex geometry and varying hydrological conditions. Accordingly, the mathematical model is adopted to predict landslide deformation accurately. However, the displacement of rainfall-triggered landslides cannot be easily predicted in the temporal domain due to 1) complicated interactions of precipitation with the landslide body 2) lagging effects of the hydrological response, 3) and variation in failure mechanisms in different evolutional phases. High-precision prediction can provide a scientific guide for the early warning and forecast of landslide activity. Numerous studies have proposed many models for the prediction of landslide displacement, such as the time series model (Xu et al. 2011; Zhang et al. 2014; Zhou et al. 2016), exponential smoothing model (Liu et al. 2009; Yin et al. 2007), grey model (Chen and Wang 1988; Lv and Liu 2012), functional network model (Chen et al. 2015) , extreme learning machine (Feng et al. 2004; Lian et al. 2012; Lian et al. 2013) and artificial neural network model (Jiang and Chen 2016; Lian et al. 2015; Nefeslioglu et al. 2008). Among those proposed prediction models, nonlinear models are considered to have the greatest potential for solving difficult and complicated problems. Artificial Neural Network (ANN)-based methods are effective in coping with complex problems due to their random non-linear fitting capability and data-driven feature, which have been frequently used in landslide prediction recently (Cai et al. 2015). However, in most cases, ANN-based methods have some drawbacks, such as slow convergence rate and local minimum traps (Cai et al. 2015; Jiang and Chen 2016; Lian et al. 2012). Furthermore, the precision of prediction cannot be guaranteed with limited training samples in most of the above mentioned methods.

A machine learning method, support vector machine (SVM), was proposed by (Vapnik et al. 1997; Vapnik 1995). It shows a good performance in classification, regression and function approximation. It looks for a non-linear relation between inputs and outputs through mapping the inputs to a high dimension space based on a kernel function (Su et al. 2015). Especially, the SVM can find global optimal solutions for complex non-linear problems with small training samples. However, the SVM has high computational complexity due to quadratic programming (Vapnik 1995). For better performances, (Suykens and Vandewalle 1999) proposed an improved version of SVM—the least square support vector machine (LSSVM), which runs faster and shows more adaptability because the quadratic optimization problem of SVM is transformed into a linear system of equations (Cai et al. 2015). Recently, LSSVM has been adopted to landslide deformation prediction in some researches (Cai et al. 2015; Dou et al. 2015; Liu and Zhang 2014). But it does not address the problem how to select optimal value of γ and σ^2 parameter, which may directly influence the model's performance. Genetic Algorithms (GA) is one of the modern heuristic methods that can be used to find the global optimal parameters for LSSVM model.

A WD-GA-LSSVM model for rainfall-triggered landslide displacement prediction is proposed in this study. The wavelet denoising (WD) is used in displacement time series and then the LSSVM is optimized by GA algorithm. The monitoring precipitation and displacement observed in specific antecedent time range are selected to train/ considered as inputs of WD-GA-LSSVM, while the Kualiangzi landslide in the southwest of China is chosen as a case study to verify this model. And the comparative analysis of the Back Propagation Neural Network (BPNN) model, LSSVM and WD-GA-LSSVM model predictions indicate that the WD-GA-LSSVM model significantly improves the robustness and prediction accuracy of the model.

1 Methodology

1.1 Least squares support vector machines (LSSVM)

LSSVM is the abbreviation for the Least Squares version of a classical Support Vector Machines that are a set of related supervised learning methods that analyse data and recognize patterns, and are used for classification and regression analysis. LSSVM applies the linear least squares criteria to the loss function instead of a convex quadratic programming problem used for the classical SVM.

Given a training data set of N samples $\{x_i, y_i\}_{i=1}^{N}$ with input data $x_i \in R^n$ and $y_i \in R$ being the corresponding target values, where R^n is the n-dimensional vector space and R is the one-dimensional vector space. The LSSVM carries out mapping of the samples with a linear regression function in a high-dimensional feature space.

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

where $w \in R^n$ denotes the adjustable weight vector, $b \in R$ is bias, and $\phi(x)$ is the non-linear kernel mapping function, which maps the input vector to the high-dimensional feature space.

Based on the structural risk minimization principle, the objective function of the LSSVM can be given as follows (Suykens and Vandewalle 1999):

$$\min_{w,b,\sigma} J(w,\sigma) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^2 \tag{2}$$

which is subject to the equality constraint:

$$y_i = w^T \phi(x_i) + b + e_i, \quad i = 1,2,\ldots,N \tag{3}$$

where γ is a regularization parameter, which determines the trade-off between the training error and the model fitness; e_i is the random error; $\phi(x)$ is the non-linear kernel function that maps the input data x_i into a high dimensional fature space, where a linear regression problem is established and solved.

For solving the optimization problem of LSSVM, a Lagrangian function is constructed as follows:

$$L(w,b,\sigma,\alpha) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \left[w^T \phi(x_i) + b + e_i - y_i \right] \tag{4}$$

where α_i is the Lagrange multiplier. The conditions for optimality are:

$$\frac{\partial L}{\partial w} = 0 ; \frac{\partial L}{\partial b} = 0 ; \frac{\partial L}{\partial e_i} = 0 ; \frac{\partial L}{\partial \alpha_i} = 0 \tag{5}$$

So, the solution of Eq.(4) can be described as follows:

$$\begin{cases} w = \sum_{i=1}^{N} \alpha_i \ \phi(x_i) \\ \sum_{i=1}^{N} \alpha_i = 0 \\ \alpha_i = \gamma e_i \\ w^T \ \phi(x_i) + b - e_i - y_i = 0 \end{cases} \tag{6}$$

Elimination of w and e_i will yield a linear system instead of a quadratic programming problem:

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & K + \gamma^{-1} I_N \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{7}$$

with $Y=[y_1,y_2,\ldots,y_N]^T$, $I_N=[1,\ldots,1]^T$ and $\alpha=[\alpha_1,\ldots,\alpha_N]^T$. Where I_N is an $N \times N$ identity matrix. K is a kernel function matrix, and $K_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$. There are three types of kernel functions: a polynomial function, a radial basis function (RBF), and a sigmoid function. In this study, RBF is chosen as the kernel function due to its fewer parameters that need to be set and excellent non-linear mapping performance (Cai et al. 2015; Li et al. 2015). It is given by:

$$K(x, x_i) = \exp(-\frac{1}{2\sigma^2} \|x - x_i\|^2) \tag{8}$$

where σ^2 is the parameter related to the bandwidth of the kernel in statistics, which is an important parameter for the generalization behaviour of a kernel method.

Finally, the regression function of the LSSVM model can be constructed as follows:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{9}$$

As mentioned above, two parameters (regularization parameter γ and kernel parameter σ^2) have powerful influence on the efficiency and generalization performance of the LSSVM model. Generally, the training error will be decreasing with increasing γ value, but when the value of γ is too high, the model will be very complex and time-consuming, and will cause over-training problems as well. The σ^2 value will affect the non-linear mapping performance from input space to high-dimension feature space, which relates to the generalization of the LSSVM model. At present, the determination of the parameters has no mature theory for guiding and normally a trial and error approach is conducted. For prediction it is difficult to reach the best performance of the LSSVM model. To overcome this problem, Genetic Algorithm (GA) is applied to determine the optimal values of these two free parameters.

1.2 Parameter optimization of LSSVM with GA

Up to now, there are various intelligent algorithms proposed for the optimization problem, such as a genetic algorithm (GA) (Li and Kong 2014; Wang et al. 2012), improved fly optimization algorithm (IFOA) (Si et al. 2016), particle swarm optimization (PSO) (Gholgheysari Gorjaei et al. 2015; Zhou et al. 2016). However, some studies prove that GA is a better choice to determine the parameters because it can determine the global optimal solution quickly by simulating a natural selection process, which can reduce the uncertainty of an experience-based choice and improve the predicative accuracy of the LSSVM model. The algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm randomly selects individuals from the current population and uses them as parents to produce the next generation with three genetic operators of selection, crossover, and mutation. At last, the best individual (the optimal values of LSSVM parameters) can be found with repeated evolution from generation to generation. Therefore, GA is chosen to search for the optimal parameters in our study. This algorithm can be implemented in MATLAB with a combination of the GAOT toolbox (Houck et al. 1995) and LSSVM toolbox (Suykens et al. 2002). The procedure for implementing the GA-LSSVM model is depicted in Figure 1. In addition, the wavelet de-nosing method is applied in the pre-processing of displacement time-series of landslides before training the LSSVM model, for reducing the disturbances of measuring noises. The optimal parameters of this model can be found by searching the minimum fitness value, which is defined with the root mean square error (RMSE) of the output of the model employing training samples.

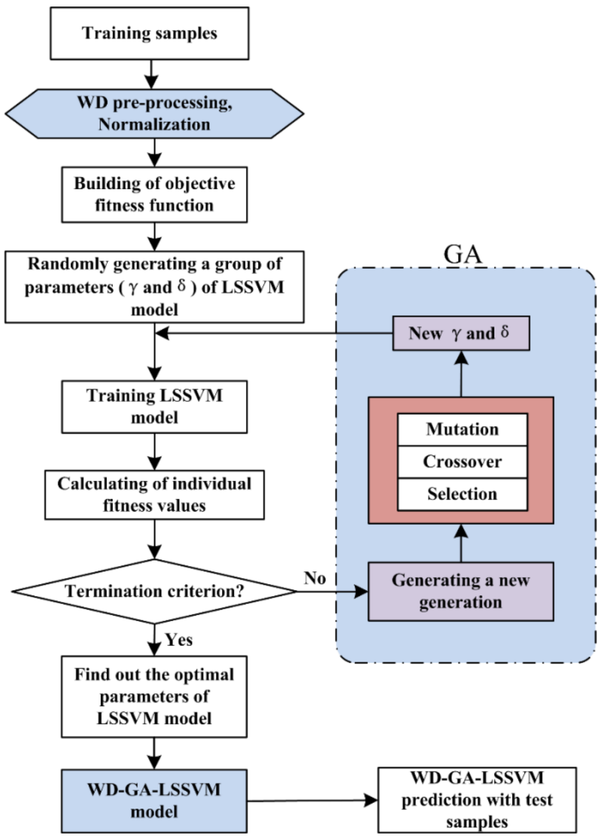

Figure 1 Flowchart of the WD-GA-LSSVM model (modified from Cai et al. 2015).

1.3 Prediction performance criteria

The performance of the proposed models is evaluated using some common methods, including average relative percentage error (ARPE), average absolute percentage error (AAPE), root mean square error (RMSE), and relation index (RI). They are designed to measure the deviation of predicted values from the actual observed values. A smaller error value indicates better prediction performance of the model in terms of ARPE, AAPE, and RMSE. RI indicates the correlation between the prediction and the observation. For a predictive model with higher accuracy, the value of *RI* should be close to 1. The above indicators can be defined as follows (Gholgheysari Gorjaei et al. 2015):

$$\mathrm{ARPE}(\%) = 100 \times \frac{1}{N}\sum_{i=1}^{N}\frac{(y_{oi}-y_{pi})}{y_{oi}} \tag{10}$$

$$\mathrm{AAPE}(\%) = 100 \times \frac{1}{N}\sum_{i=1}^{N}\left|\frac{(y_{oi}-y_{pi})}{y_{oi}}\right| \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{oi}-y_{pi})^2} \tag{12}$$

$$\mathrm{RI} = \sqrt{1-\frac{\sum_{i=1}^{N}(y_{oi}-y_{pi})^2}{\sum_{i=1}^{N}(y_{oi}-\bar{y}_o)^2}} \tag{13}$$

where, y_{oi} and y_{pi} are the actual and predicted value, respectively; $\bar{y}_o$ is the mean of displacement values; N denotes the number of time series data.

The accuracy factor (A_f), a simple multiplicative factor denoting the spread of results about the prediction, is calculated as follows (Basant et al. 2010):

$$A_f = 10^{\left(\frac{\sum\left|\log(y_{pi}/y_{oi})\right|}{N}\right)} \tag{14}$$

Here, A_f =1 indicates that there is a perfect agreement between all the predicted values and the observed values.

2 Description of the Landslide Case

The Kualiangzi landslide, which is a large and typical rain-triggered deep-seated slow moving rock slide, is located in Zhongjiang county, Deyang city, Sichuan province, China. Figure 2 shows the aerial view of this landslide in 2013 (Xu et al. 2015). The landslide developed in a nearly horizontal bedding rocky slope, with a maximum width and length of 1300 m and 420 m respectively. The

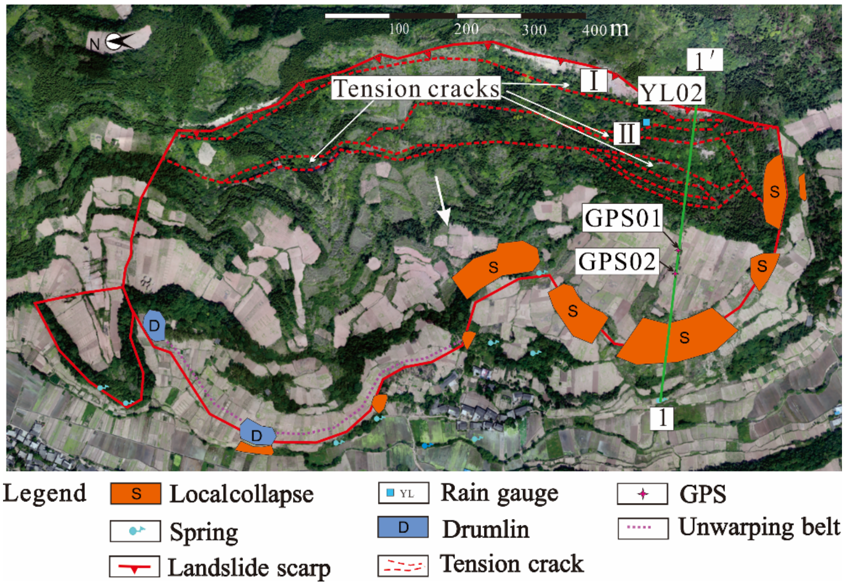

Figure 2 Aerial view of the Kualiangzi landslide and the location of monitoring devices (Xu et al. 2015).

average thickness of the landslide is about 50 m which was obtained from borehole data. Totally, it covers an area of approximately 0.51 km^2, and has an estimated volume of 2.55×10^7 m^3 (Xu et al. 2015). The Kualiangzi landslide has experienced two dramatic movements in the rainfall season of 1949 and 1981and it has a gentle dipping slip surface, which is caused during the formation of red rock and is typical for the eastern Sichuan province. Firstly, a large and long tension crack was generated at the rear edge of the landslide (I-crack in Figure 2). Intermittent slides were also triggered by several rainfall events in the next 60 years, and gradually a secondary tension trough was formed (II-crack in Figure 2) at the rear edge, together with slumps and upheaval at the leading edge. The strong rainfalls in 1981 enlarged the I-crack to a length of 1km and a width of 60m.

Figure 3 shows the engineering geological profile of this landslide. Two GPS displacement observation stations and one rain gauge were deployed in the field in 2012 to monitor the behaviour of this landslide continuously. Although the dip of the slide surface is only 2°~5°, it is still under a creeping state that is difficult to understand based on the traditional limit equilibrium theory. It is found that the rainfall is a crucial influencing factor in the recent landslide displacements. The average annual precipitation is 844.5mm, and the main rainfall occurs in the rainfall season from May to August every year. The precipitation in this area is always characterized by its high intensity and long duration.

Figure 4(a) shows the amount of rainfall from June 1st to August 15th in 2013, which is obviously characterized with abundant precipitation. Ten large rainfall events occurred with an intensity over 20 mm/d. Among those events were five rainstorms (indicated with A~E) with daily precipitation over 50 mm/d. The cumulative precipitation in this period, which is called the rainfall season, reached 660 mm. In contrast, the non-rainfall season covers the period from August 16th to December 31st during which the daily precipitation of each rainfall event was lower than 20 mm and the cumulative precipitation reached only 180 mm. Figure 4(b) shows the corresponding displacement of the landslide, which indicates that the intense seasonal precipitation is the significant factor in the landslide deformation. The displacement rate varies with precipitation intensity all the time.

3 Application to Landslide Displacement Prediction

3.1 Application results

The Kualiangzi landslide has been continuously monitored since June 2013. The monitoring data was uploaded to the monitoring service centre automatically via wireless radio techniques and the GPRS network. The information including daily precipitation and displacement time series is depicted in Figure 5(a). It can be seen that the deformation of the landslide was mainly influenced by the seasonal precipitation every year. As shown in Figure 5(b) (indicated with blue area in Figure 5(a)), the displacement rate of one day might be affected by the average rainfall intensity of several days prior to that day, rather than by only the maximum value

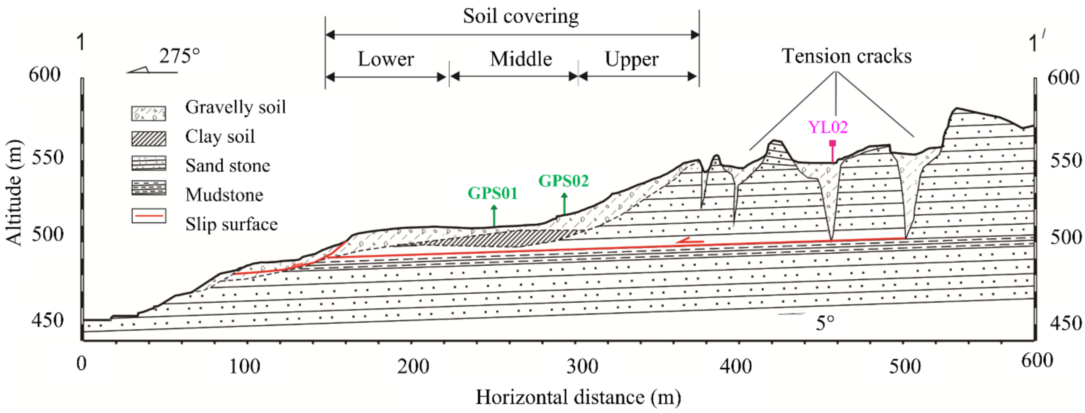

Figure 3 Engineering geological profile and the monitoring profile of section 1.1 in Figure 2 (Xu et al. 2015).

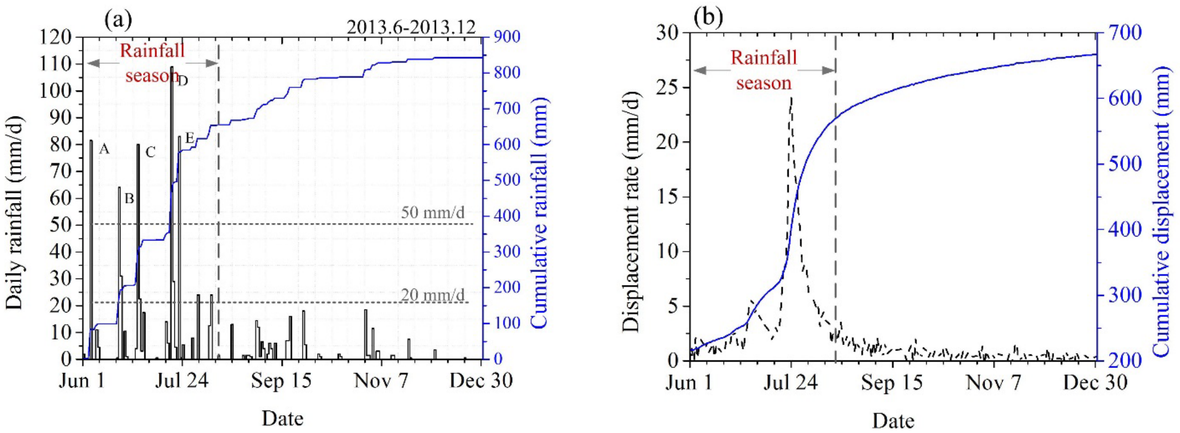

Figure 4 Monitoring results during the period from June to December, 2013. (a) Daily rainfall and cumulative precipitation; (b) Landslide displacement during this period.

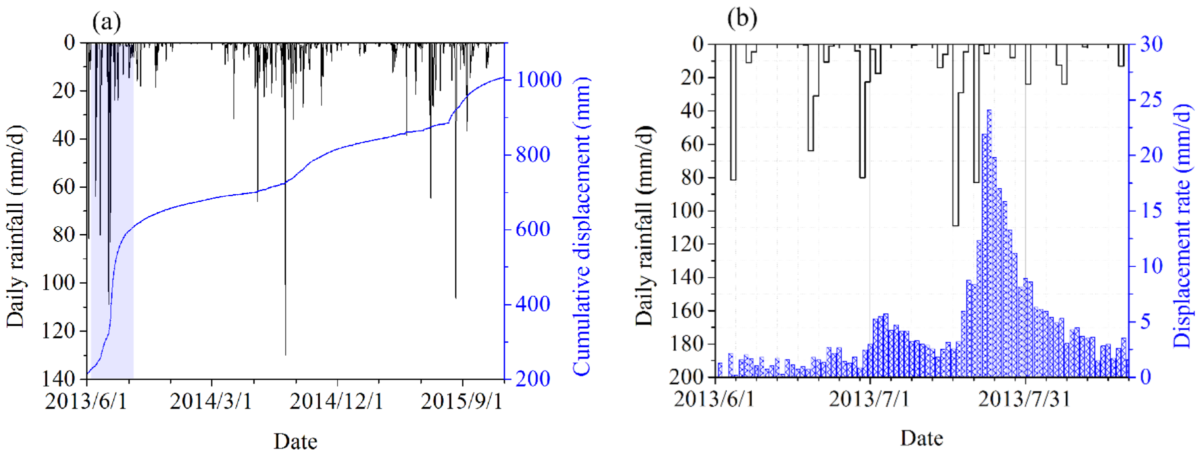

Figure 5 Monitoring data of the Kualiangzi landslide from 2013 to 2015. (a) Daily rainfall and cumulative displacement, blue area indicates the rainfall season in 2013; (b) Daily rainfall and displacement rates during the rainfall season in 2013.

of one rainfall event or one daily precipitation. This is due to the lagging effects of the infiltrating water, which is controlled by the geological and topographic characteristics of the landslide. Furthermore, the amount of infiltration into the landslide body may be different from time to time for each rainfall event, due to the variation of the degree of saturation of the geo-materials, the existing water level and other conditioning factors.

Essentially, the latest daily displacement rates can directly reflect the interaction mechanisms of the underground water level and the landslide behaviour in nearly real time. It is worth mentioning that the generalization and stable abilities of the proposed model can reach the best performance after being trained using the observed data collected during the whole rainfall season and non-rainfall season. It implies dynamic behaviour of the landslide over a complete range of different scenarios. Figure 6 shows the performances of WD-GA-LSSVM model versus the selected number of days for the average antecedent precipitation. It suggests that the average daily precipitation in the latest 20 days can perform the best accuracy in this case study. In this study, daily displacement rates in the last 3 days combining with the antecedent average precipitation over 20 days as input vector of the WD-GA-LSSVM model for displacement prediction can provide the best accuracy. However, this lagging effect would be different from case to case because of the variations of geological and topographical conditions (Lepore et al. 2013). The antecedent time should be determined during the model training phase.

The measurement data of this landslide from June to December in 2013 are selected as training data for building the model. Because of the measurement errors induced by the GPS system, the original displacement data were de-noised by the wavelet de-noising (WD) method before being used to train this model. The MATLAB built-in function *wden* is applied to smooth the original data in order to eliminate measurement errors.

The input for training the proposed model should be normalized between 0 and 1 based on Eq. (15) to make sure that the utilized input variables were independent of measurement units.

$$x_{nori} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (15)$$

where x_i and x_{nori} are the *i-th* original data and normalized data respectively , $\min(x)$ and $\max(x)$ are respectively minimum and maximum values of the training data series.

GA was used to optimize the two free parameters (γ and σ^2) of the LSSVM model. The GA has a generation number of 120 and an initial population size of 40. Figure 7 shows the search process for optimal parameters that are found after

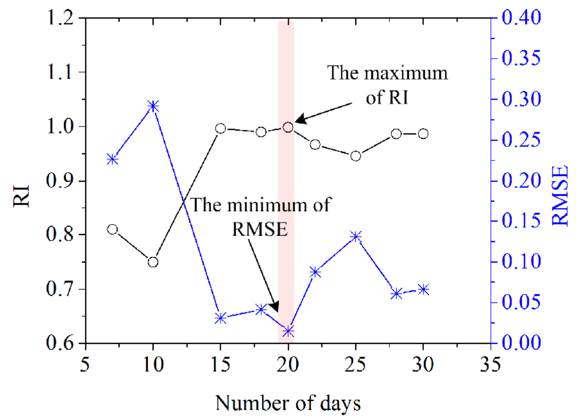

Figure 6 Performances of WD-GA-LSSVM model against the number of days for average antecedent precipitation.

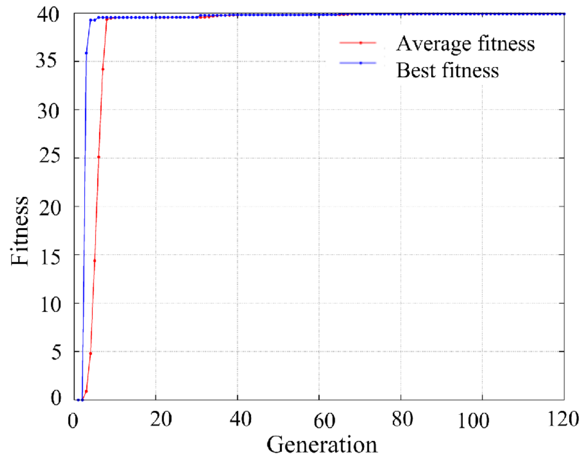

Figure 7 Fitness curve of the search for the optimal parameters for the WD-GA-LSSVM model.

60 generations using the genetic algorithm method. The training performance of the WD-GA-LSSVM model can be seen in Figure 8(a), indicating a good agreement between the output of the model and the monitored displacement. The optimized parameters γ (=751.5605) and σ^2 (= 50) for the WD-GA-LSSVM model are obtained through the GA optimization procedure and the training process. Figure 8(b) shows the training performance of GA-LSSVM without WD processing. Obviously, WD is very helpful to eliminate the influences of measurement errors to improve the precision of this model.

Subsequently, the WD-GA-LSSVM model was applied to predict the displacement variations of the landslide in 2014 and 2015, respectively. The prediction results can be seen in Figure 9(a) and Figure 9(b). The predictive results of the WD-GA-LSSVM model are in very good agreement with the observed displacements measured in the field. The figures show that the proposed model can predict the changes in daily displacement of the landslide accurately.

3.2 Comparison of GA-LSSVM with other available predictive models

In order to evaluate the performance of the above-proposed WD-GA-LSSVM model, we applied the Back Propagation Neural Network (BPNN) model, which is available in MATLAB's built-in toolbox for prediction, and using the same training samples as for the WD-GA-LSSVM model. Three layers were applied to construct the BPNN model, consisting of an input layer, a hidden layer and an output layer. Tansig and purelin transfer functions were used for the hidden and the output layer respectively. The BPNN model was compared with the WD-GA-LSSVM model, and the LSSVM

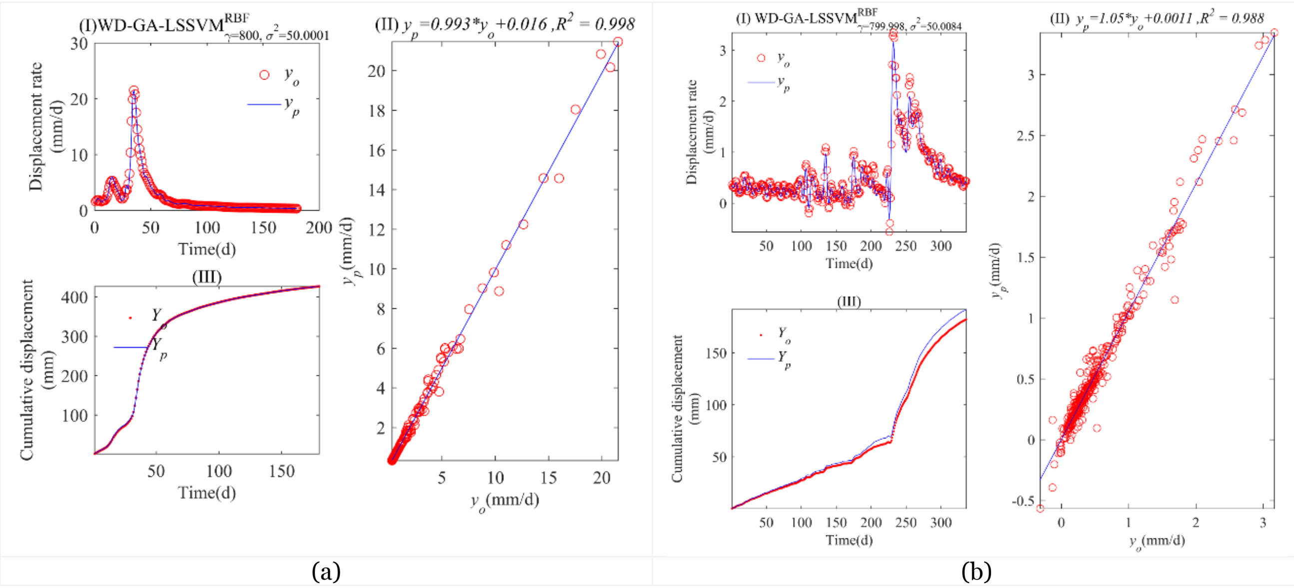

Figure 8 (a) Output of WD-GA-LSSVM model and (b) output of GA-LSSVM training with the displacement time series in 2013. (I) Estimated (y_p) and observed (y_o) displacement rate, (II) agreement in prediction by the proposed model and observed displacement rate, (III) cumulative displacement, estimated (Y_p) and observed (Y_o).

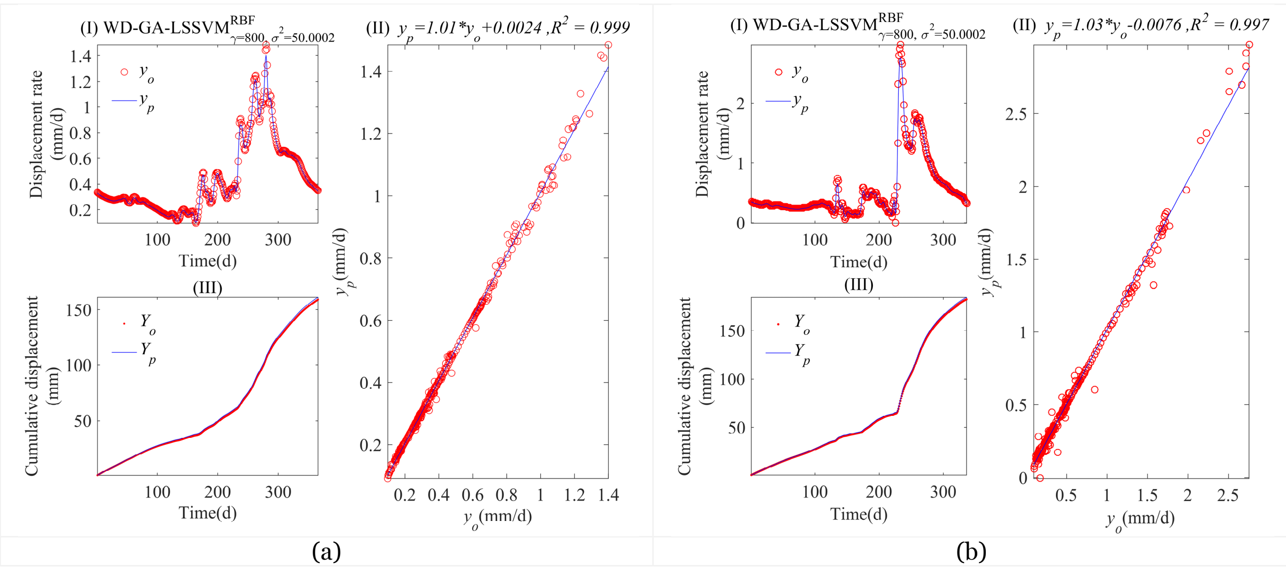

Figure 9 Prediction of landslide deformation in 2014 (a) and 2015 (b). (I) Predicted (y_p) and observed (y_o) displacement rates, (II) agreement in prediction by the proposed model and the observed displacement rate, (III) cumulative displacement, predicted (Y_p) and observed (Y_o).

model as well. The optimal parameters of γ and σ^2 were found to be 10754.904 and 17.854 respectively by using grid search approach in the stand alone LSSVM model. Table 1 shows the performances of those three models for the same testing data samples respectively. It was observed that the WD-GA-LSSVM model performed slightly better than the LSSVM model in terms of RMSE, ARPE, AAPE, relation index (RI), and accuracy factor (A_f).

4 Discussion

In order to evaluate the direct influences of antecedent precipitation on the deformation of the Kualiangzi landslide, the average daily rainfall over certain prior days was chosen as the single input factor in the WD-GA-LSSVM to predict landslide displacement. Figure 10(a) shows the training performance of the WD-GA-LSSVM model only using this single rainfall factor. The figure shows that it is difficult to predict the future displacement rate just using the average antecedent rainfall data as the single factor of landslide displacement. Figure 10(b), shows that the prediction of the WD-GA-LSSVM model, using as input both the daily displacement rate in the latest days and the antecedent precipitation is in very good agreement with the real observations in the field. The daily displacement rate in the latest days, which is directly reflecting the inner behaviour of the landslide in time proved to be of great importance for landslide displacement prediction.

The results show that the excellent performances of the WD-GA-LSSVM model for prediction of displacement of landslide triggered by only one factor related to rainfall. However, it should be revised and improved for more complicated landslide systems that responding for multiple factors like the effect of an earthquake or the effect of remedial works. And essentially, this model should also be re-trained for keeping good predictive performances, because of variations of geo-technical, morphological and geometrical characteristics over a longer time. Therefore, the influence of precipitation on the deformation of landslide is nonlinear, dynamic and complex because the amount of evapotranspiration, the amount of overland flow and the aperture of fissures will vary from time to time. The implementation of a high-precision prediction of landslide deformation, which is only based on historical rainfall data, is a challenging task.

Table 1 Comparative analysis performances of three different displacement prediction models: WD-GA-LSSVM, LSSVM, and BPNN

Model	γ	σ^2	*RMSE*	*ARPE* (%)	*AAPE* (%)	*RI*	A_f
WD-GA-LSSVM	751.561	50.000	0.015	0.37	2.45	0.999	1.007
LSSVM	10754.904	17.854	0.023	-1.33	3.07	0.998	0.967
BPNN	-	-	0.112	10.58	21.89	0.920	0.893

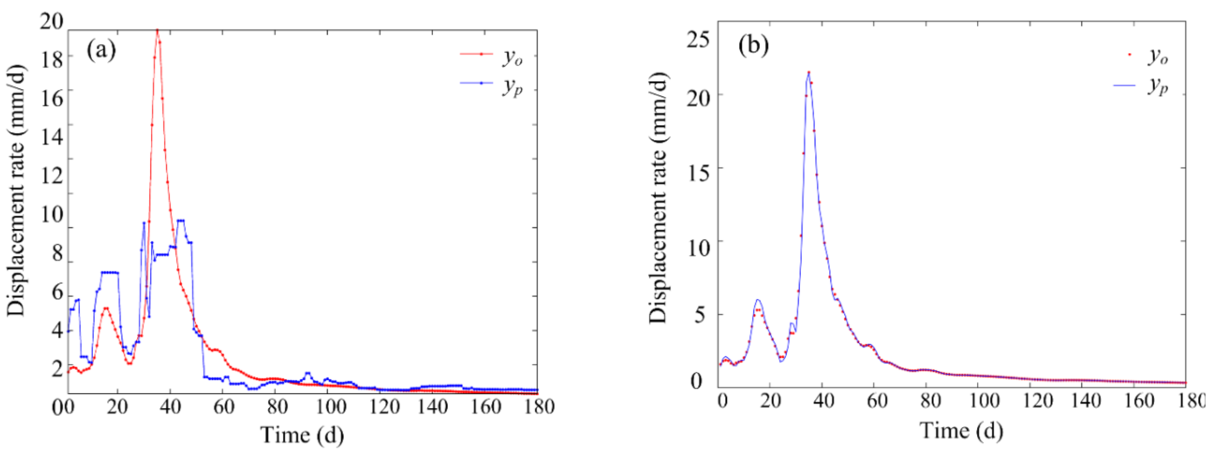

Figure 10 Performances of the GA-LSSVM model with different input data. (a) Prediction based on only the average daily rainfall in a certain prior period; (b) Prediction based on daily displacement rates of the last three days and average daily rainfall data over 20 antecedent days.

5 Conclusions

This study proposes a novel WD-GA-LSSVM model with a wavelet de-noising technique to accurately predict the displacement of a landslide. The average antecedent precipitation and the daily displacement variations are chosen as the input vector of the WD-GA-LSSVM model for the prediction of landslide displacement in the next days. A rainfall-triggered landslide is chosen as a case study. The following conclusions can be drawn from the application of the model and analysis of the results:

(1) The WD-GA-LSSVM model can provide a higher predictive accuracy, better generalization performance, and more robustness even for changes of displacement over short (daily) periods compared with other models (e.g. LSSVM, BPNN). GA is used to optimize two tuning parameters of LSSVM model, whilst BPNN needs to pre-set many parameters associated with the network's construction and to be trained with many more samples. The wavelet de-noising method can be used to effectively eliminate the measurement errors of displacement to improve the accuracy of the WD-GA-LSSVM model. The performances of the WD-GA-LSSVM model proved to be the best with the smallest RMSE of 0.015 mm, an ARPE of 0.37%, a relation index (*RI*) of 0.9986 and an accuracy factor (A_f) of 1.0073.

(2)After many trials, the current displacement variations in the latest three days perfectly reflect the response of the landslide system to the external rainfall factor, which is non-linear, dynamic and complicate. The variation in landslide displacement is mainly controlled by the rainfall trigger and the evolution state of the landslide. The modelling results indicate that rainfall data cannot be used independently to predict accurately the evolution of the landslide because of the temporal variations in interaction between rainfall and the landslide body. Therefore, apart from continuous observation of precipitation, real-time displacement monitoring of landslide plays a very important role in the high-accuracy prediction of landslide displacement.

(3) In this landslide case, it is proved that the combination of daily displacement rates of the previous 3 days and average daily rainfall data over 20 antecedent days is the best input vector of the LSSVM model to predict the displacement rate of the landslide effectively and accurately. However, the time period of the antecedent rainfall used may be varying in other landslide case because the inner structure, geomorphology, and other features are possibly different.

Acknowledgments

This study was supported by the Chinese National Natural Science Foundation (Grant No. 41502293), the National Basic Research Program (973 Program) (Grant No. 2014CB744703), the Funds for Creative Research Groups of China (Grant No. 41521002). We would like to extend special thanks to Dr. T.W.J. van Asch, for all his valuable suggestions in greatly improving the quality of this paper.

References

Basant N, Gupta S, Malik A, Singh KP (2010) Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water — A case study. Chemometrics and Intelligent Laboratory Systems 104: 172-180. https://doi.org/10.1016/j.chemolab.2010.08.005

Cai Z, Xu W, Meng Y, et al. (2015) Prediction of landslide displacement based on GA-LSSVM with multiple factors. Bulletin of Engineering Geology and the Environment. https://doi.org/ 10.1007/s10064-015-0804-z

Chen J, Zeng Z, Jiang P, Tang H (2015) Deformation prediction of landslide based on functional network. Neurocomputing 149:151-157. https://doi.org/10.1016/j.neucom.2013.10.044

Chen M, Wang L. (1988) A prediction method by grey system for slope deformation and failure. In: Proceedings of the 5th international symposium on landslides, Lausanne. pp 577-582

Dou J, Paudel U, Oguchi T, et al. (2015) Shallow and deep-seated landslide differentiation using support vector machines: a case study of the Chuetsu area, Japan. Terrestrial, Atmospheric & Oceanic Sciences 26:227-239. https://doi.org/10.3319/tao.2014.12.02.07

Feng XT, Zhao H, Li S (2004) Modeling non-linear displacement time series of geo-materials using evolutionary support vector machines. International Journal of Rock Mechanics and Mining Sciences 41:1087-1107. https://doi.org/10.1016/j.ijrmms.2004.04.003

Gholgheysari Gorjaei R, Songolzadeh R, et al. (2015) A novel

PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. Journal of Natural Gas Science and Engineering 24:228-237. https://doi.org/10.1016/j.jngse.2015.03.013
Houck C, Joines J, Kay M (1995) A genetic algorithm for function optimization: a matlab implementation. NCSU-IE TR,
Jiang P, Chen J (2016) Displacement prediction of landslide based on generalized regression neural networks with k-fold cross-validation. Neurocomputing 198:40-47. http://orcid.org/10.1016/j.neucom.2015.08.118
Lepore C, Arnone E, Noto LV, et al. (2013) Physically based modeling of rainfall-triggered landslides: a case study in the Luquillo forest, Puerto Rico. Hydrology and Earth System Sciences 17:3371-3387. https://doi.org/10.5194/hess-17-3371-2013
Li B, Li D, Zhang Z, et al. (2015) Slope stability analysis based on quantum-behaved particle swarm optimization and least squares support vector machine. Applied Mathematical Modelling 39:5253-5264. https://doi.org/10.1016/j.apm.2015.03.032
Li XZ, Kong JM (2014) Application of GA–SVM method with parameter optimization for landslide development prediction. Natural Hazards and Earth System Science 14:525-533. https://doi.org/10.5194/nhess-14-525-2014
Lian C, Zeng Z, Yao W, Tang H (2012) Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine. Natural Hazards 66:759-771. https://doi.org/10.1007/s11069-012-0517-6
Lian C, Zeng Z, Yao W, Tang H (2015) Multiple neural networks switched prediction for landslide displacement Engineering Geology 186:91-99. https://doi.org/10.1016/j.enggeo.2014.11.014
Lian C, Zeng ZG, Yao W, Tang HM (2013) Displacement prediction of landslide based on PSOGSA-ELM with mixed kernel. 2013 Sixth International Conference on Advanced Computational Intelligence (Icaci): 52-57
Liu Y, Zhang YX (2014) Application of optimized parameters SVM in deformation prediction of creep landslide tunnel. Applied Mechanics and Materials 675-677: 265-268. https://doi.org/10.4028/www.scientific.net/AMM.675-677.265
Liu ZB, Xu WY, Zhang KP, Liu K (2009) Prediction of rock slope deformation based on optimized exponential smoothing method. Journal of Hohai Univeristy (Nat Sci) 37:313-316. (In Chinese)
Lv Y, Liu H (2012) Prediction of landslide displacement using grey and artificial neural network theories Advances in Science Letter 11:511-514
Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Engineering Geology 97: 171-191. https://doi.org/10.1016/j.enggeo.2008.01.004
Si L, Wang Z, Liu X, et al. (2016) Identification of shearer cutting patterns using vibration signals based on a least squares support vector machine with an improved fruit fly. optimization algorithm. Sensors 16(1): 90. https://doi.org/10.3390/s16010090
Su C, Wang L, Wang X, et al. (2015) Mapping of rainfall-induced landslide susceptibility in Wencheng, China, using support vector machine. Natural Hazards 76:1759-1779. http://orcid.org/10.1007/s11069-014-1562-0
Suykens J, Vandewalle J (1999) Least squares support vector machines classifiers. Neural Processing Letter 9:293-300
Suykens JAK, Gestel TV, Brabanter JD, et al. (2002) Least squares support vector machines. World Scientific, Singapore.
Vapnik V, Golowich SE, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems9:281-287
Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
Wang DD, Qiu GQ, Xie WB, Wang Y (2012) Deformation prediction model of surrounding rock based on GA-LSSVM-markov. Natural Science 04:85-90. https://doi.org/10.4236/ns.2012.42013
Xu F, Wang Y, Du J, Ye J (2011) Study of displacement prediction model of landslide based on time series analysis. Chinese Journal of Rock Mechanics and Engineering 30:746-751. (In Chinese)
Xu Q, Li W, Liu H, et al. (2015) Hysteresis effect on the deep-seated landslide by rainfall: the case of the kualiangzi landslide, china: 1557-1562. https://doi.org/10.1007/978-3-319-09057-3_276
Yin G, Zhang W, Zhang D, Kang Q (2007) Forecasting of landslide displacement based on exponential smoothing and nonlinear regression analysis Rock and Soil Mechanics 28:1725-1728. (In Chinese)
Zhang QX, Wu YP, Ou GZ, et al. (2014) Displacement prediction of Liangshuijing landslide based on time series additive model. Computer modelling & new technologies 18:215-223.
Zhou C, Yin K, Cao Y, Ahmed B (2016) Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. Engineering Geology 204: 108-120. https://doi.org/10.1016/j.enggeo.2016.02.009