A New Hybrid Method for Establishing Landslide Displacement Point Forecasting, Interval Forecasting and Probabilistic Forecasting

Hong Wang
Guizhou University
Guangyu Long
Guizhou University
Jianxing Liao (jxliao@gzu.edu.cn)
Guizhou University
Yan Xu
Jilin University
Yan Lv
Jilin University

Research Article

Keywords: landslide displacement prediction, double exponential smoothing, variational mode decomposition, long short-term memory network, gaussian process regression

Posted Date: August 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-839771/v1

Version of Record: A version of this preprint was published at Natural Hazards on November 16th, 2021.
See the published version at https://doi.org/10.1007/s11069-021-05104-x.

A new hybrid method for establishing landslide displacement point forecasting, interval forecasting and probabilistic forecasting

Hong Wang[1], Guangyu Long[1], Jianxing Liao[1*], Yan Xu[2], Yan Lv[2]
[1]College of Civil Engineering, Guizhou University, Guizhou Provincial key laboratory of rock and soil mechanics and engineering safety, Guiyang, Guizhou, China 550025
[2]Construction Engineering College, Jilin University, Xi Min Zhu Street, Changchun, Jilin, China 130026
hwang23@gzu.edu.cn, longguangyu21@163.com, jxliao@gzu.edu.cn, lvyan@jlu.edu.cn, xuyan8102@jlu.edu.cn

*Corresponding author (e-mail: jxliao@gzu.edu.cn)

Abstract: In addition to the inherent evolution trend, landslide displacement contains strong fluctuation and randomness, the omni-directional landslide displacement prediction is more scientific than single point prediction or interval prediction. In this work, a newly hybrid approach composed of double exponential smoothing (DES), variational mode decomposition (VMD), long short-term memory network (LSTM) and gaussian process regression (GPR), was proposed for point, interval and probabilistic prediction of landslide displacement. The proposed model includes two parts: (i) predicting the inherent evolution trend of landslide displacement by DES-VMD-LSTM; (ii) evaluating the uncertainty in the first prediction based on the GPR model. In the first part, DES is used to predict the trend displacement, VMD is used to extract the periodic and stochastic displacement from the residual displacement, and then LSTM is used to predict them. The triggering factors of periodic and stochastic displacement are screened by maximum information coefficient (MIC), and the screened factors are decomposed into low- and high-frequency components by VMD, to predict periodic and stochastic displacement respectively. The first cumulative displacement prediction results are achieved by adding the predicted trend, periodic and stochastic displacement. By setting the first predicted displacement as input and actual displacement as expected output, the point, interval and probability prediction of displacement are realized in GPR model. The plausibility of this method was validated firstly with the data from Bazimen (BZM) and Baishuihe (BSH) landslide in the Three Gorges Reservoir area. This model has potential capacity to realize deterministic prediction of displacement and exhibit uncertainty contained in displacement. A comparing study shows that this method has a high performance at point, interval and probability prediction of displacement.

Keywords: landslide displacement prediction; double exponential smoothing, variational mode

decomposition, long short-term memory network; gaussian process regression

1. Introduction

Landslide is a geological phenomenon of slope rock and soil sliding along the through shear failure surface, which is combining results of different environmental factors such as topography, lithology, hydrology and human activities, etc. The landslide not only lead to environmental deterioration, but damage the infrastructure and cause casualties. The Three Gorges Reservoir area is one of the areas prone to landslides, over 4200 landslides were recorded there (Yin et al., 2010). In this area, the extremely unstable hydrological environment, including the excessive rainfall and periodic reservoir water level, leads to frequently landslide disasters. For landslide, the displacement evolution is often used as an indicator to predict the behavior of failure process. High precision prediction can greatly reduce the loss brought by landslide disasters.

Accurate prediction of landslide displacement is very challenging, as the failure process of landslide is dynamic, nonlinear and stochastic (Cai et al., 2016). To date, there are three main approaches for landslide displacement prediction, involving physical model, statistical model and intelligent model (Ma et al., 2017). In the physical model, the displacement is predicted by using the model based upon creep theory and specific physical properties of geotechnical materials, for instance tertiary creep (Saito, 1965), Hayashi model (Setsuo et al., 1988) and Fukuzono model (Fukuzono, 1985). Statistical models use to predict landslide displacement with principle of time series analysis, like exponential smoothing model (Gould et al., 2008), autoregressive integrated moving average (Carlà et al., 2016), etc. Intelligent model developed on the basis of machine learning or deep learning algorithm is a promising method for landslide displacement prediction, such as back propagation neural network (BP) (Mayoraz and Vulliet, 2002), support vector regression (SVR) (Feng et al., 2004), extreme learning machine (ELM) (Lian et al., 2014). Recently, many hybrid models have been developed to improve the prediction accuracy of landslide displacement (Wen et al., 2017; Zhou et al., 2016).

The aforementioned methods mainly focus on how to obtain accurate point prediction, while ignore the uncertainty accumulated in landslide displacement prediction. Landslide displacement prediction contains various uncertainties, including measurement error, systematic error, inherent randomness, model uncertainty and so on (Ma et al., 2018). These uncertainties will weak the prediction accuracy in traditional point prediction model. Therefore, it is necessary to establish the probability prediction model for landslide displacement. The pre-existing probabilistic prediction models of landslide displacement

are established dominantly based on the prediction interval (PI) technology. For example, Lian et al. (Lian et al., 2016a) designed a PI model on foundation of switched neural network for landslide displacement prediction, by using bootstrap, kernel based extreme learning (KELM) machine and artificial neural network (ANN). Their follow-up study replaced KELM with random vector functional link networks (RVLFNs) to improve the quality of PI (Lian et al., 2018). Furthermore, by combining the lower and upper bound estimation (LUBE) model with other methods, a PI model is established to explain the uncertainty in landslide displacement prediction (Lian et al., 2016b; Wang et al., 2019). Although these models based on PI have been successfully applied to landslide displacement prediction, there are still some shortcomings. For example, the PI models based on bootstrap technology need massive resampling process, which significantly rise the computing complexity and the data volume. Moreover, these models focus mainly on the uncertainty of landslide displacement prediction namely interval prediction, while ignores the point prediction, which could not give a complete probability prediction result.

Gaussian process regression (GPR) is a new machine learning method based on Bayesian theory and statistical theory, which is suitable for complex regression problems such as high dimension, small sample size and nonlinear (Anjishnu et al., 2013). GPR performs well in landslide displacement prediction in different fields (Liu et al., 2014; Liu et al., 2012). Unlike the traditional method, GPR can directly predict the distribution of target sequence, wherein the point prediction, interval prediction and probability prediction results can be achieved at the same time. Since GPR focus major on the error of prediction subject, it alone is difficult to obtain ideal prediction during landslide displacement prediction. Combining with other methods could improve its unique probabilistic prediction characteristics.

The total landslide displacement curve in the Three Gorges Reservoir area contains trend, periodic and stochastic components (Huang et al., 2016; Li et al., 2012; Lian et al., 2015). To achieve high-precision prediction results, decomposing the total displacement into these three components and predicting individually is necessary. The combination of DES and VMD is able to decompose superior time series. DES can decompose landslide displacement into linear and nonlinear components. The linear component is trend displacement that can be predicted directly based on historical displacement, The nonlinear component including periodic and stochastic item are accurately separated with the help of VMD, and are predicted independently by using LSTMs. The LSTM is widely favored by researchers because of its high performance in dealing with time series problems. The LSTM model has a excellent

performance in landslide displacement prediction (Xu and Niu, 2018).

In this paper a new landslide displacement prediction model DES-VMD-LSTM-GPR is proposed. The main research contents are as follows: (1) establishing a more accurate point prediction model based on DES-VMD-LSTM; (2) introducing a GPR model to assess the uncertainty in the first prediction, as well as to achieve interval and probability prediction; (3) the proposed method is applied to study BZM and BSH landslide, and its effectiveness is validated through comparing with different prediction methods.

2. Materials and methods

2.1 Methodology

2.1.1 Double exponential smoothing

The DES is an important method of time series prediction, which can accurately extract the trend characteristics of time series (Holt, 2004). The basic idea of DES that all historical data have an impact on the forecast data, but the recent data has a greater impact than older data, and the influence changes geometrically with time (Wu et al., 2016). The three main formulas of DES are as follows:

$$\begin{cases} S_t = \alpha Y_t + (1-\alpha)(S_{t-1} - b_{t-1}) \\ b_t = \beta(S_t - S_{t-1}) + (1-\beta)b_{t-1} \\ T_{t+m} = S_t + mb_t \end{cases} \quad (1)$$

Where S_t and b_t represent stable and trend component of series at time t; Y_t is the is the observation; α and β are smoothing parameters within $[0,1]$; T_{t+m} is the prediction value at time $t+m$, where $\mathrm{m} = 1$.

2.1.2 Variational mode decomposition

The VMD is an adaptive, completely non recursive modal decomposition and signal processing method, which can decompose the time series with high complexity and strong nonlinearity into the specified number of intrinsic mode functions (IMFs) (Dragomiretskiy and Zosso, 2014). The core idea of VMD is to construct and solve variational problems. Firstly, the variational problem is constructed, i.e., assuming that the original sequence Y is decomposed into K specified IMF components, each IMF sequence is guaranteed to have a central frequency and a finite bandwidth, and the sum of estimated bandwidth for IMF is the minimum, and the constraint condition is that the sum of all IMF is equal to the original sequence Y, then the corresponding constraint variational expression is as follows:

$$\begin{cases} \min\limits_{\{u_k\},\{w_k\}} \left\{ \sum\limits_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \\ s.t. \sum\limits_{k=1}^{K} u_k = Y \end{cases} \tag{2}$$

Where K is the number of specified mode decomposition, $\{u_k\} = \{u_1, u_2, \cdots, u_K\}$ and $\{w_K\} = \{w_1, w_2, \cdots, w_K\}$ the K mode components IMF and their corresponding central frequencies after decomposition, $\delta(t)$ is the Dirac delta function.

To solve Eq. (2), the Lagrange multiplication operator λ is introduced to transform the constrained variational problem into the unconstrained variational problem, and the expression is as follows:

$$L(\{u_k\},\{w_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 + \left\| Y(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), Y(t) - \sum_k u_k(t) \rangle \tag{3}$$

Where α is the quadratic penalty factor, which can reduce the interference of gaussian noise. By using the alternating direction multiplier iterative algorithm combined with Parseval/Plancherel and Fourier equidistant transform, the u_k, w_k and λ are obtained by alternating optimization iteration:

$$\hat{u}_k^{n+1}(w) \leftarrow \frac{\hat{Y}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \tag{4}$$

$$w_k^{n+1} \leftarrow \frac{\int_0^\infty w |\hat{u}_k^{n+1}(w)|^2 dw}{\int_0^\infty |\hat{u}_k^{n+1}(w)|^2 dw} \tag{5}$$

$$\hat{\lambda}^{n+1}(w) \leftarrow \hat{\lambda}^n(w) + \gamma \left(\hat{Y}(w) - \sum_k \hat{u}_k^{n+1}(w) \right) \tag{6}$$

Where γ the noise tolerance, which meets the fidelity requirements of signal decomposition, $\hat{u}_k^{n+1}(w)$, $\hat{u}_i(w)$, $\hat{Y}(w)$ and $\hat{\lambda}(w)$ correspond to Fourier transform of $u_k^{n+1}(t)$, $u_i(t)$, $Y(t)$ and $\lambda(t)$ respectively.

Based on the above principles, the solution process of VMD algorithm is summarized as follows:

(1) Initialization $\hat{u}_k^1$, w_k^1, λ^1 and the maximum number of iterations N, n ← 0;

(2) Update $\hat{u}_k$ and w_k with formulas (4) and (5);

(3) Update $\hat{\lambda}$ with formulas (6);

(4) Repeat steps (2) and (3) until $\frac{\sum_k \|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} < e$ is satisfied, where criterion e > 0

2.1.3 Maximum information coefficient

Maximum information coefficient (MIC) is an important unsupervised feature extraction method, which can be used to reveal the degree of association between two random variables, including linear and non-linear relationship, which has the characteristics of breadth and fairness (Reshef et al., 2011). Assuming that there are two random variables $\mathrm{A} = [a_i, i = 1,2,\cdots,n]$ and $\mathrm{B} = [b_i, i = 1,2,\cdots,n]$, where n is the sample number. Then, the MIC between A and B is calculated as follows:

(1) The data of A and B are taken out to form data set $\mathrm{D} = [(a_i, b_i), i = 1,2,\cdots,n]$, and the data set D is sorted in a certain order;

(2) The data set D is mapped to a two-dimensional plane. Then, the A is divided into x portion and the B is divided into y portion to form the grid G. For a fixed D, different partition methods can obtain G with different grid number $x * y$, the grid number should satisfy $x * y < n^{0.6}$. Based on this, different distribution $D|_G$ of dataset D can be obtained under different partition conditions;

(3) For different $D|_G$, the corresponding MI, i.e. $I(D|_G)$, is calculated by Eq. (7), where $P_{AB}(a,b)$ is the joint probability distribution of A and B; $P_A(a)$ and $P_B(b)$ are the marginal probability distribution of A and B, respectively;

$$I(A;B) = \sum_{a\epsilon A}\sum_{b\epsilon B} P_{AB}(a,b)\log\frac{P_{AB}(a,b)}{P_A(a)P_B(b)} \quad (7)$$

(4) Find out the maximum value $\max I(D|_G)$ of MI in all cases of partition, let $I'(D,x,y) = \max I(D|_G)$, and standardize it as follows:

$$M(D)_{xy} = \frac{I'(D,x,y)}{\log\min\{x,y\}} \quad (8)$$

According to $M(D)_{xy}$, the MIC of random variables A and B can be obtained as follows:

$$\mathrm{MIC}(D) = \max\{M(D)_{xy}\} \quad (9)$$

When $\mathrm{MIC}(D) = 0$, it means that variable A and B do not depend on each other. On the contrary, when $\mathrm{MIC}(D)$ value is close to 1, there is a strong linear or nonlinear correlation between these two variables.

2.1.4 Long short-term memory network

The LSTM is a variant of recurrent neural network (RNN) proposed by Hochreiter and Schmiduber (Hochreiter and Schmidhuber, 1997). By improving the traditional RNN, LSTM solved the problem of

vanishing and explosion gradient of RNN and realized the long-term memory of information. Due to its excellent performance in time series, LSTM has been successfully applied in many fields, such as speech recognition, natural language processing, wind power prediction and so on(Devi and Thongam, 2020; Yu et al., 2019; Zhang et al., 2019).

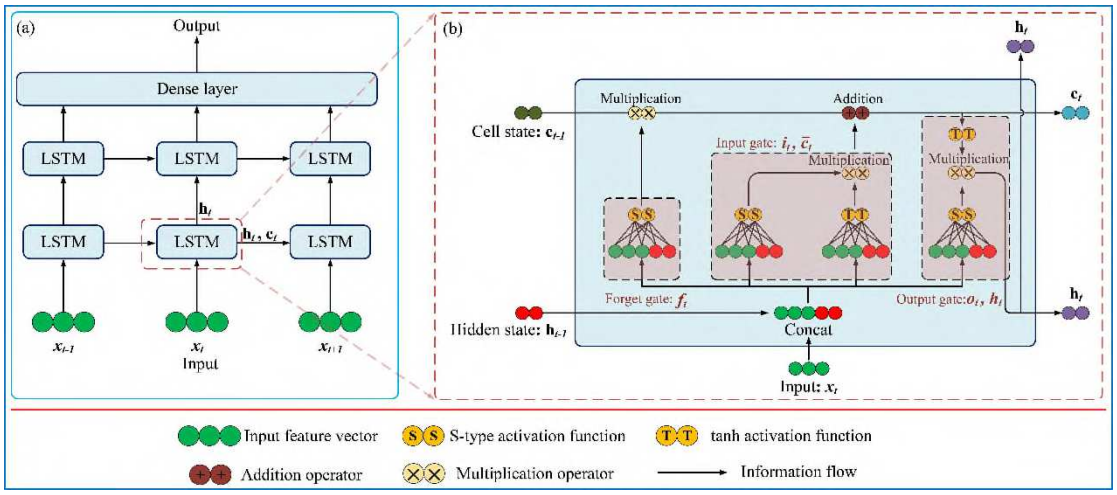


Fig 1. (a) LSTM network; (b) LSTM cell structure

The reason why LSTM can deal with the long-term dependencies problem is that it is equipped with several gates to control the magnitude based on the standard RNN, and these gates can control the flow of information. As shown in Fig. 1(b), the input gate controls how much input information can enter the memory cell at the current time, and the output gate controls the output information entering other cells or as the final result. The forget gate controls the retention degree of data information from the previous moment. Three gates interaction controls memory cell output status. The functions of the three gates are defined as follows:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \tag{11}$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \tag{12}$$

where w_i, w_f, w_o represents the weights; b_i, b_f, b_o are their corresponding bias values; h_{t-1} are output value of the memory cell at the previous moment; x_t are input value at the current time; i_t, f_t, o_t are the values of the input, forget, and output gates, respectively.

2.1.5 Gaussian process regression

GPR is an uncertainty model, which describes the regression problem from the perspective of probability(Carl Edward and Williams, 2005). Different from the traditional deterministic regression model, GPR directly models the function $f(x)$ and obtains the distribution of function $f(x)$. Therefore,

GPR can explore more possible regions of data than traditional models, so as to solve the problem of interval and probabilistic prediction. The general model of regression and prediction problem is shown as follows:

$$Y = f(x) + \varepsilon, \varepsilon \sim N(0, \sigma_n^2 I_n) \tag{13}$$

where Y is the observation with noise and X is the input eigenvector, ε is the observation noise with gaussian distribution, σ_n^2 is the noise variance and I_n is the unit matrix.

GPR can accurately describe $f(x)$ by mean function $m(x)$ and kernel function (covariance function) $k(x, x)$ after learning observation data. The expression is as follows:

$$f(x) \sim GP\big(m(x), k(x, x)\big) \tag{14}$$

Where:

$$\begin{cases} m(x) = E[f(x)] \\ k(x, x) = E[\big(f(x) - m(x)\big)\big(f(x) - m(x)\big)^T] \end{cases} \tag{15}$$

$f(x)$ is a random variable composed of multi-dimensional gaussian distribution. According to the properties of multi-dimensional gaussian distribution, the prior distribution of observation Y and the joint distribution of observed value Y and predicted value y can be obtained by Eq. (13), (14) and (15), as follows:

$$Y = N(0, k(x, x) + \sigma_n^2 I_n) \tag{16}$$

$$\begin{bmatrix} Y \\ y \end{bmatrix} \sim N\left(0, \begin{bmatrix} k(x, x) + \sigma_n^2 I_n & k(x, x_*) \\ k(x_*, x) & k(x_*, x_*) \end{bmatrix}\right) = N\left(0, \begin{bmatrix} k + \sigma_n^2 I_n & k_*^T \\ k_* & k_{**} \end{bmatrix}\right) \tag{17}$$

In general, the mean is subtracted when GPR data is preprocessed, that is, $m(x) = 0$. In Eq. (16) and (17), $k(x, x) = [k_{ij}]$ is a symmetric positive definite covariance matrix, k_{ij} represents the correlation between x_i and x_j, which can be calculated by a kernel function. $k(x, x_*) = k(x_*, x)^T$ is the covariance matrix between training set x and test set x_*, and $k(x_*, x_*)$ represents the covariance matrix of test set itself.

When the training set $D = [x_i, Y_i]$ is known, the posterior distribution of prediction value y corresponding to test set x_* can be obtained as follows:

$$p(y|D, x_*) \sim N\big(m(y), k(y, y)\big) \tag{18}$$

Where:

$$\begin{cases} m(y) = \bar{y} = k_*[k + \sigma_n^2 I_n]^{-1} Y \\ k(y, y) = \sigma_y^2 = k_{**} - k_*[k + \sigma_n^2 I_n]^{-1} k_*^T \end{cases} \tag{19}$$

In Eq. (19), $\bar{y}$ is the mean value of predicted value y corresponding to test set x_*, σ_y^2 is the variance of y. The interval prediction results correspond to 95% confidence level, i.e. $[\bar{y} - 1.96\sigma_y, \bar{y} +$

$1.96\sigma_y$]. The expression of probability prediction result for i-th prediction value y_i is as follows:

$$p(y_i) = \frac{1}{\sqrt{2\pi}\sigma_{y_i}} \exp\left(-\frac{(y_i - \bar{y}_i)^2}{2\sigma_{y_i}^2}\right) \tag{20}$$

According to the above process, it can be found that the kernel function plays a decisive role in the prediction performance of GPR. Common kernel functions include Radial-basis function kernel (RBF), Matern kernel, Rational quadratic kernel, Exp-sine-squared kernel, etc. The expression of RBF kernel is as follows, l_1 and l_2 are the hyperparameters to be optimized.

$$k_{ij} = l_1 \exp\left(-\frac{(x_i - x_j)^2}{2l_2}\right) \tag{21}$$

2.2 Materials

2.2.1 Bazimen landslide

Bazimen landslide (BZM) is located at the estuary of the right bank of Xiangxi River, a tributary of the north bank of the Yangtze River in Guizhou town, Zigui County, 31 km away from the Three Gorges dam (Fig. 2(a)). The BZM landslide mass is dustpan-shaped with the west higher than the east, and the main sliding direction is about N110°. The plane shape of the landslide is irregular fan-shaped, with homologous gullies developed on both sides of the landslide, and the back edge is in the shape of an armchair. See Yang et al. (2019) for more information.

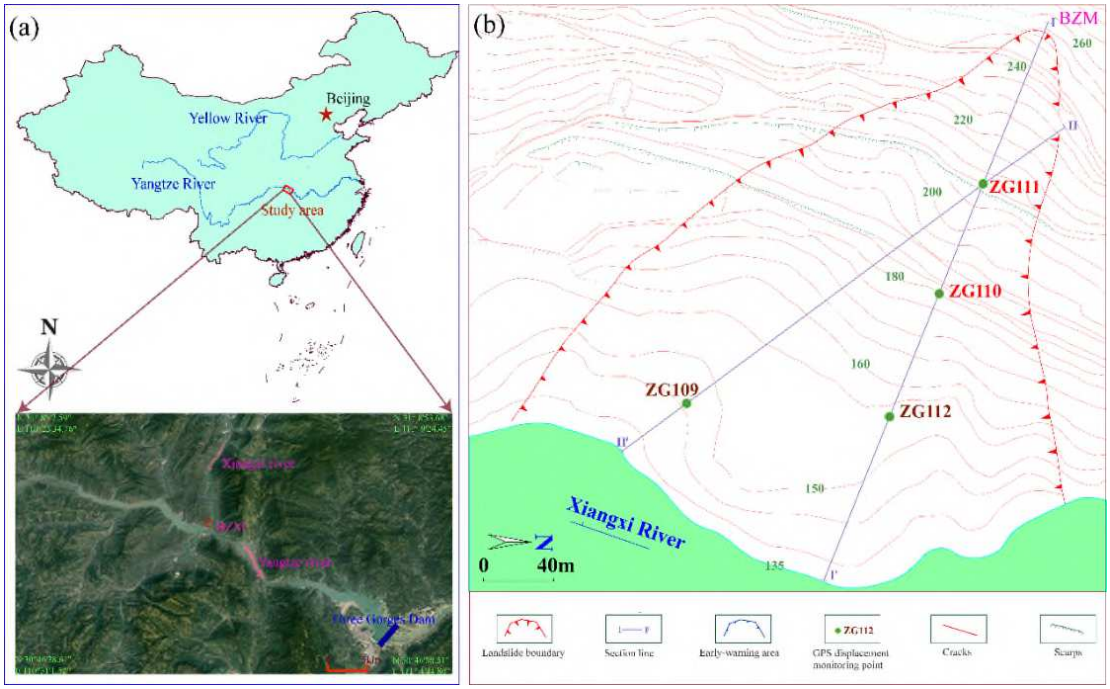


Fig 2. Geographical location and topographic map of Bazimen landslide (BZM)

In order to monitor the displacement of bank landslide in real time in Three Gorges Reservoir area,

many GPS stations are deployed on the slope surface. As shown in Fig. 2(b), 4 GPS stations were set up in the sliding area of BZM. In the past few years, a large number of displacement monitoring data have been collected through these GPS stations. These data were collected and stored in the National Cryosphere Desert Data Center (Wu, 2016; Wu and Haifeng, 2016). Fig. 3 showed the displacement, rainfall and reservoir level data for BZM from January 2007 to December 2012. According to the displacement curve, it can be found that BZM landslide are still in unstable state and the deformation is increasing continuously during the entire monitoring period. The deformation of BZM landslide increases in stepwise, which is relate to the periodic change reservoir water level and rainfall. The displacement of the four monitoring points increased rapidly from May to July, which was physically consistent with the annual flood period and reservoir water level falling time. On the contrary, the displacement of BZM changes little in non-flood and reservoir impoundment period, and their deformation tends to be stable.

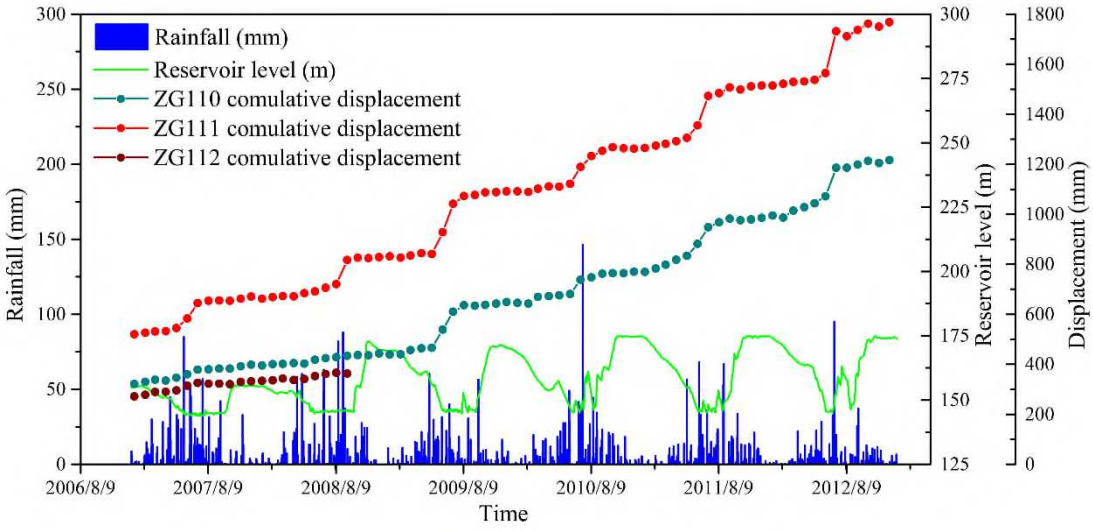


Fig 3. Displacement monitoring curve of Bazimen (BZM) landslide

2.2.2 Landslide displacement decomposition

The landslide displacement curve is a non-stationary time series that increases obviously with time. Many researches have shown that the landslide displacement curve can be divided into three components, namely trend component $T(t)$, periodic component $P(t)$ and stochastic component $S(t)$, as shown in Eq. (17). Among them, the $T(t)$ increases monotonically with time and is controlled by lithology, topography and other factors; the $P(t)$ is approximately a periodic function with time, which is affected by seasonal rainfall and periodic regulation of reservoir water level; the $S(t)$ is a near white noise sequence, which is affected by wind load, non-seasonal rainfall and temporary regulation of reservoir

water level, etc.

$$Y(t) = T(t) + P(t) + S(t) \quad (22)$$

In order to clarify the physical meaning of each component in the landslide displacement time curve and accurately reflect the evolution characteristics of each component, this paper uses the combination of DES and VMD to decompose the landslide displacement curve into trend, periodic and stochastic component. The DES method is used to extract and predict trend displacement. In this work, the model parameters α= 0.95, β= 0.05, and the stable variable S_1 and trend variable b_1 are initialized as Y_2 and $Y_2 - Y_1$ respectively. As shown in Fig 4, the prediction results of trend displacement are basically consistent with the evolution characteristics of cumulative displacement, increasing monotonically with time and having a certain lag. The residual displacement fluctuates with time and has the features of periodicity and randomness. Then the VMD method (α=300, γ =0.005， K=2) is used to correctly decompose periodic and stochastic components from the residual displacement.

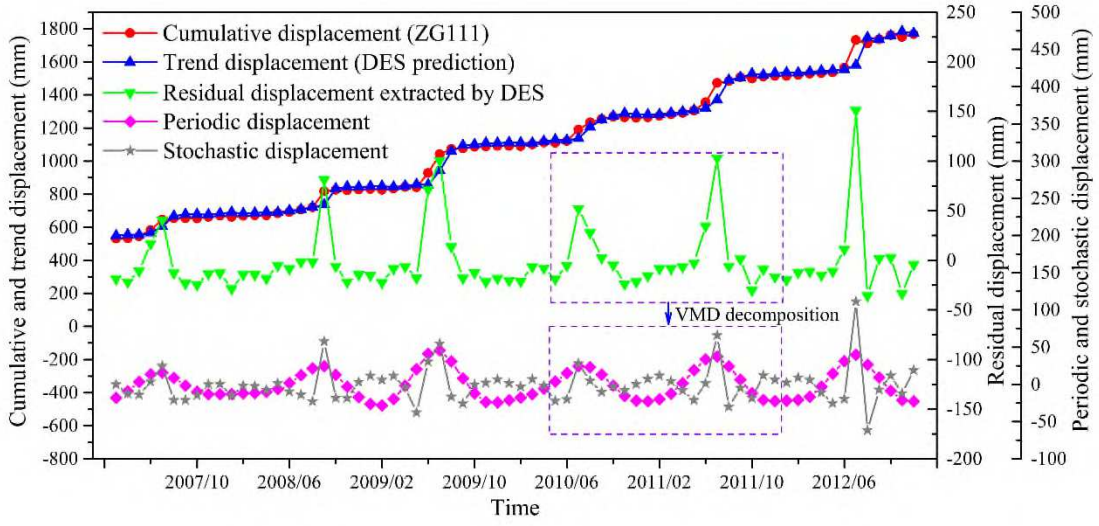


Fig 4. The decomposition results of trend displacement and residual displacement based on DES

2.2.3 Triggering factor selection and decomposition

Selecting appropriate triggering factors is the precondition for high prediction accuracy. First, the candidate factors are determined, and then the triggering factors are filtered from the candidate factors according to the MIC value. As described in section 2.2.2, the deformation of landslide is mainly related to historical deformation, rainfall and reservoir level fluctuation. Furthermore, Fig 2 indicated that the rainfall and reservoir level factors exert a hysteresis effect on landslide displacement. Therefore, 12 variables (I_1-I_{12}) including the maximum continuous precipitation, cumulative precipitation and reservoir

level average elevation of the current month, the previous month and the first two months are selected as candidate factors, as shown in Table 1.

Table 1. Candidate factors and its MIC value with periodic and stochastic displacement

Factor	Connotation	MIC	
		Periodic displacement	Stochastic displacement
I_1	Maximum continuous precipitation in current month	0.654	0.406
I_2	Maximum continuous precipitation in the previous month	0.528	0.300
I_3	Maximum continuous precipitation in the first two months	0.372	0.278
I_4	Cumulative precipitation in current month	0.654	0.392
I_5	Cumulative precipitation in the previous month	0.523	0.263
I_6	Cumulative precipitation in the first two months	0.373	0.304
I_7	Variation range of reservoir level in a single month	0.395	0.349
I_8	Variation range of reservoir level in two months	0.364	0.247
I_9	Average elevation of reservoir level in the current month	0.601	0.532
I_{10}	Average elevation of reservoir level in the previous month	0.666	0.389
I_{11}	Average elevation of reservoir level in the first two months	0.423	0.357
I_{12}	Cumulative displacement increment in the previous month	0.483	0.308

To screen the trigger factors from these candidate factors, the MIC value between candidate factors and the target displacement is calculated. For periodic displacement, MIC value greater than 0.4 is used as trigger factor, and for stochastic displacement, MIC value greater than 0.3 is regarded as trigger factor. Finally, I_1, I_2, I_4, I_5, I_9, I_{10}, I_{11}, I_{12} are chosen for the periodic displacement prediction task, and I_1, I_2, I_4, I_6, I_7, I_9, I_{10}, I_{11}, I_{12} are chosen for the stochastic displacement prediction task.

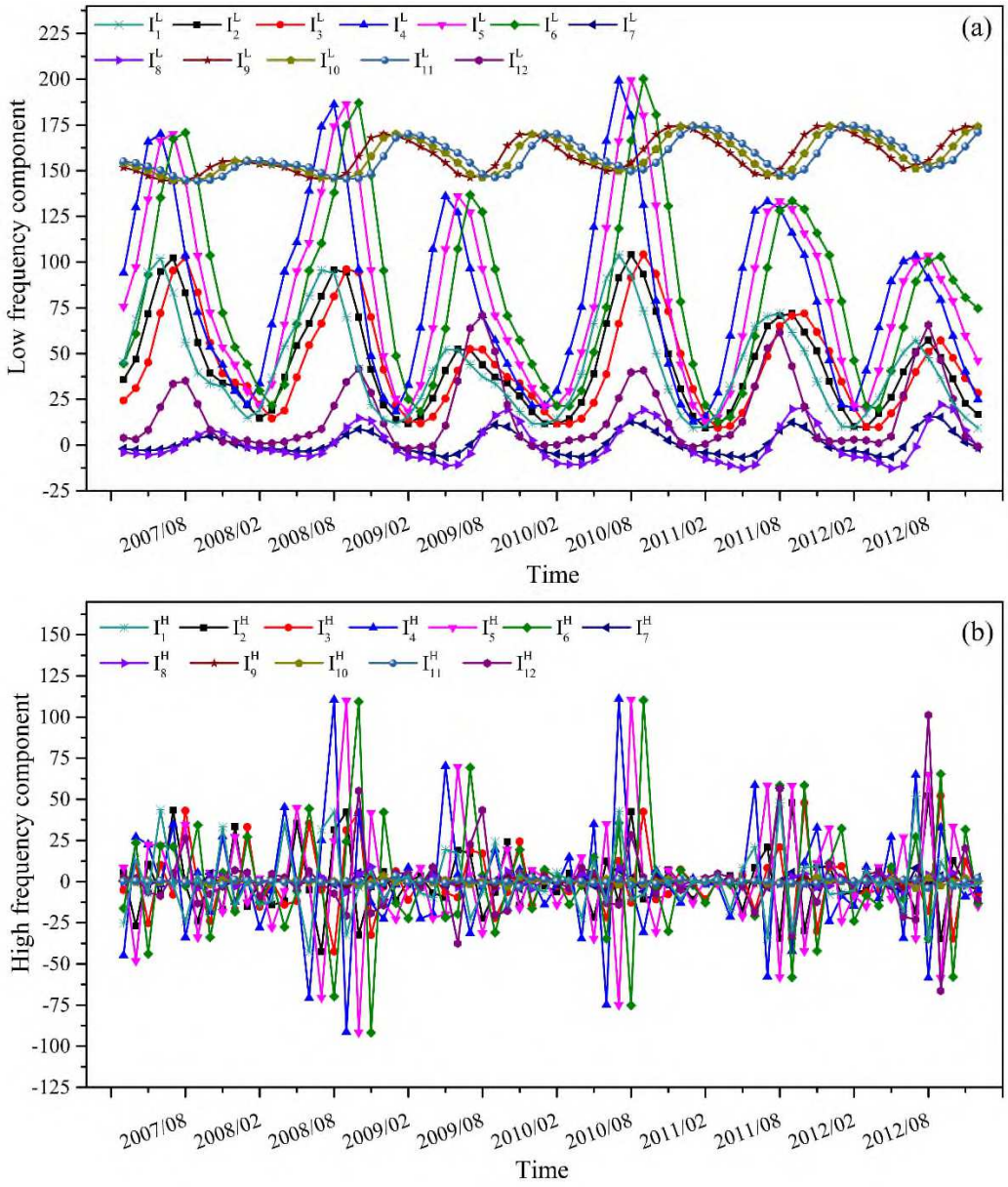


Fig 5. Factor I_1-I_{12} VMD decomposition. (a): low frequency component, (b): high frequency component

Corresponding to the periodic and stochastic displacement, the time series of trigger factors can also be decomposed into two subsequences by VMD method, as shown in Fig 5. The high-frequency component has the characteristics of periodic variation, while the low-frequency component fluctuates with time. Fig. 6 shows the relationship between low and high frequency components of I_1 and periodic and stochastic displacement. The variation characteristics of low-frequency component and periodic displacement are generally the same, and the evolution trend of high- frequency component and stochastic displacement is similar. Therefore, this paper will use the high-frequency component of the trigger factor to predict the periodic displacement and its low-frequency component to predict the stochastic displacement.

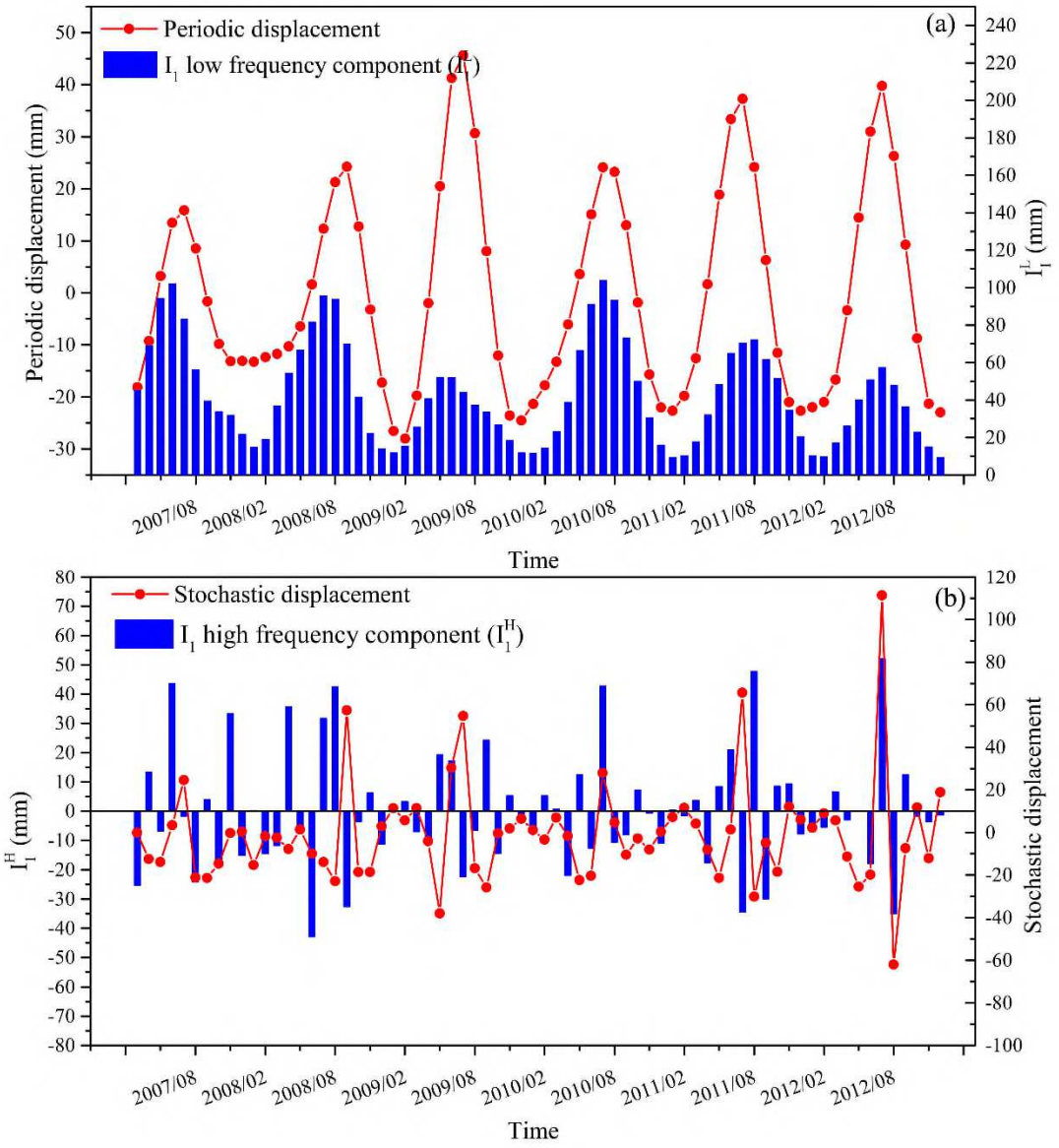


Fig 6. (a) Relationship between low frequency component of factor I_1 and periodic displacement; (b) relationship between high frequency component of factor I_1 and stochastic displacement.

2.3 Prediction process

Fig. 7 describes the analytical process of the predict model in the form of a flowchart. The main steps are as follows:

(1) Collecting landslide displacement monitoring data (Y);

(2) Prediction the trend displacement (T) based on DES;

(3) Decomposing periodic displacement (P) and stochastic displacement (S) from residual displacement based on VMD;

(4) Screening triggering factors based on MIC, and decomposition these factors by VMD;

(5) Establishing periodic and stochastic displacement prediction models based on LSTM;

(6) The first cumulative displacement prediction result is obtained by summing the prediction results of trend, periodic and stochastic displacement;

Finally, the GPR model is established to realize the point, interval and probability prediction of landslide displacement by taking the predicted results of the sixth step as the input feature and the actual monitoring displacement as the expected output.

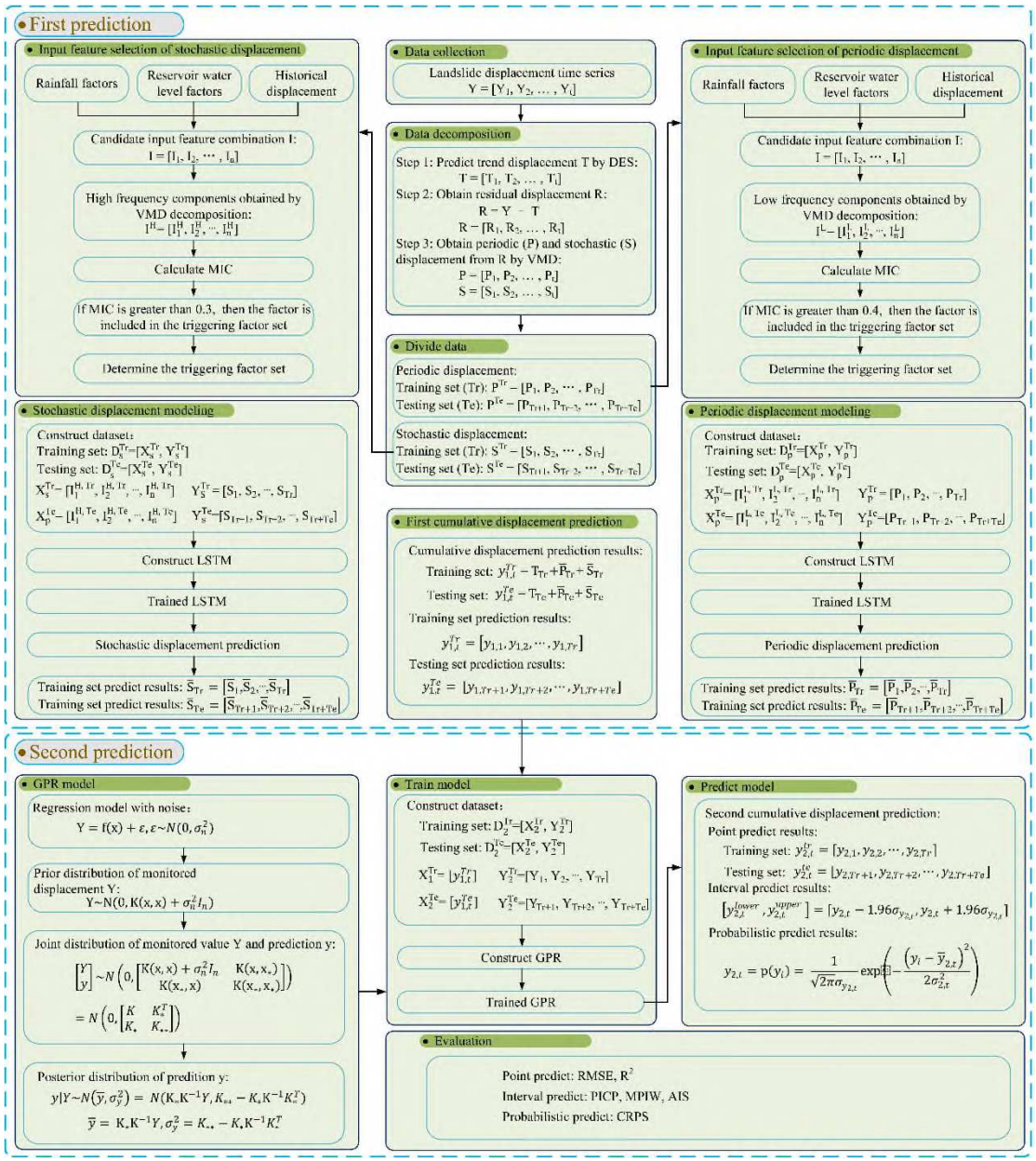


Fig 7. Flow chart of landslide displacement point, interval and probabilistic prediction

2.3.1 Method parameter setting

DES, VMD, LSTM and GPR are the main components of the hybrid method proposed. To test the effectiveness, BP, SVR and GRU are selected to replace LSTM for comparative analysis. In these

methods, the parameter values of DES and VMD methods have been determined in Section 2.2.2 and 2.2.3, and the main parameters and their values of the remaining five methods are listed in Table 2. Among them, the epochs of training (Ep) and layers (L) are determined by multiple debugging, and other parameters are mainly determined by optimization algorithm. For example, the number of hidden layer nodes (n_h), number of and batch size (Ba) in LSTM, GRU, BP are determined by Gridsearch algorithm; the epsilon parameter in SVR is obtained by Randomsearch algorithm; and the kernel function parameter l_1 and l_2 in GPR are mainly optimized by maximizing the log-marginal-likelihood. The BP, LSTM and GRU method are trained by Adam optimization algorithm, and Gridsearch and Randomsearch use 10-fold cross validation to obtain optimal parameters. All methods are implemented in Python version 3.7, and the main application packages include Scikit-learn, Keras, Minepy and Statsmodels.

Table 2. Method parameter

Method	Parameter	Meaning	Value		Reason
			Periodic	Stochastic	
LSTM	L	number of layers	2	4	common value $[1,2,3,4,\cdots]$
	n_i	number of input layer nodes	-	-	number of the input variable
	n_h	number of hidden layer nodes	8,4	16,12,8,4	common value $[2,4,6,8,\cdots]$
	n_o	number of output layer nodes	1	1	time series regression
	Ba	batch size	16	16	common value $[8,16,32,\ldots]$
	Ep	epochs of training	2000	2000	converged
BP	L	number of layers	2	4	the same as LSTM
	n_i	number of input layer nodes	-	-	the same as LSTM
	n_h	number of hidden layer nodes	8,4	24,12,8,4	the same as LSTM
	n_o	number of output layer nodes	1	1	the same as LSTM
	Ba	batch size	16	16	the same as LSTM
	Ep	epochs of training	2000	2000	the same as LSTM
GRU	L	number of layers	2	4	the same as LSTM
	n_i	number of input layer nodes	-	-	the same as LSTM
	n_h	number of hidden layer nodes	8,4	16,12,8,4	the same as LSTM
	n_o	number of output layer nodes	1	1	the same as LSTM
	Ba	batch size	16	16	the same as LSTM
	Ep	epochs of training	2000	2000	the same as LSTM
SVR	k	kernel function	Poly	Poly	a sklearn kernel function
	c	parameter in RBF function	20	400	debug acquisition
	degree	parameter in RBF function	3	5	common value $[2,3,4,5,\cdots]$
	epsilon	parameter in RBF function	0.00005	0.00005	obtained by Randomsearch in $[-5,5]$
GPR	k	kernel function	RBF		a competitive kernel function
	l_1	parameter in gaussian function	3.1e3		value in $[1e-4, 1e4]$
	l_2	parameter in gaussian function	0.01		value in $[1e-4, 1e4]$

2.3.2 Prediction performance measure

2.3.2.1 Point prediction metrics

The root mean square error (RMSE) and coefficient of determination (R^2) are used to assess the performance of point predict accuracy, which can be expressed by the following:

$$\mathrm{RMSE} = \sqrt{\frac{1}{Te}\sum_{t=1}^{Te}(y_t - Y_t)^2} \tag{23}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{Te}(y_t - Y_t)^2}{\sum_{t=1}^{Te}(Y_t - \bar{Y}_t)^2} \tag{25}$$

Where *Te* is the sample number of test set, y_t is predicted value, Y_t and $\bar{Y}_t$ are actual monitoring value and its mean value, respectively.

2.3.2.2 Interval prediction metrics

Two common indexes are used to evaluate the interval prediction quality of GPR model, including coverage probability (PICP) and mean prediction interval width (MPIW). The expression is as follows:

$$\mathrm{PICP} = \frac{1}{Te}\sum_{t=1}^{Te} C_t \tag{26}$$

$$C_t = \begin{cases} 1 & Y_i \in [L_t^\alpha, U_t^\alpha] \\ 0 & Y_i \notin [L_t^\alpha, U_t^\alpha] \end{cases} \tag{27}$$

$$\mathrm{MPIW} = \frac{1}{Te}\sum_{t=1}^{Te}[U_t^\alpha - L_t^\alpha] \tag{28}$$

In Eq. (26), (27) and (28), L_t^α and U_t^α represent the upper and lower limit of the prediction interval at t time step under the confidence level of α; C_t is a Boolean variable, which represents the number of observations falling into the prediction interval $[L_t^\alpha, U_t^\alpha]$.

In addition to PICP and MPIW, another comprehensive index average interval score (AIS) is used to measure the quality of the forecast interval (Winkler, 1972). The interval score (IS) of prediction interval at time t is defined as follows:

$$S^{(\alpha)}(t) = \begin{cases} -2\alpha I_t^\alpha - 4(L_t^\alpha - Y_t), & if\ Y_t < L_t^\alpha \\ -2\alpha I_t^\alpha & if\ Y_t \in [L_t^\alpha, U_t^\alpha] \\ -2\alpha I_t^\alpha - 4(Y_t - U_t^\alpha), & if\ Y_t > U_t^\alpha \end{cases} \tag{29}$$

Where I_t^α is the width of the prediction interval at time t:

$$I_t^\alpha = U_t^\alpha - L_t^\alpha \tag{30}$$

Thus, the AIS is calculated as follows:

$$\overline{S^{(\alpha)}} = \frac{1}{Te}\sum_{t=1}^{Te} S^{(\alpha)}(t) \tag{31}$$

2.3.2.3 Probabilistic prediction metrics

Continuous ranked probability score (CRPS) is an effective index to evaluate the accuracy of probabilistic prediction (Alessandrini et al., 2015). CRPS can be defined as follows:

$$\begin{cases} \mathrm{CRPS} = \dfrac{1}{Te}\displaystyle\sum_{t=1}^{Te}\int_{-\infty}^{+\infty}[F(y_t) - H(y_t - Y_t)]^2 dy_t \\ F(y_t) = \displaystyle\int_{-\infty}^{y_t} p(x)dx \end{cases} \tag{32}$$

$$H(y_t - Y_t) = \begin{cases} 1 & (y_t < Y_t) \\ 0 & (y_t \geq Y_t) \end{cases} \tag{33}$$

Where $p(y_t)$ and $F(y_t)$ are the probability density function and cumulative distribution function of t time step prediction value, respectively; $H(y_t - Y_t)$ is a Boolean variable. The relatively low values of CRPS indicate high-performance of probabilistic prediction.

3. Results and discussion

72 displacement data points were collected from 2007 to 2012, which are divided into two parts: 60 displacement data in 2007-2011 as training set and 12 data in 2012 as testing set. Since the prediction results of trend displacement have been analyzed in section 2.2.2, the following part mainly discusses the prediction results of periodic, stochastic and cumulative displacement. To maximally preserve the distribution of the data, the original data is normalized to $[-1,1]$ using linear normalization for training and prediction.

3.1 Periodic and stochastic displacement prediction

Due to the results of each run are slightly different, the averaged value of run 5 times is taken as the final results. Fig. 8 shows the prediction results of periodic and stochastic displacement based on LSTM, GRU, SVR and BP. Overall, the prediction effect is considerable. Table 3 listed the prediction accuracy and running time of the four methods. The RMSE and R^2 values of these predictors are very close, which demonstrate that the four methods are highly competitive. Considering all evaluation criteria, the dynamic approaches (LSTM and GRU) is higher accuracy than that of static approaches (BP and SVR), as the memory function make LSTM and GRU more professional in dealing with time series problem. Besides, the prediction accuracy of GRU is slightly higher than LSTM, because GRU is extended from LSTM through combining forget gate and input gate, and the model parameters are reduced by about 1/3. Although the prediction accuracy obtained by the dynamic algorithm is high, it is at the expense of time. Table 3 shows that the running time of LSTM for periodic displacement prediction is about 29 times that

of SVR and 8 times that of BP.

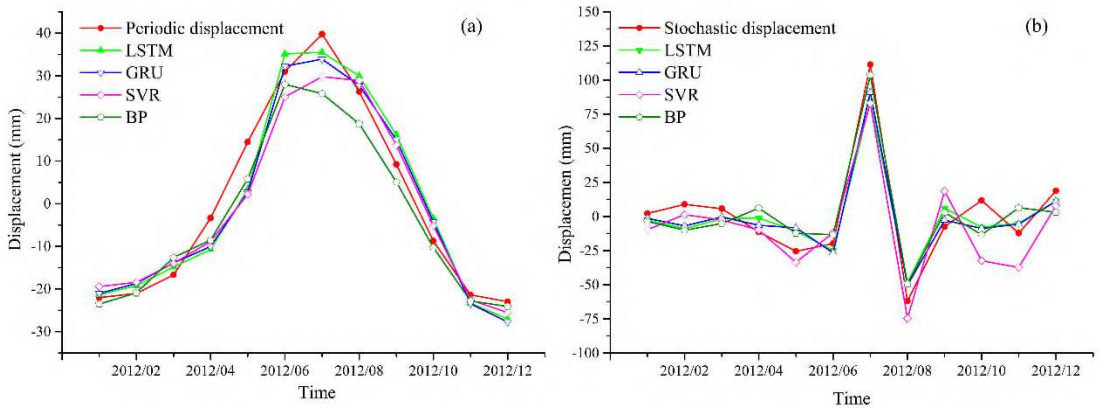


Fig 8. Prediction results of periodic and stochastic displacement based on LSTM, GRU, SVR, and BP

Table 3. Prediction accuracy of periodic and stochastic displacement

Method	Statistics	Periodic displacement			Stochastic displacement		
		RMSE/mm	R^2	Time/s	RMSE/mm	R^2	Time/s
LSTM	Mean	5.344	0.944	29.506	21.075	0.437	51.718
	Max	5.910	0.957	31.680	24.631	0.825	53.250
	Min	4.602	0.928	26.980	16.045	0.140	50.490
GRU	Mean	5.204	0.940	26.874	20.331	0.615	47.036
	Max	6.183	0.954	27.950	23.048	0.683	47.720
	Min	4.134	0.919	25.600	16.045	0.546	45.470
SVR	Mean	5.675	0.918	<1	25.744	0.624	<1
	Max	5.809	0.922	<1	26.913	0.707	<1
	Min	5.407	0.915	<1	25.095	0.570	<1
BP	Mean	5.899	0.898	8.666	24.881	0.395	10.596
	Max	6.346	0.951	8.930	29.233	0.630	12.240
	Min	4.393	0.884	8.420	21.548	0.170	10.110

Another phenomenon can be found from Table 3, that is, the prediction accuracy of periodic displacement is higher than that of stochastic displacement, and the prediction stability of stochastic displacement is poor. The possible reasons could be: (1) Stochastic displacement is affected by many factors, such as human activities, observation error, temporary reservoir water level regulation, non-seasonal rainfall, etc. which are not entirely considered in this work; (2) Stochastic displacement contains many strong randomness and uncertainty, the existing static or dynamic algorithms could not capture the potential laws accurately.

3.2 Point prediction

The cumulated displacement point prediction include two parts, the first is the sum of trend, periodic and stochastic component prediction, and the second is GPR modeling prediction. As illustrated in Fig.

9(a) and (b), these two prediction results are basically comparable. The evaluation criterions in Fig. 9 (c) displays that the accuracy of GPR second prediction is slightly improved, but it is not apparent, which means that GPR will slightly improve the point prediction performance of the cumulative displacement.

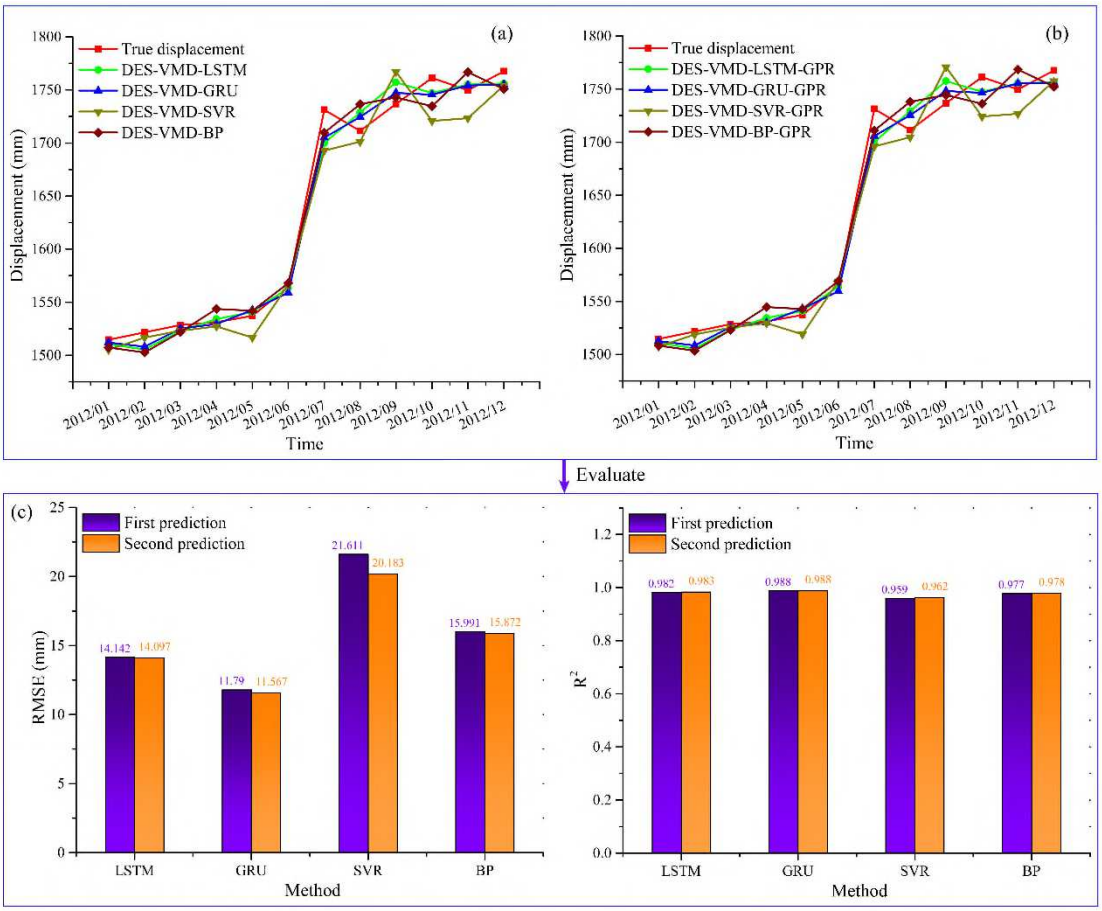
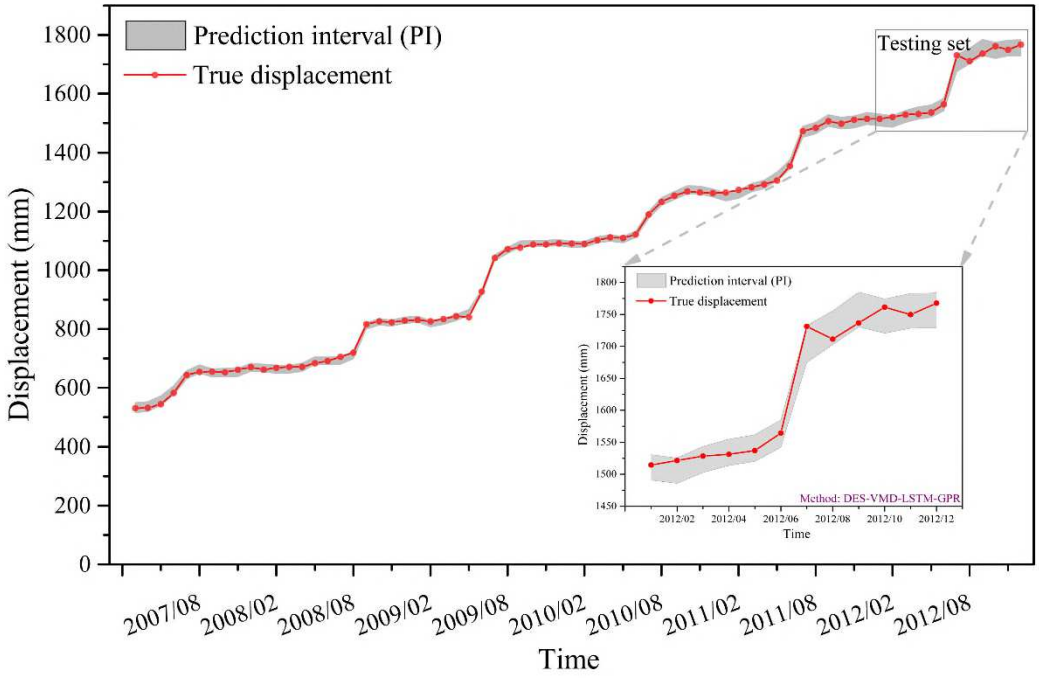
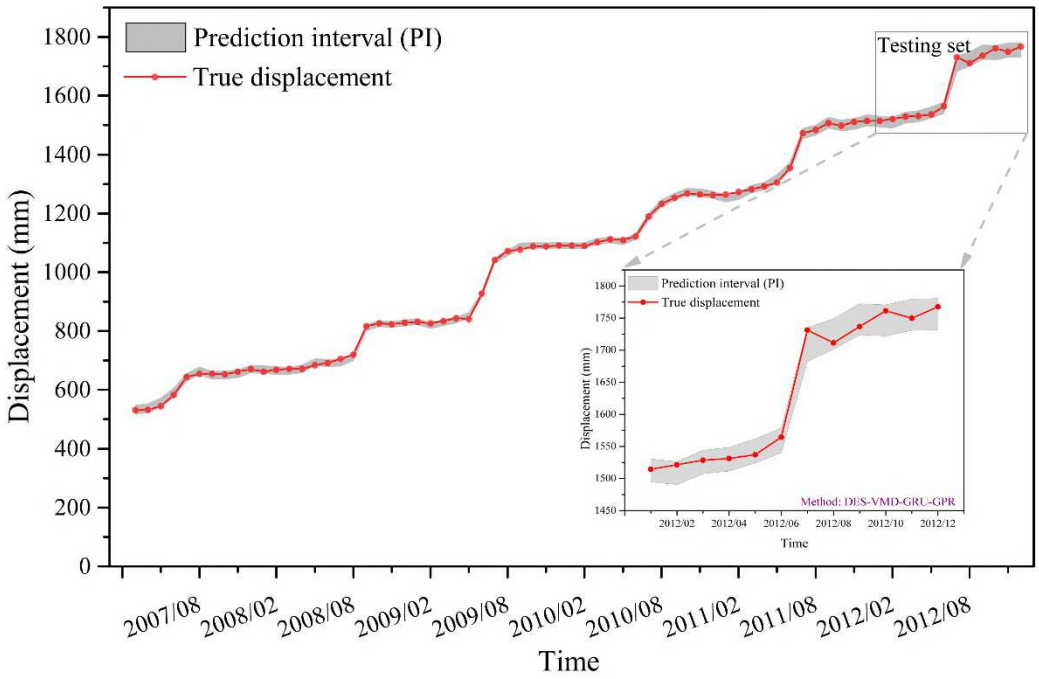


Fig 9. (a) First cumulative displacement prediction result; (b) Second cumulative displacement prediction result; (c) Point prediction accuracy metrics of two predictions before and after GPR modeling.

3.3 Interval prediction

The interval prediction results based on DES-VMD-LSTM-GPR, DES-VMD-SVR-GPR, DES-VMD-BP-GPR and DES-VMD-GRU-GPR method are shown in Fig. 10. At the 95% confidence level, it can be found that the prediction interval obtained by the four methods can cover the true displacement well, whether in the training set or in the test set. It proves the feasibility of these methods for landslide displacement interval prediction.


Prediction interval (PI)
True displacement
Testing set
Displacement (mm)
Time
Method: DES-VMD-LSTM-GPR


Prediction interval (PI)
True displacement
Testing set
Displacement (mm)
Time
Method: DES-VMD-GRU-GPR

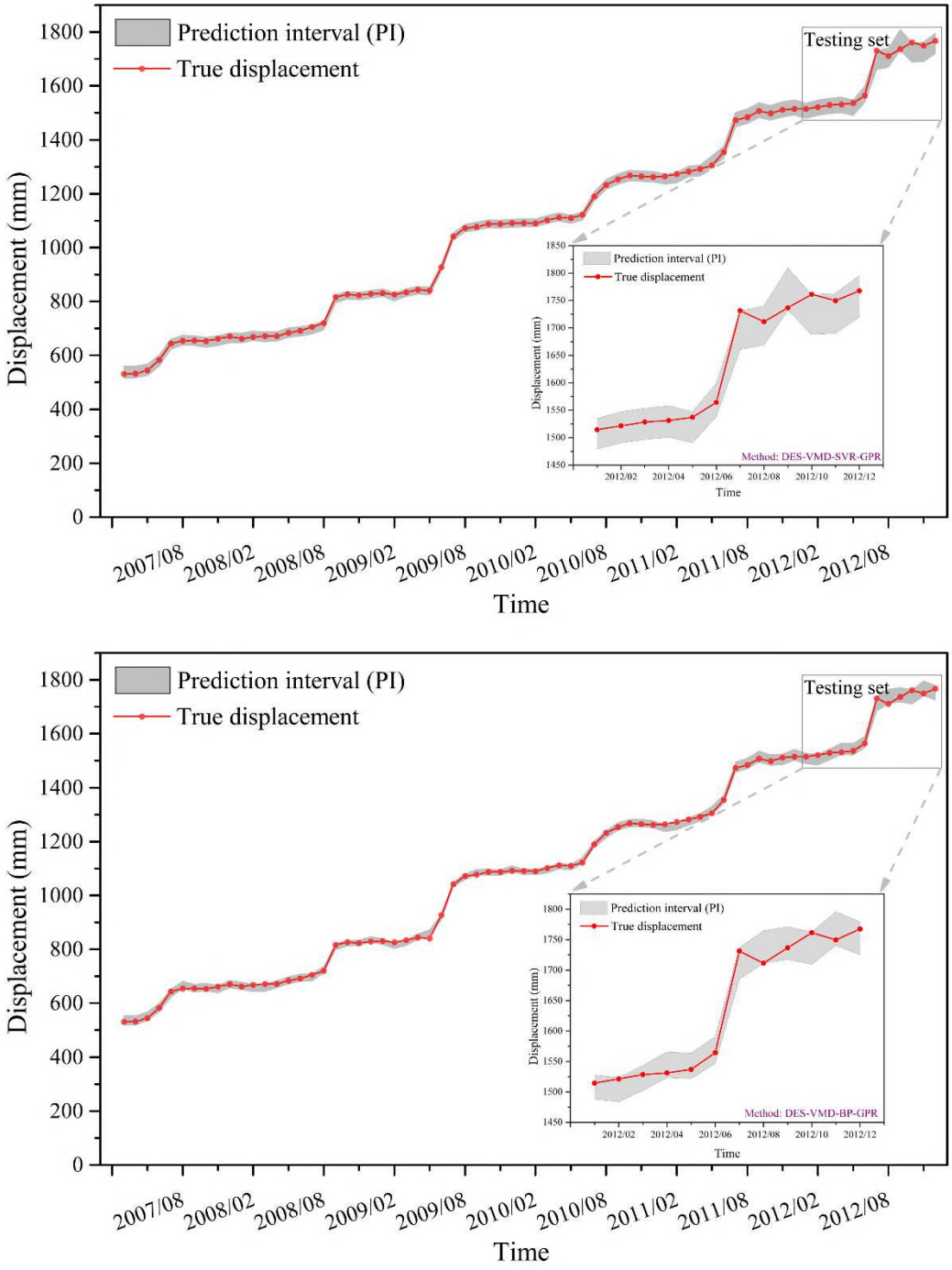


Fig 10. The interval prediction results of four methods in BZM landslide

Table 5 listed the evaluation metrics of interval prediction quality of the four methods. The PICP values of the four methods were 100%, 100%, 91.67% and 100%, respectively. A good interval coverage probability is proved by the confidence level close to or over 95%. Under the same PICP condition, higher AIS value and lower MPIW value could achieve better prediction interval. The AIS and MPIW values measured in DES-VMD-LSTM-GPR, DES-VMD-GRU-GPR and DES-VMD-BP-GPR are close, while in DES-VMD-SVR-GPR is the worst. Combined with the cumulative displacement point

prediction results, it can be found that the point prediction accuracy measured in DES-VMD-SVR is also the lowest. The interval prediction of GPR model is mainly used to explain the uncertainty in landslide displacement prediction, namely the error contained in the first prediction result. When the first prediction accuracy is high, GPR model achieve a better interval prediction. Of course, low accuracy of the first point prediction will deteriorate the interval prediction quality in GPR model.

Table 5. Interval prediction performance of all studied methods

Method	Interval prediction metrics		
	AIS	PICP	MPIW
DES-VMD-LSTM-GPR	-91.87	1.0000	47.39
DES-VMD-GRU-GPR	-81.59	1.0000	42.53
DES-VMD-SVR-GPR	-124.35	0.9167	65.29
DES-VMD-BP-GPR	-90.20	1.0000	47.47

3.4 Probabilistic prediction

The hybrid methods based on GPR can not only provide accurate point prediction and reliable interval prediction, but provide probability prediction results. The Fig. 11 visualized the probability distribution curve of BZM landslide at different time points. The six probability density curves are very complete, without unordinary value, which indicates that the probability density curves obtained by the proposed method is appropriate. In January, March, May and November, the monitored displacement is close to the center of the curve, which indicates a high prediction accuracy of these points. In July and September, the monitored displacement line is deviated from the center of the curve, caused by increasing prediction error. For the probability prediction of test set, the monitored displacements distribute evenly on both sides of the curve center, which proves the reliability of the probability prediction result. If all monitored displacements concentrate at the center or locate far from the center, it may reduce the reliability of probability prediction.

The evaluating indicator of the probabilistic prediction results i.e. CRPS index are summarized in Table 6. The CRPS is used to evaluate the entire probability density distribution curve, including the quality of point and interval prediction, as well as the performance of probability prediction. The CRPS of BZM landslide based on the four methods are similar and ranges 6.431 ~ 11.128. And the CRPS rank are consistent with the results of point prediction and interval prediction.

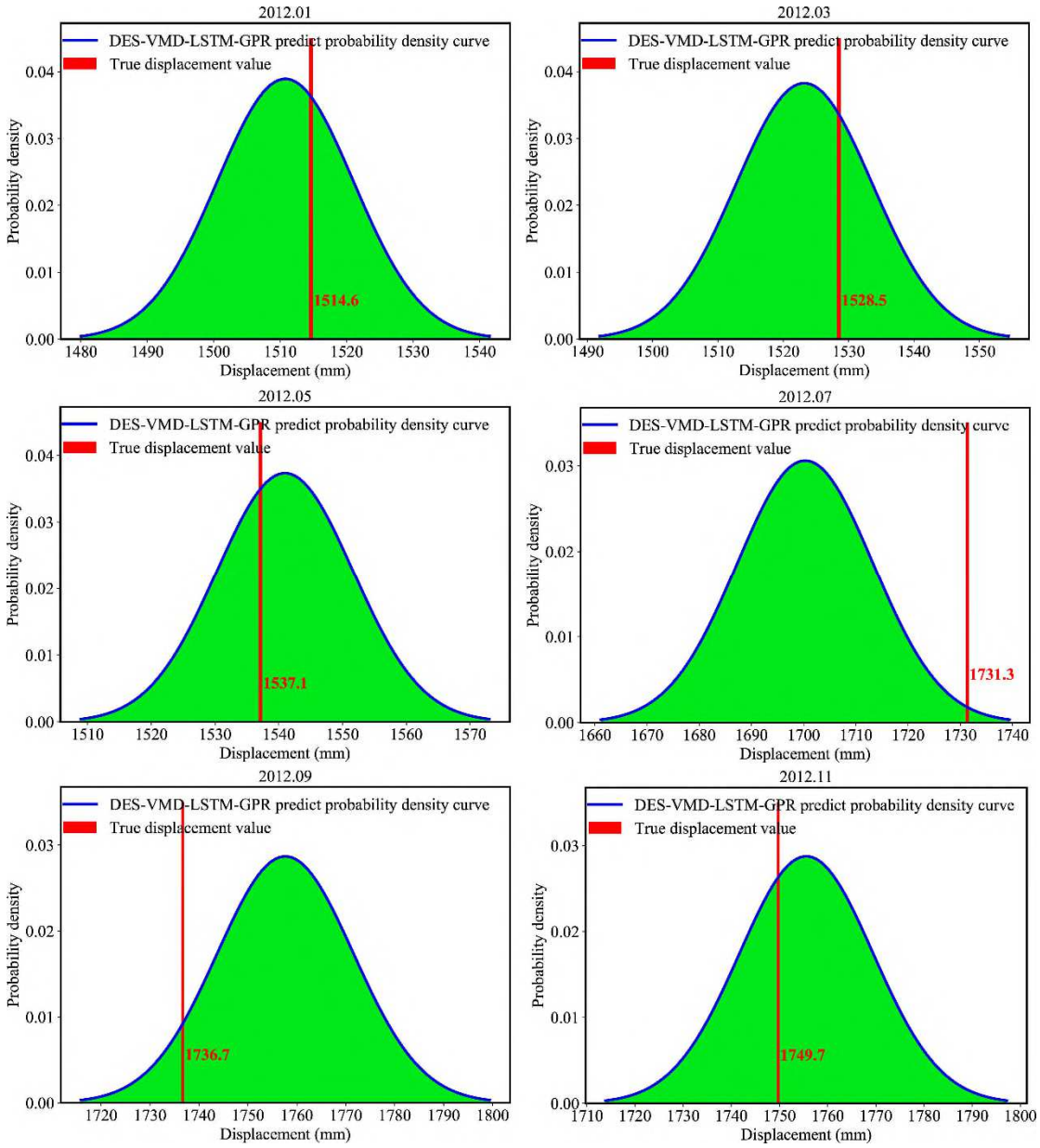


Fig 11. The probabilistic prediction results by DES-VMD-LSTM-GPR method.

Table 6. The probabilistic prediction metric of four method

Metric	DES-VMD-LSTM-GPR	DES-VMD-SVR-GPR	DES-VMD-BP-GPR	DES-VMD-GRU-GPR
CRPS	7.809	6.431	11.128	9.371

3.5 Application and comparative analysis

To further test the method, the proposed method is used to analyze the Baishuihe landslide (BSH) from Three Gorges Reservoir area. GPS monitoring point ZG118 in BSH landslide is used to test the model, and the prediction results are shown in Fig. 12 (testing set: 2012.01-2012.12). Firstly, the prediction quality by the proposed method is very good, whether point prediction (RMSE：7.161；R^2：

0.982), interval prediction (PICP：100%； AIS：-87.85) or probability prediction (CRPS: 4.151), which means that the proposed method has great potential for the early warning in stepwise landslide.

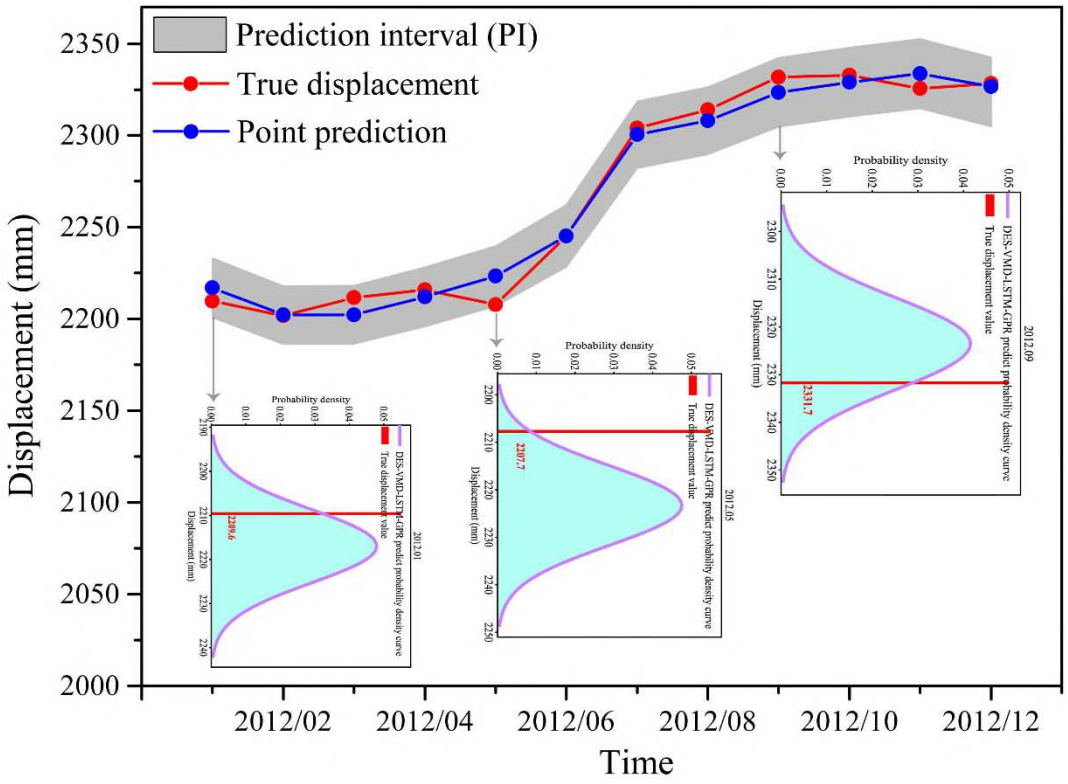


Fig 12. Point, interval and probabilistic prediction results of BSH landslide (ZG118) based on the proposed method

The point prediction and interval prediction results are compared with the calculation results in previous literature to verify the effectiveness of the proposed method. For point prediction, the proposed method is compared with GA-SVM (Zhou et al., 2016), ABC-ELM (Yan et al., 2019), LSTM (Yang et al., 2019) and PSO-SVR (Zhang et al., 2015). The prediction metrics are listed in Table 7. The LSTM based predictor performs better than other methods, and the prediction accuracy of the proposed method and LSTM in Yan et al. (2019) is almost the same, which are mainly attributed to the unique time processing advantages of LSTM. In the architecture of LSTM, the forget gate can control how much information at the previous time is used as the input of the current time, and it can make full use of historical information to learn rules. For static models like SVM and BP, there is no connections in different time steps, and rules can only be learned from one time point, and the response between output and input cannot be fully learned. Another reason why the proposed method has high prediction accuracy is the reasonable trigger factor selection. Reservoir level and rainfall are the dominant triggering factors of reservoir bank slope deformation. In traditional predictors, a pseudo periodic displacement combining

471 the periodic and random displacement is predicted according to rainfall and reservoir level factors. These
472 method neglects the periodic and random characteristics of trigger factors. In this work, the periodic and
473 stochastic displacement are accurately separated from the residual displacement by using VMD model,
474 and the trigger factor is also decomposed into low- and high-frequency components, which have similar
475 trend respectively with periodic and stochastic displacement (Fig. 6), which is the key to improve the
476 prediction performance of the model.

477 Table 7. Comparisons of point prediction constructed using the proposed method and these methods in
478 the previous references

Landslide	Method	Metrics		Reference
		RMSE/mm	R^2	
BZM (ZG111)	DES-VMD-LSTM-GPR	14.097	0.983	This paper
	GA-SVM	27.220	-	Zhou et al., (2016)
	ABC-ELM	23.010	0.940	Yan et al., (2019)
	LSTM	17.650	-	Yang et al., (2019)
BSH (ZG118)	DES-VMD-LSTM-GPR	7.161	0.982	This paper
	PSO-SVR	40.760	0.850	Zhang et al., (2015)
	LSTM	7.110	-	Yang et al., (2019)

479 The interval prediction feasibility of the proposed approach is assessed by comparison with other
480 four methods: DES-PSO-ELM (Wang et al., 2019), SBS-RVFLN (Lian et al., 2018), SB-KELM (Lian et
481 al., 2016a) and Bootstrap-KELM-BPNN (Li et al., 2019a). Three aspects need to be discussed: (1) all the
482 methods show satisfactory coverage probabilities. The PICP values of most methods are 100%, except
483 for DES-PSO-ELM and Bootstrap-KELM-BPNN (approaching or exceeding 95% confidence). These
484 results indicated that all models are reliable and have good extrapolation prediction ability in landslide
485 displacement interval prediction. (2) Regarding the interval prediction, the proposed method performs
486 better or similarly to the other methods in the two applications. It is known from Table 5, the smaller the
487 absolute value of AIS indicate the narrower width of prediction interval. The AIS absolute value of the
488 proposed method is higher than DES-PSO-ELM and SBS-RVFLN but lower than SB-KELM and
489 Bootstrap-KELM-BPNN, which shows that the proposed method tends to establish a narrower prediction
490 interval, namely better interval prediction quality. (3) Compared with other methods, the proposed
491 method is simple and can obtain more comprehensive prediction information. As the SBS-RVFLN, SB-
492 KELM and Bootstrap-KELM-BPNN is generated based on bootstrap technology, it requires at least B
493 times of repeated sampling calculation (B = 50 in SB-KELM). The prediction accuracy of SBS-RVFLN

494 and SB-KELM also depends strongly on the accuracy of K-means classifier. These models are too
495 complex and time-consuming. Although DES-PSO-ELM is simple, in which only single interval
496 prediction can be obtained.

497 Table 8. Comparisons of interval prediction at the 95% confidence level constructed using the proposed
498 method and the method in the previous references

Landslide	Method	Interval prediction metrics			Reference
		AIS	PICP	MPIW	
BZM (ZG111)	DES-VMD-LSTM-GPR	-91.87	1.0000	47.39	This study
	DES-PSO-ELM	-76.77	0.9167	-	Wang et al., (2019)
	SBS-RVFLN	-69.15	1.0000	-	Lian et al., (2018)
	SB-KELM	-101.72	1.0000	-	Lian et al., (2016a)
BSH (ZG118)	DES-VMD-LSTM-GPR	-87.85	1.0000	46.19	This study
	Bootstrap-KELM-BPNN	-	0.9806	94.25	Li et al., (2019a)

499 Temporal deformation of landslide is a noisy and non-stationary process, which is impacted by many
500 internal and external factors, such as lithology, geological structure, rainfall, reservoir level, monitoring
501 error and so on. Due to the complex nonlinear relationship between various influencing factors and
502 landslide deformation, accurate landslide displacement prediction is very difficult. Different from the
503 previous blindly pursuing high prediction accuracy, this paper combines the respective advantages of
504 LSTM and GPR, and propose a new idea of landslide displacement prediction. Based on this model, not
505 only the deterministic prediction of landslide displacement can be realized, but also the variability and
506 uncertainty related to landslide displacement prediction can be estimated.

507 Although the DES-VMD-LSTM-GPR model achieved a good performance in landslide
508 displacement prediction, it has some drawbacks. As a deep learning algorithm, LSTM needs to adjust
509 more model parameters than the classical machine learning algorithm, and has a strong dependence on
510 the amount of data. If the amount of training data is insufficient, the prediction accuracy will be affected.
511 Landslide displacement monitoring is based on month. It takes years or even decades to obtain a large
512 amount of data. Therefore, the method of developing larger datasets in a limited monitoring cycle should
513 be considered in the later stage. Another drawback is the gaussian distribution assumption of GPR model.
514 The landslide displacement has strong volatility and randomness. The gaussian distribution assumption
515 may not be the optimal distribution. More distributions need to be tested in the future, such as Weibull,
516 Rayleigh and Beta distribution.

517 4. Conclusions

518 Landslide displacement prediction is the key of landslide early warning ability. To overcome the

519 shortcomings of point prediction and consider the inherent uncertainty in such prediction, the GPR is
520 used to quantify the uncertainty related to point prediction. A new hybrid method, namely DES-VMD-
521 LSTM-GPR, is proposed to realize landslide displacement point prediction, interval prediction and
522 probability prediction. The DES method is used to predict trend displacement. The periodic and
523 stochastic displacement are decomposed from the residual displacements based on VMD, and then
524 predicted by LSTM model. The first cumulative displacement prediction results are obtained by adding
525 the trend, periodic and stochastic displacement, and then the GPR model is established based on this, and
526 the second prediction is carried out to explain the uncertainty in the first prediction. To verify the
527 prediction ability of the proposed method, we replace the LSTM part in the hybrid method with SVR,
528 BP and GRU, and apply them to BZM landslide. The results show that the dynamic models LSTM and
529 GRU have higher prediction accuracy than BP and SVR. The secondary modeling of GPR model does
530 not significantly improve the first point prediction results, but obtains reliable interval and probability
531 prediction results. Its uncertain prediction performance depends on the first prediction accuracy. Further,
532 we extend the proposed model to Baishuihe landslide and compare it with the calculation results in the
533 previous literature. The application results show that both point prediction and interval prediction show
534 superior prediction performance, and compared with other methods, this model is more concise, can
535 obtain more comprehensive prediction information, and can be used as a potential landslide disaster early
536 warning tool.

537 **Conflicts of interest:** None

538 **Acknowledgments:** This work was supported by the Doctoral fund of Guizhou University (Grant No.
539 X2021011 and X2021010). The authors are grateful to the National Cryosphere Desert Data Center for
540 providing data and material. Finally, the authors thank the reviewers for their valuable suggestions.

541 References

542 Alessandrini, S., Delle Monache, L., Sperati, S., Cervone, G., 2015. An analog ensemble for short-
543 term probabilistic solar power forecast. Applied Energy 157, 95-110.
544 Anjishnu, B., Dunson, D.B., Tokdar, S.T., 2013. Efficient Gaussian process regression for large
545 datasets. Biometrika, 75-89.
546 Cai, Z., Xu, W., Meng, Y., Shi, C., Wang, R., 2016. Prediction of landslide displacement based on
547 GA-LSSVM with multiple factors. Bulletin of Engineering Geology & the Environment 75,
548 637-646.
549 Carl Edward, R., Williams, C.K.I., 2005. Guassian process for machine learning. MIT Press.
550 Carlà, T., Intrieri, E., Di Traglia, F., Casagli, N., 2016. A statistical-based approach for determining
551 the intensity of unrest phases at Stromboli volcano (Southern Italy) using one-step-ahead
552 forecasts of displacement time series. Natural Hazards 84, 669-683.
553 Devi, K.J., Thongam, K., 2020. Automatic speaker recognition from speech signal using
554 bidirectional long-short-term memory recurrent neural network. Computational Intelligence.
555 Dragomiretskiy, K., Zosso, D., 2014. Variational Mode Decomposition. IEEE Transactions on
556 Signal Processing 62, 531-544.
557 Feng, X.T., Zhao, H., Li, S., 2004. Modeling non-linear displacement time series of geo-materials
558 using evolutionary support vector machines. International Journal of Rock Mechanics &
559 Mining ences 41, 1087-1107.
560 Fukuzono, T., 1985. A new method for predicting the failure time of a slope, Proceedings of 4th
561 International Conference and Field Workshop on Landslides. Tokyo University Press, Tokyo,
562 pp. 145-150.
563 Gould, P.G., Koehler, A.B., Ord, J.K., Snyder, R.D., Hyndman, R.J., Vahid-Araghi, F., 2008.
564 Forecasting time series with multiple seasonal patterns. European Journal of Operational
565 Research 191, 207-222.
566 Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735-
567 1780.
568 Holt, C.C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages.
569 International Journal of Forecasting 20, 5-10.
570 Huang, F., Yin, K., Zhang, G., Gui, L., Yang, B., Liu, L., 2016. Landslide displacement prediction

using discrete wavelet transform and extreme learning machine based on chaos theory. Environmental Earth Sciences 75, 1376.

Li, L., Wu, Y., Miao, F., Zhang, L., Xue, Y., 2019. Landslide displacement interval prediction based on different Bootstrap methods and KELM-BPNN model. Chinese Journal of Rock Mechanics and Engineering 38, 912-926.

Li, X., Kong, J., Wang, Z., 2012. Landslide displacement prediction based on combining method with optimal weight. Natural Hazards 61, p.635-646.

Lian, C., Chen, C.L.P., Zeng, Z., Yao, W., Tang, H., 2016a. Prediction Intervals for Landslide Displacement Based on Switched Neural Networks. IEEE Transactions on Reliability 65, 1483-1495.

Lian, C., Zeng, Z., Yao, W., Tang, H., 2014. Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis. Neural Computing and Applications 24, 99-107.

Lian, C., Zeng, Z., Yao, W., Tang, H., 2015. Multiple neural networks switched prediction for landslide displacement. Engineering Geology 186.

Lian, C., Zeng, Z., Yao, W., Tang, H., Chen, C.L.P., 2016b. Landslide Displacement Prediction With Uncertainty Based on Neural Networks With Random Hidden Weights. IEEE Trans Neural Netw Learn Syst 27, 2683-2695.

Lian, C., Zhu, L., Zeng, Z., Su, Y., Yao, W., Tang, H., 2018. Constructing prediction intervals for landslide displacement using bootstrapping random vector functional link networks selective ensemble with neural networks switched. Neurocomputing 291, 1-10.

Liu, Z., Shao, J., Xu, W., Chen, H., Shi, C., 2014. Comparison on landslide nonlinear displacement analysis and prediction with computational intelligence approaches. Landslides 11, 889-896.

Liu, Z., Xu, W., Shao, J., 2012. Gauss Process Based Approach for Application on Landslide Displacement Analysis and Prediction. Computer Modeling in Engineering & ences 84, 99-122.

Ma, J., Tang, H., Liu, X., Hu, X., Sun, M., Song, Y., 2017. Establishment of a deformation forecasting model for a step-like landslide based on decision tree C5.0 and two-step cluster algorithms: a case study in the Three Gorges Reservoir area, China. Landslides 14, 1275-1281.

Ma, J., Tang, H., Liu, X., Wen, T., Zhang, J., Tan, Q., Fan, Z., 2018. Probabilistic forecasting of

landslide displacement accounting for epistemic uncertainty: a case study in the Three Gorges Reservoir area, China. Landslides, 1-9.

Mayoraz, F., Vulliet, L., 2002. Neural Networks for Slope Movement Prediction. International Journal of Geomechanics 2, 153-173.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., Sabeti, P.C., 2011. Detecting Novel Associations in Large Data Sets. Science 334, 1518-1524.

Saito, M., 1965. Forecasting the time of occurrence of a slope failure. Proceedings of the 6th International Conference on Soil Mechanics and Foundation Engineering 2, 537-541.

Setsuo, HAYASHI, Bo-Won, PARK, Fujiya, KOMAMURA, Tsuyoshi, YAMAMORI, 1988. On the Forecast of Time to Failure of Slope (II) Approximate Forecast in the Early Period of the Tertiary Creep. Landslides 23, 1-16.

Wang, Y., Tang, H., Wen, T., Ma, J., 2019. A hybrid intelligent approach for constructing landslide displacement prediction intervals. Applied Soft Computing 81, 105506.

Wen, T., Tang, H., Wang, Y., Lin, C., Xiong, C., 2017. Landslide displacement prediction using the GA-LSSVM model and time series analysis: a case study of Three Gorges Reservoir, China. Natural Hazards & Earth System ences 17, 1-20.

Winkler, R.L., 1972. A Decision-Theoretic Approach to Interval Estimation. Journal of the American Statistical Association 67, 187-191.

Wu, L., Liu, S., Yang, Y., 2016. Grey double exponential smoothing model and its application on pig price forecasting in China. Applied Soft Computing 39, 117-123.

Wu, Y., 2016. Basic characteristics and monitoring data of Baishuihe landslide in Zigui County, Three Gorges Reservoir area, 2007-2012. National Cryosphere Desert Data Center.

Wu, Y., Haifeng, H., 2016. Deformation monitoring data of Bazimen Landslide in Zigui County, Three Gorges Reservoir area, 2007-2012. National Cryosphere Desert Data Center.

Xu, S.L., Niu, R.Q., 2018. Displacement prediction of Baijiabao landslide based on empirical mode decomposition and long short-term memory neural network in Three Gorges area, China. Computers & Geosciences 111, 87-96.

Yan, H.， Li , S.H.， Wu, L.Z., 2019． Landslide displacement prediction based on multiple data-driven model methods. Journal of engineering geology 27(2): 459-465.

Yang, B., Yin, K., Lacasse, S. et al. 2019. Time series analysis and long short-term memory neural

network to predict landslide displacement. Landslides 16, 677–694.

Yin, Y., Wang, H., Gao, Y., Li, X., 2010. Real-time monitoring and early warning of landslides at relocated Wushan Town, the Three Gorges Reservoir, China. Landslides 7, 339-349.

Yu, X.-m., Feng, W.-z., Wang, H., Chu, Q., Chen, Q., 2019. An attention mechanism and multi-granularity-based Bi-LSTM model for Chinese Q&A system. Soft Computing 24, 5831-5845.

Zhang, J., Yan, J., Infield, D., Liu, Y., Lien, F.-s., 2019. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. Applied Energy 241, 229-244.

Zhang, J., Yin, K.L., Wang, J.J., Huang, F.M., 2015. Displacement prediction of Baishuihe landslide based on time series and PSO-SVR model. Chinese Journal of Rock Mechanics and Engineering 34(2), 383-391.

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.