Research Article

Point and Interval Predictions for Tanjiahe Landslide Displacement in the Three Gorges Reservoir Area, China

Yankun Wang,[1] **Huiming Tang**,[1,2] **Tao Wen**,[3] **Junwei Ma**,[2] **Zongxing Zou**,[2] **and Chengren Xiong**[2]

[1]*Faculty of Engineering, China University of Geosciences, Wuhan, Hubei 430074, China*
[2]*Three Gorges Research Center for Geo-Hazards of Ministry of Education, China University of Geosciences, Wuhan, Hubei 430074, China*
[3]*School of Geosciences, Yangtze University, Wuhan, Hubei 430100, China*

Correspondence should be addressed to Huiming Tang; tanghm@cug.edu.cn and Tao Wen; wentao200840@yangtzeu.edu.cn

Accurate landslide displacement prediction has great practical significance for mitigating geohazards. Traditional deterministic forecasting methods can provide only a single point value and cannot give the degree of uncertainty associated with the forecast, thereby failing to provide information on predictive confidence. This study applied interval prediction for landslide displacement. Taking the Tanjiahe landslide of the Three Gorges Reservoir Area as an example and considering the impact of seasonal variations in reservoir level and rainfall, the uncertainties associated with landslide displacement prediction were quantified into prediction intervals (PIs) by a bootstrapped least-square support vector machine (LSSVM) method (B-LSSVM). The proposed method consists of three steps: First, the LSSVM and bootstrapping were combined to estimate the true regression means of landslide displacement and the variance with respect to model misspecification uncertainties. Second, a new LSSVM model optimized by a genetic algorithm (GA) was implemented to estimate the noise variance. Finally, the point prediction was derived from the regression means, and the PIs were constructed by combining the regression mean, the model variance, and the noise variance. We applied the proposed method to predict the displacement of four GPS monitoring points of the Tanjiahe landslide, and we comprehensively compared the prediction accuracy and the quality of the constructed PIs with benchmark methods. A simulation and performance comparison showed that the proposed method is a promising technique for providing accurate and reliable prediction results for landslide displacement.

1. Introduction

Landslides are the most frequent geological hazard in China. These events seriously threaten infrastructure and the safety of human life and result in extensive casualties and property losses every year. The Three Gorges Reservoir Area (TGRA) is one of the areas in China hardest hit by landslide disasters [1]. A displacement time series provides a direct representation of the landslide evolution process. Accurate landslide displacement predictions can help inform people about the future evolutionary dynamics of landslides.

The evolution process of a landslide is characterized by nonlinearity [2]. A range of causes, such as geological factors, climatic factors, earthquakes, and human activities, contributes to the formation of landslides, making it challenging to accurately forecast landslide displacement [3]. Due to the complexity of landslide systems, it is hard to build explicit mathematical expressions between soil/rock mechanical properties and displacement using physical methods. Compared with physical methods, data-driven methods require only the available monitoring data and not the physical parameters of the landslide. These methods have the outstanding capabilities of capturing nonlinear mapping between input features and outputs. In recent years, data-driven methods have been widely applied and are becoming increasingly attractive in landslide displacement prediction. Du et al. [4] used a back propagation neural network to predict landslide displacement, and the effectiveness of the

method is verified by studying the Baishuihe and Bazimen landslides in TGRA. Cao et al. [3], Huang et al. [5, 6], and Lian et al. [7] applied an ELM-based model for point prediction of landslide displacement, and the performance of proposed models is tested in reservoir landslides of the TGRA. Cai et al. [8], Cao et al. [3], Miao et al. [9], Ren et al. [10], and Wen et al. [11] proposed SVM-based hybrid models to predict the displacement of reservoir landslides, and the numerical results demonstrate that their hybrid method outperformed artificial neural network-based methods. In [12, 13], a hybrid approach is proposed based on two-step clustering and decision tree C5.0 algorithms for a step-like landslide, and the usefulness of the hybrid approach is explored by the Zhujiadian landslide in TGRA. Most of data-driven methods can predict landslide displacement well. However, most predictions are deterministic (or point) predictions; i.e., they can only provide a crisp displacement value and cannot provide the uncertainty associated with the predictions.

In practice, the accuracy of point estimates can be affected by uncertainties. These uncertainties stem mainly from two important sources: (1) model uncertainties caused by the misspecification of the structure and parameters of the machine learning technique (e.g., number of nodes in hidden layers of artificial neural networks (ANNs), hyperparameters in SVMs, and unreasonable section of input variables) and (2) noise variance of measured data due to the stochastic and chaotic characteristics of landslide displacement. For practical applications, decision-makers require accurate point prediction results as well as quantitative estimates of the inherent uncertainty of the forecasts [14]. Thus, to improve the reliability and credibility of model outputs, it is necessary to incorporate prediction uncertainties into point predictions to quantify the uncertainties. Prediction intervals (PIs) are tools that are commonly used to estimate likely uncertainties in deterministic forecasting. Reliable PIs allow decision-makers to efficiently perceive the degree of uncertainty and make more informed decisions [15]. Due to their application in risk management, PIs have successfully been applied to probabilistic predictions in many fields, such as electricity market price prediction [16], wind power prediction [17], and flood prediction [18].

However, current studies on the probability prediction of landslide displacement are very limited. Lian et al. [19] established an ANN model with random hidden weights for interval prediction of landslide displacement. In addition, Lian et al. [20] and Ma et al. [21] used a hybrid method based on bootstrapping, ELM, and ANN to construct PIs in landslide displacement prediction. Wang et al. [22] proposed a hybrid method that combines double exponential smoothing and lower and upper bound estimation models to construct the PIs of landslide displacement. All these studies constructed PIs of landslide displacement by ANNs. Although ANNs are powerful tools for approximating highly nonlinear systems, their disadvantages include faulty theory foundation, local minimum, and overfitting [23], which may result in large prediction error, leading to unnecessarily wide PIs. Therefore, more reliable methods need to be developed to obtain high-quality PIs in landslide displacement. SVMs are

considered the best regression method because of their strong inference capacity, excellent generalization, and accurate prediction ability [24], and they outperform many ANNs by avoiding the overfitting problem [25]. Moreover, SVMs can still obtain satisfactory prediction results from small training sets. However, the training process of a SVM is time consuming for large datasets because a quadratic programming problem needs to be solved under an inequality constraint. To reduce the computational cost of SVMs, an improved version, i.e., the least-square support vector machine (LSSVM), was proposed by Suykens and Vandewalle [26]. LSSVMs have the advantages of excellent generalization and high prediction accuracy of SVMs as well as rapid computation. Considering their superiority, LSSVMs have successfully been applied in the deterministic prediction of landslide displacement [8, 11, 27].

In this paper, a LSSVM model in combination with bootstrapping was used for the probabilistic prediction of landslide displacement. In this method, the true regression means and model misspecification uncertainties were first estimated using a bootstrapped LSSVM (B-LSSVM) model, then a new LSSVM model optimized by a genetic algorithm (GA) was implemented to predict the noise variance. After proper estimation of the regression means and uncertainties, the point prediction results and PIs were obtained by combining the regression means, the model variance, and the noise variance. The hybrid method, namely, B-LSSVM, was employed for point and interval estimations of the Tanjiahe landslide in the TGRA of China, and its performance was compared to that of several other data-driven methods.

2. Methodology

2.1. PI Formulation. PI is a powerful approach for quantifying the uncertainties in point prediction. A PI consists of upper and lower limits between which a future unknown value is expected to occur with a prescribed probability called a confidence level ($100(1 − \alpha)$%) (usually 95%) [15]. Considering a stochastic time series, the ith measured value t_i can be expressed as follows:

$$t_i = g(x_i) + \varepsilon(x_i), \tag{1}$$

where x_i is the vector of inputs, $\varepsilon(x_i)$ denotes noise with zero mean, and $g(x_i)$ is the nonlinear function denoting the true regression mean.

The goal of point prediction is to approximate the true regression mean $g(x_i)$. Data-driven methods are usually used for the point prediction model. However, prediction errors are unavoidable because of the inherent uncertainties in the forecasting process. Such errors can be caused by the misspecification of the structure and parameters of the data-driven methods. Let a trained data-driven model $\hat{g}(x_i)$ denote an estimation of the true regression mean $g(x_i)$. Accordingly, the prediction bias can be represented as follows:

$$t_i − \hat{g}(x_i) = [g(x_i) − \hat{g}(x_i)] + \varepsilon(x_i). \tag{2}$$

PIs address the variance of the left side of Equation (2). Confidence intervals (CIs) manage the variance of the first term on the right side of Equation (2). PIs should be distinguished from CIs. CIs address the accuracy of the estimate of the true model $g(x_i)$, which is estimated by the probability distribution $P(g(x_i) \mid \widehat{g}(x_i))$, whereas PIs address the accuracy of the estimate of the target t_i, which is related to the probability distribution $P(t_i \mid \widehat{g}(x_i))$. Therefore, the PI will be wider than the CI. In practical applications, PIs are more useful than CIs because PIs are concerned with the accuracy with which we can predict the observed target value itself and not just the accuracy of our estimate of the true regression [14].

Assuming that the two components on the right side of Equation (2) are statistically independent, the total prediction variance $\sigma_t^2(x_i)$ can be expressed as follows:

$$\sigma_t^2(x_i) = \widehat{\sigma}_g^2(x_i) + \widehat{\sigma}_\varepsilon^2(x_i), \qquad (3)$$

where $\widehat{\sigma}_g^2(x_i)$ is the variance of the model misspecification uncertainties related to the data-driven model and $\widehat{\sigma}_\varepsilon^2(x_i)$ is the noise variance. Once these values are properly estimated, the $100(1 - \alpha)\%$ $(0 < \alpha < 1)$ confidence level PIs can be defined as follows:

$$L_t^{(\alpha)}(x_i) = \widehat{g}(x_i) - z_{1-\alpha/2}\sqrt{\sigma_t^2(x_i)}, \qquad (4)$$

$$U_t^{(\alpha)}(x_i) = \widehat{g}(x_i) + z_{1-\alpha/2}\sqrt{\sigma_t^2(x_i)}, \qquad (5)$$

where $L_t^{(\alpha)}(x_i)$ and $U_t^{(\alpha)}(x_i)$ are the lower and upper limits of the ith PI, respectively, and $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution.

2.2. LSSVM for Regression Analysis. The LSSVM is an improved formulation of the original SVM based on structural risk minimization [26, 28]. In this study, the LSSVM was applied to conduct a regression analysis between influential factors and landslide displacement. Given a training dataset $\{(x_i, t_i)\}_{i=1}^N$, where $x_i \in R^m$ denotes m influential factors and $t_i \in R$ is the measured landslide displacement, the formulation of the LSSVM for regression analysis can be represented using the following constrained optimization problem:

$$\min_{w,b,\xi} \quad J(w, \xi) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N \xi_i, \qquad (6)$$
$$\text{s.t.} \quad t_i = w^T \phi(x_i) + b + \xi_i, \quad i = 1, 2, \cdots, N,$$

where γ is a regularization parameter, ξ_i represents random errors, w^T is the weight vector, $\varphi(x)$ is the kernel space function, and b is the threshold.

The resulting LSSVM model for regression analysis can be constructed as follows:

$$t(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i K(x_i, x_k) + b\right), \qquad (7)$$

where α_i is the Lagrange multiplier and $K(\cdot)$ is a kernel function matrix. In this study, the radial basic function (RBF) is applied as the kernel function of the LSSVM because it has fewer parameters and excellent nonlinear mapping performance. The RBF can be expressed as follows:

$$K(x_i, x_k) = \exp\left(-\frac{\|x_i - x_k\|^2}{2\delta^2}\right), \qquad (8)$$

where δ is the bandwidth of the RBF $(\delta > 0)$.

In this study, the LSSVM was implemented by the LS-SVMlab toolbox [29] developed in MATLAB. The regularization parameter γ and kernel parameter δ in LSSVM algorithms are known to affect performance. Therefore, these two parameters must be optimized. The optimization process is described in Section 2.4.

2.3. Bootstrapping. Bootstrapping [30] is essentially a resampling technique for estimating the distribution of a statistic. In practice, sampling the global individuals of a statistic is often difficult and impossible to achieve; therefore, the actual distribution of a statistic is not known a priori. A common alternative approach is to sample only a small component of the statistic and to then apply bootstrapping to infer the approximate distribution of the statistic. By uniformly resampling limited sample data B times, bootstrapping can approximately measure the index of the distribution of a statistic (e.g., mean and variance). In the construction of PIs, bootstrapping is the most commonly used technique and has been found to be quite reliable compared with other approaches. Bootstrap sampling can be based on pairs or residuals. In this study, bootstrapping of residuals was used. The sampling process is described as follows:

(1) Train the LSSVM model using the training set to obtain the optimal hyperparameters (γ and δ) and then apply the LSSVM to both the training and testing sets to obtain the forecast value

(2) Calculate the forecast residuals of the training set and recenter it

(3) Uniformly sample the recentered residuals using bootstrapping to obtain the resampling dataset

(4) Generate new targets by summing the forecast value and recentered residuals to acquire the resampling training set

(5) Apply the LSSVM model trained in step (1) to the qth resampling samples and the testing set to obtain the estimated value
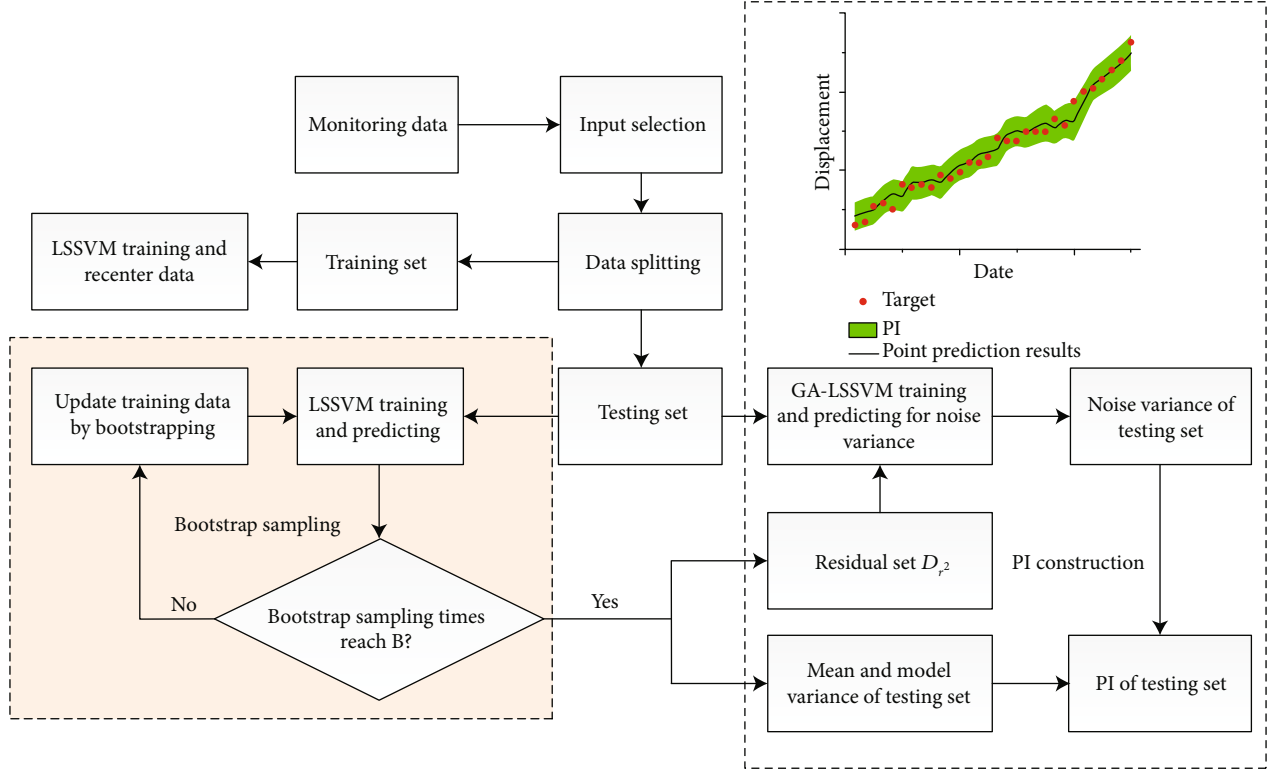
FIGURE 1: Generalized framework of the point and prediction interval estimations based on the B-LSSVM method.

(6) Repeat steps 3–5 B times to derive B bootstrap replications

2.4. PI Construction Based on Bootstrapping and LSSVM.
The flowchart of PI construction based on bootstrapping and LSSVM is shown in Figure 1. The flowchart consists of six main steps: (1) input variable selection, (2) data splitting, (3) bootstrap sampling and LSSVM training, (4) regression mean and model variance estimation, (5) noise quantification, and (6) PI construction. The details of each step are described as follows.

2.4.1. Input Variable Selection. In this step, the selection of LSSVM input variables is mainly based on empirical guidelines. These guidelines are summarized from other studies of landslide displacement prediction in the TGRA.

2.4.2. Data Splitting. The whole dataset was split into two sets: (a) a training set and (b) a testing set. The training set is used to determine the optimal LSSVM structure and hyperparameters. The testing set is used to validate the performance of the proposed method.

2.4.3. Bootstrap Sampling and LSSVM Training. In the process of bootstrapping, B training datasets are derived from the residual dataset and LSSVM training is performed for each generation of the bootstrap samples. In this step, the two hyperparameters of LSSVM were optimized through the coupled simulated annealing (CSA) algorithm and the simplex method, which are the default optimization algorithms embedded in the LS-SVMlab toolbox. The tuning

process consisted of two steps. First, the CSA was applied to determine suitable starting points of the two parameters within the search limits ($e^{(-10)}$, $e^{(10)}$). Second, these starting points were transferred to the simplex method to search for optimal values.

To prevent an additional source of uncertainty arising from the hyperparameters of the LSSVM and improve the computational efficiency, once the hyperparameters (γ and δ) of the LSSVM were determined in step (1) of the bootstrapping process, they were fixed to train and predict all bootstrapped and test samples. After all of the B bootstrap samples have been trained by the LSSVM, an ensemble of point values predicted by the B-LSSVM models is obtained and the distribution characteristics of these point values can be estimated using the following procedure.

2.4.4. Regression Mean and Model Variance Estimation. The ensemble model formed by the B-LSSVM will produce a less biased estimate of the true regression of the targets. The mean of the B-LSSVM model output $\hat{g}(x_i)$ represents the approximation of the true regression mean $g(x_i)$, and it can be regarded as the point prediction result.

$$\hat{g}(x_i) = \frac{1}{B} \sum_{q=1}^{B} g_q(x_i), \qquad (9)$$

where $g_q(x_i)$ is the forecast value derived from the qth LSSVM model.

Following the calculation of the regression mean, the variance of B-LSSVM predictions can be used to estimate the variance of model uncertainty:

$$\widehat{\sigma}_g^2(x_i) = \frac{1}{B-1} \sum_{q=1}^{B} \left[g_q(x_i) - \widehat{g}(x_i) \right]^2. \tag{10}$$

2.4.5. Noise Quantification. In the construction of PIs, both the variance of model uncertainty $\widehat{\sigma}_g^2$ and the noise variance $\widehat{\sigma}_\varepsilon^2$ must be estimated. From Equation (3), $\widehat{\sigma}_\varepsilon^2$ can be obtained as follows:

$$\widehat{\sigma}_\varepsilon^2 = E\left[(t - \widehat{g})^2\right] - \widehat{\sigma}_g^2. \tag{11}$$

According to Equation (11), the squared residual is calculated as follows:

$$r^2(x_i) = \max \left([t_i - \widehat{g}(x_i)]^2 - \widehat{\sigma}_g^2(x_i), 0 \right), \tag{12}$$

where $\widehat{g}(x_i)$ and $\widehat{\sigma}_g^2(x_i)$ can be calculated from Equations (9) and (10), respectively.

Combining the residuals and the corresponding inputs, a new residual dataset D_{r^2} is formed as follows:

$$D_{r^2} = \left\{ \left(x_i, r^2(x_i) \right) \right\}_{i=1}^N. \tag{13}$$

Then, the unknown values $\widehat{\sigma}_\varepsilon^2$ of the testing set can be estimated by implementing a new LSSVM model to train the residual dataset D_{r^2}. The training cost function is usually defined using the following form:

$$C_N = \frac{1}{2} \sum_{i=1}^{N} \left[\frac{r^2(x_i)}{\widehat{\sigma}_\varepsilon^2(x_i)} + \log\left(\widehat{\sigma}_\varepsilon^2(x_i) \right) \right]. \tag{14}$$

Since the cost function (14) is not embedded in the LS-SVMlab toolbox, the new LSSVM cannot be easily trained with the default algorithm in this toolbox. Therefore, we applied the GA to minimize the cost function (14) in this step. The GA is an optimization algorithm that simulates the process of natural evolution in biology, and it has excellent efficiency in finding the global optimization. The details of the GA can be found in Davis [31]. In the GA, the initial population evolves toward the new population through selection, crossover, and mutation until the iteration termination conditions are reached. The best individual (the optimal parameters γ and δ of LSSVM) can be derived. Then, the optimal parameters were transferred to the LSSVM model to predict the noise variance $\widehat{\sigma}_\varepsilon^2$ of the testing set. To ensure that the estimated variance is positive, the cost function was set to infinity once the estimated variance is negative in the process of GA optimization.

2.4.6. PI Construction. After estimating the true regression mean $\widehat{g}(x_i)$, the variance of model uncertainty $\widehat{\sigma}_g^2(x_i)$, and the noise variance $\widehat{\sigma}_\varepsilon^2$, the point predictions and PIs with a $(1 - \alpha)100\%$ confidence level can be obtained using Equations (9), (4), and (5).

2.5. Performance Criterion. The root-mean-square error (RMSE) and mean absolute percentage error (MAPE) are the evaluation indices most commonly used for assessing point prediction accuracy. These indices measure the deviation degree between the predicted displacement and the observed displacement and can be expressed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}[t_i - \widehat{g}(x_i)]^2}{N}}, \tag{15}$$

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|t_i - \widehat{g}(x_i)|}{t_i}. \tag{16}$$

The prediction interval coverage probability (PICP) is a commonly used index for assessing PI, and it indicates the possibility that the target values lie within the upper and lower limits. Given N test samples, the PICP can be expressed as follows:

$$\mathrm{PICP} = \frac{1}{N} \sum_{i=1}^{N} c_i, \tag{17}$$

where

$$c_i = \begin{cases} 1, & t_i \in \left[L_t^{(\alpha)}(x_i), U_t^{(\alpha)}(x_i) \right], \\ 0, & t_i \notin \left[L_t^{(\alpha)}(x_i), U_t^{(\alpha)}(x_i) \right]. \end{cases} \tag{18}$$

The PICP is strongly related to the width of the PI. A high PICP can be readily achieved by broadening the width of the PI. However, such PIs are certainly meaningless in practice. Therefore, a measure termed the normalized mean PI width (NMPIW) is used to quantify the width of the PI, and it is expressed as follows:

$$\mathrm{NMPIW} = \frac{1}{N\zeta} \sum_{i=1}^{N} \left(U_t^{(\alpha)}(x_i) - L_t^{(\alpha)}(x_i) \right) * 100\%, \tag{19}$$

where ζ is the range of the actual measured values.

In general, a PI with a high PICP and low NMPIW is considered high quality. However, both measures evaluate the quality of the PI from one aspect. A comprehensive index that combines PICP and NMPIW is required. In this study, we use the modified coverage width-based criterion (CWC) [19] for the comprehensive assessment of PIs.

$$\mathrm{CWC} = (\mathrm{NMPIW} + \psi) e^{\left(\gamma (\mathrm{PICP} - \mu)^2 \right)/2\delta^2}, \tag{20}$$

where the ψ is a small positive value within the range of 0.1% to 0.5% and μ and δ are the two hyperparameters. The value of μ is usually set to $1 - \alpha$, and δ is set to a small positive value less than 1. In this study, ψ and δ are set to 0.001 and 0.05, respectively. Generally, γ is set to 1 during the
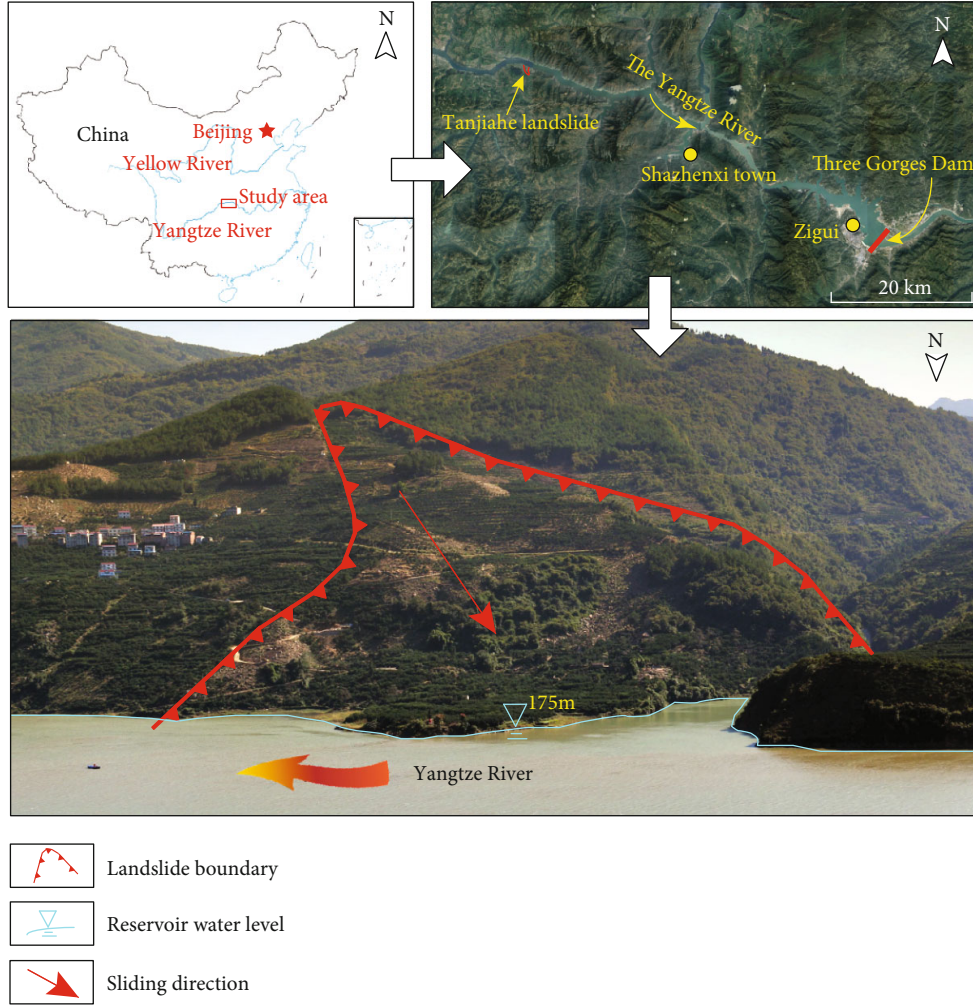
FIGURE 2: Location and geomorphological delimitation of the Tanjiahe landslide.

training of LUBE, and for testing, γ is given by the following step function:

$$\gamma = \begin{cases} 0, & \text{PICP} \geq \mu, \\ 1, & \text{PICP} < \mu. \end{cases} \quad (21)$$

In the evaluation of the constructed PIs, the aim of the CWC is to find a trade-off between the informativeness (NMPIW) and validity (PICP) of PIs. According to the definition of CWC, the smaller the CWC value, the higher the quality of PIs. If the PICP is not less than the assigned μ, γ is equal to 0. Therefore, the exponential term in Equation (20) is eliminated and the CWC depends on the NMPIW. Otherwise, γ is equal to 1 and the exponential term penalizes the violation of the coverage probabilities.

3. Case Study: Tanjiahe Landslide

3.1. Geological Features. The Tanjiahe landslide is situated in Zigui County, Hubei Province, China, on the southern bank of the Yangtze River 52 km upstream from the Three Gorges

Dam (Figure 2). A three-dimensional topographical map of the Tanjiahe landslide is shown in Figure 3(a), which indicates that the Tanjiahe landslide is fan-shaped from an in-plane view. The landslide has a longitudinal dimension of approximately 1000 m and a mean width of approximately 400 m. The area of the landslide is approximately 0.4 km², and the mean depth of slip surface is approximately 40 m. The accumulation of the landslide is estimated at 16 million m³. The elevation of the toe of the landslide is 135 m above sea level (a.s.l.), and the elevation of the top of the landslide is 432 m a.s.l. The average slope of the foot of the landslide is approximately 10°, and the slope varies from 25° to 30° in the middle and at the head of the landslide. Flat terrain occurs at the foot of the landslide; the elevation of this terrain extends from 164 m to 188 m a.s.l., and its area is approximately 0.046 km².

Figure 3(b) shows a schematic geological profile of cross-section A–A′ of the Tanjiahe landslide. The figure shows that the main sliding direction is orientated at 342°. The materials of the landslide consist of loose debris soil. The bedrock consists of thin- to medium-thickness layered quartz sandstone and carbonaceous siltstone, with an inclination of 10° and a
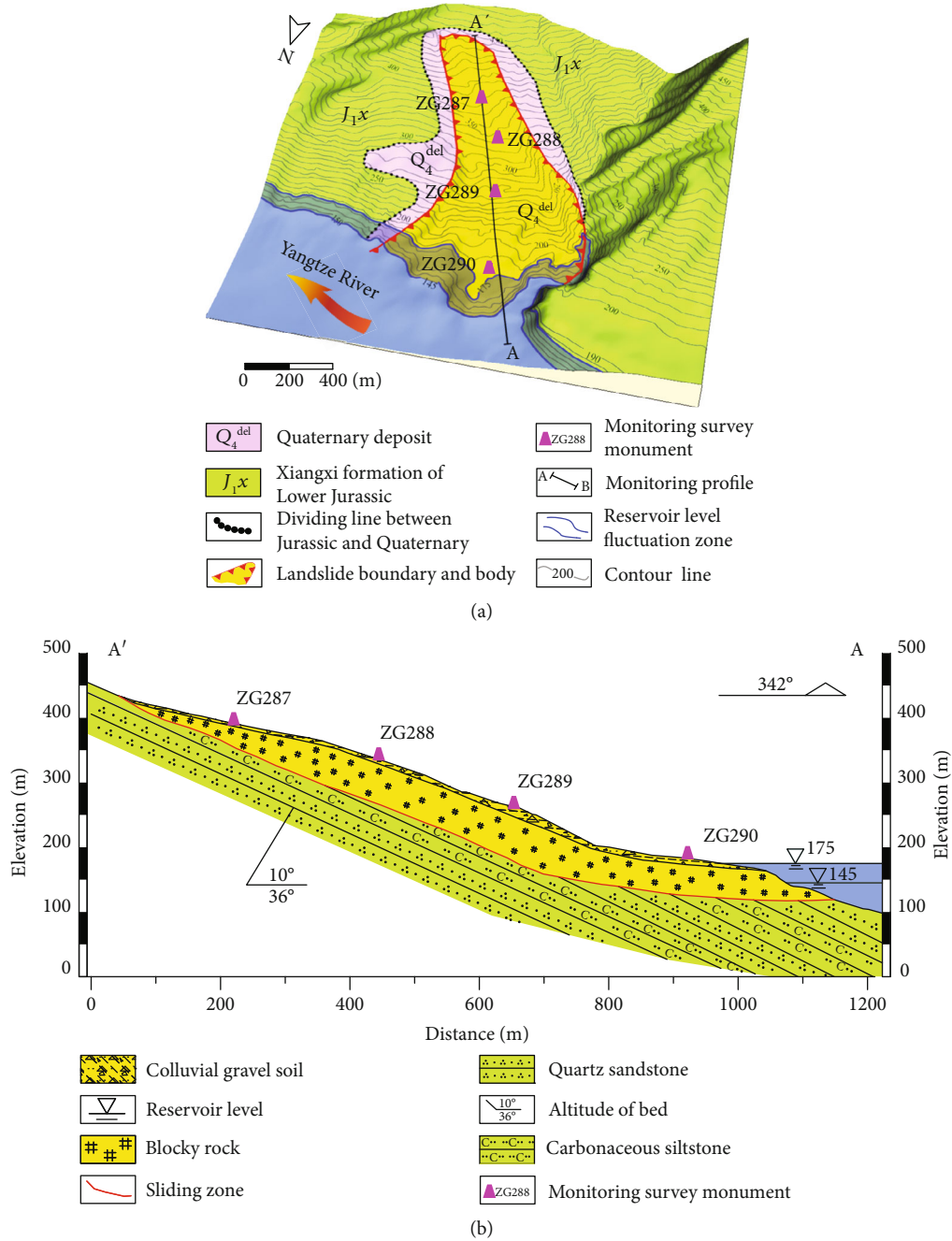
(a)

(b)

Figure 3: (a) Three-dimensional topographical map of the Tanjiahe landslide and (b) schematic geological cross-section (A–A′) of the Tanjiahe landslide.

dip angle of 36°. The material of the sliding zone is silty clay mixed with little gravel. The head of the slip surface has developed along the top surface of the bedrock and a dip angle equal to that of the bedrock, whereas the foot of the slip surface developed across the bedrock.

3.2. Deformation Characteristics. In October 2006, the reservoir level increased from 135 m to 156 m a.s.l. During this period, the head of the Tanjiahe landslide became dramatically deformed. In July 2007, a collapse with a volume of approximately 300 m³ occurred at an elevation of 350 m

a.s.l. in the middle of the landslide. Many surface cracks with an average length of approximately 30 m, width of approximately 0.2 m, and relative displacement ranging from 0.15 to 0.25 m successively appeared in the middle and head of the landslide during the period from July to September 2007. Additionally, some small collapses with volumes of approximately 3 m³ appeared near the cracks. To monitor the deformation process of the landslide, four GPS monitoring stations (ZG287, ZG288, ZG289, and ZG290) were installed along the main sliding direction of the landslide. The frequency of monitoring was once a month.
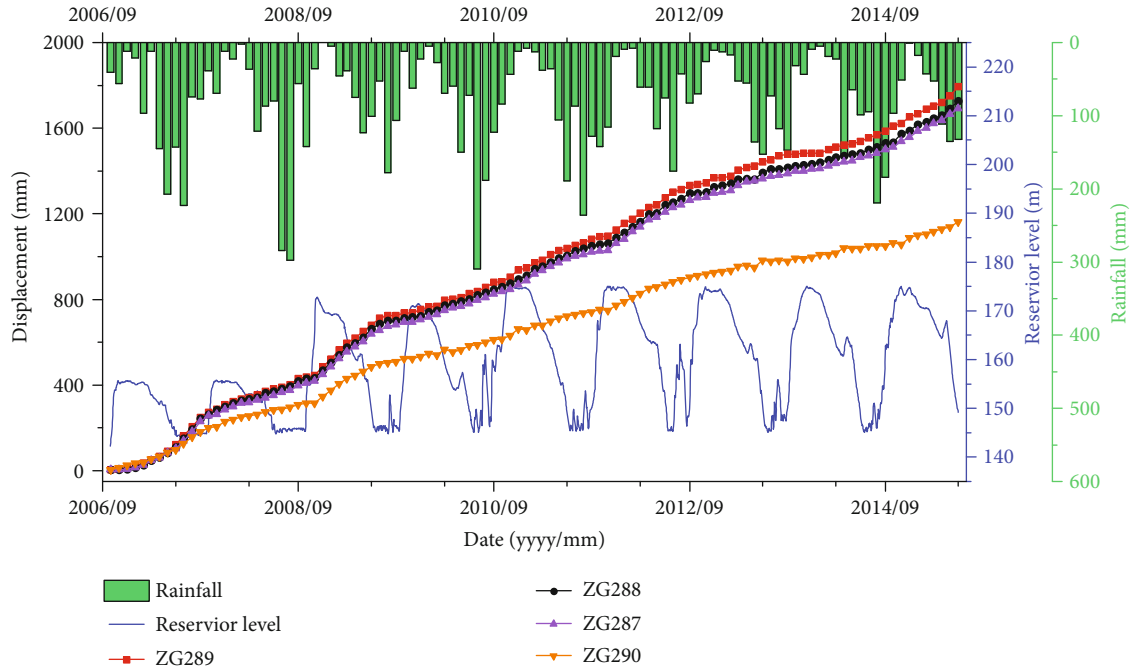
Figure 4: Time series of Tanjiahe landslide monitoring data spanning the period from October 2006 to June 2015.

Figure 4 displays the monitoring results of the water level, rainfall, cumulative displacement, and displacement velocity over a ten-year period from October 2006 to June 2015 [32]. An analysis of the monitoring data from all four permanent GPS stations indicated similar patterns (step-like and steady growth tendency) of landslide movements but different deformation sizes. Station ZG289, which was located in the middle of the main body, measured the maximum displacement, and the cumulative displacement of stations ZG287, ZG288, and ZG289, which were located at the head and foot of the main body, was significantly larger than that of the frontal station ZG290.

Figure 5 shows the correlation analysis between the displacement velocity and reservoir level and rainfall and indicates that the deformation velocity exhibited fluctuations due to the periodic fluctuations in rainfall and the reservoir water level. The shaded areas represent the dry season within a particular year. The dry season generally occurred between October of the current year and April of the following year, during which the reservoir water level was high, remaining between approximately 160 m and 175 m. Almost all periods of rapid movement occurred during the dry season. Considering the lag effect of the increase in the reservoir level on landslide displacement deformation, it can be inferred that the reservoir level rose significantly under the influence of landslide deformation. Compared with that in the dry season, the landslide deformation velocity in the rainy season was relatively low except in 2007, when the deformation velocity measured at the ZG288 monitoring station reached the maximum value of 51.3 mm/month. This finding may be attributed to the initial filling of the reservoir from 135 m to 156 m between September and November 2006, which resulted in a longer period of landslide stress adjustment and, combined with the heavy rainfall from April to Septem-

ber 2007, resulted in severe landslide deformation. Based on these findings, it can be stated that the reservoir level and heavy rainfall were crucial trigger factors that influenced the deformation of the Tanjiahe landslide and an increase in the reservoir level was the main influential factor that significantly increased the deformation. The material of the sliding mass within the water-level fluctuation zone is composed of blocky rock with high permeability, which makes the groundwater in the landslide discharge in time when the reservoir level fluctuates. Meanwhile, the sliding surface is steep in the upper part and gentle in the lower part; the gentle sliding surface is more conducive to resist sliding. The rising reservoir water level increases the positive pore water pressures, which induces a decrease of the effect stress and resistance force in the sliding surface, thereby accelerating the deformation of the landslide.

3.3. Probabilistic Forecasting of Landslide Displacement. The proposed B-LSSVM method was applied to predict the displacement of the four monitoring points (ZG287, ZG288, ZG289, and ZG290) of the Tanjiahe landslide. The main trigger factors, such as the reservoir level and rainfall, were also considered in the prediction. Both point prediction results and PIs were obtained by the B-LSSVM method. The process of prediction is described as follows.

(1) Input Variable Selection. The evolution process of landslide displacement is dominated by both internal factors (e.g., geological conditions) and trigger factors (e.g., reservoir water-level fluctuations and rainfall). Based on previous studies on landslide displacement forecasting conducted in the TGRA [3, 9, 12, 21, 23], displacements over the previous one, two, and three months were selected as the input variables related to internal factors. The average elevation of
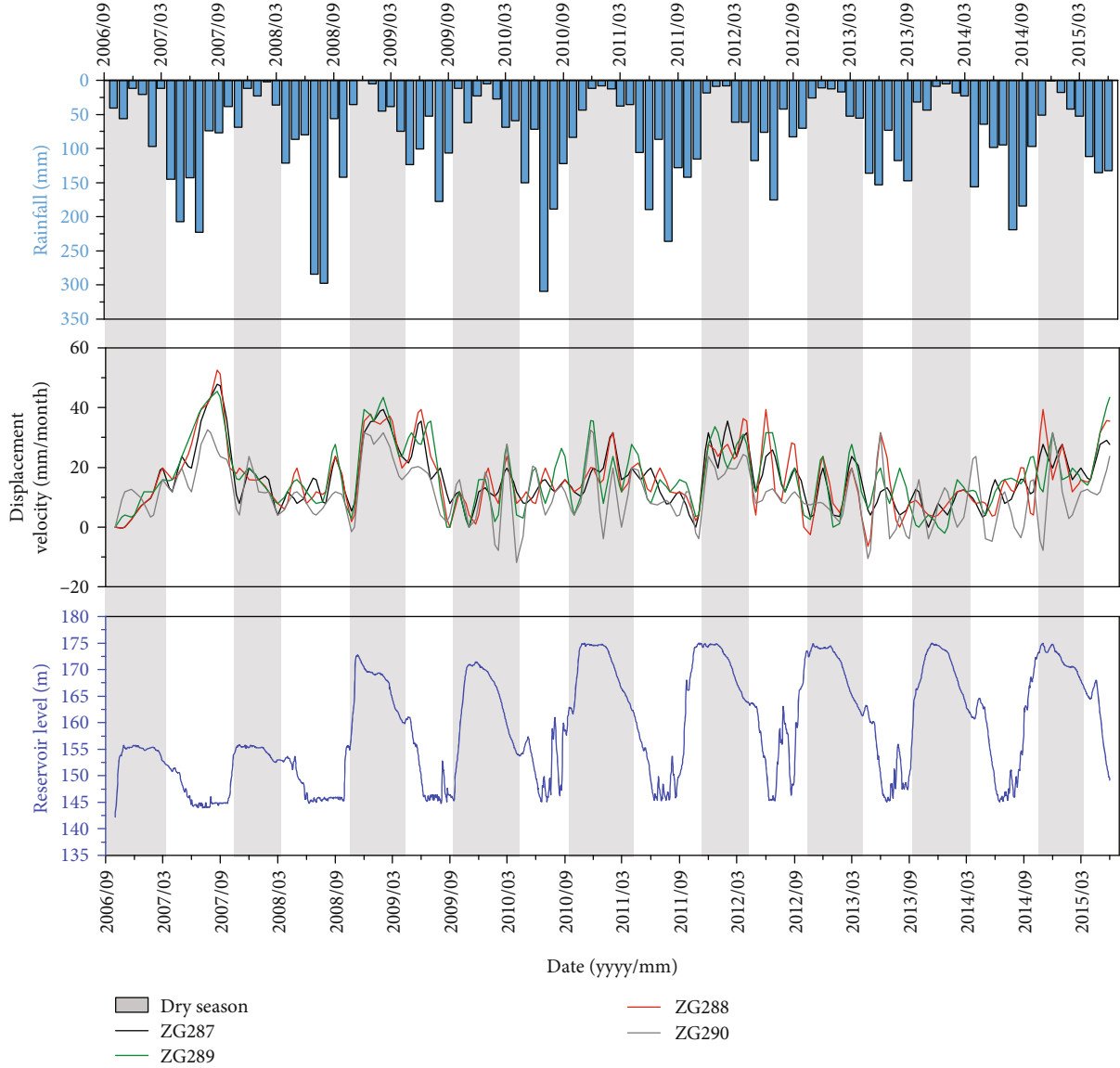
FIGURE 5: Correlation analysis between the displacement velocity and reservoir level and rainfall.

the reservoir water level in the current month, the rainfall over the previous one and two months, and the variation in the reservoir level during the previous one month were selected as the input variables related to trigger factors. In this study, these seven factors were also chosen as the input variables to conduct landslide displacement one-step-ahead forecasting.

(2) Data Splitting. In this step, 70% of the monitoring data (from October 2006 to December 2012) were treated as the training dataset and the remaining 30% (from January 2013 to June 2015) were selected as the testing dataset. To reduce the dimensional effect of the data on the predictive performance, the training and testing datasets were normalized within the limits (-1, 1). The normalized datasets were then applied to the LSSVM model for training and forecasting.

(3) Bootstrap Sampling and LSSVM Training. The number of bootstrap replications may influence the quality of PIs. To determine the appropriate number of bootstrap replications, 50, 100, 200, 500, 1000, 2000, and 5000 bootstrap replications were implemented for each monitoring point case. Thus, different ensemble numbers of LSSVM models were formed, and they were used for forecasting both the training set and the testing set.

(4) Regression Means and Model Variance Estimation. Based on the prediction results of the ensemble LSSVM model, the regression means and model variance of the training set and the testing set can be estimated by Equations (9) and (10).

(5) Noise Quantification. After the regression means and model variance of the training set are calculated, the residual set can be built. A new LSSVM model tuned by a GA was

TABLE 1: PI performance index of ZG287.

B	PICP		NMPIW		CWC	
	B-LSSVM	B-ELM	B-LSSVM	B-ELM	B-LSSVM	B-ELM
50	0.9333	0.9667	0.0835	0.2543	0.0893	0.2699
100	0.9333	1.0000	0.0805	0.2425	0.0861	0.2435
200	0.9667	1.0000	0.0765	0.2536	0.0775	0.2546
500	0.9667	1.0000	0.0782	0.2685	0.0792	0.2695
1000	0.9667	1.0000	0.0736	0.2681	0.0746	0.2691
2000	0.9333	1.0000	0.0774	0.2585	0.0829	0.2595
5000	0.9333	1.0000	0.0735	0.2430	0.0788	0.2440

TABLE 2: PI performance index of ZG288.

B	PICP		NMPIW		CWC	
	B-LSSVM	B-ELM	B-LSSVM	B-ELM	B-LSSVM	B-ELM
50	0.9333	1.0000	0.0936	0.3110	0.1000	0.3120
100	0.9000	1.0000	0.1206	0.2952	0.2005	0.2962
200	0.9333	1.0000	0.1005	0.2632	0.1073	0.2642
500	0.9333	1.0000	0.1034	0.3213	0.1104	0.3223
1000	0.9333	1.0000	0.0938	0.3195	0.1002	0.3205
2000	0.9333	1.0000	0.0944	0.2916	0.1008	0.2926
5000	0.9000	1.0000	0.0980	0.2653	0.1633	0.2663

used to train the residual set. In the GA, the searching ranges of γ and δ are $e^{(-10)}$ and $e^{(10)}$, and the number of iterations, population size, crossover fraction, and migration fraction were set to 200, 50, 0.8, and 0.2, respectively. By minimizing the cost function (14) in the GA, the optimal hyperparameters of LSSVM could be obtained. Then, the noise variance of the testing set was predicted by the optimal LSSVM.

(6) PI Construction. The confidence level was set to 95% (α equal to 0.05). Once the training and testing processes were completed, the predicted displacements were derived after unnormalizing the predicted data. Then, the point predictions and PIs with a 95% confidence level were obtained.

3.4. Results and Analysis. The purpose of this paper is to accurately predict the displacement of the Tanjiahe landslide using the B-LSSVM method and quantify the uncertainties in the prediction to further establish a high-quality PI for landslide displacement. To verify the efficacy and superiority of the B-LSSVM method, multiple methods, including BP, ELM, and LSSVM tuned by the default optimization algorithm of the LS-SVMlab toolbox, were applied for point prediction comparison, and a hybrid model of ANN-based PIs, namely, B-ELM [16], was used for probabilistic forecast comparison. B-ELM is a hybrid method that combines the bootstrap, ELM, and ANN methods. Ma et al. [21] applied the B-ELM method for interval prediction of landslide displacement. In B-ELM, bootstrapping of residuals is performed and the bootstrap replicate number is set to 200. The number of hidden nodes in the ELM is determined by

a 10-fold cross-validation method using only the training set. The node numbers are determined to be 29, 31, 30, and 31 for ZG287, ZG288, ZG289, and ZG290, respectively. The trained ELMs are applied for 200 bootstrap samples to estimate the regression mean and variance of model uncertainty, and an ANN model is used to estimate the noise variance. The prediction results for the proposed method and other methods are presented in Tables 1–5 and Figures 6 and 7.

Tables 1–4 show a comparison of the PI performance between the B-ELM method and the B-LSSVM method. Different bootstrap replications were implemented for each monitoring point. The PICP, NMPIW, and CWC were computed to assess the quality of the PIs. As shown in Tables 1–4, both methods are reliable and can provide satisfactory predictions for the test samples. The PICPs of B-LSSVM and B-ELM are all very close to the 95% confidence level, and the PICP of B-ELM is even higher than the 95% confidence level. For example, the PICP of B-ELM mostly reached 100% in Tables 1–3. This result seems to indicate that the B-ELM method is more reliable than the B-LSSVM method. However, compared with the B-LSSVM method, the B-ELM method performed poorly with respect to the NMPIW. In Tables 1–4, the mean NMPIW of the B-ELM was approximately 2~5 times larger than that of the B-LSSVM. This result shows that the B-ELM method tended to construct broad PIs to achieve a high PICP compared with the B-LSSVM method. Overfitting may occur in the training phase of the ELM model, which leads to a low generalization ability and large prediction error in the testing phase. Therefore, the variance of model uncertainty, i.e., $\widehat{\sigma}_g^2(x_i)$ in

TABLE 3: PI performance index of ZG289.

B	PICP		NMPIW		CWC	
	B-LSSVM	B-ELM	B-LSSVM	B-ELM	B-LSSVM	B-ELM
50	0.9333	1.0000	0.0890	0.2315	0.0951	0.2325
100	0.9333	1.0000	0.0853	0.2478	0.0912	0.2488
200	0.9333	1.0000	0.0857	0.2541	0.0917	0.2551
500	0.9333	1.0000	0.0867	0.2571	0.0928	0.2581
1000	0.9667	0.9667	0.0863	0.2792	0.0873	0.2802
2000	0.9333	1.0000	0.0862	0.2805	0.0922	0.2815
5000	0.9333	1.0000	0.0861	0.2735	0.0921	0.2745

TABLE 4: PI performance index of ZG290.

Replications	PICP		NMPIW		CWC	
	B-LSSVM	B-ELM	B-LSSVM	B-ELM	B-LSSVM	B-ELM
50	0.9333	0.9333	0.1669	0.3152	0.1775	0.3343
100	0.9333	0.9667	0.1641	0.3275	0.1746	0.3285
200	0.9333	0.9667	0.1612	0.3387	0.1715	0.3397
500	0.9333	0.9667	0.1634	0.2990	0.1738	0.3000
1000	0.9333	1.0000	0.1561	0.3267	0.1661	0.3277
2000	0.9333	0.9667	0.1582	0.3076	0.1683	0.3086
5000	0.9333	0.9667	0.1593	0.3186	0.1695	0.3196

TABLE 5: Summary of the point prediction results of the displacement of the four GPS monitoring points.

Method	MAPE				RMSE			
	ZG287	ZG288	ZG289	ZG290	ZG287	ZG288	ZG289	ZG290
BP	1.621	1.363	1.431	1.664	31.979	23.730	27.835	22.977
ELM	0.966	1.264	1.334	1.252	18.497	23.531	30.776	16.884
LSSVM	0.430	0.567	0.583	0.869	8.537	10.997	11.283	10.528
B-LSSVM	0.425	0.555	0.547	0.868	8.162	10.910	10.697	10.508

Equation (10), was large, resulting in a wide PI. In contrast, the B-LSSVM method can not only provide a satisfactory PICP but also yield a narrower NMPIW. From Tables 1–4, under all the bootstrap replications, the CWC of B-LSSVM was always significantly lower than that of B-ELM at different bootstrap replications. Since the CWC is a comprehensive index balancing the PICP and NMPIW, it can be stated that the B-LSSVM method outperforms the B-ELM model and can generate more high-quality PIs in Tanjiahe landslide displacement prediction.

Moreover, a significant improvement in the quality of PIs was not observed with an increase in bootstrap replications. The B-LSSVM with 200 bootstrap replications can always obtain a satisfactory prediction performance, indicating that the chosen number of bootstrap replications is reasonable. At each monitoring point, the CWCs of both the B-LSSVM and B-ELM methods exhibited no major vibration at different bootstrap replications. The lowest CWC obtained was less than or equal to 1000 bootstrap replications. The

B-LSSVM method constructed optimal PIs of ZG287, ZG288, ZG289, and ZG290 at 1000, 50, 100, and 1000 bootstrap replications, respectively. The B-ELM method constructed optimal PIs of ZG287, ZG288, ZG289, and ZG290 at 50, 200, 1000, and 500 bootstrap replications, respectively. Figure 6 depicts the optimal PIs with a 95% confidence level using the two methods. As the figure shows, the four constructed PIs mostly covered the actual displacements, while the $PI_{B-LSSVM}$ values were much narrower than the PI_{B-ELM} values. The figure clearly shows that the $PI_{B-LSSVM}$ values were encompassed by the PI_{B-ELM} values.

Table 5 shows a summary of the point prediction results of the displacement of the four GPS monitoring points obtained by the BP, ELM, single LSSVM, and B-LSSVM methods. The RMSE and MAPE were calculated for a quantitative assessment and comparison of the different methods' performance. As shown in Table 5, the B-LSSVM method generates the highest point prediction accuracy. The RMSE and MAPE of the B-LSSVM method are the smallest of the
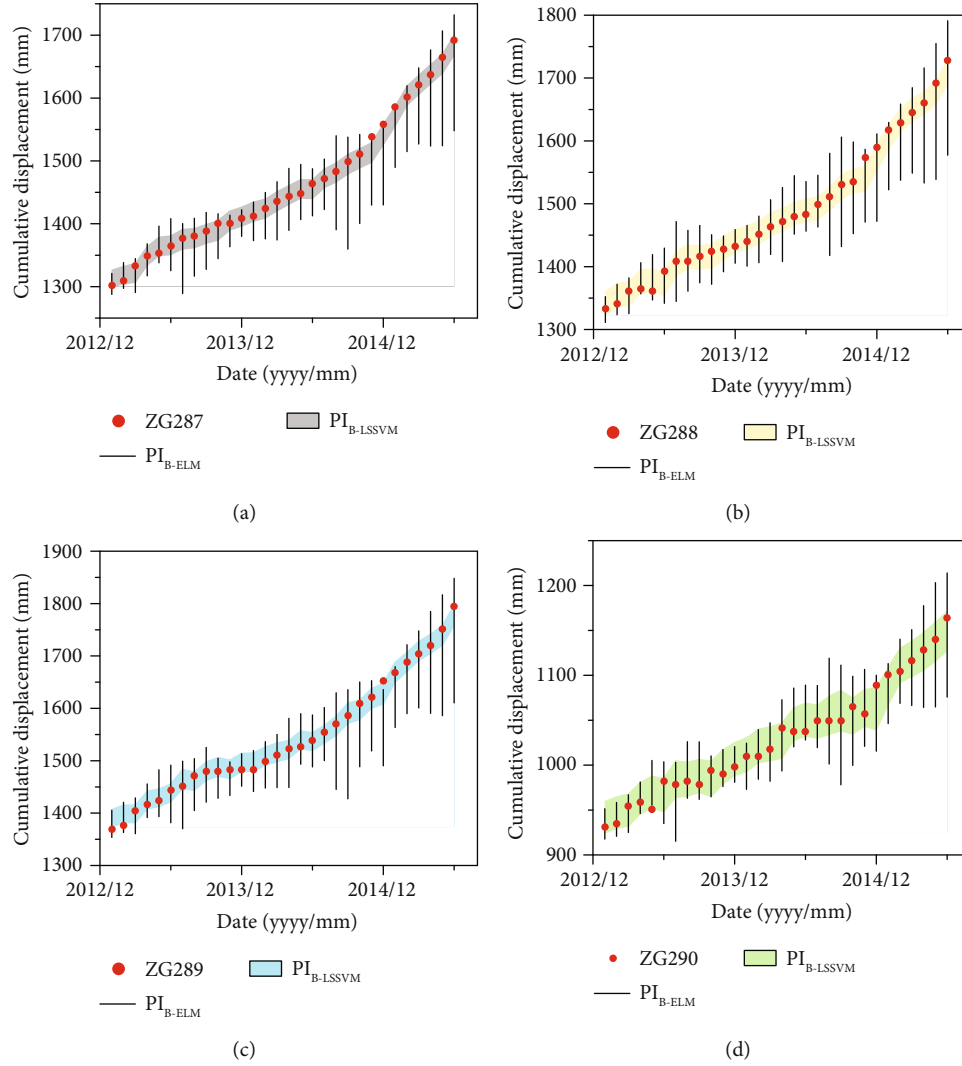
(a)

(b)

(c)

(d)

FIGURE 6: Optimal PIs of the displacement of four monitoring points with a 95% confidence level using the B-LSSVM and B-ELM methods (PI$_{B\text{-}ELM}$ denotes the PI constructed using the B-ELM method and PI$_{B\text{-}LSSVM}$ denotes the PI constructed using the B-LSSVM method).

four GPS monitoring points, indicating that the proposed method performs the best among the compared methods and is capable of providing accurate point prediction for landslide displacement. The RMSEs for ZG287, ZG288, ZG289, and ZG290 obtained by the B-LSSVM method were 0.425, 0.555, 0.547, and 0.868, respectively, and the MAPEs were 8.162, 10.910, 10.697, and 10.508, respectively. These values are slightly lower than those obtained using the original LSSVM model, which denotes that the ensemble LSSVM prediction method can improve the performance of the single LSSVM method. The prediction accuracy of the SVM-based methods (i.e., LSSVM and B-LSSVM) was obviously higher than that of the ANN-based methods (i.e., BP and ELM). For instance, the MAPE and RMSE of B-LSSVM in ZG289 are approximately three times smaller than those of the BP and ELM. The good performance of the SVM-based methods may be attributed to the excellent generalization and prediction ability with a small sample dataset, which is less likely to result in overfitting compared with ANN-

based methods. Figure 7 shows a comparison between the point forecast results of the B-LSSVM method and the measured values, and the prediction curves of the B-LSSVM method were in good agreement with reality in both the training set and the testing set. Based on the above results and analysis, it can be stated that the B-LSSVM method provides accurate point predictions and yields reliable and high-quality PIs in landslide displacement prediction.

4. Conclusions

In this paper, a hybrid method for providing accurate point and reliable PI estimations of landslide displacement was proposed based on the combination of bootstrapping and LSSVM. With this method, the uncertainties associated with the traditional deterministic prediction can be quantified. The Tanjiahe landslide was used to test the performance of this method. The results indicate that the B-LSSVM is stable and the number of bootstrap replications has less effect
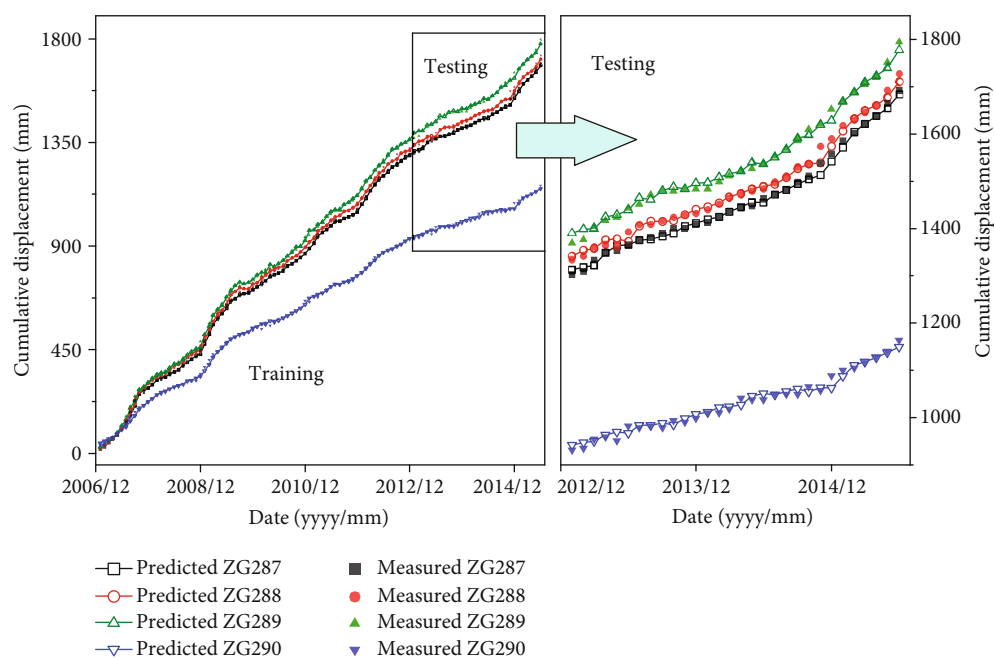
Figure 7: Actual value and point prediction value of the B-LSSVM method.

on the quality of PIs constructed by the B-LSSVM. The B-LSSVM method performs better than the BP, ELM, and LSSVM methods in point displacement prediction, and it is also superior to the B-ELM method in interval prediction of landslide displacement. Therefore, the B-LSSVM method is a promising tool for providing more valuable information and further confidence for decision-makers in mitigation decisions.

It should be noted that the conclusions are obtained based on the site-specific Tanjiahe landslide. The applicability of the B-LSSVM method to other landslides needs to be verified in future studies. In addition, the prediction performance of B-LSSVM can be affected by the selected input factors. Future research should fully investigate the effect of the influencing factors on landslide displacement prediction.

Data Availability

The data used to support the findings of this study are available from the corresponding authors upon request.

Disclosure

The authors thank the 2018 European Geosciences Union General Assembly for publishing the abstract of this paper, which can be found in the following link: https://meetingorganizer.copernicus.org/EGU2018/EGU2018-3600.pdf.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] H. Tang, J. Wasowski, and C. H. Juang, "Geohazards in the Three Gorges Reservoir Area, China - lessons learned from decades of research," *Engineering Geology*, vol. 261, p. 105267, 2019.

[2] S. Qin, J. J. Jiao, and S. Wang, "A nonlinear dynamical model of landslide evolution," *Geomorphology*, vol. 43, no. 1-2, pp. 77–85, 2002.

[3] Y. Cao, K. Yin, D. E. Alexander, and C. Zhou, "Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors," *Landslides*, vol. 13, no. 4, pp. 725–736, 2016.

[4] J. Du, K. Yin, and S. Lacasse, "Displacement prediction in colluvial landslides, Three Gorges Reservoir, China," *Landslides*, vol. 10, no. 2, pp. 203–218, 2013.

[5] F. Huang, J. Huang, S. Jiang, and C. Zhou, "Landslide displacement prediction based on multivariate chaotic model and extreme learning machine," *Engineering Geology*, vol. 218, pp. 173–186, 2017.

[6] F. Huang, K. Yin, G. Zhang, L. Gui, B. Yang, and L. Liu, "Landslide displacement prediction using discrete wavelet transform

and extreme learning machine based on chaos theory," *Environment and Earth Science*, vol. 75, no. 20, pp. 1–18, 2016.

[7] C. Lian, Z. Zeng, W. Yao, and H. Tang, "Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level," *Stochastic Environmental Research and Risk Assessment*, vol. 28, no. 8, pp. 1957–1972, 2014.

[8] Z. Cai, W. Xu, Y. Meng, C. Shi, and R. Wang, "Prediction of landslide displacement based on GA-LSSVM with multiple factors," *Bulletin of Engineering Geology and the Environment*, vol. 75, no. 2, pp. 637–646, 2016.

[9] F. Miao, Y. Wu, Y. Xie, and Y. Li, "Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model," *Landslides*, vol. 15, no. 3, pp. 475–488, 2018.

[10] F. Ren, X. Wu, K. Zhang, and R. Niu, "Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China," *Environmental Earth Sciences*, vol. 73, no. 8, pp. 4791–4804, 2015.

[11] T. Wen, H. Tang, Y. Wang, C. Lin, and C. Xiong, "Landslide displacement prediction using the GA-LSSVM model and time series analysis: a case study of Three Gorges Reservoir, China," *Natural Hazards and Earth System Sciences*, vol. 17, no. 12, pp. 2181–2198, 2017.

[12] J. Ma, H. Tang, X. Hu et al., "Identification of causal factors for the Majiagou landslide using modern data mining methods," *Landslides*, vol. 14, no. 1, pp. 311–322, 2017.

[13] J. Ma, H. Tang, X. Liu, X. Hu, M. Sun, and Y. Song, "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*, vol. 14, no. 3, pp. 1275–1281, 2017.

[14] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Networks*, vol. 19, no. 2, pp. 225–235, 2006.

[15] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

[16] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 463–470, 2014.

[17] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 303–315, 2014.

[18] H. Zhang, J. Zhou, L. Ye, X. Zeng, and Y. Chen, "Lower upper bound estimation method considering symmetry for construction of prediction intervals in flood forecasting," *Water Resources Management*, vol. 29, no. 15, pp. 5505–5519, 2015.

[19] C. Lian, Z. Zeng, W. Yao, H. Tang, and C. L. P. Chen, "Landslide displacement prediction with uncertainty based on neural networks with random hidden weights," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 12, pp. 2683–2695, 2016.

[20] C. Lian, C. L. P. Chen, Z. Zeng, W. Yao, and H. Tang, "Prediction intervals for landslide displacement based on switched neural networks," *IEEE Transactions on Reliability*, vol. 65, no. 3, pp. 1483–1495, 2016.

[21] J. Ma, H. Tang, X. Liu et al., "Probabilistic forecasting of landslide displacement accounting for epistemic uncertainty: a case study in the Three Gorges Reservoir area, China," *Landslides*, vol. 15, no. 6, pp. 1145–1153, 2018.

[22] Y. Wang, H. Tang, T. Wen, and J. Ma, "A hybrid intelligent approach for constructing landslide displacement prediction intervals," *Applied Soft Computing*, vol. 81, p. 105506, 2019.

[23] C. Zhou, K. Yin, Y. Cao, and B. Ahmed, "Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China," *Engineering Geology*, vol. 204, pp. 108–120, 2016.

[24] M.-Y. Cheng and N.-D. Hoang, "Interval estimation of construction cost at completion using least squares support vector machine," *Journal of Civil Engineering and Management*, vol. 20, no. 2, pp. 223–236, 2014.

[25] J. H. Zhao, Z. Y. Dong, Z. Xu, and K. P. Wong, "A statistical approach for interval forecasting of the electricity price," *IEEE Transactions on Power Apparatus and Systems*, vol. 23, no. 2, pp. 267–276, 2008.

[26] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[27] X. Zhu, Q. Xu, M. Tang, W. Nie, S. Ma, and Z. Xu, "Comparison of two optimized machine learning models for predicting displacement of rainfall-induced landslide: a case study in Sichuan Province, China," *Engineering Geology*, vol. 218, pp. 213–222, 2017.

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[29] K. BrabanterDe, P. Karsmakers, F. Ojeda et al., *LS-SVMlab Toolbox User's Guide: Version 1.8*, Katholieke Universiteit Leuven, Leuven, Belgium, 2010.

[30] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, CRC press, 1994.

[31] L. Davis, *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York, 1991.

[32] P. Huang, *The Failure Characteristics of Rock Bedding Landslide in Three Gorges Reservoir Area of Shazhenxi*, China Three Gorges University, Chinese, 2016.