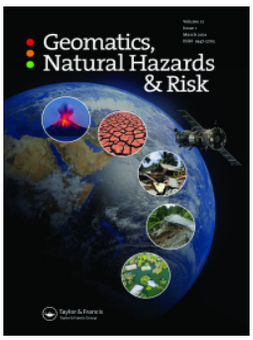

Geomatics, Natural Hazards and Risk

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgnh20

Modelling and predicting landslide displacements and uncertainties by multiple machine-learning algorithms: application to Baishuihe landslide in Three Gorges Reservoir, China

Yanan Jiang, Qiang Xu, Zhong Lu, Huiyuan Luo, Lu Liao & Xiujun Dong

To cite this article: Yanan Jiang, Qiang Xu, Zhong Lu, Huiyuan Luo, Lu Liao & Xiujun Dong (2021) Modelling and predicting landslide displacements and uncertainties by multiple machine-learning algorithms: application to Baishuihe landslide in Three Gorges Reservoir, China, Geomatics, Natural Hazards and Risk, 12:1, 741-762, DOI: 10.1080/19475705.2021.1891145

To link to this article: https://doi.org/10.1080/19475705.2021.1891145

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 02 Mar 2021.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tgnh20

GEOMATICS, NATURAL HAZARDS AND RISK
2021, VOL. 12, NO. 1, 741–762
https://doi.org/10.1080/19475705.2021.1891145

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Modelling and predicting landslide displacements and uncertainties by multiple machine-learning algorithms: application to Baishuihe landslide in Three Gorges Reservoir, China

Yanan Jiang[a,b,c], Qiang Xu[a], Zhong Lu[c], Huiyuan Luo[b], Lu Liao[d] and Xiujun Dong[a]

[a]State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu, China; [b]College of Earth Science, Chengdu University of Technology, Chengdu, China; [c]Huffington Department of Earth Sciences, Southern Methodist University, Dallas, TX, USA; [d]Technology Service Center of Surveying and Mapping, Sichuan Bureau of Surveying, Mapping and Geoinformation, Chengdu, China

ABSTRACT

Time series landslide displacement is the most critical data set to understand landslide characteristics and infer its future development. To predict landslide displacements and their quantitative uncertainties, a mathematical description of the landslide evolution should be established. This paper proposes a novel hybrid machine-learning model to predict landslide displacements and quantify their uncertainties using prediction intervals (PIs). First, wavelet de-noising and Hodrick-Prescott (HP) filters are applied to decompose the original landslide displacement into periodic, trend, and noise components. Second, a module built on the framework of bootstrap and extreme learning machine (ELM) with a hybrid grey wolf optimizer (HGWO) is used to derive a formula for modelling the periodic component of the landslide motion. Another formula for predicting the trend component of the displacement is derived by double exponential smoothing (DES). Grey relational analysis and kernel principal component analysis (KPCA) are used to select the main factors controlling the landslide motions. Finally, the two constructed formulas are used to generate the predictions of landslide displacements and the PIs. Validation and comparison experiments have been carried out on the Baishuihe landslide in the Three Gorge Reservoir of China. Results demonstrate the proposed method can achieve better performance with higher-quality PIs than other state-of-the-art methods.

ARTICLE HISTORY
Received 30 October 2020
Accepted 11 February 2021

KEYWORDS
Landslide displacement; machine learning; prediction intervals; hybrid grey wolf optimizer; extreme learning machine

CONTACT Qiang Xu xq@cdut.edu.cn; Zhong Lu zhonglu@smu.edu

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Landslide is a prevalent and recurrent geological phenomenon worldwide. It poses serious threats to the local communities and could disrupt the roads, tunnels, bridges, farmlands, and the environment (Huang 2009). In the Three Gorges Reservoir (TGR) of China, over 4200 landslides are distributed throughout this region, and the majority of these landslides present characteristics of multiple triggers and reactivations (Yin et al. 2010). Accurate modelling and predicting landslide displacements are essential for the prevention of landslide hazards.

Numerous approaches have been developed for modelling landslide based on physics-based models and statistics-based models. Physics-based models are generally built on creep theory describing the constitutive relationship of rock and soil (Ma et al. 2018; Miao et al. 2018); these models require a wide range of actual observations and laboratory experiments to determine the physical and mechanical parameters, and thus limit their wide application. Statistical models predict landslide displacement based on regression analyses of the historical displacements (Wang et al. 2019). In comparison, statistical models tend to develop the response relationship between landslide displacement and its associated causal factors using a variety of statistical analyses and machine learning methods, which do not require the determination of the physical parameters. Consequently, statistical models are generally easier to implement and have gained increasing popularity.

In previous literature, statistical models have been vigorously discussed and widely used in forecasting landslide displacement. Representative models include the grey system model (Deng 1982), the Verhulst model (Tianbin and Mingdong 1996), multivariate regression models (Jibson 2007), autoregressive-integrated moving average (ARIMA) (Carla et al. 2016), and others. In the past decades, some great progress has been made since the involvement of machine learning (ML) methods in landslide research. ML was first directly used as a landslide prediction model, but got unsatisfactory outputs, mostly because of the suboptimal parameter settings, complicated impact factors, and insufficient observations. Therefore, optimized ML models with hybrid parameters were created.

In recent years, ML models such as artificial neural networks (ANN) (Lian et al. 2015; Ma et al. 2018; Guo et al. 2019), support vector machine (SVM) (Pradhan 2013; Zhou et al. 2016), and various kinds of optimized coupling models (Li et al. 2018; Wang et al. 2019) have gradually become the mainstream as a more powerful approach in landslide research. Generally, SVM has a better performance than the ANN, but the model itself has some defects, e.g. having difficulty in parameter selection (Cao et al. 2016). Nowadays, the extreme learning machine (ELM) has overcome the challenge of parameter initialization and has advantages on global minimum optimization and strong generalization; therefore, it has been successfully tested in many other fields with promising prediction results (Miche et al. 2015; Cao et al. 2016; Tramèr et al. 2017). However, as a single-hidden-layer feed-forward neural network, ELM may introduce additional prediction errors if trained with a small data set (Li et al. 2018). To achieve a better performance, ensemble-based bootstrap sampling and hybrid parameter optimization methods are introduced to ensure the optimal performance of ELM in this study.

The optimization algorithms can be divided into the deterministic algorithms and the stochastic algorithms (Mirjalili 2015), the latter is capable of avoiding local solutions due to randomness. Among the stochastic algorithms, grey wolf optimizer (GWO) proposed by Mirjalili et al. (2014) is a new biological intelligence algorithm that is inspired by the behaviour of grey wolves, gives better quality solution than other old metaheuristic algorithms (Faris et al. 2018; Teng et al. 2019), such as genetic algorithm (GA), ant colony optimization (ACO), particle swarm optimization (PSO), and others (Mirjalili et al. 2014; Gao and Zhao 2019; Wang and Li 2019). However, as one of the swarm intelligence algorithms, GWO is still likely to fall into stagnation when predation suffers slower convergence during some of the long sequential execution time (Zhu et al. 2015). Therefore, a meta-heuristic search method named differential evolution (DE) is introduced to further accelerate the convergence and increase the optimization accuracy of GWO, called the hybrid GWO (HGWO). The HGWO will update the previous best position of each grey wolve to help GWO jump out of the stagnation through DE's strong searching ability.

Besides model building, the selection of input variables also has a big impact on the prediction accuracy in machine learning. The recent researches began to analyse the role of impact factors in landslide evolution during ML modelling (Guo et al. 2019; Miao et al. 2018; Zhou et al. 2018; Zhu et al. 2018); however, only a few have established the response relationship between causal factors and landslide deformation. In consequence, the physical meanings of various components of the landslide displacement time series are unclear, causing the associated causal factors cannot effectively be linked with the corresponding components. What's more, previous researches on landslide displacement forecasting are mainly focused on deterministic prediction (H. Li et al. 2019; Wang et al. 2019; Zhou et al. 2018), which provides only one fixed forecast value at a given time point for each target. Although it can estimate the forecasting error at the point, point prediction gives limited consideration on stochastic behaviours of the landslides system and therefore could not efficiently represent the dynamic uncertainty of landslides.

To solve these problems, this paper proposes a new composite model integrating multiple ML and statistical models to produce unbiased reliable forecasts for landslide displacement. The model combines multi-algorithms to optimize the ML models and quantifies the uncertainty by prediction intervals (PIs). The model aims to establish a more accurate response relationship between causal factors and landslide deformation. The landslide displacement time series consists of a long-term trend dominated by internal geological conditions, a short-term periodical fluctuation affected by external triggering factors, and random noise components. Time series landslide displacements are first denoised and decomposed. Then, a novel composite framework involving bootstrap resampling, double exponential smoothing (DES), and hybrid grey wolf optimizer (HGWO) optimized ELM is built to obtain formulas to predict landslide displacements. Experiments on the Baishuihe landslide in TGR from 2003 to 2013 have confirmed the effectiveness and utility of the proposed model for long-term practical monitoring and forecasting of landslide displacements.

2. Methodology

2.1. Time series analysis of landslide displacements

Generally, the dynamic movement of a landslide is subject to internal geological conditions and external environmental factors. The displacement dominated by geological conditions (e.g. topography, geotechnical properties) is generally found to be approximately monotonic over time, indicating the long-term trend. The displacement affected by external triggering factors (e.g. rainfall and reservoir water variation) can be expressed as a proximate periodic function, leading to different deformation features. The pattern of the landslides displacement in the TGR of China is normally controlled by joint efforts between geological conditions and environmental factors. The accumulated displacement can be decomposed as follows:

$$D_t = T_t + P_t + n_t \tag{1}$$

where D_t is the original accumulated displacement, T_t stands for the trend displacement, P_t is the periodic displacement, and n_t represents the random noise of the displacement.

In this study, a wavelet-based denoising method, which preserves important signals while removing noise by diagnosing features of data at different frequencies, is applied to remove the random noise in original displacement. Then, the de-noised displacement time series is divided into the periodic and trend components by the Hodrick-Prescott (HP) filter. The HP filter has been widely used in macroeconomics due to its strength in deriving smoothed-curve expression of the time series observation. Previous studies have proven that HP is more sensitive to the long-term trend evolution than to the short-term fluctuations (Ravn and Uhlig 2002). The sensitivity of the long-term trend to short-term variation can be adjusted by modifying a parameter λ. Then, the trend component of a time series can be obtained as follows:

$$\min_{\gamma} \left\{ \sum_{t=1}^{T} (y_t - \gamma_t)^2 + \lambda \sum_{t=3}^{T-2} [(\gamma_t - \gamma_{t-1}) - (\gamma_{t-1} - \gamma_{t-2})]^2 \right\} \tag{2}$$

where y_t ($t = 1, 2, \ldots, \mathrm{T}$) denotes the logarithms of data time series; γ represents the trend component; λ is the smoothing parameter, controlling the smoothness of the trend component. The objective of Equation (2) is to obtain a smoothed trend component by minimizing both 1) the misfit between the observed and the trend component (i.e. term 1 in Equation (2)) and 2) the second-order difference of the trend component (i.e. term 2 in Equation (2)). Thus, if the smoothing parameter λ is 0, no smoothing takes place. The larger the value of λ is, the smoother the result is. When the value of λ becomes large enough, for example, $\lambda =$ infinity, the trend component could be considered as linear. In principle, the appropriate value of λ also depends on the periodicity of the original data. For yearly data, $\lambda = 100$ is suggested, while for monthly data $\lambda = 14400$ is suggested (Zhu et al. 2018). In this study, the landslide displacement with sharp acceleration occurred once a year. Consequently, the smoothing parameter $\lambda = 100$ is assigned.

2.2. Analysis of model uncertainty

In the landslide displacement prediction model, the predicted displacement can be expressed as follows:

$$F_i = f(x_{Ti}) + \varepsilon_i = T(x_{Ti}) + P(x_{Pi}) + \varepsilon_i \tag{3}$$

where F_i is the ith observable target values, $T(x_{Ti})$ and $P(x_{Pi})$ represent the true trend and periodic regression outputs of a forecasting model, x_i is the vector of inputs, and ε is the random error assumed to be normally distributed with mean zero and variance σ_ε^2. In practice, the goal of regression is to produce an estimate of the true $\hat{T}(x_{Ti})$ and $\hat{P}(x_{Pi})$. Given all terms in Equation (1) have associated sources of uncertainty, and assuming they are independent, the total variance of prediction is given by the following:

$$\sigma_F^2 = \sigma_f^2 + \sigma_\varepsilon^2 = \sigma_{T\text{model}}^2 + \sigma_{P\text{model}}^2 + \sigma_\varepsilon^2 \tag{4}$$

where $\sigma_{T\text{model}}^2$ and $\sigma_{P\text{model}}^2$ stand for the model uncertainty of the forecasting trend and periodic displacement separately, and σ_ε^2 is the random noise variance.

So far, a diverse set of approaches have been developed to quantify the model uncertainty, ranging from Delta (Chryssolouris et al. 1996), mean-variance estimation (MVE) (Nix and Weigend 1994), Bayesian (Dybowski and Roberts 2011), lower upper bound estimation (LUBE) (Khosravi et al. 2011b), and bootstrap (Khosravi, Nahavandi, Creighton, Srinivasan 2011). Bootstrapping is a statistics-based technique focusing on the simplification of the original time series, allowing estimation of the sampling distribution of almost any statistic (Wan et al. 2014). Besides, this method has good compatibility with other training methods of ML (Pearce et al. 2018); it requires no complex matrices and derivatives during calculations, ensuring efficient implementation.

Model uncertainty can be attributed to several factors including model bias, training data variance, and parameter uncertainty (Pearce et al. 2018). The model bias occurs in two main forms: firstly, how comprehensive are the employed factors affecting typical landslide systems; secondly, how representative is the model representing the landslide process. Accordingly, to mitigate the bias, grey relational analysis (GRA) and kernel principal component analysis (KPCA) are coupled to analyse the possible triggering factors and figure out the potential contribution of each component affecting landslide displacement. Then, an optimized ELM model is applied to acquire the global minimum optimization by minimizing the output errors; consequently, the model misspecification could be assumed zero. The parameter uncertainty results from sub-optimal parameters are assigned for the forecasting model. Besides, ML related algorithm is sensitive to parameter setting; various weighting parameters lead to varying outputs even if with the same inputs. Therefore, an ensemble-based ELM trained with different parameter initializations (parameter resampling) and an HGWO parameter method are proposed to derive the optimum parameters. As for the training data uncertainty, it is solved by a bootstrapping. The final resulting

ensembles of ELMs contain a certain level of diversity, which can be used to estimate the model uncertainty and construct the PIs given a certain confidence level.

2.3. Forecast model and evaluations

In this study, an integrated model, based on the framework of bootstrap and ELM optimized by the HGWO, is used to obtain the prediction formula for the periodic component and calculate the trend prediction formula aided by DES. The corresponding predictions are summated to get the forecasted cumulative displacements and the prediction intervals (PIs). The main processes are illustrated in Figure 1.

2.3.1. Double exponential smoothing

The DES is designed to process time series exhibiting a certain trend (Holt 2004). This approach works by applying a weighted combination of the past observations with recent observations given relatively higher weight than the older ones (Zhu et al. 2018), and updating the slope component through exponential smoothing. In this study, DES is employed to predict the trend displacement of a landslide with an obvious pattern.

Given an observed time series $\{x_i\}$, we begin at the time $i = 0$ to indicate an observation sequence. $\{s_i\}$ denotes the smoothed value of the DES at the time i, $\{b_i\}$ indicates the best estimate of the trend displacement at time i. Thus, the output of DES can be written as F_{i+m}, an estimate of x at the time $i+m$, where $m \geq 1, m \in N$, based on time series data available before time i. Then, DES can be performed based on the following formulas:

$$F_{i+m} = s_i + mb_i \tag{5}$$

$$s_i = \zeta x_i + (1-\zeta)(s_{i-1} + b_{i-1}) \tag{6}$$

$$b_i = \xi(s_i - s_{i-1}) + (1-\xi)b_{i-1} \tag{7}$$

where $i = 1, 2, 3..., \zeta, s_1 = x_1$, $b_1 = x_1 - x_0$, ζ is the data smoothing factor with $0<\zeta<1$, ξ is the trend soothing factor with $0 < \xi < 1$.

2.3.2. Extreme learning machine

ELM is a hidden-layer feed-forward neural network training/learning method. ELM has advantages on strong generalization and can learn thousands of times faster than networks trained using backpropagation (Huang et al. 2006). In this algorithm, the initial parameter including the weight matrix connecting the input to the hidden layer, the threshold of hidden layer neurons is randomly assigned. During the model training, the output weights of the hidden layer and the global optimal solution can be obtained when the number of hidden neurons is set.

Given a set of training data (x_i, t_i), where $(x_i, t_i) \in R^n \times R^m (i = 1, 2, ..., N)$. A standard ELM with $\tilde{N}$ hidden neurons with an activation function $f(x)$ can be written as follows:

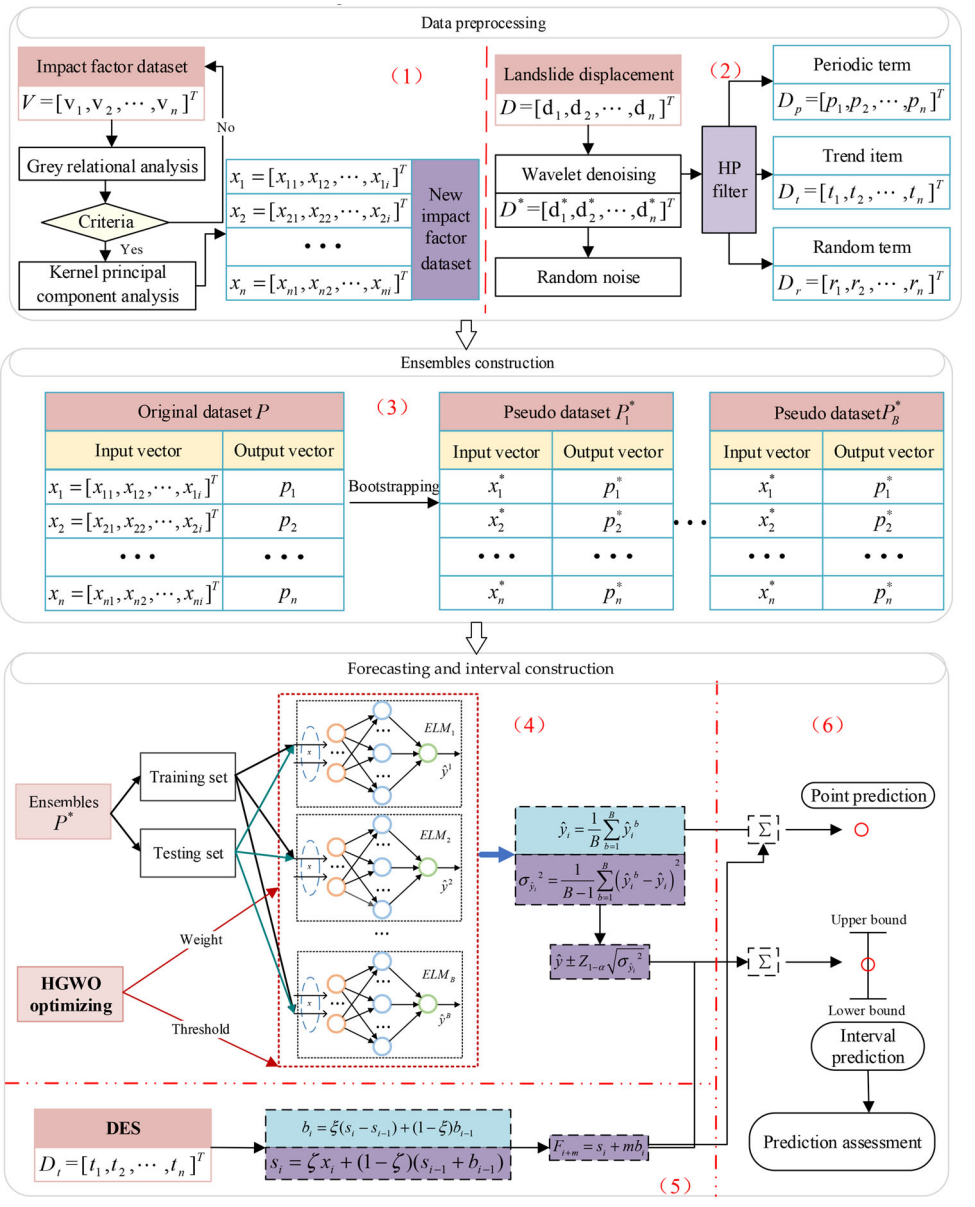

Figure 1. Flowcharts of the forecasting model.

$$\sum_{i=1}^{\tilde{N}} \beta_i f_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i f_i(a_i \cdot x_j + b_i) = t_j, \quad j = 1, \ldots, N \tag{8}$$

where $a_i = [a_{i1}, a_{i2}, \ldots, a_{in}]^T$ is the weight matrix connecting the input to the hidden layer,b_i is the threshold of hidden layer neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weights matrix connecting the hidden to output neurons. We rewrite the above equation compactly as $H\beta = T$, where H is the output matrix of the hidden layer, β is output weigh matrix, and T is the targets matrix. These variables can be expressed as follows:

$$H = \begin{bmatrix} f(a_1 \cdot x_1 + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ f(a_1 \cdot x_N + b_1) & \cdots & f(a_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}, \quad \beta = \begin{bmatrix} {\beta_1}^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m}, \quad T = \begin{bmatrix} {t_1}^T \\ \vdots \\ {t_N}^T \end{bmatrix}_{N \times m} \tag{9}$$

The ELM training process is equivalent to finding the least-squares solution $\hat{\beta}$ of Equation (10),

$$\|H(a_1, ..., a_{\tilde{N}}, b_1, ..., b_{\tilde{N}})\hat{\beta} - T\| = \min_{\beta} \|H(a_1, ..., a_{\tilde{N}}, b_1, ..., b_{\tilde{N}})\beta - T\| \tag{10}$$

The smallest norm least-squares solution is $\hat{\beta} = H^{\dagger}T$, where $H^{\dagger}$ is the Moore-Penrose generalized inversion of the matrix H.

2.3.3. Hybrid grey wolf optimizer

Grey wolf optimizer (GWO) is a recent meta-heuristic algorithm that is inspired by the behaviour of grey wolves (Mirjalili et al. 2014). This algorithm embeds the biological evolution and the 'survival of the fittest' (SOF) principle of biological updating of nature, which leads to favourable convergence. However, the GWO is likely to fall into stagnation when predation suffers slower convergence due to the long sequential execution time (Zhu et al. 2015). Thus, a meta-heuristic search method named differential evolution (DE) algorithm is introduced to help the GWO to jump out of the stagnation, this hybrid GWO is named HGWO.

Assuming the population size is N, the ith wolf in a search dimension d can be written as

$$X^i = \{X_1^i, X_2^i, ..., X_d^i\} \tag{11}$$

Thus, the position of the p th $(p = 1, 2, ...d)$ component of the ith individual can be expressed as

$$X_p^k = X_{p(\text{low})}^k + \left(X_{p(\text{up})}^k - X_{p(\text{low})}^k\right) \times \text{rand}(0, 1) \tag{12}$$

where $k = 1, 2, ..., N$ $X_{p(\text{low})}^k$ and $X_{p(\text{up})}^k (k = 1, 2, ..., N)$ denote the lower and upper boundaries of individual wolves in the search landscape, respectively, $\text{rand}(0, 1)$ represents a random number in $[0, 1]$.

The encircling behaviour of grey wolves can be model as follows:

$$D(t) = \left|C \cdot X_p(t) - X(t)\right| \tag{13}$$

$$X(t + 1) = X_p(t) - A \cdot D \tag{14}$$

where t represents the current iteration, A and C are coefficient vectors, X indicates the position vector of a grey wolf, X_p stands for the position vector of the prey. The coefficient vectors $A = 2a \cdot r_1 - a, C = 2 \cdot r_2$, r_1 *and* r_2 are random variables in the

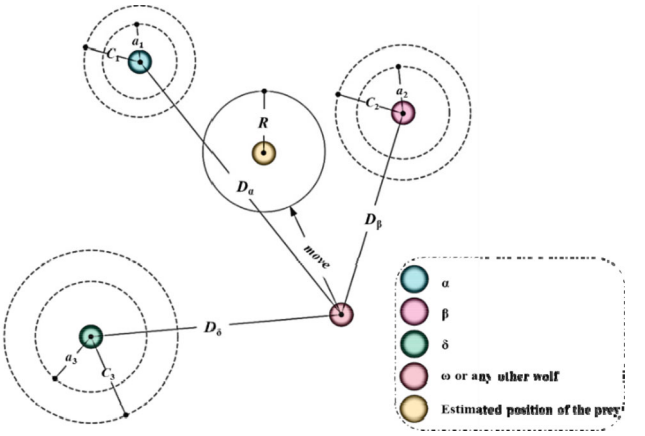

Figure 2. Position updating in HGWO.

range of [0,1], a is the control coefficient, which linearly decreases from 2 to 0 throughout the iterations.

The HGWO allows relocating a solution around another in an n-dimensional search space to simulate chasing and encircling prey by grey wolves in nature. During the simulation, the grey wolves will first search and track the prey, and then encircle it. However, in fact, the optimal prey position is unknown. To mimic wolves internal leadership hierarchy, the wolves are divided into four types: alpha (α), beta (β), delta (δ), and omega (ω). The objective is to find the minimum in this search landscape; thus in HGWO, it assumes that positions of α, β, and δ are likely to be in the prey (optimum) positions. In the iterative search, the best three individuals obtained are recorded respectively as α, β, and δ. Other wolves denoted as ω relocate their positions according to the locations of α, β, and δ. Figure 2 shows how a search agent updates its position according to α, β, and δ in a 2D search space. It can be observed that the final position would be in a random place within a circle, which is defined by the positions of α, β, and δ in the search space. In other words, α, β, and δ estimate the position of the prey, and other wolves update their positions randomly around the prey. The following mathematical formulas are used to update the positions of wolf ω.

$$D_\alpha = |C_1 \cdot X_\alpha - X|, \quad D_\beta = |C_2 \cdot X_\beta - X|, \quad D_\delta = |C_3 \cdot X_\delta - X| \tag{15}$$

where X_α, X_β, and X_δ represents the position of wolf α, β, and δ respectively;C_1, C_2, and C_3 are randomly generated vectors. Equations (15)–(17) calculate the distances between the position of the current individual and that of individualα, β, and δ. The final locations of the current individual can be calculated by

$$X_1 = X_\alpha - A_1 \cdot D_\alpha, \quad X_2 = X_\beta - A_2 \cdot D_\beta, \quad X_3 = X_\delta - A_3 \cdot D_\delta \tag{16}$$

$$X_{(t+1)} = (X_1 + X_2 + X_3)/3 \tag{17}$$

DE, first proposed by Storn and Price (Storn and Price 1997), is a stochastic algorithm solving global optimization. It starts with a random population generation, whose next-generation population is produced based on mutation, crossover, and selection operations. DE adopts a typical novel strategy to produce the variation of individuals: firstly, randomly select three different individuals; then zoom the difference vector of two individuals; finally, synthesize the zoomed vector with the third individual to achieve the mutation operation and remain the diversity of the population as follows:

$$V^{i}{}_{(g)} = X^{r_1}{}_{(g)} + F \cdot \left(X^{r_2}{}_{(g)} - X^{r_3}{}_{(g)}\right) \tag{18}$$

$$U^{k}_{j\,(g)} = \begin{cases} V^{k}_{j\,(g)}, & \text{if } \operatorname{rand}(0,1) \le CR \parallel j = j_{\text{rand}} \\ X^{k}_{j\,(g)}, & \text{otherwise} \end{cases} \tag{19}$$

where $r_1 \neq r_2 \neq r_3 \neq i$; g is the number of iteration; F is the scaling factor;CR represents the crossover probability; j_{rand} denotes a random integer between 1; d is the number of the dimension of the solutions (individuals). In DE, the greedy strategy is adopted to select individuals for the next generation using Equation (20).

$$X^{k}(g+1) = \begin{cases} U^{k}(g), & \text{if } f\left(U^{k}(g)\right) \le f\left(X^{k}(g)\right) \\ X^{k}(g), & \text{otherwise} \end{cases} \tag{20}$$

2.3.4. Prediction intervals (PIs) and evaluation indices

PIs are common tools used to quantify the uncertainty related to prediction models (Mazloumi et al. 2011). Before we elaborate on PIs, it is worth distinguishing confidence intervals (CIs) from PIs because they are not the same parameter but unfortunately are often confused. A CI is an estimate of an interval computed from the statistics of the observed data, e.g. population mean. CIs consider the distribution $P_r(f(x)|\hat{f}(x))$ in Equation (3), and hence only require estimation of σ_f^2. A PI is an estimate of an interval in which a future observation will fall, with a certain probability, given what has already been observed. PIs consider $P_r(F|\hat{f}(x))$ in Equation (3) and must also consider σ_ε^2 (Dybowski and Roberts 2011; Khosravi et al. 2011a). Therefore, PIs are necessarily wider than CIs.

For a landslide process model in ensemble form, given a set of pairs $\{(x_i, F_i)\}_{i=1}^{N}$, where x_i represents model inputs related to influenced factors of the landslide displacement, F_i denotes the outputs associated with displacement prediction. A PI construction based on the bootstrap, with a given confidence level of $(1 - \alpha) \times 100\%$, can be expressed as follows:

$$\left[\hat{F}_i - Z_{1-\alpha}\sqrt{\sigma_{\hat{F}_i}{}^2}, \hat{F}_i + Z_{1-\alpha}\sqrt{\sigma_{\hat{F}_i}{}^2}\right] = \left[\hat{L}\left(\hat{F}_i\right), \hat{U}\left(\hat{F}_i\right)\right] \tag{21}$$

where $\hat{L}(\hat{F}_i), \hat{U}(\hat{F}_i)$ are the lower and upper bounds of the ith PI, respectively, and α is the quartile of the standard normal distribution.

In this study, based on the bootstrap framework, we use an HWGO-optimized ELM model to construct the PIs of the landslide displacements. Figure 1 gives a schematic of the model used in the bootstrap method to generate PIs. PIs not only provide intervals where targets are highly likely to occur but also indicate the coverage probabilities. From the perspective of prediction, the constructed PIs should be as narrow as possible, whilst capturing a specified portion of data. Here, two performance indices, the PI coverage probabilities (PICPs) (Dybowski and Roberts 2011; Khosravi et al. 2011a) and the PI normalized root-mean-square width (PINRWs) (Lian et al. 2016) are used to assess the performance of a PI. These indices can be expressed by the following:

$$\mathrm{PICP} = \frac{1}{N}\sum_{i=1}^{N} c_i,\ \ c_i = \begin{cases} 1, & F_i \in \left[L(x_i), U(x_i)\right] \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

$$\mathrm{PINRW} = \frac{1}{R}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{U}(F_i) - \hat{L}(F_i)\right)^2} \tag{23}$$

where N is the number of samples, and R equals to the maximum minus minimum of the target value.

By definition, PICP is a value that ranges from 0 to 1. However, a perfect PICP (100% confidence level) with an extremely wide PINRWs is meaningless for decision-makers; thus, large values of PICP with small values of PINRWs indicate high-quality PIs. Given a confidence level of PICP, the objective was to find out the narrowest PINRWs. Therefore, a combined index that can balance the PICP and PINRW is required to provide a comprehensive assessment of PIs. Here, we use a criterion named coverage width-based criterion (CWC) based on a Gaussian function as the comprehensive index. CWC is given by (Lian et al. 2016),

$$\mathrm{CWC} = (\mathrm{PINRW} + \psi)e^{\frac{\kappa(\mathrm{PICP}-\mu)^2}{2\delta^2}} \tag{24}$$

where ψ denotes a small positive value range from 0.1% to 0.5%, and μ and δ are two hyperparameters of which the values should be set before the learning process. Generally, μ is set to $1-\alpha$ and δ to a small positive value less than 1. For the testing, κ is given by

$$\kappa = \begin{cases} 0, & \mathrm{PICP}, \mu \\ 1, & \mathrm{PICP}, \mu \end{cases} \tag{25}$$

CWC aims to find a trade-off between the informativeness (PINRW) and validity (PICP) of PIs. According to Equation (25), the smaller the value of the CWC, the higher the quality of PIs.

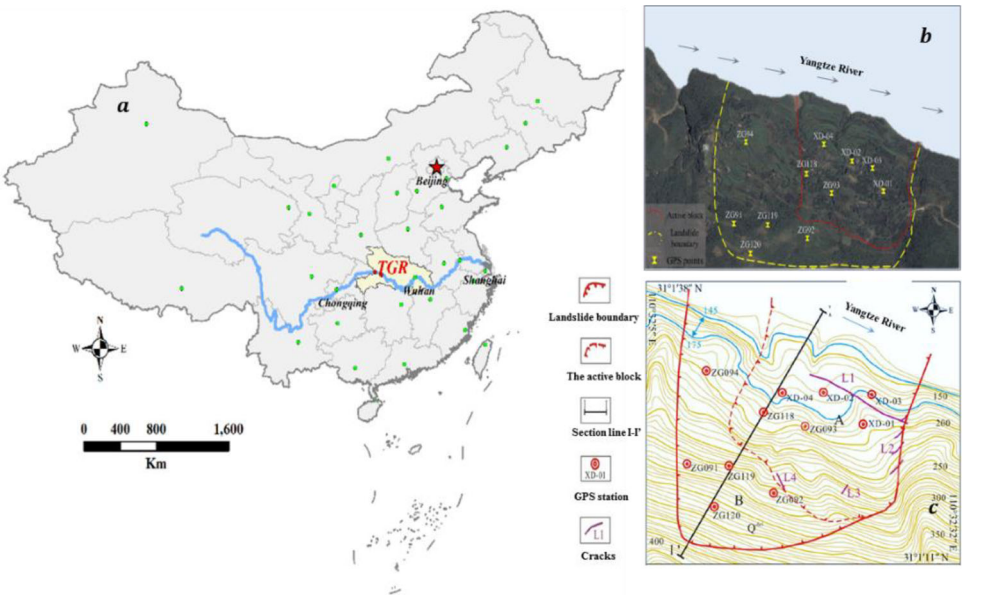

Figure 3. (a) Location of the landslide, (b) an overall view of the landslide, (c) topographic map of the landslide.

3. Application to Baishuihe landslide

3.1. Study area and monitoring data set

The Baishuihe landslide sits in the town of Zigui, Hubei province. It is located on the south side of the Yangtze River and spread into the Yangtze River. The landslide is about 56 km away from the Three Gorges Dam, China (Figure 3(a)). As a typical recurrence reservoir landslide, the Baishuihe landslide is fan-shaped (Figure 3(b-c)), with the main sliding direction of 20° NE. The maximal dimensions in north-south and east-west are 780 and 700 m, respectively. The volume of the landslide is about 1260×10^4 m^3 with an average slide thickness of about 30 m (Li et al. 2019).

The materials of this landslide are mainly quaternary deposits, including silty clay and fragmented rubble with a loose and disorderly structure. The underlying bedrock is composed of Jurassic Xiangxi Formation, overlain by Quaternary deposits which contain silty mudstone and sand muddy siltstones (Li et al. 2010). As shown in Figure 4, the elevation of the current slide ranges from 75 to 390 m, with a relatively flat central part, while larger gradients in the upper and lower parts of the landslide. As suggested by the field investigation and monitoring data, the landslide can be categorized into a relatively stable block (block-B), and an active back (block-A) (Figures 3 and 4).

As an activated landslide, the Baishuihe landslide was first triggered by the first reservoir impoundment in June 2003. Since then, several cracks have been found on the landslide (Figure 3(c)). To monitor its movement, 11 professional GPS stations and three borehole inclinometers were deployed on-site (Figures 3 and 4) with synchronizing measurements once a month. Besides, the water level of the reservoir has been collected by on-site water level gauge, and the precipitation has been obtained

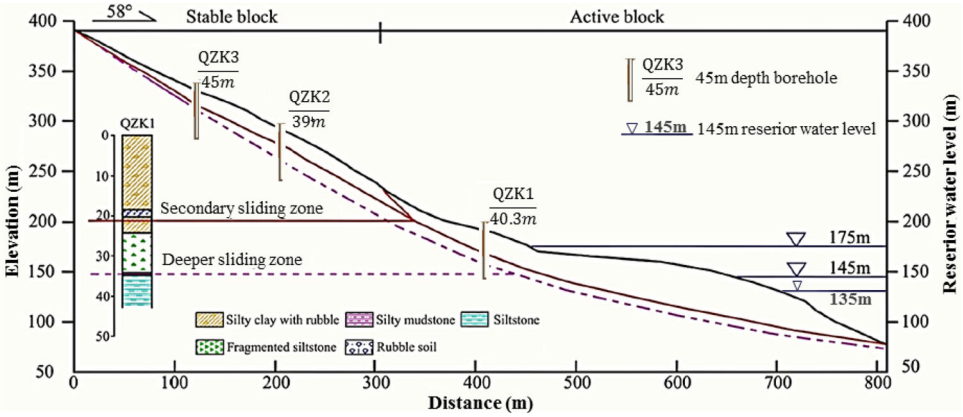

Figure 4. Geological profile (I-I′ in Figure 3(c)) of the Baishuihe landslide.

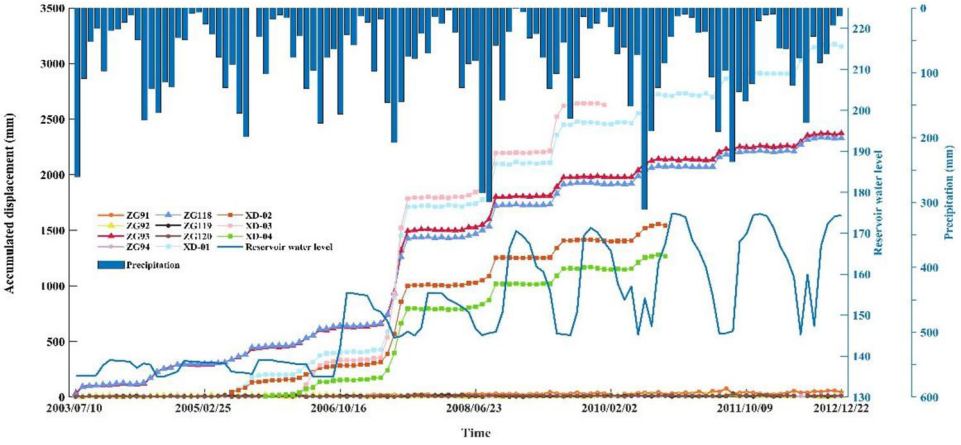

Figure 5. Monitoring data in time series.

from a monitoring site installed at 9.5 km away from the landslide. The monitoring data are presented in Figure 5.

As shown in Figure 5, the displacement of the Baishuihe landslide exhibited an obvious monotone feature over time as well as seasonal patterns, increasing from April to September each year and remaining stable from October to April in the next following year. Thus, the period of the seasonal characteristic is approximately a year, which is a joint effort of the heavy seasonal precipitation and the fluctuation of reservoir water level. It can also be inferred from Figure 5 that these combined effects of the triggering factors on the landslide displacements show significant time lags. Among the displacement monitoring stations, ZG93 and ZG118 deployed on the active Block-A of the slope, have complete records to represent the dynamic behaviour of the landslide, and are selected for the following analysis in this study.

3.2. Time series analysis

The raw data set is first processed by outlier removal and missing value imputation. Then, a wavelet-based de-noising is adapted to remove random noise from time

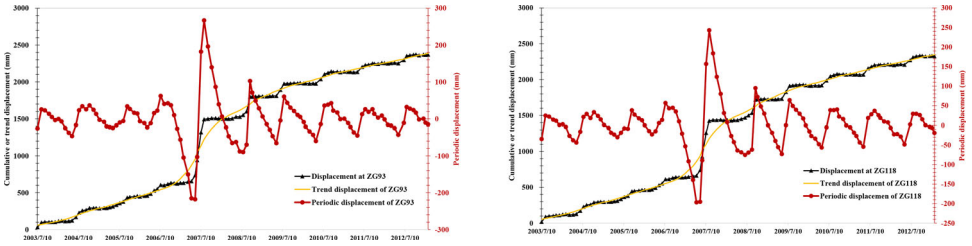

Figure 6. Decomposition of time series displacements at ZG193 (left) and (b) ZG118 (right) stations.

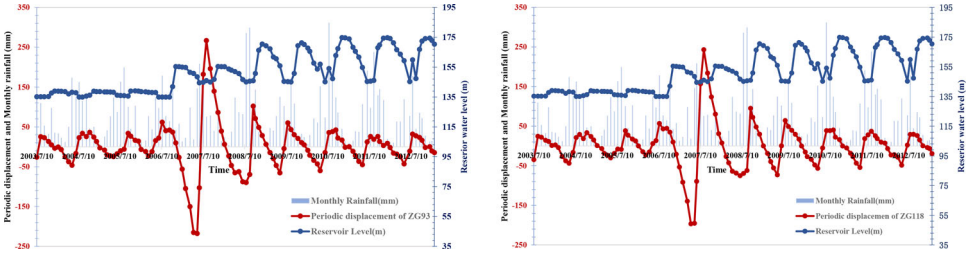

Figure 7. Seasonal pattern of the periodic displacements, the precipitation, and the reservoir water levels.

series displacements. As shown in Figure 5, the landslide displacement pattern comprises a high-frequency seasonal displacement dominated by external triggering factors and a low-frequency monotone displacement affected by geological conditions. Thus, a decomposition of the time series displacements is conducted to establish a response relationship separately, and the result is shown in Figure 6.

As mentioned, the reservoir water level and precipitation have a seasonal trend that generates the landslide with a seasonal displacement pattern. Here, the variation of the precipitation, the reservoir water levels, and the periodic displacements, illustrated in Figure 7, are used to analyse the potential response relations. As shown in Figure 7, significant seasonality patterns of the two major triggering factors in a concordance with the periodic displacements, and the obvious time lags between the factors and the instant displacement are observable.

To evaluate the complex nonlinear relations, the GRA is first used to calculate the significance of two different data sequences. Factors with a grey relational degree (GRD) value greater than 0.6 are believed to have a significant impact on the instant displacements. Consequently, six indicators are selected as the initial influencing factors of the landslide displacement, including monthly average reservoir water level, monthly cumulative rainfall, the previous two-month cumulative rainfall, the amount of reservoir water level fluctuation per month, the amount of reservoir water level fluctuation per two months and the reservoir water level rate per month.

After correlation assessment using GRA, the Gaussian KPCA is adopted to select efficient features (without redundancy) for landslide displacement modelling and forecasting through analysing the six influencing factors and the time series displacements. KPCA performed in a reproducing kernel Hilbert space involves multivariate analyses to understand complicated nonlinear relationships between variables and

Table 1. Variables selection by GRA-KPCA as model inputs.

	ZG118 (c = 800)			ZG93(c = 800)		
Principal components	Eigenvalue	Variance contribution %	Total variance contribution %	Eigenvalue	Variance contribution %	Total variance contribution %
1	0.772	45.425	45.425	0.767	45.104	45.104
2	0.518	30.482	75.907	0.518	30.448	75.552
3	0.246	14.486	90.393	0.251	14.788	90.340

their relevance to the problem being studied. Principal components of KPCA, converting from a set of possibly correlated variables to uncorrelated orthogonal information, can realize the dimensionality reduction effectively. Generally, the first few components contain much of the signal with a high signal-to-noise ratio, while the later components are dominated mainly by noise and can be directly disposed of without great loss. As shown in Table 1, the first three principal components concentrate ∼90% of the total variance contribution, with the first component the highest contribution (∼45%), following by ∼30% of the second component. Therefore, these three components are used as input-influencing factors for model training and forecasting.

3.3. Results and comparative analysis

3.3.1. Results

As mentioned, ZG93 and ZG118 on the active Block-A have complete records to represent the dynamic behaviour of the landslide and are selected to establish and evaluate the prediction model. Displacements acquired monthly from 2003 to 2013 and the associated daily reservoir water level and precipitation are used in the study. The first 100 groups of the data set (acquired from July 2003 to November 2011) are used as the training data set and the last 16 groups as the testing data set to evaluate the performance of our proposed method. During the prediction of landslide trend displacement, the parameters ζ and ξ of DES are set to 0.99 and 0.98, respectively, to obtain smooth results. As illustrated in Figure 8, the predicted trend displacement is close to those on-site measurements, with the absolute error of less than 3 mm, suggesting that the DES is a promising trend forecasting model.

Before modelling the periodic component of the landslide motion, the principal triggering components and the periodic displacements are normalized into the range of [0, 1] to eliminate dimensional effects on ML models and improve the reliability of the forecasts. It will be renormalized back after HGWO-ELM training processing. The bootstrap replicate number is set to 20, and the hidden layer neurons of ELM are set to 12. To further mitigate the effects of the model parameter uncertainty, an ensemble-based ELM is trained with different parameter initializations, and the HGWO will iterate 100 times to get the optimum parameters for each ELM. Then, the regression mean and the variance of the nth bootstrapped HGWO-ELM can be estimated as indicators of the model uncertainty, and the PIs can be constructed given a certain confidence level of $(1-\alpha)\times 100\%$ using Equation (21). The predicted

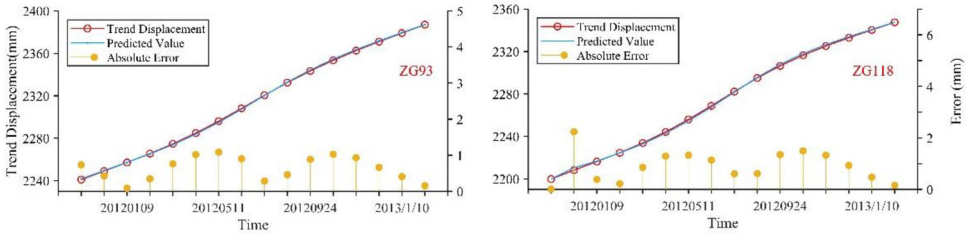

Figure 8. Forecasting results of the trend displacement of the landslide.

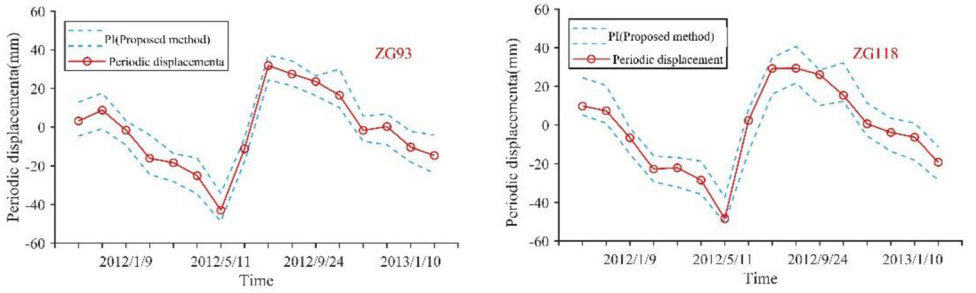

Figure 9. Forecasting results of the periodic displacement of the landslide with 95% confidence level.

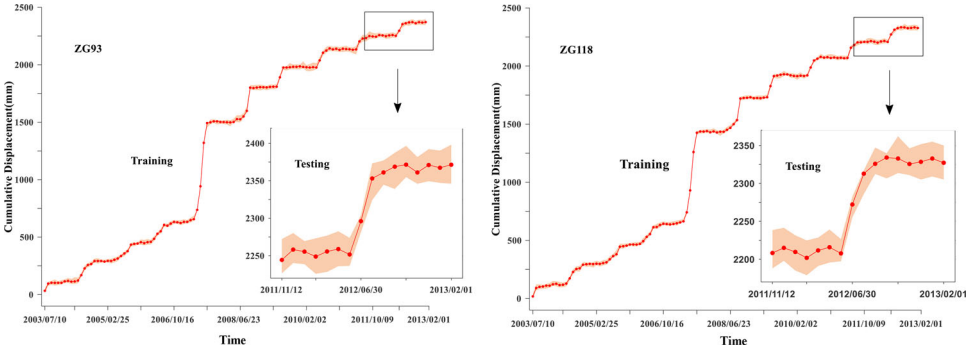

Figure 10. Forecasting results with PIs at a nominal confidence level of 95%.

periodic displacements with PIs at a nominal confidence level of 95% are shown in Figure 9.

The predicted cumulative displacements with PIs can be derived by combining the two predicted components, and the constructed PIs with a nominal confidence level of 95% are illustrated in Figure 10. As can be seen, the on-site monitoring displacements used as either training or testing are distributed in the constructed PIs with a 95% confidence level. Even in the periods of faster seasonal variations, the results also display satisfactory coverage probability. Thus, the value of PICP obtained using the proposed method is 100%, demonstrating the predictive ability of our proposed method in landslide displacement. The value of PINRWs is 0.2116 and 0.2266 for the

Table 2. Prediction performance of the studied methods.

		Evaluation indices		
Monitoring stations	Methods	PICP	PINRW	CWC
ZG118	DES-HGWO-ELM	1.0000	0.2266	0.2276
	DES-GWO-ELM	1.0000	0.4702	0.4712
	DES-PSO-ELM	1.0000	0.7161	0.7171
ZG93	DES-HGWO-ELM	1.0000	0.2116	0.2126
	DES-GWO-ELM	1.0000	0.4665	0.4675
	DES-PSO-ELM	1.0000	0.7345	0.7355

two on-site stations, respectively. In general, high-quality PIs always have a higher value of PICP and a smaller value of PINRW. Thus, as a trade-off between the two, CWC is used; the smaller the value of the CWC, the higher the quality of PIs. In this study, the results of CWC are 0.2126 and 0.2276, respectively, indicating that the proposed model exhibits satisfactory performance in the interval prediction of landslide displacement. Furthermore, the proposed approach has accounted for the existing uncertainties during the model training and forecasting process and thus could generate reliable PIs.

3.3.2. Comparative analysis

To illustrate the superiority of the proposed method over existing state-of-art methods such as hybrid DES and PSO-ELM, hybrid DES, and GWO-ELM, we compare the results of these models based on the same framework of bootstrap. Parameters and other inputs associated with Bootstrap and DES are uniform to facilitate comparative analysis, e.g. the number of bootstrap replicates. Besides, the number of hidden layer nodes and the initial parameters of the ELM model are also consistent with those in the proposed method. During the displacement prediction, the cumulative displacements are first denoised and decomposed; then, the trend term is predicted via the DES, while the periodical displacement is predicted by the HGWO-ELM (our method), the PSO-ELM, and the GWO-ELM algorithms, separately. During this process, three different optimization algorithms, the HGWO, the PSO, and the GWO, are used to optimize the weight matrix connecting the input and hidden layers of ELM. Thus, different weight matrices generated by three different optimization algorithms are used during ELM training and predicting.

Evaluation indices defined in Section 2.3.4 are used to quantify the performance of the proposed method and the comparative methods. Table 2 gives the indices of all the studied methods. As can be seen, the PIs generated by the proposed method and comparison methods display satisfactory coverage probabilities. All the training and testing data lie in the constructed PIs with 95% confidence level. In fact, in the landslide prediction, it is better to cover all the training data with a high confidence level (e.g. 95%) PIs (means high PICP). If not, it is hard to believe that the forecasting model is well-trained, because it is a minor probability event with 5% probability of the data that would fall outside of the PIs. Besides, if there are several data in sequence falling outside of the PIs, model performance would be even worse, because it is too hard to tell whether they are caused by the prediction error or that they indicate a new state of motion of the landslide.

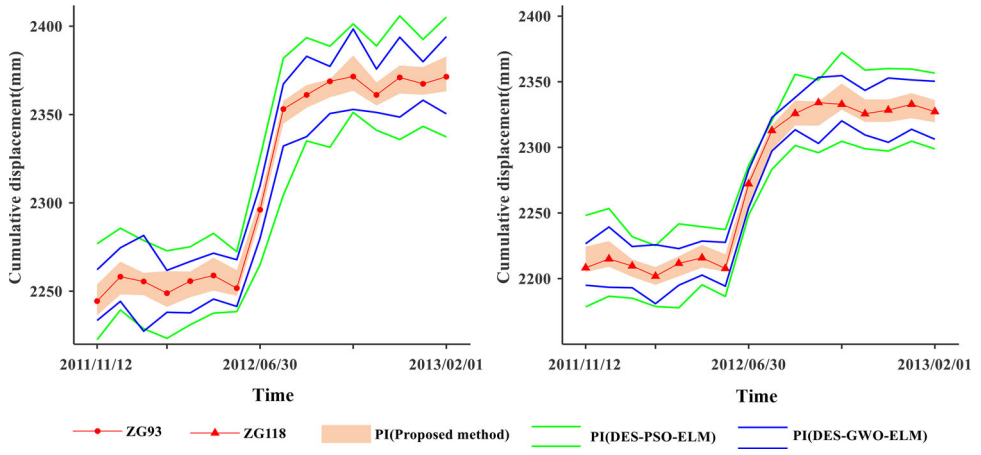

Figure 11. Comparative results with PIs at a nominal confidence level of 95%.

Although the CWC is a comprehensive index that encompasses the PICP and the PINRW, given a certain value of PCIP, the CWC values in Table 2 are mainly controlled by the PINRW. The smaller the value of the PINRW, the smaller CWC, and the higher the quality of the PIs. Results in (Wang et al. 2019) showed that if the DES model was not used in landslide trend displacement prediction, the value of the PINRWs would be several times larger, illustrating the advantage of the DES. More importantly, the reason for using DES in this study is that it was designed to process time series exhibiting a certain trend and therefore could model the trend displacement of a landslide better than other methods (e.g. polynomial fitting and moving average).

As shown in Table 2, at the monitoring station of ZG118, the values of the CWC of the DES-HGWO-ELM, DES-PSO-ELM, and DES-GWO-ELM methods are 0.2276, 0.4712, and 0.7171, and the corresponding values for ZG93 are 0.2126, 0.4675, and 0.7355, respectively. These indicators suggest that the performance of the proposed DES-HGWO-ELM is better than the other two methods, and the DES-PSO-ELM is the worst among the three. That is, the HGWO gives the best parameter optimization for the ELM compared with the GWO and PSO, ensuring that our proposed method exhibits satisfactory predictive ability in the interval prediction of landslide displacement. The comparative results are shown in Figure 11.

From Figure 11, we can see that the widths of the PIs vary with time, which indicates that the uncertainty of the displacement prediction varies with time and that the bootstrap-based methods could quantify these uncertainties. The PIs generated by the proposed method are the narrowest, followed by the DES-GWO-ELM, and the DES-PSO-ELM ranks the last. These results demonstrate the superiority of our proposed method for interval prediction of landslide displacement.

4. Discussion

Predicting landslide displacements lays the foundation for the prevention of landslide hazards. The behaviour of a landslide can be characterized by its transient displacements that can be modelled mathematically. ML-related methods trained by historical monitoring data have gained increasing popularity in landslide displacement

prediction. However, most of the published works provide only the deterministic point forecasting result and therefore could not accurately represent the dynamic uncertainty of landslide displacement.

We propose a novel hybrid machine-learning model to predict the landslide displacement and quantify the uncertainties in terms of PIs. Through considering the uncertainty, our approach enables users to make a more informed and appropriate decision. Besides, the variation of the PIs width is also important in decision making, because a wide PI represents a high level of uncertainty of the forecasts. Although PIs can be used as a complementary source of information along with point predictions, the calculation of PIs is computationally intensive when dealing with large data sets.

The failure of a slope is often characterized by acceleartion in deformation rate. Hence, the ML models, such as the HGWO-ELM powered by strong learning and generalization capacity, can be practically useful in assisting the prevention and mitigation of landslide hazards. For example, the predicted displacements help set warning thresholds and detect anomalous displacements, which might be a sign of an unprecedented acceleration of a typical landslide and usually trigger the early warning procedures. If the detected anomalous displacements occur at the acceleration stage, landslide models built on creep theory, such as Saito model (Saito 1965), could be used to predict a landslide.

Besides, displacement monitoring and prediction based on *in situ* measurements is still an important field worthy of further study in landslide research. However, the cost of the ground monitoring devices in terms of time and manpower may limit the application of this method. The developing technologies of the earth observation from the space open a new era for landslide displacement monitoring, e.g., the satellite radar interferometry with monitoring accuracy of a millimeter can also provide displacement measurements coving a wide range every 6 days using open-access Sentinel-1 data set (Hu et al. 2017), and will someday serve as the fundamental data set to fulfil landslide prediction.

5. Conclusions

Accurate prediction of landslide displacement is a vital part of the early warning of landslide hazard and directly provides technical support for decision-makers in emergency responses. However, prediction errors are often inevitable considering the dynamic uncertainties of landslide evolution and can be very significant in the deterministic point forecasting methods. In this paper, a hybrid machine-learning model has been established to predict the landslide displacements and quantify their uncertainties using prediction intervals (PIs). Through considering the displacement forecasting uncertainty, this approach enables users to make more informed and appropriate decisions in disaster management.

The displacements data set and the associated reservoir water level and precipitation over the Baishuihe landslide in TGR, China, have been utilized in this study. After outlier removal and missing value imputation, the original landslide displacement time series is decomposed into the periodic, trend, and noise components by wavelet de-noising and Hodrick-Prescott (HP) filters. The trend displacement that exhibits an obvious monotone feature over time is predicted by the DES, while the

periodic displacement induced by the seasonal triggering factors is predicted by the ELM. Then, the predicted cumulative displacement can be obtained.

During modelling, bootstrap sampling and a HGWO are also applied to ensure the optimal performance of the ELM. Besides, the input variables for model training also have a big impact on prediction accuracy. Thus, GRA and KPCA are also combined to generate efficient triggering features for the ELM. Results validate the predictive modelling capacity of HGWO-ELM in forecasting landslide seasonal displacement.

To further illustrate the effectiveness and reliability of the proposed method, comparative analysis of other state-of-art methods including hybrid DES and PSO-ELM, hybrid DES, and GWO-ELM has also been conducted. Results show that our proposed method can guarantee a perfect PICP with the narrowest PIs (PINRWs) and the minimum CWC value. Thus, the proposed model has the capacity of effectiveness and can give a more satisfactory performance than other state-of-art methods in the interval prediction of the landslide displacement.

Acknowledgments

The authors would like to thank the National Service Center for Specialty Environmental Observation Stations for providing the *in situ* monitoring data set. The first author would like to thank the China Scholarship Council for funding her research at the Southern Methodist University, USA. Finally, our thanks go to the anonymous reviewers of this paper for their constructive comments and suggestions.

Data availability statement

The raw data used in this study can be downloaded for scientific research after approval at http://www.crensed.ac.cn/.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was financially supported by the National Key Research and Development Program of China [2017YFC1501004], the Key Research and Development Program of Sichuan Province [2020YFS0353, 2018SZ0339, and 2019YFS0074], the Key Program of the Education Department of Sichuan (18ZA0054).

ORCID

Zhong Lu http://orcid.org/0000-0001-9181-1818

References

Cao Y, Yin K, Alexander DE, Zhou C. 2016. Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors. Landslides. 13(4): 725–736.

Carla T, Intrieri E, Di Traglia F, Casagli N. 2016. A statistical-based approach for determining the intensity of unrest phases at Stromboli volcano (Southern Italy) using one-step-ahead forecasts of displacement time series. Nat Hazards. 84(1):669–683.

Chryssolouris G, Lee M, Ramsey A. 1996. Confidence interval prediction for neural network models. IEEE Trans Neural Netw. 7(1):229–232.

Deng JL. 1982. Control problems of grey systems. Syst Control Lett. 1(5):288–294. http://www.sciencedirect.com/science/article/pii/S016769118280025X.

Dybowski R, Roberts SJ. 2011. Confidence intervals and prediction intervals for feed-forward neural networks. In: Dybowski R, Gant V, editors. Clinical applications of artificial neural networks. Cambridge, U.K.:Cambridge University Press, 1–27.

Faris H, Aljarah I, Al-Betar MA, Mirjalili S. 2018. Grey wolf optimizer: a review of recent variants and applications. Neural Comput Appl. 30(2):413–435.

Gao Z-M, Zhao J. 2019. An improved grey wolf optimization algorithm with variable weights. Comput Intell Neurosci. 2019:2981282.

Guo Z, Chen L, Gui L, Du J, Yin K, Do HM. 2019. Landslide displacement prediction based on variational mode decomposition and WA-GWO-BP model. Landslides. 1–17.

Holt CC. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. Int J Forecasting. 20(1):5–10.

Hu X, Oommen T, Lu Z, Wang T, Kim J-W. 2017. Consolidation settlement of Salt Lake County tailings impoundment revealed by time-series InSAR observations from multiple radar satellites. Remote Sens Environ. 202:199–209..

Huang G-B, Zhu Q-Y, Siew C-K. 2006. Extreme learning machine: theory and applications. Neurocomputing. 70(1–3):489–501.

Huang RQ. 2009. Some catastrophic landslides since the twentieth century in the southwest of China. Landslides. 6(1):69–81.

Jibson RW. 2007. Regression models for estimating coseismic landslide displacement. Eng Geol. 91(2–4):209–218.

Khosravi A, Nahavandi S, Creighton D, Atiya AF. 2011a. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Trans Neural Netw. 22(9):1341–1356.

Khosravi A, Nahavandi S, Creighton D, Atiya AF. 2011b. Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE Trans Neural Netw. 22(3):337–346.

Khosravi A, Nahavandi S, Creighton D, Srinivasan D. 2011. Optimizing the quality of bootstrap-based prediction intervals. The 2011 International Joint Conference on Neural Networks; 31 July–5 August 2011

Li D, Yin K, Leo C. 2010. Analysis of Baishuihe landslide influenced by the effects of reservoir water and rainfall. Environ Earth Sci. 60(4):677–687..

Li H, Xu Q, He Y, Deng J. 2018. Prediction of landslide displacement with an ensemble-based extreme learning machine and copula models. Landslides. 15(10):2047–2059.

Li H, Xu Q, He Y, Fan X, Li S. 2019. Modeling and predicting reservoir landslide displacement with deep belief network and EWMA control charts: a case study in Three Gorges Reservoir. Landslides. 1–15.

Li S, Wu L, Chen J, Huang R. 2019. Multiple data-driven approach for predicting landslide deformation. Landslides. 1–10.

Lian C, Zeng Z, Yao W, Tang H. 2015. Multiple neural networks switched prediction for landslide displacement. Eng Geol. 186:91–99.

Lian C, Zeng Z, Yao W, Tang H, Chen CLP. 2016. Landslide displacement prediction with uncertainty based on neural networks with random hidden weights. IEEE Trans Neural Netw Learn Syst. 27(12):2683–2695.

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. 15(6):1145–1153..

Mazloumi E, Rose G, Currie G, Moridpour S. 2011. Prediction intervals to account for uncertainties in neural network predictions: methodology and application in bus travel time prediction. Eng Appl Artif Intell. 24(3):534–542.
Miao F, Wu Y, Xie Y, Li Y. 2018. Prediction of landslide displacement with step-like behavior based on multialgorithm optimization and a support vector regression model. Landslides. 15(3):475–488.
Miche Y, Akusok A, Veganzones D, Björk K-M, Séverin E, Du Jardin P, Termenon M, Lendasse A. 2015. SOM-ELM—self-organized clustering using ELM. Neurocomputing. 165: 238–254.
Mirjalili S. 2015. The ant lion optimizer. Adv Eng Softw. 83:80–98.
Mirjalili S, Mirjalili SM, Lewis A. 2014. Grey wolf optimizer. Adv Eng Softw. 69:46–61.
Nix DA, Weigend AS. 1994. Estimating the mean and variance of the target probability distribution. Proceedings of 1994 ieee international conference on neural networks (ICNN'94).
Pearce T, Zaki M, Brintrup A, Neely A. 2018. High-quality prediction intervals for deep learning: a distribution-free, ensembled approach. arXiv Preprint arXiv:1802.07167.
Pradhan B. 2013. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput Geosci. 51:350–365.
Ravn MO, Uhlig H. 2002. On adjusting the Hodrick-Prescott filter for the frequency of observations. Rev Econ Stat. 84(2):371–376.
Saito M. 1965. Forecasting the time of occurrence of a slope failure. Proceedings of 6th International Congress of Soil Mechanics and Foundation Engineering, Montreal.
Storn R, Price K. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim. 11(4):341–359.
Teng Z-j, Lv J-l, Guo L-w. 2019. An improved hybrid grey wolf optimization algorithm. Soft Comput. 23(15):6617–6631.
Tianbin L, Mingdong C. 1996. Time prediction of landslide using Verhulst inverse-function model. J Geol Hazards Enveronment Preserv. (3)
Tramèr F, Kurakin A, Papernot N, Goodfellow I, Boneh D, McDaniel P. 2017. Ensemble adversarial training: attacks and defenses. arXiv Preprint arXiv:1705.07204.
Wan C, Xu Z, Wang Y, Dong ZY, Wong KP. 2014. A hybrid approach for probabilistic forecasting of electricity price. IEEE Trans Smart Grid. 5(1):463–470.
Wang J-S, Li S-X. 2019. An improved grey wolf optimizer based on differential evolution and elimination mechanism. Sci Rep. 9(1):7181..
Wang Y, Tang H, Wen T, Ma J. 2019. A hybrid intelligent approach for constructing landslide displacement prediction intervals. Appl Soft Comput. 81:105506.
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(3):339–349.
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. Eng Geol. 204:108–120.
Zhou C, Yin K, Cao Y, Intrieri E, Ahmed B, Catani F. 2018. Displacement prediction of step-like landslide by applying a novel kernel extreme learning machine method. Landslides. 15(11):2211–2225.
Zhu A, Xu C, Li Z, Wu J, Liu Z. 2015. Hybridizing grey wolf optimization with differential evolution for global optimization and test scheduling for 3D stacked SoC. J Syst Eng Electron. 26(2):317–328.
Zhu X, Xu Q, Tang M, Li H, Liu F. 2018. A hybrid machine learning and computing model for forecasting displacement of multifactor-induced landslides. Neural Comput Appl. 30(12): 3825–3835.