Adaptive hybrid machine learning model for forecasting the step-like displacement of reservoir colluvial landslides: a case study in the three Gorges reservoir area, China

Li Linwei1 · Wu Yiping2 · Huang Yepiao3 · Li Bo1 · Miao Fasheng2 · Deng Ziqiang1

Abstract
Establishing an accurate and dependable displacement prediction model is essential to building an early warning system for landslide hazards. This study proposes an adaptive hybrid machine learning model for forecasting the step-like displacements of reservoir colluvial landslides. Firstly, candidate factors are determined based on the landslide deformation response. Then, the cumulative displacement and candidate factors are decomposed using the optimized variational mode decomposition algorithm. Second, the sensitivity analysis of the gray wolf optimizer-based kernel extreme learning machine (GWO-KELM) models to each factor component is analyzed using the PAWN method. Then, the factors are optimized based on the analysis results. Third, based on the optimized factors, GWO-KELM models of different displacement components are established and integrated to predict the cumulative displacement. The Baishuihe landslide was taken as an example. The raw data of its three monitoring sites were employed to verify the performance of the proposed model. The results indicate that the model can decompose the cumulative displacement and factors with the adaptively determined parameters. In addition, the model performed well over a three-year prediction of the landslide displacement.

Keywords Reservoir colluvial landslide · Step-like displacement · Displacement prediction · Variational mode decomposition · Input factor reduction · Kernel extreme learning machine

1 Introduction
As the most widely distributed geological hazard worldwide, landslides endanger human safety and property (Petley 2012; Haque et al. 2016; Moayedi et al. 2019). Since the Three Gorges Reservoir area (TGRA) began to impound in 2003, many ancient colluvial landslides have been reactivated (Zhang et al. 2015; Iqbal et al. 2018). It is imperative to develop accurate early warning systems (EWSs) for landslide disasters in this district (Yin et al. 2010). Thus, the displacement prediction models of landslides, as a critical element of EWSs (Casagli et al. 2010; Intrieri et al. 2013), are receiving increasing attention with the growing demand for disaster prevention and mitigation (Kirschbaum et al. 2010).

Landslide displacement decomposition is the crucial part affecting the performance of prediction models. The decomposition methods can be divided into partial decomposition (PD) and complete decomposition (CD) methods. The PD method combines the time series analysis (TSA) model and moving average technique to separate the trend and periodic displacements from the cumulative displacement (Du et al. 2013; Zhou et al. 2016). However, it cannot obtain the random displacement components in the TSA model, leading to the incomplete decomposition problem. In the CD method, the TSA and the complex signal decomposition technique, such as wavelet transform (WT), empirical...
mode decomposition (EMD), and variational mode decomposition (VMD), are integrated to decompose the cumulative displacement (Lian et al. 2013; Shihabudheen and Peethambaran 2017; Li et al. 2018). This method offers a solution to the above incomplete decomposition problem. However, due to the complexity of these complex signal decomposition techniques, decomposition parameter determination has become the main obstacle restricting the application of this method. For example, Li et al. (2018) and Guo et al. (2020) proposed a TSA-VMD approach with excellent performance to realize cumulative displacement decomposition. However, the settings of the VMD parameters in their study still depend on the trial and calculation, which is not conducive to the promotion of their method. Thus, considering the differences in the entropy characteristics, central frequency characteristics, and physical meanings between the displacement components after decomposition, we propose an optimized VMD (OVMD) method to solve the above problem by applying the VMD algorithm.

The development of prediction models has experienced a leap from empirical models to neural network models (Saito 1965; Fukuzono 1985; Li et al. 2009; Intrieri et al. 2019; Criss et al. 2020). As a neural network prediction model with strong generalization ability, the kernel extreme learning machine (KELM) has been widely utilized (Zhou et al. 2018). Like other nonlinear models, such as the backpropagation neural network (BPNN), the support vector regression (SVR), and the extreme learning machine (ELM), the selection of input factors is still a crucial part of affecting the prediction performance of models. In most cases (Zou et al. 2020; Li et al. 2021; Zhang et al. 2021; Ma et al. 2022), a univariate correlation analysis, such as the Pearson correlation analysis or the gray correlation analysis (GRA), is applied to conduct the correlation analysis between factors and displacements before building the prediction model. However, no factors are removed in the model construction process after the analysis. By comparison, a more reasonable quantitative method for factor selection is to propose candidate factors based on the landslide deformation response at first. Then, based on the sensitivity analysis of the initially trained model to candidate factors, the factors are optimized accordingly. Thus, a global sensitivity analysis (GSA) method called the PAWN method (Pianosi and Wagener 2015) is introduced in this study. The PAWN method can be easily used to characterize the significance of the input factors (Wang et al. 2020).

In summary, in this study, we propose an adaptive hybrid model to solve the parameter optimization problem of the VMD algorithm and the factor selection problem of the prediction model above. In the model, the OVMD method based on two criteria and one comprehensive index was first proposed and combined with the TSA model to decompose the cumulative displacement with adaptively determined decomposition parameters. Then, based on the landslide deformation response analysis, some candidate factors were proposed for each decomposed displacement component. Next, an integrated approach based on the Latin hypercube sampling (LHS) and PAWN algorithms was first proposed to optimize these factors according to the sensitivity analysis of KELM models to candidate factors. Finally, according to the optimized factors, the Grey Wolf Optimizer-based (Mirjalili et al. 2014) KELM models (namely, the GWO-KELM models) of each displacement component were built to conduct multistep-ahead prediction. Their prediction results are superposed to realize the prediction of the cumulative displacement of landslides.

To verify the performance of the proposed model, we selected the Baishuihe landslide as an example to predict the displacement of its three most representative monitoring sites ZG93, ZG118, and XD01. Fifty predictions for each monitoring site were implemented using this model. The root mean square error (RMSE), mean absolute percentage error (MAPE), R-squared ($R^2$), and Akaike information criterion (AIC) were used to compare the average result of these fifty operations with the results of four different GWO-KELM models.

2 Methodology

2.1 TSA model of cumulative displacement

The cumulative displacement of landslides is nonstationary and increases obviously with time. The factors that cause the variation in the cumulative displacement can be divided into trend, periodic, and random factors (Yang 1992). Correspondingly, the cumulative displacement can be characterized as follows:

$$X(t) = x(t) + \beta(t) + \gamma(t)$$

where $X(t)$ represents the observed value of landslide displacement, $x(t)$ represents the trend displacement, $\beta(t)$ represents the periodic displacement, and $\gamma(t)$ represents the random displacement.

Due to the limitations of the monitoring means, these kinds of displacement components in the cumulative displacement cannot be directly obtained by current monitoring instruments. Therefore, obtaining the displacement components from the original cumulative displacement has become a necessary foundation for accurately establishing the nonlinear mapping between factors and displacements. The various displacement components were obtained in this study using the OVMD technique.
2.2 OVMD

VMD is a complicated signal-decomposition method with excellent performance. The basic definitions and computational process of the VMD method can be found in the literature published by Dragomiretskiy and Zosso (2013). Before using the method, three parameters (namely, the number of recovered modes \( K \), the balance parameter of the data-fidelity constraint \( \zeta \), and the timestep of the dual ascent \( \tau \)) need to be specified appropriately. However, how to set the parameters \( K \) and \( \zeta \) is still extremely complicated, which is the critical link to effectively decomposing landslide displacement time series. As shown in Fig. 1, the OVMD method is proposed, which can adaptively determine \( K \) and \( \zeta \) to achieve automatic decomposition of displacement and factor time series. The specific calculation procedure of the method is as follows.

1. **Optimization of \( K \)** The interval of \( K \) (i.e., \( K_{\text{interval}} = \left[1, 6\right] \), where \( K = 1 \) means no decomposition of the input signal) is determined, and \( \zeta \) is temporarily set to 1. Then, the input signal is decomposed according to the initial parameter setting (i.e., \( K = 2 \), and \( \zeta = 1 \)). The optimal value \( K_{\text{optimal}} \) is found according to the following two criteria. If these two criteria are not met, iterative calculations are performed until these two criteria are reached:

   - **Criterion 1** is whether there is a medium or high correlation between different decomposed components (i.e., whether the distance correlation coefficient (DCC) between different components is greater than 0.4). This ensures that the decomposition components obtained by using the OVMD method have different physical meanings.

   - **Criterion 2** is whether there is a mode aliasing between decomposed components after decomposition (i.e., is the difference in the center frequencies (DCF) between the decomposed components higher than half of their sum of 3 dB bandwidths (HSB) as shown in Fig. 2). Criterion 2 is proposed to supplement Criterion 1 to prevent excessive decomposition of signals.

2. **Optimization of \( \zeta \)** After \( K_{\text{optimal}} \) is found, the optimization interval of \( \zeta \) is set by equal spaces between logarithms (i.e., \( \zeta_{\text{interval}} = \left[10^0, 10^4\right] \)). Then, the input signal is temporarily decomposed according to the initial parameter setting (i.e., \( K = K_{\text{optimal}} \) and \( \zeta = 1 \)). The comprehensive determination indexes (CDIs) corresponding to different \( \zeta \) are calculated according to the signal-to-noise ratio (SNR), approximate entropy, center frequency, and 3 dB bandwidth of the decomposed components. The definition of the CDI is as follows:

---

Fig. 1 Calculation flowchart of the OVMD method
where \( CDI \) is the comprehensive determination index; \( MAPEN \) is the mean value of the approximate entropies of decomposed components; \( NaN \) is a null variable; and \( t \) is the logical variable of the determination matrix of modal aliasing (DMMA), which is calculated based on the DCF and HSB between the decomposed components. If there is a mode aliasing between decomposed components, \( t = 0 \); otherwise, \( t = 1 \).

The smaller the CDI index is, the more regular the variation in each component is, the smaller the uncertainty according to the optimized parameters. Finally, the input time series is decomposed into \( n \) decomposed components. If there is a mode aliasing between decomposed components, \( t = 0 \); otherwise, \( t = 1 \).

### 2.3 GWO-KELM

The KELM model is an improved neural network model proposed by Huang et al. (2006) based on the ELM model and the kernel function. The regression function based on the KELM model can be expressed as follows:

\[
\begin{bmatrix}
K(x,x_1) \\
\vdots \\
K(x,x_N)
\end{bmatrix}^T (I/C + \Omega_{ELM})^{-1} T
\]

where \( I \) is the diagonal matrix, \( C \) is the penalty parameter, \( HC \) is the bias constant of the diagonal elements in the symmetric matrix \( \Omega_{ELM} \), \( \Omega_{ELM} \) is the kernel function matrix of ELM, and \( T \) is the output target vector–matrix \( (T = [t_1, t_2, \ldots, t_l]^T) \). The function matrix \( \Omega_{ELM} \) is defined as follows:

\[
\Omega_{ELM} = HH^T : \Omega_{ELM(I,D)} = h(x_i)h(x_j) = K(x_i, x_j)
\]

where \( H \) is the output matrix of the hidden layer (also known as the randomization matrix, and \( H = [h(x_1), h(x_2), \ldots, h(x_l)] \)), \( h(x_i) \) is the output vector of the hidden layer corresponding to the input vector \( x_i \), and \( K(x_i, x_j) \) is the kernel function.

For the model parameter optimization of the KELM model, the penalty parameter and the kernel function parameter have significant effects on its performance. In order to obtain more accurate and stable prediction results, we used the grey wolf optimizer algorithm (GWO), which is proposed by Mirjalili et al. (2014), to optimize the penalty parameter and the kernel function parameter in this study.

### 2.4 PAWN

The PAWN, as a distribution-based GSA methodology, was proposed by Pianosi and Wagener (2015) to calculate the sensitivity \( S_i \) of the input factor \( x_i \). More concretely, the sensitivity of the model output \( y \) to \( x_i \) can be calculated according to the maximum distance between the unconditional output distribution \( F_y(y) \) and the conditional output distribution \( F_{y|x_i}(y|x_i) \). \( F_y(y) \) can be obtained by varying all the input factors \( x \) simultaneously. \( F_{y|x_i}(y|x_i) \) is generated by changing all the input factors except \( x_i \). The calculation of the maximum distance between \( F_y(y) \) and \( F_{y|x_i}(y|x_i) \) is implemented by the Kolmogorov–Smirnov (KS) test. Thus, for \( x_i \), \( S_i \) is defined as follows:

\[
S_i = \text{statKS}(x_i) \text{ where } KS(x_i) = \max \left| F_y(y) - F_{y|x_i}(y|x_i) \right|
\]

where \( \text{stat} \) denotes a statistic determined by the user, such as the maximum, median, or mean. In this study, we set \( S_i = \text{mean}(KS(x_i)) \), in which the value range of \( S_i \) is \([0,1]\). The smaller \( S_i \) is, the smaller the influence of \( x_i \) on \( y \).

The KS statistic \( \hat{S}_{dummy} \) of the dummy factor is introduced into the PAWN method to identify the inputs whose measured sensitivities are too low to be distinguishable from approximation errors (Zadeh et al. 2017; Pianosi and Wagener 2018). The dummy factor is an input factor that is artificially generated in the calculation process of the PAWN and does affect the final analysis results. If the \( S_i \) of an input factor \( x_i \) is more significant than \( \hat{S}_{dummy} \), it suggests that this input factor \( x_i \) is indeed influential.
2.5 Prediction process and hyperparameter setting of the OVMD-GWO-KELM model

Figure 3 shows a schematic of the OVMD-GWO-KELM model for forecasting the cumulative displacement of landslides. This model consists of the following three parts.

2.5.1 Data processing and adaptive decomposition

1. **Data cleaning** It involves filling in the missing values of the raw data, reducing noise in the raw data, identifying outliers in the raw data, and correcting inconsistencies in the raw data.

2. **Isochronous processing** The raw data of the landslide displacements, reservoir water levels, precipitations, and so on must be preprocessed to unify the data sampling frequency at a fixed value (i.e., monthly).

3. **Normalization processing** All the raw data were normalized to $[-1, 1]$ by different dimensions of the factors.

4. **Decomposition of the input and output variables based on the OVMD method** The OVMD was applied to decompose the cumulative displacements and their factors adaptively. The detailed calculation flow of the OVMD method is described in Sect. 2.2.

5. **Dataset construction** The training datasets of the different displacement components were formed by combining the displacement component and its decomposed factors.

2.5.2 Factor adaptive selection

1. **Response surface modeling** Based on the constructed datasets, GWO-KELM models for different displacement components were established and trained as

![Schematic diagram of the adaptive hybrid machine learning model](image-url)
response surface models to represent the real response of the landslide. To fully use each training sample, we used the leave-one-out method ten times during the parameter optimization for each GWO-KELM model.

2. **Data resampling and response fitting** LHS was used to generate $N$ samples from the factors of each displacement component. All the input factors were assumed to follow uniform distributions. The generated samples are taken in the response surface model of each displacement component to obtain their corresponding predicted values.

3. **Calculation of the KS statistic based on the PAWN method** Based on the resampled factors and their corresponding displacements, the PAWN was applied to calculate the mean KS statistics of the factors. The KS statistic of the dummy factor was also obtained. Then, factors whose mean KS statistics were less than the mean KS statistics of their corresponding dummy factors were removed. The initial optimized factors were obtained based on the remaining factors.

4. **Reduction of redundant factors**

   1. The sorted mean KS statistics of the initial optimized factors were fit with an appropriate exponential fitting function as follows:

   \[
   y_{ks} = Ae^{Bx_{order}} + Ce^{Dx_{order}}
   \]

   where $y_{ks}$ is the fitted value of all factors’ mean KS statistics. $x_{order}$ is the order number of the factors sorted in descending order according to the mean KS statistics. $A$, $B$, $C$, and $D$ are the function parameters that need to be fitted.

   2. The second-order differences of $y_{ks}$ (i.e., $\Delta(\Delta y_{ks})$) are calculated.

   3. All the factors whose $\Delta(\Delta y_{ks})$ are less than 0.001 are eliminated.

   4. The retained factors were used as the final optimized factors to reconstruct the datasets of the different displacement components.

### 2.5.3 Model training, prediction, and validation

1. **Model training and prediction** GWO-KELM prediction models of the different displacement components were rebuilt based on the reconstructed datasets. Then, the optimal input factors corresponding to the period needed to be predicted were introduced into the well-trained models to predict the displacement components. The predicted cumulative displacement was obtained by summing all the prediction results of the displacement components.

2. **Performance validation of the model** In this study, four frequently used performance indexes (namely, $R^2$, MAPE, RMSE, and AIC) are used to validate the performance of models from the perspective of the average prediction performance, extreme value prediction performance, and model complexity.

---

**Table 1 Hyperparameter settings of the proposed model**

| Hyperparameter | Notes |
|----------------|-------|
| $K_{interval} = [1, 6]$ | $K_{interval}$ is the optimization interval for the number of recovered modes $K$ used in the OVMD method to decompose displacement and factor time series |
| $z_{interval} = [10^0, 10^4]$ | $z_{interval}$ is the optimization interval for the balance parameter of the data-fidelity constraint $z$ used in the OVMD method to decompose displacement and factor time series |
| $\tau = 0$ | $\tau$ is the timestep of the dual ascent used in the OVMD method to decompose displacement and factor time series |
| $N = 10,000$ | $N$ is the total number of calculations used in the LHS-PAWN model to optimize input factors |
| $W = 10$ | $W$ is the number of conditioning intervals used in the LHS-PAWN model to optimize input factors |
| method = ‘uniform sampling’ | method is the sampling method used in the LHS-PAWN model to optimize input factors |
| $N_{agent} = 40$ | $N_{agent}$ is the number of search agents used in the GWO algorithm to optimize the penalty factor $C$ and the kernel function parameter $\lambda$ of KELM models |
| $N_{iteration} = 200$ | $N_{iteration}$ is the maximum iteration used in the GWO algorithm to optimize the penalty factor $C$ and the kernel function parameter $\lambda$ of KELM models |
| $N_{convergence} = 10^{-7}$ | $N_{convergence}$ is the convergence error used in the GWO algorithm to optimize the penalty factor $C$ and the kernel function parameter $\lambda$ of KELM models |
| $F_{kernel} = ‘RBF kernel’$ | $F_{kernel}$ is the type of kernel function used in the KELM models |
| $I_C = [1 \times 10^{-13}, 1 \times 10^{13}]$ | $I_C$ is the optimization interval for the penalty factor $C$ of KELM models used in the GWO algorithm |
| $I_\lambda = [1 \times 10^{-13}, 1 \times 10^{13}]$ | $I_\lambda$ is the optimization interval for the kernel function parameter $\lambda$ of KELM models used in the GWO algorithm |
| $v = ‘tenfold cross validation’$ | $v$ is the cross-validation method used in the GWO algorithm to optimize the penalty factor $C$ and the kernel function parameter $\lambda$ of KELM models |
2.5.4 Hyperparameter setting

As with other machine learning models, the prediction performance of the proposed OVMD-GWO-KELM model in forecasting unknown data is controlled by hyperparameter settings. Table 1 shows the hyperparameter settings used in the proposed model.

Fig. 4  a Geographic location;  b topographical map;  c panoramic image

Fig. 5 Monthly cumulative displacements, daily precipitation, and daily reservoir water level
3 Materials

3.1 Baishuihe landslide

The Baishuihe landslide is located on the south bank of the Yangtze River (Fig. 4a). Figure 4b shows that the toe of the landslide is below the reservoir water level of 135 m, and its rear part is located at the geotechnical boundary with an elevation of 410 m. The main sliding direction, north–south length, and east–west width are 20°NE, 600 m, and 700 m, respectively. The front and back parts of the landslide are steeper than its middle part, with an overall slope of approximately 30°. The average thickness is 30 m, and its estimated volume is \(1.26 \times 10^7\) m³. The monitoring data of inclinometer QZK1 show two sliding zones (Li et al. 2010). The shallow sliding zone is the contact zone between Quaternary loose deposits and cataclastic rocks, with a depth of 12–25 m. The deep sliding zone is the contact zone between the bottom of the cataclastic rock and the underlying bedrock, with a depth of 18.9–34.1 m.

From the monitoring report proposed by the China National Cryosphere Desert Data Center (http://www.ncdc.ac.cn/), we can find that the monthly cumulative displacements of the landslide were obtained using GPS monitoring technology from 2003 to 2007. After the strong deformation occurred in 2007, the monthly displacements were monitored by the total electronic station. A strict manual audit controls the quality of the monitoring data in the report, and no additional preprocessing method is required before using it. According to both the location (Fig. 4b) and monitoring displacement change (Fig. 5) of each site, the early warning area of the landslide was determined (Fig. 4c).

3.2 Selection of the candidate factors

Precipitations, fluctuations in reservoir water level, and earthquakes are the main factors affecting slope deformation (Crosta 2004; Lee et al. 2008; Tang et al. 2019). No massive seismic disaster has been reported in the TGR area recently. Abundant precipitation occurs year-round in this area (Yao et al. 2015). Additionally, the Baishuihe landslide is a typical riverside landslide in this region. Thus, the precipitations and the fluctuations in the reservoir water level should be the significant factors of the landslide deformation.

Figure 5 shows that the sharp annual increases in displacements and heavy precipitation processes are concentrated in June, July, and August. There is a significant correlation between the occurrence of antecedent heavy precipitation processes before 1 or 2 months and the sharp increase in the monthly landslide displacement. Thus, it is appropriate to consider the cumulative effect of precipitation infiltration within two months on the landslide deformation. Some earlier research (Du et al. 2013; Krkač et al. 2017; Bogaard and Greco 2018) was consistent with this analysis. Therefore, the monthly cumulative antecedent...
precipitation, the bimonthly cumulative antecedent precipitation, and the monthly maximum continuous effective precipitation with four different reduction factors (i.e., \( r = 1, 0.8, 0.6, \) and \( 0.4 \)) were selected as the candidate factors. These six precipitation factors were defined as \( CF_1, CF_2, \ldots, CF_6 \), respectively.

Figure 5 also shows that the fluctuations in the reservoir water level have both positive and negative effects on the increase in landslide displacement because the pore water pressure is difficult to dissipate (Tang et al. 2019). Hence, considering the hysteresis effect of the groundwater level caused by the fluctuation of the reservoir water level (Miao et al. 2018; Xue et al. 2020), the influence of the current and past water levels on landslide displacements should be considered. Therefore, we selected the monthly and bimonthly average elevation of reservoir levels, the monthly and bimonthly increase in reservoir levels, the monthly and bimonthly decrease in reservoir levels, the monthly and bimonthly relative variation in reservoir levels, and the monthly and bimonthly absolute variation in reservoir levels as the candidate factors. These ten water-level factors were defined as \( CF_7, CF_8, \ldots, CF_{16} \).

There is a remarkable difference in the deformation responses of landslides between their different evolution stages (Glade and Crozier 2005). For landslides with progressive failure characteristics, even extremely heavy precipitation may result in only minor deformation when they are in the primary or secondary creeping stage. In contrast, slight precipitation may easily cause extremely intense deformation in the tertiary creeping stage (Yang et al. 2019). Thus, it is necessary to consider the influence of historical deformation on the current evolution of landslides (Cao et al. 2016). Thus, the cumulative displacements and their increments over the preceding 1–3 months were adopted as the candidate factors. These six historical displacement factors were defined as \( CF_{17}, CF_{18}, \ldots, CF_{22} \).

4 Results

Compared with the other eight monitoring sites, the monitoring sites ZG93, ZG118, and XD01 have two advantages: (1) the monitoring period is the longest, and the monitoring data are the most complete; and (2) they are all located in the middle of the landslide and can better reflect the whole landslide movement process. Therefore, we selected the data of these three monitoring sites as the prediction targets. Specifically, for the prediction of each monitoring site, the data from June 2006 to December 2013 (i.e., the data of 100 months) were used as the training set. The data from January 2014 to December 2016 (i.e., the data of 36 months) were used as the testing set. The ratio of testing data to training data is \( 36/100 = 0.36 \). The multi-step-ahead prediction method was used in the prediction of
the proposed model. In the following text, we took the data of monitoring site ZG93 as an example to introduce the details of the prediction process and its results obtained using the proposed model.

4.1 Decomposition of landslide displacements and factors based on the OVMD method

4.1.1 Decomposition parameters and decomposition results of landslide displacements

1. The optimal number of modes

Figure 6 shows the variations in the DCC between the decomposed components with the increasing of \( K \). When \( K = 4 \), two groups of medium-correlation or high-correlation decomposed components appeared. With the increase in \( K \) from 4 to 6, the number of medium-correlation and high-correlation components increased from 2 to 3. Clearly, \( K = 3 \) is the threshold at which high-correlation or medium-correlation components appeared in the displacement decomposition result of monitoring site ZG93.

Figure 7 shows the variations in the DCF and HSB between different components with the increase of \( K \). When \( K = 4 \), the HSB between different components is larger than the DCF between different components for the first time. Accompanying the increase of \( K \) from 4 to 6, the number of points with HSBs higher than the DCFs also increased from 3 to 6. It is clear that \( K = 3 \) is also the threshold at which the aliasing components appeared in the decomposition result of monitoring site ZG93.
According to the definition of the convergence conditions for the optimization of $K$ described in Sect. 2.5.1, $K = 3$ is the underdecomposition and overdecomposition threshold of the cumulative displacement for monitoring site ZG93, which means that $K_{\text{optimal}} = 3$. Monitoring sites ZG118 and XD01 have the same $K_{\text{optimal}}$ as monitoring site ZG93.

2. The optimal balancing parameter

After determining $K_{\text{optimal}}$, $\alpha_{\text{optimal}}$ was searched immediately. Figure 8 shows that the SNR and MAPEN
gradually decrease with the increase of $a$. This finding means that the larger $a$ is, the better the regularity of displacement components is, and the more accurate the reconstructed displacement is. However, when $a$ is more significant than a specific value, mode aliasing would appear in the decomposed components. Therefore, according to the definition of the CDI and its corresponding optimization principle, the $a_{\text{optimal}}$ for monitoring sites ZG93, ZG118, and XD01 were determined, as shown in Fig. 8d.

3. Displacement decomposition results

The decomposition of the cumulative displacements for monitoring sites ZG93, ZG118, and XD01 was conducted based on $K_{\text{optimal}}$ and $a_{\text{optimal}}$ described above. The cumulative displacements, decomposed displacement components, and residuals are shown in Fig. 9.

4.1.2 Decomposition parameters and the decomposition results of factors

OVMD was also used to decompose the candidate factors. The decomposition parameters of these factors are shown in Table 2. We can find that not all the candidate factors need to be decomposed. In addition, the decomposition parameters are almost the same for the same type of factors.

The representative decomposition results of the different types of factors for monitoring site ZG93, including $CF_3$, $CF_9$, $CF_{11}$, $CF_{14}$ and $CF_{17}$, are shown in Fig. 10. For the candidate factors that need to be decomposed, their components obtained from the decomposition have different variation characteristics.

4.2 Optimization of the input factors based on the PAWN method

4.2.1 GSA of candidate factors for prediction models

After establishing the prediction models for different displacement components, the validity of each model’s candidate factors was analyzed based on the PAWN. Figure 11 shows the mean KS statistics of the candidate factors. The red dotted line in Fig. 11 represents the mean KS statistics of the dummy factors. Most of the mean KS statistics of candidate factors are less than or remarkably close to those of dummy factors. This finding means that except for only a few valid candidate factors, most of the input factors have

| Candidate factors | Decomposition parameter | Decomposition or not | Reordered number of adaptively decomposed factors |
|-------------------|-------------------------|----------------------|-----------------------------------------------|
| $CF_1$            | 1                       | No                   | $F_1$                                         |
| $CF_2$            | 1                       | No                   | $F_2$                                         |
| $CF_3$            | 5                       | Yes                  | $F_{12}$, $F_{23}$, $F_{34}$, $F_{44}$, $F_{51}$ |
| $CF_4$            | 4                       | Yes                  | $F_{13}$, $F_{24}$, $F_{35}$, $F_{45}$         |
| $CS$              | 4                       | Yes                  | $F_{14}$, $F_{25}$, $F_{36}$, $F_{46}$         |
| $CF_6$            | 4                       | Yes                  | $F_{15}$, $F_{26}$, $F_{37}$, $F_{47}$         |
| $CF_7$            | 1                       | No                   | $F_3$                                         |
| $CF_8$            | 1                       | No                   | $F_4$                                         |
| $CF_9$            | 2                       | Yes                  | $F_{16}$, $F_{27}$                            |
| $CF_{10}$         | 1                       | No                   | $F_5$                                         |
| $CF_{11}$         | 4                       | Yes                  | $F_{17}$, $F_{28}$, $F_{38}$, $F_{48}$         |
| $CF_{12}$         | 1                       | No                   | $F_6$                                         |
| $CF_{13}$         | 1                       | No                   | $F_7$                                         |
| $CF_{14}$         | 5                       | Yes                  | $F_{18}$, $F_{29}$, $F_{39}$, $F_{49}$, $F_{52}$ |
| $CF_{15}$         | 1                       | No                   | $F_8$                                         |
| $CF_{16}$         | 4                       | Yes                  | $F_{19}$, $F_{30}$, $F_{40}$, $F_{50}$         |
| $CF_{17}$         | 3                       | Yes                  | $F_{20}$, $F_{31}$, $F_{41}$                  |
| $CF_{18}$         | 3                       | Yes                  | $F_{21}$, $F_{32}$, $F_{42}$                  |
| $CF_{19}$         | 3                       | Yes                  | $F_{22}$, $F_{33}$, $F_{43}$                  |
| $CF_{20}$         | 1                       | No                   | $F_9$                                         |
| $CF_{21}$         | 1                       | No                   | $F_{10}$                                      |
| $CF_{22}$         | 1                       | No                   | $F_{11}$                                      |
little or no influence on the results of the prediction models. Additionally, there is a certain similarity between their valid candidate factors for the prediction models of the same type of displacement components.

4.2.2 Selection of valid candidate factors

Figure 12 shows the optimized candidate factors obtained using the reduction method described in Sect. 2.5.2. The red dotted line represents the initial dividing line of the optimized factors. The green dotted line represents the final dividing line of the optimized factors. By using this approach, the number of candidate factors is reduced. For instance, the number of candidate factors is reduced from 52 to 7 within the trend displacement prediction process of monitoring site ZG93. Moreover, their candidate factors after reduction are quite different for the prediction models of different displacement components.

4.2.3 Computational stability of factor selection

The candidate factors of the different displacement component prediction models were reduced 50 times to evaluate the stability of the PAWN method. Figure 14 shows the statistical results of factor selections. The reduced result of candidate factors is very stable, and the remaining factors are concentrated on a few factors. In addition, in terms of the prediction model of the same displacement component for different monitoring sites, the 50 times factor selection results have some similarities in the selection of several core factors.

4.3 Cumulative displacement prediction based on the proposed model

To verify the accuracy of the proposed model, we established fifty integrated models to predict the cumulative displacements for each monitoring site. The final prediction results were the mean values of the prediction results provided by these fifty models. Four different VMD-GWO-KELM models were taken as the comparison models to compare the model performances. These comparison models were separately established based on the initial candidate factors, the candidate factors optimized by the GRA method, the adaptive decomposed factors, and the adaptive decomposed factors optimized by the GRA method. For optimizing the input factors by using the GRA method, please refer to the research published by Miao et al. (2017) and Yang et al. (2019).

Figure 15 and Table 3 show that the $R^2$ and MAPE of the proposed model are slightly higher or remarkably close to those of the comparison models. For the RMSE, the proposed models have relatively low RMSEs in the training process. They also have the lowest RMSEs in the prediction process. For the AIC, all the proposed models have the minimum AIC, whether in the process of prediction or training. In addition, the standard deviations of
the prediction results for the proposed model changed slightly with the increase of the prediction period.

Table 3 also shows the variations in each performance index caused by the change in the number of factors. In most cases, optimizing input factors is conducive to improving each model’s performance, whether in prediction or training. Only in very few cases does the optimization of factors lead to the decline in the $R^2$ and MAPE of the proposed model in prediction. However, compared with the number of factors that need to be increased to improve these two indexes in the comparison models, this decrease in the $R^2$ and MAPE of the proposed model is entirely acceptable to users.

5 Discussion

5.1 Component extraction of the landslide displacement and its factors

As shown in Fig. 9, the trend displacements retain the changing trend in the original cumulative displacements.
For the periodic displacements, there is a good correspondence between their periodic variations and the step-like change in the original cumulative displacements. Their variations have a particular uncertainty and randomness for the random displacements. These results are consistent with the definition of the landslide displacement components in the TSA, which supports that the proposed OVMD can effectively decompose the landslide displacements. Moreover, the residuals between the reconstructed cumulative displacements and the original cumulative displacements account for only approximately 1/1000 of the original cumulative displacements. Using the reconstructed cumulative displacements obtained by the OVMD for prediction would not yield a misjudgment of the original cumulative displacements of landslides.

The decomposition of factors has been proposed only in recent years. Its purpose is to extract the components in each factor that are more related to their corresponding displacement components to build prediction models with better performance. In this kind of research (Deng et al. 2017; Li et al. 2018; Guo et al. 2020), it was usually assumed that there were two components with different frequencies in the factors. Then the correlations between the two components and the prediction objects were analyzed. According to the analysis results, the prediction model based on the high-correlation components was built to achieve a more accurate prediction. However, as shown in Table 2, it is found that not all the factors have the above two components. In addition, a few factors have more than two components with different physical meanings. This fact reminds us that we must pay more attention to avoid performance improvements based on fake factor components, thus increasing the risk of incorrect predictions.
5.2 Influence of factor coupling on the prediction results

Extracting valid factors has become the key to improving the overall performance of prediction models. In past research, GRA was a widely used method to determine the input factors of the model with the advantages of strong generality (Yang et al. 2019). It is very suitable for selecting input factors to quickly remove some irrelevant and low-correlation input factors without the need for training models.

Figure 13 shows that only a portion of the optimization factors obtained by the proposed PAWN-based method has high correlations with the prediction outputs. This result is significantly different from the optimization factors obtained by the GRA method. This is because the coupling influence between low-correlation factors on the prediction outputs has been considered when the PAWN-based method is used.

Table 3 shows that compared with no optimization of factors, using the GRA method can improve the prediction performance of models. However, the performance improvements obtained using the GRA method are still lower than those obtained using the proposed PAWN-based method. This indicates that considering the impact of factor coupling on the prediction results is beneficial to improving the overall performance of prediction models.
the prediction performance. Thus, in constructing prediction models, we should choose the high-sensitivity factors as the input factors instead of the high-correlation factors. The proposed PAWN-based method has apparent advantages in finding high-sensitivity factors compared with the GRA method.

5.3 Performance of the proposed model and its application prospects

Table 3 shows that the average prediction performance of the proposed OVMD-GWO-KELM model is slightly higher than that of the comparison models in most cases. For predicting the extreme values of displacements (i.e., when the displacements with significant growth are predicted), the proposed model has a better generalization ability than the comparison models. Meanwhile, the optimized input factors make the complexity of the proposed model significantly lower than those of the comparison models. In addition, the computational stability of the proposed model is also confirmed according to the variation characteristics of the results of the fifty multistep-ahead predictions. The above results show that adopting the proposed model is remarkably effective for improving the performance of the EWS for landslide hazards.

However, it should be noted that when using the proposed model to predict the displacement of other landslides, it is vital to propose more targeted candidate factors based on the deformation response to establish a more accurate prediction model. In addition, if the proposed model cannot be dynamically updated with the growth of prediction time, its prediction accuracy will inevitably decrease. Therefore, in future research or applications of the proposed model, its adaptive updating function should be subjected to more targeted optimization.
6 Conclusion

This study proposed an adaptive hybrid machine learning model, namely, the OVMD-GWO-KELM model. Taking the Baishuihe landslide as a case study, the process of applying this model for landslide cumulative displacement prediction with multistep-ahead prediction was discussed in detail, and the following conclusions were drawn from this study:

1. The adaptive decomposition results obtained using OVMD show three displacement components with clear physical meaning and different variation characteristics in the cumulative displacement of the Baishuihe landslide. This finding is consistent with the conclusion of the TSA model of landslide displacement for step-like landslides.

2. Compared to the GRA method, the LHS-PAWN method can analyze the influence of the interaction between various low-correlation factors on the results of prediction models. Thus, considering the redundancy and lack of information in raw data, the low-sensitivity factors and combinations of low-sensitivity factors could be correctly reduced by applying this LHS-PAWN method.

3. Compared with four different VMD-GWO-KELM models, the proposed model can significantly improve its generalization without significantly reducing its training effect, so the overfitting and underfitting...
Performance comparison of prediction models for the cumulative displacement of monitoring sites ZG93, ZG118, and XD01

| Monitoring site | Model | Evaluation index | Total number of input factors |
|----------------|-------|------------------|------------------------------|
|                |       | RMSE Training    | Testing                      |                               |
|                |       |                  |                              |                               |
| ZG93           | ①    | 17.4751          | 16.3650                      |                               |
|                | ②    | 38.6489          | 40.2726                      |                               |
|                | ③    | (−0.5804)        | (−0.6554)                    |                               |
|                | ④    | 39.2250          | 27.9394                      |                               |
|                | ⑤    | (−2.5648)        | (−1.3649)                    |                               |
|                | ⑥    | 14.5159          | 34.3427                      |                               |
|                | ⑦    | (+0.0190)        | (−0.1152)                    |                               |
|                | ⑧    | 16.0672          | 34.0923                      |                               |
|                | ⑨    | (+0.0151)        | (−0.1896)                    |                               |
| ZG118          | ①    | 16.6814          | 23.1888                      |                               |
|                | ②    | 33.8939          | 32.2191                      |                               |
|                | ③    | (−0.4027)        | (−0.2113)                    |                               |
|                | ④    | 36.6687          | 24.9150                      |                               |
|                | ⑤    | (−1.2698)        | (−0.1097)                    |                               |
|                | ⑥    | 13.7103          | 46.0783                      |                               |
|                | ⑦    | (+0.0224)        | (−0.1724)                    |                               |
|                | ⑧    | 14.8868          | 23.7909                      |                               |
|                | ⑨    | (+0.0170)        | (−0.0057)                    |                               |
| XD01           | ①    | 25.6475          | 23.6041                      |                               |
|                | ②    | 49.0062          | 58.4528                      |                               |
|                | ③    | (−0.5899)        | (−0.8800)                    |                               |
|                | ④    | 40.3796          | 57.8299                      |                               |
|                | ⑤    | (−1.1692)        | (−2.7163)                    |                               |
|                | ⑥    | 14.4541          | 35.0616                      |                               |
|                | ⑦    | (+0.0864)        | (−0.0848)                    |                               |
|                | ⑧    | 20.6926          | 30.0310                      |                               |
|                | ⑨    | (+0.0478)        | (−0.0620)                    |                               |

① denotes the proposed OVM-D-GWO-KELM model based on optimized factors; ② denotes the VMD-GWO-KELM model based on candidate factors; ③ denotes the VMD-GWO-KELM model based on candidate factors that are optimized by the GRA method; ④ denotes the VMD-GWO-KELM model based on adaptively decomposed candidate factors; ⑤ denotes the VMD-GWO-KELM model based on adaptively decomposed candidate factors that are optimized by the GRA method. Taking the proposed model as the comparison target, the positive or negative values in brackets respectively represent the improvement or decrease of the model performance indexes with each addition of an input factor.
phenomena can be effectively suppressed. In addition, the complexity of the model is typically low, and the stability of its calculation results is still high with varying time.

Acknowledgements The authors thank the colleagues in our laboratory for their constructive comments and assistance.

Author contributions Li Linwei and Wu Yiping contributed to the conception of the study and wrote the manuscript. Li Linwei and Huang Yeping performed the experiment and the data analyses. Li Bo and Miao Fasheng contributed significantly to the analysis and manuscript preparation. Deng Ziqiang assisted Li Linwei in revising the manuscript and polishing the language. All authors have read and approved the final manuscript.

Funding The study is funded by the National Natural Science Foundation of China (No. 41977244), the Doctoral Research Foundation of Guizhou University (GDRJH 2021 [25]), and the Project of Guizhou Provincial Department of Science and Technology (Guizhou science and technology cooperation platform for talents [2021] 5626).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bogaard T, Greco R (2018) Invited perspectives: hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydrometeorological thresholds. Nat Hazard 18(1):31–39
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
Casagli N, Catani F, Del Ventisette C, Luzi G (2010) Monitoring, prediction, and early warning using ground-based radar interferometry. Landslides 7(3):291–301
Criss RE, Yao W, Li C, Tang H (2020) A predictive, two-parameter model for the movement of reservoir landslides. J Earth Sci 31(6):1051–1057
Crosa GB (2004) Introduction to the special issue on rainfall-triggered landslides and debris flows. Eng Geol 3(73):191–192
Deng DM, Liang Y, Wang LQ, Wang CS, Sun ZH, Wang C, Dong MM (2017) Displacement prediction method based on ensemble empirical mode decomposition and support vector machine regression-a case of landslides in Three Gorges Reservoir area. Rock Soil Mech 38(12):3660–3669 (in Chinese with English Abstract)
Dragomiretskiy K, Zosso D (2013) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544
Du J, Yin K, Lacasse S (2013) Displacement prediction in colluvial landslides, Three Gorges Reservoir, China. Landslides 10(2):203–218
Fukuzono T (1985) A new method for predicting the failure time of a slope. In Proceedings of 4th international conference and field workshop on landslide, 1985, pp 145–150
Glade T, Crozier MJ (2005) Landslide hazard and risk: concluding comment and perspectives. Landslide hazard and risk. Wiley, Chichester, pp 767–774
Guo Z, Chen L, Gui L et al (2020) Landslide displacement prediction based on variational mode decomposition and WA-GWO-BP model. Landslides 17:567–583
Haque U, Blum P, da Silva PF, Andersen P, Pilz J, Chalov SR, Malet JP, Auflüg MJ, Andres N, Poyiadji E, Lamas PC, Zhang W, Peshkevski I, Petursson HG, Kurt T, Dobrev N, Garcia-Davalillo JC, Halkia M, Ferri S, Gaprindashvili G, Engström J, Keelings D (2016) Fatal landslides in Europe. Landslides 13(6):1545–1554
Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
Intrieri E, Gigli G, Casagli N, Nadim F (2013) Brief communication: landslide early warning system: toolbox and general concepts. Nat Hazard 13:85–90
Intrieri E, Carlà T, Gigli G (2019) Forecasting the time of failure of landslides at slope-scale: a literature review. Earth Sci Rev 193:333–349
Iqbal J, Dai F, Hong M, Tu X, Xie Q (2018) Failure mechanism and stability analysis of an active landslide in the Xiangjiang reservoir area, southwest China. J Earth Sci 29(3):640–661
Kirschbaum DB, Adler R, Yang H, Hill S, Lerner-Lam A (2010) A global landslide catalog for hazard applications: method, results, and limitations. Nat Hazards 52(3):561–575
Krmkaš M, Špoljaric D, Bernat S, Arbanas SM (2017) Method for prediction of landslide movements based on random forests. Landslides 14(3):947–960
Lee CT, Huang CC, Lee JF, Pan KL, Lin ML, Dng JJ (2008) Statistical approach to earthquake-induced landslide susceptibility. Eng Geol 100(1–2):43–58
Li C, Tang H, Hu X, Li D, Hu B (2009) Landslide prediction based on wavelet analysis and cusp catastrophe. J Earth Sci 20(6):971–977
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 L, Wu Y, Miao F, Liao K, Zhang L (2018) Displacement prediction of landslides based on variational mode decomposition and GWO-MIC-SVR model. Chin J Rock Mech Eng 37(6):1395–1406
Li L, Wu Y, Miao F, Xue Y, Huang Y (2021) A hybrid interval displacement forecasting model for reservoir colluvial landslides with step-like deformation characteristics considering dynamic switching of deformation states. Stoch Environ Res Risk Assess 35(6):1089–1112
Lian C, Zeng Z, Yao W, Tang H (2013) Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine. Nat Hazards 66(2):759–771
Ma J, Wang Y, Niu X, Jiang S, Liu Z (2022) A comparative study of mutual information-based input variable selection strategies for the displacement prediction of seepage-driven landslides using optimized support vector regression. Stoch Environ Res Risk Assess 36:3109–3129
Miao F, Wu Y, Xie Y, Yu F, Peng L (2017) Research on progressive failure process of Baishuihe landslide based on Monte Carlo model. Stoch Environ Res Risk Assess 31(7):1683–1696
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
Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
