



## Article

# Predicting the Deformation of a Slope Using a Random Coefficient Panel Data Model

Zhenxia Yuan <sup>1,2,\*</sup>, Yadong Bian <sup>1,2,3</sup>, Weijian Liu <sup>1</sup>, Fuzhou Qi <sup>1</sup>, Haohao Ma <sup>1</sup>, Sen Zheng <sup>4</sup> and Zhenzhu Meng <sup>5</sup>

<sup>1</sup> School of Architecture and Civil Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China; 5756@zut.edu.cn (Y.B.); 6533@zut.edu.cn (W.L.); 6851@zut.edu.cn (F.Q.); 6441@zut.edu.cn (H.M.)

<sup>2</sup> Henan Environmental Geotechnical Engineering and Underground Engineering Disaster Control Engineering Research Center, Zhengzhou 450007, China

<sup>3</sup> School of Civil Engineering, Henan Polytechnic University, Jiaozuo 454000, China

<sup>4</sup> Laboratory of Hydraulic Engineering, School of Civil Engineering, 1015 Lausanne, Switzerland; sen.zheng@epfl.ch

<sup>5</sup> School of Water Conservancy and Environment Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China; mengzhzh@zjweu.edu.cn

\* Correspondence: 6570@zut.edu.cn

**Abstract:** Engineering constructions in coastal areas not only affect existing landslides, but also induce new landslides. Variation of the water level makes the coastal area a geological hazard-prone. Prediction of the slope displacement based on monitoring data plays an important role in early warning of potential landslide and slope failure, and supports the risk management of hazards. Given the complex characteristic of the slope deformation, we proposed a prediction model using random coefficient model under the frame of panel data analysis, so as to take the correlation among monitoring points into consideration. In addition, we classified the monitoring data using Gaussian mixture model, to take the temporal-spatial characteristics into consideration. Monitoring data of Guobu slope was used to validate the model. Results indicated that the proposed model have a better performance in prediction accuracy. We also compared the proposed model with the BP neural network model and temporal – temperature model, and found that the prediction accuracy of the proposed model is better than those of the two control models.

**Keywords:** landslide; slope deformation; random coefficient model; monitoring data; clustering; Gaussian mixture model



**Citation:** Yuan, Z.; Bian, Y.; Liu, W.; Qi, F.; Ma, H.; Zheng, S.; Meng, Z. Predicting the Deformation of a Slope Using a Random Coefficient Panel Data Model. *Fractal Fract.* **2024**, *8*, 429. <https://doi.org/10.3390/fractalfract8070429>

Academic Editor: Carlo Cattani

Received: 17 May 2024

Revised: 10 July 2024

Accepted: 16 July 2024

Published: 22 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Landslides and slope failures are worldwide natural disasters [1–3]. In coastal areas, the construction of hydraulic constructions may change the geological environment of banks slopes, and thus influence the stability of slopes. Hazards such as landslides and slope failures are also affected by variations in water level. Water not only activate existing landslides on the slope, but also may increase the possibility of new landslides, making the water area a geological high-risk area [4]. For example, resulting from the construction of the Grand Coulee Dam in Roosevelt Lake, America, from 1941 to 1953, more than 500 landslides occurred [5]. The same problems also occurred in China’s Three Gorges Reservoir [6–8]. Since the Three Gorges Reservoir has been constructed, due to the rise in water level in the reservoir, more than 4 thousand geological hazards have occurred from Yichang to Jiangjin along the Yangtze River, most of which were landslides, with a total volume of more than 4 billion m<sup>3</sup>, such as the famous Qianjiangping landslide, the Shuping landslide, and the Muyubao landslide [9–11].

Once a slope near water areas such as reservoirs and lakes fails, it may induce a series of secondary disasters such as disastrous waves, and destroy the nearby hydraulic constructions [12–14]. For instance, a historical event occurred at the Vajont Reservoir in

Italy in 1963 [15,16]. The landslide formed an impulse wave that over-topped the dam and damaged two villages downstream. A recent example was occurred in Greenland in 2017, where a huge landslide impacted the Karrat Fjord and generated a wave which traveled 32 km to Nuugaatsiaq [17,18]. Monitoring systems and early warning systems have been widely used to control the risk of such natural hazards. By analyzing the monitoring data recorded from the monitoring systems, previous studies have found that the possibility of failure of bank slope in coastal areas is not only relevant to the geological factors, but also to the water level fluctuations. Therefore, in this paper, we gave a specific insight to the reservoir bank slope, with consideration of fluctuation of the water level.

Slope deformation is the most intuitive indicator for characterizing the inherent evolution of landslides. Predicting the deformation of slope is of great importance for detecting potential slope failures and landslides. Previous studies have attempted a number of methods to predict the slope deformation, including numerical simulations and statistical analyses based on monitoring data. The current numerical methods for evaluating the slope stability include the finite element method [19–21], the distinct element method [22,23], the limit equilibrium method [24], and so on. Researchers have also carried out large deformation analyses for slope instability. For example, Esposito et al. proposed a numerical method based on practical equivalent continuum approach, so as to analyze the deformation of a gravitational slope in the Lepini Mountains, Italy [25]. Guallini et al. conducted a structural and kinematic analysis for the deformational system observed in the South Polar Layered Deposits [26]. However, due to the complexity of the structure of slopes, numerical simulations can hardly reflect the actual internal mechanisms of slope deformation.

Statistical analysis based on monitoring data quantify the relation between the slope deformation and the relative explanatory variables obtained monitoring system [27]. The advantage of statistical methods is that the prediction model uses real-time monitoring data, and can easily update up-to-the-minute monitoring data in the model [28–30]. Previous studies have attempted numbers of monitoring methods. For instance, Liu et al. used a InSAR technology to identify landslides in Badong, and found the correlation between landslide movement and water level variation [31].

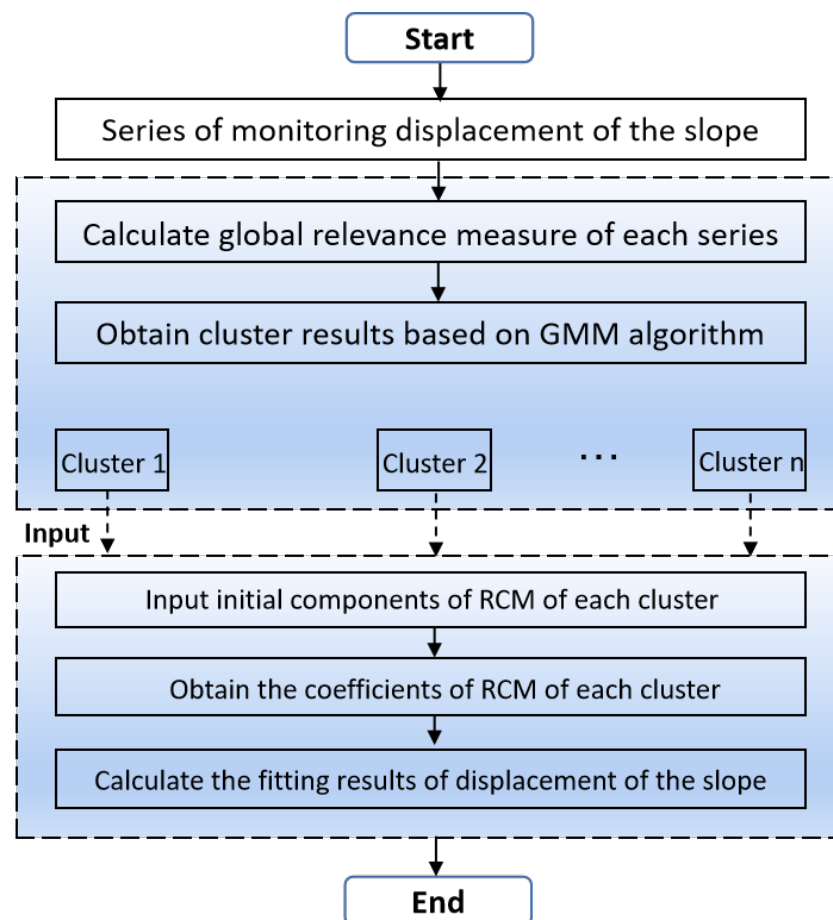
With the increase in monitoring sensors installed in the slope, mass monitoring data are obtained from the monitoring system. With the developing of computing science, intelligent computing methods, such as gray theory, artificial neural networks, and support vector machines, are applied to monitoring data modeling [32–35]. Also, prediction models are improved from single monitoring point models to multiple monitoring points models. It should be noted that the displacement at each monitoring point relates to the displacement in the surrounding area. However, none has consider the correlations among different monitoring points in prediction models of slope displacement.

In this paper, to consider the correlations among monitoring points, we proposed a prediction method under the frame of panel data analysis, combining random coefficient model and Gaussian mixture model. We first classified the monitoring points into four groups, based on their data characteristics using a Gaussian mixture model, so as to take the temporal-spatial characteristics of monitoring points into account. With the clustering results, we established the random coefficient model for each group. A case study of the Guobu slope [36], a bank slope at the Laxiwa reservoir, was conducted to validate the model. Specifically, as the stability of slope near water areas is affected by the fluctuation of water level, the reservoir water level was selected as an explanatory variable in the proposed model. We also compared the prediction accuracy of the proposed model with the BP neural network model and temporal-temperature model, which are one of the most commonly used artificial models and statistical models, respectively.

The paper is organized as follows: Section 2 introduces the model development, including the random coefficient model and the clustering method. Section 3 presents the geological information of the Guobu slope study area and the monitoring dataset selected in this paper. Prediction results of the proposed model, as well as a comparison with other methods, are given in Section 4. Section 5 concludes the paper with highlighted remarks.

## 2. Model Development

In this section, to consider the correlations among monitoring points, we developed the prediction model based on the random coefficient model (RCM) and Gaussian mixture model (GMM) under the frame of panel data analysis. As the first step, we classified the monitoring points into several groups according to the data characteristics using the Gaussian mixture model, in order to take the temporal-spatial characteristics of monitoring points into account. With the clustering results, we established the random coefficient model for each data group. Specifically, as the stability of slope near water areas is affected by the fluctuation of water level, the reservoir water level was selected as an explanatory variable in the proposed model. The flow chart of the proposed model for predicting the displacement of slope is shown in Figure 1.



**Figure 1.** Flow chart of the proposed model for predicting the displacement of slope.

### 2.1. Data Clustering Based on Gaussian Mixture Model

To correlations among data at different monitoring points, we classified the monitoring points into several groups according to the data characteristics using Gaussian mixture model (GMM).

The GMM is a probabilistic model that assumes all the data points are generated from a mixture of several Gaussian distributions with unknown parameters, which means the data can be fitted by mixtures of several Gaussian distributions using maximum likelihood estimation method. The model aims to identify the underlying subpopulations within the overall population. Using one-dimensional dataset as an example, the probability density function of variable  $x$  obeys a mixture of two Gaussian distributions:

$$P(x|\mu_1, \mu_2, \sigma) = \sum_{k=1}^2 p_k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu_k)^2}{2\sigma^2}\right) \quad (1)$$

where  $k = 1, 2$  counts the two Gaussian distributions, the  $k$ th prior probability is  $\{p_1 = 1/2, p_2 = 1/2\}$ ,  $\{\mu_k\}$  denotes the mean value, and  $\sigma$  denotes the standard deviation of the two Gaussian distributions.  $\{x_n\}_{n=1}^N$  are independent samples in the dataset which follow Gaussian distributions.  $k_n$  denotes the unknown cluster tag of the  $n$ th test.

In the case where  $\sigma$  and  $\{\mu_k\}$  are known, the posterior probability of  $n$ th test  $k_n$  can be written as

$$P(k_n|x_n, \theta) = \frac{1}{1 + \exp[-(\omega_1 x_n + \omega_0)]}. \quad (2)$$

with  $\theta = \{\{\mu_k\}, \sigma\}$ . In the case where  $\{\mu_k\}$  is unknown and  $\sigma$  is known,  $\{\mu_k\}$  can be obtained from  $\{x_n\}_{n=1}^N$ . Then, the iterative algorithm of  $\{\mu_k\}$  that maximizes the likelihood estimation can be determined:

$$P(\{x_n\}_{n=1}^N | \{\mu_k\}, \sigma) = \prod_n P(x_n | \{\mu_k\}, \sigma). \quad (3)$$

Considering  $L$  as the natural logarithm of the likelihood, the derivative of logarithm likelihood of  $\mu_k$  is:

$$\frac{\partial}{\partial \mu_k} L = \sum_n p_{k|n} \frac{x_n - \mu_k}{\sigma^2}, \quad (4)$$

with  $p_{k|n} = P(k_n = k | x_n, \theta)$  being the Gaussian density. Ignoring the items in  $\frac{\partial}{\partial \mu_k} P(k_n = k | x_n, \theta)$ , the second derivative can be approximated as:

$$\frac{\partial^2}{\partial \mu_k^2} L = - \sum_n p_{k|n} \frac{1}{\sigma^2}. \quad (5)$$

Then, the initial  $\mu_1, \mu_2$  can be iterated to  $\mu'_1, \mu'_2$  using the approximate Newton–Raphson steps:

$$\mu'_k = \frac{\sum_n p_{k|n} x_n}{\sum_n p_{k|n}}. \quad (6)$$

For a multi-dimensional dataset, the mixture density of Gaussian distributions can be written as:

$$p_{k|n} = \frac{\pi_k \frac{1}{\prod_{i=1}^I \sqrt{2\pi} \sigma_i^{(k)}} \exp\left(-\sum_i^I (\mu_i^{(k)} - x_i^{(n)})^2 / 2(\sigma_i^{(k)})^2\right)}{\sum_{k'} \pi_{k'} \frac{1}{\prod_{i=1}^I \sqrt{2\pi} \sigma_i^{(k')}} \exp\left(-\sum_i^I (\mu_i^{(k')} - x_i^{(n)})^2 / 2(\sigma_i^{(k')})^2\right)}, \quad (7)$$

where  $k$  counts the Gaussian distribution,  $i$  counts the dimension of data,  $n$  counts the data sequence,  $I$  denotes the total number of the data's dimension,  $\pi_k$  the weighting,  $\mu_i^{(k)}$  and  $\sigma_i^{(k)}$  are the mean value and variance of Gaussian distribution, and  $x_i^{(n)}$  is the variance. The iterative formulas of  $\sigma_i^{(k)}$  and  $\pi_k$  are:

$$\sigma_i^{2(k)} = \frac{\sum_n p_{k|n} (x_i^{(n)} - \mu_i^{(k)})^2}{\sum_n p_{k|n}} \quad (8)$$

$$\pi_k = \frac{\sum_n p_{k|n}}{\sum_k \sum_n p_{k|n}} \quad (9)$$

These equations form the basis of the Expectation-Maximization algorithm used to estimate the parameters of the GMM iteratively. The GMM uses the principle of maximum likelihood estimation to iteratively update the parameters of multiple Gaussian distributions that best fit the observed data. By classifying data points into different Gaus-



sian components, GMM captures the underlying structure of the data, which can exhibit complex dependencies and multi-scale patterns often seen in fractional processes.

### 2.2. Random Coefficient Model

For the panel data analysis, the dataset can be considered as a three dimensional dataset, which contains a time series and a data panel. Figure 2 illustrates the structure of a panel data model. The random coefficient panel data model (RCM) allows regression coefficients to vary across different individuals or entities. This model is particularly useful for panel data, which contains multiple observations over time for the same entities, capturing individual-specific effects. Equation (10) expresses a two dimensional panel regression model:

$$y_{it} = \sum_{k=1}^K \beta_{ki} x_{kit} + u_{it} = \sum_{k=1}^K (\beta_k + \gamma_{ki}) x_{kit} + u, \tag{10}$$

where  $y_{it}$  is the displacement data panel,  $x_{kit}$  denotes the explanatory variables,  $t$  is the time index,  $i$  is the cross section index,  $k$  is the explanatory variables index,  $\beta_{ki}$  includes  $\beta_k$  and  $\gamma_{ki}$ ,  $\beta = (\beta_1, \dots, \beta_K)'$  is the common mean coefficient vector,  $\gamma = (\gamma_{1i}, \dots, \gamma_{Ki})'$  is the operator between each individual and common mean value, and  $u$  is a random interference term.

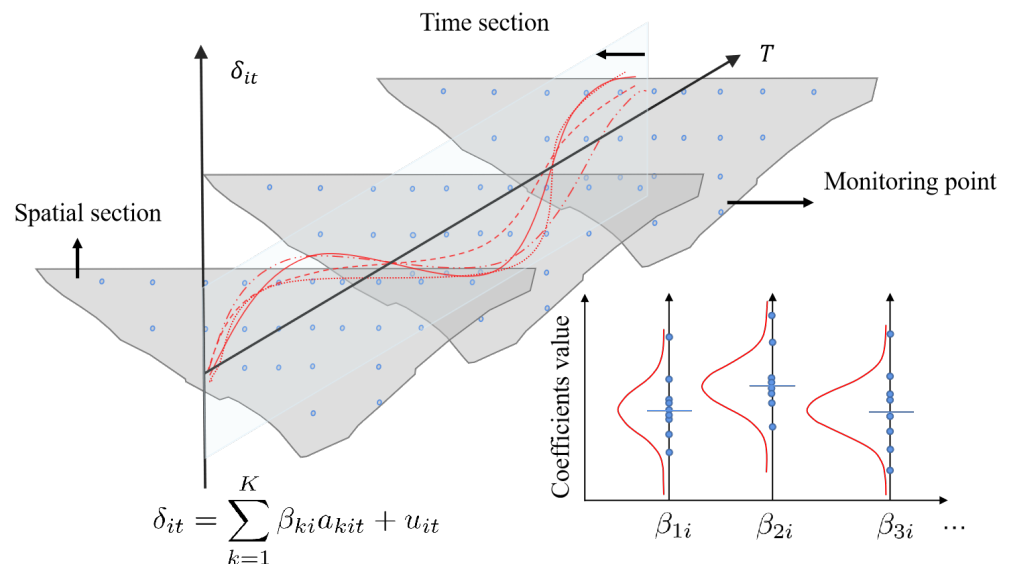


Figure 2. Sketch of the random coefficient panel data model.

Assuming  $\beta_i = \beta + \gamma_i$  is a random variable:

$$\begin{aligned} E(\gamma_i) &= 0, \\ E(\gamma_i \gamma_j') &= \begin{cases} \Delta, & i = j, \\ 0, & i \neq j, \end{cases} \\ E(x_{it} \gamma_j') &= 0, \\ E(u_i u_j') &= \begin{cases} \sigma_i^2 I_T & i = j, \\ 0, & i \neq j. \end{cases} \end{aligned} \tag{11}$$

Integrating the  $NT$ th observation data point, the following equation can be obtained:

$$y = X\beta + \tilde{X}\gamma + u, y_{NT \times 1} = (y'_1, \dots, y'_N)' \tag{12}$$

where  $X_{NT \times K} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$ ,  $\tilde{X}_{NT \times NK} = u = (u'_1, \dots, u'_N)'$ ,  $\gamma = (\gamma'_1, \dots, \gamma'_N)$ ,  $N$  counts

the number of panel,  $T$  is time series, the  $i$ -th diagonal block is  $\psi_i = X_i \Delta X'_i + \sigma_i^2 I_T$ , and  $\tilde{X} \gamma + u$  is the compound error term.

The estimator of  $\beta$  with the ordinary least squares is biased. Once  $\frac{1}{NT} X'X$  converges into a non-zero constant matrix, we can hence obtain a consistent ineffective estimation. The optimal linear unbiased estimator of  $\beta$  can be removed from the GLS:

$$\begin{aligned} \hat{\beta}_{GLS} &= \left( \sum_{i=1}^N X'_i \psi_i^{-1} X_i \right)^{-1} \left( \sum_{i=1}^N X'_i \psi_i^{-1} y_i \right) = \sum_{i=1}^N W_i \hat{\beta}_i, \\ W_i &= \left\{ \sum_{i=1}^N \left[ \Delta + \sigma_i^2 (X'_i X_i)^{-1} \right]^{-1} \right\}^{-1} \left[ \Delta + \sigma_i^2 (X'_i X_i)^{-1} \right]^{-1}, \\ \hat{\beta}_i &= (X'_i X_i)^{-1} X'_i y_i. \end{aligned} \quad (13)$$

$\hat{\beta}_{GLS}$  is an effective estimation of  $\beta$  which follows an asymptotic normal distribution. The variance of the estimator is:

$$\text{Var}(\hat{\beta}_{GLS}) = \left( \sum_{i=1}^N X'_i \psi_i^{-1} X_i \right)^{-1} = \left\{ \sum_{i=1}^N \left[ \Delta + \sigma_i^2 (X'_i X_i)^{-1} \right]^{-1} \right\}^{-1}. \quad (14)$$

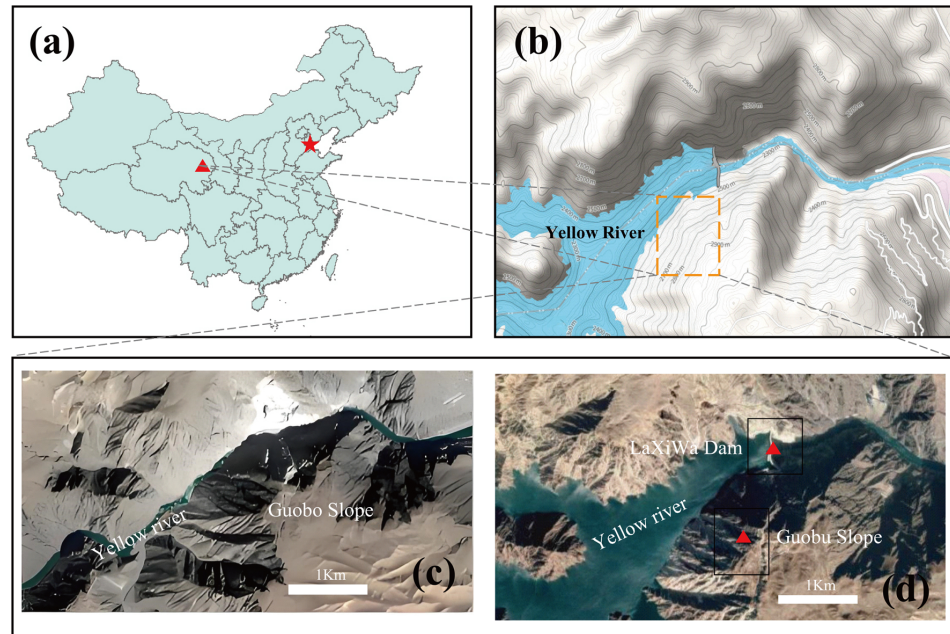
### 3. Study Area and Dataset

#### 3.1. Study Area

The Laxiwa hydropower station is located in Laxiwa valley, Qinghai Province, China. It is known as the largest hydro-power station constructed on the Yellow River. The construction of the station has been prepared since October 2001. The excavation on the dam's abutment began in September 2004, and the impoundment started in March 2009. After the impoundment, the water level rose by about 200 m from 2254 m. The first two electric generators were put into function in May of the same year [37,38].

The Guobu slope is a northwest-oriented slope with an average angle of  $43^\circ$ , and is located at the right bank of upstream of the dam at Laxiwa station. Figure 3 illustrates the geographical location of the Guobu slope, and the relative location between Laxiwa hydropower station and Guobu slope. Remarkable deformation, frequent rockfalls, and small scale collapses have been observed on the Guobu slope in 3 months after the impoundment of the dam at Laxiwa station [39].

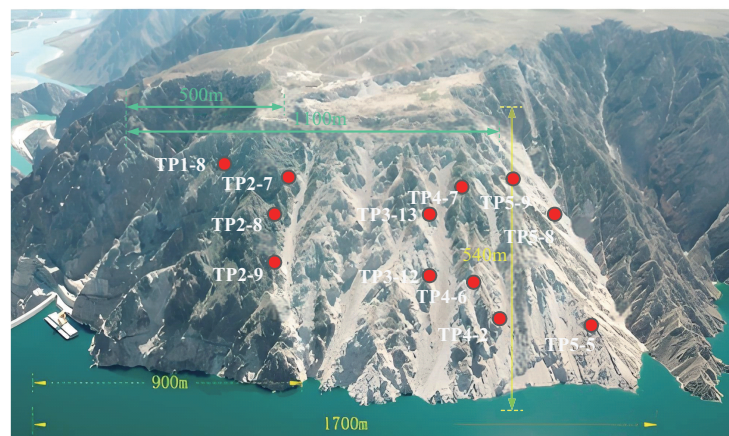
As the deforming slope is very close to the dam site (about 700 m from the dam) and has a large volume (about 700 m high and 1000 m wide in average), once the slope fails, the event will destroy the dam and the hydropower station. Also, the associated waves created by the induced landslide would flood the downstream area, and threaten people and constructions located downstream. Similar disasters had occurred in history; in 1993 in Vajont, Italy, failure of the left bank slope of the Vajont reservoir over-topped the dam and caused more than 2000 casualties downstream [40]. According to the characteristics of the slope surface, previous research has estimated that the total volume of potentially unstable mass can be 120 to 150  $\text{Mm}^3$ ; that is, about one half of that of the Vajont landslide [36].



**Figure 3.** Geographical location of Guobu slope: (a) geographical location, (b) contour map, (c) distant view, (d) relative location between Guobu slope and Laxiwa dam.

### 3.2. Dataset

The monitoring system of deformation of the slope was built quickly after the deformation had been observed. The deformation of the slope was monitored since then. Different techniques were used to monitor the situation of the slope: displacement monitoring sensors, boreholes, exploration tunnels, and so on. All monitoring results reveal that the slope is deforming continuously, and the risk of slope failure exists. Due to space limitations, in this paper, we only selected the surface displacement monitoring dataset of 12 monitoring points to establish the prediction model, i.e., TP1-8, TP2-7, TP2-8, TP-9, TP3-12, TP3-13, TP4-2, TP4-6, TP4-7, TP5-5, TP5-8, and TP5-9. The data were collected from 29 May 2011, to 13 November 2013. The sensor that recorded the displacement data was a prism. Figure 4 shows the geological positions of selected monitoring points on the slope. The scale of the picture is marked by the yellow line at the bottom of the figure. The monitoring points are evenly distributed on the topping of the slope. Table 1 exhibits the coordinates  $(x, y)$  and altitude  $z$  of the 12 monitoring points selected in this paper. The Cartesian coordinate system is used to describe the location of monitoring points. The origin of  $(x, y)$  is  $36.006^\circ$  N,  $101.149^\circ$  E.



**Figure 4.** The positions of the selected monitoring points.

**Table 1.** The spatial coordinates ( $x, y, z$ ) of the selected monitoring points.

Monitoring Point	Coordinate $x$ (m)	Coordinate $y$ (m)	Coordinate $z$ (m)
TP1-8	3053.13	6483.37	2830.24
TP2-7	2896.17	6369.15	2802.69
TP2-8	2978.88	6274.93	2701.89
TP2-9	3088.51	6164.12	2602.38
TP3-12	2845.30	5946.82	2592.63
TP3-13	2720.19	6070.24	2703.45
TP4-2	2705.62	5789.85	2547.86
TP4-6	2693.06	5859.02	2602.79
TP4-7	2542.74	6053.20	2783.44
TP5-5	2442.54	5621.43	2497.97
TP5-8	2458.04	5956.94	2732.45
TP5-9	2515.49	6103.51	2825.94

The displacement data of the selected 12 monitoring points are used to develop the prediction model. Figure 5 shows the time variation of monitored slope displacement at each monitoring point. In general, the displacement data at all monitoring points increase over time. The varying tendencies of monitoring data at monitoring points TP1-8, TP2-7, TP2-8 and TP2-9 fluctuate significantly, whereas the varying tendencies of monitoring data at other monitoring points are relatively smooth. It can be seen from the figure that the displacements of TP1-8, TP2-7, TP2-8 and TP2-9 are much smaller than those of other monitoring points. The maximum displacements of these four monitoring points are lower than 250 mm, the maximum displacements at other monitoring points are larger than 2500 mm. To ensure the visual presentation quality, we here presented data of the 12 monitoring points using two subfigures. Resulting from the differences of data varying range and sensors installed in the slope, the accuracy of monitoring data at each monitoring points are different. It should be noted that the characteristics of monitoring data may affect the performance and accuracy of prediction model.

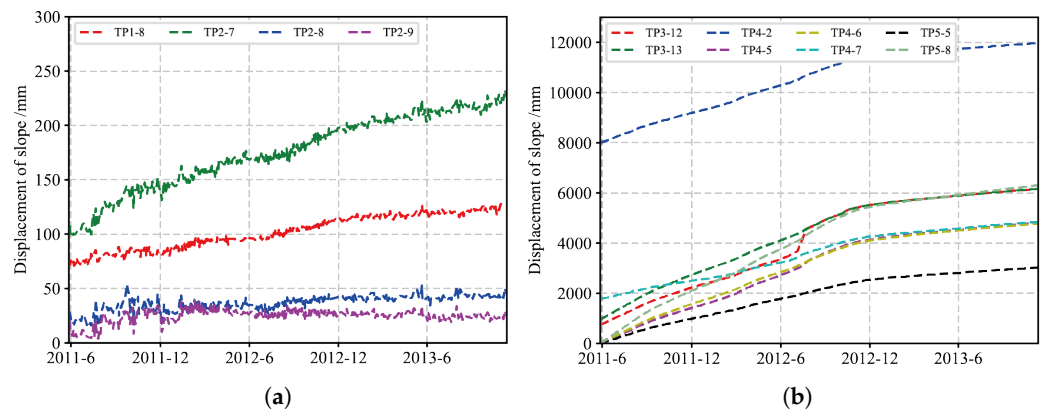
**Figure 5.** Time variation of monitored slope displacement at (a) monitoring points TP1-8, TP2-7, TP2-8, TP2-9 and (b) monitoring points TP3-12, TP3-13, TP4-2, TP4-5, TP4-6, TP4-7, TP5-5, TP5-8.

Figure 6 shows the time variation of the water level in the reservoir. The water level was relatively stable during two periods: the water level fluctuated around 2430 m during June 2011 to December 2011; and the water level fluctuated around 2445 m during August 2012 to November 2013. The water level increased gradually from 2430 m to 2445 m from December 2011 to August 2013. In addition, it can be seen from Figure 5 that the displacements at most monitoring points during the water filling period were more significant than the period with stable water level, which verified the necessity of considering the variation of water level in the model development.



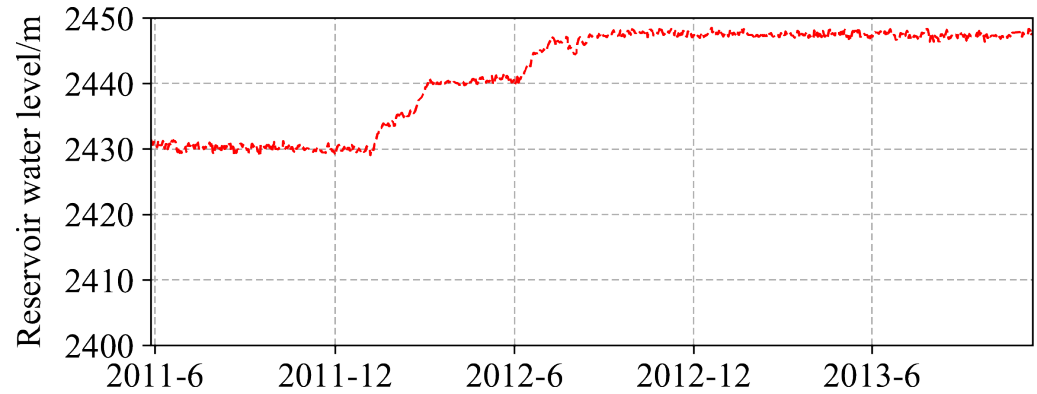


Figure 6. Time variation of the water level in the reservoir.

#### 4. Results and Discussions

##### 4.1. Prediction Using RCM-GMM Model

We first classified the monitoring points into several groups according to the characteristics of monitoring data using the Gaussian mixture model, and then modeled the displacement of monitoring points at each group using the random coefficient model. Figure 7 shows the clustering results of the monitoring points using the Gaussian mixture model. The scale of the picture is marked by the yellow line at the bottom of the figure. Mathematical details of the clustering methods have been presented in Section 2.1. The 12 monitoring points are classified into four groups, based on similarities of characteristics of data sequence: cluster 1 contains TP1-8, TP2-7, and TP2-8; cluster 2 contains TP2-9; cluster 3 contains TP3-12, TP4-2, and TP5-5; cluster 4 contains TP3-13, TP4-6, TP4-7, TP5-8, and TP5-9.

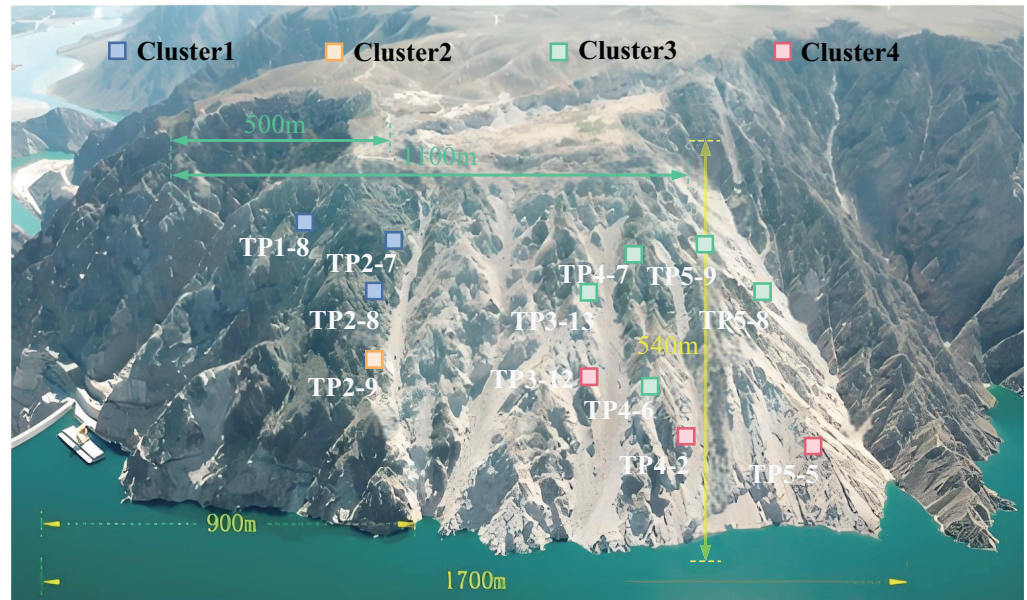
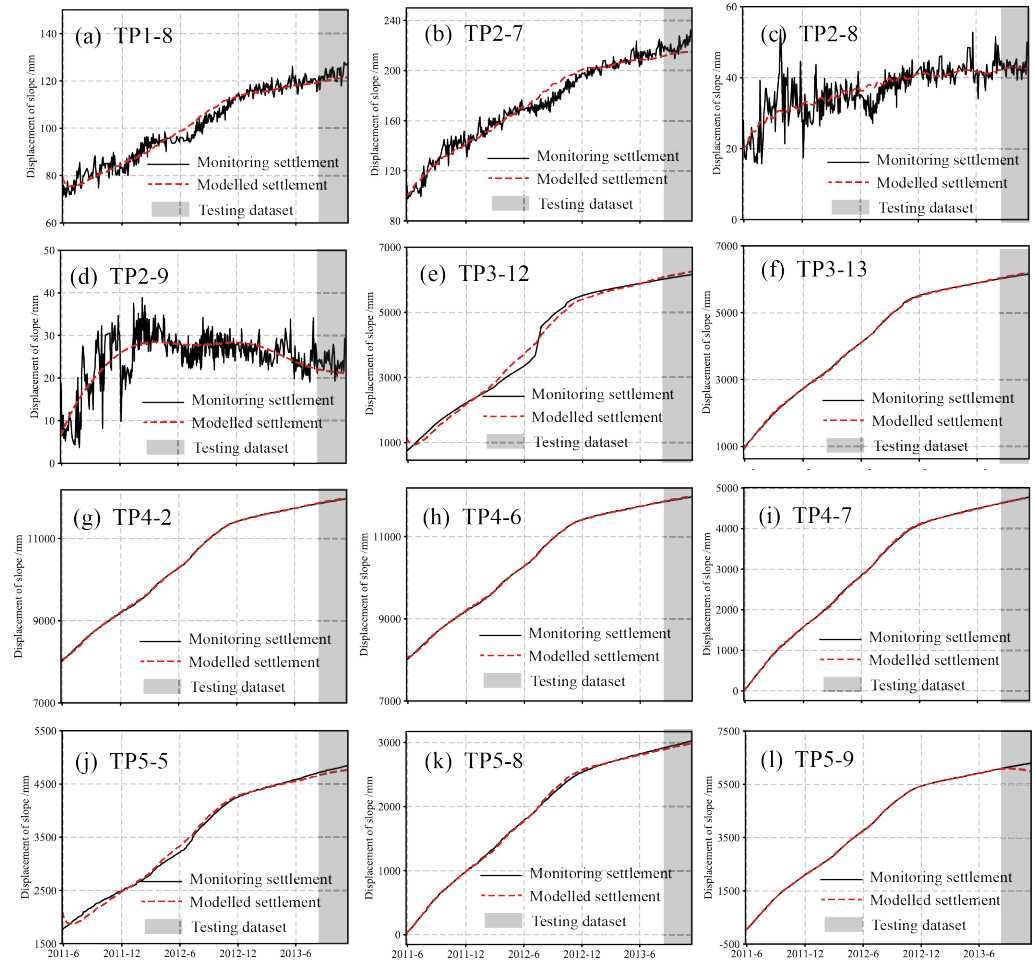


Figure 7. Clustering results of the monitoring points.

According to the clustering results estimated using the Gaussian mixture model, we then modeled the monitoring data of monitoring points in each cluster using the random coefficient model. The mathematical details of the random coefficient model were introduced in Section 2.2. In total, 90% of the dataset was selected as training data, and 10% was selected as testing data. We first developed the prediction model using the training dataset, and then validated the model using the testing dataset.

Figure 8 displays the monitored and modeled displacement of the slope for each monitoring point. The black solid line represents the monitoring data, and the red dashed

line denotes the modeled dataset. The test dataset is highlighted by the gray area in the figures. For monitoring points TP3-12, TP3-13, TP4-2, TP4-6, TP4-7, TP5-5, TP5-8, and TP5-9, the time evolution of monitoring data varies smoothly, and the modeled data fits quite well with the monitoring data. For monitoring points TP1-8, TP2-7, TP2-8 and TP2-9, the time evolution of monitoring data has significant fluctuations, the modeled data generally fits the midline of monitoring data.



**Figure 8.** Comparison of monitored and modeled displacement of slope for each monitoring point: (a) TP1-8, (b) TP2-7, (c) TP2-8, (d) TP2-9, (e) TP3-12, (f) TP3-13, (g) TP4-2, (h) TP4-6, (i) TP4-7, (j) TP5-5, (k) TP5-8, (l) TP5-9.

Figure 9 shows the coefficient of determination and root mean square error (RMSE) of results modeled by RCM for each monitoring point. The blue line denotes the coefficient of determination and the green bar denotes the RMSE. The coefficient of determination represents the correlation between monitored data and predicted data of the training dataset, and reflects the prediction precision of the prediction model. The expression of the coefficient of determination  $R$  and the

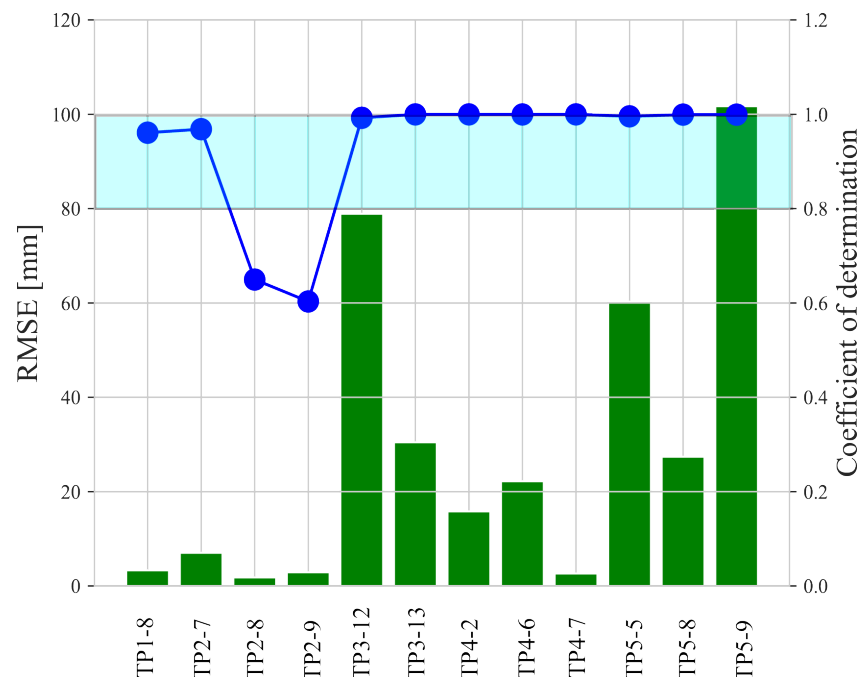
$$R = \frac{\sum_{i=1}^N (\hat{\delta}_i - \bar{\delta})}{\sum_{i=1}^N (\delta_i - \bar{\delta})} \quad (15)$$

where  $\delta_i$  denotes the monitored data of the training dataset, where  $i$  counts the series number of the training dataset,  $\bar{\delta}$  is the average of the monitored data of the training dataset,  $\hat{\delta}$  denotes the modeled data of the training dataset, and  $N$  counts the total number of data in the training dataset. It can be seen from the figure that the coefficients of determination of most monitoring points are higher than 0.9 and close to 1, which represents a fairly

good prediction accuracy. However, the coefficients of determination of monitoring points TP2-8 and TP2-9 are significant lower than those of other monitoring points. The root mean square error (RMSE) represents the derivation between the estimator and the observer of the testing dataset, whose expression can be written as:

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^M (\delta_j - \hat{\delta}_j)^2} \quad (16)$$

where  $j$  is the series number of the testing dataset,  $M$  is the total number of data in the testing dataset,  $\delta_j$  is the monitored data of the testing dataset, and  $\hat{\delta}_j$  the modeled data of the testing dataset. As shown in Figure 9, the RMSE of the monitoring points TP1-8, TP2-7, TP2-8, TP4-7, and TP-9 are lower than 10 mm. The RMSE of TP3-12, TP5-5, and TP5-9 ranges between 50 mm and 100 mm, and they are obvious larger than those of other monitoring points. The lower error at monitoring points TP2-8 and TP2-9 are attributed to a combination of measurement source, interference, quality historical data, and specific geographic or structural conditions, etc. The historical data available for these two points are more consistent comparing with other monitoring points which are influenced by structural vibrations or geographic variations, providing a better basis for accurate predictions. The displacement of slope of these three monitoring points are larger than 2000 mm, that is, the ratios of RMSE to observed dataset are still very small. For all monitoring points, the varying range of the ratio of RMSE to observed dataset is between 0.086% and 11.866%, which demonstrates the high prediction accuracy of the proposed model. Understanding the specific reasons would require a detailed analysis of the monitoring setup and the conditions at each point. This paper focus on the global performance of proposed model rather than the specific performance at each monitoring point, we hence do not enter the discussion here.



**Figure 9.** The coefficient of determination and RMSE of results modeled by RCM for each monitoring point.

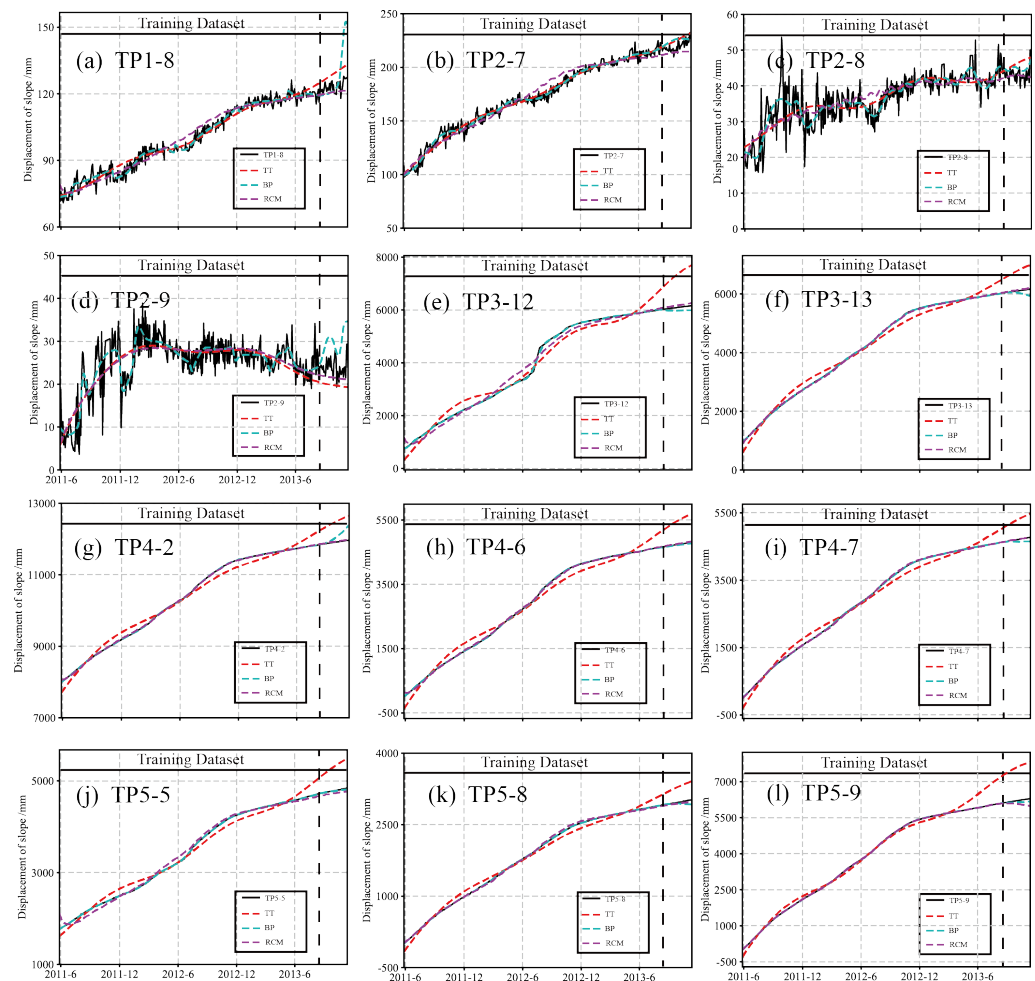
#### 4.2. Comparison of the Prediction Performance of RCM with BP and TT

In addition to the random coefficient model proposed in this paper, we also predicted the displacement of slope for all 12 monitoring points using temporal–temperature model



(TT) and BP neural network model (BP), which are one of the most common statistical model and artificial methods, respectively. The choice of the random coefficient model, BP neural network model, and temporal-temperature model balances demonstrating the added value of the proposed model, covering a methodological spectrum, ensuring domain relevance, and maintaining practical feasibility. The proposed model is developed by combining the random coefficient model and Gaussian mixture model. Comparing it with the traditional random coefficient model demonstrates the benefits of classifying the dataset using the Gaussian mixture model. BP neural network model is a well-established method in machine learning for time series prediction, and it is often used as a benchmark due to its strong performance in capturing non-linear relationships. Temporal-temperature model is the most commonly used statistical model. Hence, these two models are selected as representations of machine learning methods and statistical methods, respectively. These comparisons provide a clear and compelling case for the effectiveness and innovation of the proposed model.

Figure 10 shows the predicted displacement of slope determined by TT, BP and RCM for each monitoring point. The black solid line represents the monitored data. The red dashed line, green dashed line, and pink dashed line denote the predicted data estimated by TT, BP and RCM, respectively. The black dashed line separates the training dataset and testing dataset into two parts: data at the left side is the training dataset, and data at the right side is the testing dataset. In general, the predicted dataset of TT has a significant deviation from those of BP and RCM, and the predicted dataset of BP and RCM are fairly close.



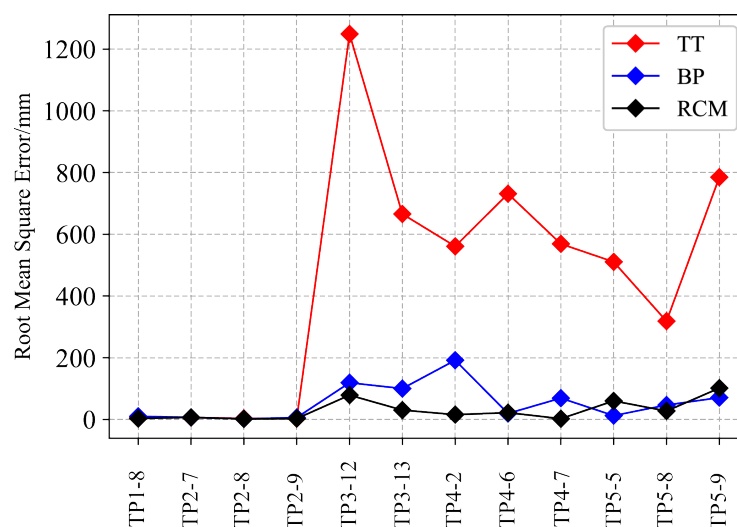
**Figure 10.** The predicted displacement of slope determined by TT, BP and RCM for each monitoring point: (a) TP1-8, (b) TP2-7, (c) TP2-8, (d) TP2-9, (e) TP3-12, (f) TP3-13, (g) TP4-2, (h) TP4-6, (i) TP4-7, (j) TP5-5, (k) TP5-8, (l) TP5-9.

To quantify the prediction precision of the temporal–temperature model (TT), BP neural network model (BP), and random coefficient model (RCM), we calculated the coefficients of determination of these three models for each monitoring point, which are exhibited in Table 2. For all three models, the coefficients of determination of most monitoring points are greater than 0.9. For monitoring points TP2-8 and TP2-9, the coefficients of determination of all these three models are relatively low, which results from the high fluctuation of the monitoring data. For monitoring points TP2-7, TP2-8, TP2-9, TP3-12, TP5-5, TP5-8, and TP5-9, BP has the largest coefficient of determination; for monitoring points TP1-8, TP3-13, TP4-2, TP4-6, and TP4-7, RCM has the largest coefficient of determination. In general, BP and RCM have better performance in prediction precision than TT.

Figure 11 shows the root mean square error of each monitoring points of the temporal–temperature model, BP neural network model and random coefficient model, respectively. The root mean square error of the temporal–temperature model is significantly higher than those of BP neural network model and random coefficient model; the two models withdrawal have similar values of root mean square error. That is, the BP neural network model and random coefficient model have similar prediction accuracies, and both have better performances than temporal–temperature model. In addition, the accuracies for each monitoring points are different. The prediction performances for TP1-8, TP2-7, TP2-8, and TP2-9 are fairly good for all these three prediction models.

**Table 2.** The coefficients of determination of TT, BP and RCM for each monitoring point.

Monitoring Point	TT	BP	RCM
TP1-8	0.9594	0.9452	0.9610
TP2-7	0.9834	0.9875	0.9684
TP2-8	0.6445	0.7926	0.6496
TP2-9	0.6032	0.6866	0.6032
TP3-12	0.9285	0.9994	0.9926
TP3-13	0.9713	0.9996	0.9998
TP4-2	0.9656	0.9976	0.9998
TP4-6	0.9618	0.9998	0.9998
TP4-7	0.9745	0.9998	0.9999
TP5-5	0.9608	0.9999	0.9957
TP5-8	0.9797	0.9997	0.9993
TP5-9	0.9722	0.9999	0.9993



**Figure 11.** Root mean square error of TT, BP and RCM for each monitoring point.

To further estimate the prediction abilities of these three models, we selected 5%, 10%, and 15% of the dataset as test data, and calculated the error indicators of the prediction

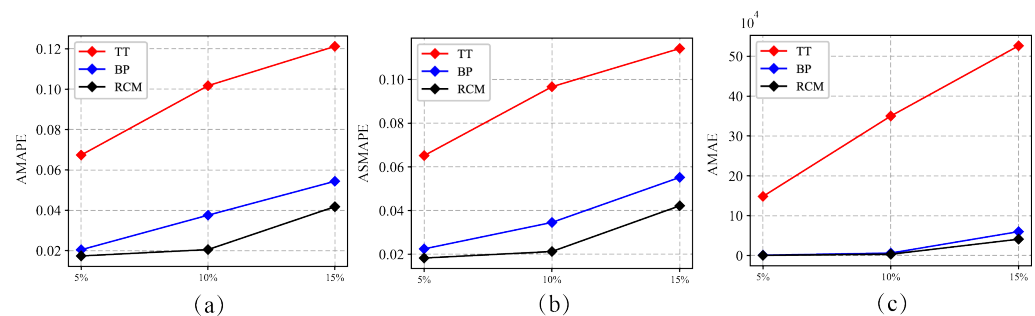
ability under different prediction length. The error indicators used here include: the average of mean square error (AMSE), the symmetric mean absolute percentage error (ASMAPE), and the mean absolute percentage error (AMAPE) of the prediction results of all monitoring points. The expressions of AMSE, ASMAPE, and AMAPE are as follows:

$$\text{AMSE} = \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M (\delta_{ij} - \hat{\delta}_{ij}) \quad (17)$$

$$\text{ASMAPE} = \frac{100\%}{N \times M} \sum_{i=1}^N \sum_{j=1}^M \frac{|\hat{\delta}_{ij} - \delta_{ij}|}{(|\hat{\delta}_{ij}| + |\delta_{ij}|)/2} \quad (18)$$

$$\text{AMAPE} = \frac{100\%}{N \times M} \sum_{i=1}^N \sum_{j=1}^M \frac{|\hat{\delta}_{ij} - \delta_{ij}|}{|\delta_{ij}|} \quad (19)$$

where  $\delta_{ij}$  is the predicted data;  $\hat{\delta}_{ij}$  is the monitored data;  $i$  and  $j$  count the monitoring point and the data sequence of each monitoring point, respectively;  $N$  and  $M$  denote the total number of monitoring points and the total quantity of data of each monitoring point, respectively. Figure 12 illustrates the AMSE, ASMAPE, and AMAPE of the above three models under the prediction length of 5%, 10%, and 15%. The varying range of AMSE, ASMAPE, and AMAPE of these three models are: 505–40995, 0.018–0.00422, and 0.0174–0.0418, respectively. It can be seen from the figure that the error indicators of all these three models increase with the increase in the prediction length, that is, the prediction accuracies decrease with the increase in the prediction length. In addition, the prediction accuracy of the random coefficient model is higher than those of BP neural network model and temporal–temperature model.



**Figure 12.** Comparison of the (a) AMSE, (b) ASMAPE, and (c) AMAPE of the three prediction models under different predicting length.

## 5. Conclusions

Predicting the displacement of slope based on monitoring data is an effective method to evaluate the risk of slope failure and landslide. The displacement at each monitoring point relates to the displacement in the surrounding area. The objective of this study was to take the correlations between monitoring data at different monitoring points into account in the prediction model.

With this objective in mind, we provide a prediction method under the frame of panel data modeling based on Gaussian mixture model and random coefficient model.

Many existing models used time series data from individual monitoring points without considering the cross-sectional dimension (i.e., different monitoring points) simultaneously. However, monitoring points within a slope may be influenced by common factors and hence are not entirely independent. Employing panel data analysis incorporates the correlation among monitoring points, which enhances the model's robustness by capturing both the temporal dynamics and spatial heterogeneity in slope deformation data. In addition, by classifying the monitoring data using a Gaussian mixture model, the proposed model effectively captures and distinguishes different temporal-spatial patterns in the

deformation data. This classification enables a more nuanced understanding and modeling of the deformation process, allowing the model to account for complex and heterogeneous behaviors in the slope.

Using the Gaussian mixture model, we first classified the monitoring data of different monitoring points into 4 clusters considering the temporal-spatial characteristics of monitoring points. With the clustering results, we established the prediction model for each cluster using the random coefficient model. We use monitoring data of Guobu slope to validate the model, which is famous slope located on the bank of Yellow river near Laxiwa hydro station and with a high risk of slope failure. 90% of the dataset are selected as training data, and 10% are used as testing data. Results show that the coefficients of determination of the proposed model are larger than 0.9 and the ratios of RMSE to recorded data is lower than 1%, indicating the high prediction performance of the proposed model. We also compared the prediction accuracy of the proposed model (RCM) with BP neural network model (BP) and temporal-temperature model (TT), which are one of the most commonly used artificial model and statistical model, respectively. Results showed that the prediction accuracy of RCM is slightly higher than that of BP, and significantly better than that of TT.

**Author Contributions:** Conceptualization, Z.Y. and Y.B.; methodology, Z.Y., S.Z. and Z.M.; validation, Y.B., W.L. and F.Q.; formal analysis, Z.Y. and H.M.; writing—original draft preparation, Z.Y.; writing—review and editing, Z.Y. and Z.M.; project administration, Y.B.; funding acquisition, Y.B. and Z.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant No. LZJWY24E090005.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data available upon request.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Meng, Z. Experimental study on impulse waves generated by a viscoplastic material at laboratory scale. *Landslides* **2018**, *15*, 1173–1182. [[CrossRef](#)]
- Wasowski, J.; Bovenga, F. Investigating landslides and unstable slopes with satellite Multi Temporal Interferometry: Current issues and future perspectives. *Eng. Geol.* **2014**, *174*, 103–138. [[CrossRef](#)]
- Liu, X.; Liu, Y.; Yang, Z.; Li, X. A novel dimension reduction-based metamodel approach for efficient slope reliability analysis considering soil spatial variability. *Comput. Geotech.* **2024**, *172*, 106423. [[CrossRef](#)]
- Raucoules, D.; De Michele, M.; Malet, J.P.; Ulrich, P. Time-variable 3D ground displacements from high-resolution synthetic aperture radar (SAR). Application to La Valette landslide (South French Alps). *Remote Sens. Environ.* **2013**, *139*, 198–204. [[CrossRef](#)]
- Schuster, R. Reservoir-induced landslides. *Bull. Eng. Geol. Environ.* **1979**, *20*, 8. [[CrossRef](#)]
- Chen, H.; Huang, S.; Xu, Y.P.; Teegavarapu, R.S.; Guo, Y.; Nie, H.; Xie, H. Using baseflow ensembles for hydrologic hysteresis characterization in humid basins of Southeastern China. *Water Resour. Res.* **2024**, *60*, e2023WR036195. [[CrossRef](#)]
- Perissin, D.; Prati, C.; Rocca, F.; Wang, T. PSInSAR analysis over the Three Gorges Dam and urban areas in China. In Proceedings of the 2009 Joint Urban Remote Sensing Event, Shanghai, China, 20–22 May 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–5.
- Shi, X.; Zhang, L.; Tang, M.; Li, M.; Liao, M. Investigating a reservoir bank slope displacement history with multi-frequency satellite SAR data. *Landslides* **2017**, *14*, 1961–1973. [[CrossRef](#)]
- Shi, X.; Jiang, H.; Zhang, L.; Liao, M. Landslide displacement monitoring with split-bandwidth interferometry: A case study of the shuping landslide in the three gorges area. *Remote Sens.* **2017**, *9*, 937. [[CrossRef](#)]
- Tang, H.; Wasowski, J.; Juang, C.H. Geohazards in the three Gorges Reservoir Area, China—Lessons learned from decades of research. *Eng. Geol.* **2019**, *261*, 105267. [[CrossRef](#)]
- Zhou, C.; Cao, Y.; Yin, K.; Wang, Y.; Shi, X.; Catani, F.; Ahmed, B. Landslide characterization applying sentinel-1 images and InSAR technique: The muyubao landslide in the three Gorges Reservoir Area, China. *Remote Sens.* **2020**, *12*, 3385. [[CrossRef](#)]
- Meng, Z.; Ancey, C. The effects of slide cohesion on impulse-wave formation. *Exp. Fluids* **2019**, *60*, 151. [[CrossRef](#)]
- Mohammed, F.; Fritz, H.M. Physical modeling of tsunamis generated by three-dimensional deformable granular landslides. *J. Geophys. Res. Ocean.* **2012**, *117*, C11015. [[CrossRef](#)]

14. Meng, Z.; Hu, J.; Zhang, J.; Zhang, L.; Yuan, Z. The Momentum Transfer Mechanism of a Landslide Intruding a Body of Water. *Sustainability* **2023**, *15*, 13940. [[CrossRef](#)]
15. Ciabatti, M. *La Dinamica Della Frana del Vajont*; Museo geologico Giovanni Capellini: Bologna, Italy, 1964.
16. Genevois, R.; Ghirotti, M. The 1963 vaiont landslide. *G. Geol. Appl.* **2005**, *1*, 41–52.
17. Gylfadóttir, S.S.; Kim, J.; Helgason, J.K.; Brynjólfsson, S.; Höskuldsson, Á.; Jóhannesson, T.; Harbitz, C.B.; Løvholt, F. The 2014 Lake Askja rockslide-induced tsunami: Optimization of numerical tsunami model using observed data. *J. Geophys. Res. Ocean.* **2017**, *122*, 4110–4122. [[CrossRef](#)]
18. Gauthier, D.; Anderson, S.A.; Fritz, H.M.; Giachetti, T. Karrat Fjord (Greenland) tsunamigenic landslide of 17 June 2017: Initial 3D observations. *Landslides* **2018**, *15*, 327–332. [[CrossRef](#)]
19. Pariseau, W.G.; Puri, S.; Schmelter, S.C. A new model for effects of impersistent joint sets on rock slope stability. *Int. J. Rock Mech. Min. Sci.* **2008**, *45*, 122–131. [[CrossRef](#)]
20. Griffiths, D.; Marquez, R. Three-dimensional slope stability analysis by elasto-plastic finite elements. *Geotechnique* **2007**, *57*, 537–546. [[CrossRef](#)]
21. Bui, H.H.; Fukagawa, R.; Sako, K.; Wells, J.C. Slope stability analysis and discontinuous slope failure simulation by elasto-plastic smoothed particle hydrodynamics (SPH). *Géotechnique* **2011**, *61*, 565–574. [[CrossRef](#)]
22. Cundall, P.; Damjanac, B. A comprehensive 3D model for rock slopes based on micromechanics. In Proceedings of the Slope Stability 2009, Santiago, Chile, 9–11 November 2009.
23. Gurocak, Z.; Alemdag, S.; Zaman, M.M. Rock slope stability and excavatability assessment of rocks at the Kapikaya dam site, Turkey. *Eng. Geol.* **2008**, *96*, 17–27. [[CrossRef](#)]
24. Qi, S.; Wu, F.; Yan, F.; Lan, H. Mechanism of deep cracks in the left bank slope of Jinping first stage hydropower station. *Eng. Geol.* **2004**, *73*, 129–144. [[CrossRef](#)]
25. Esposito, C.; Martino, S.; Mugnozza, G.S. Mountain slope deformations along thrust fronts in jointed limestone: An equivalent continuum modelling approach. *Geomorphology* **2007**, *90*, 55–72. [[CrossRef](#)]
26. Guallini, L.; Brozzetti, F.; Marinangeli, L. Large-scale deformational systems in the South Polar Layered Deposits (Promethei Lingula, Mars): “soft-sediment” and deep-seated gravitational slope deformations mechanisms. *Icarus* **2012**, *220*, 821–843. [[CrossRef](#)]
27. Meng, Z.; Wang, Y.; Zheng, S.; Wang, X.; Liu, D.; Zhang, J.; Shao, Y. Abnormal Monitoring Data Detection Based on Matrix Manipulation and the Cuckoo Search Algorithm. *Mathematics* **2024**, *12*, 1345. [[CrossRef](#)]
28. Tang, X.S.; Li, D.Q.; Chen, Y.F.; Zhou, C.B.; Zhang, L.M. Improved knowledge-based clustered partitioning approach and its application to slope reliability analysis. *Comput. Geotech.* **2012**, *45*, 34–43. [[CrossRef](#)]
29. Meng, Z.; Hu, Y.; Ancey, C. Using a data driven approach to predict waves generated by gravity driven mass flows. *Water*. **2020**, *12*, 600. [[CrossRef](#)]
30. Ren, F.; Wu, X.; Zhang, K.; Niu, R. Application of wavelet analysis and a particle swarm-optimized support vector machine to predict the displacement of the Shuping landslide in the Three Gorges, China. *Environ. Earth Sci.* **2015**, *73*, 4791–4804. [[CrossRef](#)]
31. Liu, P.; Li, Z.; Hoey, T.; Kincal, C.; Zhang, J.; Zeng, Q.; Muller, J.P. Using advanced InSAR time series techniques to monitor landslide movements in Badong of the Three Gorges region, China. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 253–264. [[CrossRef](#)]
32. Sezer, E. A computer program for fractal dimension (FRACEK) with application on type of mass movement characterization. *Comput. Geosci.* **2010**, *36*, 391–396. [[CrossRef](#)]
33. Li, X.; Kong, J. Application of GA-SVM method with parameter optimization for landslide development prediction. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 525–533. [[CrossRef](#)]
34. Yiqing, L.; Hongfu, L. Prediction of landslide displacement using grey and artificial neural network theories. *Adv. Sci. Lett.* **2012**, *11*, 511–514. [[CrossRef](#)]
35. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. *Eng. Geol.* **2016**, *204*, 108–120. [[CrossRef](#)]
36. Zhang, D.; Wang, G.; Yang, T.; Zhang, M.; Chen, S.; Zhang, F. Satellite remote sensing-based detection of the deformation of a reservoir bank slope in Laxiwa Hydropower Station, China. *Landslides* **2013**, *10*, 231–238. [[CrossRef](#)]
37. Lin, P.; Liu, X.; Hu, S.; Li, P. Large deformation analysis of a high steep slope relating to the Laxiwa Reservoir, China. *Rock Mech. Rock Eng.* **2016**, *49*, 2253–2276. [[CrossRef](#)]
38. Li, M.; Zhang, L.; Shi, X.; Liao, M.; Yang, M. Monitoring active motion of the Guobu landslide near the Laxiwa Hydropower Station in China by time-series point-like targets offset tracking. *Remote Sens. Environ.* **2019**, *221*, 80–93. [[CrossRef](#)]
39. Pang, L.; Li, C.; Liu, D.; Zhang, F.; Chen, B. Response of Guobu Slope Displacement to Rainfall and Reservoir Water Level with Time-Series InSAR and Wavelet Analysis. *Appl. Sci.* **2023**, *13*, 5141. [[CrossRef](#)]
40. Semenza, E.; Ghirotti, M. History of the 1963 Vaiont slide: The importance of geological factors. *Bull. Eng. Geol. Environ.* **2000**, *59*, 87–97. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.