 CMES, vol.84, no.2, pp.99-122, 2012

Gauss Process Based Approach for Application on Landslide Displacement Analysis and Prediction

Zaobao Liu[1,2], Weiya Xu[1] and Jianfu Shao[2]

Abstract: In this paper, the Gauss process is proposed for application on landslide displacement analysis and prediction with dynamic crossing validation. The prediction problem using noisy observations is first introduced. Then the Gauss process method is proposed for modeling non-stationary series of landslide displacements based on its ability to model noisy data. The monitoring displacement series of the New Wolong Temple Landslide is comparatively studied with other methods as an instance to implement the strategy of the Gauss process for predicting landslide displacement. The dynamic crossing validation method is adopted to manage the displacement series so as to give more precise predictions. Different covariance functions are illustrated to give predictive results which show that different covariance functions result in varying levels of prediction accuracy. Comparisons with other methods are also discussed in this study. The results show that the Gauss process can perform better than the RBF network and the SVM methods in this problem in view of the trends according to the original data. Finally, the landslide criterion is given for creep-typed slopes that landslide event would occur imminently if the cross angle at the intersection point of displacement curve changes more than 45°.

Keywords: landslide prediction; time series modeling; Gauss process; landslide criteria

[1] Geotechnical Research Institute of Hohai University, 210098 Nanjing China. E-mail: forward@hhu.edu.cn; wyxu@hhu.edu.cn. Partially supported by the project sponsored by National Program on Key Basic Research Project (973 Program) (Grant No.2011CB013504), Postgraduate Research Innovation Plan of Jiangsu Higher Education of China (Grant No.CX09B_158Z).

[2] La Laboratoire de Mecanique de Lille, Genie Civil, Universite de Science et Technologie de Lille, 59655 Villeneuve d'Ascq France. E-mail: zaobao.liu@polytech-lille.fr; jianfu.shao@univ-lille1.fr.

1 Introduction

1.1 Research background and related work

Time series are commonly used in many fields to reveal the characteristics of complicated systems. The observation of landslide displacement is typical time series data which represents the overall features of landslides. Taking away the complicated inducing factors, landslide displacement is the comprehensive external outcome of the underlying dynamic evolutionary process for potential landslides. The displacement has been used as one of the main means to indicate the status of slopes. Study on landslide displacement has been recognized as an effective way to know the potential for a landslide event for a long time. Researchers have been studying proper prediction models for monitoring series of landslide displacement in the past decades. Landslide displacement prediction is the fundamental work for the prevention of landslide disasters despite the fact that it is far from being able to completely forecast them. Studying landslide displacements is an effective tool for better understanding landslide movements. Models or methods have been proposed with proper criteria on the issue of modeling specific landslide displacement series, mainly including the grey forecasting model, the neural network and the support vector machine, etc.

For the grey forecasting model, Zhou and Hu(2008) presented an effective hybrid approach for forecasting gyro drift based on grey theory and ARMA model. Using the view of treating landslide as a grey system, Liu, Xu, Meng and Chen (2009) conducted a study on landslide displacement prediction with a modified GM(1.1) model for unequal interval observation series. While, the predictive accuracy of the GM model depends largely on the number of samples for modeling; the exact sample size was difficult to reasonably identify.

Many researchers have reported the outperformance of different ANNs for study of slope displacement and movement. Feng, Wang and Yao (1996) stated a real-time prediction model for roof pressure in coal mines using a multilayer feed-forward neural network and achieved a satisfactory level of accuracy. Sakellariou and Ferentinou (2005) promoted the study on the estimation of slope stability using neural networks based on a collective data set of historical slopes worldwide and they also studied the relative importance of the parameters affecting slope stability. Wang and Xu (2005) presented the Back Propagation Neural Networks (BPNN) with five input nodes, two hidden layers, and two output nodes to evaluate slope instability by using a training data set of landslide samples throughout regional observations. Ferentinou and Sakellariou(2007) presented a study on the prediction of slope performance obtained by using the back-propagation algorithm, the theory of Bayesian neural networks and the Kohonen self-organizing maps. The results indicated that

this method was promising and should be further explored.

Feng, Zhang, and Xu (1999, 2004); Matías, Taboada and Ordótez (2010); XU F. and XU W. (2010); Samui and Kothari (2011) investigated the implementation of support vector machines for modeling the displacement and movement of slopes to give satisfactory predictive results. Feng, Zhang, and Xu (1999, 2004) presented the support vector machine (SVM) to obtain a global optimization model for the evaluation of the non-linear displacement behavior of geo-materials under the conditions of large project dimensions, small sample sizes and nonlinearity. They found that the SVM can appropriately describe the evolutionary law of the deformation of geo-materials at depth and provide predictions for the future 6-10 time steps with acceptable accuracy and confidence. Matías, Taboada and Ordótez (2010) proposed the PLSVM method with the kernel composed of a linear kernel and a nonlinear kernel. They found that the PLSVM had improved results over the other autoregressive approaches for predicting the monthly movement of a mine slope. This had significant impact on the safety of mining operations. XU F. and XU W. (2010) conducted the prediction of slope displacement series using a hybridization of the SVMs and Markov chains and found that the integrated model provided consistently accurate predictive results. Samui and Kothari (2011) examined the capability of the LSSVM model for slope stability analysis and they also carried out a comparative study between the LSSVM and ANN. Their study concluded that the developed LSSVM is a robust model for slope stability analysis. Various models were presented by former researchers for time series prediction.

Some of the models mentioned above gave the estimation of slope stability according to the collection of historical data on slope cases, however precise data acquisition is quite a difficult task for some of the parameters of the slopes, especially the descriptive data. Thus some others presented models for prediction of landslide displacement based on observations, which is another way to analyze slopes and can be called the phenomenal way. However, these methods are not sufficient to prevent the searching of new methods to study landslide displacement since they all have some restrictions. For example, the performance of grey model is closely related to the amount of data that is modeled. And no consensus has been made on the best quantity of data for modeling. ANNs are thought to be powerful for their ability to deal with nonlinear problems while the parameters and results of ANNs are sometimes difficult to be physically explained. The SVM is proposed for machine leaning on the principle of minimizing the empirical risk. Its performance is greatly affected by the types of kernel function and other parameters. Mehdi and Mehdi (2011) reviewed broadly the literature on the time series prediction when they proposed a hybridization of neural networks and ARIMA models for time series forecasting. They concluded that improving forecasting accuracy, especially

time series forecasting, was an important yet often difficult task facing decision makers in many areas. The research for improving the effectiveness of forecasting models would never stop despite the numerous time series models available. Thus work is still needed to present a prediction method which can take into account of the previous knowledge of the landslide system.

In Neal's work (1996) on Bayesian learning for neural network analysis inspired a new regression method. The Gauss process has proved to be an attractive method for modeling noisy data based on prior over functions. Problems with noise-free and stationary series have been studied using the Gauss Process [Williams and Rasmussen (1996); Brahim-Belhouari and Vesin (2001); Brahim-Belhouari and Bermak (2004)]. In this study modeling the noisy and non-stationary landslide displacement series was discussed.

The Gauss process is based on Bayesian leaning. The advantages lie in that it utilizes not only the model information and data message, but also makes the best of prior knowledge about the studied object. What's more, the prior knowledge can be free of special restrictions. The prior distribution need not be objective as it can partially or completely depend on the subjectivities. Thus it gains superiority in discussing strong empirical systems like landslides. And this has increased its development and applications in many fields [Li and Sun (2009); Tarek and Nizar (2011)].

Conclusions can be made from the literature that a predictive model may be able to generate satisfactory results for one set of training samples, but that it might not be able to outperform for other input data with different behaviors. The clue may lie in that some models are linear, while others are nonlinear. Linear models can perform well for linear systems and nonlinear ones can generate quite good results for nonlinear systems, however linear models could not adequately account for nonlinear systems and vice-versa. Landslide displacement is complicated with typical nonlinear features. Despite all of the achievements, landslide displacement modeling is still needed to be further studied since it deals with natural data, which concerns with not only calculations and numerical analysis, but also concepts, perception, judgment and employment of experience that cannot be strictly represented numerically.

1.2 Content of this study

This study first introduced slope prediction problem with noisy observations and then presented the explanation of Gauss process including the definition, model selection, model training and the decision making. In Section 4, the observed displacement data of Wolong Temple landslide is illustrated to apply the approach introduced in this study. Comparisons on different models are also discussed in

Section 4 as well as the discussion on the landslide criteria. Finally conclusions are made based on the work.

2 Prediction problems of slope displacement

A slope is a nonlinear open system coupling affected by things such as the geo-environment of the rock mass, the hydrological condition, underground water, rainfall, manual activities, and even the seismic activities. It is extremely difficult to uncover the complicated evolutionary process for slopes in complex conditions. The displacement behavior of slopes is aggravated by its material structures, reinforcements, excavation blasting, human activities, tectonic activities, seismic forces, high stresses, high water pressure, temperature gradient, strong geo-chemical reaction and their coupled effects [Feng, Zhao, Li (2004)].The measured displacement series is noisy, non-stationary and variable over time. Modeling the measured displacement is important and feasible for analysis and prediction of the slope stability status. The prediction of slope displacement is aimed at estimating and predicting its future displacement in magnitude and tendency based on the historical displacement time series.

Once a certain model is proved to be suitable for prediction of landslide displacement, then the forecasting of a landslide occurrence could be considered within proper thresholds and the criteria for potential slope instability. The predictive model is processed as the main issue for the prediction problem of slope displacement. Regard a set of nodes as d such that the observed target is denoted as $X = \left\{x_{k-d+1}, x_{k-(d-1)+1}, \cdots, x_k\right\}$ and the next $x_{\mathrm{k}+1}$ is the prediction target or the modeling output. Since the measuring conditions are usually influenced by multitudinous complicated factors, let ε, y^i, $f(x,\theta)$ be the overall effect of noise corrupting the data, the observation and the predictive distribution of slope displacement, respectively. Then the aim of prediction is to obtain the mapping of observations using observed input data $x = [x_{k-d+1}, x_{k-(d-1)+1}, \cdots, x_k]$. Thus the prediction problem for the slope displacement can be denoted as

$$y^i = f(x^i,\theta) + \varepsilon^i \quad i = 1, \cdots, n. \tag{1}$$

The goal of prediction is to obtain the non-linear mapping $f(x,\theta)$, where θ denotes the unknown parameters to be determined by training with the input data using proper techniques. In fact, as an important application field of neural networks, the mapping $f(x,\theta)$ can be given by a specified network. The outcome of the RBF network is computed as a linear superposition [Bishop (1995)] as

$$f(x,\theta) = \sum_{j=1}^{M} w_j g_j(x) \tag{2}$$

where w_j and $g_j(x)$ are the weights of the output layer and the Gaussian basis functions, respectively. And $g_j(x)$ are defined as

$$g_j(x)=\exp\left\{-\frac{\|x-\mu_j\|^2}{2\sigma_j^2}\right\} \tag{3}$$

where μ_j and σ_j denote means and variances respectively.

Thus the parameters θ can be defined as $\theta=\left[w_j,\mu_j,\sigma_j^2\right]$, $(j=1,\cdots,M)$, which would be estimated by a special training algorithm like the Back Propagation Algorithm.

Unlike the neural network method, predictions by non-parameter methods, for example the support vector machine [XU F. and XU W. (2010)], are gained without representing the unknown slope system as an explicit parameterized mapping. A new method for regression was inspired by Bayesian learning and an attractive method for modeling non-stationary noisy landslide displacement data is proposed hereafter based on prior over function. And the Gauss Process is applied with the proper prior covariance and the dynamic crossing validation method.

3 Methodology: Gauss process based approach

The Gauss process represents the posterior distribution over functions based on training data and prior distribution. The graphical model of the Gauss process is shown in Fig.1 to give a visualized view, where the squares denote the known variables and circles the unknown ones.

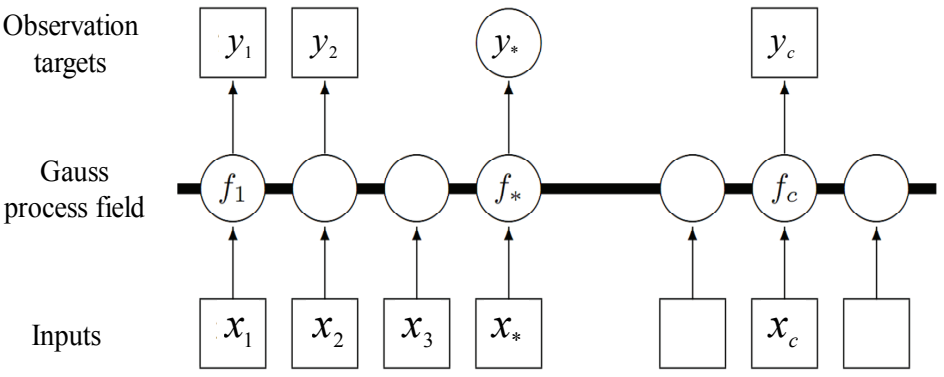

Figure 1: Graphical description of Gauss Process

3.1 Basic equations

A Gaussian process $f(x)$ is a collection of random variables, any finite set of which have a joint Gaussian distribution (Rasmussen and Williams 2006). Its statistical characteristics are completely specified by its mean function $m(x)$ and covariance function $k(x,x')$, i.e.

$$\begin{aligned} &f(x) \sim GP(m(x),k(x,x')) \\ &m(x) = \mathrm{E}[f(x)] \\ &k(x,x') = \mathrm{E}[(f(x)-m(x))-(f(x')-m(x'))] \end{aligned} \tag{4}$$

Given the observations $D = \left\{x^{(i)},y^{(i)}|i = 1,2,\cdots,n\right\}$ and the prediction input x_* (also named test input), the goal of Gauss process modeling is to obtain the output y_* for the distribution $P(y_*|D,x_*)$. Suppose the prior distribution of observation target y satisfies $y \sim N(0,k(x,x'))$ and the independent noise ε obeys $\varepsilon \sim N(0,\sigma_n^2)$, thus the covariance of noisy observations is obtained

$$\begin{aligned} &cov(y_p,y_q) = k(x_p,x_q) + \sigma_n^2\delta_{pq} \text{ or} \\ &cov(y) = K(X,X\pounds + \sigma_n^2 I \end{aligned} \tag{5}$$

where $K(X,X)$ is a positive definite covariance matrix with the size $n \times n$ and its elements denote the correlations of different observation samples. Consequently the joint distribution of the observed targets and the predictions can be signified as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X)+\sigma_n^2 & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{6}$$

For notation simplicity, if $K = K(X,X), K_* = K(X,X_*)$, the regression equation of the Gauss process for noisy observed target is obtained

$$f_*|X,y,X_* \sim N(\bar{f}_*, cov(f_*)) \tag{7}$$

$$\bar{f}_* \triangleq \mathrm{E}[f_*|X,y,X_*] = K_*^{\mathrm{T}}[K+\sigma_n^2 I]^{-1}y \tag{8}$$

$$cov(f_*) = K(X_*,X_*) - K_*^{\mathrm{T}}[K+\sigma_n^2 I]^{-1}K_*. \tag{9}$$

The output of regression eq. (7) is not a single value but a probability distribution of predictions. This advantage can be used to obtain the prediction intervals that describe a degree of confidence in the predictions.

3.2 Covariance functions

Covariance functions specify the relationships between the input data and output elements. Note that although eq.(7) gives the predictive distribution of the observed target, the lateral prior function $K(X,X)$ need to be specified in advance. Although there are many possible choices of prior covariance functions, an arbitrary function of input pairs (X,X_*) will not, in general, be a valid covariance function. The covariance function is a crucial element for Gaussian process regression, as it encodes pre-assumptions about the lateral function which we wish to learn and it defines the similarity between the inputs and test point with the predictive output. But from the modeling point of view, the goal is to specify prior covariance functions that contain our prior beliefs on the structure of the lateral function we are modeling. Formally, it aims to specify a function which will generate a positive definite covariance matrix for any set of input data and represent the relationships between the input data and the output predictions.

One covariance function used universally has the square exponential term

$$K_{\text{se}}(x_p,x_q)=\sigma_f^2\exp\left(-\frac{1}{2}(x_p-x_q)^T M(x_p-x_q)\right)+\sigma_n^2\delta_{pq} \tag{10}$$

where σ_f^2, σ_n^2 and M denote the observed target variance, the noise variance and the length scale, respectively. And $\theta=(\sigma_f^2,\sigma_n^2,\{M\})^T$ are the hyper-parameters to be adjusted by model training. The properties of the covariance function depend on the values of the hyper-parameters. This covariance function expresses the idea that nearby inputs will have highly correlated outputs. One extreme case is x_p-x_q=0. The simplest non-stationary covariance function is the one corresponding to a linear trend so that $K(x_p,x_q)\neq K(|x_p-x_q|)$, which is

$$K_{\text{ns}}(x_p,x_q)=v_0+v_1\sum_{l=1}^{d}x_l^p x_l^q \tag{11}$$

It has been proved that the addition and multiplication of simple covariance functions are powerful in constructing various covariance functions [Brahim-Belhouari and Bermak(2004)]. There are a variety of other covariance functions such as the Matérn class, the polynomial, and the rational quadratic [Rasmussen and Williams (2006)].

For landslide observation modeling on non-stationary displacement series, the hyper-parameters of prior functions is denoted as $\theta=(\sigma_f^2,\sigma_n^2,\{M\},v_1,v_2,\cdots)^T$, which will be specified by model training.

3.3 Model training

Model training is aimed at obtaining the values of the hyper-parameters in eq.(7), or more precisely in the prior distributions, based on the observed data of the landslide displacement series. Noting the predictive distribution of the Gauss process regression, the value of the hyper-parameters can be achieved in a maximum likelihood framework by adjusting the hyper-parameters so as to maximize the log likelihood of hyper-parameters. The partial derivatives of the log marginal likelihood with respect to each hyper parameter can be obtained (See Appendix)

$$\frac{\partial}{\partial \theta_j} \log p(y|X) = \frac{1}{2} \mathrm{y}^T K_y^{-1} \frac{\partial K_y}{\partial \theta_j} K_y^{-1} \mathrm{y} - \frac{1}{2} tr\left(K_y^{-1} \frac{\partial K_y}{\partial \theta_j} \right) = \frac{1}{2} tr\left(\left(K_y^{-1} y (K_y^{-1} y)^T - K_y^{-1} \right) \frac{\partial K}{\partial \theta_j} \right) \quad (12)$$

To adjust the values of the hyper-parameters, the initial values are first given randomly in a regular scope within the hyper-parameters space. After this training takes place on the input data with an iteration method, such as the conjugate gradient [Steihaug(1983)]or particle swarm optimization algorithm [Chau(2006)], to search for the optimal values of the hyper-parameters.

3.4 Loss function

It can be perceived from the previous section that, given the observed data set and test point, Gauss process regression implements model training and searches for the optimal hyper-parameters in the theoretical frame of maximum marginalization using the prior distributions. After this it computes the predictive distribution of the observation target with the optimized hyper-parameters by eq. (7). However in practical applications, the decisions must be made about how to act, for instance, a point-like prediction can be optimal in some sense. To this end, a loss function $L(y_{true}, y_{guess})$ is needed to specify the loss incurred by guessing the true value y_{true} with y_{guess}. For example, the loss function could be an absolute deviation or a relative deviation between the guess value and true value.

The predictive goal is to obtain the point prediction value y_{guess}. It is impossible to estimate directly what the state with the minimum decision loss is for the reason that the true value y_{true} was previously unknown. Loosely speaking, the loss function can be defined by the expected loss and optimized by minimization of the expected loss function [Rasmussen and Williams (2006)], i.e.

$$\begin{aligned} \tilde{R}(y_{guess}|X_*) &= \int L(y_*, y_{guess}) p(y_*|X_*, D) dy_* \\ y_{optimal}|X_* &= \arg\min \tilde{R}(y_{guess}|X_*) \end{aligned} \quad (13)$$

Those decision loss functions commonly used are absolute error loss function ($AEL = |y_{guess} - y_*|$) and square error loss function ($SEL = (y_{guess} - y_*)^2$), also with their variant versions. It has been proved in statistics that the estimation obtained by minimizing absolute loss AEL is the median of probability $p(y_*|X_*, D)$ and that obtained by minimizing the square error loss is the mean value of $p(y_*|X_*, D)$. And in this study, the variant versions of loss functions were also used for point predictions.

Attention should be paid that Gauss process regression derives the predictive distribution without any reference to the loss function. It just depends on the prior and the marginalization over the functions with data input of the observation target. This is the fundamental difference between the Bayesian based methods and the non-Bayesian paradigms. In non-Bayesian methods, model training is typically implemented by minimizing the empirical loss, for example, the support vector machine [Chau (2006)]. In contrast, there is a clear separation in the Bayesian method between the loss and likelihood function. The likelihood function expresses how the noisy measurements are assumed to deviate from the underlying noise free function. In contrast, the loss function captures the consequences of making a speci?c choice of guessing value, given an actual true state. The likelihood and loss function need not have anything in common [Barber and Saad(1996)].

4 Using Gauss process to model landslide displacement series

Discussion about analysis and prediction of the New Landslide of the Wolong Temple were presented in this section to illustrate the attractiveness of the Gauss process in modeling the non-stationary displacement series of landslides. Some extra data managing skills was also proved to be effective in the promotion of predictive accuracy and tendency.

4.1 Observations of Wolong Temple landslide

The New Landslide in the Wolong Temple occurred in a loess tableland. Tear cracks were noticed from the beginning of 1971 and it was monitored since the 11^{th} of March by a pile driven into the earth. The landslide occurred in the early morning of May 5^{th} with severe destruction because of a sliding movement. The dataset listed in Tab. 1 is the observed displacements of the cracks labeled No.5, which has been recognized as the key monitoring point to indicate the stability status of the slope. As can be drawn from the column of Dis. in Tab.1, the displacements of the landslide developed slowly initially but increased dramatically by the end of the observed displacement series. The characteristics of the displacement series were studied with different methods. YUAN, XU and GUO (2005) presented the Negative Selection Algorithm for identifying the mutation point of

the displacement curve and found that the jump spot occurred at the 49th point of the displacement series. DONG, FU and LENG (2007) presented a model using Takens theory for predicting the landslide displacement based on support vector machines (SVM) and they concluded that RBF kernel function had the priority in promotion of generalization accuracy combined with SVM. Now our purpose was modeling the observations using the Gauss process and the dynamic cross validation to give more accurate predictions.

Table 1: Observed displacements of the crack labeled No.5

Day/d	Disp./mm	Day/d	Disp./mm	Day/d	Disp./mm	Day/d	Disp./mm
15	1.0	28	8.2	41	12.0	54	23.0
16	1.5	29	8.4	42	13.0	55	24.0
17	1.7	30	8.7	43	13.4	56	25.2
18	2.5	31	9.0	44	14.0	57	26.0
19	3.2	32	9.2	45	15.0	58	27.0
20	4.0	33	9.4	46	16.1	59	28.2
21	4.4	34	10.0	47	16.4	60	30.0
22	5.1	35	10.1	48	17.2	61	31.0
23	5.9	36	10.3	49	17.6	62	32.0
24	6.3	37	10.4	50	18.2	63	33.0
25	7.0	38	10.5	51	19.0	64	42.0
26	7.3	39	10.8	52	19.2	65	47.0
27	7.8	40	11.1	53	20.0	66	61.0

4.2 Dynamic crossing validation

The dynamic crossing validation was implemented to promote the predictive performance of Gauss process regression. The observed displacement series were divided into two independent datasets: the training set and the test set. The predictive performance was checked to represent the generalization error for the derived regression model. The procedure of the cross validation was implemented as follows:

(a)Supposing the observed dataset be $S = (s_{,s_2,\cdots s_n)}$, divide S into a series of subsets $S_{iTr} = (s_{,s_{i+1},\cdots s_{i+d})}$, $i = 1,2,\cdots,(n-d)$, with the same size d; then $(n-d)$ subsets would be generated;

(b) Let m be the size of test sets, thus the test sets could be denoted as $S_{iTe} = s_{i+d+1}, s_{i+d+2}, \cdots s_{i+d+m}$;

(c) Generate models on the dataset S using the Gauss process regression method and implement the training process with training set to adjust and optimize the

hyper-parameters; and then with the test inputs output the results of the test sets $S^p_{iTe} = s^p_{i+d+1}, s^p_{i+d+2}, \cdots, s^p_{i+d+m}$;
(d) Repeat the step (a) ∼(c) for $(n-d)$ rounds, then the predictive output for the test sets would be $S_{predict} = (s^p_{d+1}, s^p_{d+2} \cdots £ s^p_n)$.

Loosely the parameter d can be recognized as the length of the training set, and m that of the test set. Generally speaking, for effectiveness of the regression model, the value of d would be no less than 12 so as to keep the generated model eligible for catching patterns underlying the observed datasets. Meanwhile, m should not be too large for reduction of the generalization error.

4.3 Modeling preparations

Generally, two problems should be settled before the Gauss process can give predictions: the prior covariance function and the corresponding hyper-parameters. The prior covariance can be specified manually based on empirical experiences or expert judgments which provide structural information about covariance functions. And the numerical values of the hyper-parameters make the characteristics of the covariance differ broadly and will be optimized by model training with proper techniques.

Here another prior covariance function was also chosen: the Matérn class functions with isotropic distance measure, i.e.

$$K_{\mathrm{Mc}}(x_p, x_q) = \sigma_f^2 f(\sqrt{d} * r) * \exp(-\sqrt{d} * r) \tag{14}$$

Where $f(t) = 1+t$, $r = \sqrt{(x^p - x^q)^T * P^{-1}(x^p - x^q)}$, P is l times the unit matrix and σ_f^2 is the signal variance, The corresponding hyper-parameters are l and σ_f^2 with the initial value $l = 1/4, \sigma_f^2$=1. The training and test process were implemented using the displacement series in Tab. 1 with $d = 15, m = 1$.

In order to test the performance of modeling on the non-stationary displacement series using Gauss process regression, dynamic crossing validation was applied to strengthen the ability of the model to give more precise predictions. The hyper-parameters are adjusted by the strategy of model training techniques introduced in Section 3.3. The conjugate gradient method is applied for iteration computing and optimizing values of the hyper-parameters. To avoid local minimum problem during the training process, we would randomly initialize several selective values within the space of the hyper-parameters during the implementation procedures. If the values of hyper-parameters are specified, we can obtain the predictions of the test input by substituting the values of the hyper-parameters into eq. (7).

The data for modeling often has a large size in length in practical application for analysis and prediction of long-term monitoring displacement series of landslides.

This causes approximation problems for these large datasets in data processing. There are several choices for the approximation methods, such as the subset of regressors, the Nyström method, the subset of data points and the Bayesian committee machine [Rasmussen and Williams (2006)]. In this study we perform short-term point predictions and use Subset of Data Points method which is indicated in the process of dynamic crossing validation.

4.4 Results and analysis

In this section, comparison and discussion are organized into three stages. The former is the comparison of the Gauss process introduced in this study with the previous work; the second is a result comparison of different covariance functions using the Gauss process; the third part is result discussion on different methods applied for the predictive results based on the observation of the landslide.

4.4.1 Comparison with a previous work

The relative error loss (*REL*) is defined to evaluate the performances of different covariance functions $REL = \left|\frac{y_i - y_i^*}{y_i}\right| \times 100\%, (i = 1,2,\ldots,n)$,$y_i$,$y_i^*$ is the i^{th} observation and prediction.

Predictive results and the corresponding *RELs* are comparatively shown in Tab.2 for the strategy of the present work in this study and the previous work in the literature [LIU and XU (2009)]. The distributive characteristics of the predictive *REL* are both shown in Fig.2 for the present and previous works. It can be seen that the previous method is moderate since the height of error histogram of the previous work are much larger than that of the present strategy using GPR with the dynamic crossing validation technique for most of the predictive points, which was especially highlighted at the turning point of the observed curve of the displacement. Also, it can be seen that the error histograms jumps swiftly at the turning point of the observed data curve for both methods which implies that it is detective for predictions on catastrophe points using continuous methods. Whereas the predictive errors present in Fig.2 shows that the prediction strategy proposed in this research can grasp and track the displacement fluctuations and turnings in a very short time.

Besides the relative error loss for each point prediction, two other overall loss functions are also promoted to evaluate quantitatively the overall performance of the regression strategy presented above. The average relative error loss (AREL) and the average square error loss (ASEL) are defined as $AREL = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i^* - y_i}{y_i}\right| \times 100\%$, $ASEL = \frac{1}{n}\sum_{i=1}^{n}(y_i^* - y_i)^2$ to give comparisons with the previous work.

The *AREL* and *ASEL* of the previous and present strategies are shown in Tab.3. As

Table 2: Comparisons of predictions and the errors of the present and previous work

Time /d	*a* /mm	predictions/mm		relative error loss%		Time /d	*a* /mm	predictions/mm		average error loss %	
		b	*c*	*b*	*c*			*b*	*c*	previous	present
30	8.7	8.65	8.69	0.57	0.17	49	17.6	17.72	18.02	0.68	2.40
31	9.0	8.78	8.93	2.44	0.78	50	18.2	17.89	17.73	1.70	2.58
32	9.2	8.85	9.24	3.80	0.48	51	19.0	18.31	18.64	3.63	1.90
33	9.4	9.26	9.43	1.49	0.29	52	19.2	18.33	19.50	4.53	1.58
34	10.0	9.31	9.59	6.90	4.07	53	20.0	18.13	19.72	9.35	1.38
35	10.1	9.32	10.27	7.72	1.70	54	23.0	20.26	20.43	11.91	11.15
36	10.3	10.22	10.40	0.78	0.97	55	24.0	20.62	24.09	14.08	0.38
37	10.4	10.51	10.52	1.06	1.17	56	25.2	20.93	25.46	16.94	1.04
38	10.5	10.77	10.55	2.57	0.49	57	26.0	28.92	26.42	11.23	1.63
39	10.8	10.66	10.59	1.30	1.94	58	27.0	29.65	26.84	9.81	0.59
40	11.1	10.73	10.90	3.33	1.79	59	28.2	28.92	27.68	2.55	1.84
41	12.0	10.78	11.25	10.17	6.23	60	30.0	29.82	28.94	0.60	3.54
42	13.0	12.76	12.55	1.85	3.49	61	31.0	31.28	30.99	0.90	0.02
43	13.4	13.61	13.90	1.57	3.77	62	32.0	32.80	32.11	2.50	0.35
44	14.0	14.27	13.88	1.93	0.88	63	33.0	33.80	32.94	2.42	0.17
45	15.0	15.57	14.44	3.80	3.71	64	42.0	35.26	33.82	16.05	19.48
46	16.1	16.74	15.67	3.98	2.70	65	47.0	36.76	41.70	21.79	11.28
47	16.4	17.95	17.07	9.45	4.07	66	61.0	61.15	51.85	0.25	15.00
48	17.2	17.28	16.40	0.47	4.66						

Hint: the column *a* represents the original observing data; column *b* represents previous results; column *c* represents the present results of this study.

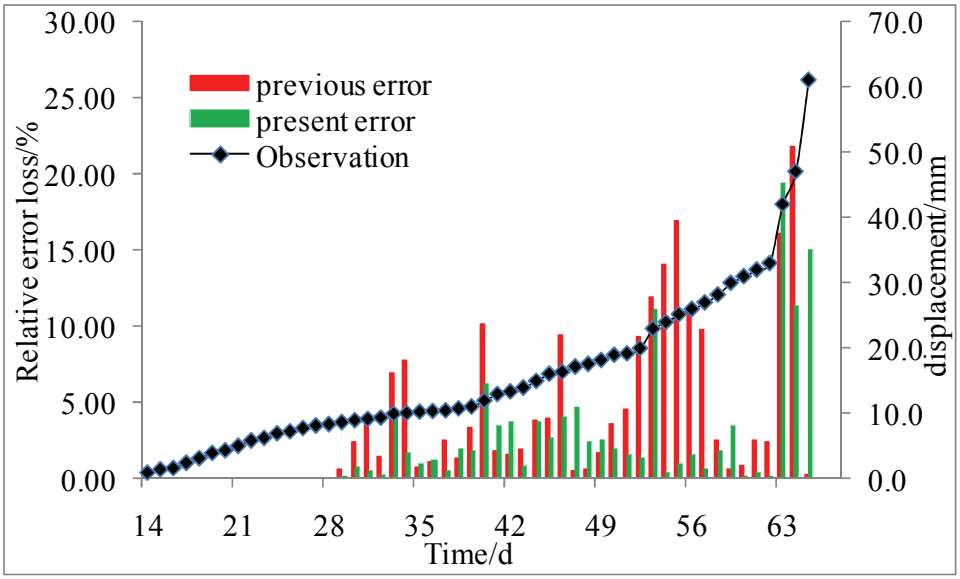

Figure 2: Comparisons of relative error loss for each prediction

can be drawn from Tab.3, the *AREL* of the present strategy is 3.23%, about 60.9% of that of the previous work; the *ASEL* of the present strategy is 4.68, about 80.1% of that of the previous method. Thus the performance of predictions using the present method in this study has been greatly promoted in contrast to that of the previous work.

The reason for moderate performance of the previous work in the literature lies in that the Gauss process is one of Bayesian learning methods but it was confused in the literature. In the Bayesian theory model learning is implemented to compute and adjust the hyper-parameters in the frame of marginalization over the likelihood function. It is different in nature with those learning methods in the frame of minimizing empirical loss. The former is implemented with probabilistic characteristics. And prior knowledge about the studied system is embedded with a certain form, for example the covariance functions. These are the key points of the Gauss process. As for the literature [LIU and XU (2009)], it described that the Genetic Algorithm was used to optimize the hyper-parameters. However, prudential readers would notice that it applied Genetic Algorithm to adjust the hyper-parameters using a fitness function (essentially a loss function) after the training and prediction for each processing step and carried out computations using an iteration method until the termination conditions(e.g.100 processing steps) were satisfied. It did not implement the model learning process in the frame of marginal likelihood but

in the form of loss function. Thus it could not generate the same type of satisfactory results as the strategy that is introduced in this work. Actually, those biological algorithms could be applied to optimize the hyper-parameters instead of the commonly-used conjugate gradient method for iteration computations but they cannot be mixed with loss functions.

Table 3: Comparisons of predictive loss functions

method	AREL/%	ASL
previous	5.30	5.84
present	3.23	4.68

4.4.2 Comparisons of different covariance functions

Several terms of covariance functions introduced above, $K_{\mathrm{se}}, K_{\mathrm{Mc}}, K_{\mathrm{ns}}$and the composites, are adopted in this study in order to investigate the influences of different prior covariance functions to the strategy of the Gauss process. The corresponding predictive results are listed in Tab.4. The *RELs* for each point prediction is also shown in Tab.4.

Several conclusions can be obtained from the predictive performances shown in Tab.3. Predictive results of the Gauss process with different covariance functions all had good accuracy, with *RELs* less than 3%, for point predictions at the non-fluctuation positions on the displacement curve. However, predictive performances at the fluctuation positions were moderate, mostly less than 10%, though not very good. Results did not differ very much from those obtained by the single covariance (K_{se} or K_{mc}) and the composite covariance functions. Also, K_{se} and K_{mc} performed similarly in the modeling prediction problem of the Wolong Temple landslide displacement. The compositions of K_{se} or K_{mc} with K_{ns} did not improve the *RELs* in our predictive modeling problem, while the composition of K_{se} and K_{mc} slightly improved the overall predictive performances. The reason may lie in that K_{ns} is not suitable for accounting the characteristics of the displacement series of the Wolong Temple and the existence of K_{ns} reduces the predictive performance. Thus it can be summarized here that the compositions of different covariance functions do not necessarily give improvement in predictive performances for a specified modeling problem. In addition to this covariance functions should be in accordance with the characteristics of the modeling system so as to give satisfactory predictive results.

4.4.3 Comparison with other methods

In the previous section, we discussed the comparative results of different Gauss process strategies. Comparisons with other methods such as the SVM and ANNs will

Table 4: Predictive performances of Gauss process with different covariance functions

Day/d	Observation	predictions of different covariance functions /mm						relative error loss of different covariance functions /%			
		K_{se}	K_{Mc}	$K_{se}+K_{ns}$	$K_{ns}+K_{Mc}$	$K_{se}+K_{Mc}$	K_{se}	K_{Mc}	$K_{se}+K_{ns}$	$K_{ns}+K_{Mc}$	$K_{se}+K_{Mc}$
30	8.70	8.56	8.65	9.50	8.73	8.68	0.14	0.05	0.80	0.03	0.02
31	9.00	8.80	8.88	9.85	8.95	8.89	0.20	0.12	0.85	0.05	0.11
32	9.20	9.11	9.17	9.15	9.20	9.17	0.09	0.03	0.05	0.00	0.03
33	9.40	9.28	9.36	9.36	9.41	9.36	0.12	0.04	0.04	0.01	0.04
34	10.00	9.44	9.55	9.61	9.60	9.55	0.56	0.45	0.39	0.40	0.45
35	10.10	10.29	10.21	10.32	10.28	10.21	0.19	0.11	0.22	0.18	0.11
36	10.30	10.47	10.42	10.54	10.48	10.42	0.17	0.12	0.24	0.18	0.12
37	10.40	10.62	10.56	10.68	10.61	10.56	0.22	0.16	0.28	0.21	0.16
38	10.50	10.56	10.58	10.58	10.61	10.58	0.06	0.08	0.08	0.11	0.08
39	10.80	10.66	10.54	10.66	10.67	10.54	0.14	0.26	0.14	0.13	0.26
40	11.10	10.89	10.89	11.12	11.12	10.94	0.21	0.21	0.02	0.02	0.16
41	12.00	11.14	11.16	11.46	11.44	11.22	0.86	0.84	0.54	0.56	0.78
42	13.00	11.77	11.79	12.29	12.26	12.52	1.23	1.21	0.71	0.74	0.48
43	13.40	12.59	14.10	13.93	13.81	14.03	0.81	0.70	0.53	0.41	0.63
44	14.00	13.57	13.98	13.47	13.54	13.80	0.43	0.02	0.53	0.46	0.20
45	15.00	13.99	14.33	14.22	14.30	14.25	1.01	0.67	0.78	0.70	0.75
46	16.10	15.69	15.56	15.31	15.49	15.83	0.41	0.54	0.79	0.61	0.27
47	16.40	16.84	16.88	16.68	16.92	17.02	0.44	0.48	0.28	0.52	0.62
48	17.20	16.95	17.11	17.00	16.21	16.52	0.25	0.09	0.20	0.99	0.68
49	17.60	17.78	17.79	17.72	17.76	18.04	0.18	0.19	0.12	0.16	0.44
50	18.20	17.99	18.07	18.05	18.08	18.00	0.21	0.13	0.15	0.12	0.20
51	19.00	18.39	18.55	18.52	18.60	18.64	0.61	0.45	0.48	0.40	0.36
52	19.20	19.17	19.35	19.27	19.41	19.49	0.03	0.15	0.07	0.21	0.29
53	20.00	19.57	19.66	19.61	19.72	19.56	0.43	0.34	0.39	0.28	0.44
54	23.00	20.41	20.37	20.34	20.41	20.40	2.59	2.63	2.66	2.59	2.60
55	24.00	22.12	22.20	22.21	22.19	24.33	1.88	1.80	1.79	1.81	0.33
56	25.20	23.84	25.76	23.90	25.65	23.83	1.36	0.56	1.30	0.45	1.37
57	26.00	25.81	26.76	25.86	26.35	26.55	0.19	0.76	0.14	0.35	0.55
58	27.00	27.25	26.87	27.28	26.58	26.91	0.25	0.13	0.28	0.42	0.09
59	28.20	28.51	27.48	28.50	27.48	28.21	0.31	0.72	0.30	0.72	0.01
60	30.00	29.61	28.73	28.41	28.74	29.30	0.39	1.27	1.59	1.26	0.70
61	31.00	30.94	30.75	30.74	30.71	31.21	0.06	0.25	0.26	0.29	0.21
62	32.00	32.07	31.99	31.91	31.98	31.97	0.07	0.01	0.09	0.02	0.03
63	33.00	32.95	32.98	32.85	33.00	33.09	0.05	0.02	0.15	0.00	0.09
64	42.00	34.13	34.02	33.91	34.00	34.06	7.87	7.98	8.09	8.00	7.94
65	47.00	38.87	39.28	39.03	39.26	42.51	8.13	7.72	7.97	7.74	4.49
66	61.00	53.60	53.07	52.89	51.78	48.22	7.40	7.93	8.11	9.22	12.78

be now discussed. We first introduces the variant terms of the covariance functions in the kernel form to show the consistency of the Gauss process with other learning methods. Then it compares the predictive results obtained by different predictive strategies for the last seven point predictions of the landslide displacement series.

It can be noticed that the coefficient $K_*^{\mathrm{T}}[K+\sigma_n^2 I]^{-1}$ in eq.(8) does not depend on the observed target y, but only on the input X hence the predictive mean $\bar{f}_*$ is a linear combination of the observed target y; this is the property of the Gauss process. The covariance in eq.(9) is the difference between $K(X_*,X_*)$ $K(X_*,X_*)$(the prior covariance) and a positive term representing the information that the observations give us about the function From another point of view, the mean $\bar{f}_*$ in eq.(8) is the linear combination of n kernel functions each of which is centered on a training point i.e. $\bar{f}(x_*)=\sum_{i=1}^{n}\alpha_i k(x_i,x_*)$ where $\alpha=(K+\sigma_n^2 I)^{-1}y$.

Therefore it can be concluded that the Gauss process is consistent with other kernel learning methods. In fact it has been proved that most kernel learning methods could be used in accordance with the Gauss process with specific restrictions [Rasmussen and Williams (2006)]. Thus, comparisons of the Gauss process with ANNs and SVM are to be discussed hereafter in regards to the predictive performance for modeling the landslide displacement. Only the last seven point predictions are discussed for comparison since it has been shown that the Gauss process can obtain quite satisfactory results for the other predictions as given in Tab.2 and Tab.4.

The predictive results obtained by RBF network and SVM [DONG, FU and LENG (2007)] are shown and compared with that of the Gauss process in Fig. 3. It can be seen that the three strategies (GPR, RBF network and SVM) all give very good point predictions for days from 60 to 62 but there are divergent results for days 63 to 66 until the landslide occurred. The observations start jumping from day 63, which leads to moderate performances for all three strategies. Fortunately, the results of the Gauss process regression show that the predictive results it generates have the same trends with that of the observations even though it cannot perform very well for point values. However the results given by the RBF and the SVM are inferior for trend keeping even though they give better point predictions for some days. The ultimate goal of modeling slope displacements is for forecasting landslide occurrence. Trend prediction is an essential element for both long-term and short-term forecasting of landslides, as well as for imminent warnings. It can be concluded in this end that the GPR performs better than the RBF and the SVM on the predictive modeling of the landslide displacement.

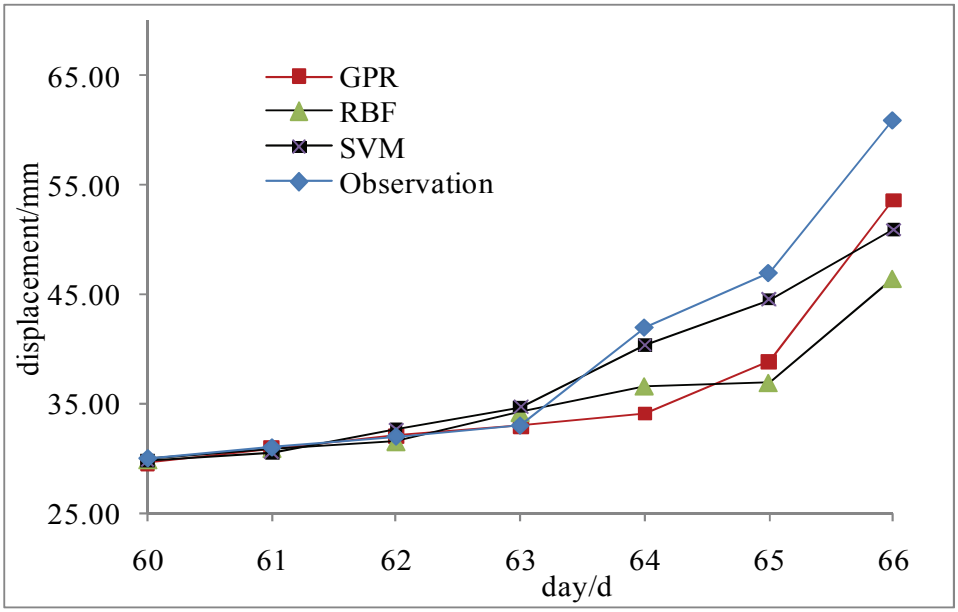

Figure 3: Predictive performances of different strategies

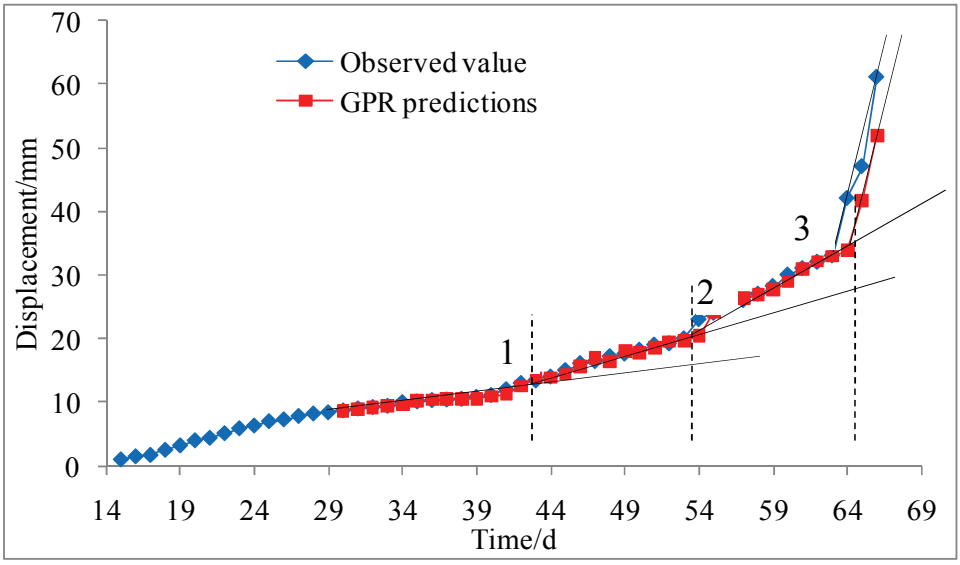

Figure 4: Actual displacements and the GPR predictions

4.5 Further discussion

Predictive modeling on landslide displacement is an effective way for revealing the future features of landslides. It can provide reference for landslide occurrences in advance with proper warning criteria. In contrast to the previous works, now concentration can be made on the lateral information that the GPR predictions expresses for the warning criteria of landslides. We can see the observed displacement process and the corresponding predictions by the GPR of the landslide in Fig.4. As can be seen from Fig.4, the displacement of this landslide is stepped over four stages divided by three apparent inflections before the sliding movement occurred. The linear trends of the displacement curve at each inflection points are drawn with continuous solid lines and the boundary lines of each stage are marked with dash lines for a clearer understanding of the involution stages. It must be recognized that the linear trend at each inflection point is derived approximately by a line that can pass through as many data points as possible. Also, one can find that the intersection angle at each inflection point is different from the others and that they enlarge as the landslide displacement develops with time. The values of the intersection angles were no less than 10°, and the intersection angle at the last inflection (labeled as 3) was more than 45°, after which the landslide occurred. Thus, this characteristic can be considered as a criterion for forecasting such kinds of landslide.

There are two main indexes which evaluate a model's ability to forecast landslides from the observed displacement series: the trends of the underlying displacement and the retardation time of forecasting, given the thresholds value for the landslide occurrence. Drawing back to the attractive performance of the GPR strategies, it can be noted that the predictive strategy of the paper could adapt well with the displacement curves even at the inflection points. In addition to this, the trends close to sliding time were nearly the same with that of the observations since the trend line for both the observation and GPR curves are somewhat parallel to each other. On the other hand, once the observed displacement jumped to a new range, the GPR model could adapt to it swiftly in the next prediction. At this point, if the intersection angle on the predictive curve is 45°, it would be a signal for the forecasting of the landslide occurrence when modeling the observed displacement series. Thus it can be regarded as a threshold value of criterion for landslide forecasting. Whereas one could also recognize that despite its attractive ability on tendency tracking, the displacement predictions of the GPR with dynamic crossing validation is somewhat hysteretic to the observations at the sudden turnings of the curves. Also, the landslide forecasting criteria here will be responsible only for the slopes with typical creep-typed displacements curves like Wolong Temple New Landslide.

Actually, some criteria have been proposed for landslides forecasting, but the corresponding threshold values for each criterion are difficult to identify. A portion

of the criteria is the displacement rate, the rainfall thresholds, the safety factors and the like. With the increasing sophistication of the site-investigation of slopes, criteria with good reliability, feasibility and efficacy will intensively utilize the information from the monitoring systems. The landslides discussed in this article ware the creep-typed slope in which the displacement curve changes slowly. In addition to this, there are landslides with stepped or shock-type displacement curves. The stepped displacement curves often exist in slopes formed by deposits. And the threshold values for the forecasting criteria are to be further discussed. Using a reliable and precise prediction model for slope displacement prediction like the Gauss process in this research, landslide forecasting can be implemented loosely with proper threshold values of displacement criterion.

5 Conclusions

The Gauss process utilizes monitored data for model training. It can cover expert prior knowledge in prior functions, which has increased its application in the fields in which empirical experience is important. Its hyper-parameters are optimized by model training to minimize the likelihood except for empirical risk functions. These are why it can outperform than other techniques only dealing with data series. Based on the results of this study, conclusions can be drawn as follows:

(1) The Gauss process is a good technique for modeling of landslide displacement series. It has superior ability in point predictions for landslide displacement modeling with proper prior covariance functions. Also, it can provide satisfactory results for tendency tracking of the displacement series. Thus it can be regarded as a good strategy for tendency predictions of phenomenal data of complex systems.

(2) The covariance functions and the relevant hyper-parameters are the main causes for moderate or good performances of the Gauss process strategy with dynamic crossing validation in this study. Model training of the Gauss process is executed by means of minimizing the likelihood of hyper-parameters. The predictive *AREL* and *ASEL* of the strategy in this study is about 60.9% and 80.1% better than that in previous works. Thus training target function cannot be confused with loss functions. Also, comparative works show that the strategy introduced in this study is superior to the RBF and the SVM since only this strategy can accurately track the tendency of the displacement curve of Wolong Temple Slide even there is a sudden inflection. This ability is very important for a model in order to predict landslide occurrences.

(3) It would be a signal for forecasting a landslide occurrence if the intersection angle reaches 45° at the turnings on the predictive curve when modeling the observed displacement of the creep-typed slope. A landslide is a complex system involving

many disciplines. Displacement modeling is a phenomenal analytical method for landslide study; further studies need to continue on landslide predictions by a composition of displacement study and evolutionary mechanism analysis. Meanwhile, studies should be also carried out on the topic of warning criteria for slopes with different kinds of displacement characteristics such as the mutant type and stepped type.

Appendix

Starting with Bayesian theory, the marginal likelihood $p(y|X)$ is the integral of the likelihood times the prior distribution, referring to the marginalization over the lateral function f

$$p(y|X) = \int p(y|f,X)p(f|X)df \tag{15}$$

The prior of the Gauss process is Gaussian distribution, $f|X \sim N(0,K)$

$$\log p(f|X) = -\frac{1}{2}f^T K^{-1} f - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi \tag{16}$$

The likelihood is a factorized Gaussian, $y|f \sim N(f,\sigma_n^2 I)$. Let $K_y = K + \sigma_n^2 I$, the log marginal likelihood is obtained as follows since the factorization of Gaussians is also of Gaussian.

$$\log p(y|X) = -\frac{1}{2}\mathrm{y}^T (K_y)^{-1}\mathrm{y} - \frac{1}{2}\log|K_y| - \frac{n}{2}\log 2\pi \tag{17}$$

The hyper-parameters θ are implied potentially in the log marginal likelihood eq.(15). Since

$$\frac{\partial}{\partial\theta}K_y^{-1} = -K_y^{-1}\frac{\partial K_y}{\partial\theta}K_y^{-1}; \quad \frac{\partial}{\partial\theta}\log|K_y| = tr\left(K_y^{-1}\frac{\partial K_y}{\partial\theta}\right) \tag{18}$$

Then eq.(12) is obtained.

Acknowledgement: We are grateful for Carl Edward Rasmussen and Hannes Nickisch for sharing their fundamental code in MATLAB.

References

Barber D.; Saad D. (1996): Does Extra Knowledge Necessarily Improve Generalisation? *Neural Computation* , vol.8, pp.202-214.

Bishop C.M. (1995): *Neural Networks for Pattern Recognition*. Clarendon press, Oxford.

Brahim-Belhouari S.; Vesin J.M.(2001): Bayesian learning using Gaussian process for time series prediction. In:*11th IEEE Statistical Signal Processing Workshop*, pp.433-436.

Brahim-Belhouari S.; Bermak A. (2004): Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, vol.47, no.4, pp.705-712.

Chau K.W. (2006): Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. *Journal of Hydrology,* vol.329, no.3-4, pp.363-367.

Dong H.; Fu H.L.; Leng W.M. (2007): Landslide Displacement Prediction Based on Takens Theory and SVM. *China Journal of Highway and Transport*, vol.20, no.5, pp.13-18. (in Chinese)

Feng X.T.; Wang Y.; Yao J. (1996): A neural network model on real-time prediction of roof pressure in coal mines. *Int. J. of Rock Mech. and Min. Sci.*, vol.33, no.6, pp.647-53.

Feng X.T.; Zhang Z.; Xu P. (1999): Adaptive and intelligent prediction of determination time series of high rock slope. *Transactions of Nonferrous Metals Society of China*, vol.4, pp.842-847.

Feng X.T.; Zhao H.B.; Li S.J. (2004): Modeling non-linear displacement time series of geo-materials using evolutionary support vector machines. *International Journal of Rock Mechanics and Mining Sciences,* vol.41, no.7, pp.1087-1107. (in Chinese)

Ferentinou M.D.; Sakellariou M.G. (2007): Computational intelligence tools for the prediction of slope performance. *Computers and Geotechnics*, vol.34, no.5, pp. 362-384.

Li H.; Sun J. (2009): Gaussian case-based reasoning for business failure prediction with empirical data in China. *Information Sciences*, vol.179, pp.89-108.

Liu K.; Liu B.; Xu C. (2009): Intelligent analysis model of slope nonlinear displacement time series based on genetic-Gaussian process regression algorithm of combined kernel function. *Chinese Journal of Rock Mechanics and Engineering*, vol.28, no.10, pp.2128-2134. (in Chinese)

Liu Z.B.; Xu W.Y.; Meng Y.D. (2009): Modification of GM (1, 1) and its Application in Analysis of Rock-slope Deformation. *Proc. of 2009 IEEE Int. Con. on GSIS*. Nanjing, China, pp. 415-419.

Matías J.M.; Taboada J.; Ordótez C. (2010): Partially linear support vector ma-

chines applied to the prediction of mine slope movements. *Mathematical and Computer Modelling*, vol.51, no.3-4, pp.206-215.

Mehdi K.; Mehdi B.(2011): A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*, vol.11, no.2): 2664-2675.

Neal R.M. (1996): *Bayesian Learning for Neural Networks*. Springer, New York.

Rasmussen C. E.; Williams C. K. I. (2006): *Gaussian Processes for Machine Learning*, the MIT Press.

Sakellariou M.; Ferentinou M. (2005): A study of slope stability prediction using neural networks. *Geotechnical and Geological Engineering*, vol. 24, no.3, pp.419-445.

Samui P.; Kothari D.P. (2011): Utilization of a least square support vector machine (LSSVM) for slope stability analysis. *Scientia Iranica*, vol.18, no.1, pp.53-58.

Steihaug T. (1983): The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. *SIAM Journal on Numerical Analysis*, vol.20, no.3, pp.626-637.

Tarek E.; Nizar B.(2011): Bayesian learning of finite generalized Gaussian mixture models on images. *Signal Processing*, vol.91, no.4, pp.801-820.

Wang H.B.; Xu W.Y.; Xu R.C. (2005): Slope stability evaluation using Back Propagation Neural Networks. *Engineering Geology*, vol.80, no.3-4, pp. 302-315.

Williams C.K.I.; Rasmussen C.E.(1996): Gaussian process for regression. In: *Advances in Information Processing Systems*, MIT Press, Cambridge, MA, 8th Edition, pp.111-116.

Xu F.; Xu W.Y. (2010): Prediction of displacement time series based on support vector machines-Markov chain. *Rock and Soil Mechanics*, vol.31, no.3, pp. 944-948. (in Chinese)

Yuan Y.; Xu Q.; Guo K. (2005): Search for the jump spot of a landslide using the nsmutation algorithm. *Journal of Engineering Geology*, vol.13, no.4, pp.447-450. (in Chinese)

Zhou Z.J.; Hu C.H. (2008): An effective hybrid approach based on grey and ARMA for forecasting gyro drift, *Chaos, Solitons and Fractals*, vol.35, pp.525-529.