Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

doi: 10.3850/978-981-18-5182-7_03-006-cd

ISGSR 2022

Efficient Reliability Analysis of Slopes in Spatially Variable Soils Based on Active-Learning Multivariate Adaptive Regression Spline

Zhi-Ping Deng[1*], Min Zhong[1], Min Pan[1], Jing-Tai Niu[1] and Ke-Hong Zheng[1]

[1]College of Water Conservancy and Ecological Engineering, Nanchang Institute of Technology, 289 Tianxiang Road, Nanchang 330099, People's Republic of China
E-mail: dengzhiping@nit.edu.cn.

Abstract: Reliability analysis of slopes in spatially variable soils is often faced with a huge computational burden. To effectively reduce the number of numerical models in reliability analysis and alleviate the calculation pressure, this paper proposes an active-learning multivariate adaptive regression spline (MARS) method for reliability analysis of spatially variable slopes. Firstly, Latin hypercube sampling is used to obtain the initial training samples. The active-learning function is then used to update the MARS surrogate model by searching for new training samples around the limit state surface. The failure probability of the slope can be calculated using Monte Carlo simulation (MCS) after obtaining an appropriate surrogate model. A typical slope with spatially variable soils is used as an example to verify the efficiency of the proposed method. The results show that the proposed method not only reduces the quantity of calculations required, but also achieves high accuracy in estimating slope failure probability.

Keywords: Slope reliability; Spatial variability; Random field; Multivariate adaptive regression splines; Active-learning.

1 Introduction

Due to geological, environmental, and physicochemical factors on natural soil sedimentation, soil characteristics tend to be spatially variable (Phoon and Kulhawy 1996; Dasaka and Zhang 2012). Even with various exploration and testing methods, the resulting soil properties remain uncertain (Li et al. 2017; Deng et al. 2020; Deng et al. 2022). The traditional slope stability safety factor evaluation method does not consider the spatial variability of soil parameters. To this end, a considerable number of studies have been conducted in recent years on the spatial variability of slope soil parameters (Cho, 2010; Li et al. 2017; Deng et al. 2017; Wang et al. 2020; Deng et al. 2021).

Many methods have been proposed for slope reliability analysis, the simplest and most effective of which is direct Monte Carlo simulation (MCS). Its advantage is that when the number of samples is large enough, an unbiased estimation of failure probability with high accuracy can be achieved. Nevertheless, this method necessitates a significant amount of computation. When using the finite element or finite difference strength reduction method (SRM), the cost of the MCS method will be very high (Zhang et al. 2011). Especially for the case with a small failure probability, hundreds of thousands of numerical simulation calculations (Kang et al. 2015) are often needed, such that the computational workload is too large to meet the requirements of practical application. To reduce the calculation workload, some advanced MCS methods are proposed, such as the important sampling method and subset simulation method, but these methods still need to call a certain amount of slope stability calculation times. To better reduce the computational workload, some scholars have proposed to construct an efficient explicit surrogate model to replace the numerical calculation model with high workload, such as support vector machines (SVM) (Samui et al., 2013), Artificial Neural Network (ANN) (Cho, 2009), and Gaussian Process Regression (GPR) (Kang et al., 2015). Nevertheless, these studies only consider finding a more robust algorithm to reduce the requirement of sample points, rather than developing a method to provide more sample points for the construction of response surface models. The number of sample points is the number of calls to a deterministic analysis, which can lead to computationally heavy and long computation times, especially when using finite difference methods to obtain safety factors (He et al. 2020).

In recent years, surrogate models proposed by some scholars still require a certain number of training samples to obtain models with higher prediction accuracy. To improve the efficiency and accuracy of slope reliability research, this paper proposes a spatial slope reliability analysis method based on active-learning MARS-MCS (AMARS-MCS), which can obtain a surrogate model with high accuracy in a small number of training samples. First, the Karhunen-Loève (K-L) expansion method is used to discretize the random field of soil parameters, and a large sample pool and basic training samples are generated. The active learning function is used to select the optimal sample, and the selected optimal sample is substituted into the training sample. On this basis, the MARS surrogate model is established. Finally, the slope reliability analysis is carried out by using the MCS method. This paper uses a typical slope in spatially variable soils as an example to illustrate the effectiveness of the proposed method.

2 Simulation of Spatial Variability of Soil Parameters Using Karhunen-Loève Expansion Method

Due to the influence of geology and physicochemical action in the formation process of soil, the parameters of soil shear strength show spatial variability. The K-L expansion requires fewer random variables, and has high computational accuracy and efficiency. Therefore, this method is used in this paper to discretize the random field, and the shear strength parameter correlation logarithmic random field can be expressed as:

$$H_i^{LN}(x,y) = exp(\mu_{lni} + \sum_{j=1}^{n} \sigma_{lni} \sqrt{\lambda_i} f_j(x,y) \chi_{i,j}) \quad (for\ i = c, \phi) \tag{1}$$

where, n is the number of truncation terms which needs to be determined according to the discrete accuracy ; λ_j are the eigenvalues corresponding to the autocorrelation function; $f_j(x,y)$ is the characteristic function corresponding to the autocorrelation function; $\mu_{\ln i}$ and $\sigma_{\ln i}$ are the mean and standard deviation of the corresponding normal distribution parameter ln$i$ respectively.

In view of the fact that the gauss type correlation function is continuous and derivable everywhere, the Gauss type correlation function is chosen as the spatial autocorrelation of soil parameters in this paper(Cami et al. 2020 and Zhou 2021):

$$\rho(\tau_m, \tau_n) = 1/\exp\left[\frac{L_m^2}{\delta_h^2} + \frac{L_n^2}{\delta_v^2}\right] \tag{2}$$

where, δ_h and δ_v are horizontal and vertical correlation distance respectively; L_m and L_n are the difference values of the corresponding coordinates of two points in the soil space, respectively; $\chi_{i,j}$ is a standard normal random variable.

3 Multivariate Adaptive Regression Spline Method

The multivariate adaptive regression spline method is a nonlinear and nonparametric regression method based on a piecewise strategy proposed by Friedman (1991). This method can effectively use the sample data to establish the relationship between the input variables and their responses. In the system, the real function $f(\xi)$ can be expressed as follows:

$$f(\xi) \approx \hat{f}(\xi) = a_0 + \sum_{m=1}^{M} a_m B_m(\xi) \tag{3}$$

where, $\hat{f}(\xi)$ is the predicted value of the MARS model output; $\xi = (\xi_1, \xi_2, ..., \xi_\pi)$ is a set of input variables; a_m is the coefficient of the mth basis function; coefficient a_m can be obtained by least squares fitting of training samples.

The MARS algorithm has two processes, forward selection and backward pruning. The forward process algorithm starts with only one constant term basis function $B_0(x)$=1; each iteration generates two new basis functions, then identifies the new basis functions, and finally obtains the approximate model. The model updated in each iteration is as follows:

$$\hat{f}(\xi) = a_0 + \sum_{m=1}^{M} a_m B_m(\xi) + \hat{a}_{m+1} B_i(\xi) \max(0, x_j - t) + \hat{a}_{m+2} B_i(\xi) \max(0, t - x_j) \tag{4}$$

where, $\hat{a}_{m+1}$ and $\hat{a}_{m+2}$ can be used for the least squares fitting calculation of training samples; $B_i(\xi)$ is the previously determined basis function, 0≤1≤M. In general, the forward process algorithm will produce an over-fitting model, because the MARS algorithm allows the use of the previous basis function to construct a new basis function, resulting in a decrease in the contribution of the initial basis function. Therefore, the backward process algorithm should be adopted to improve the prediction ability of the model.

The backward process algorithm takes every cycle, deletes the original basis function and obtains the sub-model, using generalized cross-validation (GCV). The specific form is as follows:

$$GCV(M) = \frac{1}{N}\sum_{i=1}^{N}\left[y_i - \hat{f}_M(x_i)\right]^2 \bigg/ \left[1 - \frac{C(M)}{N}\right]^2 \tag{5}$$

where, $C(M) = trace\left(B\left(B^T B\right)^{-1} B^T\right) + 1 + dM$; $trace\left(B\left(B^T B\right)^{-1} B^T\right) + 1$ is the number of effective coefficients of the model. $\hat{f}_M(x_i)$ is the corresponding prediction value at the sample point; d For a penalty factor between 2 and 4, generally 3.

4 Active Learning Function

After the initial sample points are determined, the surrogate model is iteratively updated. The active learning function is based on the current selected MARS model, and each time an optimal sample point (obtained by each simulation analysis) is selected to join the training sample set to achieve the purpose of iterative updating of the MARS model. In order to select the optimal sample point, the sampling method is based on MCS pool (Echard et al. 2011; Pan, and Dias, 2017).

The active learning function construction is always a difficult point (Xiao et al., 2018). In recent years, scholars have proposed an empirical active learning function based on sampling principles (Echard et al. 2011; Xiao et al. 2018). On the basis of predecessors, Li et al (2018) proposed an efficient active learning function considering a distance penalty term:

$$u_c = \arg\min \frac{(1 + \|u_T\|)\,|\,\hat{G}(u_T)\,|}{\left[\dfrac{d(u_T, S)}{1 + \exp(-20(d(u_T, S) - d(S)))}\right]} \tag{6}$$

where, $\hat{G}(\cdot)$ is a functional proxy model. For the classifier, only the prediction result is given instead of the numerical value. $\hat{G}(\cdot)$ can be replaced by F(**u**) . For other models, $\hat{G}(\cdot)$ need not be changed; $d\left(u_T, S\right)$ represents the minimum distance between the point of the sample set and u_T in the Independent Standard Normal Space; d(S) is a minimum distance limit that varies and can be calculated as follows:

$$d(S) = \lambda \max_{u_S^{(i)}}\left(\min_{u_S^{(j)}}\left\|u_S^{(i)} - u_S^{(j)}\right\|\right) : u_S^{(i)}, u_S^{(j)} \in S \tag{7}$$

where, λ is a constant coefficient term between 0.1 and 0.5, and λ=0.2 is used. The constant coefficient term is followed by a distance value, which represents the distance between two sample points in the sampling area.

5 The Steps of the AMARS-MCS Method Execution

To consider both stratigraphic uncertainty and spatial variability of soil parameters in the slope, a full probabilistic design method combining the AMARS-MCS model with the random field model is proposed. The implementation procedure is presented, as shown in Figure 1.

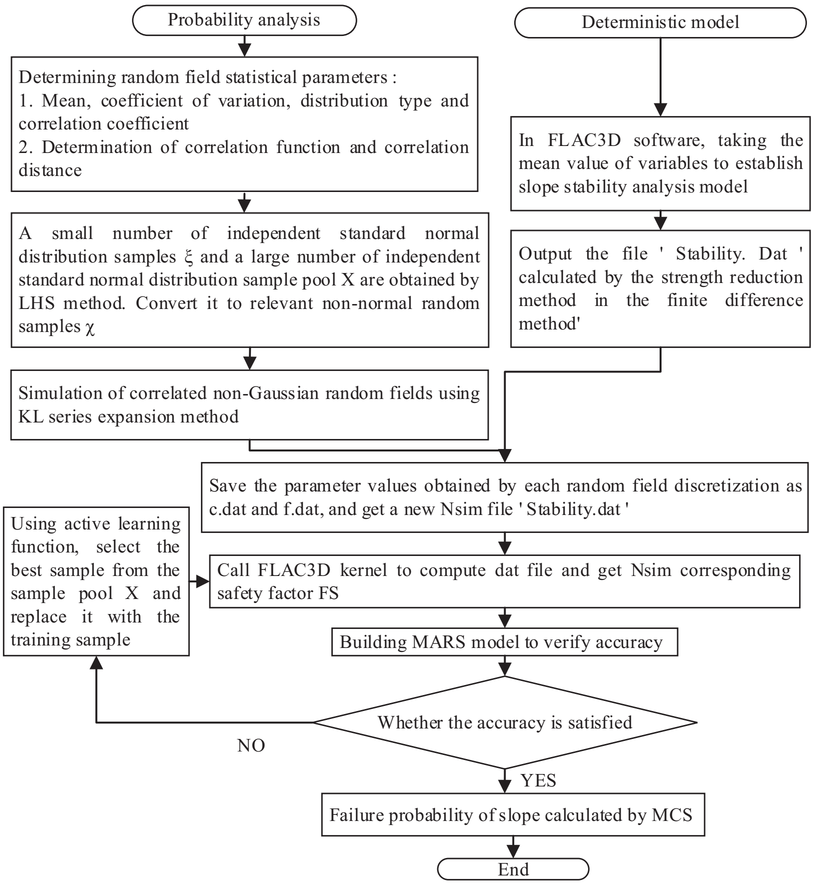

Figure 1. Flowchart of slope reliability analysis based on the AMARS-MCS method

6 Illustrative Example

This section will study the reliability of a single-layer *c*-*φ* slope. Previous studies (Cho, 2010; Li et al., 2015; Liu et al., 2017b; Liu, 2019) also analyzed the slope. Therefore, previous research results can be directly used to test the accuracy of the proposed method.

6.1 Description and certainty analysis of slope

The slope model studied in this section is shown in Figure 2. The slope height is 10 m and the slope toe is oriented at 45°. The soil parameters for the probabilistic stability analysis are shown in Table 1. This study only focuses on the spatial variability of cohesion and internal friction angle of the soil. The correlation coefficient $\rho_{c\varphi}$ is assumed to be -0.5. Subsequently, the K-L expansion method with the assumed SOFs is used to discrete the random fields.and other parameters are regarded as constants. Assuming horizontal fluctuation range δ_h to 40, vertical fluctuation range δ_v to 4, the random field is discretized, so that the discrete data corresponding to the centroid of each random field unit can characterize the shear strength of the soil body of the unit. According to the known parameters, the model is initially established by FLAC3D, and the safety factor is 1.228 as calculated by SRM, which is very close to the results of 1.204 and 1.205 calculated by Cho (2010) and Liu et al. (2019), indicating the correctness of the model.

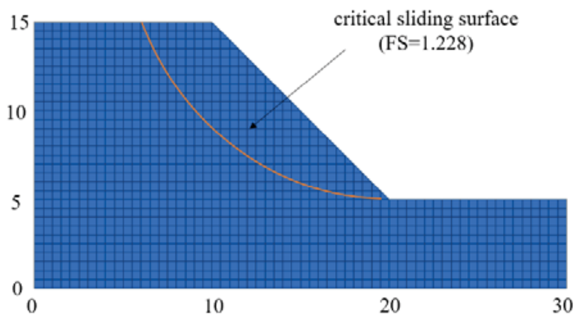

Figure 2 Slope model

Table 1. Statistical characteristics of soil parameters

Soil parameters	Mean	Cov	Distribution	SOF	$\rho_{c\varphi}$
c	10 kPa	0.3	Lognormal	δ_h=40 m, δ_v=4 m	-0.5
φ	30°	0.2	Lognormal	δ_h=40 m, δ_v=4 m	
γ	20 kN/m^3	—	—	—	—

The symbol "–" means not applicable

The established model shown on Figure 2 includes 1190 quadrilateral elements and 20 triangular elements. In order to compare with the results of Li et al. (2015) and Liu et al. (2019), a consistent grid division has been employed.

6.2 Analysis of slope reliability results

Firstly, according to the random field statistical parameters and model centroid coordinates, the parameters are dispersed to each centroid coordinate point, and the spatial variability is simulated to generate thirty slope models with parameters. Then, the finite difference SRM is used to calculate the slope safety factor. Based on the thirty training samples calculated by the SRM, the number of training samples is repeatedly increased according to the active learning function. Figure 3 shows the variation of slope failure probability with the number of samples. It can be seen from the figure that the failure probability tends to stabilize quickly with the increase of the number of iterative samples. The relationship between the function prediction G1 and the calculated value G2 is shown in Figure 4, which also shows the accuracy of the model. Table 2 lists the results (including Mean value of safety factor μ_{FS}, standard deviation of safety factor σ_{FS}, and coefficient of variation COV_{FS}) obtained by different methods. It can be seen from Table 2 that the corresponding results of different methods are consistent. The proposed method requires only 60 samples to obtain more accurate results, while the MARS-MCS method proposed by Liu et al. (2019) requires at least 280 samples. This indicates that the proposed method can obtain sufficiently accurate reliability calculation results with a small computational cost.

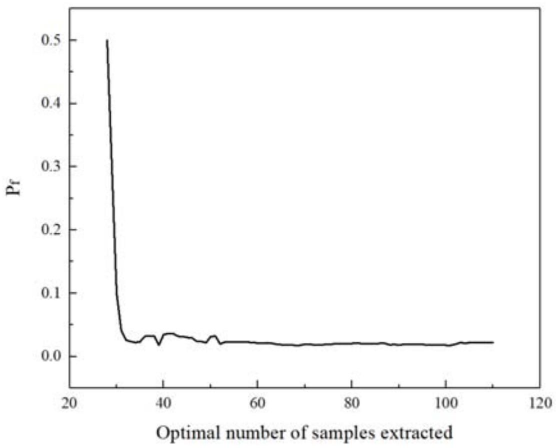

Figure 3. P_f changes with the number of training samples

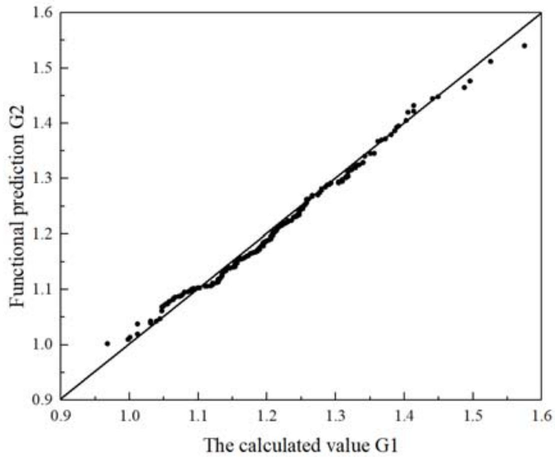

Figure 4. Comparison of predicted safety factor and calculated safety factor

Table 2. Reliability calculation results of different methods

Method	μ_{FS}	σ_{FS}	COV_{FS}	P_f
MARS-MCS（Liu et al. 2019）	1.198	0.114	0.096	2.92×10^{-2}
1×10^4 LHS（this study）	1.184	0.103	0.087	2.72×10^{-2}
AMARS-MCS（this study）	1.204	0.108	0.900	2.50×10^{-2}

7 Conclusion

This study presents a spatially variable slope reliability analysis method based on AMARS-MCS, which combines the advantages of an active learning function and MARS. In this study, a spatial slope is used as an example to illustrate the effectiveness of the proposed method. The main conclusions are as follows:

(1) The AMARS model as used to analyze slopes with spatial variability can effectively establish the relationship between the slope shear strength parameters and the safety factor.

(2) For spatially variable slopes, the proposed method can obtain sufficiently accurate reliability calculation results with a small computational cost.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Project No. 52009054), and the Science and Technology Project Funded by the Department of Education of Jiangxi Province (Project No. GJJ211932). The financial supports are gratefully acknowledged.

References

Cami, B., Javankhoshdel, S., Phoon, K. K., & Ching, J. (2020). Scale of fluctuation for spatially varying soils: estimation methods and values. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering,* 6(4), 03120002.

Cho, S. E. (2009). Probabilistic stability analyses of slopes using the ANN-based response surface. *Computers and Geotechnics,* 36(5), 787-797.

Cho S. E. (2010). Probabilistic assessment of slope stability that considers the spatial variability of soil properties. *Journal of Geotechnical and Geoenvironmental Engineering*, 136(7):975–984.

Dasaka, S. M., and Zhang, L. M. (2012). Spatial variability of in situ weathered soil. *Géotechnique*, 62(5), 375-384.

Deng, Z. P., Li, D. Q., Qi, X. H., Cao, Z. J., and Phoon, K. K. (2017). Reliability evaluation of slope considering geological uncertainty and inherent variability of soil parameters. *Computers and Geotechnics*, 92, 121-131.

Deng, Z. P., Jiang, S. H., Niu, J. T., Pan, M., & Liu, L. L. (2020). Stratigraphic uncertainty characterization using generalized coupled Markov chain. *Bulletin of Engineering Geology and the Environment*, 79(10), 5061-5078.

Deng, Z. P., Pan, M., Niu, J. T., Jiang, S. H., and Qian, W. W. (2021). Slope reliability analysis in spatially variable soils using sliced inverse regressionbased multivariate adaptive regression spline. *Bulletin of Engineering Geology and the* Environment, 80(9), 7213-7226.

Deng, Z. P., Pan, M., Niu, J. T., & Jiang, S. H. (2022). Full probability design of soil slopes considering both stratigraphic uncertainty and spatial variability of soil properties. *Bulletin of Engineering Geology and the Environment*, 81(5), 1-13.

Echard, B., Gayton, N., and Lemaire, M. (2011). AK-MCS: an active learning reliability method combining Kriging and Monte Carlo simulation. *Structural Safety*, 33(2), 145-154.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1), 1-67.

Goh, A. T., and Zhang, W. G. (2014). An improvement to MLR model for predicting liquefaction-induced lateral spread using multivariate adaptive regression splines. *Engineering Geology*, 170, 1-10.

Huang, J., Lyamin, A. V., Griffiths, D. V., Krabbenhoft, K., and Sloan, S. W. (2013). Quantitative risk assessment of landslide by limit analysis and random fields. *Computers and Geotechnics*, 53, 60-67.

He, X., Xu, H., Sabetamal, H., and Sheng, D. (2020). Machine learning aided stochastic reliability analysis of spatially variable slopes. *Computers and Geotechnics*, 126, 103711.

Jiang, S. H., Li, D. Q., Zhou, C. B., and Zhang, L. M. (2014). Capabilities of stochastic response surface method and response surface method in reliability analysis. *Structural Engineering and Mechanics*, 49(1), 111-128.

Jiang, S. H., Li, D. Q., Cao, Z. J., Zhou, C. B., and Phoon, K. K. (2015). Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. *Journal of Geotechnical and Geoenvironmental Engineering*, 141(2), 04014096.

Samui, P. (2013). Multivariate adaptive regression spline (Mars) for prediction of elastic modulus of jointed rock mass. *Geotechnical and Geological Engineering*, 31(1), 249-253.

Kang, F., Han, S., Salgado, R., and Li, J. (2015). System probabilistic stability analysis of soil slopes using Gaussian process regression with Latin hypercube sampling. *Computers and Geotechnics*, 63, 13-25.

Li, D. Q., Jiang, S. H., Cao, Z. J., Zhou, W., Zhou, C. B., and Zhang, L. M. (2015). A multiple response-surface method for slope reliability analysis considering spatial variability of soil properties. *Engineering Geology*, 187, 60-72.

Li, X. Y., Zhang, L. M., Gao, L., and Zhu, H. (2017). Simplified slope reliability analysis considering spatial soil variability. *Engineering Geology*, 216, 90-97.

Li, X., Gong, C., Gu, L., Gao, W., Jing, Z., and Su, H. (2018). A sequential surrogate method for reliability analysis based on radial basis function. *Structural Safety*, 73, 42-53.

Liu, L., Zhang, S., Cheng, Y. M., and Liang, L. (2019). Advanced reliability analysis of slopes in spatially variable soils using multivariate adaptive regression splines. *Geoscience Frontiers*, 10(2), 671-682.

Pan, Q., and Dias, D. (2017). Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions. *Reliability Engineering and System Safety*, 167, 484-493.

Pan, Q., and Dias, D. (2017). An efficient reliability method combining adaptive support vector machine and Monte Carlo simulation. *Structural Safety*, 67, 85-95.

Phoon, K. K., & Kulhawy, F. H. (1996). On quantifying inherent soil variability.

Kulhawy, F. H., and Phoon, K. K. (1996, August). Engineering judgment in the evolution from deterministic to reliability-based foundation design. *In Proceedings of Uncertainty* (Vol. 96, pp. 29-48).

Xiao, N. C., Zuo, M. J., and Zhou, C. (2018). A new adaptive sequential sampling method to construct surrogate models for efficient reliability analysis. *Reliability Engineering and System Safety*, 169, 330-338.

Qi, X. H., and Li, D. Q. (2018). Effect of spatial variability of shear strength parameters on critical slip surfaces of slopes. *Engineering Geology*, 239, 41-49.

Wang, L., Wu, C., Gu, X., Liu, H., Mei, G., and Zhang, W. (2020). Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. *Bulletin of Engineering Geology and the Environment*, 79(6), 2763-2775.

Zhang, J., Zhang, L. M., and Tang, W. H. (2011). New methods for system reliability analysis of soil slopes. *Canadian Geotechnical Journal*, 48(7), 1138-1148.

Zhou, Z., Li, D. Q., Xiao, T., Cao, Z. J., and Du, W. (2021). Response surface guided adaptive slope reliability analysis in spatially varying soils. *Computers and Geotechnics*, 132, 103966.