Engineering Geology 181 (2014) 27–37

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Geotechnical reliability analysis with limited data: Consideration of model selection uncertainty

J. Zhang [a], H.W. Huang [a,*], C.H. Juang [b], W.W. Su [a]

[a] Key Laboratory of Geotechnical and Underground Engineering of Ministry of Education, Department of Geotechnical Engineering, Tongji University, Siping Road 1239, Shanghai 200092, China
[b] Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form 20 May 2014
Accepted 1 August 2014
Available online 14 August 2014

Keywords:
Probabilistic method
Failure probability
Model uncertainty
Slope stability

ABSTRACT

The limited amount of data available in geotechnical practice makes it difficult to identify a unique probability model for the joint distribution of uncertain variables. Yet, the calculated failure probability can be sensitive to the probability model used, even if different models are calibrated based on the same data. The model selection uncertainty is a poorly understood area of research in current geotechnical practice. In this study, we show how to construct candidate probability models based upon the copula theory to more realistically model the soil data with explicit consideration of the possible non-linear dependence relationship between random variables. The authors used a Bayesian method to quantify the model selection uncertainty and to compare the validity of the candidate models. A model averaging method that combines predictions from competing models was then developed to deal with the situation when the effect of model selection uncertainty cannot be neglected. Averaging over the reliability index seems more plausible than averaging over the failure probability in geotechnical reliability analyses. To reduce the computational work, models with significantly less model probabilities can be removed from the model averaging process without an obvious effect on the prediction accuracy.

1. Introduction

As probability-based methods provide a more rational basis for treating uncertainties, the probabilistic modeling of geotechnical problems has been the subject of much research (e.g., Tang, 1984; Whitman, 1984; Rackwitz, 2000; Ching et al., 2009; Griffiths et al., 2011; Zhang et al., 2011; Ahmed and Soubra, 2012; Uzielli and Mayne, 2012; Duncan, 2013; Zhu and Zhang, 2013). While its usefulness is now well known, the dilemma of the practical application of probabilistic methods is that the site-specific data available for geotechnical reliability analysis are often limited (Christian, 2004; Zhang et al., 2004; Wang and Cao, 2013), which could be insufficient for identifying a unique probability model for the joint distribution of uncertain soil properties. In practice, the selection of a specific probabilistic model often involves subjective decisions with several options (Beer et al., 2013). If the lack of information is very severe, for example, when only bounds are known for parameters involved in the mechanical problem, non-probabilistic methods such as interval analysis and fuzzy modeling have been sought for uncertainty quantification and processing (Degrauwe et al, 2010; Luo et al., 2011; Beer et al., 2013).

In geotechnical engineering, the joint distribution of random variables is often determined based on the marginal distributions and a correlation matrix. In recent years, substantial progress has been made in modeling multivariate data based on the copula theory (e.g., Cherubini et al., 2004; Trivedi and Zimmer, 2005; Nelsen, 2006), in which a copula function instead of the correlation matrix is used to represent the dependence relationship among random variables. As shown in Boardman and Vann (2011), as the copula function varies, there could be multiple joint distributions corresponding to the same marginal distributions and the same correlation matrix. Li and his coauthors pioneered the application of the copula theory in a number of geotechnical reliability problems (Li et al., 2013; Tang et al., 2013a, b,c). Their study shows that previous probability models used in geotechnical engineering such as multivariate normal distribution is indeed based on the Gaussian copula, which can only consider the linear dependence relationship between random variables and may not always be optimal. Therefore, it is important to consider other copula functions for constructing probability models in geotechnical reliability analysis.

While the copula theory provides a flexible tool to model the geotechnical data more realistically, it further complicates the model selection problem, i.e., it is necessary to select not only the marginal distributions but also the copula function. In practice, a single probability model is often chosen based the best available knowledge, and the failure probability is thus calculated based on the selected probability model ignoring the model selection uncertainty. Yet, the calculated failure probability could be sensitive to the probability model adopted (e.g., Tang et al., 2013a). However, the model selection uncertainty is

* Corresponding author.
E-mail addresses: cezhangjie@gmail.com (J. Zhang), huanghw@tongji.edu.cn (H.W. Huang), hsein@clemson.edu (C.H. Juang).

http://dx.doi.org/10.1016/j.enggeo.2014.08.002
0013-7952/

rarely considered in current geotechnical reliability analysis, and how to address the model selection uncertainty problem remains a problem.

The cohesion and the friction angle are two common uncertain variables in many geotechnical problems. Using shear strength data as an example, we propose a method of reliability analysis that explicitly considers the model selection uncertainty. The results from a reliability analysis could be the statistics of a system response or the failure probability/reliability index. As an illustration, in this study our focus is on failure probability/reliability index. We will first introduce our method of constructing the probability models for shear strength parameters based on the copula theory. Next, a description of our method of evaluating and incorporating model selection uncertainty in the failure probability calculation is presented. Finally, we provide two examples of our proposed method for investigating the model selection uncertainty and its effect on reliability analysis. Although we use the shear strength data to illustrate the proposed method, it is equally applicable to other soil data types when model selection uncertainty needs to be addressed.

2. Constructing of candidate probability models

2.1. Modeling a joint distribution based on the copula theory

In many reliability problems, such as the bearing capacity of shallow foundations (e.g., Wang, 2011; Juang and Wang, 2013), the stability of earth slopes (e.g., Zhang et al., 2010, 2013; Huber et al., 2011; Salgado and Kim, 2014), the design of retaining structures (e.g., Zevgolis and Bourdeau, 2010; Low et al., 2011), the cohesion and the friction angle are often treated as random variables. It is a common practice to model the cohesion and the friction angle as either normal or lognormal variables (Lumb, 1970; Cherubini, 2000). A correlation coefficient is often used to describe the dependence between the cohesion and the friction angle (e.g., Harr, 1987; Cherubini, 1997). Though a negative correlation between the cohesion and the friction angle is the usual outcome in such modeling, a positive correlation between the cohesion and the friction angle also occurs (e.g., Wolff, 1985). A single correlation coefficient cannot describe the possible non-linear dependence relationship between random variables, however (e.g., Boardman and Vann, 2011). In this study, the authors constructed the joint distribution of the cohesion and the friction angle based on the copula theory, as described below.

Let x_1 and x_2 denote the cohesion and the friction angle, respectively. Let $F_1(x_1)$ and $F_2(x_2)$ denote the cumulative distribution function (CDF) of x_1 and x_2, respectively. Let $F(x_1, x_2)$ denote the joint CDF of $\{x_1, x_2\}$. Based on Sklar's theorem (Sklar, 1959), if x_1 and x_2 are continuous variables, $F(x_1, x_2)$ can be written as follows

$$F(x_1, x_2) = K[F_1(x_1), F_2(x_2), \theta] \tag{1}$$

where $K(u_1, u_2, \theta)$ is a copula function with $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$, and θ is a parameter of the copula function. Eq. (1) shows that the joint distribution of x_1 and x_2 decomposes into two parts: the marginal distributions and the copula function, as several commonly used copula functions in Table 1 indicate. In this table, the Gaussian copula is the copula associated with the bivariate normal distribution, and thus is the dependence function implicitly assumed in the bivariate normal distribution. The Clayton copula can be used to model data with strong left tail dependence and relatively weak right tail dependence. Compared with the Gaussian copula, the dependence in the tails of the Frank copula tends to be relatively weak, and the strongest dependence is centered. The FGM is often attractive due to its simplicity, and it is most useful when dependence between the two marginal distributions is modest in magnitude. The Plackett copula exhibits less tail dependence than the Gaussian copula. In addition to the above general descriptions, different copula models differ in detail as represented by their mathematical expressions. One can refer to Cherubini et al. (2004), Trivedi and Zimmer (2005), and Nelsen (2006) for more about the characteristics of different copula models.

Based on Eq. (1), the joint probability density function (PDF) of x_1 and x_2 can be expressed as:

$$f(x_1, x_2) = k[F_1(x_1), F_2(x_2), \theta] f_1(x_1) f_2(x_2) \tag{2}$$

where $k(u_1, u_2, \theta)$ is the density function of $K(u_1, u_2, \theta)$ defined as follows

$$k(u_1, u_2, \theta) = \frac{\partial^2 K(u_1, u_2, \theta)}{\partial u_1 \partial u_2}. \tag{3}$$

The density functions of several copula functions are also shown in Table 1. Eq. (2) can be used to construct the PDF of the joint distribution of $\{x_1, x_2\}$. For instance, if it is assumed that the marginal distributions are normal and that copula function is Clayton, the joint PDF of x_1 and x_2 is expressed as

$$f(x_1, x_2) = (1+\theta)[F_1(x_1) \cdot F_2(x_2)]^{-\theta-1} \left\{ [F_1(x_1)]^{-\theta} + [F_2(x_2)]^{-\theta} - 1 \right\}^{-2-\frac{1}{\theta}} f_1(x_1) f_2(x_2) \tag{4}$$

where $f_i(x_i)$ and $F_i(x_i)$ in this case are the PDF and CDF of x_i $(i = 1, 2)$, respectively, and both are PDF of normal variables.

2.2. Maximum likelihood calibration of model parameters

Let μ_i and σ_i denote the mean and the standard deviation of x_i, $(i = 1, 2)$, respectively. In the bivariate distribution constructed based on the copula theory for shear strength parameters, the parameters to be calibrated include μ_1, σ_1, μ_2, σ_2, and θ. Let $\mathbf{d} = \{d_1, d_2\}$ denote a measurement of $\{x_1, x_2\}$. Let $\mathbf{d}^1, \mathbf{d}^2, \mathbf{d}^3, \ldots, \mathbf{d}^n$ denote n measurements of $\{x_1, x_2\}$. For ease of presentation, let $\mathbf{\Theta} = \{\mu_1, \sigma_1, \mu_2, \sigma_2, \theta\}$ and $\mathbf{D} =$

Table 1
Summary of several commonly used copula functions.

Copula	$k(u_1, u_2, \theta)$	$k(u, u_2, \theta)$	Range of θ
Gaussian	$\Phi[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta]$	$\frac{1}{\sqrt{1-\theta^2}} \exp\left[-\frac{\theta^2\xi_1^2 - 2\theta\xi_1\xi_2 + \theta^2\xi_2^2}{2(1-\theta^2)}\right]$ $\xi_1 = \Phi^{-1}(u_1)$ $\xi_2 = \Phi^{-1}(u_2)$	$(-1, 1)$
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$	$(1+\theta)(u_1 u_2)^{-\theta-1}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-2-\frac{1}{\theta}}$	$(0, +\infty)$
Frank	$-\frac{1}{\theta} \ln\left\{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\right\}$	$\frac{\theta(1-e^{-\theta})e^{-\theta(u_1+u_2)}}{\left[(1-e^{-\theta}) - (1-e^{-\theta u_1})(1-e^{-\theta u_2})\right]^2}$	$(-\infty, +\infty)\backslash\{0\}$
FGM	$u_1 u_2[1 + \theta(1 - u_1)(1 - u_2)]$	$1 + \theta(1 - 2u_1)(1 - 2u_2)$	$[-1, 1]$
Plackett	$\frac{S - \sqrt{S^2 - 4u_1 u_2 \theta(\theta-1)}}{2(\theta-1)}$ $S = 1 + (\theta-1)(u_1 + u_2)$	$\frac{\theta[1 + (\theta-1)(u_1 + u_2 - 2u_1 u_2)]}{[S^2 - 4u_1 u_2 \theta(\theta-1)]^{\frac{3}{2}}}$	$(0, +\infty)\backslash\{1\}$