Computers and Geotechnics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers and Geotechnics

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

Research Paper

Bayesian model comparison and characterization of bivariate distribution for shear strength parameters of soil

Lei Zhang[a], Dian-Qing Li[a], Xiao-Song Tang[a,*], Zi-Jun Cao[a], Kok-Kwang Phoon[b]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering (Ministry of Education), Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China

[b] Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07-03, 1 Engineering Drive 2, Singapore 117576, Singapore

ARTICLE INFO

Keywords:
Shear strength parameters
Joint probability distribution
Correlation coefficient
Bayesian approach
MCMC simulation
Copula

ABSTRACT

This paper develops a Bayesian approach for model comparison and characterization of the bivariate distribution of c' and ϕ' using limited site-specific data. The copula approach is presented to model the bivariate distribution of c' and ϕ'. The Bayesian model comparison method is developed to select the most probable bivariate distribution model of c' and ϕ'. The most probable model is used to characterize the joint probability density function (PDF) of c' and ϕ' under the Bayesian framework. The developed approach is illustrated and validated using real data of c' and ϕ' for clays from the core wall of Xiaolangdi rockfill dam in China.

1. Introduction

The shear strength parameters of soil [i.e., effective cohesion (c') and effective friction angle (ϕ')] are important parameters for evaluating stability and deformation of geotechnical structures, such as slopes (e.g., [41,9,27,28,29,13]), retaining walls (e.g., [25,16]) and foundations (e.g., [26,7,8]). In reliability analysis of these geotechnical structures, c' and ϕ' are typically treated as uncertain variables [9,16,7]. Furthermore, it is widely accepted that c' and ϕ' are negatively correlated parameters (e.g., [18,17,24,27,28,29,12,13]). To achieve a realistic evaluation of geotechnical reliability, the joint probability distribution of c' and ϕ' should be constructed since ignoring the negative correlation between c' and ϕ' would lead to a substantial overestimate of the probability of failure (e.g., [17,7,27,28,29,13]).

Recently, the copula approach (e.g., [22]) provides a general and flexible way for modeling the joint probability distribution of c' and ϕ' (e.g., [28,29,38,39,44,21,12,13,40]). A copula refers to a function that couples a multivariate distribution to its one-dimensional marginal distributions (e.g., [22]). There are many copulas in the literature to characterize the dependence structure among variables such as Gaussian, *t*, Plackett, Frank, Clayton and Gumbel copulas (e.g., [22]). Each copula has its own characterized dependence structure. The copula approach constructs the joint probability distribution of c' and ϕ' by combing the marginal distributions of c' and ϕ' with a copula function. Specifically, the construction of the joint probability distribution of c' and ϕ' includes: (1) identification of the best-fit marginal distributions and copula, and (2) estimation of distribution parameters of the best-fit marginal distributions and copula parameters of the best-fit copula.

Conventional statistical methods are commonly adopted to construct the joint probability distribution of c' and ϕ'. For instance, AIC (Akaike Information Criterion) scores are often used to identify the best-fit marginal distributions and copula (i.e., a marginal distribution or copula with the minimum AIC score among the set of candidate distributions or copulas is the best-fit marginal distribution or copula). Furthermore, sample mean, standard deviation and correlation coefficient are used to estimate the distribution parameters and copula parameters. Note that AIC scores as well as sample mean, standard deviation and correlation coefficient are statistics derived from the site-specific laboratory or field test data of c' and ϕ' [12]. Therefore, the accuracy of the constructed joint probability distribution of c' and ϕ' using conventional statistical methods depends on the reliability of the derived statistics, which lies on the sample size of the site-specific data of c' and ϕ'. For example, a minimum sample size of 30 is required to achieve a meaningful estimation of the mean, standard deviation and correlation coefficient [1]. The minimum sample size increases dramatically to more than 100 to achieve a reliable identification of the best-fit marginal distributions and copula using AIC scores (e.g., [14,15,30]). However, small sample size for geotechnical data is a common feature in geotechnical practice. The sample size of geotechnical data in a specific site is typically less than 30 for common geotechnical parameters (e.g., [23,31,37,6]). The joint probability distribution of c' and ϕ' derived from a small sample size has large statistical uncertainty and may be seriously biased (e.g., [30]). Therefore, characterization of the joint probability distribution of c' and ϕ'

* Corresponding author.
E-mail address: xstang@whu.edu.cn (X.-S. Tang).

http://dx.doi.org/10.1016/j.compgeo.2017.10.003
Received 7 December 2016; Received in revised form 19 July 2017; Accepted 5 October 2017
0266-352X/

based on limited site-specific data is a challenging problem in geotechnical engineering.

In addition to the limited site-specific data, there are often some prior knowledge (e.g., engineering judgement, local experience, published reports and studies) about the site. To make up for the lack of site-specific data, it is useful to take advantage of these prior knowledge so that a reasonable characterization of geotechnical variability can be achieved. The Bayesian approach has been proven to be a robust and practical approach for combining limited site-specific data with prior information in geotechnical engineering (e.g., [20,42,43]). The Bayesian approach can also quantify the statistical uncertainty in geotechnical variability derived from limited site-specific data. As the sample size of site-specific data increases, the statistical uncertainty in geotechnical variability decreases, and the geotechnical variability converges to its true value (e.g., [6]). The Bayesian approach has been widely applied to characterize geotechnical variability (e.g., [31,33,34,37,5,2,3,4,32,35,36]). For example, the Bayesian approach has been used for model comparison in geotechnical engineering such as the selection of appropriate regression models for predicting rock and soil parameters [3,35,36], the selection of spatial correlation function for characterizing the spatial variability of soil parameters [4], and the identification of soil stratification [2,33,34]. Furthermore, the Bayesian approach has been adopted to characterize the univariate probability distribution of soil parameters such as sand friction angle [31], Young's modulus of clay [32] and undrained shear strength [3].

The aforementioned studies using the Bayesian approach focused on one geotechnical parameter only. To characterize the variability of two geotechnical parameters, Wang and Aladejare [36] adopted the Bayesian approach to derive the site-specific joint probability distribution of uniaxial compressive strength (*UCS*) and Young's modulus (*E*) of rock. However, the study of Wang and Aladejare [36] did not perform model comparison to select the most probable bivariate distribution model (including the most probable marginal distributions and copula function) of *UCS* and *E*, but used the bivariate normal distribution model directly in the characterization of the site-specific joint probability distribution of *UCS* and *E*. The bivariate normal distribution assumes that both *UCS* and *E* follow univariate normal distributions, and their dependence structure can be characterized by a Gaussian copula. It is well known that geotechnical parameters do not necessarily follow univariate normal distributions. They may follow univariate lognormal distributions, Gumbel distributions, Gamma distributions, and among others. Similarly, geotechnical parameters also do not necessarily have a dependence structure of the Gaussian copula. They may have a dependence structure of the Plackett copula, Frank copula, No.16 copula, and among others. More importantly, the selection of the marginal distributions and copula for geotechnical parameters has a significant impact on the calculated geotechnical reliability (e.g., [28,13]). Past studies (e.g., [28,13]) showed that the probabilities of failure for geotechnical structures produced by different marginal distributions and copulas can differ in several orders of magnitude. Therefore, prior to the characterization of the joint probability distribution of geotechnical parameters using the Bayesian approach, the model comparison among the various bivariate distribution models using the copula approach should be conducted, and is the topic of the present study.

This paper aims to develop a Bayesian approach for model comparison and characterization of the bivariate distribution of c' and ϕ' using limited site-specific data. To achieve this goal, this article is organized as follows. In Section 2, the copula approach for modeling the bivariate distribution of c' and ϕ' is briefly introduced. In Section 3, the Bayesian model comparison method for selecting the most probable bivariate distribution model of c' and ϕ' using limited site-specific data and prior knowledge is developed. In Section 4, Bayesian characterization of the joint probability density function (PDF) of c' and ϕ' using the most probable bivariate distribution model is presented. The implementation procedure is provided in Section 5, followed by the illustration and validation of the developed approach using real data of c' and ϕ' for clays from the core wall of Xiaolangdi rockfill dam in China in Section 6.

2. Bivariate distribution of c' and ϕ' using copulas

The copula approach (e.g., [22,12]) is adopted in this study to model the bivariate distribution of c' and ϕ'. Let $F(c', \phi' \mid p_{c'}, q_{c'}, p_{\phi'}, q_{\phi'}, \theta)$ be the joint cumulative distribution function (CDF) of c' and ϕ'. The marginal CDFs of c' and ϕ' are denoted as $F_1(c' \mid p_{c'}, q_{c'})$ and $F_2(\phi' \mid p_{\phi'}, q_{\phi'})$, respectively. According to Sklar's theorem (e.g., [22]), $F(c', \phi' \mid p_{c'}, q_{c'}, p_{\phi'}, q_{\phi'}, \theta)$ can be expressed in the following general form:

$$F(c',\phi'|p_{c'},q_{c'},p_{\phi'},q_{\phi'},\theta) = C(F_1(c'|p_{c'},q_{c'}),F_2(\phi'|p_{\phi'},q_{\phi'})|\theta) = C(u,v|\theta) \tag{1}$$

where $(p_{c'}, q_{c'})$ are the distribution parameters for c'; $(p_{\phi'}, q_{\phi'})$ are the distribution parameters for ϕ'; $C(u, v \mid \theta)$ is a bivariate copula function, and θ is a copula parameter describing the dependency between c' and ϕ'. As shown in Eq. (1), the marginal CDFs $F_1(c' \mid p_{c'}, q_{c'})$ and $F_2(\phi' \mid p_{\phi'}, q_{\phi'})$ are usually denoted as u and v ranging from 0 to 1, respectively. Therefore, both u and v are standard uniform variables, and $C(u, v \mid \theta)$ is essentially a bivariate probability distribution on $[0, 1]^2$ with uniform marginal probability distributions on [0, 1]. By taking derivatives of Eq. (1), the joint PDF of c' and ϕ', $f(c', \phi' \mid p_{c'}, q_{c'}, p_{\phi'}, q_{\phi'}, \theta)$, can be obtained as:

$$f(c',\phi'|p_{c'},q_{c'},p_{\phi'},q_{\phi'},\theta) = f_1(c'|p_{c'},q_{c'})f_2(\phi'|p_{\phi'},q_{\phi'})D(u,v|\theta) \tag{2}$$

where $f_1(c' \mid p_{c'}, q_{c'})$ and $f_2(\phi' \mid p_{\phi'}, q_{\phi'})$ are the marginal PDFs of c' and ϕ', respectively; $D(u, v \mid \theta)$ is a copula density function, which is given by

$$D(u,v|\theta) = \partial^2 C(u,v|\theta)/\partial u \partial v \tag{3}$$

Sklar's theorem states that the bivariate distribution of c' and ϕ' can be expressed in terms of a copula function and their marginal distributions. Given the marginal distributions of c' and ϕ', and the copula function describing the dependence structure between c' and ϕ', the joint CDF and PDF of c' and ϕ' can be obtained by using Eqs. (1) and (2). Note that $p_{c'}$ and $q_{c'}$ are related to the mean $\mu_{c'}$ and standard deviation $\sigma_{c'}$ of c', while $p_{\phi'}$ and $q_{\phi'}$ are related to the mean $\mu_{\phi'}$ and standard deviation $\sigma_{\phi'}$ of ϕ'. Similarly, the copula parameter θ relates to the Kendall rank correlation coefficient τ between c' and ϕ' as follows (e.g., [12]):

$$\tau = 4\int_0^1 \int_0^1 C(u,v|\theta)\mathrm{d}C(u,v|\theta) - 1 \tag{4}$$

Therefore, $F_1(c' \mid p_{c'}, q_{c'})$ and $f_1(c' \mid p_{c'}, q_{c'})$ can be replaced by $F_1(c' \mid \mu_{c'}, \sigma_{c'})$ and $f_1(c' \mid \mu_{c'}, \sigma_{c'})$, respectively. Similarly, $F_2(\phi' \mid p_{\phi'}, q_{\phi'})$ and $f_2(\phi' \mid p_{\phi'}, q_{\phi'})$ are replaced by $F_2(\phi' \mid \mu_{\phi'}, \sigma_{\phi'})$ and $f_2(\phi' \mid \mu_{\phi'}, \sigma_{\phi'})$, respectively. For $C(u, v \mid \theta)$ and $D(u, v \mid \theta)$, they are replaced by $C(u, v \mid \tau)$ and $D(u, v \mid \tau)$, respectively. As a result, $F(c', \phi' \mid \mu_{c'}, \sigma_{c'}, \mu_{\phi'}, \sigma_{\phi'}, \tau)$ and $f(c', \phi' \mid \mu_{c'}, \sigma_{c'}, \mu_{\phi'}, \sigma_{\phi'}, \tau)$ are used instead of $F(c', \phi' \mid p_{c'}, q_{c'}, p_{\phi'}, q_{\phi'}, \theta)$ and $f(c', \phi' \mid p_{c'}, q_{c'}, p_{\phi'}, q_{\phi'}, \theta)$, respectively, and expressed as:

$$F(c',\phi'|\mu_{c'},\sigma_{c'},\mu_{\phi'},\sigma_{\phi'},\tau) = C(F_1(c'|\mu_{c'},\sigma_{c'}),F_2(\phi'|\mu_{\phi'},\sigma_{\phi'})|\tau) = C(u,v|\tau) \tag{5}$$

$$f(c',\phi'|\mu_{c'},\sigma_{c'},\mu_{\phi'},\sigma_{\phi'},\tau) = f_1(c'|\mu_{c'},\sigma_{c'})f_2(\phi'|\mu_{\phi'},\sigma_{\phi'})D(u,v|\tau) \tag{6}$$

The key tasks for modeling the bivariate distribution of c' and ϕ' using copulas are to select appropriate marginal distribution functions for c' and ϕ', and a copula function describing the dependence structure between c' and ϕ'. Since there exists a negative correlation between c' and ϕ', the copulas allowing a wide range of negative correlation coefficients are selected to fit the dependence structure between c' and ϕ'. For this reason, the Gaussian copula, Plackett copula, Frank copula and No.16 copula (e.g., [28,29,13,12]) are selected as the set of candidate copulas to fit the dependence structure between c' and ϕ'. These four copulas can describe negative dependences, and the values of