Soils and Foundations 2015;55(5):1243–1258

HOSTED BY
ELSEVIER
CrossMark

The Japanese Geotechnical Society

Soils and Foundations

www.sciencedirect.com
journal homepage: www.elsevier.com/locate/sandf

Modelling dependence structures of soil shear strength data with bivariate copulas and applications to geotechnical reliability analysis

Xing Zheng Wu*

Department of Applied Mathematics, School of Applied Science, University of Science and Technology Beijing, 30 Xueyuan Road, Haidan District, Post code: 100083, Beijing, PR China

Received 24 October 2014; received in revised form 22 April 2015; accepted 5 June 2015
Available online 1 October 2015

Abstract

Accurate estimates of the dependence of soil shear strength parameters (including cohesion and friction angle) play a crucial role in decision making by civil engineers in terms of geotechnical engineering safety. With increased site-specific information comes the need for joint soil strength models to account for the correlation characteristics between shear strength properties. In this study, using 16 sets of soil shear strength observations (consisting of 391 samples) as examples, the suitability of the dependence structure for these experimental observations is firstly identified by a goodness-of-fit test based on the Bayesian Information Criterion (BIC) with the normal, Student's *t*, Clayton, Frank, Gumbel, and Plackett copulas. The dependence structure between shear strength components is found to be asymmetric in most cases. Secondly, a set of paired samples of shear strength simulated from the different bivariate copulas, which contributed to various dependencies, is implemented as input for two typical geotechnical probabilistic analyses, e.g., infinite slope stability against a single sliding plane and the bearing capacity of a shallow foundation. The impact of the different choices for these dependence structures on the resulting reliability index is discussed. In both illustrative examples, the normal copula leads to an overestimation of the reliability index, whereas the Gumbel copula yields the lowest reliability index. Conservative reliability indices are obtained when the joint behaviour of the soil shear strength follows a bivariate normal distribution.

Keywords: Copula; Shear strength; Cross-correlation; Cohesion; Friction angle; Soil

1. Introduction

The two components of soil shear strength, cohesion and inner friction angle or friction angle, are the fundamental parameters determined by the classical Mohr–Coulomb failure criterion and applied to describe and explore the stability of geotechnical engineering (Burland, 1990). The inner friction, an angle of the failure surface or envelope, is defined on a stress plane (normal verse shear stress), and the cohesion is determined as a vertical intercept of the stress plane with the failure surface. The correlation between cohesion and friction angle can result from the mechanism for the parameters measured simultaneously using the same test or the physical link between the suction (related to cohesion) and mineralogy (related to friction angle) of the materials. During recent decades, a number of studies have emphasised that a cross-correlation may exist in these shear strength parameters (Lumb, 1970; Matsuo and Kuroda, 1974; Wolff, 1985; Cherubini, 2000; Fenton and Griffiths, 2003; Forrest and Orr, 2010; Fellin and Oberguggenberger, 2012). As examined

*Tel.: +86 18611869118; fax: +86 10 82610539.
E-mail address: xingzhengwu@gmail.com
Peer review under responsibility of The Japanese Geotechnical Society.

http://dx.doi.org/10.1016/j.sandf.2015.09.023

by several researchers (Chowdhury and Xu, 1992; Cherubini, 2000; Soubra and Mao, 2012; Wu, 2013a), the cross-correlation coefficient is a crucial factor in the probabilistic prediction of the response of geostructures. Wu (2013a) claimed that ignoring the correlations will most likely lead to underestimated or overestimated failure probabilities. Even so, some other investigators (Alonso, 1976; Tobutt and Richards, 1979; Li and Lumb, 1987; Lee and Chi, 2011) ignored the possible cross-correlation in their probabilistic analyses. Nevertheless, more evidence should be provided on the dependence structure between soil strength, especially for their non-symmetric and nonlinear features.

Considering the fact that the dispersion of the observed shear strength results is large, a linear regression line overlapping the plotted data is not appropriate, although this technique is often used (Parker et al., 2008; Hata et al., 2008; Sanchez Lizarraga and Lai, 2014). Alternatively, through the use of a linear Pearson's correlation coefficient, a bivariate normal distribution is popular for dealing with the joint probability distribution of two correlated random variables (Lumb, 1970; Matsuo and Kuroda, 1974; Wolff, 1985; Cherubini, 2000; Soubra and Mao, 2012). However, difficulties arise when random variables are not confirmed to follow jointly normal functions defined by the non-normal marginal distributions (Clemen and Reilly, 1999; Embrechts et al., 2003), which are typical situations in the reliability analysis of geostructures (Lumb, 1970; Kulatilake and Fuenkajorn, 1987). To overcome such difficulties, the Nataf transformation is used to transform non-normal data to normal data (Nataf, 1962). Der Kiureghian and Liu (1986) developed conversion factors to transform the linear correlation coefficient from non-normal space to normal space. Fundamentally, the Nataf transformation adopts a normal copula (Dutfoy and Lebrun, 2009) for the modelling dependence structure between variables. Copula functions are based on Sklar's Theorem (Sklar, 1959), and they provide a very flexible approach for modelling joint distributions in terms of univariate marginal functions and quantifying the dependence structure between variables (Joe, 1997; Nelsen, 2006). Hence, a joint distribution of cohesion and friction angle can be constructed by linking together any two marginal distributions of them (Clemen and Reilly, 1999; Lambert and Vandenhende, 2002; Poulin et al., 2007; Salvadori et al., 2007). Consequently, the impacts of the dependence asymmetries and the changes in correlation associated with real soil shear properties can be incorporated into the reliability analyses with ease.

There has recently been a surge of interest in applications of copula functions in civil engineering (Genest and Favre, 2007; Marchant et al., 2011; Li et al., 2012; Tang et al., 2013; Wu, 2013a, 2015a). Particularly, in applications to geotechnical engineering practice, studies of the cross-correlation characteristics of soil shear strength using copulas (Tang et al., 2013; Wu, 2013a, 2013b) are becoming increasingly important due to the urgent need to accurately quantify geotechnical risks with an acceptable degree of credibility. For instance, a copula-based direct integration method (Tang et al., 2013) is used to calculate the probability of failure to explore the impact of the copula selection on the probabilistic analysis. In their studies, less information on the marginal distribution is incorporated into the models. Wu (2013a, 2013b) proposed a copula-based sampling method (CBSM) to evaluate the impact of the correlation coefficient of the shear strength on the calculated reliability index of geostructures. However, these earlier works on copula-based models were based on very limited data from only a few sites, the issues for modelling the dependence characteristics of more actual observations of soil shear strength tests remain a challenge to be solved, especially when the non-normal marginal distribution is underlying. In this context, a large database should be compiled to explore the uncertainties in shear strength parameters. The best-fitting copulas of each soil will then be demystified; thus, the differences between these copulas and the conventional joint functions can be evaluated. Such refined attempts are likely to be fruitful and will provide better performances of a probabilistic analysis when integrating the copula models into the determination of a reliability index.

Therefore, a primary aim of this work is to validate the following: (1) Does individual soil with high cohesion (or a high friction angle) have a low friction angle (or cohesion) linked to the shear strength characteristics? (2) Which copula can be well-fit to the scatterplot of the shear strength observations, especially when the joint probability density functions are asymmetric? (3) How does the copula choice of soil shear strength impact the reliability indices of geostructures? In this study, the estimated distribution of 16 compiled laboratory datasets will be chosen to calibrate the copula models. The reliability index of geostructures is calculated using the CBSM when the correlations and dependence structures of shear strength are represented by different copulas, and its adequacy is determined from comparisons with the conventional bivariate normal distribution. The paper is organised as follows: Section 2 presents mathematical details of the joint bivariate distribution of soil strength and the marginal distributions. Section 3 identifies the marginal distribution and copula fitted to the collected experimental observations. The application of copula-based samples to the reliability analysis of the infinite slope stability and bearing capacity problem is illustrated in Section 4, and the impact of dependence structures on the reliability index is discussed. Section 5 draws the conclusions.

2. Theoretical framework

2.1. Mechanical background of shear strength parameters

In terms of the Mohr–Coulomb failure criterion, the mechanical and strength behaviour of cohesive granular materials involves two factors, namely, cohesion c and friction angle ϕ. These parameters are generally determined by means of direct shear or triaxial compression tests. For each test, a Mohr circle corresponding to shear failure can be drawn in Mohr's plane ($\hat{\sigma}$, $\hat{\tau}$), where $\hat{\sigma}$ is the normal stress and $\hat{\tau}$ is the shear stress. This procedure is executed for different confining pressures (at least three levels) and then the corresponding

Mohr circles can be plotted on the same graph. Consequently, an envelope of Mohr circles is approximated by a straight line to define the relation between shear stress and normal stress (Lumb, 1970). In fact, this relation determines c (shear stress at $\hat{\sigma} = 0$) and ϕ (slope of the line).

Due to the geological soil formation process and highly anisotropic condition, soil properties naturally exhibit heterogeneity and variation even within an apparently homogeneous profile. Additionally, testing procedures, both in the laboratory and in situ, are subject to many human factors which can introduce both systematic and random errors. A series of repeated tests should be performed for each type of soil to achieve a statistical description of the parameters. Considering that the measured results lie in a range or the discrepancies among them can be seen to be relatively large, these parameters may be described in terms of random variables. Such a reasonable reproducible test is helpful for exploring the real strength behaviour of materials, including their variabilities and probability density functions. Also, these paired parameters of the shear strength property show a correlation over a relatively large space or within a similar soil profile at a field site. This is reasonable since the two parameters are obtained by the same test simultaneously. In other words, they are functional for the same observation. Generally, a negative cross-correlation implies that when one strength parameter increases, the other decreases and vice versa. As noticed by Embrechts et al. (2003) and Wu (2013a), the application of the conventional Pearson's correlation coefficient ρ to describe such cross-correlation characteristics has been criticised. Copulas have proved to be a powerful tool when the observed data are limited and the dependence structures of the two components are complex (McNeil et al., 2005), which will be discussed briefly below.

2.2. Copulas

Considering a pair of 2-dimensional variables, $u_1 \in [0, 1]$ and $u_2 \in [0, 1]$, as an example, a copula $C(u_1, u_2)$ is defined conventionally as a bivariate cumulative distribution function with uniform marginal distributions. A probabilistic method for defining this copula is based on Sklar's Theorem (Sklar, 1959). For random variables Z_1 and Z_2 with continuous distribution functions $F_{Z_1} = P(Z_1 \le z_1)$ and $F_{Z_2} = P(Z_2 \le z_2)$, respectively, which can represent soil strength pairs, Sklar's Theorem states that a 2-dimensional distribution function $H_c(z_1, z_2)$ of the random vector (Z_1, Z_2) is determined by a unique copula function C, written as

$$H_c(z_1, z_2) = C\left(F_{Z_1}(z_1), F_{Z_2}(z_2)\right) \tag{1}$$

To construct the copula function, a corollary of Sklar's Theorem can be applied, whereby the copula is represented as a 2-dimensional distribution function with continuous marginal distributions (because their integral transforms are uniform distributions) and evaluated at the inverse functions or quantile function $F_{Z_1}^{-1}(u_1)$ and $F_{Z_2}^{-1}(u_2)$ (Wu, 2015a), defined as $F_{Z_1}[F_{Z_1}^{-1}(z_1)] = z_1$, i.e.,

$$\begin{aligned} C(u_1, u_2) &= H_c\left(F_{z_1}^{-1}(u_1), F_{z_2}^{-1}(u_2)\right) \\ &= P\left\{F_{Z_1}(z_1) \le u_1, F_{Z_2}(z_2) \le u_2\right\} \end{aligned} \tag{2}$$

By taking the derivative of Eq. (2), the joint probability density function is obtained as

$$f(z_1, z_2) = c\left(F_{Z_1}(z_1), F_{Z_2}(z_2) \middle| \rho \text{ or } \theta\right) f_{Z_1}(z_1) f_{Z_2}(z_2) \tag{3}$$

where $c(F_{Z_1}(z_1), F_{Z_2}(z_2)) = \partial^2 C(F_{Z_1}(z_1), F_{Z_2}(z_2)) / \partial F_{Z_1}(z_1) \partial F_{Z_2}(z_2)$, ρ is Pearson's correlation coefficient, θ is the copula dependence parameter, and $f_{Z_1}(z_1)$ and $f_{Z_2}(z_2)$ are the marginal probability density functions.

Having associated n pairs of observations, $(Z_1{}^i, Z_2{}^i), i = 1$ to n, copulas provide a convenient way to separately fit each variable to a marginal distribution and then join them together. As discussed below, multivariate normality is only one option in a wide range of copula-based models that can capture the principal features of shear strength data, such as non-symmetry, nonlinear dependence, or heavy-tail behaviour. There are numerous different copulas to choose from, with various correlation properties such as symmetry, tail dependence, and range of dependence (Joe, 1997; Nelsen, 2006). Considering the correlation characteristics between soil strength parameters, we chose the normal copula and Student's t copula from the elliptical class of copulas, the Clayton, Frank, and Gumbel copulas from the Archimedean family, and the Plackett copula, which is in a class of its own, as summarised in McNeil et al. (2005) and Wu (2013a). Some of these copulas may not allow negative correlations, but negating the values of one variable can produce a positive correlation. Nelsen (2006) has commented on choosing copulas and given additional details about them.

A probability density contour (PDC) is usually defined to characterise the different spread patterns of the observed data in all directions, which can visualise the dependence implications of the different models, as described in the literature (Zani et al., 1998; Wu, 2013a). Fig. 1 shows the PDC of bivariate distributions with the identical marginal distributions, but different copulas. These distributions all have a standard normal as both marginal distributions, and the parameters of copulas are chosen to give ρ is -0.5. Therefore, joint distributions of soil shear strength with the identical marginal distributions and the same correlation coefficient may also exhibit fundamentally different properties depending on the copula family selected to represent their dependence structures. In addition, the one-standard-deviation ellipse with the different ρ (-0.5, 0, 0.5) is shown centred at the mean (0, 0), indicating the ellipse is tilted with ρ. Consequently, such differences in PDC may lead to the different results in the failure probability of geostructures.

The normal copula is the dependence function of a joint normal distribution, but a normal copula does not necessarily restore a joint normal distribution unless the marginal distributions are also normal (Clemen and Reilly, 1999; Lambert and Vandenhende, 2002). Student's t copula has two parameters:

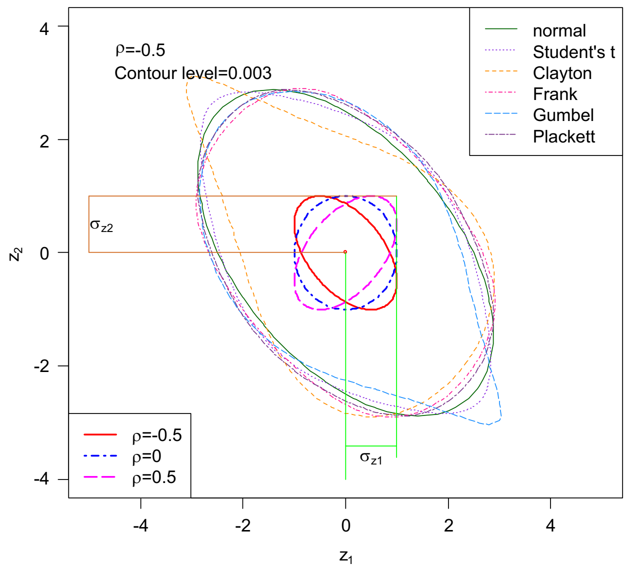

Fig. 1. PDCs of bivariate distributions with the same marginal distributions but different copulas. All marginal distributions are standardised normal (mean $\mu_{z1}=0$ and standard deviation $\sigma_{z1}=1$; $\mu_{z2}=0$ and $\sigma_{z2}=1$). The parameters of the copulas are chosen to give a $\rho=-0.5$. One standard deviation (ellipse) is overlapped with different ρ.

one corresponds to the dependence parameter and the other corresponds to the degrees of freedom λ (>0). The number of degrees of freedom controls the heaviness of the tails, and as it increases, Student's *t* copula approaches the normal one. Both the normal and Student's t copulas are symmetrical, and the normal copula is a limiting case of the Student-*t* copula when λ is infinite. The λ will not be considered as a parameter to be estimated for simplicity and its value is fixed to 4 (e.g., Embrechts et al., 2003; Moosbrucker, 2006; Yan, 2007; Silva and Lopes, 2008). This value is approximately the estimated degrees of freedom from all the compiled data in this study. The advantage of the Student's *t* copula is that it can capture the lower and upper tail dependence of data, i.e., the joint non-exceedance and the exceedance probabilities for rare events (see McNeil et al. (2005) for details). For copulas from the Archimedean family, the generator function solely characterises the dependence structure of random variables and is often described by a univariate function with model parameter θ. In this family, the Clayton copula usually shows a strong association in the left tail, while the Gumbel copula shows a strong association in the right tail. In this sense, the Clayton and Gumbel copulas describe the asymmetric dependence, while no clear association in the tails can be observed for the Frank copula. The Plackett copula is the most well-known example of an algebraically constructed copula, and association θ is determined by the odds-ratio based on observed frequencies in the four quadrants rather than on the correlation of random variables (Nelsen, 2006).

2.3. Estimation of marginal distributions

One of the key steps in fitting copulas for shear strength characteristics lies in the determination of the appropriate marginal distribution for each strength component. The soil strength typically shows large scattering, and Lumb (1970) demonstrated that normal distributions are suitable for soil strength properties. The applicability of the normal distribution of soil properties is also supported by Tobutt (1982), Liang et al. (1999), and Baecher and Christian (2003). Several researchers (Brejda et al., 2000; Fenton and Griffiths, 2003) noted that most soil properties are strictly non-negative, and in that case, a log-normal distribution is more suitable than a normal distribution. Other distributions, such as triangular, versatile beta, and generalised gamma distributions are gaining popularity (Baecher and Christian, 2003). The gamma distribution does not accept negative values, which is suitable for soil applications. The marginal distribution functions tested in this study are the normal, log-normal, Gumbel, Weibull, and gamma distributions. For the expressions of the probability density functions of these distributions and their domains, readers are asked to refer to Montgomery and Runger (1999) or Ang and Tang (2007). Selection of the best soil strength parameter model can be established by goodness-of-fit diagnostics. The best-fitting criteria for marginal distributions are identified initially by the Anderson–Darling (Anderson and Darling, 1954; AD) test. However, because the AD test does not account for the estimated number of parameters, the Bayesian Information Criterion (BIC) values are considered. The BIC measure is given by

$$\begin{aligned}\mathrm{BIC} = & -2\times\ \ln(\text{maximized likelihood for the model}) \\ & +\ln(\text{the sample size})\times \text{number of fitted parameters}\end{aligned} \tag{4}$$

A smaller BIC value corresponds to a better fit. In addition, the well-established Akaike Information Criterion (Akaike, 1974; AIC) is used commonly in the statistical literature. The relative merits of these criteria are discussed by Burnham and Anderson (2004).

2.4. Copula parameter estimation

The next step in fitting copulas involves estimating the parameters of the copula. The inference functions for marginal distribution method (IFM) proposed by Joe (1997) is employed for this purpose, which estimates the marginal parameters in the first stage. Given observed data Z_j, the parametric probability distribution can be determined by taking the logarithm of the likelihood function, namely,

$$L(\boldsymbol{\eta}_j) = \sum_{i=1}^{n}\left\{\log\left[f_{Z_j}(Z_{ji}\middle|\boldsymbol{\eta}_j)\right]\right\} \tag{5}$$

where $\boldsymbol{\eta}_j$ is the vector of the parameters for marginal distribution F_{Z_j}, $j=1$ and 2.

In the second stage, given estimates for $\hat{\boldsymbol{\eta}}_1$ and $\hat{\boldsymbol{\eta}}_2$, the unknown copula parameter θ of a specified parametric copula

$C(z_1, z_2; \theta)$ is determined using the following log-likelihood:

$$L(z_1, z_2; \theta) = \sum_{i=1}^{n} \log\left\{ c\left(\hat{F}_{Z_1}(z_{1i}; \boldsymbol{\eta}_1), \hat{F}_{Z_2}(z_{2i}; \boldsymbol{\eta}_2); \theta \right) \right\} \tag{6}$$

The IFM is a fully parametric method; and thus, the misspecification of the marginal distributions may affect the performance of this estimator. Consequently, the joint distribution is estimated expeditiously by this two-step procedure where the marginal distributions and the parameters in the copula function are estimated using these maximum likelihood values individually. As noted by Joe (2005), use of the technique can prevent computational difficulty for high dimensional models in the maximum likelihood estimation where the marginal distribution parameters and the copula parameters are estimated simultaneously. To quantitatively identify the appropriate copula for a set of observed data, the calculated values of the BIC are compared. The copula with the smallest BIC value can be considered as the best-fitting copula.

2.5. Implementation in R

The procedures for estimating the distribution parameters of the above-mentioned copulas, implemented in R (R Development Core Team, 2013), can be summarised by the following steps:

[1] Estimates of parameters $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ of marginal distributions F_{Z_1} and F_{Z_2} are obtained by separately maximising the corresponding likelihoods specified in Eq. (5) of the univariate marginals. The package, 'fitdistrplus', is employed in choosing and fitting a parametric univariate distribution to a dataset (Pouillot and Delignette-Muller, 2010).

[2] The cumulative distribution function of each marginal distribution is used to transform the observations into pseudo-observations with uniform marginal distributions. The copula parameter θ is estimated by fitting empirical distribution functions to the marginals and maximum likelihood to the copula model, as given in Eq. (6). In fact, the model parameters, including the marginal and copula ones, are estimated by the two-step estimation procedure (first marginals, then dependence). The R package, 'copula', is utilised to facilitate building multivariate modelling for fitting copulas (Yan, 2007; Yan and Kojadinovic, 2010). In this package, the command 'fitCopula' facilitates the goodness-of-fit testing, by imposing the maximum likelihood estimation, and returns the estimation of the unknown copula parameter. An 'mvdc' class is designed to define multivariate distributions via copula (MVDC), which contains the following items: (1) copula name, (2) copula parameter, and (3) marginal parametric distributions for all variables and their estimated parameters. Next, the simulated random variables can be generated easily through a function of 'rmvdc' for the 'mvdc' object if the size of the simulation is assigned. A function of 'dmvdc' is then applied in conjunction with drawing contour plots, which determines the joint probability density in Eq. (3) of cohesion and friction angle based on the 'mvdc' object. Here, the name of the 'mvdc' is prefixed with [r] for random number generation and [d] for density. For more insight into the embedded algorithms and illustrative examples of these functions, interested readers may consult Yan (2007).

These packages are free and open sources, which can help engineers facilitate statistical and reliability-based analyses with ease.

3. Modelling dependence structures of soil shear strength data

3.1. Data sets

In this study, 16 datasets collected from seven journals and two conference proceedings are used to examine the marginal distributions to interpret the true nature of the dependence structures and to fit the joint behaviour of the soil shear strength. The source data used here (consisting of 391 samples) are accessible via an R package 'GeoRiskR' released by Wu (2015b). Each dataset consists of a number of observations within the same soil horizons. A summary of the compiled tests, including the laboratory testing procedures, conditions, soil characteristics, sampling methods, and the original applications, is listed in Table 1, as explained below.

3.1.1. Matera Blue clay, Italy

Cherubini (2000) reported 16 paired data for c and ϕ by drained triaxial tests on Matera Blue clay, as shown in Fig. 2, which was used to perform a probabilistic analysis of the bearing capacity problem.

3.1.2. Dublin Boulder clay, Ireland

Forrest and Orr (2010) presented the relationship of c and ϕ of 22 triaxial tests performed on Dublin Boulder clay (DBC), as shown in Fig. 2, which were sampled during the construction of the Dublin Port Tunnel. This dataset was used to assess the reliability index of a square foundation.

3.1.3. Airport embankment, Japan

In laboratory experiments, Hata et al. (2008) investigated in-situ soil with the triaxial unconsolidated–undrained tests for embankments at airport C. The relationships between c and ϕ are shown in Fig. 2, which illustrates that the variation in cohesion is larger than the variation in inner friction angles. Considering the heterogeneity of the soil strength, a seismic reliability analysis of the residual displacements was performed using the Newmark sliding block method through Monte Carlo simulations.

3.1.4. Mal dam, Hungary

Kádár (2013) collected a series of shear strength of the fly ash and high plasticity clay from the Mal tailing dam, as illustrated in Fig. 3, whose tests were performed by the direct shear box.

3.1.5. Decomposed granite or volcanic rocks

Lumb (1970) reported an average correlation coefficient -0.24 for the data from strength tests on residual decomposed

Table 1
Summary of the compiled tests, including the test methods, conditions, soil characteristics, sampling method, and their original application.

Case	Size	Test method	Test condition	Soil characteristic	Sampling method	Original application	Source (Journal or Proceeding)
Cherubini MBC	16	Triaxial	Drained	Matera Blue clay	–	Footing bearing-capacity	Canadian Geotechnical Journal
Forrest & Orr DBC	22	Triaxial	–	Dublin Boulder clay	Undisturbed	Foundation bearing-capacity	Georisk
Hata et al. Airport C	23	Triaxial	Unconsolidated-undrained	Sand mixed gravel	–	Embankment slope stability	Georisk
Kadar Flash-Ash	20	Direct shear	–	Emissions from coal-fired power plant	–	Dam stability	Proceedings of the Second Conference of Junior Researchers in Civil Engineering
Kadar High-plasticity	19	Direct shear		High plasticity clay			
Lumb BL-1	55	Triaxial	Consolidated-undrained	Clayey coarse sand	Disturbed	Compacted fill in embankment and earth dam	Canadian Geotechnical Journal
Lumb BL-2	45						
Lumb BL-3	17						
M & K Soil 1	22	Triaxial	Unconsolidated-undrained	Unsaturated silty clay	Sampled from Watarase river bank	Sliding stability	Soils and Foundations
M & K Soil 2	41			Unsaturated silt, sandy silt			
M & K 10%	15	Direct shear	–	Silty clay	–	Effect of moisture content on the correlation	
M & K 15%	15						
M & K 20%	16						
Onodera et al. Masado	19	Direct shear	Consolidated drained or undrained	Clayey and silty Masado	Undisturbed	Shear strength	Soils and Foundations
Schultze Rhineland	23	NA	–	Clayey sand	Disturbed	Slope stability	ICASP-1
Speedie Melbourne	23	Direct shear	–	–	Undisturbed	Stability of borrow pits	New Zealand Engineering

Note: – denotes for not addressed. ICASP-1 denotes for the Proceedings of the 1st international conference on applications of statistics and probability in soil and structural engineering.

granite from Hong Kong, Penang (Malaysia), and Snowy Mountains (Australia), and on residual decomposed volcanic rocks from Hong Kong. Fig. 4 shows the values of the observed c and ϕ for three types of soil at Class B sites.

3.1.6. Unsaturated soils, Japan

Matsuo and Kuroda (1974; denoted by M & K) investigated the statistical characteristics of several types of unsaturated soil from embankments in Japan. The dataset of c against ϕ is re-plotted and shown in Fig. 5 (soils 1 and 2 were used for the triaxial compression tests under the unconsolidated–undrained condition) and Fig. 6 (soil 4 was tested by the direct shear apparatus under different water contents, 10%, 15%, and 20%). Moreover, the probability of sliding failure of the embankment was presented under consideration of the randomness of the soil strength properties.

3.1.7. Masado, Japan

To explore the influence of the porosity on the shear strength, Onodera et al. (1976) reported the strength tests on 19 undisturbed residual soils (named Masado in Japan) with their improved direct shear apparatus, as shown in Fig. 3.

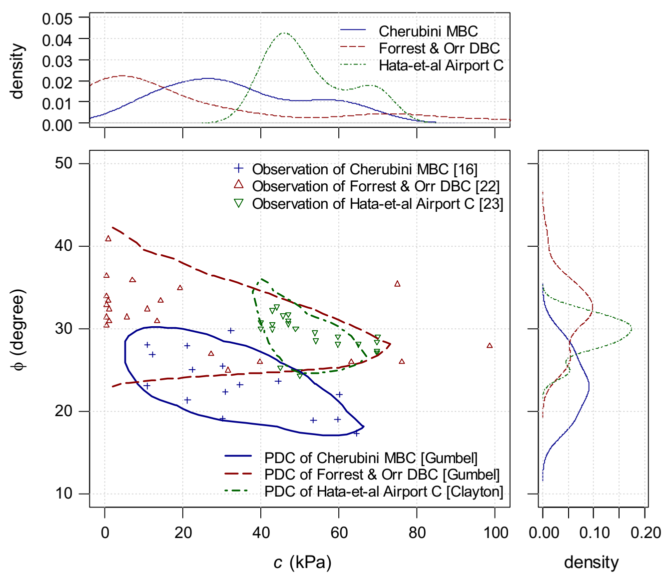

Fig. 2. Paired data of c and ϕ for observations adapted from Cherubini (2000), Forrest and Orr (2010), and Hata et al. (2008).

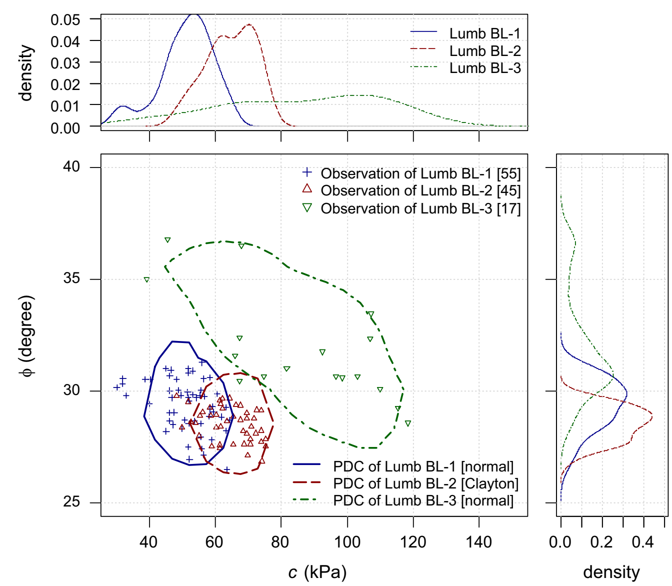

Fig. 4. Paired data of c and ϕ of soil BL-1, BL-2, and BL-3 (obtained from Lumb, 1970).

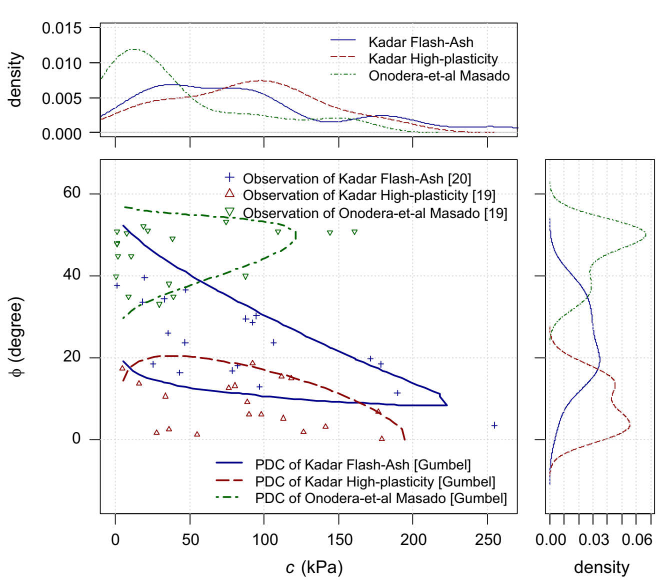

Fig. 3. Paired data of c and ϕ for soils at Mal tailing dams in Hungary (obtained from Kádár, 2013) and for Masado in Japan (adopted from Onodera et al., 1976).

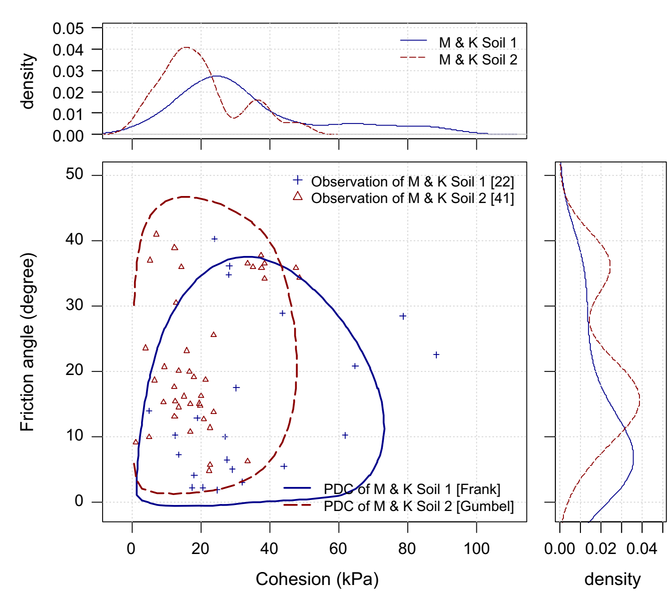

Fig. 5. Paired data of c and ϕ for soils 1 and 2 (after Matsuo and Kuroda, 1974).

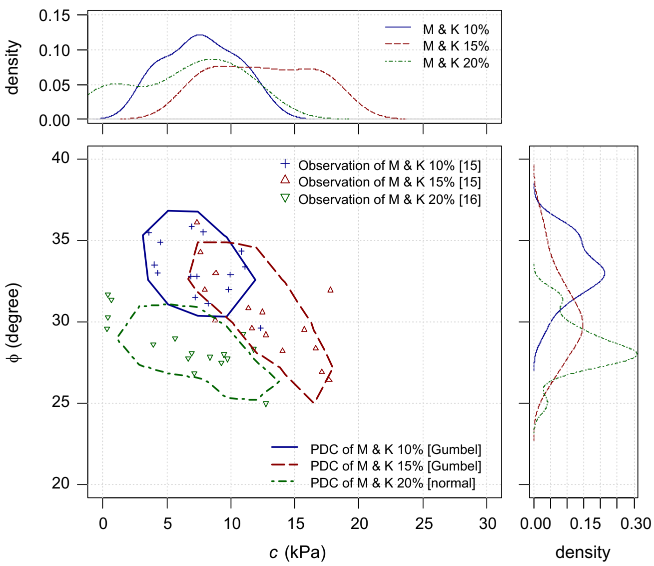

Fig. 6. Paired data of c and ϕ for soil 4 under different water contents (after Matsuo and Kuroda, 1974).

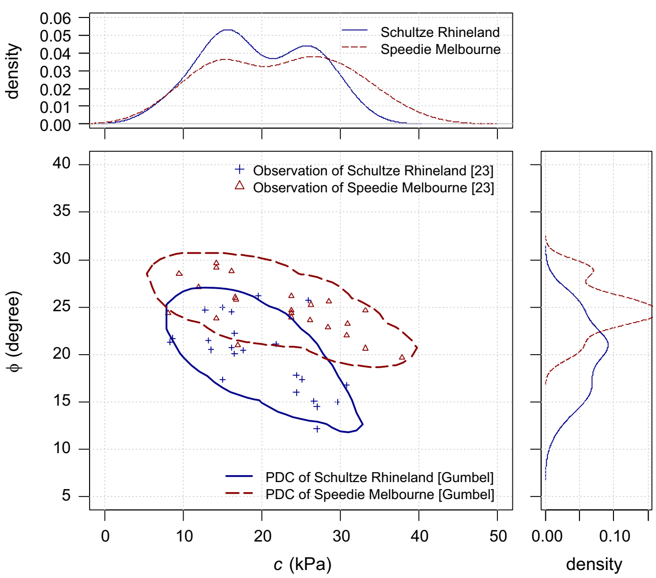

Fig. 7. Paired data of c and ϕ for soils in Germany (after Schultze, 1971) and Australia (after Speedie, 1955).

3.1.8. Rhineland, Germany

Schultze (1971) collected numerous samples from several regions of the Rhineland to investigate the frequency distribution and the correlation of the soil properties. The scattered shearing strength parameters used for the bank stability and settlement analyses are illustrated in Fig. 7.

3.1.9. Borrow pits, Australia

To select design values for shear strength against slope sliding, Speedie (1955) reported his 23 tests on soils from excavated borrow pits to determine their variations, as shown in Fig. 7. The importance of considering the wide range in values in the planning of major structures was emphasised.

3.2. Marginal distributions and their correlation coefficients

The mean and standard deviation of the above data are calculated and listed in Table 2. The variation is obviously larger for c than for ϕ.

Several types of distributions, including the normal, log-normal, Gumbel, Weibull, and generalised gamma, are evaluated as models to individually fit c and ϕ data from these samples. The kernel density curves of these measurements are indicated along with their scatter graphs in Figs. 2–7 using the method given by Sheather and Jones (1991). The density curve depends largely on the choice of the bandwidth parameter; and thus, the default kernel bandwidth function is selected by an algorithm based on Silverman's rule-of-thumb (Silverman, 1986, page 48, Eq. (3.31)). The graphic evidence exhibits substantially the skewness of these curves in most cases. The presence of skewness in the distribution of shear strength implies that the standard deviation may not be an appropriate measure of their variabilities.

The values for BIC of these candidate marginal distributions are listed in Table 3. In these tables, the best-fitting is denoted by a symbol 'a', although the difference between the BIC results for each data set is not significant in some cases. The normal distribution cannot adequately provide the best-fitting in most cases, while the Weibull and generalised gamma distributions provide the best-fitting for some data sets. The estimated parameters for the best-fitting marginal distributions are given in Table 4. Obviously, the asymmetric marginal model performs substantially better than the normal and symmetric models in most cases.

In the studies for geotechnical engineering problems, Pearson's correlation coefficient ρ was most commonly used (Lumb, 1970; Matsuo and Kuroda, 1974; Cherubini, 2000). Compared with Pearson's ρ, Kendall's τ does not assume that the relationship between two random variables is linear, and that it measures the correspondence of rankings between correlated random variables. Hence, τ is more suitable for checking nonlinear co-movements in data, and it receives wide acceptance in the construction of copulas (Joe, 1997; Nelsen, 2006). The estimated correlation coefficients in terms of Pearson's ρ and Kendall's τ for each of the datasets are listed in Table 2. Regardless of the drainage conditions and the type of soil, most of these results are negative and some of them exhibit a strong dependence. For soil 4 by Matsuo and Kuroda (1974), under different water contents of 10%, 15%, and 20%, the values of the correlation coefficients increase as the water content is increased, and such a claim requires further investigation. However, the results of the correlation coefficients are positive for soils 1 and 2 by Matsuo and Kuroda (1974) and for Masado by Onodera et al. (1976). To quantify the statistical uncertainties associated with different sample sizes, 95% confidence intervals of the various correlation indices are given in Table 2. In most cases, the absolute values for Pearson's ρ are approximately 35% larger than the absolute values for Kendall's τ, as shown in Fig. 8, which can be explained by an expression that relates these two parameters in the case of a bivariate normal population (Frees and Valdez,

Table 2
Mean, standard deviation, and correlation coefficient of soils.

Case	c (kPa)		ϕ (deg.)		Correlation	
	Mean	Standard deviation	Mean	Standard deviation	ρ (95% interval)	τ (95% interval)
Cherubini MBC	33.66	17.94	23.37	3.68	−0.65 (−0.87,−0.23)	−0.43 (−0.76,0.09)
Forrest & Orr DBC	22.16	29.8	31.61	4.05	−0.51 (−0.76,−0.11)	−0.33 (−0.66,0.11)
Hata et al. Airport C	52.41	10.35	29.23	2.27	−0.42 (−0.7,0)	−0.41 (−0.71,−0.01)
Kadar Flash-Ash	85.02	67.3	23.88	9.75	−0.72 (−0.88,−0.4)	−0.47 (−0.75,−0.03)
Kadar High-plasticity	87.5	50.34	8.35	5.98	−0.28 (−0.65,0.2)	−0.21 (−0.6,0.28)
Lumb BL-1	51.17	8.14	29.4	1.14	−0.37 (−0.58,−0.12)	−0.24 (−0.47,0.03)
Lumb BL-2	64.44	7.36	28.53	0.76	−0.38 (−0.61,−0.1)	−0.24 (−0.5,0.06)
Lumb BL-3	85.73	24.24	31.89	2.35	−0.69 (−0.88,−0.31)	−0.47 (−0.78,0.01)
M & K Soil 1	33.53	21.7	14.76	12.18	0.32 (−0.12,0.65)	0.19 (−0.25,0.56)
M & K Soil 2	19.89	11.85	21.91	10.86	0.37 (0.07,0.61)	0.06 (−0.25,0.36)
M & K 10%	7.66	2.76	33.27	1.77	−0.47 (−0.79,0.05)	−0.3 (−0.7,0.26)
M & K 15%	12.51	3.85	30.49	2.66	−0.75 (−0.91,−0.38)	−0.66 (−0.87,−0.22)
M & K 20%	6.52	4.25	28.57	1.67	−0.78 (−0.92,−0.47)	−0.48 (−0.79,0.02)
Onodera et al. Masado	41.61	50.05	45.44	6.63	0.26 (−0.22,0.64)	0.13 (−0.35,0.55)
Schultze Rhineland	19.61	6.63	19.91	3.93	−0.6 (−0.81,−0.24)	−0.45 (−0.73,−0.04)
Speedie Melbourne	22.1	8.39	24.83	2.68	−0.65 (−0.84,−0.33)	−0.49 (−0.75,-0.1)

Table 3
BIC values of marginal distributions of c and ϕ.

Case	c					φ				
	Normal	Log-normal	Gumbel	Weibull	Gamma	Normal	Log-normal	Gumbel	Weibull	Gamma
Cherubini MBC	142.29	140.91	141.02	**140.24**[a]	140.27	91.61	91.62	92.41	92.12	**91.54**[a]
Forrest & Orr DBC	216.95	173.29	206.62	**173.27**[a]	173.34	129.15	129.21	130.8	130.81	**129.1**[a]
Hata Airport C	178.01	175.33	**173.52**[a]	180.01	176.08	108.17	109.36	114.57	**105.92**[a]	108.92
Kadar Flash-Ash	230.08	230.52	224.93	**222.54**[a]	222.98	152.51	160.03	155.86	**153.15**[a]	156.36
Kadar High-plasticity	207.69	215.45	208.78	**207.48**[a]	209.44	126.71	133.48	126.23	**123.76**[a]	124.36
Lumb BL-1	393.77	403.36	413.61	**387.82**[a]	399.65	177.58	178.75	192.23	**173.67**[a]	178.34
Lumb BL-2	313.93	316.38	323.29	**311.15**[a]	315.42	**109.36**[a]	109.64	116.3	109.37	109.54
Lumb BL-3	161.26	164.34	164.71	**160.38**[a]	162.94	81.95	80.93	**78.23**[a]	86.37	81.25
M & K Soil 1	202.99	195.19	195.63	196.32	**195.04**[a]	177.59	167.47	172.46	167.21	**167.1**[a]
M & K Soil 2	325.48	326.37	319.31	**318.4**[a]	319.36	318.36	316.93	315.86	**314.09**[a]	314.55
M & K 10%	77.42	78.13	78.23	**76.85**[a]	77.46	**64.06**[a]	64.29	66.94	64.09	64.2
M & K 15%	87.36	87.42	87.74	**86.89**[a]	87.15	76.32	75.82	**75.65**[a]	78.89	75.96
M & K 20%	**96.25**[a]	103.26	98.71	96.84	97.45	66.31	66.3	68.72	68.37	**66.28**[a]
Onodera et al. Masado	207.47	182.91	199.49	180.34	**179.93**[a]	130.64	132.63	135.43	**127.67**[a]	131.89
Schultze Rhineland	157.57	159.01	158.98	**156.85**[a]	157.84	133.43	134.71	136.45	133.06	**132.05**[a]
Speedie Melbourne	168.37	169.92	169.61	**167.43**[a]	168.55	**115.92**[a]	116.09	118.35	117.3	115.96

Note:
[a]Denotes best-fitting.

1998; Lindskog et al., 2003), e.g., $\tau_c = \frac{2}{\pi}\arcsin(\rho)$. In most cases, the computed values for τ_c, based on ρ, have some discrepancies with the values for τ obtained directly from the observations. This implies again that the joint distribution of shear strength holds non-normal characteristics.

3.3. Parameter estimate for fitted copulas

The estimated copula parameters θ of different copulas are presented in Table 5. As listed in this table, the BICs quantitatively distinguish the best copulas. For instance, the normal copula provides a much better fit of shear strength pairs for the datasets of soils BL-1 and BL-3 by Lumb (1970), while the Clayton copula captures the dependence feature of the strength pairs of soil BL-2 by Lumb (1970). Therefore, there is convincing evidence that dependence parameter θ is site-specific. It is seen that the normal and Frank copulas are identified as the best-fitting copulas in some cases. However, the Gumbel and Clayton copulas are suitable in some other cases, which presents further evidence of asymmetric dependence between the two parameters.

Furthermore, the PDC of the best-fitting copula is overlapped to the scattered plot, as shown in Figs. 2–7. The values of the density level for these PDCs are not shown individually,

Table 4
Estimated parameters of best-fitting marginal distributions.

Case	c (kPa)			ϕ (deg.)		
	Best-fitting	Parameters		Best-fitting	Parameters	
Cherubini MBC	Weibull	2.07	38.14	Gamma	42.54	1.82
Forrest & Orr DBC	Weibull	0.57	14.42	Gamma	63.25	2
Hata Airport C	Gumbel	47.76	7.51	Weibull	16.33	30.21
Kadar Flash-Ash	Weibull	1.22	90.4	Weibull	7.94	54.43
Kadar High-plasticity	Weibull	1.72	97.11	Weibull	1.2	8.78
Lumb BL-1	Weibull	7.94	54.43	Weibull	31.52	29.92
Lumb BL-2	Weibull	10.76	67.62	Normal	28.53	0.75
Lumb BL-3	Weibull	4.37	94.46	Gumbel	30.86	1.7
M & K Soil 1	Gamma	2.76	0.08	Gamma	1.42	0.1
M & K Soil 2	Weibull	1.75	22.34	Weibull	2.19	24.83
M & K 10%	Weibull	3.22	8.57	Normal	33.27	1.71
M & K 15%	Weibull	3.84	13.89	Gumbel	29.27	2.18
M & K 20%	Normal	6.52	4.12	Gamma	28.54	1.44
Onodera et al. Masado	Gamma	0.55	0.01	Weibull	9.43	48.12
Schultze Rhineland	Weibull	3.4	21.9	Gamma	5.99	21.49
Speedie Melbourne	Weibull	3.01	24.82	Normal	24.83	2.62

Note: normal (mean; standard deviation), log-normal (mean in log; standard deviation in log), Gumbel (a,b), Weibull (shape; scale), gamma (shape; rate).

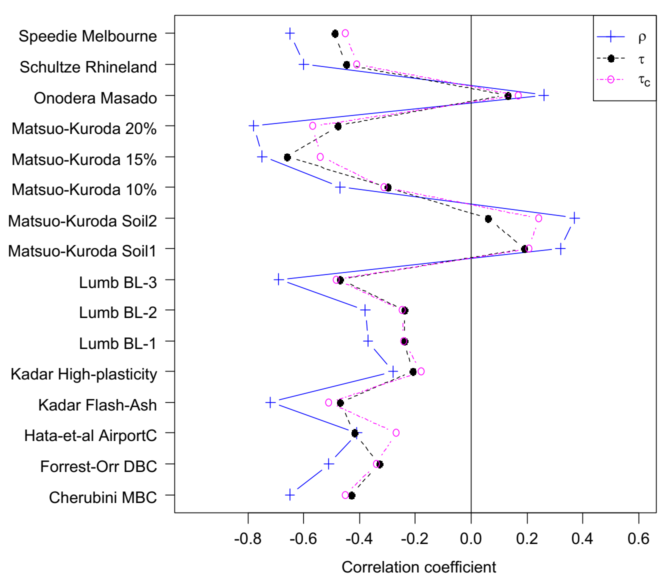

Fig. 8. The correlation coefficients (including ρ, τ, and τ_c) of c and ϕ.

which corresponds to a level that 95 percent observations fall in the region. To reveal the configurations of different copulas thoroughly, the PDCs of two soils are chosen randomly as illustrative examples.

For the soils reported by Hata et al. (2008), the mean values for c and ϕ are 52.41 kPa and 29.23°, respectively. The standard deviations of c and ϕ are 10.35 kPa and 2.27°, respectively. Correlation coefficient ρ is -0.42, as listed in Table 2. The Gumbel and Weibull are the best-fitting distributions for c and ϕ, respectively, as given in Table 4. Fig. 9 illustrates a comparison of the PDCs for the bivariate distribution of (c, ϕ) through different copulas. As shown in this figure, the PDC of the bivariate normal model (can be considered as a normal copula with the identical normal marginal distributions) is elliptical. The non-symmetric feature of the scattering pattern of the observed data indicates that the traditional multivariate linear model has difficulty providing a reasonable fit. The level curves of the density distribution through the copulas with the best-fitting marginal distributions have a significantly different shape. Graphically, the Clayton copula with the best-fitting marginal distributions provides a better fit of bivariate shear strength pairs than the fit by the other copulas, as chosen based on their BICs. This copula exhibits asymmetry in the sense that there is a clustering of values in the left tail of the joint distribution.

For the soils reported by Schultze (1971), the mean and standard deviation of c are 19.61 kPa and 6.63 kPa, respectively. The mean and standard deviation of ϕ are 19.91° and 3.93°, respectively. Correlation coefficient ρ is -0.6, as listed in Table 2. The Weibull and gamma are the best-fitting marginal distributions for c and ϕ, respectively, as given in Table 4. Fig. 10 illustrates several PDC plots under the best-fitting marginal distributions to fit the scattered data. The PDCs of most copulas are dissimilar, which implies that the choice of the copula significantly affects the resulting bivariate distribution. The Gumbel copula is the best-fitting copula on the basis of BICs, and numerous observations in the boundary regions make the other copula approaches less robust and less powerful in this situation. Graphically, this copula exhibits a strong clustering of values in the right tail.

Above all, it is safe to say that there are no universally accepted copulas for modelling the joint characteristics. The fitting results presented here are applicable only to the specific site conditions of each type of soil.

4. Influence of dependence structures on geotechnical reliability analysis

Two illustrative examples are presented to demonstrate the effect of the copula selection on the geotechnical reliability: an infinite slope filled with the soils reported by Hata et al. (2008) is evaluated with a probabilistic stability analysis against sliding, and a shallow foundation resting on the Rhineland soils reported by Schultze (1971) is discussed for a probabilistic analysis of its bearing capacity. For simplicity, only uncertainties in the soil shear strength parameters are considered, regardless of the fact that the geometries and the loading conditions are usually taken as random variables too.

4.1. Reliability analysis approaches

If a large number of samples from these bivariate copulas of shear strength is generated using the R function 'rmvdc', c and ϕ will then be mutually dependent uncertain variables. Finally, the probability of failure P_f can be estimated by the ratio of the running sum of the failed cases (performance function $g \leq 0$) m to the running sum of the total samples n_{sim}, i.e., $P_f = m/n_{sim}$, which has been denoted as the CBSM (Wu, 2013a). The technique follows a simple algorithm using the

Table 5
BIC and parameter values of fitted copulas.

Case	BIC						Copula parameter					
	Normal	Student's *t*	Clayton	Frank	Gumbel	Plackett	Normal	Student's *t*	Clayton[a]	Frank	Gumbel[a]	Plackett[a]
Cherubini MBC	−5.43	−5.52	−2	−5.02	**−7.11**[b]	−4.92	−0.62	−0.63	−1.51	−4.35	−1.82	−7.13
Forrest & Orr DBC	−4.69	−3.98	−0.44	−4.43	**−6.53**[b]	−4.25	−0.53	−0.52	−0.99	−3.39	−1.62	−4.51
Hata Airport C	0.2	3.01	**−5.07**[b]	−1.79	2.85	−0.98	−0.4	−0.38	−1.45	−3.06	−1.15	−3.51
Kadar Flash-Ash	**−10.36**[b]	−8.07	−3.45	−7.36	−9.6	−7.1	−0.7	−0.67	−1.77	−4.78	−1.89	−6.84
Kadar High-plasticity	1.35	1.16	1.51	2.13	**0.81**[b]	2.01	−0.28	−0.21	−0.53	−1.23	−1.27	−2.02
Lumb BL-1	**−4.37**[b]	3.66	−0.55	−2.8	−0.16	−1.84	−0.39	−0.34	−0.63	−2.22	−1.19	−2.49
Lumb BL-2	−3.82	−2.17	**−4.06**[b]	−2.55	−3.21	−2.32	−0.39	−0.37	−0.63	−2.06	−1.3	−2.63
Lumb BL-3	**−6.93**[b]	−5.59	−4.73	−4.66	−6.66	−4.66	−0.67	−0.59	−1.77	−4.46	−1.88	−6.54
M & K Soil 1	0.58	4.06	2.24	**0.3**[b]	1.67	0.82	0.33	0.31	0.47	1.97	1.18	2.19
M & K Soil 2	−0.04	−2.79	3.09	−0.09	**−4.04**[b]	−1.27	0.3	0.32	0.13	1.93	1.34	3.28
M & K 10%	−1.14	−0.33	0.34	0.12	**−1.36**[b]	0.35	−0.47	−0.43	−0.86	−2.33	−1.38	−2.76
M & K 15%	−7.92	−7.98	−5.36	−7.78	**−8.6**[b]	−8.49	−0.73	−0.74	−3.88	−6.21	−2.17	−13.33
M & K 20%	**−12.52**[b]	−11.91	−11.65	−11.55	−9.78	−11.95	−0.79	−0.78	−1.85	−7.34	−2.19	−15.85
Onodera et al. Masado	2.39	4.5	2.92	2.51	**2.06**[b]	2.56	0.16	0.14	0.30	0.75	1.15	1.39
Schultze Rhineland	−6.64	−6.68	−0.68	−8.5	**−9.63**[b]	−8.37	−0.56	−0.58	−1.64	−4.27	−1.76	−6.81
Speedie Melbourne	−9.31	−11.01	−5.57	−9.16	**−12.36**[b]	−10.73	−0.64	−0.67	−1.92	−4.98	−1.95	−11.1

Note:
[a]Denotes that the negative symbol in the parameter of the Clayton, Gumbel, Plackett copulas is obtained through negating the values of one variable and then to achieve a positive correlation.
[b]Denotes best-fitting.

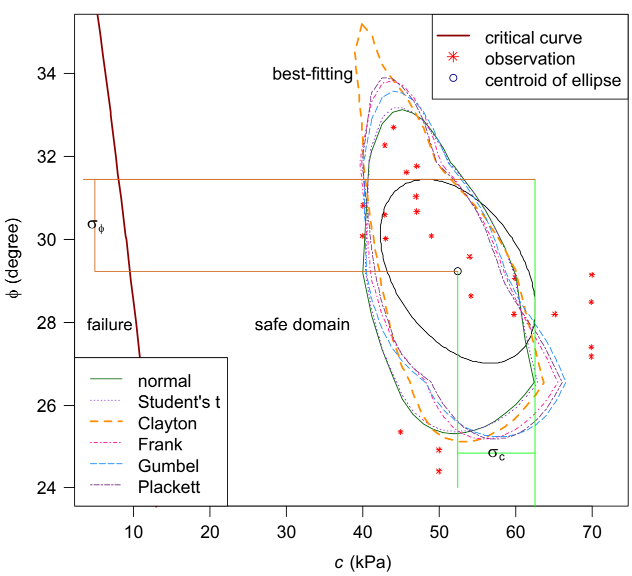

Fig. 9. Geometric interpretation of the various copulas to fitting the observed data by Hata et al. (2008).

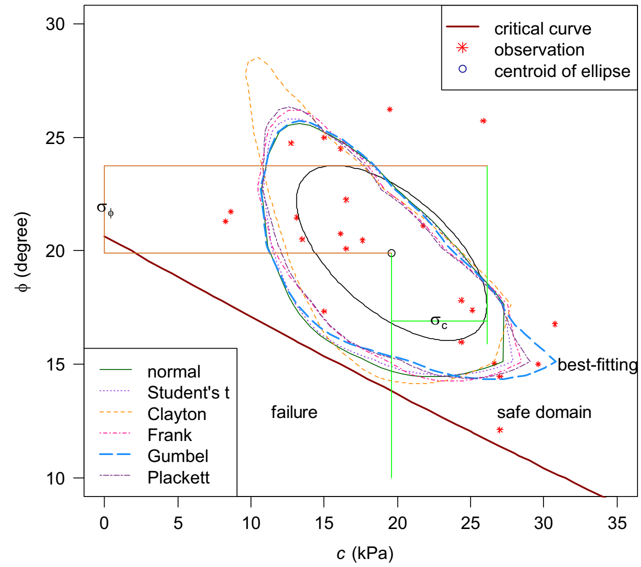

Fig. 10. Geometric interpretation of the various copulas to fitting the observed data by Schultze (1971).

Monte Carlo simulation method to generate random variables and integrates the copula function to fit the observed results for the soil strength parameters. In general, when the simulating number n_{sim} is more than $100/P_f$, the accuracy may be satisfactory (Tobutt, 1982; Husein Malkawi et al., 2000), and the probability of failure P_f can represent a deterministic solution. In the following analysis, after 50,000 trials, convergence of the results can be achieved. The simulation-based method is a computationally expansive tool for a reliability assessment of large structures (Tobutt, 1982). The calculation of P_f can appear difficult especially when its value is low. To avoid this problem, once the simulated factors of safety F_s are assumed to follow the normal or lognormal distribution, the reliability index β can be evaluated in terms of the statistics of F_s, as detailed in the literature (Baecher and Christian, 2003; Wu, 2013a).

For the sake of comparison, the first-order reliability method (FORM) and the second-order reliability method (SORM), including the formulae proposed by Breitung (1984) and Tvedt (1990), are used to compute the reliability index. Additional

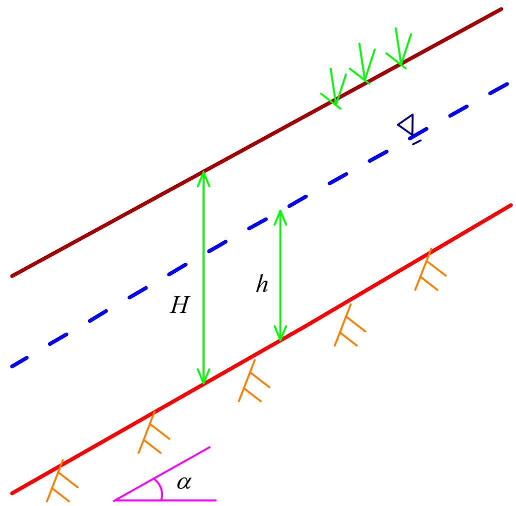

Fig. 11. Geometry of an infinite slope.

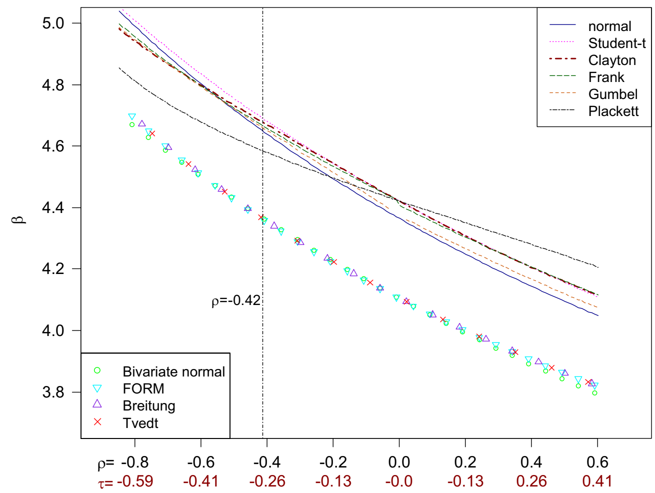

Fig. 12. Reliability index under various joint distributions for the infinite slope: lines describe β computed from different copulas with the best-fitting marginal distributions while symbols represent the computed values of β from different methods with normal marginal distributions.

details of the FORM and SORM approaches can be found elsewhere (Zhao and Ono, 1999; Baecher and Christian, 2003). In these analytical approximation methods, the original problem is transformed firstly into a standard normal probability space and then the probability of failure is calculated by converting the original hyperplane failure surface into the tangential or quadratic approximation. The reliability index β in this space is defined as the shortest distance between the hyperplane and the origin (Hasofer and Lind, 1974). Subsequently, the failure probability can then be approximately estimated from β (Hohenbichler and Rackwitz, 1983) as

$$P_f = 1 - \Phi(\beta) = \Phi(-\beta) \tag{7}$$

where Φ represents the cumulative distribution of the standard normal variate. More refined alternatives, such as the importance sampling simulation technique (Melchers, 1989), the copula-based direct integration method (Li et al., 2012), and the response surface method (Bucher and Bourgund, 1990), can also be adopted to determine the failure probability.

Furthermore, using these approaches, a parametric study on the resulting reliability index of these geostructures will be performed, as the copula parameters (e.g., correlation coefficient and copula type) are varied when fitting the soil shear strength.

4.2. An infinite slope

As the first example, Fig. 11 shows the geometry of an infinite slope. The classical performance function for the stability of this slope with a stable groundwater table is written as (Wu and Abdel-Latif, 2000; Zhang et al., 2011)

$$g = \frac{c + \left[\gamma_m (H - h_w) + (\gamma_{sat} - \gamma_w) h_w\right] \cos^2\alpha \tan\varphi}{\left[\gamma_m (H - h_w) + \gamma_{sat} h_w\right] \sin\alpha \cos\alpha} - 1.0 \tag{8}$$

where α is the inclination of the slope (deg.), H is the thickness of the slope (m), h_w is the level of groundwater (m), and γ_w, γ_m, and γ_{sat} are the unit weights of water, soil, and saturated soil above the water table, respectively. The values adopted for γ_w, γ_m, and γ_{sat} are 9.8, 17, and 19 kN/m^3, respectively. Slope angle α is supposed to be 35°, $H = 3$ m, and $h_w = 1.22$ m (Zhang et al., 2011).

The shear strength parameters of the soils are adopted from Hata et al. (2008) for the sand mixed gravel at an actual airport embankment. Clearly, shear strength parameter ϕ is expressed by a nonlinear relation in this performance function. The performance function can be drawn explicated in the plane of c and ϕ as a critical curve, shown in Fig. 9 with the thick solid line. Wu (2013a, 2015c) presented relatively simple procedures for the definition of this curve. The values for β, evaluated using CBSM for various copulas with ρ from -0.85 to 0.6 between soil shear strengths, are presented in Fig. 12. The range in τ (from -0.65 to 0.41) is determined from ρ and illustrated at the bottom of the horizontal axis. These copula parameters of the soil (Hata et al., 2008) are given in Table 5. A vertical line is drawn where the observed correlation ($\rho = -0.42$) is given in Table 2. Even when these copulas are imposed by the same marginal distributions (best-fitting as listed in Table 4 and described in Section 3.3) and the same correlation coefficient, the discrepancy between the computed reliability indices is apparent, as shown in Fig. 12. These differences resulted from the scatter in the shapes of the observations and the corresponding PDCs. This necessitates additional extensive research to gather more reliable experimental data and to explore their dependence structures. It can be seen that the joint normal copula overestimates the reliability index and the Gumbel copula leads to underestimating the reliability index, which shows some agreement with the results of other investigators, such as Tang et al. (2013).

In the following analyses, c and ϕ are further assumed to be normally distributed. The mean and standard deviations of c and ϕ of the soil (Hata et al., 2008) are given in Table 2. The reliability index given by CBSM (denoted as bivariate normal)

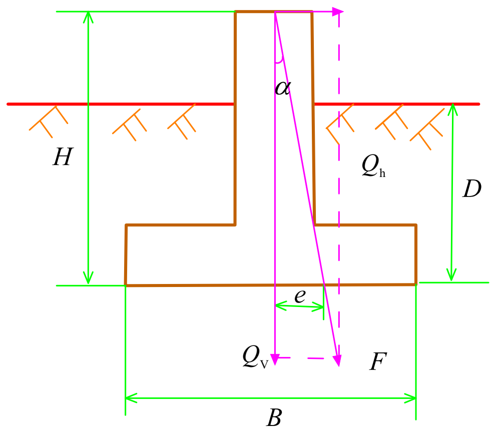

Fig. 13. Geometry of shallow foundation.

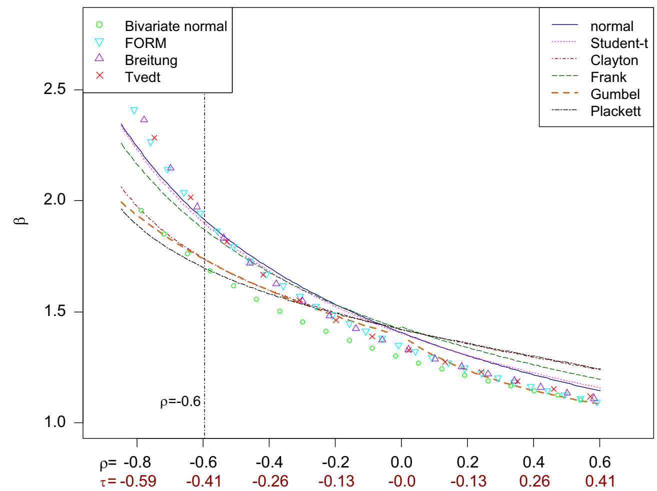

Fig. 14. Reliability index under various joint distributions for the shallow foundation: lines describe the computed values of β from different copulas with the best-fitting marginal distributions while symbols represent the computed values of β using different methods with normal marginal distributions.

is consistently validated against the results obtained from the analytical methods, e.g., FORM and SORM (denoted as Breitung and Tvedt individually), despite the fact that the results from CBSM are slightly smaller. The matched results prove that it works well in determining the reliability index using CBSM. Obviously, the values for β are smaller under the bivariate normal distribution than the results for β where the joint distribution follows the best-fitting Clayton copula.

In addition, the analysis of these results illustrates the importance of the value of β relative to the value of ρ between c and ϕ. A small variation in the correlation coefficients for c and ϕ has a large impact on the β of the slope stability in this example, due to the fact that the tilting angle of the PDCs changes with ρ. An evaluation of the effect of cross-correlation on β by CBSM has shown some similar conclusions with the results of other investigators, such as Nguyen (1985) and Wu (2013a).

4.3. A shallow foundation

The reliability assessment of the bearing capacity for a shallow foundation is considered as another example problem, as shown in Fig. 13 for its geometry. A concrete foundation, 5 m wide (B) and 25 m long (L), is founded at a depth (D) of 1.8 m in a deposit of silty sand. This footing carries a horizontal load Q_h of 300 kN/m running at a point 2.5 m above the base (H) and a centrally applied vertical load Q_v of 1100 kN/m running in the middle. Inclined loading is applied with an eccentricity e_x with respect to the centroid of the foundation, thus, the effective breadth $B'=B-2e_x$. The performance function of the shallow foundation can be calculated as (Tomlinson, 1995) follows:

$$g = q_L - q_u \tag{9}$$

where q_L is the bearing pressure for the effective foundation width $\left(=\sqrt{Q_h^2+Q_v^2}/B'\right)$, and q_u is the ultimate bearing capacity, written as

$$q_u = \gamma D N_q s_q d_q i_q + c N_c s_c d_c i_c + 0.5 B' \gamma N_\gamma s_\gamma d_\gamma i_\gamma \tag{10}$$

where N_q, N_c, and N_γ are the bearing capacity factors according to the following expressions: $N_q = (\tan(45+\frac{\varphi}{2}))^2 e^{\pi \tan\varphi}$, $N_c = \frac{N_q - 1}{\tan\varphi}$, and $N_\gamma = 2(N_q+1)\tan\varphi$; i_q, i_c, and i_γ are the loading inclination factors with $i_q = 1-(\alpha/90)^2$, $i_c = i_q$, and $i_\gamma = (1-\alpha/\varphi)^2$; s_q, s_c, and s_γ are the shape factors with $s_q = 1+\sin\varphi i_q B'/L$, $s_c = 1+0.2 i_c B'/L$, and $s_\gamma = 1-0.4 i_\gamma B'/L$; d_q, d_c, and d_γ are the depth factors when $D<B'$, $d_c = 1+0.4D/B'$ and $d_q = 1+2\tan\varphi(1-\sin\varphi)^2 D/B'$; when $D \geq B'$, $d_c = 1+0.4\arctan(D/B')$ and $d_q = 1+2\tan\varphi(1-\sin\varphi)^2 \arctan(D/B')$; and $d_\gamma = 1$. The footing is assumed to be based on the soils investigated by Schultze (1971) who presented 23 shear strength parameters to consider their variabilities. The unit weight (γ) of soils is assumed to be a constant with 21 kN/m^3.

Although ϕ expressed in Eq. (10) implicates a higher nonlinear relation, the critical curve is flattened off to an almost linear relation in the plane of c and ϕ, as illustrated in Fig. 10.

Fig. 14 presents the computed reliability indices with different copulas (presented in Section 3.3) using CBSM for ρ in the range of -0.85 to 0.6 between soil shear strength. The values for β are significantly sensitive to ρ, and the values for β decrease as those for ρ increase. Plots with various copulas show a similar pattern, but β is influenced by the choice of copula owing to the asymmetry of their PDCs. The normal copula significantly overestimates β, while the Gumbel copula underestimates it. The computed reliability indices can also be compared to the case with the specified correlation coefficient ($\rho=-0.6$) of the soil (Schultze, 1971) given in Table 2, indicated by a vertical line in Fig. 14. Analytical solutions using FORM and SORM (denoted as Breitung and Tvedt individually) are also included in this figure to compute β under the assumption that both marginal distributions follow a normal one. The mean and standard deviations of this soil are listed in Table 2. Both analytical results agree well with the CBSM results (denoted as bivariate normal), especially for the positive values for ρ, while the results from CBSM are smaller than the ones using CBSM for the negative values for ρ. Note that the computed reliability indices using SORM are quite close to those obtained by FORM, as shown in Fig. 14. In other words, there is no significant improvement over SORM, potentially due to the fact that the performance function is not strongly nonlinear. Compared with the β obtained in the case of the best-fitting Gumbel copula, the smaller β is achieved under the bivariate normal distribution because of the different patterns of the PDC.

5. Discussion

The illustrative application demonstrates that the developed joint probability distribution of c and ϕ can be an important factor influencing the reliability assessment of damaged structures. The following steps are required for practical reliability analyses associated with a copula-based method: [1] Estimate the best-fitting marginal distribution. This requires a goodness-of-fit test for a dataset. [2] Evaluate the correlation coefficient of paired shear strength. A visual interpretation by the scatterplot can be drawn out to investigate the dependence characteristics in a graphical form. [3] Establish the best-fitting copula for a dataset through another goodness-of-fit test. [4] Incorporate the reliable copula-based simulations into the limit state equation to calculate the probability of failure or reliability index. [5] Evaluate the effect of the chosen copula. A comparison can be performed using the correlated shear strength parameters simulated with the other copulas.

The copula approach is applied to generate the correlated shear strength characteristics. With this tool, any dependence can easily be integrated into an existing probabilistic analysis model, because a copula procedure permits the description of the dependencies between many random variables, independent of their marginal distributions. Establishing these relationships will greatly advance our knowledge on how c and ϕ of the soil shear strength are integrated. The joint probability distribution and dependence structure is derived from real data so that once an observed dataset is available, its shear strength characteristics can be derived from this joint distribution, i.e., a set of simulated shear strength values can be drawn at random from a population with the corresponding joint distribution. The mature R packages are extremely useful for assisting civil engineers in identifying and facilitating such joint functions.

A strong asymmetry in the marginal probability density functions of the shear strength can be found in some datasets. For instance, small cohesion for the sandy clay is held in the case of Forrest and Orr (2010) and Onodera et al (1976). Or relatively small friction angles are investigated in the case of Kádár (2013) for high plasticity clay. Such a truncated normal, log-normal, or Weibull marginal distribution can provide a better fit to avoid unrealistic negative values of shear strength. Obviously, the configuration of the corresponding joint PDC of shear strength will be imposed lower bound that truncates the tail of the PDC in some occasions to ensure a strictly positive output.

Owing to the experimental cost of soil shear strength tests, the available number of observations is usually rather small, which can easily conceal the real nature of the correlations between variables. As shown in Table 1, the number of available observations is in the range of 15–55. Obviously, from a statistical point of view, the sample size is relatively small for exploring a multivariate problem. In practice, the number of tests, between 15 and 50, should be appropriate to achieve a reliable description on the complicated patterns of the variation in conjunction with soil shear strength, as described by Genest and Favre (2007) and Wang et al. (2010) for different aspects to identify the appropriate copula model using a small sample. Therefore, the collection of over 10 samples becomes the criterion for the datasets in 'GeoRiskR'. Apparently, high-quality shear strength data, when available, can provide additional refinement to the copula-based dependence structure between strength model parameters. This study indicates that there is a genuine need for further investigations either in the field or the laboratory on correlations of geotechnical parameters because the reliability index, and hence, the probability of failure, depends on the correlation structure characteristics of the variables.

Each pair of measurements should be collected randomly from locations or sites at a project contract level to make sure the samples independently drawn for the bivariate distribution, and thus, the samples for each dataset, are assumed to be statistically independent and identically distributed. However, the spatial soil shear strength data of testing sites a certain distance from each other have the tendency to be dependent, and such a spatial correlation should be avoided in order that the proposed procedure not be confounded (e.g., Marchant et al., 2011). Readers should refer to Fenton and Griffiths (2003) and Ching et al. (2014) for a comprehensive discussion of the spatial correlation between cohesion and friction angle, which falls outside the scope of this study.

Uncertainty in the shear strength parameters can arise from the loading conditions, the sampling techniques, and the procedural control maintained during laboratory or field testing. Once the inherent characteristics of a soil property at a given site (layer) tend to be similar, they will be treated as data from a homogeneous deposit, which can be the case for the compiled data here. In some occasions, the statistical characteristics of one single soil cannot be accommodated for the presence of different soil layers, and then the proposed statistical estimates of the variability of soil properties should be extended for multi-layer soils (Wu, 2013a).

6. Conclusions

The bivariate copula functions of soil shear strengths permit a non-symmetric dependence structure between strength components; and hence, it leads to accurate probabilistic assessments of geostructures if the correlations between the input variables are well understood. In this study, the joint distribution function of correlated soil strength properties for 16 sets of experimental observations is examined. The dependence structure between shear strength components is found to be mostly negative and asymmetric. The dependence structures are site-specific and may vary with soils, so it is difficult to identify a universally accepted copula for various soils. Computed reliability indices differ because of non-symmetry scattering in the observations under various copulas, even when the identical marginal distributions and the same correlation coefficients are adopted. For both illustrative examples, the normal copula leads to an overestimation of the reliability index, while the Gumbel copula achieves the lowest reliability index. In this context, information about the dependence structures of soil shear strength parameters must be identified correctly to perform a reliability assessment of geotechnical structures.

In both examples, although conservative reliability indices are obtained when the joint behaviour of soil shear strength follows a bivariate normal distribution, this distribution can be inspired by the convenience of modelling for a wide variety of soils. The negative correlation is demonstrated to improve the structural reliability index compared to ignoring the correlation (zero cross-correlation). The computed reliability index using SORM yields results close to that using FORM; thus, the real advantage of a SORM evaluation is not realised when the critical curve is approximately linear.

Acknowledgements

This research was partially supported by the National Key Basic Research Program of China as part of Grant no. 2013CB036405, for which the authors are grateful. The authors also greatly appreciate the investigators who generously made their source data on soil shear strength publicly available. Thanks are also extended to the referees for their comments which enhanced the presentation of this study.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control AC 19 (6), 716–723.

Alonso, E.E., 1976. Risk analysis of slopes and its application to slopes in Canadian sensitive clays. Geotechnique 26, 453–472.

Anderson, T.W., Darling, D.A., 1954. A test of goodness-of-fit. J. Am. Stat. Assoc. 49, 765–769.

Ang, A.H.-S., Tang, W.H., 2007. Probability concepts in engineering. Emphasis on Applications to Civil and Environmental Engineering, 2nd Ed. John Wiley and Sons, New York.

Baecher, G.B., Christian, J.T., 2003. Reliability and Statistics in Geotechnical Engineering. John Wiley and Sons, England.

Breitung, K., 1984. Asymptotic approximations for multinormal integrals. J. Eng. Mech. (ASCE) 110 (3), 357–366.

Brejda, J., Moorman, J., Smith, T.B., Karlen, J.L., Allan, D.L., Dao, T.H., 2000. Distribution and variability of surface soil properties at a regional scale. Soil Sci. Soc. Am. J. 64, 974–982.

Bucher, C.G., Bourgund, U., 1990. A fast and efficient response surface approach for structural reliability problems. Struct. Saf. 7, 57–66.

Burland, J.B., 1990. On the compressibility and shear strength of natural soils. Geotechnique 40 (3), 329–378.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociol. Methods Res. 33, 261–304.

Cherubini, C., 2000. Reliability evaluation of shallow foundation bearing capacity on c, ϕ soils. Can. Geotech. J. 37, 264–269.

Ching, J., Phoon, K.K., Kao, P.H., 2014. Mean and variance of mobilized shear strength for spatially variable soils under uniform stress states. J. Eng. Mech. 140 (3), 487–501.

Chowdhury, R.N., Xu, D.W., 1992. Reliability index for slope stability assessment-two methods compared. Reliab. Eng. Syst. Saf. 37 (2), 99–108.

Clemen, R.T., Reilly, T., 1999. Correlations and copulas for decision and risk analysis. Manag. Sci. 45 (2), 208–224.

Der Kiureghian, A., Liu, P., 1986. Structural reliability under incomplete probability information. J. Eng. Mech. 112 (1), 85–104.

Dutfoy, A., Lebrun, R., 2009. Practical approach to dependence modelling using copula. Proc. Inst. Mech. Eng., Part O: J. Risk Reliab. 223 (4), 347–361.

Embrechts, P., Lindskog, F., McNeil, A.J., 2003. Modelling dependence with copulas and applications to risk management. In: Rachev, S.T. (Ed.), Handbook of Heavy Tailed Distributions in Finance, vol. 1. Elsevier, Amsterdam, Netherlands, pp. 329–384.

Fellin, W., Oberguggenberger, M., 2012. Robust assessment of shear parameters from direct shear tests. Int. J. Reliab. Saf. 6 (1–3), 49–64.

Fenton, G.A., Griffiths, D.V., 2003. Bearing capacity prediction of spatially random c-ϕ soils. Can. Geotech. J. 40 (1), 54–65.

Forrest, W.S., Orr, T.L.L., 2010. Reliability of shallow foundations designed to Eurocode 7. Georisk: Assess. Manag. Risk Eng. Syst. Geohazards 4 (4), 186–207.

Frees, E.W., Valdez, E.A., 1998. Understanding relationships using copulas. N. Am. Actuar. J. 2 (1), 1–25.

Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modelling but were afraid to ask. J. Hydrol. Eng. 12 (4), 347–368.

Hasofer, A.M., Lind, N.C., 1974. Exact and invariant second-moment code format. J. Eng. Mech. ASCE 100, 111–121.

Hata, Y., Ichii, K., Tsuchida, T., Kano, S., Yamashita, N., 2008. A practical method for identifying parameters in the seismic design of embankments. Georisk: Assess. Manag. Risk Eng. Syst. Geohazards 2 (1), 28–40.

Hohenbichler, M., Rackwitz, R., 1983. First-order concepts in system reliability. Struct. Saf. 1, 177–188.

Husein Malkawi, A.I., Hassan, W.F., Abdulla, F., 2000. Uncertainty and reliability analysis applied to slope stability. Struct. Saf. J. 22, 161–187.

Joe, H., 1997. Multivariate Models and Dependence Concept. Chapman and Hall, New York.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. J. Multivar. Anal. 94 (2), 401–419.

Kádár, I., 2013. Some characteristic values of the stability analysis of MAL dams. Proceedings of the Second Conference of Junior Researchers in Civil Engineering. Budapest, Magyarország, 100–104.

Kulatilake, P.H.S.W., Fuenkajorn, K., 1987. Factor of safety of tetrahedral wedges: a probabilistic study. Int. J. Min. Reclam. Environ. 1 (2), 147–153.

Lambert, P., Vandenhende, F., 2002. A copula-based model for multivariate nonnormal longitudinal data: analysis of a dose titration safety study on a new antidepressant. Stat. Med. 21, 3197–3217.

Lee, Y.F., Chi, Y.Y., 2011. Rainfall-induced landslide risk at Lushan, Taiwan. Eng. Geol. 123, 113–121.

Li, D.Q., Tang, X.S., Zhou, C.B., Phoon, K.K., 2012. Uncertainty analysis of correlated non-normal geotechnical parameters using Gaussian copula. Sci. China Technol. Sci. 55 (11), 3081–3089.

Li, K.S., Lumb, P., 1987. Probabilistic design of slope. Can. Geotech. J. 24, 520–535.

Liang, R.Y., Nusier, O.K., Malkawi, A.H., 1999. Reliability based approach for evaluating the slope stability of embankment dams. Eng. Geol. 54 (3), 271–285.

Lindskog, F., McNeil, A., Schmock, U., 2003. Kendall's tau for elliptical distributions. In: Bol, Nakhaeizade (Ed.), Credit Risk—Measurement, Evaluation and Management. Physica-Verlag, Heidelberg.

Lumb, P., 1970. Safety factors and the probability distribution of soil strength. Can. Geotech. J. 7 (3), 225–242.

Marchant, B.P., Saby, N.P.A., Jolivet, C.C., Arrouays, D., Lark, R.M., 2011. Spatial prediction of soil properties with copulas. Geoderma 162, 327–334.

Matsuo, M., Kuroda, K., 1974. Probabilistic approach to design of embankments. Soils Found. 14 (2), 1–17.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton.

Melchers, R.E., 1989. Importance sampling for structural reliability. Struct. Saf. 6 (1), 3–10.

Montgomery, D.C., Runger, G.C., 1999. Applied Statistics and Probability for Engineers. Wiley-Interscience.

Moosbrucker, T., 2006. Pricing CDOs with correlated variance gamma distributions. J. Fixed Income 16 (2), 62–75.

Nataf, A., 1962. Détermination des distributions de probabilité dont les marges sontdonnées. Comp Rend. Acad. Sci. 225, 42–43.

Nelsen, R.B., 2006. An Introduction to Copulas, 2nd ed. Springer, New York.

Nguyen, V.U., 1985. Reliability index in geotechnics. Comput. Geotech. 1 (2), 117–138.

Onodera, T., Oda, M., Minami, K., 1976. Shear strength of undisturbed sample of decomposed granite soil Source data. Soils Found. 16 (1), 17–26.

Parker, C., Simon, A., Thorne, C.R., 2008. The effects of variability in bank material properties on riverbank stability: Goodwin Creek, Mississippi. Geomorphology 101, 533–543.

Pouillot, R., Delignette-Muller, M.L., 2010. Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. Int. J. Food Microbiol. 142, 330–340.

Poulin, A., Huard, D., Favre, A., Pugin, S., 2007. Importance of tail dependence in bivariate frequency analysis. J. Hydrol. Eng. ASCE 12 (4), 394–403.

R Development Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN:3-900051-07-0. ⟨http://www.R-project.org⟩.

Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R., 2007. Extremes in Nature. Springer: Dordecht, The Netherlands.

Sanchez Lizarraga, H., Lai, C.G., 2014. Mechanical behaviour and failure of cohesive granular materials. Soil Dyn. Earthq. Eng. 64, 113–128.

Schultze, E., 1971. Frequency distributions and correlations of soil properties. In: Lumb, P. (Ed.), Statistics and Probability in Civil Engineering (Proceedings of the 1st international conference on applications of statistics and probability in soil and structural engineering, ICASP-1). Hong Kong University Press, pp. 372–387.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. R. Stat. Soc. Ser. B 53 (3), 683–690.

Silva, R.S., Lopes, H.F., 2008. Copula, marginal distributions and model selection: a Bayesian note. Stat. Comput. 18, 313–320.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Sklar, A., 1959. Fonction de répartition à n dimensions et leurs marges, 8. Publications de L'Institute de Statistique, Université de Paris, 229–231.

Soubra, A.-H., Mao, N., 2012. Probabilistic analysis of obliquely loaded strip foundations. Soils Found. 53 (3), 524–538.

Speedie, M.G., 1955. Selection of design value from shear test results. N. Z. Eng. 10 (1), 377–378.

Tang, X.S., Li, D.Q., Rong, G., Phoon, K.K., Zhou, C.B., 2013. Impact of copula selection on geotechnical reliability under incomplete probability information. Comput. Geotech. 49, 264–278.

Tobutt, D.C., 1982. Monte Carlo simulation methods for slope stability. Comput. Geosci. 8 (2), 199–208.

Tobutt, D.C., Richards, E.A., 1979. The reliability of earth slopes. Int. J. Num. Anal. Methods Geomech. 3, 232–354.

Tomlinson, M.J., 1995. Foundation Design and Construction, 6th ed. Longman Scientific.

Tvedt, L., 1990. Distribution of quadratic forms in normal space: Application to structural reliability. J. Eng. Mech. (ASCE) 116 (6), 1183–1197.

Wang, X., Gebremichael, M., Yan, J., 2010. Weighted likelihood copula modeling of extreme rainfall events in Connecticut. J. Hydrol. 390, 108–115.

Wolff, T.F., 1985. Analysis and Design of Embankment Dam Slopes: A Probabilistic Approach. Purdue University, Lafayette, Ind Ph.D. thesis.

Wu, T.H., Abdel-Latif, M.A., 2000. Prediction and mapping of landslide hazard. Can. Geotech. J. 37, 781–795.

Wu, X.Z., 2013a. Probabilistic slope stability analysis by a copula-based sampling method. Comput. Geosci. 17 (5), 739–755.

Wu, X.Z., 2013b. Trivariate analysis of soil ranking-correlated characteristics and its application to probabilistic stability assessments in geotechnical engineering problems. Soils Found. 53 (4), 540–556.

Wu, X.Z., 2015a. Assessing the correlated performance functions of an engineering system via probabilistic analysis. Struct. Saf. 52 (PA), 10–19.

Wu, X.Z., 2015b. Package 'GeoRiskR'. Data and functions for Geotechnical Risk Assessment. Version 2.1. ⟨https://r-forge.r-project.org/R/?group_id=1777⟩.

Wu, X.Z., 2015c. Development of fragility functions for slope instability analysis. Landslides 12 (1), 165–175.

Yan, J., 2007. Enjoy the Joy of Copulas: With a Package copula. J. Stat. Softw. 21 (4), 1–21.

Yan, J., Kojadinovic, I., 2010. copula: Multivariate Dependence with Copulas. R package version 0.9-5, URL http://CRAN.R-project.org/package=copula.

Zani, S., Riani, M., Corbellini, A., 1998. Robust bivariate boxplots and multiple outlier detection. Comput. Stat. Data Anal. 28, 257–270.

Zhang, J., Zhang, L.M., Tang, W.H., 2011. Slope reliability analysis considering site-specific performance information. J. Geotech. Geoenviron. Eng. 137 (3), 227–238.

Zhao, Y.G., Ono, T., 1999. A general procedure for FORM/SORM. Struct. Saf. 21 (2), 95–112.