Proc. of the 6th Intl. Symposium on Reliability Engineering and Risk Management (6ISRERM)
31 May – 1 June 2018, Singapore
Editor(s) Xudong Qian, Sze Dai Pang, Ghim Ping Raymond Ong, Kok-Kwang Phoon

ASYMMETRIC DEPENDENCIES AMONG SOIL PARAMETERS - THE COPULA APPROACH

YI ZHANG[1], ANTÓNIO TOPA GOMES[2], MICHAEL BEER[3], INGO NEUMANN[1], UDO NACKENHORST[4] and CHUL-WOO KIM[5]

[1] Geodetic Institute, Leibniz Universität Hannover, Germany.
E-mail: yi.zhang@gih.uni-hannover.de; neumann@gih.uni-hannover.de
[2] CONSTRUCT, Faculty of Engineering (FEUP), University of Porto, Portugal.
E-mail: atgomes@fe.up.pt
[3] Institute for Risk and Reliability, Leibniz Universität Hannover, Germany.
E-mail: beer@irz.uni-hannover.de
[4] Institute of Mechanics and Computational Mechanics, Leibniz Universität Hannover, Germany.
E-mail: nackenhorst@ibnm.uni-hannover.de
[5] Department of Civil and Earth Resources Engineering, Graduate School of Engineering, Kyoto University, Japan.
E-mail: kim.chulwoo.5u@kyoto-u.ac.jp

Multivariate descriptions of soil parameters are quite important for the design and risk assessment of geotechnical engineering problems. A reliable and realistic statistical multivariate model is essential to produce a representative estimate of the soil properties for evaluating the soil conditions. Therefore, an advanced modeling of soil parameters helps towards improving the geotechnical and civil engineering practices. In this paper, we introduce the concepts of asymmetric copulas for the modeling of geotechnical data. Several asymmetric copula functions, capable of modeling nonlinear asymmetric dependence structures, are tested and analyzed. To demonstrate the advantages of asymmetric copulas, the asymmetric copula concept is compared with the traditional copula approaches from the real collected site soil data. The performance of these asymmetric copulas is discussed and compared, based on data fitting and tail dependency characterizations.

Keywords: Geotechnical engineering, soil parameters, joint distribution, asymmetric copula.

1 Introduction

Geotechnical engineering problems require frequently multivariate data analysis. Deficiencies in modeling their joint relationship may largely underestimate the failure probability of geotechnical structures, hence may lead to expensive engineering loss. In real practice, the soil parameters are often observed to be dependent. However, the question is about how to define this relationship between the soil data. The definition of “dependencies” in this context can have various meanings. When addressing different dependencies for the soil parameters, the typical concept of correlation is commonly used to construct the joint distribution models. The applicability of this concept may be problematic when the dependencies are not perfectly linear. Many former works have addressed this issue (L’Heureux & Long, 2017). In contrast to the traditional joint model, the copula model has shown its advantage and attracted significant attention from many geotechnical engineering researchers (Tang et al., 2015). The key feature of a copula approach is its flexibility in modeling the dependence structure, which can be separated from the modeling of individual behavior. However, there are various types of complicated dependencies and potential biases that could affect the quality of a multivariate model. Specifically, the uncertainties related to asymmetric dependencies are one of the most influencing factors. Fortunately, asymmetric copulas, developed recently, provide a feasible solution to this problem (Kazianka & Pilz, 2010). The use of asymmetric copulas can significantly improve the functionality of traditional copula approaches in fitting the asymmetrically dependent variables. This is particularly useful for cases where physical limits force the asymmetry, which is quite common in geotechnical properties. Nevertheless, the modeling of soil data using the asymmetric copula has never been studied in detail. Therefore, this work aims to close this gap providing a real case study for demonstrating and highlighting the merits and limitations regarding the use of asymmetric copulas.

This paper is divided into five sections. A general review of the fundamental copula theory as well as the procedures of constructing asymmetric copula models is presented in Section 2. Section 3 provides the detailed information of the collected soil data. A comparative study between symmetric and asymmetric copula approaches for modeling the collected soil data is presented in Section 4. The final concluding remarks are summarized in Section 5.

2 Methodology

2.1 *Copula Theory*

Copula is a model which could connect univariate marginal distributions to a multivariate distribution (Nelsen, 2006). A copula function can be expressed as

$$C: [0,1]^n \rightarrow [0,1] \quad and \quad H(x_1, \ldots, x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big) \tag{1}$$

where $H(\cdot)$ is the cumulative joint distribution function and $F_i(\cdot)$ is the individual cumulative marginal distribution function for the ith variable. Specifically, copula C is a cumulative distribution function which connects the one-dimensional probability distributions $F_1(x_1),\ldots,$ $F_n(x_n)$ to a multivariate probability distribution $H(\cdot)$. A copula model can describe various kinds of dependencies which include association concepts such as concordance, linear correlation and the related dependence measures. However, the traditional copulas have many weaknesses (e.g. Archimedean copulas) when they are applied to model soil parameters. A key drawback is that most well established copulas can only model symmetric dependent variables whereas the soil data usually display non-symmetric dependencies.

2.2 *Asymmetric Copulas*

In order to have a more accurate modeling of asymmetrically dependent soil variables, several groups of asymmetric copulas are introduced herein.

One of the most popular ways of constructing asymmetric copulas is by means of a product of copulas (Liebscher, 2008). The general form for constructing this type of asymmetric copula is given as following

$$C_{product}(u_1, \ldots, u_n) = \prod_{i=1}^{m} C_i\big(f_{i1}(u_1), \ldots, f_{in}(u_n)\big) \tag{2}$$

where $C_1, \ldots, C_m$ are all copulas for the n-dimensional variables, $f_{ij}: [0,1] \rightarrow [0,1]$ for i=1,…,m, j=1,…,n are the individual functions for describing the individual variable's behavior which should be strictly increasing or identically equal to 1. As for the individual functions f_{ij}, many candidate functions which are suitable for the copula construction have been proposed by Liebscher (2008). This is given in Table 1.

Table 1. Examples of individual functions.

Individual function	Parameters	Value range
I. $f_{ij}(u) = u^{\theta_{ij}}$	$\sum_{i=1}^{m} \theta_{ij} = 1$	$\theta_{ij} \in [0,1]$
II. $f_{ij}(u) = u^{\theta_{ij}} e^{(u-1)\alpha_{ij}}$	$\sum_{i=1}^{m} \theta_{ij} = 1$, $\sum_{i=1}^{m} \alpha_{ij} = 0$	$\theta_{ij} \in (0,1)$, $\alpha_{ij} \in (-\infty,1)$, $\theta_{ij} + \alpha_{ij} \geq 0$
III. * $f_{1j}(u) = \exp\left(\theta_j - \sqrt{\lvert\ln u\rvert + \theta_j^2}\right)$, $f_{2j}(u) = u\exp(-\theta_j + \sqrt{\lvert\ln u\rvert + \theta_j^2})$	θ_j for $j \in \{1,\ldots,n\}$	$\theta_j \geq \frac{1}{2}$

Another way of constructing an asymmetric copula could be through the linear convex combinations of copulas. One may use the following copula to model asymmetric properties in multiple variables:

$$C_{addition}(u_1,\ldots,u_n) = \sum_{h=0}^{n} p_h \breve{C}_h(u_1,\ldots,u_n) \tag{3}$$

where p_h is a weighting factor satisfying the conditions $0 \leq p_h \leq 1$ and $\sum_{h=0}^{n} p_h = 1$, and $\breve{C}_h(u_1,\ldots,u_n)$ is the flipped copula in the h dimension. In other words, this developed copula is in fact linear convex combinations of copulas. With such combination, the copula can model different individual tail dependencies based on the same base copula.

Despite the algebraic construction methods, another convenient way of constructing asymmetric copulas is by means of the skewed copula. The idea of this approach is from the skewed multivariate Guassian distribution which allows non-zero skewness. A general n-dimensional skew Gaussian copula can be written as:

$$C_{skew-Gaussian}(u_1,\ldots,u_n;\mu,\Sigma,\beta) = F_{n,skew}\left(F_{1,skew}^{-1}(u_1;\mu_1,1,\beta_1),\ldots,F_{1,skew}^{-1}(u_n;\mu_n,1,\beta_n);\mu,\Sigma,\boldsymbol{\beta}\right) \tag{4}$$

where $F_{n,skew}(.)$ is the n-dimensional skew normal distribution with mean parameter μ, $F_{1,skew}^{-1}(.)$ is the inverse of the univariate skew normal distribution $SN(\mu_i, 1, \beta_i)$, $\boldsymbol{\beta}$ are the shape parameters and $\sum$ is the covariance matrix. There are no previous works done on its application for modeling real collected soil data. The following will provide a case study to demonstrate the key advantages of using the asymmetric copulas in modeling soil data.

3 Case Study – Site Soil Data

The soil data used in this paper results from tests performed in a residual soil from Porto granite. This material is defined as a sandy silt with clay and gravel, presenting also a certain degree of cementation. Pinheiro Branco et al. (2014) completed an extensive characterization of a residual soil, deriving index properties, stiffness and strength parameters, nut also focusing attention on the variability of the parameters. The authors collected 40 samples, all carefully collected in situ by cutting the residual soil around the sampler (0.1×0.1×0.03 m^3), isolated and transported them to the geotechnical laboratory. The samples were subjected to direct shear tests and it was determined its unit weight, γ, the saturated unit weight, γ , the dry unit weight, γ_d, the void ratio, e, the peak shear strength, ϕ'_s, and the constant volume friction angle, ϕ'_{cv}.

4 Analysis and Discussion

A total sample size of 10000 soil data is selected for the analysis in this study. All of these data are obtained from the same site and therefore are believed to have the same statistical characteristics. In this study, we are focusing on the modeling of ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$), corresponding ϕ'_s to the secant friction angle and ϕ'_{cv} to the constant volume friction angle. A general scatter plot of the data set ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$) is shown in Fig. 1 and Fig. 2.

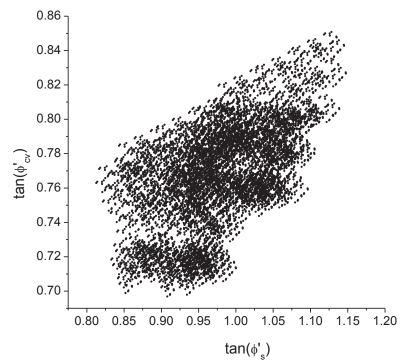

Figure 1. Scatter plot of ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$).

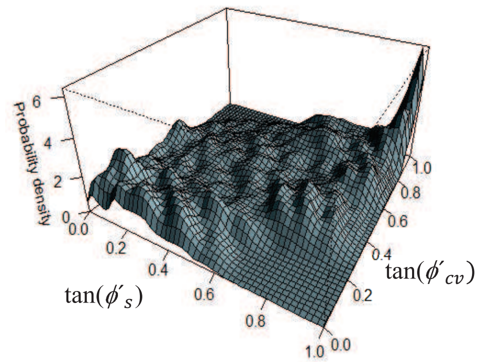

Figure 2. Empirical probability density of ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$) in the copula domain.

Several asymmetric copulas, as introduced in Section 2, are utilized here to model the soil data. These include:

1. *Symmetric copulas*: The original symmetric Archimedean copulas are considered herein. They are parametric copulas, Gumbel, Clayton and Frank copulas.
2. *Asymmetric copulas constructed by products*: We adopt the Khoudraji's device for the construction of the asymmetric copulas. Three combinations Gumbel-Clayton, Gumbel-Frank and Clayton-Frank with type I individual functions are selected.
3. *Asymmetric copulas constructed by linear convex combinations*: This group of asymmetric copulas is constructed by the rules introduced in Section 2. The selected base copulas for constructing this asymmetric copula are Gumbel, Clayton and Frank copulas.
4. *Skew Gaussian copula*: The last asymmetric copula has its exact formulation as given in Section 2.

The results for the parameter estimates, log-likelihood and AIC statistics for all the considered models fitting to ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$) are reported in Table 2. The model parameters are estimated by the method of minimization of Cramer-von Mises statistic. The best models among all the candidate models are marked in the table. The results show that the best copula models for ($\tan(\phi'_s)$, $\tan(\phi'_{cv})$) is Gumbel-Frank Type I asymmetric copula. Generally, all the asymmetric copulas show an AIC value lower than the symmetric copula. However, it does not show a better performance when the data exhibit symmetric dependences. The quality of asymmetric copulas highly relies on the utilized base copulas. On the other hand, Gumbel and Frank copulas give better representations in the data dependences. Therefore, when combining the Gumbel and Frank copula in the asymmetric copula, it shows much better features than the remaining combinations. The asymmetric copula could utilize the features from both copulas. Despite the selection of base copulas, the construction rules are also a dominant factor for the quality of asymmetric copulas. It is observed the asymmetric copulas constructed by products can generally fit the data very well in all the cases. The AIC values show that the overall performance of asymmetric copulas constructed by Khoudraji's device is quite prominent. However, the asymmetric copulas constructed by linear convex combinations present poorer results. In most of the cases, the AIC values for this type of asymmetric copulas are even higher than the symmetric Archimedean copulas. This indicates the way of constructing the asymmetric copulas by addition is not adequate for modeling the soil data dependences in this case. Compared to these combined asymmetric copulas, skewed Gaussian copula gives moderate performance. However, the key feature of using skewed Gaussian copula is that no base copulas are needed. It does not need to consider the selections of base copulas which might not be appropriate for the data.

Table 2. Parameter estimates and likelihood values for fitted soil data.

Copula type		Parameter estimate	Total log-likelihood	No. of parameters	AIC
1. One parameter copula	Gumbel	θ=1.555	35978	5	-71946
	Clayton	θ=0.6239	34957	5	-69904
	Frank	θ=3.769	35752	5	-71494
2. Asymmetric copulas constructed by products	Gumbel-Clayton Type I	γ_1=2.346, γ_2=9.296, θ_{11}=0.237, θ_{12}=0.758, θ_{21}=0.763, θ_{22}=0.242	36070	8	-72124
	Gumbel-Frank Type I	γ_1=2.566, γ_2=11.718, θ_{11}=0.206, θ_{12}=0.734, θ_{21}=0.794, θ_{22}=0.266	36165	8	-72314*
	Frank-Clayton Type I	γ_1=4.046, γ_2=5.491, θ_{11}=0.999, θ_{12}=0.924, θ_{21}=0.001, θ_{22}=0.076	35745	8	-71474
3. Asymmetric copulas constructed by linear convex combinations	Gumbel-LCC	γ=1.619, p_0=0.998, p_1=0.001, p_2=0.001	35929	7	-71844
	Clayton-LCC	γ=1.170, p_0=0.998, p_1=0.001, p_2=0.001	34465	7	-68916
	Frank-LCC	γ=3.761, p_0=0.998, p_1=0.001, p_2=0.001	35715	7	-71416
4. Skewed copula	Skew-Gaussian	β_1=0.333, β_2=-0.517, $\boldsymbol{\beta}$=[0.755, -0.980]	35867	8	-71718

*Minimum AIC value indicates the best model in each type.

To further check the quality of fitted asymmetric copulas, a comparison is made between the empirical data and the simulated data from the established models. Based on the best copula model in Table 2, the simulated data for (tan(ϕ'_s), tan(ϕ'_{cv})) are plotted in Fig. 3. It can be seen the simulated data and the original data can fit each other very well in the scatter plot. The concentrations of the simulated data generally overlap the concentrations of original data in all the plots. Even the nonlinear dependences between the variables are also well handled by the asymmetric copula. A clearer view of the fitting quality can be seen from the contour plots of the probability densities of the empirical data and the simulated data. Figure 4 shows four levels of the probability density function values for both the original data and the simulated data in all the bivariate soil data. As expected, the quality of the model fitting to the empirical data is decreasing with the drop of contour level values. Nevertheless, the similarities of the contour lines are still quite high in all the cases.

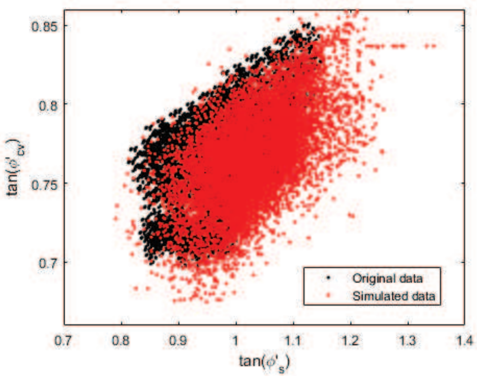

Figure 3. Scatterplot of simulated $(\tan(\phi'_s), \tan(\phi'_{cv})$

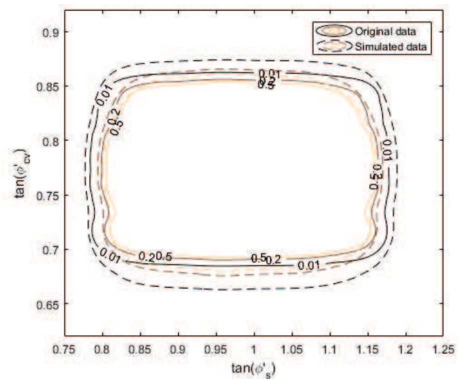

Figure 4. Contour plot of $(\tan(\phi'_s), \tan(\phi'_{cv}))$

5 Conclusion

In this paper, soil data had been analyzed by means of the asymmetric copulas in a multivariate setting. The fundamental formulation and theoretical basics of asymmetric copulas have been reviewed in detail regarding the modeling of soil parameters. The asymmetric copulas were then compared with traditional symmetric copulas on the modeling of soil parameters collected from a site located in Portugal. The copula models were constructed for the soil data group and compared based on the goodness-of-fit statistics. The results showed that the asymmetric copula can provide more appropriate characterizations of the asymmetric dependences and tail dependences in the soil data. It was found the asymmetric copula can also provide accurate predictions of extreme values from the empirical data. Future works seems necessary to investigate the ways of selecting base copulas and individual functions in the construction of asymmetric copulas.

Acknowledgments

This study is supported by grants from the Alexander von Humboldt Foundation. The first author, Yi Zhang, is sponsored by "Humboldt Research Fellowship for Postdoctoral Researchers" Program. Such financial aids are gratefully acknowledged.

References

Kazianka, H., and Pilz, J., Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment*, 24(5), 661-673, 2010.

Liebscher, E., Construction of asymmetric multivariate copulas. *Journal of Multivariate analysis*, 99(10), 2234-2250, 2008.

L'Heureux, J. S., and Long, M., Relationship between Shear-Wave Velocity and Geotechnical Parameters for Norwegian Clays, *Journal of Geotechnical and Geoenvironmental Engineering*, 143(6), 2017.

Nelsen, R. B., *An Introduction to Copulas*. Springer New York, 2006.

Pinheiro Branco, L.; Topa Gomes, A; Silva Cardoso, A; Pereira, C. S. Natural Variability of Shear Strength in a Granite Residual Soil from Porto. *Geotechnical and Geological Engineering,* 32 (4), 911-922, 2014.

Tang, X. S., Li, D. Q., Zhou, C. B., and Phoon, K. K., Copula-based approaches for evaluating slope reliability under incomplete probability information, *Structural Safety*, 52, 90-99, 2015.