Contents lists available at ScienceDirect

# Structural Safety

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

# Copula-based approaches for evaluating slope reliability under incomplete probability information

Xiao-Song Tang [a,b], Dian-Qing Li [a,b,*], Chuang-Bing Zhou [a,b,c], Kok-Kwang Phoon [d]

[a] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China
[b] Key Laboratory of Rock Mechanics in Hydraulic Structural Engineering (Ministry of Education), Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China
[c] School of Civil Engineering and Architecture, Nanchang University, Nanchang 330031, PR China
[d] Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07-03, 1 Engineering Drive 2, Singapore 117576, Singapore

## ABSTRACT

Slope reliability under incomplete probability information is a challenging problem. In this study, three copula-based approaches are proposed to evaluate slope reliability under incomplete probability information. The Nataf distribution and copula models for characterizing the bivariate distribution of shear strength parameters are briefly introduced. Then, both global and local dispersion factors are defined to characterize the dispersion in probability of slope failure. Two illustrative examples are presented to demonstrate the validity of the proposed approaches. The results indicate that the probabilities of slope failure associated with different copulas differ considerably. The commonly used Nataf distribution or Gaussian copula produces only one of the various possible solutions of probability of slope failure. The probability of slope failure under incomplete probability information exhibits large dispersion. Both global and local dispersion factors increase with decreasing probability of slope failure, especially for small coefficients of variation and strongly negative correlations underlying shear strength parameters. The proposed three copula-based approaches can effectively reduce the dispersion in probability of slope failure and significantly improve the estimate of probability of slope failure. In comparison with the Nataf distribution, the copula-based approaches result in a more reasonable estimate of slope reliability.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

It is well known that the shear strength parameters [cohesion ($c$) and tangent of friction angle ($\tan\phi$)] are important parameters for slope reliability analysis [6,12,5,20,25]. Furthermore, it is widely accepted that $c$ and $\tan\phi$ are negatively correlated (e.g., [26,31,15]). To achieve a realistic evaluation of slope reliability, the joint cumulative distribution function (CDF) or probability density function (PDF) of the shear strength parameters should be known. In geotechnical engineering practice, however, the joint CDF or PDF is often unknown due to limited data from field test or laboratory test. Based on these limited data, only the marginal distributions and covariance underlying the shear strength parameters can be determined. It is concluded that the joint probability distribution of the shear strength parameters based on these limited data cannot be determined uniquely [7,3,32].

Traditionally, the Nataf distribution is employed to construct the joint probability distribution of correlated non-normal variables based on incomplete probability information that refers to the case where only marginal distributions and covariance are available (e.g., [8,24,22,16]). For instance, Li et al. [16] investigated the rock slope reliability involving correlated non-normal variables using Nataf distribution. Although the Nataf distribution provides a convenient way for dealing with the correlated non-normal variables, it essentially adopts a Gaussian copula for modeling the dependence structure among variables [22,23,17–19]. In other words, there is an implicit assumption that the Gaussian copula is adequate for characterizing the dependence structure. Unfortunately, this commonly used assumption is not validated in a rigorous way for most applications. Furthermore, the Nataf distribution produces only one of the various possible solutions of probability of slope failure and such a probability may be biased towards the unconservative side [32]. Hence, it is of practical interest to question if there are any other models that can be used to characterize the dependence structure between the two shear strength parameters and provide a relatively reasonable estimate of probability of slope failure.

* Corresponding author at: State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, PR China. Tel.: +86 27 6877 2496; fax: +86 27 6877 4295.
*E-mail address:* dianqing@whu.edu.cn (D.-Q. Li).

Recently, the copula theory (e.g., [29]) has found wide applications in constructing the joint probability distribution of multivariate data. The copula theory provides a general and flexible way for modeling nonlinear dependence among multivariate data in isolation from their marginal probability distributions [11,33,34,14,39]. In recent years, the copula theory has been applied to geotechnical engineering. For example, Uzielli and Mayne [35] investigated the dependence among load-displacement model parameters underlying vertically loaded shallow footings on sands using copula. Tang et al. [32] investigated the impact of copula selection on slope and retaining wall reliability. Wu [36] employed the Gaussian and Frank copulas to model the trivariate distribution among cohesion, friction angle and unit weight of soils. Wu [37] investigated the series system reliability of a retaining wall using a copula-based approach. Tang et al. [32] concluded that the probabilities of slope failure associated with different copulas differ considerably. Therefore, a robust evaluation of slope reliability under incomplete probability information needs to be further studied. However, this problem is difficult due to the following reasons. First, since the slope reliability under incomplete probability information cannot be determined uniquely, it is hard to make a quantitative estimate of the slope reliability in this situation. Second, to achieve a robust estimate of slope reliability, the dispersion in probability of slope failure should be reduced as low as possible. Unfortunately, it is still a challenging problem because the probability of slope failure under incomplete information varies over a wide range.

This paper aims to propose three copula-based approaches to evaluate slope reliability in the presence of incomplete probability information. To achieve this goal, this article is organized as follows. In Section 2, the Nataf and copula models for constructing the bivariate distribution of the shear strength parameters are first introduced. Then, a global and a local dispersion factors to represent the dispersion in probability of slope failure are defined in Section 3. In Section 4, three copula-based approaches are developed to provide a robust estimate of probability of slope failure under incomplete probability information. Two illustrative examples, namely an infinite slope and the Jinping slope in China are presented in Section 5 to demonstrate the validity of the proposed approaches.

## 2. Bivariate distribution of shear strength parameters

### 2.1. The Nataf distribution for modeling bivariate distribution of shear strength parameters

As stated in the introduction, when the information on shear strength parameters is available only in terms of marginal distributions and covariance, the Nataf model is usually employed to construct the joint probability distribution of shear strength parameters for slope reliability analysis (e.g., [16,31]). To facilitate the understanding of the proposed approaches in the subsequent sections, the Nataf model is first introduced as below.

Let the random vector $\mathbf{X} = (X_1, X_2)$ denote the shear strength parameters $(c, \tan\phi)$. Assume that the marginal CDFs of $X_1$ and $X_2$, and the correlation coefficient, $\rho$, between $X_1$ and $X_2$ are known. Then, the standard normal random vector $\mathbf{Z} = (Z_1, Z_2)$ can be obtained using the following transformations:

$$Z_i = \Phi^{-1}[F_i(X_i)], \quad i = 1, 2 \tag{1}$$

where $\Phi^{-1}(.)$ is the inverse standard normal CDF. $F_i(X_i)$ is the marginal CDF of $X_i$. Following Liu and Der Kiureghian [24], a joint probability distribution is assigned to $\mathbf{X} = (X_1, X_2)$ such that $\mathbf{Z} = (Z_1, Z_2)$ are jointly normal. Using the rules of probability transformation, the joint PDF of $X_1$ and $X_2$, $f(x_1, x_2)$, is derived as

$$f(x_1, x_2) = \frac{f_1(x_1)f_2(x_2)}{\varphi(z_1)\varphi(z_2)}\varphi_2(z_1, z_2, \rho_0) \tag{2}$$

where $f_1(x_1)$ and $f_2(x_2)$ are the marginal PDFs of $X_1$ and $X_2$, respectively; $\varphi(z_1)$ and $\varphi(z_2)$ are the standard normal PDFs of $Z_1$ and $Z_2$, respectively; $\varphi_2(z_1, z_2, \rho_0)$ is the bivariate normal PDF with zero means, unit standard deviations and correlation coefficient $\rho_0$. Generally, this distribution model is referred to as the Nataf distribution. The Pearson correlation coefficient $\rho_0$ is expressed in terms of $\rho$ through the following integral relation:

$$\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)\frac{f_1(x_1)f_2(x_2)}{\varphi(z_1)\varphi(z_2)}\varphi_2(z_1, z_2, \rho_0)dx_1dx_2 \tag{3}$$

where $\mu_1$ and $\mu_2$ are the means of $X_1$ and $X_2$, respectively; $\sigma_1$ and $\sigma_2$ are the standard deviations of $X_1$ and $X_2$, respectively. For the given marginal distributions and correlation coefficient $\rho$ of $X_1$ and $X_2$, the above equation can be solved iteratively to find $\rho_0$. The Nataf distribution can be easily generalized to $N$-dimensions. This is one reason that the Nataf distribution is widely used in structural reliability analysis [8,24]. The Nataf distribution has been the standard for more than 20 years because it is not always possible to find a joint PDF with prescribed marginal distributions that is consistent with given linear correlations. Engineers and researchers have used the Nataf distribution though since nothing else was available. Recently, the copula based approach provides a new insight into the joint distribution with prescribed marginal distributions and correlation coefficient (e.g., [29,19,33,34]), which will be presented in the following.

### 2.2. Copula based approach for modeling bivariate distribution of shear strength parameters

Copulas are functions that couple a multivariate distribution to its one-dimensional marginal distributions. Alternatively, copulas are multivariate distribution functions whose one-dimensional marginal distributions are uniform on the interval of [0, 1] (e.g., [29]). There are many copulas in the literature such as Gaussian, t, Plackett, Frank, Gumbel and Clayton copulas. Each copula is characterized by its own dependence structure. According to Sklar's theorem (e.g., [29]), a bivariate distribution, $F(x_1, x_2)$, of $X_1$ and $X_2$ can be expressed in terms of a copula function $C(u_1, u_2; \theta)$ and the marginal distributions $u_1 = F_1(x_1)$ and $u_2 = F_2(x_2)$:

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2); \theta) = C(u_1, u_2; \theta) \tag{4}$$

where $\theta$ is a copula parameter describing the dependency between $X_1$ and $X_2$. From Eq. (4), the bivariate PDF, $f(x_1, x_2)$, of $X_1$ and $X_2$ can be obtained as (e.g., [29])

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)c(F_1(x_1), F_2(x_2); \theta) \tag{5}$$

where $c(F_1(x_1), F_2(x_2); \theta)$ is a copula density function, which is given by

$$c(F_1(x_1), F_2(x_2); \theta) = c(u_1, u_2; \theta) = \partial^2 C(u_1, u_2; \theta)/\partial u_1 \partial u_2 \tag{6}$$

It is evident that both the copula function $C(u_1, u_2; \theta)$ and the copula density function $c(u_1, u_2; \theta)$ are related to the copula parameter $\theta$. Like $\rho_0$ in the Nataf distribution, the copula parameter $\theta$ can be determined through the correlation coefficient $\rho$ between $X_1$ and $X_2$. According to the definition of Pearson correlation coefficient (e.g., [1]), the following integral relationship between $\rho$ and $\theta$ can be obtained:

$$\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\left(\frac{x_2 - \mu_2}{\sigma_2}\right)f_1(x_1)f_2(x_2)c(F_1(x_1), F_2(x_2); \theta)dx_1dx_2 \tag{7}$$

For prescribed marginal distributions and correlation coefficient $\rho$ of $X_1$ and $X_2$, the preceding integral equation can be solved iteratively. For example, Li et al. [21] developed a two-dimensional

**Table 1**
Summary of the adopted bivariate copula functions and their parameter domains.

| Copula | Copula function, $C(u_1, u_2; \theta)$ | Copula density function, $c(u_1, u_2; \theta)$ | Range of $\theta$ |
|---|---|---|---|
| Gaussian | $\Phi_\theta\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$ | $\frac{1}{\sqrt{1-\theta^2}}\exp\left[-\frac{\varsigma_1^2\theta^2-2\theta\varsigma_1\varsigma_2+\varsigma_2^2\theta^2}{2(1-\theta^2)}\right], \varsigma_1 = \Phi^{-1}(u_1), \varsigma_2 = \Phi^{-1}(u_2)$ | $[-1, 1]$ |
| Plackett | $\frac{S-\sqrt{S^2-4u_1u_2\theta(\theta-1)}}{2(\theta-1)}, S = 1 + (\theta-1)(u_1+u_2)$ | $\frac{\theta[1+(\theta-1)(u_1+u_2-2u_1u_2)]}{\left\{[1+(\theta-1)(u_1+u_2)]^2-4u_1u_2\theta(\theta-1)\right\}^{\frac{3}{2}}}$ | $(0, \infty)\backslash\{1\}$ |
| Frank | $-\frac{1}{\theta}\ln\left[1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$ | $\frac{-\theta(e^{-\theta}-1)e^{-\theta(u_1+u_2)}}{[(e^{-\theta}-1)+(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)]^2}$ | $(-\infty, \infty)\backslash\{0\}$ |
| No.16 | $\frac{1}{2}\left(S+\sqrt{S^2+4\theta}\right),$ $S = u_1 + u_2 - 1 - \theta\left(\frac{1}{u_1}+\frac{1}{u_2}-1\right)$ | $\frac{1}{2}\left(1+\frac{\theta}{u_1^2}\right)\left(1+\frac{\theta}{u_2^2}\right)S^{-\frac{1}{2}}\left\{-S^{-1}\left[u_1+u_2-1-\theta\left(\frac{1}{u_1}+\frac{1}{u_2}-1\right)\right]^2+1\right\},$ $S = \left[u_1+u_2-1-\theta\left(\frac{1}{u_1}+\frac{1}{u_2}-1\right)\right]^2+4\theta$ | $[0, \infty)$ |

*Note:* $\Phi^{-1}$ denotes the inverse of standard normal distribution function.

Gaussian-Hermite integral technique to solve the above integral equation. This general technique is also adopted in this study. The joint CDF and PDF of $X_1$ and $X_2$ can be directly determined using Eqs. (4) and (5) with a selected copula and the known marginal distributions of $X_1$ and $X_2$.

As mentioned previously, when the probability information on shear strength parameters is only limited to marginal distributions and covariance, a large number of copulas that are consistent with such information can be used to characterize the dependence structure. Since there exists a negative correlation between $c$ and $\tan\phi$, thus, the copulas that allow a wide range of negative correlation coefficients are selected to characterize the dependence between $c$ and $\tan\phi$. A review of the literature reveals that the Gaussian copula, Plackett copula, Frank copula and No.16 copula

(e.g., [29]) are appropriate for describing the dependence structure between $c$ and $\tan\phi$. The aforementioned four copulas, along with the domains of the $\theta$ parameter are summarized in Table 1. Among the four copulas, the Gaussian copula is an elliptical copula. The Plackett copula is a member of the Plackett copula family. The Frank and No.16 copulas are commonly used Archimedean copulas. All the four copulas can describe negative dependences, and the values of the correlation coefficients between $c$ and $\tan\phi$ can approach $-1$.

It is evident that substituting the copula density function of the Gaussian copula shown in Table 1 into Eq. (5) yields Eq. (2). Thus, the bivariate distribution using the Nataf distribution is the same as that using the Gaussian copula. In other words, the well-known Nataf distribution is nothing but a joint PDF with the Gaussian
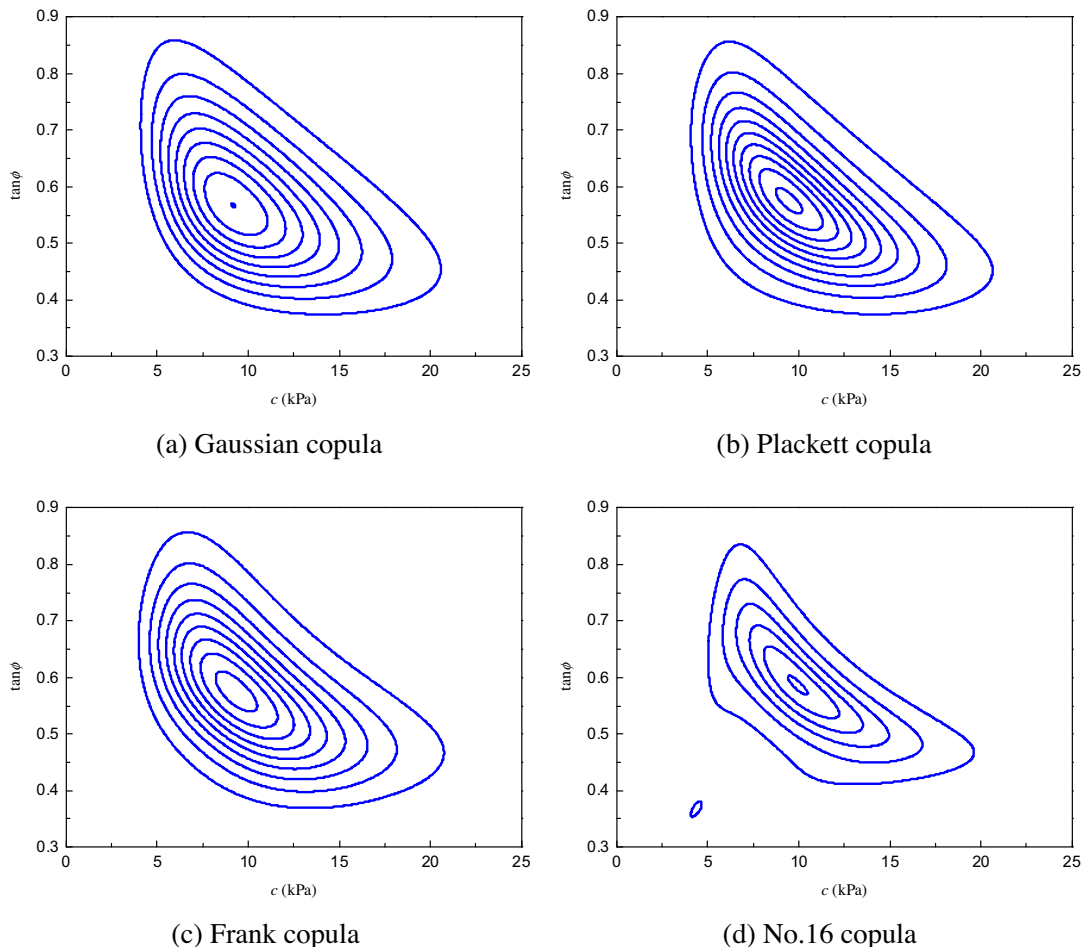


(a) Gaussian copula

(b) Plackett copula

(c) Frank copula

(d) No.16 copula

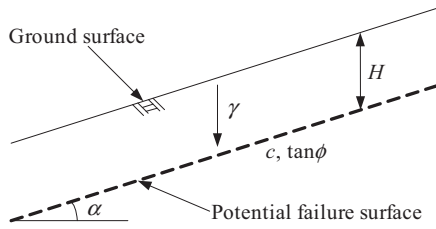**Fig. 1.** Contour plots of the joint PDFs of shear strength parameters for the four copulas.

**Fig. 2.** Infinite slope example.

copula. The Gaussian copula can be easily generalized to N-dimensions. The Archimedean copulas such as the Frank and No.16 copulas also have a multivariate PDF, but it is hard to relate the correlation coefficients between variables to their copula parameters one by one. This is because the Archimedean copulas have only up to N−1 different generator functions and thus only up to N−1 copula parameters (e.g., [29]). These parameters are commonly determined from the measured multivariate data using Maximum Likelihood Estimation (MLE) (e.g., [29]) rather than relating the correlation coefficients between variables. Notwithstanding this, as noted by Dutfoy and Lebrun [9], multivariate data are usually independent by blocks in real-life applications, each block involving only a small number of correlated variables such as two variables [13,30,38,32–34,18,19]. In this case, the multivariate data can be analyzed pair by pair using multiple bivariate copulas [36].

To visualize the dependence structures underlying different copulas, the contour plots of the bivariate PDFs of shear strength

parameters associated with the four copulas are presented in Fig. 1. In this figure, a lognormal distribution with a mean of 11 kPa and a coefficient of variation (COV) of 0.4 for cohesion $c$ and a lognormal distribution with a mean of 0.5774 and a COV of 0.2 for $\tan\phi$ are used to compare the four copulas. In addition, a correlation coefficient $\rho = -0.5$ between $c$ and $\tan\phi$ is assumed to determine the copula parameters $\theta$ underlying the four copulas. It can be seen that there is a significant difference in dependence structures associated with the four copulas even though the same marginal distributions and correlation coefficient are used. The joint PDF of shear strength parameters produced by the No.16 copula differs significantly from those produced by the other three copulas. Such a difference can lead to significant difference in probabilities of slope failure, as illustrated by Tang et al. [32].

## 3. Dispersion factor of probability of slope failure

As discussed in the previous sections, the probabilities of slope failure produced by different copulas may differ greatly [32]. To quantify the maximum possible dispersion in probability of slope failure when the dependence structure between shear strength parameters varies within the set of copulas $\varepsilon$ = {Gaussian, Plackett, Frank and No.16 copulas}, a global dispersion factor associated with probability of slope failure is introduced. Let $p_{f\min}$ = min{$p_f(C)$, $C \in \varepsilon$} and $p_{f\max}$ = max{$p_f(C)$, $C \in \varepsilon$} in which $p_f(C)$ is the probability of slope failure associated with a specific copula $C$. Following Dutfoy and Lebrun [9], the global dispersion factor of probability of slope failure, $r$, is defined as
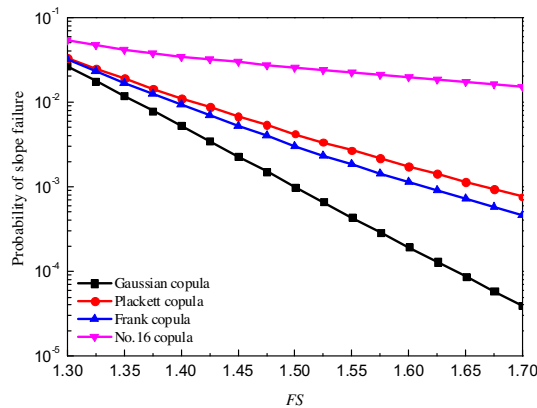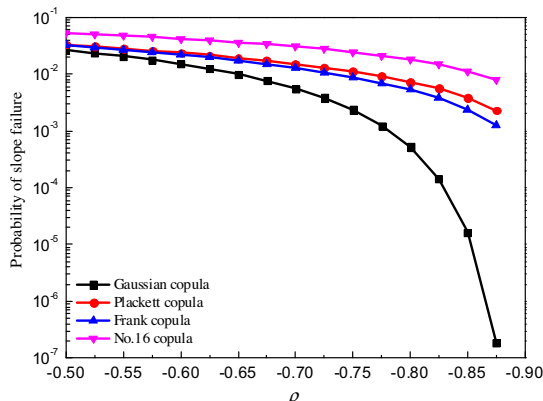


(a) $H$ varying from 5 m to 2.14 m
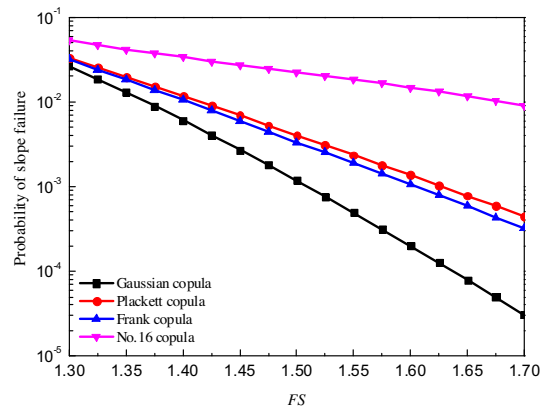


(b) $\alpha$ varying from 30° to 23.27°
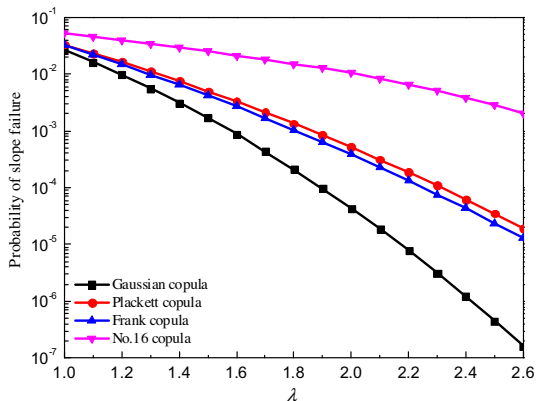


(c) $\lambda$ varying from 1 to 2.6



(d) $\rho$ varying from -0.5 to -0.88

**Fig. 3.** Probabilities of slope failure produced by different copulas for the infinite slope.

$$r = \frac{p_{f\,\max}}{p_{f\,\min}} \tag{8}$$

A large $r$ denotes the calculated probabilities of slope failure varying over a wide range, while a small $r$ denotes the calculated probabilities of slope failure varying over a narrow range. It is stated that the probability of slope failure under incomplete probability information can be evaluated quantitatively if $1 \leqslant r \leqslant 1.5$, and it can be evaluated qualitatively if $1.5 \leqslant r \leqslant 10$. If $r > 10$, the estimated probability of slope failure may exceed the actual probability of slope failure by at least one order of magnitude, which is possibly unacceptable for practical slope reliability analysis [9].

Besides the global dispersion factor of probability of slope failure, a local dispersion factor of probability of slope failure with respect to a reference copula $C$ can also be defined. The local dispersion factor of probability of slope failure with respect to a reference copula $C$, $r_0$, is defined as

$$r_0 = \max \left\{ \frac{p_{f\,\max}}{p_f(C)}, \frac{p_f(C)}{p_{f\,\min}} \right\} \tag{9}$$

where $p_f(C)$, $p_{f\max}$ and $p_{f\min}$ are the same as those defined previously.

The global dispersion factor $r$ defined in Eq. (8) can only characterize the maximum possible dispersion in probability of slope failure for the whole set of copulas $\varepsilon$, whereas the local dispersion factor $r_0$ can effectively quantify the maximum possible dispersion in probability of slope failure for a reference copula $C$ in $\varepsilon$. Hence, it is possible to obtain the maximum difference between $p_f(C)$ and $p_{f\max}$ or $p_{f\min}$ when a specific copula $C$ is used to model the dependence structure between $c$ and $\tan\phi$ for slope reliability analysis. This local dispersion factor is adopted to examine the performance of the proposed reliability approaches to evaluate the slope reliability under incomplete probability information.

## 4. Three copula-based approaches for evaluating probability of slope failure

From the copula viewpoint, the Nataf distribution adopts the Gaussian copula for modeling the dependence structure between the shear strength parameters. Consequently, the probability of slope failure associated with the Nataf distribution is only one of the various possible solutions of the slope reliability and may be biased towards the unconservative side [32]. In this section, three copula-based approaches are proposed to evaluate the slope reliability under incomplete probability information. One objective of these approaches is to provide a more reasonable estimate of probability of slope failure. With these approaches, reliability analysis of slopes is simplified to the selection of an appropriate copula for modeling the dependence structure between the shear strength parameters. These three approaches are developed in the following.

### 4.1. Copula approach 1

In the first approach, one selects a copula from the set of candidate copulas $\varepsilon$ = {Gaussian, Plackett, Frank and No.16 copulas} that results in the highest estimate of the probability of slope failure. The rationale behind this approach is that a conservative estimate of the slope reliability is generally accepted by engineers when limited information on shear strength parameters is available (e.g., [1,8,15]). When the quality of dependence information is improved, the set of acceptable candidate copulas is reduced and the accuracy in estimating probability of slope failure remains unchanged or increases. This approach requires engineers to collect more data to improve the dependence information available.

### 4.2. Copula approach 2

The second approach is developed based on the concept of the local dispersion factor of probability of slope failure as defined in Section 3. Since the local dispersion factor measures the maximum difference between $p_f(C)$ and $p_{f\max}$ or $p_{f\min}$, an effective approach for evaluating slope reliability is to choose a copula that results in the minimum value of the local dispersion factor among the set of candidate copulas $\varepsilon$ = {Gaussian, Plackett, Frank and No.16 copulas}. This copula is taken as the optimal copula to model the given dependence structure between the shear strength parameters. With this approach, the selected copula minimizes the local dispersion in probability of slope failure. It can provide a more reasonable estimate of probability of slope failure.

### 4.3. Copula approach 3

The third and more appealing approach is based on a Bayesian notion [4]. The copula for modeling the dependence structure between $c$ and $\tan\phi$ is assumed to be a weighted average of all candidate copulas in $\varepsilon$. For the dependence structure between $c$ and $\tan\phi$ characterized by $m$ candidate copulas $C_i$ $(u_1, u_2; \theta_i)$, $i = 1, 2, \ldots, m$, the Bayesian copula $C(u_1, u_2; \theta)$ is expressed as

$$C(u_1, u_2; \theta) = \sum_{i=1}^{m} p_i C_i(u_1, u_2; \theta_i) \tag{10}$$

where $p_i$ is the weight representing potential probability of each candidate copula being the true copula, and satisfies $\sum_{i=1}^{m} p_i = 1$. The Bayesian copula can be directly used to compute the probability of slope failure. Note that the copula parameters $\theta_i$ for all candidate copulas are determined using Eq. (7) with the same correlation coefficient $\rho$ between $c$ and $\tan\phi$. This approach provides a robust estimate of probability of slope failure because it accounts for the potential probability of each candidate copula being the true copula.

The potential probability of each candidate copula being the true copula could be determined by several methods, such as subjective judgment, engineering experience, and bootstrapping approach. In this study, the bootstrapping approach [10] is adopted for such a purpose. The bootstrap method is a nonparametric and straightforward approach to derive the sampling distributions of sample statistics. With the measured data set $\mathbf{X} = \{(c_i, \tan\phi_i), i = 1, 2, \ldots, N\}$, the Akaike Information Criterion (AIC) [2] is often used to identify the best-fit copula between $c$ and $\tan\phi$ [19, 32]. A copula corresponding to the smallest AIC value is considered to be the best-fit copula. It should be noted that, in geotechnical engineering practice, the best-fit copula is often identified from a very limited data set [32], which inevitably leads to uncertainty in the AIC values and the identification results. This uncertainty is characterized by the bootstrap approach. Following Luo et al. [28], a value of $N_s$ = 10000 bootstrap sample sets is adopted for bootstrapping. The sample size of each bootstrap sample set is equal to the sample size of the original data set, $N$. Based on the $N_s$ sets of bootstrap samples, the AIC values associated with the candidate copulas are calculated. Then, the best-fit copula can be identified from the AIC values, which results in $N_s$ best-fit copulas for $N_s$ bootstrap sample sets. The numbers of times being the best-fit copula for each candidate copula are obtained. In this way, the probability of each candidate copula being the true copula is obtained. Taking the CS-ET data set of shear strength parameters in Table 2 of Tang et al. [32] as an example, the numbers of times being the best-fit copula are 3783, 794, 3452 and 1971 for the Gaussian, Plackett, Frank and No.16 copulas, respectively. Thus, the corresponding probabilities being the true copula are 37.83%, 7.94%, 34.52% and 19.71%. For illustration, the above four probabilities are taken as
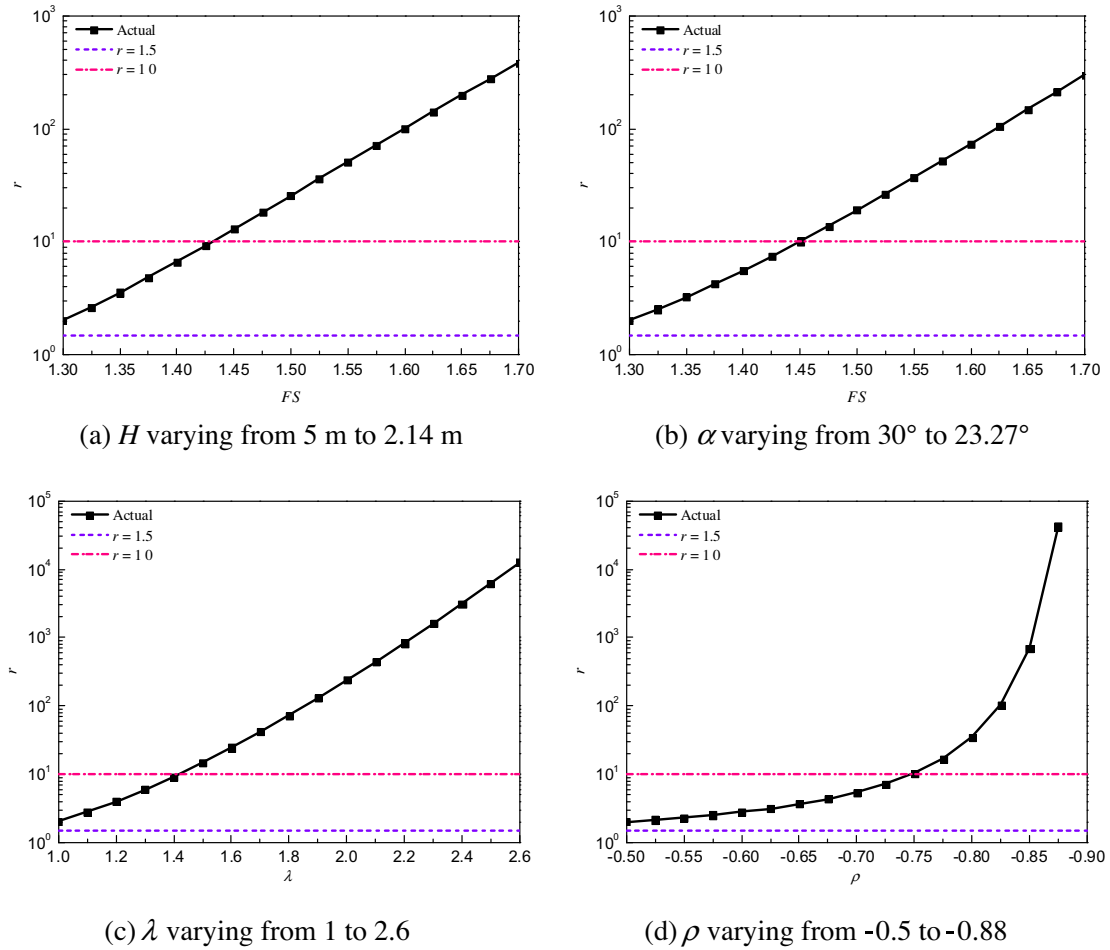
(a) $H$ varying from 5 m to 2.14 m

(b) $\alpha$ varying from 30° to 23.27°

(c) $\lambda$ varying from 1 to 2.6

(d) $\rho$ varying from -0.5 to -0.88

**Fig. 4.** Global dispersion factors of probability of slope failure for the infinite slope.

$p_i$ in Eq. (10) for the slope reliability analyses in the following section.

## 5. Illustrative examples

In this section, two slope reliability examples are studied to demonstrate the validity of the proposed copula-based approaches for evaluating the slope reliability under incomplete probability information: (1) an infinite slope example with one pair of shear strength parameters and (2) the Jinping slope example in China [31] with multiple pairs of shear strength parameters.

### 5.1. Example 1: an infinite slope with one pair of shear strength parameters

An infinite slope as shown in Fig. 2 is studied to demonstrate the validity of the proposed copula-based approaches. By assuming a deep groundwater table to the slope, the factor of safety of the infinite slope, FS, can be calculated as (e.g., [32])

$$FS = \frac{c + \gamma H \cos^2 \alpha \tan \phi}{\gamma H \sin \alpha \cos \alpha} \tag{11}$$

where $c$ and $\tan\phi$ are effective cohesion and tangent of friction angle of the soil, respectively; $H$, $\alpha$ and $\gamma$ denote the depth of the soil above bedrock, slope inclination and unit weight of the soil, respectively. In this example, $c$ and $\tan\phi$ are considered as uncertain variables. Both $c$ and $\tan\phi$ are assumed to be lognormally distributed.

The mean values of $c$ and $\tan\phi$ are 11 kPa and 0.5774, respectively. The COVs of $c$ and $\tan\phi$ are 0.4 and 0.2, respectively. Also, a correlation coefficient $\rho = -0.5$ between $c$ and $\tan\phi$ is adopted to account for the effect of correlation on slope reliability. The deterministic quantities are $\gamma = 17$ kN/m³, $H = 5$ m, and $\alpha = 30°$. These values lead to a mean factor of safety of FS = 1.30 calculated by Eq. (11).

The performance function for the infinite slope reliability problem is expressed as

$$g(c, \tan \phi) = FS(c, \tan \phi) - 1 \tag{12}$$

where $FS(c, \tan\phi)$ is evaluated by Eq. (11). Many reliability methods in the literature [27] can be used to conduct reliability analysis associated with Eq. (12). As studied by Tang et al. [32], the probability of slope failure can be computed using the direct integration method. The probability of slope failure is studied based on the following three factors: (1) geometrical parameters ($H$, $\alpha$), (2) COV scaling factor, $\lambda$ defined as COV$_c$ = 0.4/$\lambda$ and COV$_{\tan\phi}$ = 0.2/$\lambda$, and (3) correlation coefficient $\rho$. In the parametric studies as shown in Fig. 3, each factor varies over a range while the other parameters remain unchanged.

Fig. 3 compares the probabilities of slope failure on log scale produced by different copulas. To facilitate a comparison between Figs. 3(a) and (b), the changes in $H$ and $\alpha$ are transformed into the changes in FS in a uniform way. In Figs. 3(a) and (b), the FS increases from 1.30 to 1.70 when $H$ decreases from 5 to 2.14 m or $\alpha$ decreases from 30° to 23.27°. It is evident that the probabilities of slope failure produced by different copula models differ considerably. Among the four copulas, the Gaussian copula produces
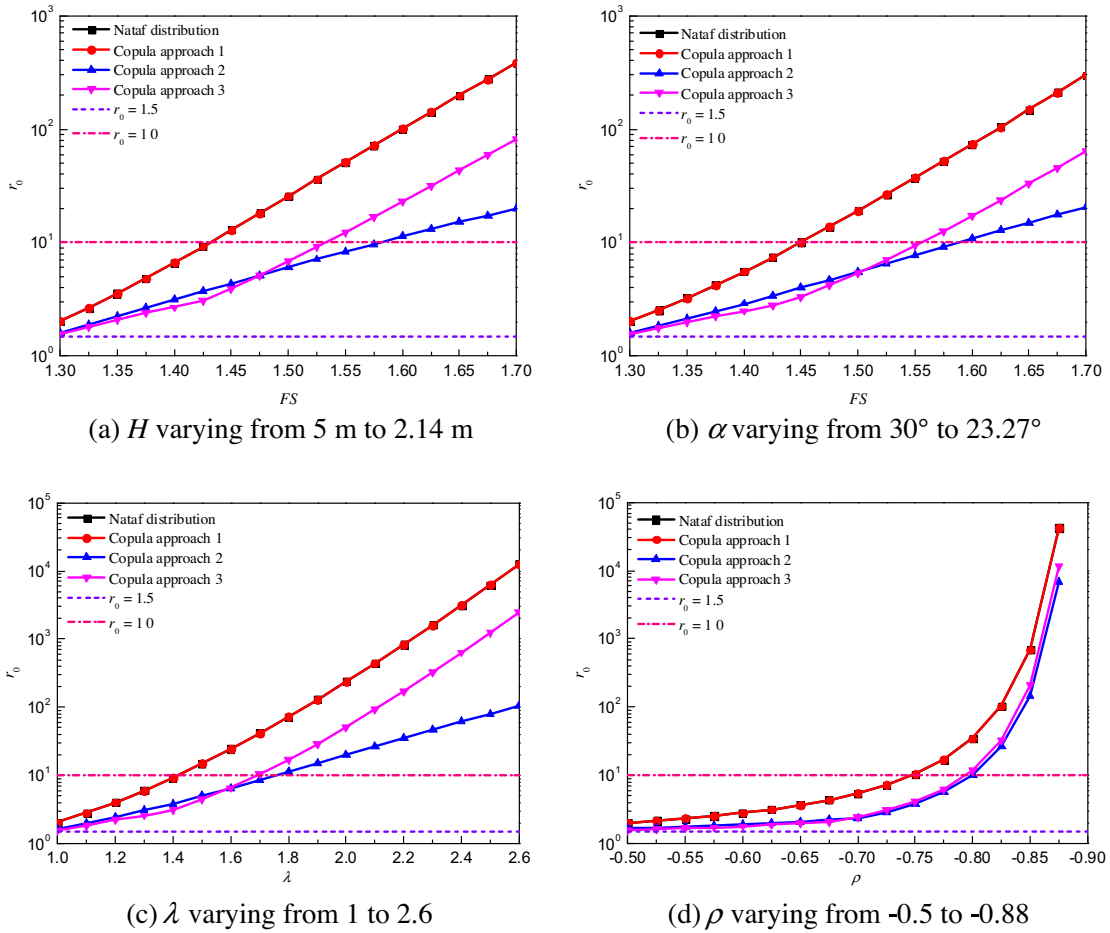
(a) $H$ varying from 5 m to 2.14 m

(b) $\alpha$ varying from 30° to 23.27°

(c) $\lambda$ varying from 1 to 2.6

(d) $\rho$ varying from -0.5 to -0.88

**Fig. 5.** Local dispersion factors of probability of slope failure for the infinite slope.

the smallest probability of slope failure, whereas the No.16 copula leads to the largest probability of slope failure. In addition, the probabilities of slope failure are more sensitive to the COVs of shear strength parameters and the negative correlation between $c$ and $\tan\phi$. These results indicate that the probability of slope failure under incomplete probability information cannot be determined uniquely. The commonly used Gaussian copula may underestimate the probability of slope failure significantly if it is inadequate to model the dependence structure between $c$ and $\tan\phi$, which is unconservative for slope safety assessment.

Based on the above results, the global dispersion factors of probability of slope failure can be obtained using Eq. (8). Here, $p_{f\text{-}max}$ denote the probabilities of slope failure produced by the No.16 copula, whereas $p_{f\min}$ are the probabilities of slope failure for the Gaussian copula. Fig. 4 shows the global dispersion factors corresponding to the four cases shown in Fig. 3. Note that the probability of slope failure exhibits large global dispersion because of the significant difference in the probabilities of slope failure produced by different copulas. The global dispersion factor increases with

decreasing probability of slope failure, which means that the error in probability of slope failure based on incomplete probability information becomes larger as the probability of slope failure decreases, especially for small COVs of the shear strength parameters or a strongly negative correlation between $c$ and $\tan\phi$. Since all the calculated values of $r$ exceed 1.5 as shown in Fig. 4, the probability of the infinite slope failure based on the marginal distributions and correlation coefficient of shear strength parameters may not be estimated quantitatively. When the probability of slope failure is larger than $1.0 \times 10^{-3}$, the calculated global dispersion factors fall within [1.5, 10], which means that a qualitative estimate of the true probability of slope failure based on incomplete probability information can be made. When the probability of slope failure is below $1.0 \times 10^{-3}$, the calculated global dispersion factors significantly exceed 10. For instance, they can be up to $1.26 \times 10^4$ in Fig. 4(c) for $\lambda = 2.6$ or $4.21 \times 10^4$ in Fig. 4(d) for $\rho = -0.88$. In this situation, the estimated probability of slope failure exceeds the true probability of slope failure by at least one order of magnitude. These results indicate that the knowledge of

**Table 2**
Comparison of local dispersion factors of probability of slope failure produced by different approaches for the infinite slope.

| Approach | $H = [5 \text{ m}, 2.14 \text{ m}]$ | | | $\alpha = [30°, 23.27°]$ | | | $\lambda = [1, 2.6]$ | | | $\rho = [-0.5, -0.88]$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $FS = 1.3$ | $FS = 1.5$ | $FS = 1.7$ | $FS = 1.3$ | $FS = 1.5$ | $FS = 1.7$ | $\lambda = 1$ | $\lambda = 1.8$ | $\lambda = 2.6$ | $\rho = -0.5$ | $\rho = -0.7$ | $\rho = -0.88$ |
| Nataf distribution | 2.02 | 25.67 | 382.45 | 2.02 | 19.08 | 299.77 | 2.02 | 72.60 | $1.26 \times 10^4$ | 2.02 | 5.42 | $4.21 \times 10^4$ |
| Copula approach 1 | 2.02 | 25.67 | 382.45 | 2.02 | 19.08 | 299.77 | 2.02 | 72.60 | $1.26 \times 10^4$ | 2.02 | 5.42 | $4.21 \times 10^4$ |
| Copula approach 2 | 1.61 | 6.06 | 19.95 | 1.61 | 5.56 | 20.39 | 1.61 | 11.32 | $1.05 \times 10^2$ | 1.67 | 2.36 | $6.72 \times 10^3$ |
| Copula approach 3 | 1.56 | 6.83 | 81.33 | 1.56 | 5.39 | 64.32 | 1.56 | 16.92 | $2.52 \times 10^3$ | 1.56 | 2.45 | $1.16 \times 10^4$ |

**Fig. 6.** Section $II_1$–$II_1$ of the Jinping slope.

**Table 3**
Statistics of the shear strength parameters in the Jinping slope example.

| Materials | Parameters | Distribution | Mean | COV |
|---|---|---|---|---|
| Lamprophyre dike $X$ | $c_1$ (kPa) | Lognormal | 20 | 0.25 |
| | $\tan\phi_1$ | Lognormal | 0.3 | 0.15 |
| Fault $f_{42\text{-}9}$ | $c_2$ (kPa) | Lognormal | 20 | 0.30 |
| | $\tan\phi_2$ | Lognormal | 0.3 | 0.20 |
| Class III2 rock mass | $c_3$ (kPa) | Lognormal | 1020 | 0.15 |
| | $\tan\phi_3$ | Lognormal | 0.9 | 0.08 |
| Class IV1 rock mass | $c_4$ (kPa) | Lognormal | 700 | 0.18 |
| | $\tan\phi_4$ | Lognormal | 0.6 | 0.10 |
| Class IV2 rock mass | $c_5$ (kPa) | Lognormal | 600 | 0.20 |
| | $\tan\phi_5$ | Lognormal | 0.4 | 0.12 |

local dispersion factors as those using the Nataf distribution for the considered infinite slope. However, in comparison with the Nataf distribution, copula approach 1 can always produce conservative reliability results. Unlike copula approach 1, both copula approaches 2 and 3 can reduce the local dispersion factors significantly, and provide a more reasonable estimate of the probability of slope failure. Table 2 summarizes the local dispersion factors produced by different approaches. Compared with copula approach 1 and the Nataf distribution, the local dispersion factors for approaches 2 and 3 are reduced substantially.

### 5.2. Example 2: the Jinping slope in China with multiple pairs of shear strength parameters

The Jinping slope studied by Tang et al. [31] is investigated again to demonstrate the validity of the proposed copula-based approaches. Fig. 6 shows a typical section, Section $II_1$–$II_1$ of the slope. In Fig. 6, faults $f_5$, $f_8$, $f_{42\text{-}9}$ and a lamprophyre dike $X$ are found. The slope is likely to slide along a potential surface that is highlighted with a thick line in Fig. 6. There are five pairs of shear strength parameters belonging to five different materials of the slope. Among them, $(c_1, \tan\phi_1)$ is the shear strength parameters for lamprophyre dike $X$; $(c_2, \tan\phi_2)$ is for fault $f_{42\text{-}9}$; $(c_3, \tan\phi_3)$ is for class III2 rock mass; $(c_4, \tan\phi_4)$ is for class IV1 rock mass; $(c_5, \tan\phi_5)$ is for class IV2 rock mass. In this example, all the five pairs of shear strength parameters are treated as random variables. The statistical properties of the shear strength parameters are summarized in Table 3. Additionally, the unit weight of the rock is treated as a deterministic quantity, $\gamma = 27$ kN/m³. For illustration, the factor of safety is calculated by the residual thrust method [31] under the natural condition. Substituting the mean values of the shear strength parameters into the slope stability model leads to a mean factor of safety of $FS = 1.18$.

the marginal distributions and covariance of the shear strength parameters is not enough to estimate the probability of slope failure accurately.

Applying copula approach 1, the No.16 copula is selected to model the dependence structure between $c$ and $\tan\phi$ because it results in the largest probability of slope failure. For copula approach 2, the Plackett copula that produces the minimum local dispersion factors among the four copulas is selected to model the dependence structure between $c$ and $\tan\phi$ in Figs. 3(a)–(c). In Fig. 3(d), the Frank copula is selected because it leads to the minimum local dispersion factors. For copula approach 3, the assumed weights 37.83%, 7.94%, 34.52% and 19.71% for Gaussian, Plackett, Frank and No.16 copulas, respectively are used to construct the Bayesian copula shown in Eq. (10).

After determining the type of copula for each copula-based approach, the local dispersion factors of probability of slope failure are obtained using Eq. (9). Fig. 5 shows the local dispersion factors $r_0$ for each copula-based approach along with the Nataf distribution. Essentially, the local dispersion factors for the Nataf distribution are the same as those for the Gaussian copula. The Nataf distribution produces the largest local dispersion factors, which implies that the probability of slope failure using the Nataf distribution may significantly deviate from the true probability of slope failure. As to be expected, copula approach 1 results in the same
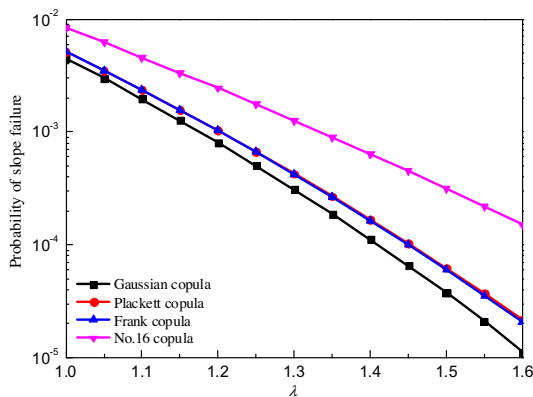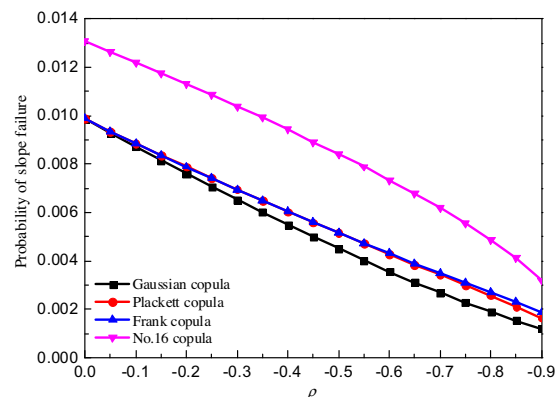


(a) $\lambda$ varying from 1 to 1.6

(b) $\rho$ varying from 0 to -0.9

**Fig. 7.** Probabilities of slope failure produced by different copulas for the Jinping slope.
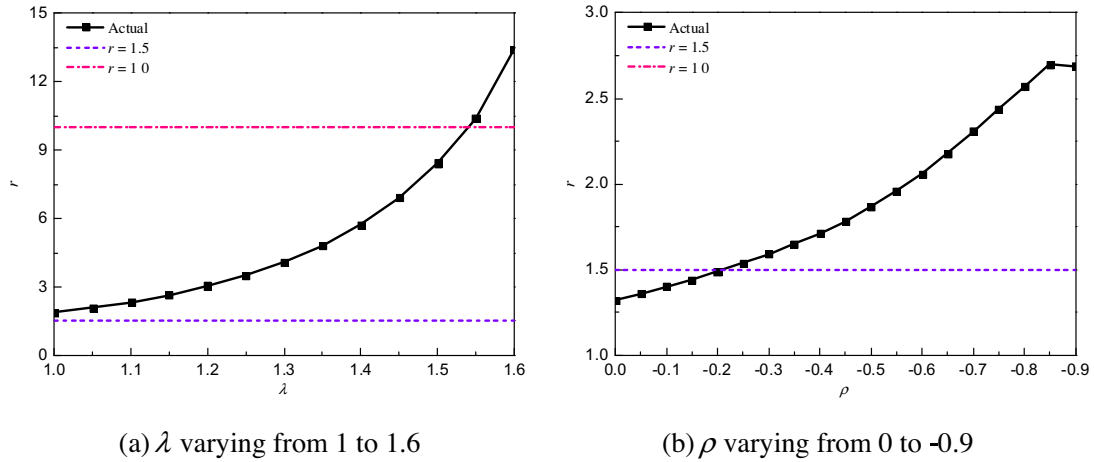
(a) $\lambda$ varying from 1 to 1.6                                    (b) $\rho$ varying from 0 to -0.9

**Fig. 8.** Global dispersion factors of probability of slope failure for the Jinping slope.



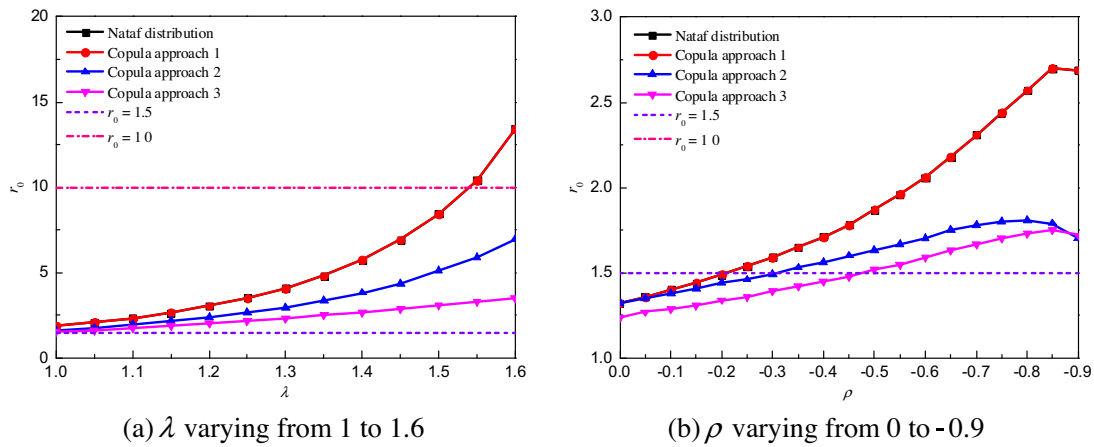(a) $\lambda$ varying from 1 to 1.6                                    (b) $\rho$ varying from 0 to - 0.9

**Fig. 9.** Local dispersion factors of probability of slope failure for the Jinping slope.

The performance function similar to Eq. (12) for the infinite slope is used again. The Monte Carlo simulation with a sample size of $10^7$ is adopted to compute the probability of slope failure. The probability of the Jinping slope failure is studied based on two factors: (1) COV scaling factor, $\lambda$ and (2) correlation coefficient $\rho$ between shear strength parameters. It is noted that there are five pairs of shear strength parameters. Hence, five bivariate copulas are employed to model the dependence structures. For simplicity, the same correlation coefficients $\rho$ are applied to all the five pairs of shear strength parameters.

Fig. 7 shows the probabilities of slope failure produced by different copula models for shear strength parameters. In Fig. 7(a), $\lambda$ varies over a range for $\rho = -0.5$. Similarly, $\rho$ varies over a range for $\lambda = 1$ in Fig. 7(b). Like the results shown in Fig. 3, the probabilities of slope failure associated with different copula models differ considerably. Again, the Gaussian copula produces the smallest probability of slope failure and the No.16 copula leads to the largest probability of slope failure. The probabilities of slope failure produced by the Gaussian copula in Fig. 7(b) are the same as those produced by the Nataf distribution in Fig. 8 of Tang et al. [31]. These results indicate that the commonly used Nataf distribution may underestimate the probability of slope failure significantly, which is unconservative for slope safety assessment.

Fig. 8 shows the global dispersion factors of probability of slope failure. The probability of slope failure exhibits large global dispersion. The global dispersion factor $r$ increases with decreasing probability of slope failure. However, when $\rho = -0.9$, the global dispersion factor becomes smaller. This is because all the selected copulas converge to the Fréchet-Hoeffding lower bound $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$ when $\rho$ approaches $-1$ (e.g., [29]). Hence, the probabilities of slope failure produced by different copulas are the same when $\rho$ approaches $-1$. In this example, most of the calculated global dispersion factors generally fall within [1.5, 10]. Thus, a qualitative estimate of the probability of slope failure under incomplete probability information can be made. These results indicate that the knowledge of the marginal distributions and covariance of the shear strength parameters is generally not enough to estimate the probability of slope failure with a sufficient accuracy.

Fig. 9 shows the local dispersion factors for the three copula-based approaches as well as the Nataf distribution. In Fig. 9, the No.16 copula is used to model the dependence structure between $c$ and $\tan\phi$ for copula approach 1. For copula approach 2, the Plackett copula and the Frank copula resulting in the minimum local dispersion factors are selected for Fig. 7(a) and (b), respectively. To construct the Bayesian copula for copula approach 3, the assumed weights 37.83%, 7.94%, 34.52% and 19.71% are used for the Gauss-

ian, Plackett, Frank and No.16 copulas, respectively. It can be observed that the Nataf distribution produces the largest local dispersion factors, and copula approach 1 results in the same values as the Nataf distribution. Compared with copula approach 1, both copula approaches 2 and 3 reduce the local dispersion factors greatly. They provide a more reasonable estimate of the probability of slope failure. The local dispersion factors become smaller when $\rho$ approaches $-1$ as discussed previously.

## 6. Summary and conclusions

This paper has proposed three copula-based approaches for evaluating slope reliability under incomplete probability information. Two illustrative examples are presented to demonstrate the validity of the proposed approaches. Several conclusions can be drawn from this study:

(1) The slope reliability under incomplete probability information cannot be determined uniquely from a theoretical viewpoint. The probabilities of slope failure produced by different copulas for modeling dependence structure between shear strength parameters differ significantly. The commonly used Nataf distribution or Gaussian copula produces only one of the various possible solutions of probability of slope failure. They may overestimate the slope reliability significantly. This finding should be noted in practical geotechnical applications.

(2) The probability of slope failure under incomplete probability information exhibits large dispersion. Both the global and the local dispersion factors increase with decreasing probability of slope failure, especially for small COVs of shear strength parameters and strongly negative correlations between $c$ and $\tan\phi$.

(3) The proposed three copula-based approaches can effectively reduce the dispersion in probability of slope failure and significantly improve the estimate of probability of slope failure. In comparison with the Nataf distribution, the proposed copula-based approaches result in a more reasonable estimate of slope reliability, which provide practical tools for evaluating the slope reliability under incomplete probability information. However, slope reliability under incomplete probability information is still a challenging problem in geotechnical engineering. More efforts on this topic should be further made.

## Acknowledgments

## References

[1] Ang AH-S, Tang WH. Probability concepts in engineering planning and design, vol. II: decision, risk, and reliability. New York: John Wiley and Sons; 1984.
[2] Akaike H. A new look at the statistical model identification. IEEE T Automat Contr 1974;19(6):716–23.
[3] Beer M, Zhang Y, Quek ST, Phoon KK. Reliability analysis with scarce information: comparing alternative approaches in a geotechnical engineering context. Struct Saf 2013;41:1–10.
[4] Cao ZJ, Wang Y. Bayesian model comparison and selection of spatial correlation functions for soil parameters. Struct Saf 2014;49:10–7.
[5] Ching JY, Phoon KK. Probability distribution for mobilised shear strengths of spatially variable soils under uniform stress states. Georisk 2013;7(3):209–24.
[6] Christian JT, Ladd CC, Baecher GB. Reliability applied to slope stability analysis. J Geotech Eng 1994;120(12):2180–207.
[7] Der Kiureghian A, Ditlevsen O. Aleatory or epistemic? Does it matter? Struct Saf 2009;31(2):105–12.
[8] Der Kiureghian A, Liu PL. Structural reliability under incomplete probability information. J Eng Mech 1986;112(1):85–104.
[9] Dutfoy A, Lebrun R. Practical approach to dependence modelling using copulas. Proc Inst Mech Eng, Part O: J Risk Reliab 2009;223(4):347–61.
[10] Efron B. Bootstrap methods: Another look at the jackknife. Ann Stat 1979;7(1):1–26.
[11] Eryilmaz S. Multivariate copula based dynamic reliability modeling with application to weighted-k-out-of-n systems of dependent components. Struct Saf 2014;51:23–8.
[12] Griffiths DV, Huang JS, Fenton GA. Probabilistic infinite slope analysis. Comput Geotech 2011;38(4):577–84.
[13] Goda K. Statistical modeling of joint probability distribution using copula: application to peak and permanent displacement seismic demands. Struct Saf 2010;32(2):112–23.
[14] Huffman JC, Stuedlein AW. Reliability-based serviceability limit state design of spread footings on aggregate pier reinforced clay. J Geotech Geoenviron Eng 2014;140(10):04014055.
[15] Jiang SH, Li DQ, Cao ZJ, Zhou CB, Phoon KK. Efficient system reliability analysis of slope stability in spatially variable soils using Monte Carlo simulation. J Geotech Geoenviron Eng 2014. http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0001227.
[16] Li DQ, Chen YF, Lu WB, Zhou CB. Stochastic response surface method for reliability analysis of rock slopes involving correlated non-normal variables. Comput Geotech 2011;38(1):58–68.
[17] Li DQ, Wu SB, Zhou CB, Phoon KK. Performance of translation approach for modeling correlated non-normal variables. Struct Saf 2012;39:52–61.
[18] Li DQ, Phoon KK, Wu SB, Chen YF, Zhou CB. Impact of translation approach for modelling correlated non-normal variables on parallel system reliability. Structure and Infrastructure Engineering 2013;9(10):969–82.
[19] Li DQ, Tang XS, Phoon KK, Chen YF, Zhou CB. Bivariate simulation using copula and its application to probabilistic pile settlement analysis. Int J Numer Anal Meth Geomech 2013;37(6):597–617.
[20] Li DQ, Qi XH, Phoon KK, Zhang LM, Zhou CB. Effect of spatially variable shear strength parameters with linearly increasing mean trend on reliability of infinite slopes. Struct Saf 2014;49:45–55.
[21] Li HS, Lv ZZ, Yuan XK. Nataf transformation based point estimate method. Chinese Sci Bull 2008;53(17):2586–92.
[22] Lebrun R, Dutfoy A. Do Rosenblatt and Nataf isoprobabilistic transformations really differ? Probab Eng Mech 2009;24(4):577–84.
[23] Lebrun R, Dutfoy A. A generalization of the Nataf transformation to distributions with elliptical copula. Probab Eng Mech 2009;24(2):172–8.
[24] Liu PL, Der Kiureghian A. Multivariate distribution models with prescribed marginals and covariances. Probab Eng Mech 1986;1(2):105–12.
[25] Lloret-Cabot M, Fenton GA, Hicks MA. On the estimation of scale of fluctuation in geostatistics. Georisk 2014;8(2):129–40.
[26] Low BK. Reliability analysis of rock slopes involving correlated non normals. Int J Rock Mech Min Sci 2007;44(6):922–35.
[27] Low BK. FORM, SORM, and spatial modeling in geotechnical engineering. Struct Saf 2014;49:56–64.
[28] Luo Z, Atamturktur S, Juang CH. Bootstrapping for characterizing the effect of uncertainty in sample statistics for braced excavations. J Geotech Geoenviron Eng 2013;139(1):13–23.
[29] Nelsen RB. An introduction to copulas. 2nd ed. New York: Springer; 2006.
[30] Phoon KK, Santoso A, Quek ST. Probabilistic analysis of soil-water characteristic curves. J Geotech Geoenviron Eng 2010;136(3):445–55.
[31] Tang XS, Li DQ, Chen YF, Zhou CB, Zhang LM. Improved knowledge-based clustered partitioning approach and its application to slope reliability analysis. Comput Geotech 2012;45:34–43.
[32] Tang XS, Li DQ, Rong G, Phoon KK, Zhou CB. Impact of copula selection on geotechnical reliability under incomplete probability information. Comput Geotech 2013;49:264–78.
[33] Tang XS, Li DQ, Zhou CB, Zhang LM. Bivariate distribution models using copulas for reliability analysis. P I Mech Eng O-J Ris 2013;227(5):499–512.
[34] Tang XS, Li DQ, Zhou CB, Phoon KK, Zhang LM. Impact of copulas for modeling bivariate distributions on system reliability. Struct Saf 2013;44:80–90.
[35] Uzielli M, Mayne PW. Load-displacement uncertainty of vertically loaded shallow footings on sands and effects on probabilistic settlement. Georisk 2012;6(1):50–69.
[36] Wu XZ. Trivariate analysis of soil ranking-correlated characteristics and its application to probabilistic stability assessments in geotechnical engineering problems. Soils Found 2013;53(4):540–56.
[37] Wu XZ. Assessing the correlated performance functions of an engineering system via probabilistic analysis. Struct Saf 2015;52:10–9.
[38] Wang Y, Rosowsky DV. Joint distribution model for prediction of hurricane wind speed and size. Struct Saf 2012;35:40–51.
[39] Zhang J, Huang HW, Juang CH, Su WW. Geotechnical reliability analysis with limited data: Consideration of model selection uncertainty. Eng Geol 2014; 181:27–37.