Reliability Engineering and System Safety xxx (xxxx) 110270

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Multivariate copula-based framework for stochastic analysis of landslide runout distance

Guotao Ma [a, b], Mohammad Rezania [a, b, *], Mohaddeseh Mousavi Nezhad [c], Kok-Kwang Phoon [d]

[a] School of Engineering, University of Warwick, Coventry, CV4 7AL, UK
[b] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China
[c] Department of Civil and Environmental Engineering, University of Liverpool, Liverpool, L69 3GH, UK
[d] Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, Singapore

ARTICLE INFO

Keywords:
Landslide
Copula theory
Non-gaussian random field
Probabilistic analysis

ABSTRACT

The increasing frequency of landslide disasters worldwide highlights the importance of accurate prediction of post-failure runout distances for risk management. However, the prediction of runout distances of gravitational mass flow remains challenging due to the inherent complex heterogeneity of geomaterials and its inter-correlated spatially varying mechanical properties. To address this challenge, this study proposes a novel multivariate stochastic method based on the combination of the Generalized Interpolation Material Point (GIMP) method and the multivariate copula-based approach. The method considers geotechnical uncertainties by simulating copula-based cross-correlated random fields from sparse field data and incorporating them into the GIMP analysis through Monte Carlo simulation (MCS). Two slope cases with similar geometries but different sources of probability information are presented to illustrate the effectiveness of the proposed method. Results show that considering the influence of multivariate random fields can significantly affect the post-failure analysis of landslides. Both slope cases show about over 40 % of all MCS samples exceed the deterministic case, which indicates that the current deterministic analysis notably underestimates the risk induced by large runout distances of landslides. This original study demonstrates the necessity to take interdependency spatial heterogeneity into consideration for post-failure modeling of landslides.

1. Introduction

Landslides are the most common natural disasters in the world, which frequently result in detrimental consequences and significant damages to human life and properties. On many occasions, with extensive runout of the landslides, substantial destruction can be observed on nearby structures. In the UK, the most significant loss of life (144 people) caused by a single landslide with runout distance of about 640 m occurred in 1966 when a waste tip formed by coal spoil failed and destroyed a school in the South Wales [4]. Thus, assessing the runout distance of landslides is significantly important as it allows for the prediction of possible catastrophic consequences and timely undertaking reasonable risk mitigation measures. However, geotechnical uncertainties (e.g., inherent spatial variability, transformation uncertainty, statistical uncertainty, etc.) [10,42,61] and large deformations of geomaterials [16,28] make the landslide post-failure runout motions extremely sophisticated and challenging to be predicted or evaluated by physical or numerical approaches.

Multiple uncertain variability and heterogeneity over space of geomaterials are widely admitted because they are formed by anthropological and uncontrolled natural process (e.g., weathering, deposition). It has been demonstrated that these uncertainties would influence the assessment of any geostructure reliability. Therefore, the quantitative characterization of the multiple uncertainties in the soil properties is the key focus in geotechnical engineering [27,36,42,31,28,29,9,13,16]). Random field (RF) theory was first adopted to mathematically characterise these spatial variability of soils [54], where the spatial distribution structure (i.e., autocorrelation) of a variable is commonly described by a theoretical autocorrelation function (Liu et al. 2017; [27,35]) and the dependence structure among different parameters is usually represented by a correlation coefficient matrix [17,24,31]. The cross-correlation matrix is regularly adopted to define the correlation intensity (interdependency) between each geotechnical parameters [11,12,19,31]. Recently, more and more research has started to combine RF theory with the large-deformation numerical methods (such as the Material Point Method, Smoothed Particle Hydro-

* Corresponding author.
E-mail address: m.rezania@warwick.ac.uk (M. Rezania).

https://doi.org/10.1016/j.ress.2024.110270
Received 2 October 2023; Received in revised form 28 May 2024; Accepted 7 June 2024
0951-8320/

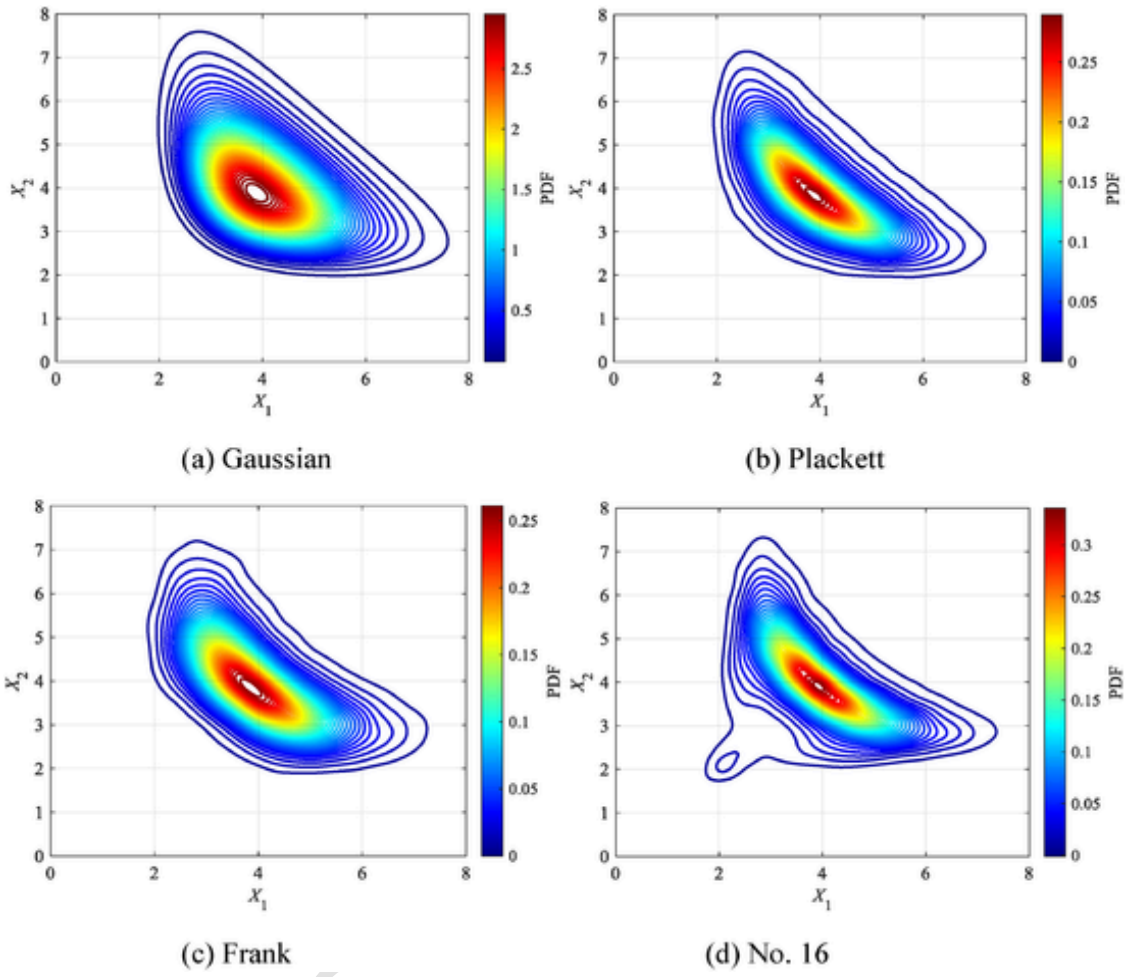


Fig. 1. Contour plots for the bivariate distributions of X_1 and X_2 using Gaussian, Plackett, Frank and No.16 copulas.

Table 1
PDFs and domains of distribution parameters associated with the selected distributions.

Distributions	$f(x,p,q)$	μ and σ^2	Range of p	Range of q
Trunc-normal	$\frac{1}{q\sqrt{2x}} exp\left(-1/2\left(\frac{x-R}{q}\right)^2\right) / \left[1-\Phi\left(-\frac{p}{q}\right)\right]$	$u = p$ $\sigma^2 = q^2$	$(-\infty, \infty)$	$(0, \infty)$
Lognormal	$\frac{1}{qx\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x-p}{q}\right)^2\right]$	$\mu = \exp\left(p+0.5q^2\right)$ $\sigma^2 = \left[\exp\left(q^2\right)-1\right]\exp\left(2p+q^2\right)$	$(-\infty, \infty)$	$(0, \infty)$
Trunc-Gumbel	$\frac{q\exp\{-q(x-p)-\exp[-q(x-p)]\}}{\{1-\exp[-\exp(pq)]\}}$	$\mu = p + 0.5772\ /\ q$ $\sigma^2 = \pi^2/\left(6q^2\right)$	$(-\infty, \infty)$	$(0, \infty)$
Weibull	$\frac{q}{p}\left(\frac{x}{p}\right)^{q-1} \exp\left[-(x/p)^q\right]$	$\mu = p\Gamma(1+1/q)$ $\sigma^2 = p^2\left[\Gamma(1+2/q)-\Gamma^2(1+1/q)\right]$	$(0, \infty)$	$(0, \infty)$

Note: Φ denotes the standard normal distribution function; Γ is the gamma function.

dynamics) in the Monte Carlo Simulation (MCS) framework to investigate the influence of spatial variability on post-failure of landslides, either univariate analysis for the undrained condition using S_u, undrained shear strength, with a solely S_u RF [25,28,29,43,55] or multivariate analysis (more than one variable) with independent RFs [14,33] or cross-correlated RFs using c, cohesion and φ, friction angle [31]. For the multivariate analysis, Nataf transformation method commonly adopted in above mentioned works, in which the off-diagonal entries quantify the correlation magnitude among each parameter. However, the Nataf transformation is regarded as a special case for Gaussian dependence structure, this implications do not certainly hold in specific cases [51]. In contrast, the consideration of various dependence structures among random variables introduces a significant impact on the features of joint probability distributions [51,57]. The interdependency among these variables can vary, leading to diverse characteristics in the joint probability distributions. Numerous studies have shown that the various dependence structures and interdependencies of multiple soil properties have significant influences on slope failures from slope stability analysis to probability of failure [16,39,57]. Nevertheless, so far, none of them has considered the interdependence characteristics in post-failure modelling of heterogeneous landslides. This research gap must be addressed. Moreover, given the complicated geological environments and limited resources, it is difficult to obtain a sufficient geotechnical data in most cases. With sparse field data,

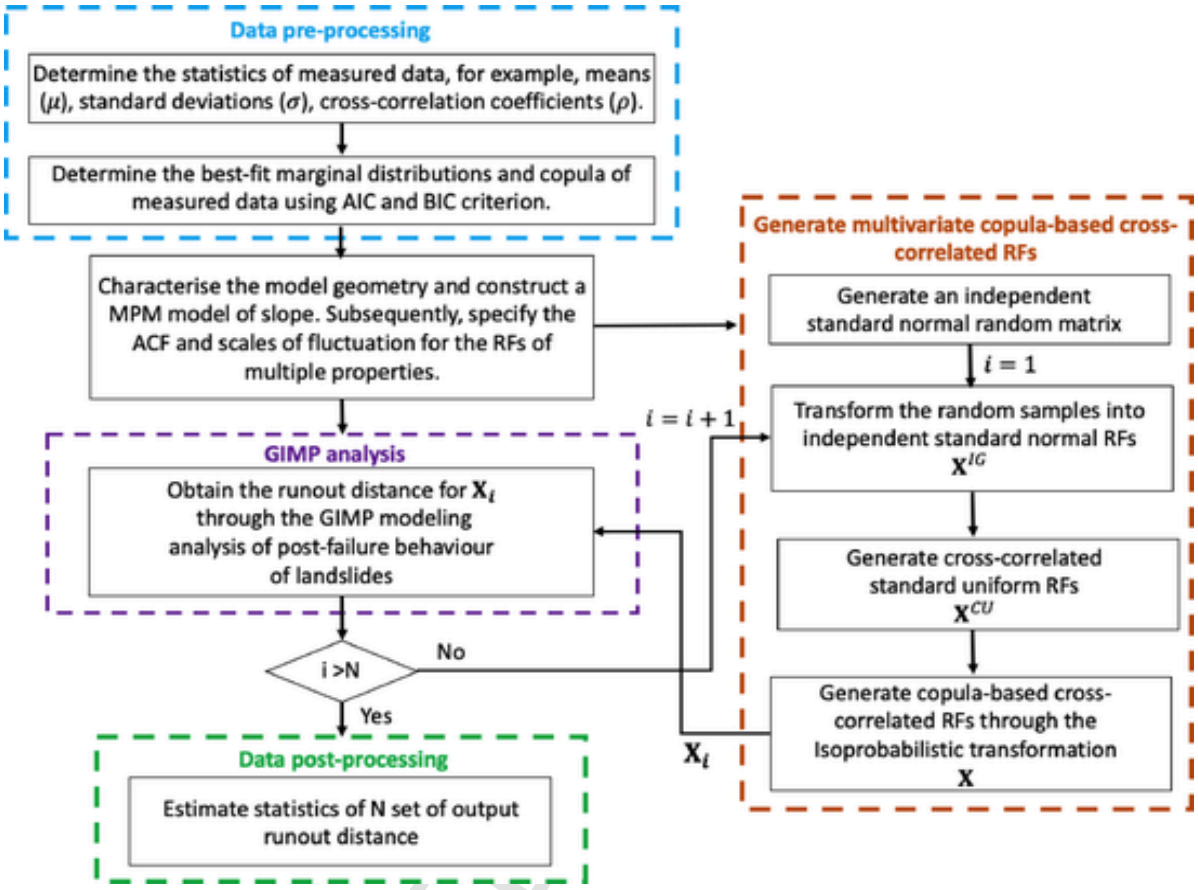


Fig. 2. Flowchart of landslide post-failure modelling with multivariate copula-based cross-correlated RF based on sparse field data.

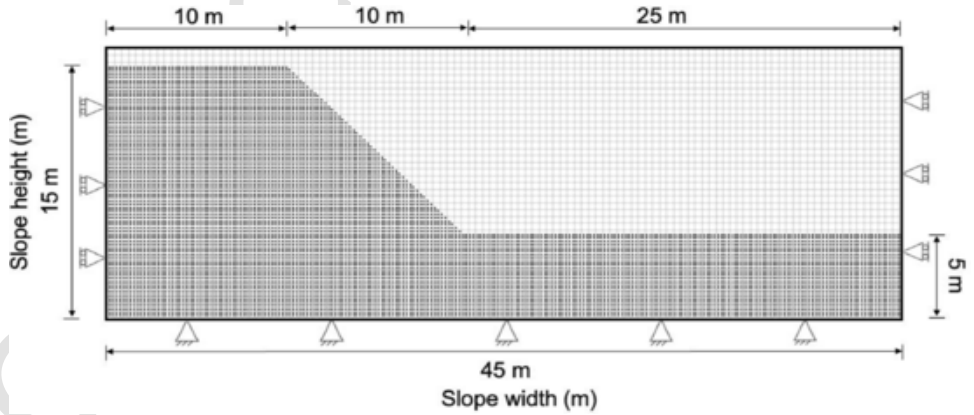


Fig. 3. Geometry and material point model for the soil slope.

probability distributions and interdependences of random variables would be approximately estimated with an inevitable uncertainty, which subsequently spreads and expands during stochastic calculation and ultimately leads to severe bias in final computing and data analysing. Consequently, a need to exactly model base distributions and dependence structures of geotechnical properties under incomplete sparse field data for reducing uncertainties in post-failure behaviours of landslides has arisen.

Fortunately, copula theory, a multivariate statistical model, has been developed to bridge the gap between marginal and joint distributions in data science [7,38,53]. By separating the characterisations of marginal distributions and interdependencies of variables, copula can integrate arbitrary probability distribution type into different dependence structures [44,63]. In recent years, several studies have explored the application of copula-based methods for statistical analysis within geotechnical engineering. For example, Masoudian et al. [32] proposed a multivariate hydro-mechanical framework with Gaussian copula-based RFs in finite element (FE) analysis to investigate the slope stability. Nguyen et al. [41] also adopted Gaussian copula for investigating the reliability of an unsaturated soil slope. However, it has been demon-

Table 2
Dataset of natural soils [49].

Test set	Cohesion, c (kPa)	Friction angle φ (°)
1	4.10	22.90
2	4.20	33.60
3	17	25
4	19.40	28.40
5	0	32.80
6	9.10	33.10
7	17	19.30
8	9.50	31.70
9	12.50	26.60
10	14.60	28.10
11	10.30	26.50
12	23.40	29.20
13	10.80	29.20
14	17.50	24.60
15	16.70	10.70
16	4.70	35
17	11	28.10
18	3.20	30.90
19	5	30.30
20	14.60	21.90
21	20	22
22	17.30	20.10
23	5.70	31
24	11	27.20
25	6.70	30.40

Table 3
Statistical data for the dataset.

Statistical parameters	cohesion, c (kPa)	Friction angle φ (°)
Mean	11.41	21.14
Standard deviation	6.1915	5.4708
Kendall correlation coefficient	−0.4672	
Pearson correlation coefficient	−0.5759	

strated that the Gaussian copula has its limitations when facing complex sparse field data or incomplete probability information which requires representation by non-Gaussian copulas [51]. This led Zhu et al. [64] to employ trivariate non-Gaussian copula-based RFs to model the spatial variability of liquid limit, plasticity index and moisture content of soils. Then, Wang et al. [57] further explored the interdependence between cohesion and friction angles by various non-Gaussian and Gaussian copulas to investigate their effect on slope reliability. Ng et al. [39] considered fabric orientation with different copula models for slope stability and risk assessment. Nguyen et al. [41] investigated the influences of copula selection on slope reliability using FE limit analysis. Among these studies, they employed the limit equilibrium method or FE with Gaussian or Non-Gaussian copula-based RFs to analyse failure or investigate reliability of hypothetical slopes for determining the stability or probability of landslide occurrence (i.e., P_f) in the initiation stage of slope failure, whereas giving very limited information about the consequences or post-failure behaviours of landslides in spatially varying soils. The post-failure stage of landslides is often oversimplified and the consequences (i.e., runout distances) are not quantitatively evaluated. While the post-failure features of landslides are particularly significant for risk assessment of slopes and hazard mitigation. To date, to the best of the author's knowledge, no research has been carried out to quantify landslide possible runout motions with considering complex interdependency multi-variables based on real field data. Therefore, modelling the post-failure of the landslide disasters considering multi-interdependencies of soil properties is still an open question. The primary challenge is incorporating complex, interdependent multivariate spatial variability of soil properties into large deformation modeling algorithms, particularly when only incomplete data are available.

In this paper, a new hybrid multivariate copula-based mechanical-statistical computational approach based on incomplete sparse field data is proposed to consider geotechnical uncertainties and quantitively evaluate the possible landslide runout distances. This research paper will present new insight through runout distance analysis of heterogeneous landslides with considering multiple uncertainties. Generalised interpolation material point method (GIMP) [2,3] is adopted to capture the runout motions of landslides. Meanwhile, the interdependent geotechnical uncertainties are featured by copula-based multivariate RFs. In considering the most appropriate copulas in the study, different copulas will be considered with the differences and similarities, such as tail-dependences features. Two slopes with similar geometries but different sources of probability information are analysed and illustrated for bivariate and multivariate evaluation respectively, which involve copula-based cross-correlated RFs with different properties based on real measured field data. Using multiple realisations of Monte Carlo simulation (MCS), the runout motion results of each fully saturated landslide are statistically analysed, and the associated risk assessment measure is also provided. This study is expected to provide an example of combining the stochastic analysis of multiple uncertain course with copulas on the large deformation geotechnical problems.

2. Copula-based model for soil interdependency

2.1. Basics of copula

Copula theory was proposed by Sklar [48], it indicates that any multivariate joint distribution function can be decomposed by a corresponding marginal distribution function and a copula function. The copula function determines the correlation among variables, including the correlation coefficient and the type of dependence structures [47]. In this theory, assuming $F\left(x_1, x_2, \ldots, x_n\right)$ is the joint cumulative distribution function (CDF) of a random vector $\boldsymbol{X} = \left[X_1, X_2, \ldots, X_n\right]$, and the marginal CDFs are $F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_n\left(x_n\right)$, a joint CDF can be defined as the probability that X is less than or equal to a specific numerical vector $[x_1, x_2, \cdots, x_n]$

$$\begin{aligned} & F\left(x_1, x_2, \ldots, x_n\right) \\ & = P\left(X_1\right. \\ & \leq x_1, X_2 \\ & \leq x_2, \ldots, X_n \\ & \left.\leq x_n\right) \end{aligned} \tag{1}$$

where $P(\bullet)$ denotes the probability. According to the *Skalar's theorem* [47,48], the joint CDF of $\mathbf{X}$ can be expressed in terms of a copula function and its univariate marginal distributions. Therefore, there exists a unique n-dimensional copula function

$$\begin{aligned} & F\left(x_1, x_2, \ldots, x_n\right) \\ & = C\left[F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_n\left(x_n\right) ; \theta\right] \\ & = C\left(u_1, u_2, \ldots, u_n ; \theta\right) \end{aligned} \tag{2}$$

where $C\left(u_1, u_2, \ldots, u_n ; \theta\right)$ is a n-dimensional variate copula function describing the dependence structure among the variables $x_1, x_2, \ldots, x_n$, and θ is a copula parameter describing the dependence among $F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_n\left(x_n\right)$. Since $F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_n\left(x_n\right)$ represent cumulative probabilities, they are usually denoted as $u_1, u_2, \ldots, u_n$ ranging from 0 to 1. Thus, $u_1, u_2, \ldots, u_n$ are standard uniform variables. By taking derivatives of Eq. (1), the joint PDF $f\left(x_1, x_2, \ldots, x_n\right)$ of the variables $x_1, x_2, \ldots, x_n$ can be given as

Table 4
AIC and BIC values for the candidate distributions.

Parameters	Normal	Lognormal	Gumbel	Weibull
	[AIC, BIC]	[AIC, BIC]	[AIC, BIC]	[AIC, BIC]
Cohesion	**163.44, 165.88**	166.58, 169.02	167.94, 170.38	174.22, 176.66
Friction angle	158.91, 161.35	170.27, 172.71	196.93, 199.37	**156.45, 158.89**

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \frac{\partial^2 C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n);\, \theta)}{\partial F_1(x_1) \partial F_2(x_2), \ldots, \partial F_n(x_n)} \frac{\partial F_1(x_1)}{\partial F_1(x_1)} \frac{\partial F_2(x_2)}{\partial F_2(x_2)}, \ldots, \frac{\partial F_n(x_n)}{\partial F_n(x_n)} \\ &= D(F_1(x_1), F_2(x_2), \ldots, F_n(x_n);\, \theta) \prod_{i=1}^{n} f_i(x_i) \\ &= D(u_1, u_2, \ldots, u_n; \theta) \prod_{i=1}^{n} f_i(x_i) \end{aligned} \quad (3)$$

where $f_1(x_1) f_2(x_2), \ldots, f_n(x_n)$ are the marginal PDFs of the variables, and $D(u_1, u_2, \ldots, u_n; \theta)$ is the multivariate copula density function related to $C(u_1, u_2, \ldots, u_n; \theta)$.

2.2. Candidate copulas

Based on literature, there are many kinds of copulas [38], each copula has its own and unique dependence structure. In this study, four widely used two-dimensional copulas, namely, Gaussian, Plackett, Frank, and No.16 are selected as the set of candidate copulas in the bivariate case [47,59,60]. Details about these bivariate copula functions $C(u_1, u_2; \theta)$ as well as their density functions $D(u_1, u_2; \theta)$ are discussed in [51] and [57]. The selected copulas are capable of describing both the positive and negative cross-correlations between cohesion and friction angle, in which the allowed values of the negative correlation coefficient can reach to -1. In Fig. 1, the contours plot the bivariate normal distributions (X_1 and X_2) by using the four copula functions. It can be observed that the bivariate distributions constructed from different copulas are particularly different, which indicates that each copula function has its own dependence structure. Especially, the Gaussian, Plackett, and Frank copulas do not have tail dependence, meaning that extreme events in one variable are not strongly correlated with extreme events in the other variable. In contrast, the No.16 copula demonstrates a unique characteristic of having only lower tail-dependent features, as illustrated in Fig. 1d. This indicates that while extreme values in one variable may not be strongly correlated with extreme values in the other variable in general, there is a significant correlation in the lower tail of the distribution. The presence of lower tail dependence, as demonstrated in previous research [57], significantly influences the assessment of geotechnical reliability. This underscores the practical importance of tail dependence in comprehending the joint behaviour of soil properties, especially in extreme conditions.

Moreover, when considering multiple variables in a copula function, the copula model as well as the inherent interdependency among variables would become more sophisticated and complicated [64]. In this work, multivariate Gaussian copula and multivariate t copula are used for trivariate analysis due to their wide applicability in geotechnical engineering [41]. The multivariate Gaussian Copula and density function can be expressed as

$$\begin{aligned} &C(u_1, u_2, \ldots, u_n; \theta) \\ &= \Phi_n\left(\Phi^{-1}(u_1) \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n); \theta\right) \end{aligned} \quad (4)$$

$$D(u_1, u_2, \ldots, u_n; \theta) = |\theta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \zeta' \left(\theta^{-1} - \mathbf{I}\right) \zeta\right] \quad (5)$$

where θ is the n-dimensional symmetric positive definite matrix, which consists of $0.5n(n-1)$ correlation coefficients $\theta_{ij} \in [-1, 1]\, (i = 1, 2, ..., n-1; j = 2, 3, ..., n)$ from $n-$dimensional Gaussian copula; Φ_n is the $n-$dimensional standard normal distribution function with the correlation coefficient matrix θ; $\Phi^{-1}(\bullet)$ is the inverse function of one dimensional standard normal distribution $\Phi(\bullet)$; $\zeta' = \left(\Phi^{-1}(u_1) \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right)$ is the standard normal distribution variables; and $\mathbf{I}$ is the identity matrix. The relation between θ_{ij} and τ_{ij} can be expressed as

$$\tau_{ij} = \frac{2 \arcsin \theta_{ij}}{\pi} \quad (6)$$

Therefore, according to above equation, θ_{ij} can be obtained by τ_{ij}. Multivariate t copula is also called Student t copula, the copula function and density function can be expressed as

$$\begin{aligned} &C(u_1, u_2, \ldots, u_n; \theta) \\ &= T_n\left(T_v^{-1}(u_1), T_v^{-1}(u_2), \ldots, T_v^{-1}(u_n); \theta, v\right) \end{aligned} \quad (7)$$

$$\begin{aligned} D(u_1, u_2, \ldots, u_n; \theta) &= |\theta|^{-\frac{1}{2}} \frac{\Gamma\left(\frac{v+n}{2}\right)\left(\Gamma\left(\frac{v}{2}\right)\right)^{n-1}}{\left(\Gamma\left(\frac{v+1}{2}\right)\right)^{n}} \\ &\times \frac{\left(1 + \frac{1}{v} \zeta' \theta^{-1} \zeta\right)^{-\frac{v+n}{2}}}{\prod_{i=1}^{n} \left(1 + \frac{\zeta_i^2}{v}\right)^{-\frac{v+1}{2}}} \end{aligned} \quad (8)$$

where θ is the n-dimensional symmetric positive definite matrix, which consists of $0.5n(n-1)$ correlation coefficients

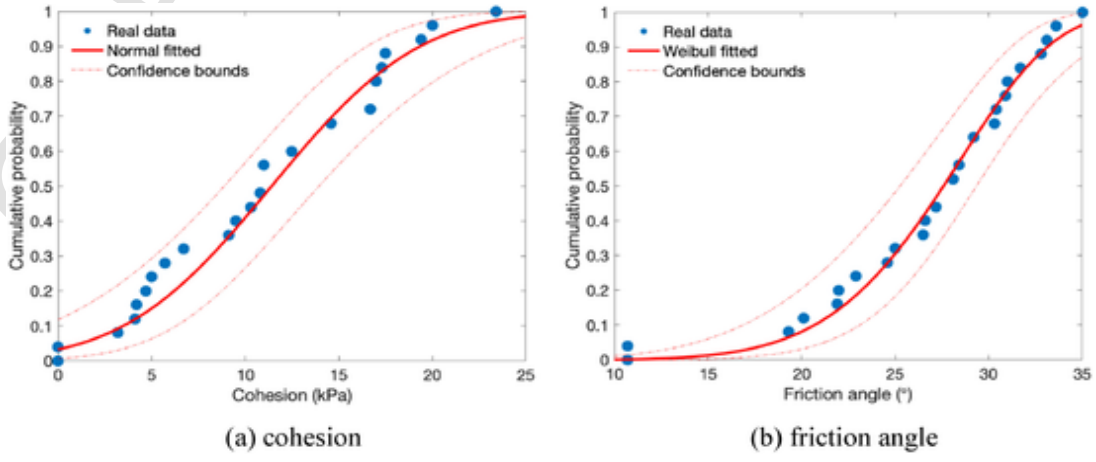


Fig. 4. Best-fit marginal CDF of the data.

Table 5
Copula parameter θ, AIC, and BIC values for the candidate copulas.

Parameters	Gaussian	Plackett	Frank	No.16
θ	−0.669	0.1055	−5.1703	0.0174
AIC	−9.166	−11.400	−11.617	**−12.832**
BIC	−7.947	−10.181	−10.398	**−11.6135**

$\theta_{ij} \in [-1, 1]\,(i = 1, 2, ..., n-1; j = 2, 3, ..., n)$ from the n-dimensional t copula; v is the degrees of freedom, $T_n(\bullet; \theta, v)$ is the n-dimensional standard t distribution function with the correlation coefficient matrix θ and degrees of freedom v.

For above bivariate and multivariate copulas, the maximum likelihood estimation (MLE) method is adopted to obtain θ, the corresponding likelihood function can be established as

$$L(\theta) = \sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}, \ldots, u_{ni}; \theta\right) \tag{9}$$

where θ is the unknown parameter of copula function; $(u_{1i}, u_{2i}, \ldots, u_{ni})$ is the empirical distribution values of the original known data $(x_{1i}, x_{2i}, \ldots, x_{ni})$, which are defined as

$$u_{1i} = \frac{rank\left(x_{1i}\right)}{N+1} \tag{10}$$

$$u_{2i} = \frac{rank\left(x_{2i}\right)}{N+1} \tag{11}$$

$$u_{ni} = \frac{rank\left(x_{ni}\right)}{N+1} \tag{12}$$

where, rank $(\bullet)$ denotes the ranking function, for instance, rank(x_{1i}) presents the rank of x_{1i} among the list $(x_{1i}, x_{2i}, \ldots, x_{ni})$ in an ascending order. By using the maximum likelihood method, the maximum value of θ in Eq. (9) is

$$\theta = \text{argmax}\, L(\theta) \tag{13}$$

Then, to realistically capture the characteristics and interdependence of geotechnical properties, the best-fit copula from the selected copulas needs to be identified. This can be done using the Akaike information criterion (AIC) [1] and Bayesian information criterion (BIC) scores [45]. The copula producing the smallest AIC and BIC score are taken as the best-fit copula.

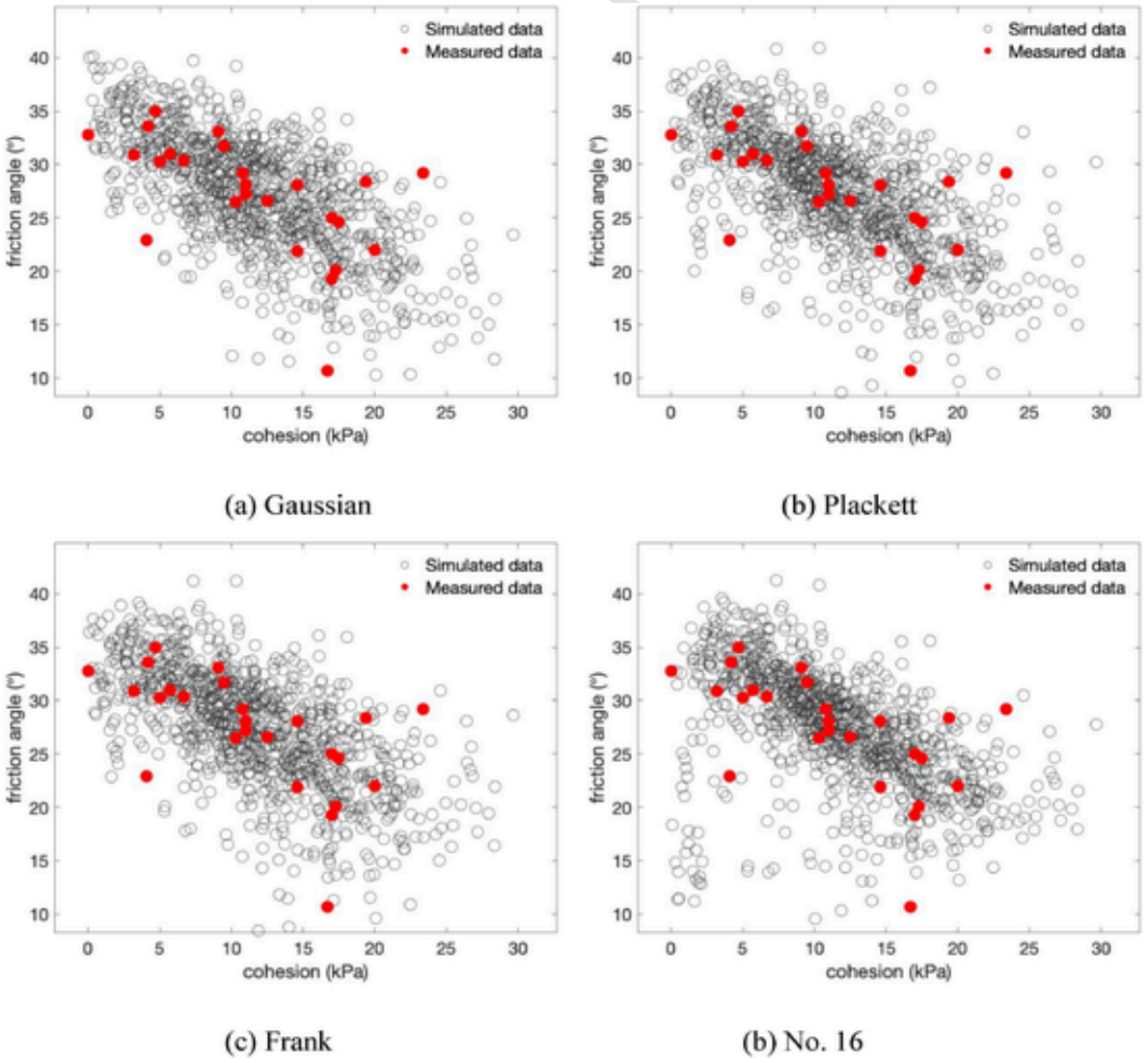


Fig. 5. Scatter plots of measured and simulated c and φ values for the dataset. (a) Gaussian copula, (b) Plackett copula, (c) Frank copula, and (d) No. 16 copula.

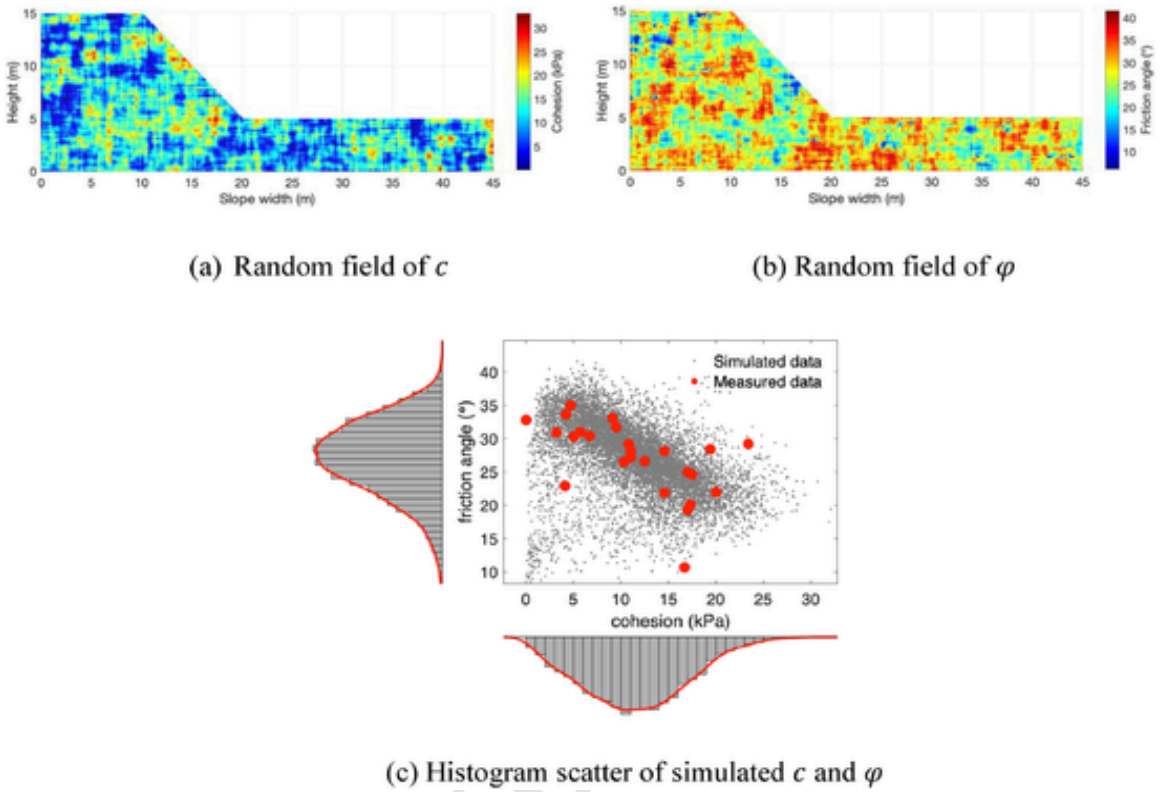


Fig. 6. A typical realization of the copula-based cross-correlated RFs of c and φ values.

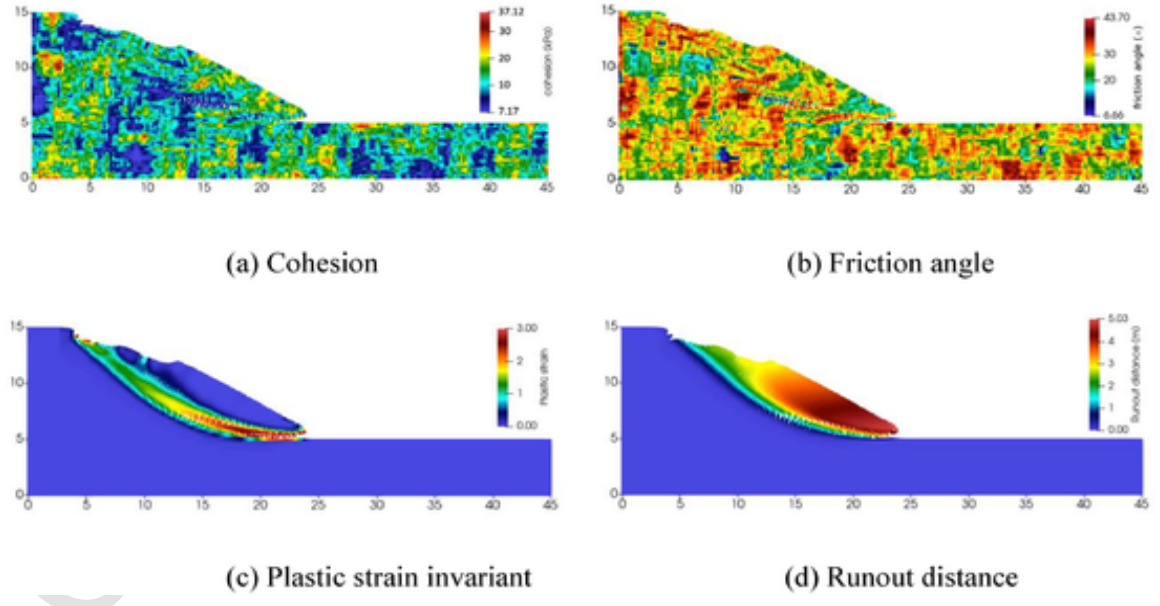


Fig. 7. A typical sample of copula-based landslide model at final stage of analysis.

$$\mathrm{AIC} = -2\sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}, \ldots, u_{ni};\ \theta\right) + 2k_2 \tag{14}$$

$$\mathrm{BIC} = -2\sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}, \ldots, u_{ni};\ \theta\right) + k_2 \ln N \tag{15}$$

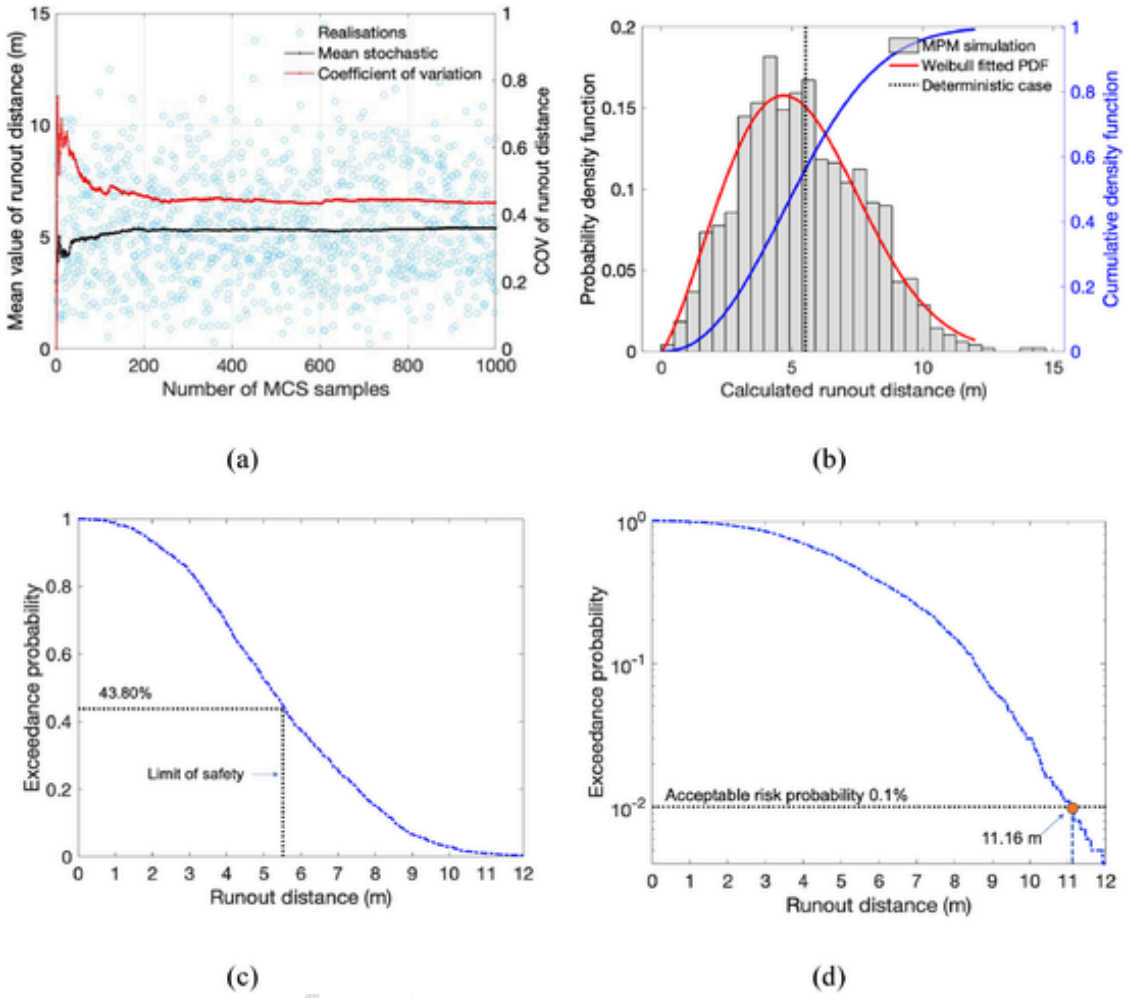


Fig. 8. Stochastic data, histogram, PDF, CDF, and exceedance probability of the runout distance.

where $(u_{1i}, u_{2i}, \ldots, u_{ni})$ is the empirical distribution values of the original known data $(x_{1i}, x_{2i}, \ldots, x_{ni})$, which can be determined by Eq. (12); k_2 is the number of copula parameters.

2.3. Best-fit marginal distribution

In this study, four commonly used marginal distributions, namely Normal truncated below zero (Trunc-normal), Lognormal, Gumbel truncated below zero (Trunc-Gumbel), and Weibull distributions are adopted to fit the distributions of geotechnical properties ([19]; Liu et al. 2017; [31]). These distributions can guarantee positive values for the simulated geotechnical data. In Table 1, the PDFs and associated distribution parameters (p, q) of the four probability distributions are illustrated. Both the method of moments and MLE are used to estimate the distribution parameters. The equations of (mean value μ, standard deviation σ) with the associated distribution parameters (p, q) are summarised. The distribution resulting in the smallest AIC and BIC scores is regarded as the best-fit marginal distribution [27].

3. Generation of copula-based multivariate random fields

The spatial variability mathematically characterized by the RF theory [54] is combined with the interdependences of soil properties featured by the copulas to generate copula-based cross-correlated multivariate RFs. The following steps are used to generate standard copula-based multivariate RFs:

1. First step is to individually generate independent standard normal RFs for each property using autocorrelation function (ACF) and Cholesky matrix decomposition method (CMD) [6,17,22,62]. In this work, CMD is utilised for its convenience in dealing with cross-correlated RFs. Single exponential ACF, the most widely used ACF model to simulate inherent spatial variability of soil properties [22,27], is adopted to build the autocorrelation matrix in the domain, which is expressed as:

$$\rho\left(\tau_{x_{ij}}, \tau_{y_{ij}}\right) = \exp\left[-2\left(\frac{\tau_{x_{ij}}}{\delta_h} + \frac{\tau_{y_{ij}}}{\delta_v}\right)\right] \tag{16}$$

where, $\rho(\tau)$ is the correlation coefficients, the $\tau_x = |x_i - x_j|$ and $\tau_y = |y_i - y_j|$ are the absolute distances between two spatial locations in horizontal and vertical directions, respectively; and δ_h and δ_v are the autocorrelation distances in horizontal and vertical directions, respectively. Note that the δ_h and δ_v can be chosen equal to each other (2.0 m), which leads to an isotropic RF. It is important to acknowledge that the anisotropy associated with spatially varying soils (i.e., δ_h and δ_v) is expected to have a significant impact on runout motions of landslides, as discussed in previous publications [5,27,29]. However, this specific aspect has

Table 6
Dataset of natural soils (Soenksen et al. 2003).

Test set	Cohesion, c (kPa)	Friction angle φ (°)	Unit weight, γ (kN/m³)
1	7.80	25.80	16.80
2	2.50	37.10	14.60
3	8.50	32.80	15.70
4	9.30	16.40	17.50
5	2.70	34.30	15.60
6	9.80	28.80	15.80
7	5	32.70	14.60
8	14	26.60	16.10
9	12	31.40	16.60
10	13	26.60	18.20
11	11.70	23	17.70
12	14.60	26	17.70
13	8.60	27	17.40
14	8.70	35.60	18
15	15.10	36.90	16.30
16	12.80	14	18
17	14.10	21.50	13.50
18	13.80	26.60	17.20
19	16.30	13	17.10
20	7.70	17.70	17.60
21	3.50	24.60	16.70
22	36	15.30	18.40
23	8.50	14	18.60
24	22.20	26.70	18.60
25	44.70	15.80	17.20
26	0	15.30	17.60
27	8.10	35.90	18.40
28	16	35.20	17.30
29	7.10	28.40	19.20
30	11.20	20.30	17.80
31	2.20	31.80	18.70
32	7	21.80	18.70
33	10	42.10	15.90
34	7	32.10	15.30
35	15.70	21.80	16.50
36	12.60	23.10	16.20
37	20.40	19.30	15.20
38	9.40	33.80	12.60
39	3.10	27.90	16.70
40	6.50	30.60	18.30
41	2.30	36.40	17.20
42	7.40	33.10	17.70

Table 7
Estimated distribution parameters.

Variables	Distributions	First parameter	Second parameter	Mean	Standard deviation
c (kPa)	Gumbel	0.15	7.13	10.92	8.42
φ (°)	Weibull	3.91	29.43	26.64	7.62
γ (kN/m^3)	Weibull	14.29	17.55	16.92	1.44

not been considered in the present paper, as its primary focus is the implementation of a multivariate copula in MPM models.

2. The second step is transformation of independent standard normal RFs into cross-correlated standard uniform RFs based on copula parameters (the cross-correlated coefficients). This step is to couple the RFs of different soil properties. In this work, the procedure for generating bivariate and multivariate RFs based on the best-fit copulas will be elaborated comprehensively for each case.
3. Finally, isoprobabilistic transformation [19] of cross-correlated uniform RFs into cross-correlated non-normal RFs is carried out by separately adding best-fit marginal distributions of each property of soil.

4. GIMP-based calculation of runout distances

In reality, stability of a dense granular assembly can be significantly reduced by a pore pressure of interstitial fluid. Water content affects the soil strength, the hydraulic pressure reduces the effective stress of the solid skeleton in the fully saturated strata, while the matrix suction strengths the particle connections in the partially saturated strata. Landslide disasters rarely occur without the presence of a partially saturated deposit and transition from a solid-like state to a fluid state. With the RFs of the soil properties determined, as explained above, the runout distance of each copula-based RF sample (i.e., a heterogeneous landslide) is calculated using a fully coupled hydro-mechanical MPM model. Darcy's law is adopted to govern the motion of interstitial water, as

$$n\left(v_i^w - v_i^s\right) = \frac{k_{ij}}{\rho^w g}\left[-p_j^w + \rho^w\left(g_j - \overset{w}{v}_j\right)\right] \tag{17}$$

where, n is the porosity, v_i is the velocity, $k_{ij}$is the tensor of permeability, ρ is the density, g_i is the gravitational acceleration, p is the hydraulic pressure, w and s represent the water and solid phases, respectively. The acceleration of the fluid can be obtained as

$$\rho^w \overset{w}{v}_j = -\frac{n\rho^w g}{k_{ij}}\left(v_j^w - v_j^s\right) - p_i^w + \rho^w g_i \tag{18}$$

The conservation of momentum of the mixture that used to govern the motion of the solid is

$$(1-n)\,\rho^s \overset{s}{v}_i + n\rho^w \overset{w}{v}_j = \sigma_{ij,j} + (1+n)\,\rho^s b_i + n\rho^w b_i \tag{19}$$

where, $\sigma_{ij} = \sigma'_{ij} - p^w\sigma_{ij}$ is the tensor of total stress, σ'_{ij} is the effective stress of the solid, σ_{ij} is the delta function and b_i is the body force. Note that the negative sign comes from the convention that the tensile force is positive. Therefore, substituting the acceleration of the fluid, the acceleration of the solid can be obtained.

The spatial discretization schemes for these equations are derived using GIMP which introduces alternative grid shape function,S_I, and particle characteristic function, $\chi_p(x)$ [3] to alleviate the grid-crossing instability of the original MPM. The implementation of computational cycle of GIMP can be found in literature, e.g., Ma et al. [27,31], which is summarized in the Appendix section in Ma et al. [31]. Additionally, the modified update stress last (MUSL) is adopted in this work as for the time when the stress is updated, which gives better computational stability [37] compared with update stress first [2] and update stress last (Sulsky et al. 1994). In this study, the soil properties were modelled as elastic-perfectly plastic materials with the Mohr–Coulomb failure criterion. More details regarding the fully coupled hydro-mechanical GIMP for saturated granular materials, such as computational validation and stability of the numerical solution can be found in Liu et al. [21] and Mroginski et al. (2020).

5. Implementation procedure

Fig. 2 shows the proposed framework, which consists of four main modules: i) data pre-processing; ii) generation of multivariate copula-based cross-correlated RFs which has been validated in the literature (e.g., [32,57,58]); iii) deterministic GIMP analysis for obtaining runout distance of landslides that has been validated from Li et al. [20] and Ma et al. [29]; iv) data post-processing.

The details of each step in this study are summarized as follows:

1. Determine the statistics of measured data, for example, means (μ), standard deviations (σ), and cross-correlation coefficients (ρ).
2. Determine the best-fit marginal distributions and copula of measured data using AIC and BIC criterion.

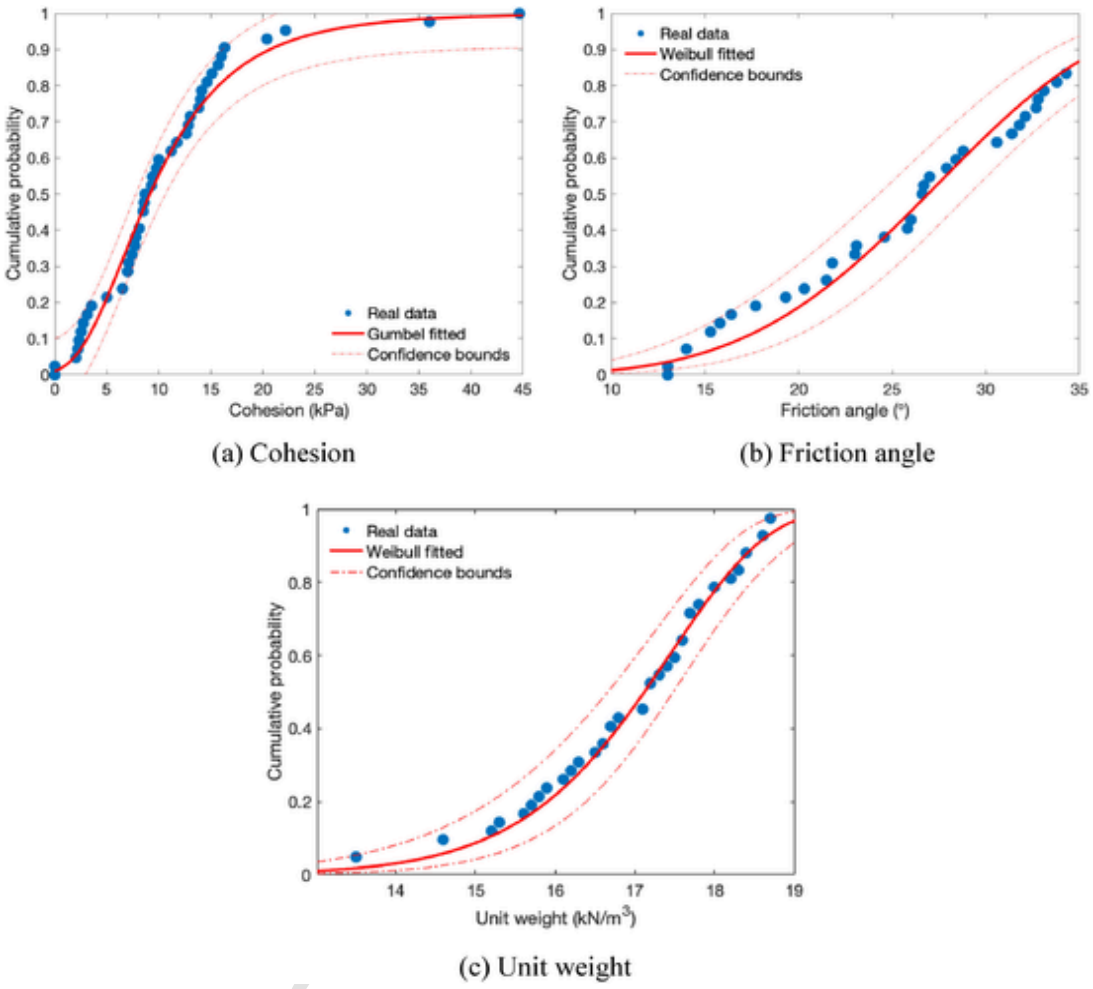


Fig. 9. Best-fit marginal CDFs of the measured data presented in Table 6.

Table 8
Estimated multivariate copula correlation coefficients for the measured data.

Variables	Cohesion	Friction angle	Unit weight
Cohesion	1	−0.3321	−0.0138
Friction angle	−0.3321	1	−0.2609
Unit weight	−0.0138	−0.2609	1

3. Characterise the model geometry, construct a material point model of the studied slope, and specify the ACF and scales of fluctuation for the RFs of multiple properties.
4. Extract centroid coordinate data (x_i, y_j) of all RF elements.
5. Perform MCS, where N independent standard normal random sample matrices of size $n \times m$ are simulated (m is the number of variates, i.e., if considering trivariate models, $m = 3$). Then, the random samples are treated as inputs to generate N realisations of the multivariate copula-based cross-correlated RFs ($\mathbf{X}$).
6. Assign the N realisations onto the MPM model to calculate runout distance of the studied model. A total of N values of runout distance are obtained.
7. Estimate the statistics of output runout distance, for instance, best-fit distributions, means, standard deviations, exceedance probability.

6. Illustrative examples

In this section, two hypothetical fully saturated slope examples with similar geometries, as referenced in [46], are investigated. But these examples differ in their sources of incomplete measured data, and they highlight the applicability of the proposed new stochastic multivariate analysis for bivariate and trivariate cases. It's worth noting that the slope models, incorporating varying probability information, have been frequently used to demonstrate the feasibility of probabilistic analysis, as indicated in earlier studies [31,39,52]. The first example aims to investigate the runout distances of heterogeneous landslides with measured cohesion c and friction angle φ data using bivariate copula-based method, while the second example aims to analyse landslide runout distances with considering trivariate cross-correlated model based on incomplete data information (c, φ, and unit weight γ) in the field. Both models incorporate probability information, inherent interdependencies, and spatial variability, using multivariate copula-based RFs at the material point level. Additionally, a deterministic case has been adopted using GIMP modeling in this study, serving as a benchmark. It's pertinent to highlight that the accuracy of the GIMP procedure has previously been validated and referred in the literature [23]. Given that the main focus here is on analysing runout responses of landslides after failure, only cases with a material point displacement larger than 0.1 m are considered [58].

As shown in Fig. 3, single-layered material point slope model has a height of 15 m and a slope gradient is 45◦. Generally, finer particle

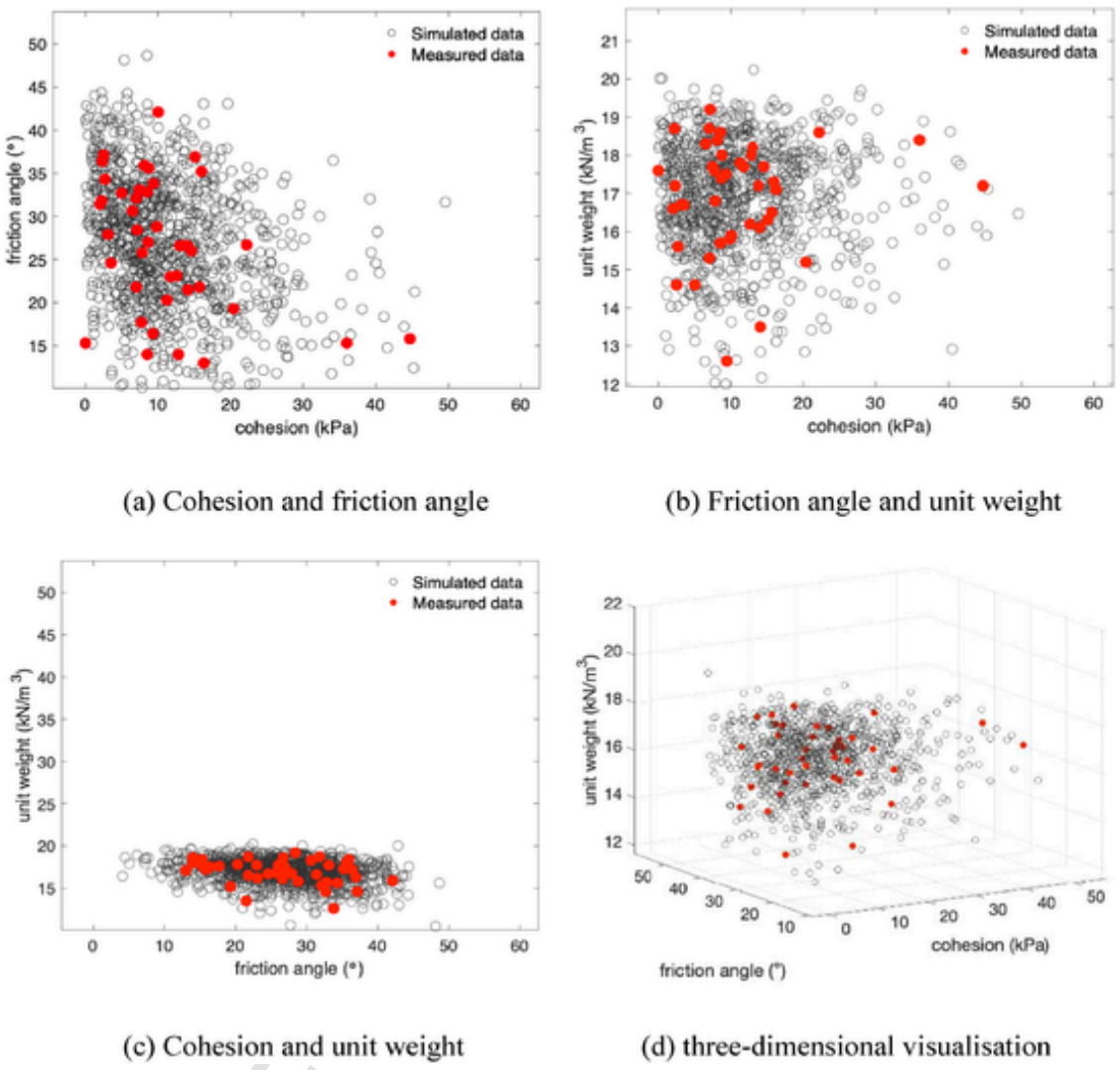


Fig. 10. A typical multi-dimensional set of cross-correlated random variables generated with multivariate Gaussian copula.

modelling can lead to more accurate simulation results, however the computational expenses will significantly increase. To balance the accuracy and computational expenses, a 4-noded square mesh with size of 0.2 m is adopted in this study, in which 4 equally spaced material points with a size of 0.1 m are placed in each mesh element. The total number of material points is 10,090. The RF is generated with a cell size equal to material point domain (0.1 m), where the properties of a single cell correspond to one material point. Therefore, as RF cell is 4 times smaller than the smallest autocorrelation distance, the spatial variability is adequately captured [19]. The bottom of the slope is a fixed boundary and roller boundary conditions are adopted for both lateral sides of the domain. The boundaries are impermeable in the normal direction, ensuring the slope remains saturated until the end of the analysis, thereby simulating undrained conditions. This study assumes a constant water table under undrained conditions and conducts hydro-mechanical analyses with the assumption that the slope is fully saturated throughout. The *in-situ* stresses are generated quasi-statically using gravitational loading. The total time for the calculation is 10 s, when soil deposits become stable according to kinematic energy and unbalanced forces of the system [18].

6.1. Example I: bivariate analysis of runout distances for landslides

In the first example, the post-failure modelling of heterogeneous landslides is conducted with measured shear strength parameters (c and φ). In this model, the density of water ρ_w is 1000 kg/m^3, the porosity n is set to 0.3, and the coefficient of permeability $k$is 1×10^{-4} m/s for the matrix, the compressible modules of water K_w is 6.146×10^9 Pa, the compressible modules of the soil skeleton $K_s = 1.724 \times 10^8$ Pa. Other parameters are considered constant, i.e., Young's modulus E = 100 MPa, Poisson's ratio ν = 0.35, and unit weight γ = 20 kN/m^3, as it has been found that their contributions on landslide runout distances are not significant [8].

Statistical analyses of measured data

An extracted dataset of c and φ measured *in-situ* on silty loam sampled from the Elkhorn River basin, north-eastern Nebraska, United States [49], is used to derive the statistical characteristics of the natural soils (as shown in Table 2).

Table 3 summarizes the means, standard deviations, and cross-correlation coefficients of c and φ values from the dataset. Four candidate marginal distributions, namely, Normal distribution, Lognormal distribution, Gumbel, and Weibull distributions are selected as the set

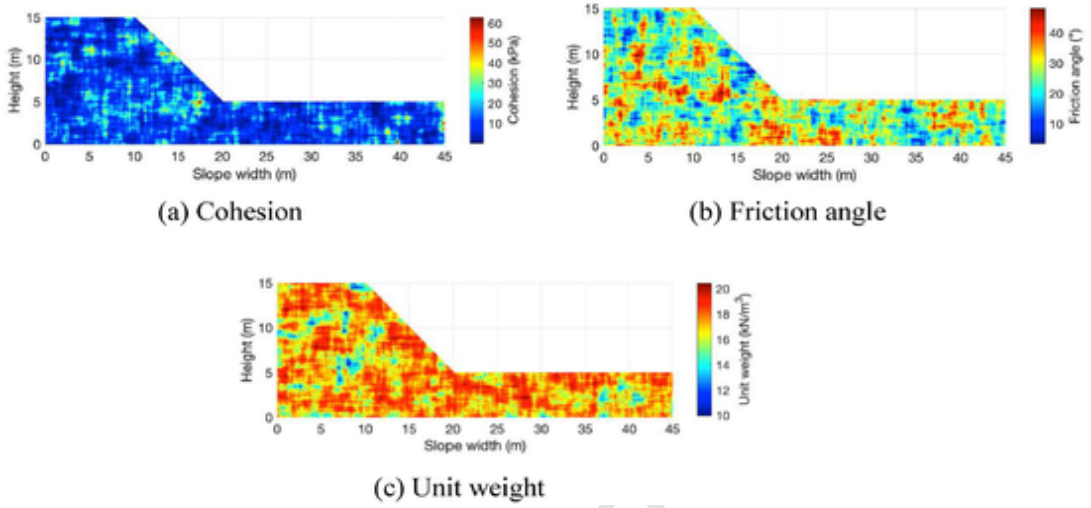


Fig. 11. A typical realisation of the copula-based multivariate RFs.

of candidate distributions to fit the marginal distributions of the measured c and φ values. Both the AIC and BIC are adopted to identify the best-fit marginal distribution of the c and φ values. Based on the dataset, the AIC and BIC scores for the candidate marginal distributions are obtained, which are summarized in Table 4.

From the values in Table 4, the normal distribution with the smallest AIC and BIC scores is identified as the best-fit marginal distribution for c, while the best-fit marginal distribution for φ is the Weibull distribution. Fig. 4 shows the corresponding identified best-fit marginal cumulative distribution functions of c and φ; for comparison, the figure also shows the empirical cumulative probabilities of the measured data. It can be found that the identified best-fit marginal distributions of c and φ match very well with their empirical cumulative probability data points.

After determining the best-fit marginal distributions of c and φ values, the next step is to identify the best-fit copula function to the dependence structure between the measured c and φ. As mentioned earlier, four two-dimensional copulas, including Gaussian copula, Plackett copula, Frank copula, and No. 16 copula are selected as the candidate copulas for characterizing dependence between the measured c and φ values. Similarly, the best-fit copula among the set of candidate copulas can be identified using the AIC and BIC scores:

$$\mathrm{AIC} = -2\sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}; \theta\right) + 2k_2 \tag{20}$$

$$\mathrm{BIC} = -2\sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}; \theta\right) + k_2 \ln N \tag{21}$$

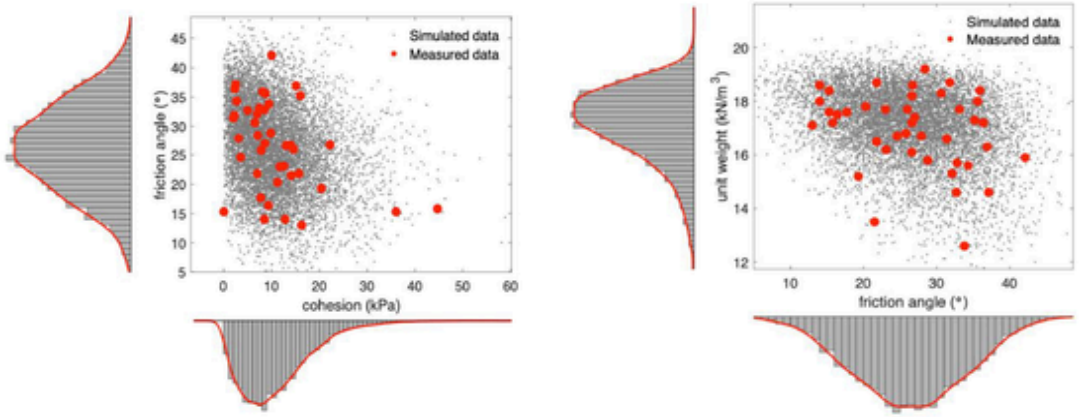
where $D\left(u_{1i}, u_{2i}; \theta\right)$ is the copula density function; $\sum_{i=1}^{N} \ln D\left(u_{1i}, u_{2i}; \theta\right)$ is the logarithm of the likelihood function of a specific copula; k_2 is the number of copula parameters, which is equal to 1 because all of selected copulas are single-parameter copula; and $\left\{\left(u_{1i}, u_{2i}\right), i = 1, 2, \ldots, N\right\}$ are the empirical distribution values of measured data defined as $u_{1i} = \frac{\mathrm{rank}(c_i)}{N+1}$, $u_{2i} = \frac{\mathrm{rank}(\varphi_i)}{N+1}$. Note that the rank($c_i$) or rank($\varphi_i$) denote the rank of c_i or φ_i in an ascending order, and $\left(u_{1i}, u_{2i}\right)$ are realisations of standard uniform variables.

Table 5 shows the θ, AIC, and BIC values associated with the copulas for the dataset. Both AIC and BIC values indicate that the No.16 copula is the best-fit copula for the measured data.

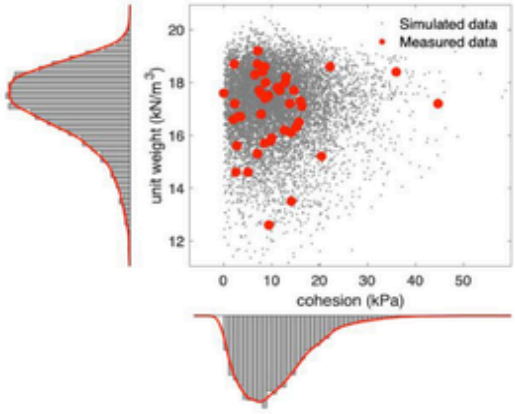
Constructing copula-based cross-correlated c-φ RFs

Based on the identified best-fit marginal distribution and copula, the copula-based cross-correlated RFs of c and φ is ready to be generated [39,41,57]. Firstly, a standard bivariate independent Gaussian RF $\mathbf{X}^{IG} = \left[\mathbf{X}_c^{IG}, \mathbf{X}_\varphi^{IG}\right]$ is generated. Subsequently, the cross-correlated standard uniform RFs $\mathbf{X}^{CU} = \left[\mathbf{X}_c^{CU}, \mathbf{X}_\varphi^{CU}\right]$ of c and φ can be generated for the considered copulas. For comparison purposes, all selected copulas are individually adopted to generate data with the best-fit marginal distributions and the same correlation coefficient. The procedures for generating data with different copula are concluded as follows:

1. ***Gaussian copula***: (a) find the upper triangular matrix $\mathbf{Q} = \begin{bmatrix} 1 & \theta \\ 0 & \left(1-\theta^2\right)^{0.5} \end{bmatrix}$ of $\mathbf{R} = \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}$ by Cholesky decomposition, in which $\mathbf{R}$ is a correlation matrix containing the Gaussian copula parameter θ; (b) based on the upper triangular matrix $\mathbf{Q}$, set $\mathbf{X}_c^{CN} = \mathbf{X}_c^{IG}$ and $\mathbf{X}_\varphi^{CN} = \mathbf{X}_\varphi^{IG}\theta + \mathbf{X}_\varphi^{IG}\left(1-\theta^2\right)^{0.5}$ the cross-correlated standard normal RFs of c and φ are obtained; (c) set $\mathbf{X}^{CU} = \Phi\left(\mathbf{X}^{CN}\right)$, in which $\Phi(\bullet)$ is the CDF of the standard normal distribution. The two cross-correlated standard uniform RFs $\mathbf{X}^{CU} = \left[\mathbf{X}_c^{CU}, \mathbf{X}_\varphi^{CU}\right]$ of c and φ are obtained by the Gaussian copula.
2. ***Plackett copula***: (a) set $\mathbf{X}^{IU} = \Phi\left(\mathbf{X}^{IN}\right)$, then, the independent standard uniform RFs of c and φ, $\mathbf{X}^{IU} = \left[\mathbf{X}_c^{IU}, \mathbf{X}_\varphi^{IU}\right]$ can be obtained; (b) set $a = \mathbf{X}_\varphi\left(1 - \mathbf{X}_\varphi^{IU}\right)$, $b = \theta + a(\theta-1)^2, c = 2a\left(\mathbf{X}_c^{IU}\theta^2 + 1 - \mathbf{X}_c^{IU}\right)$, and $d = \sqrt{\theta}\sqrt{\theta + 4a\mathbf{X}_c^{IU}\left(1-\mathbf{X}_c^{IU}\right)(1-\theta)^2}$; (c) set $\mathbf{X}_c^{CU} = \mathbf{X}_c^{IU}$ and $\mathbf{X}_\varphi^{CU} = \left[c_0\left(1 - 2\mathbf{H}_\varphi^{IU}\right)d_0\right]/2b_0$ to get $\mathbf{X}^{CU} = \left[\mathbf{X}_c^{CU}, \mathbf{X}_\varphi^{CU}\right]$.
3. ***Frank copula***: (a) set $\mathbf{X}^{IU} = \Phi\left(\mathbf{X}^{IN}\right)$, the independent standard uniform RFs of c and φ, $\mathbf{X}^{IU} = \left[\mathbf{X}_c^{IU}, \mathbf{X}_\varphi^{IU}\right]$ can be obtained; (b) set $\mathbf{X}_c^{CU} = \mathbf{X}_c^{IU}$; (c) Set $\mathbf{X}_\varphi^{IU} = C_2\left(\mathbf{X}_\varphi^{CU}\middle|\mathbf{X}_c^{CU}\right)$ where $C_2\left(\mathbf{X}_\varphi^{CU}\middle|\mathbf{X}_c^{CU}\right)$ is the conditional distribution of $\mathbf{X}_\varphi^{CU}$ given $\mathbf{X}_c^{CU}$, which is expressed as $C_2\left(\mathbf{X}_\varphi^{CU}\middle|\mathbf{X}_c^{CU}\right) = \frac{\phi^{-1(1)}\left(\left(\phi(\mathbf{X}_c^{CU};\theta)+\phi\left(\mathbf{X}_\varphi^{CU};\theta\right)\right);\theta\right)}{\phi^{-1(1)}\left(\phi(\mathbf{X}_c^{CU};\theta);\theta\right)}$, where $\phi(\bullet)$ is the generator function of an Archimedean copula. Then, $\mathbf{X}_\varphi^{CU}$ is



(a) Histogram scatter plot of simulated c and φ (b) Histogram scatter plot of simulated φ and γ



(c) Histogram scatter plot of simulated c and γ

Fig. 12. Histogram scatter plots of measured and simulateddata.

determined by solving the equation $\mathbf{X}_{\varphi}^{IU} = C_2\left(\mathbf{X}_{\varphi}^{CU}\middle|\mathbf{X}_{c}^{CU}\right)$ using the bisection method.

4. ***No.16 copula***: (a) set $\mathbf{X}^{IU} = \Phi\left(\mathbf{X}^{IN}\right)$. Then, the independent standard uniform RFs of c and φ, $\mathbf{X}^{IU} = \left[\mathbf{X}_{c}^{IU}, \mathbf{X}_{\varphi}^{IU}\right]$ can be obtained; (b) set $\mathbf{X}_{c}^{CU} = \mathbf{X}_{c}^{IU}$; (c) set $\mathbf{X}_{\varphi}^{IU} = C_2\left(\mathbf{X}_{\varphi}^{CU}\middle|\mathbf{X}_{c}^{CU}\right)$ where $C_2\left(\mathbf{X}_{\varphi}^{CU}\middle|\mathbf{X}_{c}^{CU}\right)$ is the conditional distribution of $\mathbf{X}_{\varphi}^{CU}$ given $\mathbf{X}_{c}^{CU}$. Then $\mathbf{X}_{\varphi}^{CU}$ is determined by solving the equation $\mathbf{X}_{\varphi}^{IU} = C_2\left(\mathbf{X}_{\varphi}^{CU}\middle|\mathbf{X}_{c}^{CU}\right)$.

After obtaining the cross-correlated standard uniform RFs $\mathbf{X}^{CU} = \left[\mathbf{X}_{c}^{CU}, \mathbf{X}_{\varphi}^{CU}\right]$ of c and φ from the copulas, the corresponding data in physical space is determined by applying isoprobabilistic transformation. In this way, the realisations of the copula-based cross-correlated RFs of c and φ in physical space can be generated. To examine the fitness of the copulas to dependence structures underlying the measured data, 500 data samples in uniform space are generated. From visual inspection, it is very difficult to distinguish the best-fit copula among them due to the limited number of measured data. Generally, the No.16 copula, is identified as the best-fit copula based on the AIC and BIC criteria, which can also match the measured data adequately. Through the isoprobabilistic transformation, the distribution of simulated data (c and φ) in physical space constructed using the No.16 copula agrees well with the measured data in comparison with the other three copulas as shown in Fig. 5.

Stochastic modelling of the runout distances

Fig. 6 shows one typical realization of the RFs of c and φ values and associated generated data based on the No.16 copula. The Pearson correlation coefficient is −0.4785 and the Kendall's rank correlation coefficient is −0.4357. It shows that the generated RFs of c and φ values have intermediately negative correlation. As the figure shows, the generated friction angle and cohesion values fit very well with their marginal distributions (Weibull and Normal distribution), receptively (as shown in Fig. 6c). Subsequently, the generated data is mapped onto the material points according to their spatial coordinates in the computational domain and the large-deformation simulation of the landslide post-failure process are performed to determine the final runout distance.

Fig. 7 shows the final configuration of a typical heterogeneous landslide with the best-fit copula-based cross-correlated RFs of c and φ values by MPM modelling, including the cohesion, friction angle, plastic strain invariant, and runout distance. It can be found that the landslide experienced a long runout and deposition process. Comparing with the initial state (see Fig. 6a&b), the spatial distribution of cohesion and friction angle in the slope have been notably changed since the motion of sliding mass. In Fig. 7d, green colour represents soils with rela-

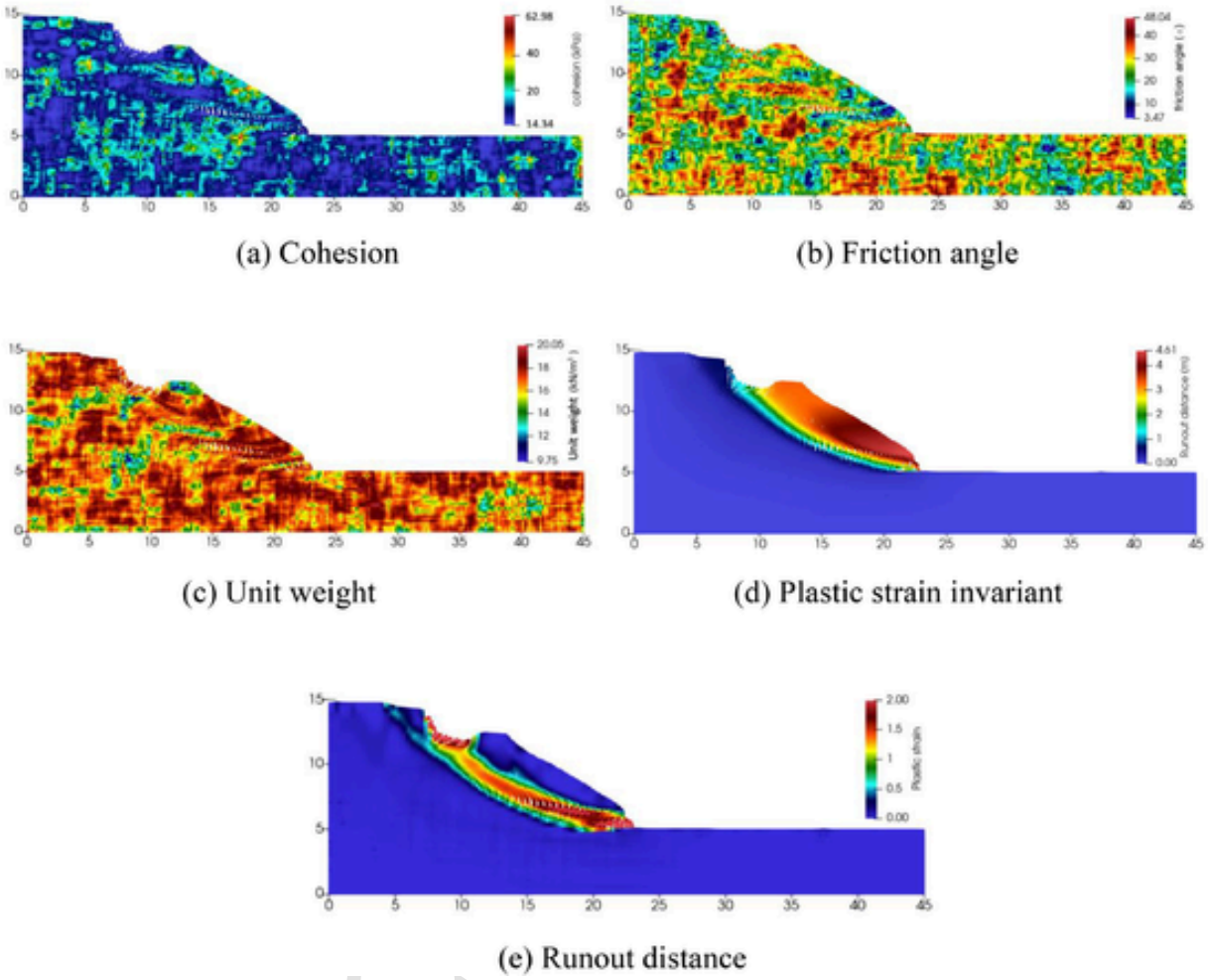


Fig. 13. A typical sample of copula-based multivariate landslide model at the final stage.

tively small displacement, and red colour represents soils with relatively large displacement. The slope may be divided into two zones, one with red for relatively large soil displacement and the other with dark blue for relatively small soil displacement. Fig. 7d illustrates the maximum soil displacement of the landslide at the final stage is 5.03 m, which is the steady state and final stage of the simulation. The soils with relatively large displacement slide downward along the slip surface, in which a circular slip surface is identified as the boundary between these two different colour zones (as shown in Fig. 7c). The plastic strain invariant contours illustrate the circular (rotational) shear band is formed. It should be noted that the landslide that poses the risks are those where the strength properties of the mobilised soils are greatly degraded during the post-failure stage. Such strain-softening behaviour of soil mass could result in very long runout distance [14,28,46,29]. As the plastic strain increase, the strength decreases from peak strength in previous elastic stage to residual strength, which is controlled by a softening parameter. This softening behaviour will cause a weak layer near the slip surface where plastic strain concentrates. Therefore, the runout distance will be larger when considering the strain-softening behaviour.

In this example, the landslides with copula-based RF model considering the spatial variabilities and uncertainties of real shear strengths and their interdependencies would result in different responses in terms of the runout distance. Each heterogeneous landslide may differ due to the variability and interdependence of c and φ spatial data. Therefore, multiple MCS are conducted to reflect the possible post-failure behaviours of landslides and investigate the geo-system responses. It should be noticed that a particular distribution of RFs may lead to small deformed (i.e., soil displacement less than 0.1 m) or stable slopes, which does not present large-deformation post-failure behaviours. This is not the focus of the study. Therefore, for a practical disaster risk assessment reason, a realisation is considered as a sample when the maximum runout distance of the landslide is greater than 0.1 m.

In the following MCS, the simulations are iteratively performed until the first and second moments of the model outputs converge. The convergence is considered achieved when the discrepancies between mean and variance values calculated in consecutive Monte Carlo iterations are less than 10^{-3}. The convergence conditions for iterative mean and iterative variance are expressed as

$$\left|\bar{X}_n - \bar{X}_{n-1}\right| < 10^{-3} \tag{22}$$

$$\left|S_n^2 - S_{n-1}^2\right| < 10^{-3} \tag{23}$$

where n is the number of iterations, the $\bar{X}_n$ is the iterative mean in the n-th iteration calculated using $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, X_i represents the model output in the i th iteration, and the iterative variance is $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2$, and corresponding COV can be obtained from the ration of mean values and variances. Fig. 8a shows all stochastic data of computed runout distance for the MCS realizations, in which the variation of the runout distance and corresponding COV values are plotted as

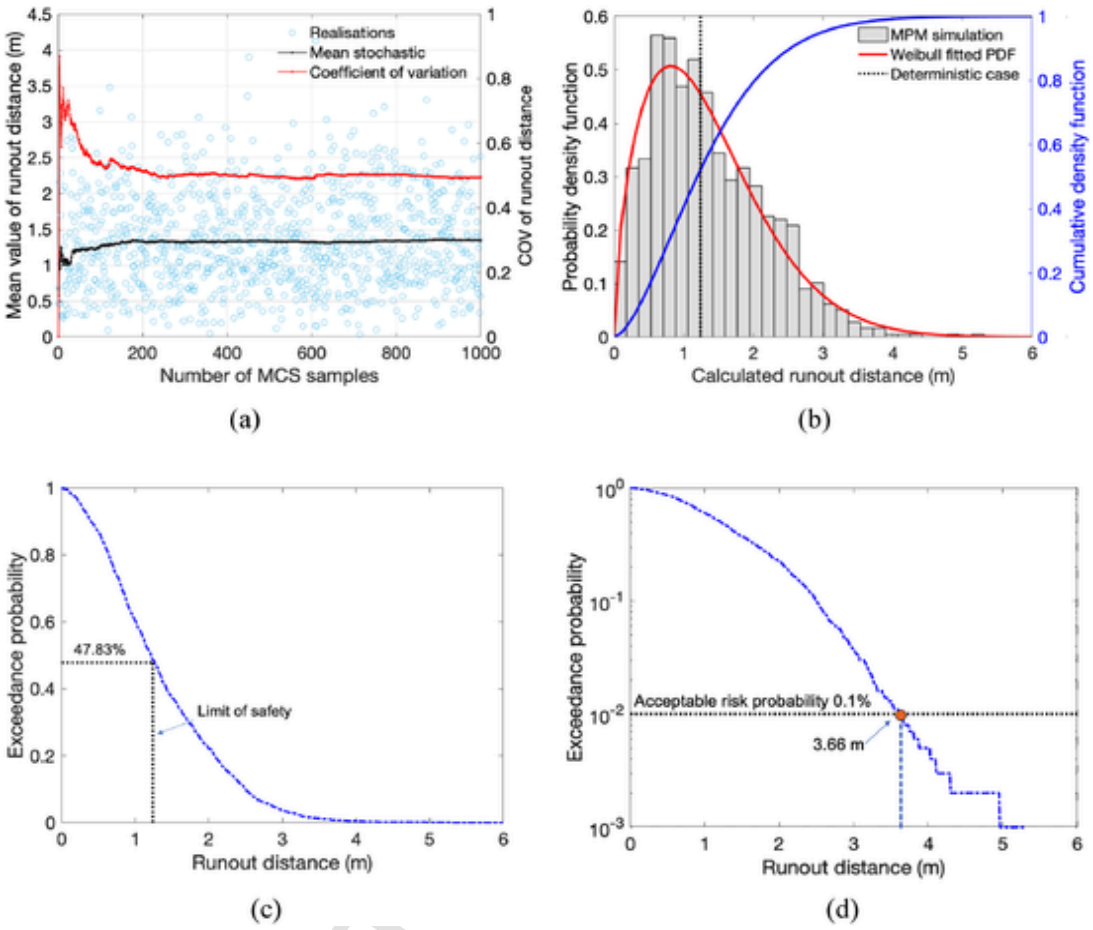


Fig. 14. Stochastic data, histogram, PDF, CDF, and exceedance probability of the runout distance.

functions of the number of MCS. It demonstrates that after about 700 simulations the statistical convergence is achieved. Therefore, 1000 realisations are performed to comfirm stable statistical results and give reliable estimates of runout distances. The computational cost for each realization, spanning from slope failure initiation to the final stage of the landslide, averaged approximately 10 to 15 min. These simulations were executed on a Mac computer powered by an Apple M1-Max chip, boasting a 10-core CPU and 64 GB of RAM. For reasonably evaluating distribution of computed runout distances from MCS, four individual candidate distributions are examined. It is found that both the AIC and BIC values indicate that the Weibull distribution is the best-fit distribution for the output runout distances. The corresponding shape parameter and scale parameter of the Weibull distribution are 2.3 and 6.0 respectively, which are determined by Maximum Likelihood Method. Fig. 8b shows the probability density histogram for the calculated runout distance of 1000 samples, where the horizontal coordinate represents the calculated runout distance, the left vertical coordinate represents the PDF, and the right coordinate represents the CDF. The red line is the PDF result fitted by the Weibull distribution based on the computed values, and the blue line is the corresponding CDF result. It is seen that the Weibull distribution can well-fit the histogram, the corresponding mean value μ and standard deviation σ can be obtained by the relevant expressions as shown in Table 1, based on which, the μ of runout distance is 5.32 m and the σ is 2.45 m. In Fig. 8, it is found that the values of calculated runout distance vary significantly, which is mostly in the range of about 1 m to 10 m, the minimum runout distance is 0.12 m, and the maximum runout distance can reach up to 14.46 m. For comparison purposes, a deterministic analysis of homogeneous landslide is conducted using the mean values of the c and φ values listed in Table 2. The results of the homogeneous case are presented as a reference for those obtained from the MCS. It illustrates that the deterministic analysis particularly underestimates the landslide risks by the runout distance with 5.52 m, which may give an unconservative estimation of runout distance. Based on the CDF curve (Fig. 8b) and exceedance probability curve (Fig. 8c), the deterministic result was considered as a limit of safety, which shows about 43.80% of all samples exceed the deterministic case (i.e., estimated runout distance > 5.52 m). This is a dramatic discrepancy in post-failure motions between the homogeneous and heterogenous landslides in terms of the runout distance. Consequently, the spatial variability of soil significantly increases the variability of output runout distance. The high degree of uncertainty in prediction of the runout distance can be observed, which necessitates incorporating the effect of natural heterogeneity of soil in modelling of post-failure process. Fig. 8d shows the exceedance probability of runout distance with logarithmic scale, if the acceptable risk probability is 0.1 % [29,58], the corresponding runout distance of landslide is 11.16 m. Therefore, any usage of lands within 11.16 m from the toe of the slope should be considered with caution given the soil spatial variability.

6.2. Example II: trivariate analysis of runout distances for landslides

Another example is presented to illustrate the use of a measured dataset of natural soils to construct trivariate copula functions and joint PDFs of random variables. Table 6 presents a dataset of measured c, φ, and γ values for soil samples from Salt creek basin in north-eastern Ne-

braska, United States (Soenksen et al. 2003), is used to derive their statistical characteristics (as shown in Table 5). Similar to the previous example, the best-fit marginal distribution for each measured data is determined based on the AIC and BIC scores of the measured data. Table 7 summarises the best-fit marginal distribution types, corresponding parameters, means, and standard deviations of the measured values.

Fig. 9 shows the corresponding identified best-fit marginal cumulative distribution functions of the measured data, together with the plots of empirical cumulative probabilities of the measured data for comparison. It is found that the identified best-fit marginal distributions of c, φ, and γ match very well with their empirical cumulative probability data points (Table 8).

The correlation matrix $\theta_{c,\varphi,\gamma}$ for the best-fit copula for these variables can be expressed in a symmetric 3×3 matrix as

$$\theta_{c,\varphi,\gamma} = \begin{bmatrix} 1 & -0.3321 & -0.0138 \\ -0.3321 & 1 & -0.2609 \\ -0.0138 & -0.2609 & 1 \end{bmatrix}$$

Based on the best-fit marginal distributions and the best-fit multivariate copula function for the measured data, the realisations of the copula-based multivariate RFs of c, φ, and γ in physical space can be generated by using CMD and isoprobabilistic transformation. To validate the fitness of the multivariate Gaussian copula on dependence structures underlying the measured data, 500 data samples are generated in physical space and plotted in Fig. 10. The figure provides comparisons between the measured data (43 sets) and the simulated data (500 samples) in two-dimensions and three-dimensions. It shows the interdependences (multiple cross-correlation) among the three variables in a realisation. The interdependent relationship among the measured soil properties has been quantitatively measured, visualised, and simulated. From a visual inspection, the dependence of simulated data agrees well with measured data.

Fig. 11 presents a typical spatial distribution of c, φ, γ within the slope model. The Pearson's correlation matrix of the simulated data is

$$\mathbf{R}_{c,\varphi,\gamma} = \begin{bmatrix} 1 & -0.3146 & -0.0049 \\ -0.3146 & 1 & -0.2613 \\ -0.0049 & -0.2613 & 1 \end{bmatrix}$$

Fig. 12 shows that the generated RFs of c and φ have negative correlation, the same can be seen between the generated RFs of c and γ, while the generated RFs of φ and γ show a very weakly negative correlation. The generated c, φ, γ fit very well with their marginal distributions (i.e., Gumbel and Weibull distributions). Subsequently, the generated values are mapped onto the material points according to their spatial coordinates in the computational domain and the simulation of the landslide post-failure process are performed to determine the final runout distances.

Fig. 13 shows the final configuration of a heterogeneous landslide computed by MPM modelling in terms of cohesion, friction angle, unit weight, plastic strain invariant, and the runout distance. A long runout and deposition process can be observed at the final state of the landslide. It can be observed that the distributions of cohesion, friction angle, and unit weight in the slope after the failure have changed since the motion of the sliding mass (Fig. 13). A quasi-circular rupture surface can be identified in Fig. 13d, where the contours of plastic strain invariant indicate the formation of a circular shear band. From Fig. 13e, the maximum soil displacement at the final stage is observed to be 4.61 m.

Similarly, MCS analysis is conducted for quantifying the uncertainties of the calculated runout distances from the heterogeneous landslide. The same convergence criterion in the first example is adopted to check the output (Fig. 14a). Based on the computed results, the Weibull distribution has been checked and used as the best-fit distribution of the output runout distance for the multivariate model. The corresponding shape parameter and scale parameter of the Weibull distribution are 1.6 and 1.5, respectively, which are determined by Maximum Likelihood Method. Fig. 14b shows the probability density histogram for the calculated runout distances from the 1000 MCS samples. It can be found that the Weibull distribution fits well with the histogram, a corresponding mean value (μ) of 1.33 m and a standard deviation (σ) value of 0.48 m are also obtained. It is observed that the runout distance mostly varies from 0.1 m to 5.2 m, and the minimum and maximum runout distances are 0.14 m and 5.32 m, respectively. For comparison, a deterministic analysis of the same example, but considering homogeneous conditions, is conducted using the mean values of c, φ, and γ (as listed in Table 7). The results of the deterministic analysis are presented as a reference for those cases obtained from the MCS. According to the results, the deterministic analysis underestimates the runout distance with 1.24 m, which reflect an unconservative estimate of the runout distances of the landslide. Based on the CDF curve and exceedance probability curve (Fig. 14c), if the deterministic result was considered as a limit of safety, which shows about 47.83 % of all samples exceed the deterministic case. Fig. 14d shows the exceedance probability of the runout distance in a semi-logarithmic plot, if the acceptable risk probability is 0.1 % [29,58], the corresponding runout distance of landslide is 3.66 m. Therefore, any buildings and infrastructure located within 3.66 m distance to the toe of the slope could be affected from a landslide.

7. Conclusions

This study proposed a novel approach based on copula methodology for generating multivariate cross-correlated geotechnical RFs and incorporating with the GIMP large-deformation analysis within a MCS framework. To this end, the main object is to investigate the probable runout distances after landslide failure. The copula-based cross-correlated RFs are simulated to capture the multiple geotechnical uncertainties and provide more reliable estimation of runout distances of landslides. The step-by-step procedure for implementing the new approach is provided, and two slope cases are presented and illustrated with different properties based on measured data. The following conclusions can be drawn:

1. Under incomplete probability information, the methodology can be considered as a general framework for stochastic post-failure analysis of landslides, which could develop a best-fit probability multivariate model for soil properties and generate statistically simulated data that could be used for large deformation analysis of landslides. This would offer notable improvements to risk assessment and mitigation of landslides.
2. The non-Gaussian distribution characteristics of soil parameters and their best-fit dependence structure (e.g., Gaussian, and non-Gaussian copulas) can be fully taken into consideration at material point level during the post-failure analysis of landslides, which is important to improve the accuracy of runout distance estimation.
3. Through the two examples, the results are compared against those from the deterministic soil slope. Both cases show about over 40 % of all MCS samples exceed the deterministic case, which indicates that the current deterministic analysis notably underestimates the risk induced by large runout distances of landslides. While considering multivariate RF may increases the complexity of modelling, its influence on the post-failure analysis of landslides should not be ignored.

Although the framework is adopted for simulation of multivariate cross-correlated RFs of a specific slope based on incomplete data information in the field, it is a general computational framework and can be extended to consider various large-deformation geotechnical problems

(e.g., tailing dam, rockburst, liquefaction-induced lateral spreading, seepage leading to suffusion, etc.). Other model geometries also can be replaced in this framework if the main consideration impacts/hits on nearby structures by extensive post-failure motions.

Uncited references

[15,26,30,34,40,50,56].

CRediT authorship contribution statement

Guotao Ma: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Mohammad Rezania:** Writing – review & editing, Visualization, Validation, Funding acquisition, Formal analysis. **Mohaddeseh Mousavi Nezhad:** Writing – review & editing, Formal analysis, Data curation. **Kok-Kwang Phoon:** Validation, Methodology.

Declaration of competing interest

The author declares no competing interests.

Data availability

Data will be made available on request.

Acknowledgements

The financial support by the European Commission 's RFCS project MINRESCUE (Contract RFCS-RPJ-899518) and the National Natural Science Foundation of China Project (Contract: 52150610492, 52350610265) are gratefully acknowledged. The first author would like to acknowledge the support from the Institute of Advanced Study, University of Warwick.

References

[1] Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19(6):716–23.
[2] Bardenhagen S.G. Energy conservation error in the material point method for solid mechanics. J Comput Phys 2002;180(1):383–403.
[3] Bardenhagen S.G, Kober E.M. The generalized interpolation material point method. Comput Model Eng Sci 2004;5(6):477–96.
[4] Bishop A, Hutchinson J, Penman A, Evans H. Geotechnical investigation into the causes and circum- stances of the disaster of 21 October 1966. A selection of technical reports submitted to the aberfan tribunal hmso. 1969. p. 1–80.
[5] Cami B, Javankhoshdel S, Phoon K.K, Ching J. Scale of fluctuation for spatially varying soils: estimation methods and values. ASCE-ASME J Risk Uncert Eng Syst, Part A: Civil Eng 2020;6(4):03120002.
[6] Chen L, Zhang W, Chen F, Gu D, Wang L, Wang Z. Probabilistic assessment of slope failure considering anisotropic spatial variability of soil properties. Geosci Front 2022;13(3):101371.
[7] Cherubini U, Luciano E, Vecchiato W. Copula methods in finance. John Wiley & Sons; 2004.
[8] Cheuk C, Ho K, Lam A. Influence of soil nail orientations on stabilizing mechanisms of loose fill slopes. Canad Geotech J 2013;50(12):1236–49.
[9] Ding J, Zhou J, Cai W. An efficient variable selection-based Kriging model method for the reliability analysis of slopes with spatially variable soils. Reliab Eng Syst Saf 2023;235:109234.
[10] Gironacci E, Nezhad M, Rezania M, Lancioni G. A non-local probabilistic method for modeling of crack propagation. Int J Mech Sci 2018;144:897–908.
[11] Griffiths D, Huang J, Fenton G. Influence of spatial variability on slope reliability using 2-D random fields. J Geotech Geoenviron Eng 2009;135(10):1367–78.
[12] Guan Z, Wang Y. Data-driven simulation of two-dimensional cross-correlated random fields from limited measurements using joint sparse representation. Reliab Eng Syst Saf 2023;238:109408.
[13] Huang S.Y, Liu L.L. New Kriging methods for efficient system slope reliability analysis considering soil spatial variability. Reliab Eng Syst Saf. 2024. 109989.
[14] Huang Y, Li G, Xiong M. Stochastic assessment of slope failure run-out triggered by earthquake ground motion. Nat Hazards 2020;101:87–102.
[15] Hungr O, Leroueil S, Picarelli L. The Varnes classification of landslide types, an update. Landslides 2014;11(2):167–94.
[16] Jiang S.H, Huang J, Griffiths D.V, Deng Z.P. Advances in reliability and risk analyses of slopes in spatially variable soils: a state-of-the-art review. Comput Geotech 2022;141:104498.
[17] Jiang S.H, Liu X, Huang J, Zhou C.B. Efficient reliability-based design of slope angles in spatially variable soils with field data. Int J Numer Anal Methods Geomech 2022;46(13):2461–90.
[18] Kafaji I. PhD thesis. University of Stuttgart; 2013.
[19] Li D, Jiang S, Cao Z, Zhou W, Zhou C, Zhang L. A multiple response-surface method for slope reliability analysis considering spatial variability of soil properties. Eng Geol 2015;187:60–72.
[20] Li X, Tang X, Zhao S, Yan Q, Wu Y. MPM evaluation of the dynamic runout process of the giant Daguangbao landslide. Landslides 2021;18(4):1509–18.
[21] Liu C, Sun Q, Jin F, Zhou G. A fully coupled hydro-mechanical material point method for saturated dense granular materials. Powder Technol 2017;314:110–20.
[22] Liu L, Cheng Y, Jiang S, Zhang S, Wang X, Wu Z. Effects of spatial autocorrelation structure of permeability on seepage through an embankment on a soil foundation. Comput Geotech 2017;87:62–75.
[23] Liu L, Liang C, Huang L, Wang B. Parametric analysis for the large deformation characteristics of unstable slopes with linearly increasing soil strength by the random material point method. Comput Geotech 2023;162:105661.
[24] Liu Q, Tang A, Huang D, Huang Z, Zhang B, Xu X. Total probabilistic measure for the potential risk of regional roads exposed to landslides. Reliab Eng Syst Saf 2022; 228:108822.
[25] Liu X, Wang Y, Li D.Q. Investigation of slope failure mode evolution during large deformation in spatially variable soils by random limit equilibrium and material point methods. Comput Geotech 2019;111:301–12.
[26] Ma G, Hu X, Yin Y, Luo G, Pan Y. Failure mechanisms and development of catastrophic rockslides triggered by precipitation and open-pit mining in Emei, Sichuan, China. Landslides 2018;15(7):1401–14.
[27] Ma G, Rezania M, Nezhad M.M. Effects of spatial autocorrelation structure for friction angle on the runout distance in heterogeneous sand collapse. Transp Geotech 2021;33:100705.
[28] Ma G, Rezania M, Nezhad M.M. Probabilistic post-failure analysis of landslides using stochastic material point method with non-stationary random fields. In: 20th International Conference on Soil Mechanics and Geotechnical Engineering (ICSMGE 2022). Sydney; 2022.
[29] Ma G, Rezania M, Nezhad M.M. Stochastic Assessment of Landslide Influence Zone by Material Point Method and Generalized Geotechnical Random Field Theory. Int J Geomech 2022;22(4):04022002.
[30] Ma G, Rezania M, Nezhad M.M, Shi B. Post-failure analysis of landslides in spatially varying soil deposits using stochastic material point method. Rock Soil Mech 2022;43(7):7.
[31] Ma G, Rezania M, Nezhad M.M, Hu X. Uncertainty quantification of landslide runout motion considering soil interdependent anisotropy and fabric orientation. Landslides 2021;1–17.
[32] Masoudian M, Afrapoli M, Tasalloti A, Marshall A. A general framework for coupled hydro-mechanical modelling of rainfall-induced instability in unsaturated slopes with multivariate random fields. Comput Geotech 2019;115:103162.
[33] Mori H, Chen X, Leung Y.F, Shimokawa D, Lo M.K. Landslide hazard assessment by smoothed particle hydrodynamics with spatially variable soil properties and statistical rainfall distribution. Canad Geotech J 2020;57(12):1953–69.
[34] Mroginski J, Castro H, Podestá J, Beneyto P, Anonis A.R. A fully coupled particle method for dynamic analysis of saturated soil. Comput Part Mech 2021;8(4): 845–57.
[35] Nezhad M, Gironacci E, Rezania M, Khalili N. Stochastic modelling of crack propagation in materials with random properties using isometric mapping for dimensionality reduction of nonlinear data sets. Int J Numer Methods Eng 2018; 113(4):656–80.
[36] Nezhad M, Rezania M, Baioni E. Transport in porous media with nonlinear flow condition. Transp Porous Media 2019;126:5–22.
[37] Nairn J.A. Material point method calculations with explicit cracks. Comput Model Eng Sci 2003;4(6):649–64.
[38] Nelsen R.B. An introduction to copulas. Springer Science & Business Media; 2007.
[39] Ng C.W, Qu C, Cheung R.W, Guo H, Ni J, Chen Y, Zhang S. Risk assessment of soil slope failure considering copula-based rotated anisotropy random fields. Comput Geotech 2021;136:104252.
[40] Nguyen T.S, Likitlersuang S, Tanapalungkorn W, Phan T.N, Keawsawasvong S. Influence of copula approaches on reliability analysis of slope stability using random adaptive finite element limit analysis. Int J Numer Anal Methods Geomech 2022;46:2211–32.
[41] Nguyen T.S, Phan T.N, Likitlersuang S, Bergado D.T. Characterization of stationary and nonstationary random fields with different copulas on undrained shear strength of soils: probabilistic analysis of embankment stability on soft ground. Int J Geomech 2022;22(7):04022109.
[42] Phoon K, Kulhawy F. Characterization of geotechnical variability. Canad Geotech J 1999;36(4):612–24.
[43] Remmerswaal G, Vardon P.J, Hicks M.A. Evaluating residual dyke resistance using the Random Material Point Method. Comput Geotech 2021;133:104034.
[44] Rodriguez J.C. Measuring financial contagion: a copula approach. J Empir Finance 2007;14(3):401–23.
[45] Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6(2):461–4.
[46] Shi B, Zhang Y, Zhang W. Analysis of the entire failure process of the rotational slide using the material point method. Int J Geomech 2018;18(8):04018092.
[47] Sklar A. Random variables, joint distribution functions, and copulas. Kybernetika 1973;9(6):449–60.
[48] Sklar M. Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris 1959;8:229–31.
[49] Soenksen P.J. Stream bank stability in eastern nebraska. US Department of the

Interior, US Geological Survey; 2003. Vol. 3.
[50] Soga K, Alonso E, Yerro A, Kumar K, Bandara S. Trends in large-deformation analysis of landslide mass movements with particular emphasis on the material point method. Geotechnique 2016;66(3):248–73.
[51] Tang X, Li D, Zhou C, Phoon K. Copula-based approaches for evaluating slope reliability under incomplete probability information. Struct Safety 2015;52:90–9.
[52] Tang X, Wang M, Li D. Modeling multivariate cross-correlated geotechnical random fields using vine copulas for slope reliability analysis. Comput Geotech 2020;127:103784.
[53] Trivedi P.K, Zimmer D.M. Copula modeling: an introduction for practitioners. Now Publishers Inc; 2007.
[54] Vanmarcke E. Random fields: analysis and synthesis. World Scientific; 2010.
[55] Wang B, Hicks M.A, Vardon P.J. Slope failure analysis using the random material point method. Géotech Lett 2016;6(2):113–8.
[56] Wang F, Li H. A non-parametric copula approach to dependence modelling of shear strength parameters and its implications for geotechnical reliability under incomplete probability information. Comput Geotech 2019;116:103185.
[57] Wang M, Tang X, Li D, Qi X. Subset simulation for efficient slope reliability analysis involving copula-based cross-correlated random fields. Comput Geotech 2020;118:103326.
[58] Wang Y, Qin Z, Liu X, Li L. Probabilistic analysis of post-failure behaviour of soil slopes using random smoothed particle hydrodynamics. Eng Geol 2019;261:105266.
[59] Wu X. 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.
[60] Wu X. Modelling dependence structures of soil shear strength data with bivariate copulas and applications to geotechnical reliability analysis. Soils Found 2015;55(5):1243–58.
[61] Xi C, Hu X, Ma G, Rezania M, Liu B, He K. Predictive model of regional coseismic landslides' permanent displacement considering uncertainty. Landslides 2022;19(10):2513–34.
[62] Yang Y, Wang P, Brandenberg S.J. An algorithm for generating spatially correlated random fields using Cholesky decomposition and ordinary kriging. Comput Geotech 2022;147:104783.
[63] Zheng Y, Zhang Y. Reliability analysis for system with dependent components based on survival signature and copula theory. Reliab Eng Syst Saf. 2023. 109402.
[64] Zhu H, Zhang L, Xiao T, Li X. Generation of multivariate cross-correlated geotechnical random fields. Comput Geotech 2017;86:95–107.