Journal of Statistical Research *https://doi.org/10.3329/jsr.v60i1.92991*
2026, Vol. 60, No. 1, pp. 213-230 *ISSN 0256 - 422 X*

MODELLING DEPENDENCY OF MULTIVARIATE SOIL PARAMETERS USING VARIOUS VINE COPULA MODELS

EDGAR JR. JOE
Department of Civil Engineering, Faculty of Engineering
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
Email: edgar@upm.edu.my

SUMMARY

Copula-based approaches have been increasingly employed to model the dependency among soil parameters. Nevertheless, existing applications are predominantly confined to conventional low-order bivariate copulas, which limits their capability to capture complex multivariate dependency structures. To overcome this limitation, this study introduces the use of vine copula models for advanced dependency modelling in geotechnical engineering. To demonstrate the applicability of various vine copula structures, a slope reliability analysis framework was developed by integrating Monte Carlo simulation with a Multilayer Perceptron regression model. The influence of different vine copula configurations on the estimated probability of failure was systematically examined, and their performance was rigorously evaluated. The results reveal that the choice of vine copula structure has a significant impact on the predicted failure probability. Moreover, the proposed vine copula models exhibit closer agreement with the observed statistical characteristics of soil parameters compared to the traditional multivariate normal distribution approach. Overall, this study provides practical insights into the selection and construction of appropriate vine copula models for high-dimensional dependency modelling, particularly in geotechnical applications. The proposed framework enables more reliable estimation of slope failure probabilities.

Keywords and phrases: Dependency; Geotechnical Reliability, Soil, Uncertainties, Vine copula

1 Introduction

Soil parameters are inherently uncertain due to measurement limitations, spatial variability, sampling disturbance, and laboratory testing errors (Davidovic et al., 2010). In reliability-based geotechnical analysis, it is essential to rigorously account for these uncertainties to avoid biased or misleading predictions of slope performance. Previous studies have demonstrated that neglecting the dependence among soil parameters can substantially distort reliability estimates, in some cases leading to failure probabilities that differ by several orders of magnitude (Wang and Aladejare, 2016, Javankhoshdel and Bathurst, 2017). Accordingly, accurate reliability assessment requires not only proper characterisation of the variability of individual soil parameters, but also robust modelling of their underlying dependency structure. In recent years, copula-based approaches have gained attention in geotechnical engineering for modelling dependence among soil parameters (Ching and Phoon, 2015, Das et al., 2018, Li et al., 2015, Lü et al., 2020). Copulas link marginal probability distributions into a joint distribution while preserving each marginal's characteristics, allowing flexible representation of dependence even when variables follow different distributions. This is particularly useful in geotechnical applications, where datasets are often limited and variables are frequently non-normal. However, most studies focus on bivariate dependence, whereas practical slope stability problems involve multiple interacting soil parameters with complex, often non-Gaussian dependencies. Conventional multivariate normal models, which assume symmetric dependence, may fail to capture such behaviour and cannot fully represent the combined influence of multiple variables.

To address these limitations, vine copula models provide a flexible framework for constructing multivariate joint distributions from interconnected bivariate copulas. Introduced by Joe (1996) and further developed by Bedford and Cooke (2001), vine copulas systematically capture complex dependency structures in an interpretable manner. Unlike conventional approaches with a single global dependence, vine copulas allow each variable pair to follow different copula families, enabling the representation of diverse features such as asymmetry and tail dependence commonly observed in geotechnical data.

Vine copula models have been widely adopted across various disciplines, including economics and finance (Tófoli et al., 2019), hydrology (Jiang et al., 2019), medicine (Nikoloulopoulos, 2019), meteorology (Maity et al., 2020), renewable energy (Chen et al., 2019), and structural engineering (Cheng et al., 2020). Despite their demonstrated versatility, their application in geotechnical engineering remains relatively limited. To date, only a small number of studies, such as Lü et al. (2020) and Tang et al. (2020), have investigated the use of vine copulas for modelling multivariate dependence among soil parameters within a reliability analysis framework.

Therefore, this study aims to introduce and systematically compares different vine copula structures in a geotechnical engineering context. Three common classes are considered: regular vine (R-vine), canonical vine (C-vine), and drawable vine (D-vine), with the latter two representing structured cases of the R-vine. The impact of vine copula selection on predicted slope failure probability is assessed through a reliability-based case study, providing a practical framework for modelling multivariate soil parameter dependence and improving the reliability of geotechnical analyses.

2 Methodology

This section presents the construction of various vine copula models based on conditional pairwise bivariate copulas. The implementation was primarily carried out using R, an open-source programming language that provides comprehensive packages for vine copula modelling (R Core Team, 2019).

2.1 Identification of Best-fit Marginal Distribution of Variable

The best-fitting marginal distribution for each variable was identified using the Akaike Information Criterion (*AIC*) and Bayesian Information Criterion (*BIC*). The *AIC* evaluates the trade-off between model fit and complexity by penalizing the inclusion of excessive parameters, thereby reducing the risk of overfitting. In contrast, the *BIC* imposes a stronger penalty for model complexity, particularly in the presence of large sample sizes. For each variable, the distribution corresponding to the minimum AIC and BIC values—computed as defined in Equations (2.1) and (2.2)—was selected as the optimal model:

$$AIC_{marginal} = -2\sum_{i=1}^{N} \ln\left(f(x_i)\right) + 2k \tag{2.1}$$

$$BIC_{marginal} = -2\sum_{i=1}^{N} \ln\left[f(x_i)\right] + k\log(N), \tag{2.2}$$

where N is the number of observations in the sample, $f(x_i)$ is the probability density function of variable x_i, and k is the number of parameters in the distribution. The set of candidate distributions included the normal, lognormal, Gumbel, and Weibull distributions.

2.2 Construction of Bivariate Copula

Following the identification of the optimal marginal distribution for each variable based on the *AIC* and *BIC*, the next step involved modelling the dependence structure among the variables. This was achieved through the construction of bivariate copulas, which allow flexible representation of joint behaviour while preserving the properties of the selected marginal distributions. According to Equation (2.3), the variables were transformed from their original domain into a standard uniform space, yielding a random vector, u [0,1]:

$$u_i = \frac{\mathrm{rank}(x_i)}{N+1}, \tag{2.3}$$

where $\mathrm{rank}(x_i)$ denotes the rank of x_i in the sample arranged in ascending order, and N is the total number of randomly collected samples.

Gumbel, Gaussian, Student's t, Clayton, and Frank copulas were considered as candidate models to characterise the dependence between pairs of variables transformed into the standard uniform space. The optimal bivariate copula for each variable pair was selected based on the minimum values of the *AIC* and *BIC*, as defined in Equations (2.4) and (2.5):

$$AIC_{bivariate} = -2\sum_{i=1}^{N} \ln\left[c(u_{1i}, u_{2i}; \hat{\theta})\right] + 2k \tag{2.4}$$

$$BIC_{bivariate} = -2\sum_{i=1}^{N} \ln\left[c(u_{1i}, u_{2i}; \hat{\theta})\right] + k\log N, \tag{2.5}$$

where u is a standard uniform random sample, N is the sample size, k is the number of model parameters, θ is the estimated copula parameter, and L is the log-likelihood function. For the Gumbel, Gaussian, Frank, and Clayton copulas, the number of parameters is $k = 1$. In contrast, the Student's t copula involves two parameters ($k = 2$), comprising the copula parameter θ and the degrees of freedom ν. The bivariate copula was constructed using the copula function and the generator function given in Tables 1 and 2.

Table 1: Copula functions for common copula families (Aas and Berg, 2009, Li et al., 2015)

Family	Copula Function, $C(u_1, u_2; \theta)$
Gaussian	$N_\theta\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$
Student's t	$T_{(\theta,\lambda)}\left(T_\lambda^{-1}(u_1), T_\lambda^{-1}(u_2)\right)$
Clayton	$\left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$
Frank	$-\frac{1}{\theta}\ln\left[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right]$
Gumbel	$\exp\left\{-\left[(-\ln u_1)^\theta + (-\ln u_2)^\theta\right]^{1/\theta}\right\}$

2.3 Construction of Vine Copula

The vine copula was subsequently constructed by assembling bivariate copulas as the building blocks of the multivariate dependence structure. In the C-vine copula, a root node, or pivot element, was chosen by selecting the node with the strongest relationship to the remaining nodes. Based on the empirical Kendall's tau matrix, the root node had the highest column sum (R Core Team, 2019). The most suitable bivariate copula was used to establish the dependence between a specific variable and the first root node in the first C-vine copula tree (Aas and Berg, 2009). Pairwise dependencies with respect to a second variable were modeled using the root node as the conditioning variable, after which the second root node was determined. This procedure was repeated to expand each tree by selecting a root node, with all pairwise dependencies constructed relative to the root node and conditioned on all previous root nodes. Equation (2.6) was then used to generate the C-vine copula:

$$f(x) = \prod_{k=1}^{n} f(x_k) \prod_{i=1}^{n-1}\prod_{j=1}^{n-i} c_{i,i+j|1:(i-1)}(F(x_i \mid x_1,\ldots,x_{i-1}),\, F(x_{i+j} \mid x_1,\ldots,x_{i-1}))\,, \tag{2.6}$$

Table 2: Generator functions, association between θ and Kendall's τ and extent of θ for common copula families (Aas and Berg, 2009, Li et al., 2015)

Family	Generator Function $\varphi(t)$	Association between θ and Kendall's τ	Extent of θ
Gaussian	NIL	$\theta = \sin\left(\frac{\pi\tau}{2}\right)$	$[-1, 1]$
Student's t	NIL	$\theta = \sin\left(\frac{\pi\tau}{2}\right)$	$[-1, 1]$
Clayton	$\dfrac{t^{-\theta} - 1}{\theta}$	$\theta = \dfrac{2\tau}{1-\tau}$	$[-1, \infty) - \{0\}$
Frank	$-\ln\left(\dfrac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$	$\tau = 1 - \dfrac{4}{\theta} + \dfrac{4}{\theta^2}\int_0^{\theta} \dfrac{t}{e^t - 1}\, dt$	$(-\infty, \infty) - \{0\}$
Gumbel	$(-\ln t)^{\theta}$	$\theta = \dfrac{1}{1-\tau}$	$[1, \infty)$

Notes: Φ denotes the cumulative distribution function (CDF) of the standard normal distribution; N_θ denotes the CDF of the standard bivariate normal distribution with Pearson correlation θ; T_λ denotes the CDF of the univariate Student's t distribution with λ degrees of freedom; $T_{(\theta,\lambda)}$ denotes the CDF of the bivariate Student's t distribution with correlation parameter θ and λ degrees of freedom.

where $f(x_k)$ represents the marginal densities for variable x_k, $k = 1, 2, \ldots, n$ and $c_{i,i+j|1:(i-1)}$ is the bivariate copula for root node i and tree j. The external product runs over the $n-1$ trees and root nodes i, while the internal product denotes the $n-i$ pair copula in each tree $j = 1, \ldots, n-i$.

In the D-vine copula framework, variable pairs were selected and sequentially arranged to maximize the total sum of pairwise Kendall's tau values. This procedure effectively corresponds to identifying a Hamiltonian path in a graph, where the variables represent vertices and the Kendall's tau values serve as edge weights. Bivariate copulas were applied at the base tree to capture the relationships between consecutive variables, namely, the first and second, the second and third, and so forth. Conditional dependencies, such as between the first and third variables given the second variable, were subsequently modelled in higher-level trees to complete the co-dependency analysis. External products were computed across $d-1$ trees, whereas internal products determined the optimal pairings within each tree. The route structure of the D-vine copula tree is formally defined by Equation (2.7):

$$\begin{aligned} f(x) = \prod_{k=1}^{n} f(x_k) \prod_{i=1}^{n-1} \prod_{j=1}^{n-i} \Big\{ & c_{j,j+i|(j+1):(j+i-1)} \\ & \times \left[F(x_j \mid x_{j+1}, \ldots, x_{j+i-1}), F(x_{x_{j+i}} \mid x_{j+1}, \ldots, x_{j+i-1})\right] \Big\}, \end{aligned} \tag{2.7}$$

where $f(x_k)$ represents the marginal densities for variable x_k, $k = 1, 2, \ldots, n$ and $c_{j,j+i|(j+1):(j+i-1)}$ is the bivariate copula for node i and tree j. For $n-1$ trees, the exterior products are determined, and the pairings for each tree are regulated using the inner product.

On the other hand, the strongest dependency in the ground-level tree of an R-vine copula is similar to that of a D-vine copula, but instead of a path-like structure, it is necessary to create a maximum tree structure with specific edge weights (Geidosch and Fischer, 2016). The technique for storing the values of paired copulas for the joint probability density in an R-vine is described in Equation (2.8):

$$f(x) = \prod_{k=1}^{n} f(x_k) \times \prod_{j=n-1}^{1} \prod_{i=n}^{j+1} c_{j,i|i+1:n} \left[F(x_j \mid x_{i+1}, \ldots, x_n), F(x_i \mid x_{i+1}, \ldots, x_n) \right], \tag{2.8}$$

where $f(x_k)$ represents the marginal densities for variable x_k, $k = 1, 2, \ldots, n$ and $c_{j,i|i+1:n}$ is the bivariate copula for node i and tree j. For $n-1$ trees, the outside products are determined, and the pairings for each tree are determined using the inner result.

The function of the conditional distribution $F(x \mid \mathbf{v})$ for the vector $\mathbf{v}$ with n-dimension can be derived based on pair copula for tree $n-1$, utilising the pair-copulas of the previous trees $1, \ldots, n$. Equation (2.9) is used step-by-step:

$$h(x \mid \mathbf{v}, \theta) := F(x \mid \mathbf{v}) = \frac{\partial C_{xv_j|\mathbf{v}_{(-j)}} \left(F(x \mid \mathbf{v}_{(-j)}), F(v_j \mid \mathbf{v}_{(-j)}) \mid \theta \right)}{\partial F(v_j \mid \mathbf{v}_{(-j)})}, \tag{2.9}$$

Here, ν_j is a random element of ν, and $\nu_{(-j)}$ represents the $(n-1)$-dimensional vector ν other than ν_j. The $C_{xv_j|\mathbf{v}_{(-j)}}$ is the function of the pair copula with the parameter θ defined in tree n.

2.4 Simulation of Random Variables from Vine Copula

Random samples from the cumulative distribution function, derived from the conditional bivariate copulas, were generated using the Rosenblatt transformation (Rosenblatt, 1952). This approach enables the mapping of vectors of standard uniform variables back into their original, independent variable space. The transformation process, which involves sequential conditioning, is expressed formally in Equation (2.10):

$$\begin{aligned} x_1 &= F_1^{-1}(u_1) \\ x_2 &= F_{2|1}^{-1}(u_2 \mid x_1) \\ &\vdots \\ x_i &= F_{i|1,2,\ldots,(i-1)}^{-1}(u_i \mid x_1, x_2, \ldots, x_{i-1}), \end{aligned} \tag{2.10}$$

where x_i denotes the random samples, u represents the standard uniform random samples in $[0, 1]$, and F_i^{-1} symbolises the inverse of the marginal cumulative distribution function.

3 Results and Discussions

A reliability study of an unsaturated slope exposed to infiltration of rainwater based on a case study in Kota Kinabalu, Sabah, Malaysia, demonstrates an illustrative application of the vine copula models.

3.1 Soil Parameters

In this study, the hydro-mechanical soil properties of the unsaturated slope are treated as random variables. The scaling parameter α and shape parameter n of Van Genuchten–Mualem's Soil Water Characteristic Curve (SWCC) (Van Genuchten, 1980), residual water content (θ_r), saturated water content (θ_s), saturated hydraulic conductivity (k_s), effective friction angle (ϕ'), effective cohesion (c'), and unit weight (γ) are among the parameters.

3.2 Slope Model

A two-dimensional slope geometry (Figure 1) was modeled in Abaqus (Dassault Systèmes, 2017), which is a general-purpose finite element program.

The boundary conditions of the finite element model were determined as follows: (i) AG and DF constrained along X-axis direction; (ii) GF restricted in X- and Y-axes direction; (iii) GH and EF represented the groundwater table; (iv) AH, DE, and FG as no-flow boundary conditions; and (v) AB, BC, and CD as surface flux where the slope was exposed to total rainfall of 347mm for 240 hours. The plane strain quadrilateral with 4 nodes was the element type used. Mesh sensitivity analysis has been executed to attain a satisfactory precision of finite element results and a rational complexity. The slope model was meshed with a global mesh size of 1 m, with a localised coarser mesh of 1.5 m employed along the slope model foundation to save computational costs. In this study, there were 3,191 elements and 3,321 nodes modelled, respectively.

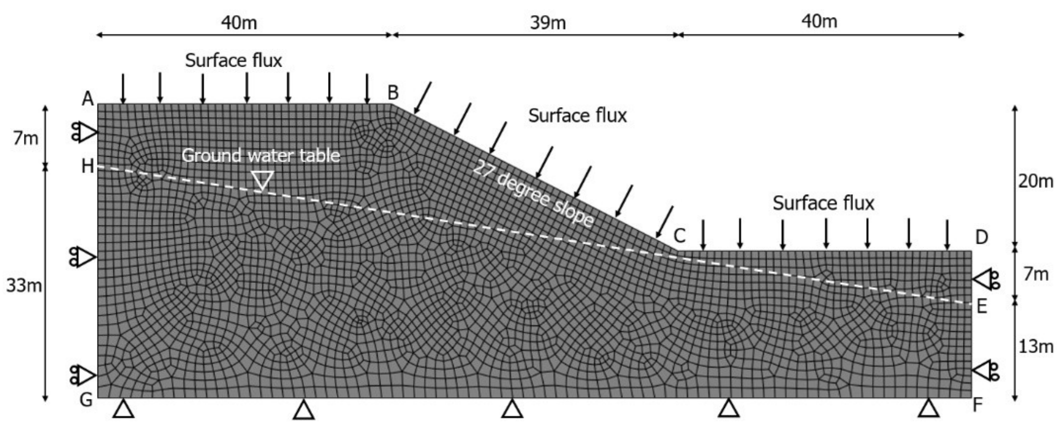

Figure 1: Modelled slope in finite element programme.

3.3 Modelling of Marginal Probability Distribution Function

The lognormal distribution is the best-fitting marginal distribution for scaling parameter α, saturated water content, shape parameter n, effective friction angle, and saturated hydraulic conductivity. The Gumbel distribution best describes the residual water content. The normal distribution best fits effective cohesion and unit weight. Table 3 summarises the outcomes using marginal distribution modelling.

Table 3: *AIC* and *BIC* values for marginal distribution fitting

	Normal		Lognormal		Gumbel		Weibull	
	AIC	*BIC*	*AIC*	*BIC*	*AIC*	*BIC*	*AIC*	*BIC*
k_s	-58.48	-55.89	-166.81^*	-164.22^*	-28.49	-25.90	-159.26	-156.67
α	-110.86	-108.27	-144.72^*	-142.13^*	-86.11	-83.52	-136.61	-134.02
n	-3.32	-0.73	-6.11^*	-3.52^*	8.04	10.63	2.44	5.04
θ_s	-105.86	-103.26	-106.09^*	-103.50^*	-101.73	-99.14	-103.52	-100.93
θ_r	-90.83	-89.56	-41.48	-40.20	-92.71^*	-91.43^*	-88.01	-86.73
γ	22.78^*	23.75^*	22.89	23.86	23.55	24.52	23.37	24.34
c'	57.60^*	58.57^*	69.06	70.03	58.32	59.29	58.17	59.14
ϕ'	74.72	75.69	73.86^*	74.83^*	77.52	78.49	75.98	76.95

Notes: For each soil parameter, an asterisk denotes the best-fit marginal distribution type.

3.4 Construction of Bivariate and Vine Copula Models

The relational dependencies among variables were modeled using bivariate and vine copulas. Initially, a conditional bivariate copula was developed to quantify the pairwise dependence among the eight nodes of the vine structure. In this framework, each pairwise bivariate copula functions as an edge connecting two variables, resulting in a hierarchical model composed of seven trees and a total of 28 edges. Each edge represents either a direct or conditional relationship between two variables and is constructed using bivariate copulas. The hierarchical arrangement allows for the modeling of both direct pairwise interactions and higher-order conditional dependencies, providing a comprehensive and accurate characterization of the joint multivariate distribution.

The conditional bivariate copula model forms the foundation for constructing C-vines, D-vines, and R-vines, as illustrated in Figures 2 to 4. These vine structures represent the propagation of dependence across successive trees and capture both sequential and hierarchical conditional relationships among variables. By providing a framework for identifying subtle interactions that are often overlooked by simple correlation measures, vine structures enhance the robustness, interpretability, and predictive performance of multivariate dependence modeling.

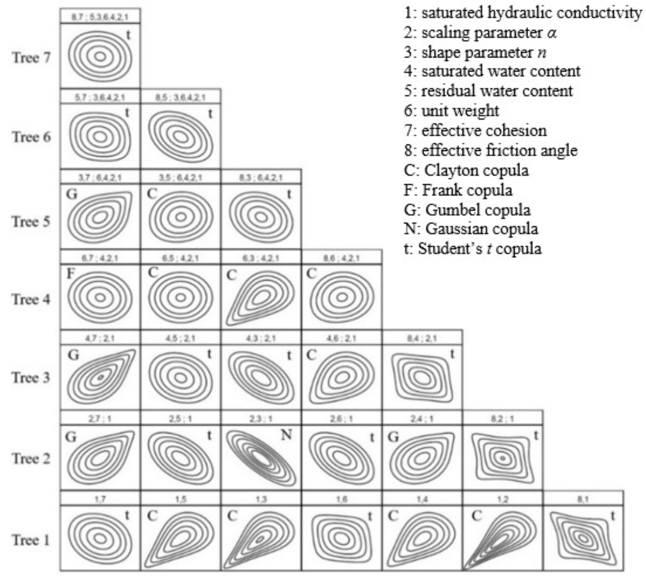

Figure 2: C-vine copula model

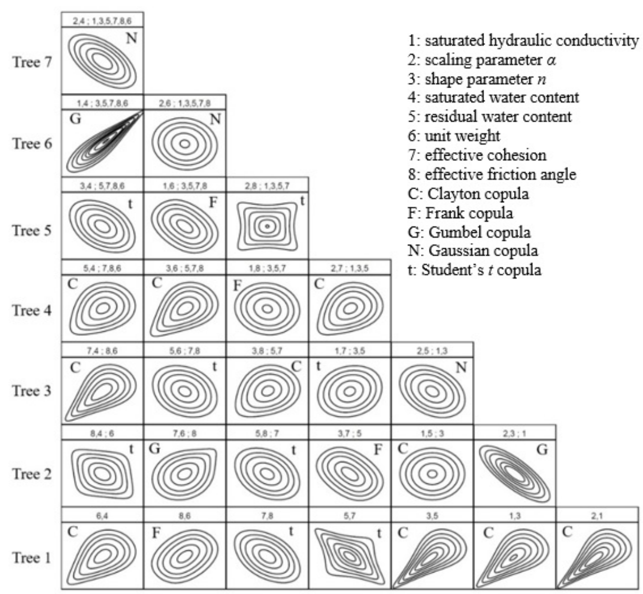

Figure 3: D-vine copula model

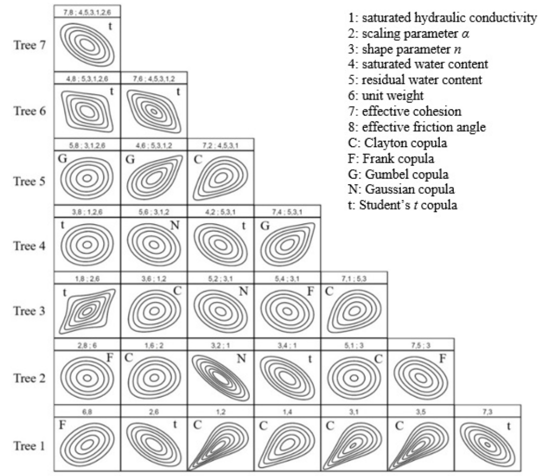

Figure 4: R-vine copula model

To capture the majority of bivariate dependencies, the Student's t and Clayton copulas were predominantly employed. In the C-vine copula, the Student's t copula appears 13 times, the Clayton copula 9 times, the Gumbel copula 4 times, and the Frank and Gaussian copulas once each. For the D-vine copula, the Clayton copula is used 10 times, the Student's t copula 8 times, the Frank copula 4 times, and the Gumbel and Gaussian copulas 3 times each. In the R-vine copula, both the Student's t and Clayton copulas are selected 9 times, the Frank copula 4 times, and the Gumbel and Gaussian copulas 3 times each. Beyond statistical goodness-of-fit, the structural distinctions among C-vine, D-vine, and R-vine models offer insights into the underlying physical organization of soil parameter dependencies.

In the C-vine structure, one central variable acts as a dominant conditioning node connected to multiple other parameters in the first tree level. When effective cohesion (c') or saturated hydraulic conductivity (k_s) appears as the central node, this suggests that these parameters exert broad influence over the dependency network. From a geotechnical standpoint, this is physically meaningful. For example, the saturated hydraulic conductivity governs rainfall infiltration rate and suction redistribution, which in turn influence shear strength mobilisation. Similarly, effective cohesion reflects bonding and structural integrity, which are closely linked to both friction angle (ϕ') and volumetric behaviour. Thus, a C-vine configuration may represent a system where one parameter—often related to pore structure or soil fabric—controls multiple hydro-mechanical responses.

In contrast, the D-vine structure represents a sequential dependency pattern, where each variable is conditionally dependent on its neighbouring variables. For example, in the first tree level, the chain begins with saturated water content, which is closely related to unit weight, reflecting the influence of moisture conditions on the bulk characteristics of the soil. Unit weight is then connected to the shear strength parameters, namely effective friction angle (ϕ') and effective cohesion (c'), highlighting the interaction between soil density and mechanical resistance. The sequence continues through residual water content and the Soil Water Characteristic Curve (SWCC) shape parameter n, which describe the soil–water retention behaviour. These hydraulic characteristics are subsequently linked to saturated hydraulic conductivity (k_s) and the scaling parameter α, both of which govern water flow and suction response in unsaturated soils. Overall, the ordering reflects a hydro–mechanical interaction pathway in which moisture-related properties, strength parameters, and flow characteristics are statistically connected in a sequential manner, consistent with the coupled behaviour observed in rainfall infiltration and slope stability problems.

Meanwhile, the R-vine structure, being the most flexible, allows for complex and non-hierarchical dependency patterns. This flexibility is particularly suitable for natural soils where multiple mechanisms operate simultaneously. In residual or tropical soils, pore geometry, weathering degree, bonding, and fissuring may influence hydraulic and strength parameters concurrently rather than in a simple hierarchical order. The R-vine model accommodates such multi-directional interactions, explaining its superior performance in reproducing the empirical joint behaviour of parameters.

3.5 Estimation of Failure Probability for Various Models of Vine Copula

Monte Carlo simulation (MCS) was employed to estimate the probability of failure using 100,000 simulated samples. This sample size was determined through a convergence analysis, which showed that both the coefficient of variation and the estimated probability of failure stabilized at 100,000 realizations. To enhance computational efficiency and avoid performing 100,000 direct finite element (FE) analyses, a surrogate modeling approach was adopted. Specifically, an implicit performance function was approximated using a Multilayer Perceptron (MLP) regression model with a single hidden layer comprising five nodes. The MLP was trained on results from 120 shear strength reduction analyses conducted in the FE program. Once trained, the surrogate model was used to evaluate the factor of safety for all Monte Carlo samples generated via the Rosenblatt transformation.

The US Army Corps of Engineers' target reliability indices (U.S. Army Corps of Engineers, 1997) can be used to assess predicted performance levels for varying failure probability. The most appropriate vine copula model in this study is the D-vine copula, which has the minimum *AIC* and *BIC* scores compared to C- and R-vine copulas (Table 4).

Table 4: Calculated *AIC* and *BIC* for various models of vine copula

Vine copula	*AIC*	*BIC*
C-vine	24.11	43.99
D-vine	-10.30	7.16
R-vine	6.68	24.62

Figure 5 presents the probability of failure versus the factor of safety for various vine copula models. As expected, the probability of failure decreases with increasing factor of safety. Among the models, the D-vine copula yielded the highest failure probabilities, ranging from 9.40×10^{-4} to 5.10×10^{-2}, corresponding to target reliability indices from "good" to "poor" performance levels.

The C-vine copula ranked second, with failure probabilities between 1.20×10^{-4} and 2.67×10^{-2}, also covering "good" to "poor" performance levels. In contrast, the R-vine copula produced lower failure probabilities, ranging from 5.00×10^{-5} to 8.80×10^{-3}, corresponding to performance levels from "good" to "below average." Both the C-vine and R-vine copulas substantially underestimated failure probability relative to the D-vine copula, with discrepancies of up to two orders of magnitude. These findings underscore that improper selection of the vine copula structure can lead to significant inaccuracies in system performance estimation.

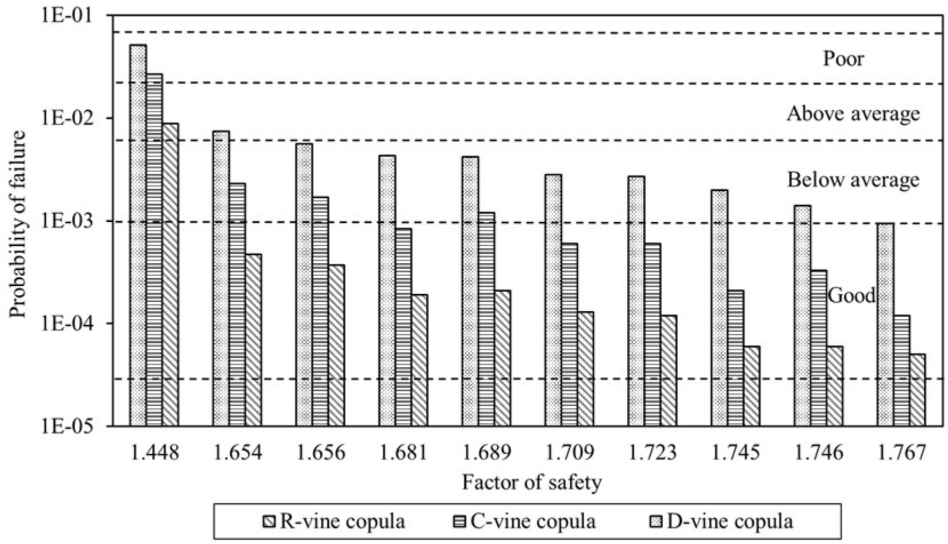

Figure 5: Probability of failure computed based on different vine copula models

3.6 The Vine Copula Model's Performance

The effectiveness of vine copula models was evaluated by comparing calculated and observed values for key statistics, including minimum, quartiles, median, mean, and maximum (Table 5). The multivariate normal distribution, a commonly used dependence model (Liu et al., 2016, Lü et al., 2020), was included for comparison. The D-vine copula emerged as the most suitable model for saturated water content, hydraulic conductivity, unit weight, effective cohesion, residual water content, and the fitting parameter n, showing the lowest absolute deviations across all statistics.

The three vine copula models were compared to identify the most accurate simulation of soil parameter fitting. For approximating the effective friction angle, the D-vine and R-vine copulas performed nearly equally. Across all tested parameters, the D-vine copula consistently yielded the most accurate predictions. The R-vine copula outperformed the C-vine copula in two cases (effective friction angle and fitting parameter α), whereas the C-vine copula outperformed the R-vine only once (fitting parameter α). Overall, the C-vine copula exhibited the lowest performance, likely due to the omission of a critical soil parameter affecting other properties. By comparison, the multivariate

Table 5: Statistics of soil parameters using different dependency models

Parameter	Model	Min.	Q1	Median	Mean	Q3	Max.
k_s (m/s)	Measured	7.86×10^{-8}	7.85×10^{-7}	1.85×10^{-6}	7.68×10^{-6}	5.06×10^{-6}	1.13×10^{-4}
	C-vine	1.14×10^{-7}	8.14×10^{-7}	2.33×10^{-6}	6.76×10^{-6}	6.63×10^{-6}	4.69×10^{-5}
	D-vine	1.15×10^{-7}	$8.15 \times 10^{-7*}$	$2.32 \times 10^{-6*}$	6.77×10^{-6}	$6.61 \times 10^{-6*}$	4.69×10^{-5}
	R-vine	$1.04 \times 10^{-7*}$	8.16×10^{-7}	2.33×10^{-6}	$6.78 \times 10^{-6*}$	6.62×10^{-6}	5.12×10^{-5}
	MNorm	2.76×10^{-8}	7.39×10^{-6}	1.58×10^{-5}	1.85×10^{-5}	2.72×10^{-5}	$8.72 \times 10^{-5*}$
α	Measured	0.01000	0.01076	0.01913	0.02889	0.03305	0.15601
	C-vine	0.00359*	0.01118*	0.02006	0.02864	0.03604	0.11133
	D-vine	0.00279	0.01119	0.02007	0.02871*	0.03601	0.15277*
	R-vine	0.00339	0.01120	0.02004*	0.02865	0.03589*	0.11805
	MNorm	0.00000	0.01518	0.03088	0.03475	0.05065	0.14544
n	Measured	1.26816	1.37592	1.49490	1.55843	1.70930	2.16880
	C-vine	1.14959	1.39552	1.53359	1.54850	1.68523	2.04740*
	D-vine	1.16452	1.39547	1.53354*	1.56131*	1.71135*	2.02162
	R-vine	1.16905*	1.39529*	1.53374	1.54849	1.68551	2.01215
	MNorm	0.77208	1.41198	1.55810	1.54850	1.68585	2.39760
θ_s	Measured	0.27915	0.31537	0.34199	0.34139	0.36990	0.41062
	C-vine	0.27955*	0.31704	0.33821	0.33975	0.36088	0.40956
	D-vine	0.24844	0.31666*	0.34242*	0.34146*	0.36360*	0.47563
	R-vine	0.27830	0.31706	0.33825	0.33975	0.36086	0.41086*
	MNorm	0.21242	0.31939	0.33824	0.33971	0.36098	0.46935
θ_r	Measured	0.00000	0.00000	0.00468	0.00884	0.01981	0.02685
	C-vine	0.00027	0.00426	0.00886	0.01108	0.01547	0.03817
	D-vine	0.00000*	0.00424	0.00886*	0.01099*	0.01547	0.03546
	R-vine	0.00040	0.00424*	0.00887	0.01108	0.01549	0.03513*
	MNorm	0.00038	0.00469	0.00975	0.01108	0.01585*	0.04539
γ (kN/m^3)	Measured	18.5500	19.7550	19.9353	19.8656	20.0141	20.6100
	C-vine	18.6479	19.3809	19.7481	19.7483	20.1159	20.8450*
	D-vine	18.5681*	19.6130*	19.7481	19.7483	20.1149*	20.9390
	R-vine	18.5826	19.3827	19.7493	19.7483	20.1150	20.9090
	MNorm	18.3890	19.3815	19.8680*	19.8705*	20.1302	21.6280
c' (kPa)	Measured	0.8400	4.0669	5.6847	5.5765	7.0810	9.8232
	C-vine	0.7079	4.0335	5.5840	5.6050	7.1363	10.8982
	D-vine	1.0586	4.0382*	5.5853*	5.6048	7.1431	10.3846*
	R-vine	0.7314*	4.0333	5.5839	5.6055	7.1415	10.8703
	MNorm	0.0019	4.0082	5.5547	5.5705*	7.0574*	14.3930
ϕ' (°)	Measured	22.3458	26.2800	28.8548	29.7909	32.8008	39.2238
	C-vine	20.2538	26.4924	29.4753	29.8330	32.7800	42.8090
	D-vine	21.3750*	26.4907	29.4719	29.8341	32.7871*	40.6292*
	R-vine	18.2371	26.4583*	29.4593*	29.8292*	32.8185	47.4296
	MNorm	11.1480	26.6340	29.7400	29.7444	32.8392	48.3880

Notes: The asterisk (*) denotes the best model for a given statistic value of a soil parameter.

normal distribution underperformed, as it assumes normality, which is often not observed in soil attributes. These results indicate that vine copulas are better suited for modelling higher-dimensional variables with diverse marginal distributions.

While this study focuses on comparisons with the multivariate normal distribution, other advanced dependency models warrant consideration. Elliptical copulas, such as the Gaussian and Student's t copulas, can capture global tail dependence, particularly in the case of the Student's t copula. However, elliptical copulas impose a single, homogeneous dependence structure across all variable pairs and assume symmetric dependence. In contrast, vine copulas allow different copula families for each variable pair, enabling modeling of heterogeneous and asymmetric dependencies commonly observed in hydro-mechanical soil parameters.

Non-parametric dependency models provide even greater flexibility by avoiding assumptions about the copula family. Nevertheless, they typically require large datasets for stable estimation and may suffer from reduced interpretability in higher dimensions. In geotechnical applications, where sample sizes are often limited, vine copulas offer a practical balance between modeling flexibility, interpretability, and computational efficiency.

3.7 Sensitivity Analysis of Soil Parameters

A sensitivity analysis was conducted to evaluate the impact of hydro-mechanical soil properties (input variables) on the factor of safety (output variable) and to identify the most critical soil parameter. The sensitivity of each variable was assessed using the weight method proposed by Garson (1991), in which the weights of hidden nodes connecting the hidden and output layers in a Multilayer Perceptron (MLP) network are decomposed into components associated with each input node. Variable importance was expressed as relative importance, ranging from 0 to 100%. The influence of hydro-mechanical soil properties on slope stability during landslide initiation is illustrated in Figure 6.

The sensitivity analysis results indicate that slope stability is primarily governed by shear strength parameters and the hydraulic behaviour of unsaturated soils. The effective friction angle exhibits the highest relative importance (30.24%), highlighting that inter-particle friction and resistance to shear deformation are the most critical factors in maintaining slope stability. This finding aligns with classical slope mechanics, where shear resistance is largely controlled by frictional strength. Residual water content is the second most influential variable (17.49%), representing the minimum moisture retained in soil pores under high suction conditions. Elevated residual water content can reduce suction-induced apparent cohesion, weakening soil shear resistance during prolonged rainfall infiltration.

The shape parameter n and scaling parameter α of the Soil Water Characteristic Curve (SWCC), following Van Genuchten (1980), also exert notable influence. These parameters govern pore-size distribution and air-entry suction, thereby controlling the rate of matric suction reduction during rainfall infiltration. Soils with smaller α or distinct n values may experience faster suction loss, leading to decreased effective stress and reduced stability. Hydraulic conductivity (12.24%) is another significant factor, as it regulates rainfall infiltration rates and pore pressure development. Higher conductivity accelerates water movement into the slope, hastening the reduction of matric suction and increasing failure susceptibility.

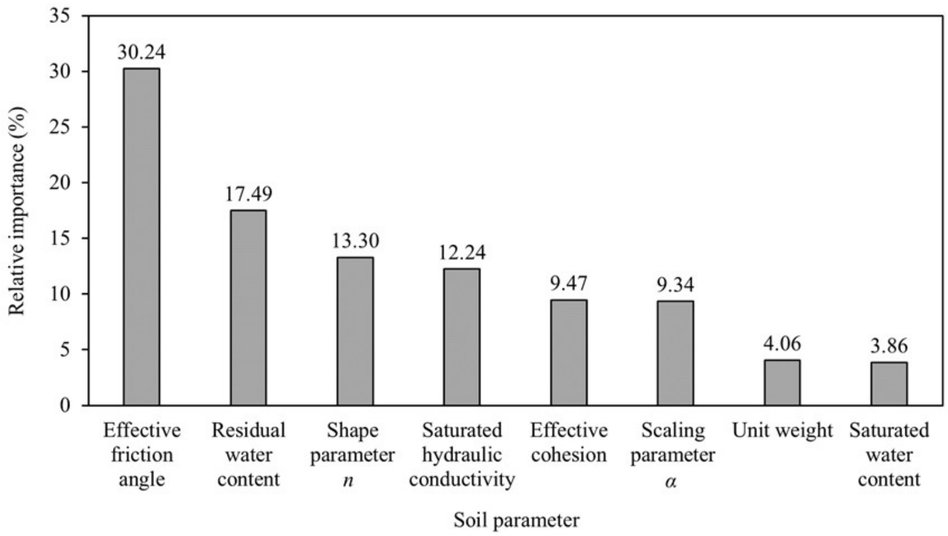

Figure 6: Influence of variables on stability of slope

Effective cohesion contributes moderately (9.47%) to slope stability. In weathered or structured soils, cohesion often reflects bonding effects and soil fabric; however, frictional resistance typically dominates in granular or partially saturated soils. Unit weight shows relatively lower influence (4.06%), as variations in self-weight are minor compared to changes in shear strength and suction effects under the studied slope conditions. Saturated water content exhibits the lowest sensitivity (3.86%), indicating that its variations have limited direct impact on the factor of safety within the parameter range considered.

Overall, the sensitivity analysis indicates that slope reliability under rainfall infiltration is predominantly controlled by mechanical strength parameters and suction-related hydraulic properties, underscoring the importance of accurately characterizing these factors in probabilistic slope stability assessments.

3.8 Generalisability of the Proposed Framework

This study presents a generalizable methodology for analyzing multivariate dependencies among soil-related parameters using a data-driven vine copula approach. While slope reliability analysis was used as an illustrative example, the methodology is adaptable to other soil types, slope geometries, and environmental conditions. The framework is entirely data-driven, relying on the identification of statistical marginal distributions, the application of information criteria such as *AIC* and *BIC* to select the most appropriate bivariate copulas, and the construction of vine structures to capture high-dimensional dependencies. As a result, the methodology does not depend on any specific geologic formation, climate, or slope, but only requires representative data for the soil parameters under study.

Although the application of vine copulas in geotechnical engineering is still in its early stages, several studies have demonstrated their effectiveness across different datasets and geotechnical contexts. For instance, Lü et al. (2020) applied a D-vine copula model to represent multivariate soil parameter distributions, achieving improved performance compared with traditional multivariate normal assumptions. Similarly, Tang et al. (2020) utilized vine copulas to model cross-correlated geotechnical random fields for slope reliability analysis.

4 Conclusion

This study demonstrated the construction and evaluation of several vine copula models, including C-vine, D-vine, and R-vine copulas. The random variables considered were effective cohesion, unit weight, effective friction angle, scaling parameter α and shape parameter n of the Soil Water Characteristic Curve (SWCC), saturated hydraulic conductivity, residual water content, and saturated water content. Among the bivariate copulas, the Clayton and Student's t copulas provided the best fit for most variable pairs.

Of the vine copula models evaluated, the D-vine copula exhibited the best overall performance, as indicated by the lowest *AIC* and *BIC* values, followed by the R-vine and C-vine copulas. In slope reliability analysis, modeling multivariate soil parameters with the D-vine copula produced higher probabilities of failure—yielding more conservative estimates—compared with the C-vine and R-vine copulas.

To ensure reliable predictions, selecting an appropriate vine copula structure is essential. Overall, the D-vine copula demonstrated the greatest capability in simulating the measured soil parameters, followed by the R-vine, C-vine, and, lastly, the multivariate normal distribution.

Acknowledgements

The author extends his gratitude to the Sabah Public Works Department and the Malaysian Meteorological Department for providing invaluable soil and rainfall data, respectively. In addition, the author deeply appreciates the constructive suggestions from the editor and anonymous reviewers, which significantly improved this work.

References

Aas, K. and Berg, D. (2009), "Models for construction of multivariate dependence: a comparison study," *European Journal of Finance*, 15, 639–659.

Bedford, T. and Cooke, R. M. (2001), "Probability density decomposition for conditionally dependent random variables modeled by vines," *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.

Chen, X., Han, J., Zheng, T., Zhang, P., Duan, S., and Miao, S. (2019), "A vine-copula based voltage state assessment with wind power integration," *Energies*, 12, 1–21.

Cheng, Y., Du, J., and Ji, H. (2020), "Multivariate joint probability function of earthquake ground motion prediction equations based on vine copula approach," *Mathematical Problems in Engineering*, 2020, 1–12.

Ching, J. and Phoon, K.-K. (2015), "Constructing multivariate distributions for soil parameters," in *Risk and reliability in geotechnical engineering*, eds. Phoon, K.-K. and Ching, J., Boca Raton, Florida: CRC Press, pp. 3–76.

Das, G. K., Hazra, B., Garg, A., and Ng, C. W. W. (2018), "Stochastic hydro-mechanical stability of vegetated slopes: an integrated copula based framework," *Catena*, 160, 124–133.

Dassault Systèmes (2017), *Abaqus/CAE User's Manual*, version 2017.

Davidovic, N., Prolovic, V., and Stojic, D. (2010), "Modeling of soil parameters spatial uncertainty by geostatistics," *Facta Universitatis - Series: Architecture and Civil Engineering*, 8, 111–118.

Garson, G. D. (1991), "Interpreting neural-network connection weights," *AI Expert*, 6, 47–51.

Geidosch, M. and Fischer, M. (2016), "Application of vine copulas to credit portfolio risk modeling," *Journal of Risk and Financial Management*, 9, 1–15.

Javankhoshdel, S. and Bathurst, R. J. (2017), "Deterministic and probabilistic failure analysis of simple geosynthetic reinforced soil slopes," *Geosynthetics International*, 24, 14–29.

Jiang, C., Xiong, L., Yan, L., Dong, J., and Xu, C. Y. (2019), "Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice," *Hydrology and Earth System Sciences*, 23, 1683–1704.

Joe, H. (1996), "Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters," in *Distributions with fixed marginals and related topics*, eds. Rüschendorf, L., Schweizer, B., and Taylor, M., Hayward, California: Institute of Mathematical Statistics, vol. 28, pp. 120–141.

Li, D.-Q., Zhang, L., Tang, X.-S., Zhou, W., Li, J.-H., Zhou, C.-B., and Phoon, K.-K. (2015), "Bivariate distribution of shear strength parameters using copulas and its impact on geotechnical system reliability," *Computers and Geotechnics*, 68, 184–195.

Liu, S., Zou, H., Cai, G., Bheemasetti, T. V., Puppala, A. J., and Lin, J. (2016), "Multivariate correlation among resilient modulus and cone penetration test parameters of cohesive subgrade soils," *Engineering Geology*, 209, 128–142.

Lü, T. J., Tang, X.-S., Li, D.-Q., and Qi, X. H. (2020), "Modeling multivariate distribution of multiple soil parameters using vine copula model," *Computers and Geotechnics*, 118, 1–14.

Maity, R., Chanda, K., Dutta, R., Ratnam, J. V., Nonaka, M., and Behera, S. (2020), "Contrasting features of hydroclimatic teleconnections and the predictability of seasonal rainfall over east and west Japan," *Meteorological Applications*, 27, 1–20.

Nikoloulopoulos, A. K. (2019), "A D-vine copula mixed model for joint meta-analysis and comparison of diagnostic tests," *Statistical Methods in Medical Research*, 28, 3286–3300.

R Core Team (2019), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rosenblatt, M. (1952), "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, 23, 470–472.

Tang, X.-S., Wang, M. X., and Li, D.-Q. (2020), "Modeling multivariate cross-correlated geotechnical random fields using vine copulas for slope reliability analysis," *Computers and Geotechnics*, 127, 1–14.

Tófoli, P. V., Ziegelmann, F. A., Candido, O., and Pereira, P. L. V. (2019), "Dynamic D-vine copula model with applications to value-at-risk (VaR)," *Journal of Time Series Econometrics*, 11, 1–34.

U.S. Army Corps of Engineers (1997), *Engineering and design: introduction to probability and reliability methods for use in geotechnical engineering*, Washington, D.C.

Van Genuchten, M. T. (1980), "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," *Journal of Soil Science Society of America*, 44, 892–898.

Wang, Y. and Aladejare, A. E. (2016), "Bayesian characterization of correlation between uniaxial compressive strength and Young's modulus of rock," *International Journal of Rock Mechanics and Mining Sciences*, 85, 10–19.

Received: November 16, 2024

Accepted: April 28, 2026