Multivariate Probabilistic Characterization of Hydraulic and Physicochemical Soil Properties

Thesis

submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Atma Prakash (Sharma)

176104007

under the supervision of

Dr. Sreedeep S.

Dr. Budhaditya Hazra

Department of Civil Engineering

Indian Institute of Technology (IIT) Guwahati

Certificate

This is to certify that the thesis entitled *"Multivariate Probabilistic Characterization of Hydraulic and Physicochemical Soil Properties"*, submitted by **Atma Prakash (176104007)**, a research scholar in the Department of Civil Engineering, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy, is a record of an original research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results presented in this thesis have not been submitted to any other university or institute for the award of any degree or diploma.

Dr. Sreedeep S.
Geotechnical Division
Department of Civil Engineering
Indian Institute of Technology

Dr. Budhaditya Hazra
Structural Division
Department of Civil Engineering
Indian Institute of Technology

Acknowledgement

I would like to cordially express my sense of gratitude to my supervisors, **Dr. Sreedeep S.** and **Dr. Budhaditya Hazra** for the valuable supervision both have provided thorough out my masters' and PhD period. I particularly thank them for imparting interest towards the field of academic research, and training me continuously towards it.

My sincere thanks to all the faculty members of the Geotechnical Division, Department of Civil Engineering of the institute for the course work done in masters' in which they introduced various concepts and provided critical insights in the subject to Geotechnical engineering.

My wholehearted thanks to **Dr. Budhaditya Hazra** for the credit courses *"Numerical methods in civil engineering"* and *"Reliability methods in structural engineering"* during my masters' which paved the way for application of probabilistic methodology in this study. I also thank **Dr. TV Bharat** for the audit courses *"Unsaturated Soil Mechanics"* and *"Expansive Soils"* during my PhD, which were extremely useful for this study.

I wholeheartedly thank my doctoral committee panel members **Dr. Anil Mishra**, **Dr. Ravi K.** and **Dr. Mallikarjun C**. for the valuable feedback provided during the oral examinations.

A special thanks to **Dr. Giovanni Spagnoli**, BASF Construction Solutions GmbH, Trostberg, Germany for contributing data to the development of a database and **Professor (Emiretus) Edward Sudicky**, Department of Earth and Environmental Sciences, University of Waterloo, Canada for providing the hydraulic conductivity data used for demonstrating implementation of a methodology presented in this study.

Finally, i would like to thank **Ministry of Human Resource and Development**, India, for providing the financial assistance and the **Indian Institute of Technology, Guwahati**, for all the necessary facilities for successful completion of this work.

Contents

Abstract

Probabilistic characterization of soil parameters is an essential pre-requisite for reliability based design (RBD) in geotechnical engineering projects. A lot of work in probabilistic characterization is focused on conventional mechanical, or, strength parameters for saturated soils. Studies accounting for hydraulic parameters e.g. soil water characteristic curve (SWCC), chemical properties such as cation exchange capacity (CEC) are relatively sparse. However these are equally important aspects for extension of RBD philosophy to unsaturated soils.

SWCC is an essential constitutive relationship for modelling unsaturated soil behaviour. Among the minimal studies for probabilistic characterization of SWCC, almost all of them are limited to soils with low clay fraction. However with increasing use of clay in various geo-environmental projects, characterization of SWCC for clays is also necessary. For this purpose, a database for SWCC of bentonite is compiled from the literature. The proposed approach entails the parametrization of SWCCs and constructing a multivariate probability distribution SWCC parameters. In the absence of measured data, the joint distribution of SWCC provides a first-hand estimate of SWCC. In case of few available site specific data, it provides useful prior information for updating site-specific limited data. For this purpose, the study also formulates a Bayesian approach integrated with copula theory. Efficacy of the proposed approach is also demonstrated for three distinct soil textures.

Apart from SWCC, CEC and specific surface area (SSA) greatly influence the engineering behaviour of clayey soils. However, the measurement of CEC and SSA is not trivial but rather a challenging and time-consuming task. Therefore, this study proposes a multivariate probabilistic approach for the estimation of CEC and SSA. A global multivariate database (labelled as CLAY/C-S/5/278) of five physico-chemical properties: liquid limit (LL), plasticity index (PI), clay fraction (CF), CEC and SSA is compiled. The joint distribution is constructed using vine copula theory. It is shown that the joint distribution can be viewed as "global" prior/unconditional PDF which can be updated to posterior/conditional PDF using Baye's rule when new data is available. Practical geotechnical examples are also shown.

Finally, as a practical RBD application for hydraulic properties, a stochastic seepage and slope stability analysis is conducted. Both the spatial and cross dependence are modelled using vine copula based multivariate random field framework. For investigating the practical engineering importance of dependence structure, stochastic seepage and slope stability analysis under steady and transient seepage conditions is conducted. It is shown that the assumption of arbitrary dependence structure e.g. Gaussian, can significantly affect the RBD results.

Overall this study contributes towards the development of probabilistic models of specific hydraulic and physico-chemical clay properties which in turn are useful inputs for application of RBD approaches in unsaturated porous media.

Keywords: Probabilistic characterization; Multivariate distribution; Copula theory; Vine copula; Soil-water characteristic curve; Specific surface area; Cation exchange capacity; CLAY/C-S/5/278; Stochastic seepage; Slope stability; Spatial dependence structure ; Cross dependence structure

List of Figures

List of Tables

Abbreviations

RBD	Reliability-based design
COV	Coefficient of variation
PDF	Probability density function
CDF	Cumulative distribution function
SWCC	Soil water characteristic curve
CEC	Cation exchange capacity
SSA	Specific surface area
LL	Liquid limit
CF	Clay fraction
PTFs	Pedotransfer functions
RMSE	Root mean squared error
AIC	Akaike information criterion
BIC	Bayesian information criterion
UNSODA	UNsaturated SOil hydraulic DAtabase
K-L	Karhunen-Lo'eve
GMZ	Gaomiaozi (China)
FEBEX	Full-scale Engineered Barriers Experiment (Spain)
MX- 80	Wyoming (USA)
AEV	Air entry value
DIC	Deviance Information Criterion
RSS	Residual sum of squares
CM	Cramer-von Mises
K-L	Kullback and Leibler
vG	van Genuchten
K-S	Kolmogorov-Smirnov
MCMC	Markov Chain Monte Carlo
M-H	Metropolis-Hastings
PI	Plasticity index
C-vine	Canonical vine
D-vine	Drawable vine
EGME	Ethylene glycol monoethyl ether
SP	Segregation potential
GEV	Generalized extreme value
FS	Factor of safety

Nomenclature

S	Degree of saturation
ψ	Suction
R^2	Coefficient of determination
S_r	Residual degree of saturation
α	van Genuchten parameter
n	van Genuchten parameter
m	van Genuchten parameter
w	Water content
θ_w	Volumetric water content
N	Sample size
$C(.)$	Copula function
$D(.)$	Copula density
$F(.)$	CDF
$f(.)$	PDF
θ	Copula parameter
$P(.)$	Probability
Φ	Multivariate standard normal function
Φ^{-1}	Inverse CDF of one-dimensional standard normal distribution
$t_v^{-1}(.)$	Inverse of student t distribution
$l(\theta)$	Log-pseudo likelihood function
$\Gamma(.)$	Gamma function.
U_ij	Pseudo observation
$D_\theta^G(.)$	Gaussian copula density
$D_\theta^t(.)$	t copula density
S_n	CM statistics
$l(\theta_n)$	Maximized log likelihood
k	Number of free parameters
$E(.)$	Expectation
τ	Kendall's rank correlation coefficient
μ	Mean
σ	Standard deviation
λ	Lognormal parameter related to mean
ζ	Lognormal parameter related to standard deviation
S_e	Effective degree of saturation

θ_w	Volumetric water content
θ_r	Residual volumetric water content
θ_s	Saturated volumetric water content
ρ	Pearson's correlation coefficient
N_s	Simulated sample size
c,	Cohesion
ϕ	Angle of internal friction
γ_t	Total unit weight
$c(\psi)$	Apparent cohesion resulting due to ψ.
H	Depth of failure
β	Slope angle
G_s	Specific gravity of solid
e	Void ratio
P_f	Failure probability
$g(X)$	Limit state function
p	Swelling pressure
ρ_d	Dry densities
Q	Quantile
T_i	Tree number
SP	Segregation potential
$\nabla(T_f)$	Temperature gradient in the frozen zone
d_{50}	Median diameter
n_o	Pore fluid ion concentration ($ions/m^3$)
K	Boltzmann's constant
u	Non-dimensional mid plane potential
T	Temperature in Kelvins
σ	Surface charge density at clay surface
e_o	Vacuum permittivity
D	Pore fluid dielectric constant
K	Diffuse-double layer parameter
d	Half the clay platelets distance
ρ_w	Water density
v	Exchangeable cations valency

Chapter 1

Introduction

1.1 General

Reliability-based design (RBD) approaches are becoming increasingly popular for geotechnical engineering projects (Phoon and Ching, 2014). A lot of them have recently been incorporated into the design codes as well (Phoon and Retief, 2016). RBD provides a rational basis for incorporating uncertainties (e.g. soil properties, model uncertainties) explicitly into the geotechnical design (Phoon and Ching, 2014). Any design or analysis procedure using RBD philosophy requires statistical characterization of uncertain parameters. The characterization can be in the form of point estimates, e.g. mean, coefficient of variation (COV), quantiles, or complete probability density function (PDF).

A common approach in the current state of the art is to treat the material parameters in the constitutive relationship as random quantities. For example, the cohesion and friction parameters in the Mohr-Columb model and soil-water characteristic curve (SWCC) parameters in the van Genuchten (1980) model. Very often, these parameters exhibit significant correlation among themselves. This aspect renders the task of probabilistic characterization essentially multivariate. Accurate characterization of correlation requires not only the correlation value but also the correlation structure. For this purpose most commonly used model is the multivariate normal distribution (Ching et al., 2017). This model assumes Gaussian marginals along with Gaussian dependence structure. Even the popular translational approach (Ching and Phoon, 2014, e.g.) is essentially a marginal transform, the Gaussian dependence structure is still inherited. Copula theory (Nelsen, 2007) provides an alternative to conventional translational models.

Copula can be considered as a function that "couples" or "joins" multiple one-dimensional marginals to their parent joint distribution. It is essentially a multivariate distribution with uniform one-dimensional marginals. Sklar (1959) theorem paved the way for the application of copula theory. It states that every multivariate cumulative distribution function (CDF) of a random vector can be expressed in terms of its marginals and a copula. Mathematically it can be expressed as follows:

$$H_{X,Y}(x,y) = C(F_X(x), F_Y(y)|\theta) \tag{1.1}$$

where $H_{X,Y}(x,y)$ is the joint CDF, $F_X(x), F_Y(y)$ are the respective one dimensional marginal

distribution and C is the copula function with dependence parameter θ. According to Sklar (1959) theorem, C is unique if the marginals $F_X(x)$ and $F_Y(y)$ are continuous. Similarly, if there exists a multivariate CDF $H_{X,Y}(x, y)$ with continuous marginals $F_X(x)$ and $F_Y(y)$ such that $u = F_X(x)$ and $v = F_Y(y)$, the copula C can be obtained as follows:

$$C(u, v|\theta) = H_{X,Y}(F_X^{-1}(u), F_Y^{-1}(v)) \tag{1.2}$$

where F_X^{-1}, F_Y^{-1} are the inverse CDFs, also known as quantile functions, for the marginal univariate distribution. It is easy to observe that a bivariate copula is defined over unit square $[0, 1]^2$ since $u \in [0, 1], v \in [0, 1]$. This implies that the copula can also be expressed in terms of joint CDF of U, V as follows:

$$C(F_X(x), F_Y(y)|\theta) = C(u, v|\theta) = P(U \leq u, V \leq v) \tag{1.3}$$

where $P(.)$ represents the joint probability. The joint probability density function (PDF) $f_{X,Y}(x, y)$ can be obtained after differentiating Eq. 1.1 as follows:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) D(u, v|\theta) \tag{1.4}$$

where $f_X(x) = dF_X(x)/dx$ and $f_Y(y) = dF_Y(y)/dy$ are the one dimensional PDFs and $D(u, v|\theta) = \partial^2 C(u, v|\theta)/\partial u \partial v$ is the copula density function. One of the main advantages that copula approach provide is that the marginals $F_X(x)$ and $F_Y(y)$ can be from any distribution family and also can be widely different from each other. It can be easily noted in Eq. 1.4 that the task of modelling the random variables ($f_X(x)$ and $f_Y(y)$) and their dependence structure ($D(u, v|\theta)$) are independent of each other. This is a significant advantage over the other conventional methods such as translational approach (Carsel and Parrish, 1988; Phoon et al., 2010) for construction of joint distribution since this allows a more general and flexible structure. This particular aspect is also one of the major reasons for its rising popularity and its use in this dissertation. Although the copula model is indeed a more general and versatile model for modelling dependency, and it encompasses the Gaussian dependency as a special case, certain disadvantages are also associated with it. For example, the use of copula in modelling high dimensional variables is not as straightforward and convenient as Gaussian. The computational tasks associated with the Gaussian model are more straightforward as the conditional distributions are Gaussian and well established analytical solution exist for Bayesian estimation.

Fig. 1.1 presents a schematic for RBD. It is essential to characterize uncertainty in model parameters or boundary conditions of the governing equations prior to performing any reliability analysis. The characterization can be in the form of mean, COV, univariate PDF or multivariate PDF depending upon the problem statement. The PDF acts as an input parameter for the governing equation(s) of the problem statement. The reliability based analysis can be achieved using various frameworks such as Monte Carlo simulation (Ang and Tang, 2007). The output of the governing equations can be coupled with performance criteria of the structure (e.g. Factor of safety (FS)) resulting in a limit-state equation. In probabilistic framework, the interest for design is usually

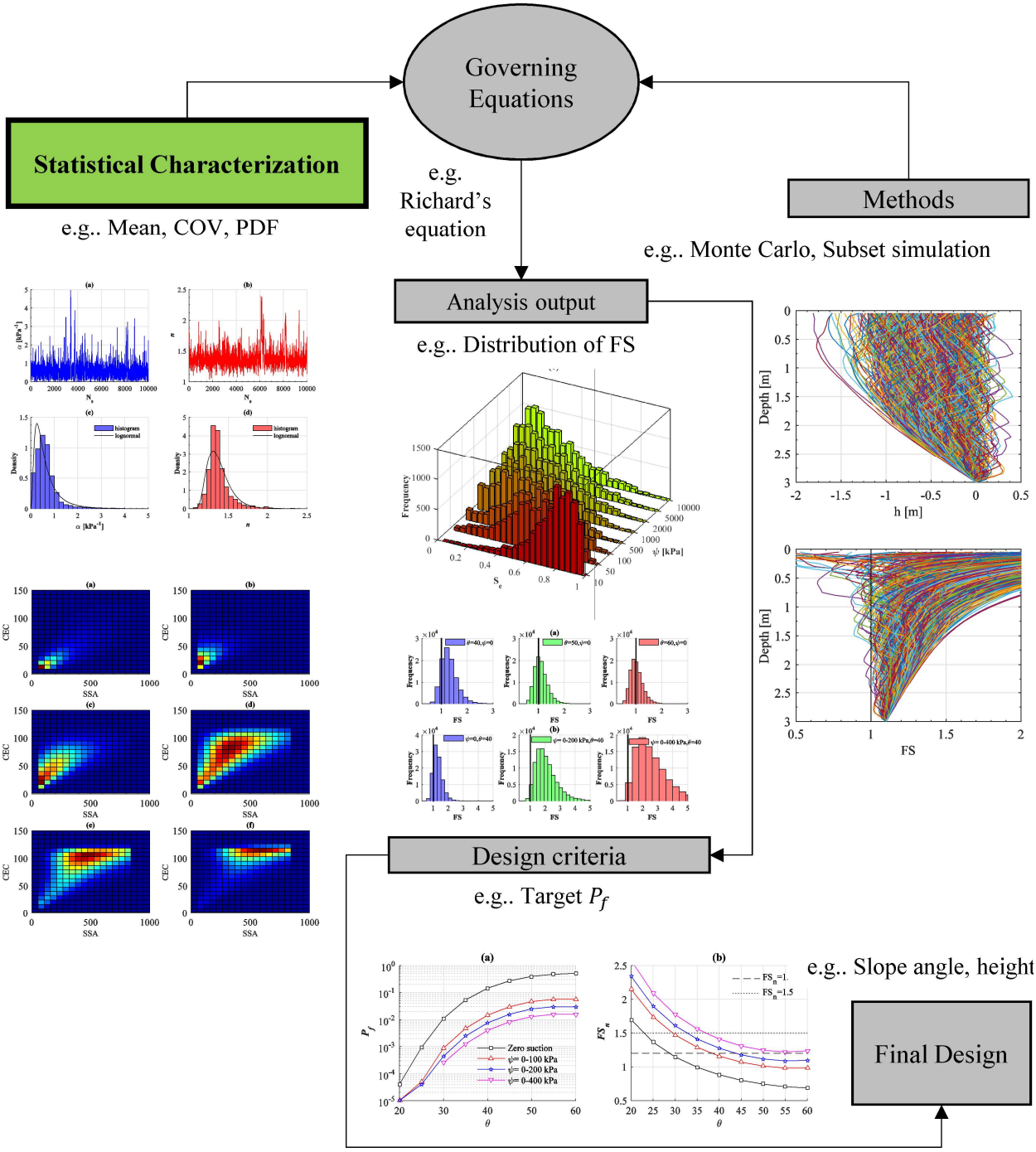


Figure 1.1: A schematic for reliability based design (RBD) with statistical characterization as the main focus of this thesis.

Table 1.1: Acceptable failure probability P_f after Salgado and Kim (2014).

Conditions	Acceptable P_f
Temporary structures: no potential life loss, low repair cost	0.1
Minimal consequences of failure: bench slope or open pit mine	0.1-0.2
Minimal consequences of failure: repair cost is less than cost of reducing P_f	0.01
Existing large cut on interstate highway	0.01-0.02
Large cut on interstate highway to be constructed	<0.01
Lives may be lost when slopes fail	0.001
Acceptable for all slopes	0.0001
Unnecessarily low	0.00001

the probability of satisfaction or, violation of the limit state function $g(X)$, e.g. $g(X) \leq 0$. The failure probability (P_f) can be estimated in terms of limit state function $g(X)$. For a slope stability problem, the limit state function can be defined as: $g(X) = FS - 1$ and P_f can be estimated in terms of input parameters $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ as follows.

$$P_f = P(g(X_1, X_2, ..., X_n) \leq 0) = \int ... \int_{g(\mathbf{X}) \leq 0} f_X(x) dx \tag{1.5}$$

In the Monte Carlo simulation framework, the above can be estimated using the standard indicator function approach as follows.

$$P_f = \frac{1}{N} \sum_{i=1}^{N} I(g(\mathbf{X}) \leq 0) = \frac{1}{N} \sum_{i=1}^{N} I(FS - 1 \leq 0) \tag{1.6}$$

where N is the number of realizations of FS and I is an indicator function for counting the number of times the slope fail i.e. $FS \leq 1$. Table 1.1 presents some commonly accepted target failure probability for various engineering structures in geotechnical engineering. Based on acceptable target P_f, design target, e.g. slope height, or, slope angle can be assigned a value. Among all the steps illustrated in Fig. 1.1, statistical characterization in the form of multivariate PDF is the main focus of this thesis. The task of constructing the multivariate PDF is achieved in this thesis using copula theory.

A lot of work in geotechnical engineering is focused on probabilistic site characterization where mostly strength parameters such as undrained shear strength, uniaxial compressive strength, Young's modulus, cohesion and friction were considered (Wang and Cao, 2013; Wang and Aladejare, 2016, e.g.). Most of the extensive works for probabilistic characterization of hydraulic parameters are for saturated hydraulic conductivity (Sudicky, 1986; Li et al., 2009; Zhang, 2001). Probabilistic studies accounting for unsaturated hydraulic parameters e.g. soil-water characteristic curve (SWCC) (Phoon et al., 2010; Chiu et al., 2012; Prakash et al., 2018*a*, 2017, e.g.), and physico-chemical properties such as cation exchange capacity (CEC) and specific surface area (SSA) are relatively sparse. However, probabilistic characterization of these parameters is important for extending RBD

philosophy in unsaturated porous media. The three parameters considered in this study are SWCC, CEC and SSA. These three properties are 3 out of the 5 important properties (the other two are soil water density and suction stress characteristic curve) identified by Lu and Mitchell (2019) for *"addressing new world challenges in geotechnical engineering"*.

1.2 Objectives

In the wake of the aforementioned discussion, objectives of this study are listed as follows.

1. Probabilistic characterization of SWCC of bentonite
2. Probabilistic characterization of SWCC from limited data
3. Probabilistic estimation of CEC and SSA
4. Stochastic seepage analysis considering non-Gaussian spatial and cross dependence structure of hydraulic parameters

1.3 Scope and organization of the thesis

Having identified the objectives, this thesis is organized chapter-wise as follows:

1. Chapter 2 covers the literature background and motivation leading to the objectives presented in this thesis.
2. Chapter 3 covers probabilistic characterization of bentonite SWCC using multivariate copula theory. For this purpose, a database for SWCC of bentonite is compiled from the literature. The proposed approach entails the parametrization of SWCCs and constructing a multivariate probability distribution for van Genuchten (1980) model parameters using copula theory.
3. Chapter 4 formulates a Bayesian approach integrated with copula theory to obtain the site-specific probability distribution of SWCC. Limited number of SWCC are incorporated with prior knowledge to obtain the updated probability distribution of SWCC.
4. Chapter 5 proposes a multivariate probabilistic approach for the estimation of CEC and SSA. A global multivariate database (labelled as CLAY/C-S/5/278) of five physico-chemical properties: liquid limit (LL), plasticity index (PI), clay fraction (CF), CEC and SSA is compiled.
5. In Chapter 6, a practical RDB application for hydraulic properties, stochastic seepage and slope stability analysis is conducted using vine copula based multivariate random field framework. Both spatial and cross dependence are modelled. For demonstrating engineering importance of dependence structure, a case study for slope stability under steady and transient seepage conditions is formulated.

6. Chapter 7 finally summarizes the major conclusions, contributions, limitations and possible future scope of this study.

Among the above chapters, the main body of this thesis is organized into four chapters, i.e. Chapters 3-6. Chapters 3 and 4 deal with probabilistic characterization of *hydraulic properties* i.e. SWCC, and Chapter 5 covers characterization of *physicochemical properties* i.e. CEC and SSA. Finally, for demonstrating a RBD application of hydraulic properties, Chapter 6 conducts a stochastic seepage and slope stability analysis. The task of multivariate probabilistic characterization in all the chapters mentioned above is achieved using copula theory. Chapter 3 utilizes multivariate copula theory for the trivariate joint distribution. Chapter 4 formulates a copula-based Bayesian approach for bivariate probabilistic characterization. Chapter 5 constructs the 5-dimensional joint distribution using vine copula theory, and Chapter 6 utilizes the vine copula theory in n dimensions for spatial data.

Chapter 2

Background, literature review and motivation

2.1 General

This chapter presents the literature background leading to motivation for the objectives of this study. Along with the literature background, limitations in the current state of the art are identified and the objectives formulated. Sections 2.2 covers probabilistic characterization of hydraulic property, SWCC. Section 2.3 deals with probabilistic estimation of physico-chemical properties, CEC and SSA. Finally, section 2.4 covers the literature background related to an application area for hydraulic properties i.e. stochastic seepage analysis.

2.2 Probabilistic characterization of SWCC

Soil-water characteristic curve (SWCC) is an essential constitutive relationship for modelling unsaturated soil behaviour. Among the very limited studies for probabilistic characterization of SWCC, almost all of them are limited to soils where clay fraction is not high. However with increasing use of clay in various geo-environmental projects, characterization of SWCC of clays is also necessary. Bentonites are increasingly used for various geo-environmental applications owing to their low hydraulic conductivity and high water retention capacity. It is used as barrier material for deep underground disposal of high-level radioactive waste (Villar, 2002; Villar and Lloret, 2004; Villar, 2005; Lloret et al., 2005; Tripathy et al., 2015), liner or cover material for municipal waste containment facility to prevent the contamination of groundwater (Kayabali, 1997; Kumar and Yong, 2002; Gates et al., 2009; Osinubi and Amadi, 2009). It is also used for remediating the contaminated soils (Gitipour et al., 1997; Ling et al., 2008). Investigating the problems mentioned above necessitate the knowledge of soil-water characteristic curve (SWCC) of bentonite, which maps the variation of water content with suction. It is the most fundamental constitutive relationship of unsaturated soils and a mandatory input for unsaturated behavioral modelling. However, the measurement of SWCC is not a straightforward task and it is often expensive, time-consuming and cumbersome (Phoon et al., 2010). The problem becomes more acute for deformable materials like bentonites where simultaneous volumetric measurements are also needed for constructing a meaningful SWCC. While the SWCC is expected to be "characteristic" or unique for a material, quite often there are

various uncertainties associated with both measurements as well as estimation techniques (Likos and Yao, 2014; Abhijit and Sreedeep, 2014).

One of the commonly used estimation methods is to use pedotransfer functions (PTFs) (Schaap et al., 2001; Wösten et al., 2001; Cornelis et al., 2005; Mbonimpa et al., 2002). The PTFs are a correlation between SWCC parameters and some easily measurable soil properties such as grain size distribution (Arya and Paris, 1981) and Atterberg limits (Thakur et al., 2006). However, many a time the constraint being used in SWCC equations like van Genuchten (1980) is either not mentioned or ignored. It is noted from the literature that the choice of constraints may have a significant impact on the parameters and hence the developed correlations (Likos and Yao, 2014). Hence, it is essential to evaluate the impact of various constraints commonly used with vG equation and identify the best possible representation of the measured data. The measurement of SWCCs is also subjected to inherent uncertainties resulting in a wide variability of SWCC for a given soil. These uncertainties can result from the improper choice of instruments and measuring range (Abhijit and Sreedeep, 2014; Prakash et al., 2017, 2018*b*), initial density of the material (Huang et al., 1998; Mbonimpa et al., 2006; Nuth and Laloui, 2008; Zhou et al., 2012) and hysteresis (Wheeler et al., 2003; Khalili et al., 2008; Nuth and Laloui, 2008; Gallipoli, 2012) to state a few.

It is apparent that an engineering estimate from the statistical generalizations of experimental data from soils of a similar class is extremely useful and has been successfully employed in the past for various classes of soils (Carsel and Parrish, 1988; Phoon et al., 2010, e.g.,). Carsel and Parrish (1988) constructed joint probability distributions for SWCC parameters using the UNSODA database (Nemes et al., 2001) for various class of soils. A preliminary statistical analysis was conducted by Sillers and Fredlund (2001) on 230 SWCCs and the first and second moments were reported for the SWCC parameters. Phoon et al. (2010) proposed a first-hand estimate from the statistical generalization of SWCC belonging to a similar class of soils based on which a joint lognormal model for SWCC parameters was proposed. Chiu et al. (2012) used Bayesian approach to obtain the updated probability density function (PDF) of uncertain parameters associated with SWCC and created confidence intervals for various classes of soil. A first-order error analysis approach was utilized by Zhai and Rahardjo (2013) to evaluate the uncertainties related to SWCC and for determining confidence intervals. Prakash et al. (2018*b*) used copula theory to construct the bivariate distribution of vG parameters α and n for quantifying instrumental and material based uncertainty in fly ash.

However, it may be noted that these databases used in the aforementioned studies do not contain a sufficient number of data points for highly expansive clays like bentonites. While the sample size for classes of soil such as sand, silt, sandy clay, clayey loam was adequate, for clays it was rather low or not available. The reason for the same can be attributed to the intricacies related to the measurement of high suction in clays and longer time duration required as compared to sandy or silty soils. For example, measuring a single data point of water content and suction for bentonite can take up to 50 days (Gatabin et al., 2016). Measurement of high suction not only requires higher equilibration time but also need multiple samples to ensure repeatability. The significant

swelling and shrinkage behaviour of bentonites necessitate the knowledge of volumetric shrinkage to be measured simultaneously for computing the degree of saturation (S) (Tripathy et al., 2014; Dieudonne et al., 2017). This aspect results in SWCC determination being more time consuming and a tedious task. Another limitation in the previously reported studies is that the choice of van Genuchten (vG) parameter constraints for subsequent statistical analysis was chosen arbitrarily for convenience, such as the residual degree of saturation (S_r)=0 and parameter $m = 1 - 1/n$ (Carsel and Parrish, 1988; Phoon et al., 2010; Prakash et al., 2018*b*, e.g.,). No rigorous evaluation of constraints was performed in previous studies in an optimal sense, specifically for bentonites. However, it is crucial as the choice of vG constraints may have a significant influence on the unsaturated properties of soils being estimated.

In this study (Chapter 3), a database is developed exclusively for the SWCC of bentonites by compiling measured S and ψ from the literature (ref. Table 3.1). Initially, the impact of various commonly used constraints in van Genuchten (1980) equation on the statistics of obtained parameters is investigated based on the developed database. Further, the performance evaluation of these constraints is done in terms of accuracy (Root mean squared error(RMSE)), linearity (Coefficient of determination R^2) and information criterion (Akaike information criterion (AIC) and Bayesian information criterion (BIC)). Using the best fit model, the scatter in the compiled database of SWCCs is reduced to a set of curve fitting parameters α, n, m. Thereafter, trivariate distribution of the same is evolved using copula theory.

A common approach in the aforementioned studies (Carsel and Parrish, 1988; Phoon et al., 2010; Chiu et al., 2012) is to use a database (e.g. UNSODA (Nemes et al., 2001)) of SWCC for a particular texture (class) of soil and thereafter construct the joint probability distribution of the SWCC parameters. These studies are mostly focused towards building a "global" multivariate distribution, which are useful for understanding the nature of variability and dependence structure among SWCC parameters. These "global" distributions also serve as useful prior information in context of probabilistic model updating (e.g., Phoon, 2018; Ching and Phoon, 2018) for site-specific SWCCs.

In some of the aforementioned studies (e.g. Chiu et al., 2012; Prakash et al., 2018*a*) unsaturated RBD has been conducted to demonstrate applicability of these "global" multivariate distribution of SWCC. These examples are useful for understanding the propagation of uncertainty in SWCC towards the RBD. However, these "global" distributions may not be completely representative of an actual site (Ching and Phoon, 2018). Ideally, conducting an actual unsaturated RBD for a project requires joint distribution of SWCC specific to that site. In theory this can always be achieved by measuring multiple SWCCs corresponding to each source of uncertainty at that site and thereafter construct the site specific SWCC distribution. However this task in not practically feasible due to the following reasons: (a) construction of joint distribution by conventional approaches requires at least 30-50 data points (Ang and Tang, 2007) (b) measurement of even a single SWCC in itself is time consuming, expensive and tedious. For example, measurement of SWCC for coarse grained soils like sandy or silty soils can take up to weeks where as for clayey soil it can take even up to

50 days (e.g. Gatabin et al., 2016). This aspect essentially implies that obtaining a site specific distribution of SWCC is "practically impossible" using conventional approach. In such cases, when there is not sufficient data for characterization of site specific variability, Bayesian approaches are useful and have successfully been utilized to construct the site specific distribution of design parameters in conventional (saturated) geotechnical engineering (e.g. Wang and Cao, 2013; Wang et al., 2015; Cao and Wang, 2014; Wang and Aladejare, 2016; Zhang et al., 2018; Ching and Phoon, 2018; Bozorgzadeh et al., 2019). A common thread among these aforementioned studies is to formulate appropriate priors using "global" distributions and "hybridize" it with the limited site specific data. The result is a "hybridized" distribution which provides a better estimate of population statistics than the one obtained only using available limited site specific data (Ching and Phoon, 2018). Given the fact that obtaining multiple SWCCs is even more difficult than the conventional (saturated) soil parameters (Phoon et al., 2010), this study hypothesizes that Bayesian approach can be utilized to construct the site-specific distribution of SWCC using limited data.

It should be noted that in most of the aforementioned studies utilizing Bayesian approach (e.g. Wang and Cao, 2013; Wang et al., 2015; Cao and Wang, 2014; Bozorgzadeh et al., 2019) are in the context of univariate distribution. In some cases where multivariate distribution was constructed (e.g. Wang and Aladejare, 2016; Ching and Phoon, 2018) multivariate normal distribution was utilized. SWCC parameters are not only significantly correlated but also the marginal distributions are skewed and highly non-normal (Phoon et al., 2010; Prakash et al., 2018*b*). However construction of non-normal joint distribution using conventional approaches such as translational methods (e.g. Phoon et al., 2010) are difficult to integrate within the Bayesian approach. Very few studies (e.g. Zhang et al., 2018) have utilized Bayesian approach to model multivariate non-normal distribution. Copula theory (e.g. Prakash et al., 2017, 2018*b*) provides a flexible and more general approach and is therefore integrated with the Bayesian approach in this study. Therefore, this study (Chapter 4) formulates a Bayesian approach to construct the joint distribution of SWCC parameters using limited number of SWCCs.

2.3 Probabilistic estimation of CEC and SSA

Surface charge plays a crucial role in influencing the physical, chemical and engineering behaviour of fine-grained soils. Cation exchange capacity (CEC) and specific surface area (SSA) are the two key factors governing the surface charge and are widely recognized as fundamental clay properties (Mitchell et al., 2005; Khorshidi and Lu, 2017; Cerato and Lutenegger, 2002; Santamarina et al., 2002; Khorshidi et al., 2016; Lu and Mitchell, 2019). CEC is essentially a measure of exchangeable cations on the negatively charged clay surface. It is commonly expressed in milliequivalents (meq) per unit 100 grams. SSA commonly expressed in m^2/g is the ratio of total (external + internal) surface area to the soil mass. It is regarded as an intrinsic soil property similar to grain-size distribution of soils (Cerato and Lutenegger, 2002). CEC along with SSA represents the surface charge density ($\sigma = CEC/SSA$) at the clay surface. It is well established that CEC and SSA

greatly influence the engineering behaviour of clayey soils such as diffusion, fluid flow, sorption and swell shrink behaviour (Santamarina et al., 2002; Low, 1980; Khorshidi et al., 2016; Khorshidi and Lu, 2017). CEC and SSA are crucial parameters in geotechnical engineering practice where physicochemical forces play a dominant role e.g. compressibility characteristics (Tripathy et al., 2014, e.g.), consolidation (Dutta and Mishra, 2016, e.g.), frost heave susceptibility (Konrad, 2005, e.g.) etc. Both these properties are directly relevant to the challenges mentioned in "*role of geotechnics in addressing new world problems*" (Lu and Mitchell, 2019, p.1) such as carbon storage and capture, nuclear energy and long term waste containment (Lu and Mitchell, 2019, p.108).

The measurement of CEC and SSA is not easy but rather a challenging and time-consuming task (Khorshidi et al., 2016; Khorshidi and Lu, 2017). The available methodologies and measurement specific errors induce various uncertainties that may lead to non-unique values of CEC and SSA (Lu and Mitchell, 2019). Some common sources of uncertainty in the determination of CEC are soil pH, saturating and displacement solution (Khorshidi and Lu, 2017). Due to the challenging and non-trivial procedure for direct determination of CEC and SSA, it is common to employ transformation models relating CEC and SSA with other geotechnical properties such as Atterberg limits (liquid limit (LL) and plasticity index (PI)), clay fraction (CF) and soil water retention curve (Farrar and Coleman, 1967; Smith et al., 1985; Yukselen and Kaya, 2006; Khorshidi et al., 2016; Khorshidi and Lu, 2017; Spagnoli and Shimobe, 2019, e.g.). However, there are certain limitations associated with most of these models that need attention. Some of them are as follows.

- These models were developed on limited sample sizes and work satisfactorily mostly on their own calibrated dataset. It is not plausible to extrapolate these models beyond the range of calibration dataset. For example, a model developed on a dataset containing smectite dominant clays cannot be expected to work satisfactorily for a kaolinite dominant dataset. This is because the behaviour of these two clay minerals is completely different.

- These models only provide a point estimate, i.e. the value at the trend line of the transformation model. There is little or no room for uncertainty quantification in the estimate. Only if the degree of correlation between variables is close to one, then the transformation uncertainty can be considered negligible, and a point estimate (e.g. mean) satisfactory.

- Most of the models are developed for pair-wise estimation only, i.e. for the estimation of CEC and SSA, only one explanatory variable at a time can be used. For example, the most popular models for both CEC and SSA rely on liquid limit(LL) or plasticity index(PI). These models cannot be used to reduce the uncertainties in the estimate when multiple clay parameters are available, e.g. CF.

- Most of the models treat SSA and CEC as independent quantities, and separate transformation models are proposed. However, it will be shown later that CEC and SSA form a random vector instead of independent random variables.

Given the aforementioned four significant limitations in the current state of the art, this study proposes (Chapter 5) a multivariate probabilistic approach for estimation of CEC and SSA. For

the construction of the multivariate model, a five-dimensional database (278×5) (labelled as CLAY/C-S/5/278) is developed. However it should be noted that the construction of 5 dimensional distribution is not a trivial task particularly for this database since three variables CF, CEC and SSA have a theoretical/practical upper bound at 100%, $135meq/100g$ and $800m^2/g$ respectively. This aspect requires the use of right truncated distribution and the popular multivariate normal or translational approach (Ching and Phoon, 2014; Ching et al., 2017; Ching, Li, Phoon and Weng, 2018, e.g.) cannot be used to evaluate the joint density easily.

Therefore this study uses copula approach as it allows for arbitrary marginals as the dependence and marginal modelling are uncoupled and can be modelled separately. However even the copula approach is popular in geotechnical literature only till 2 dimensions (Li et al., 2015; Tang et al., 2015; Prakash et al., 2018*a*, 2019, 2018*b*, e.g.). Beyond 2 dimensions, very few generalizations of bivariate copula exists and they too have their own limitations (Wu, 2013*b*, e.g.). Vine-copula approach overcomes this limitation (Bedford and Cooke, 2001; Aas et al., 2009) and is therefore utilized in this study. The constructed 5 dimensional joint distribution not only summarizes the marginals but also the conditional dependence among them. It is shown that the constructed joint distribution can be viewed as "global" prior/unconditional PDF which can be updated to posterior/conditional PDF using Baye's rule when new data is available. The proposed approach is superior and more generalized than the current pair-wise regression approach.

2.4 Stochastic seepage analysis with consideration to uncertainties in hydraulic parameters

Stochastic seepage and slope stability analysis is crucial for the reliability based design of embankments, dams, hydraulic structures, landslide early warning systems etc. (Yeh, 1989; Griffiths and Fenton, 1993; Fenton and Griffiths, 1997; Gui et al., 2000; Zhang, 2001; Griffiths and Lu, 2005; Rahardjo et al., 2007; Lu and Godt, 2008; Li et al., 2009; Srivastava et al., 2010; Zhang et al., 2011; Santoso et al., 2011; Godt et al., 2012; Cho, 2014; Dou et al., 2015; Ering and Babu, 2016; Canli et al., 2018; Yuan et al., 2019; Correa et al., 2020, e.g.). Importance of seepage induced slope instability can also be judged from the numerous case studies reported in literature. Lu and Godt (2008); Godt et al. (2012) presented a case study of slope failure at Seattle, Washington. Leshchinsky et al. (2015) investigated a case study of shallow landslide in south Korea. Muceku et al. (2016) analysed the stability of hill slope around Berati Town, Albania. Kuriqi et al. (2016) investigated seepage effect on river dike stability along the Buna river in Albania. Ering and Babu (2016) analysed the Malin landslide in india. Many more cases studies could be found in literature such as Gian et al. (2017) (Vietnam), Pham et al. (2018)(Pohang, Korea; Boso peninsula, Japan; Bologna Italy) and Wang et al. (2020) (Gansu Province, China) among many others.

For the stochastic seepage analysis, statistical characterization of hydraulic parameters is commonly done using spatially varying random fields (Vanmarcke, 1977, 2010, e.g). This aspect renders the governing partial differential equation (Darcy's law and continuity equation) essentially stochas-

tic. A key aspect in solving the stochastic partial differential equation for seepage is generating random fields of the input parameters, i.e. saturated permeability k_s and soil-water characteristic curve parameters (SWCC) (e.g. α and n) (Zhang, 2001). Popular random field simulation techniques in geotechnical engineering such as local average subdivision (Fenton and Vanmarcke, 1990; Griffiths and Fenton, 2008; Griffiths et al., 2011; Zhu et al., 2015; Zhu, Griffiths, Huang and Fenton, 2017), covariance matrix decomposition (Griffiths and Fenton, 2008; Srivastava et al., 2010; Ching et al., 2019; Li et al., 2019, e.g.) and Karhunen–Loève (K-L) expansion (Phoon et al., 2005; Li et al., 2009; Zhao and Wang, 2018; Montoya-Noguera et al., 2019, e.g.) assume a multivariate Gaussian model to characterize the joint distribution underlying the random field. This essentially implies Gaussian marginals along with Gaussian dependence structure (copula). Although translational approach (Phoon et al., 2005; Shields et al., 2011; Ferrante et al., 2005, e.g.) are commonly used to generate random fields with non-Gaussian marginals, the Gaussian dependence structure is still inherited.

An alternate to translational approach, copula theory offers flexibility to handle non-Gaussian dependence structure as the marginal and dependence modelling are uncoupled and can be done separately. However, even traditional copula approaches have some severe limitations regarding generalizations in n dimensions (Aas et al., 2009), which is essential for random fields. Some of these limitations are (a) out of umpteen families of copula only Gaussian and t copula can be generalized to arbitrary n-dimensions (b) some alternative to this such as Nested Archimedian copula do have their limitations regarding the restrictions on parameters and dependence structure (Aas et al., 2009). Due to the aforementioned reasons, a majority of work utilizing copula theory in geotechnical engineering is dedicated to the bivariate random variable in the context of cross-dependence only (Li et al., 2015; Tang et al., 2015; Prakash et al., 2018*a*; Zhang et al., 2018, e.g.). Beyond 2 dimensions ($n > 2$), very few generalizations of bivariate copula exists and they too have their own limitations regarding the dependence structure (Wu, 2013*b*; Zhu, Zhang, Xiao and Li, 2017, e.g.). This is also partly because a lot of bivariate copula families to handle various kind of dependence structure are available, and also theory and inference are comparatively well established since a long time (Aas et al., 2009).

Theoretically a random field can be viewed as an infinite-dimensional multivariate distribution. However, it is often treated in discretized form as n dimensional distribution with n very large. Therefore the construction of arbitrary n dimensional distribution is essential to characterize the joint distribution underlying the random fields. For this purpose, this study presents a vine-copula based multivariate random field framework. This approach utilizes only bi-variate copula to build multidimensional distributions. Developed originally by the works of Bedford and Cooke (2002); Aas et al. (2009), vine-copula approach has been utilized in the context of spatial dependence modelling of environmental parameters (Gräler and Pebesma, 2011; Gräler, 2014, e.g.). Non-vine copula approach, i.e. using Gaussian copula and its transforms is popular in hydrology literature for groundwater quality and solute transport parameters (Bárdossy, 2006; Bárdossy and Li, 2008; Haslauer et al., 2012) .

In geotechnical engineering, vine copula theory has only recently been used in the spatial or cross dependence context. Wang and Li (2017) using the vine copula theory demonstrated the impact of cross-dependence structure on the failure probability of a tunnel excavation. Wang and Li (2018) emphasized the importance of vine copula in modelling spatial dependence in random fields. Wang and Li (2019) utilized vine copula theory to model the non-Gaussian dependence structure of cone penetration data. Very recently, Lü et al. (2020) demonstrated the cross-dependence modelling of soil properties using a vine copula approach. All these studies have utilized vine-copula either for modelling spatial dependence (Wang and Li, 2018, 2019) or cross-dependence (Wang and Li, 2017; Lü et al., 2020) only. This study tackles both spatial and cross dependence modelling in a common vine copula framework. Also, all aforementioned studies focused only on stress-strain/strength parameters only, such as cohesion, friction angle and poison ratio. To the best of author's knowledge, there are no studies accounting for the non-Gaussian spatial as well as cross-dependence structure of hydraulic parameters. Therefore, in the context of stochastic seepage and slope stability analysis, this study (Chapter 6) presents a vine copula based approach to model both spatial and cross dependence structure among the random fields of hydraulic parameters; hydraulic conductivity and van Genuchten (1980) soil water characteristic parameters α and n.

2.5 Gaps in research and Motivation

In the view of aforementioned discussion, the gaps in research and the motivation behind it are summarized as follows:

1. **Probabilistic characterization of SWCC of bentonite:** Among the minimal studies for probabilistic characterization of SWCC almost all of them are limited to soils where clay fraction is low. However, with an increasing use of clay in various geo-environmental projects, characterization of SWCC of clays is also necessary. This is addressed in Chapter 3.

2. **Probabilistic characterization of SWCC from limited data:** Most of the studies for probabilistic characterization of SWCC are focussed for data sufficient/abundant cases. However the experimental determination of SWCC is not a routine geotechnical test, therefore case of limited data is significant for site-specific unsaturated RBD. This is addressed in Chapter 4. Chapter 4 also leverages the results of Chapter 3, by extending its basic philosophy towards addressing probabilistic characterization from limited data. The global distribution estimates from Chapter 3 are also utilized to demonstrate the applicability of the proposed approach.

3. **Probabilistic estimation of CEC and SSA:** Determination of CEC and SSA is challenging due to inherent uncertainties and difficulty in experimental measurement. Popular approach is to employ transformation models for its estimation. However, most of the existing models were developed on limited sample sizes, and quantification of uncertainty associated with the estimate is not possible. This problem of multivariate statistical characterization of index

properties of clay is addressed in Chapter 5, where the number of random variables is large. In continuation from the previous two chapters, this chapter presents a framework for random variables in higher dimensions. The high dimensional distribution can be efficiently built using lower dimensional conditional distributions that are amenable to engineering interpretations as well as practical applications.

4. **Stochastic seepage analysis considering non-Gaussian spatial and cross dependence structure of hydraulic parameters:** Conventionally the stochastic seepage analysis is performed using a random field framework for hydraulic parameters e.g. k and SWCC parameters. However, almost all the works till date are limited to Gaussian spatial and cross-dependence structure. This study (Chapter 6) builds on the framework presented in Chapter 5 and presents a more generalized approach of multivariate statistical characterization in the form of random fields. The presented approach in Chapter 6 can handle both spatial and cross dependence structure in a unified framework.

Chapter 3

Probabilistic characterization of soil-water characteristic curve of bentonite

3.1 General

In this study, a database is developed exclusively for the SWCC of bentonites by compiling measured S and ψ from the literature (ref. Table 3.1). Initially, the impact of various commonly used constraints in van Genuchten (1980) equation on the statistics of obtained parameters is investigated based on the developed database. It is found that the parameters statistics vary widely with the choice of constraints. Also, the degree of correlation among the parameters, which is essential in the context of multivariate probabilistic analysis is found to vary considerably. Further, the performance evaluation of these constraints is done in terms of accuracy (Root mean squared error(RMSE)), linearity (Coefficient of determination R^2) and information criterion (Akaike information criterion (AIC) and Bayesian information criterion (BIC)). It is found that the three-parameter vG model with constraint S_r=0 and m as an independent parameter is most suited for bentonites. Using the best fit model, the scatter in the compiled database of SWCCs is reduced to a set of curve fitting parameters α, n, m. Thereafter, trivariate distribution of the same is evolved using copula theory. Two elliptical copulas, Gaussian and 't' are explored, and the results show that elliptical copulas can be successfully used for modelling the trivariate distribution of vG parameters. For further applications, this study has developed confidence intervals for bentonite SWCC by using the evolved trivariate distribution.

The contributions of this study are threefold and can be summarized as: (1) a database exclusively for measured SWCC of bentonite is compiled from the literature, (2) performance evaluation of commonly used constraints in vG model and its impact on SWCC parameter statistics of bentonite is demonstrated, (3) trivariate distribution of vG parameters (α, n, m), is constructed using copula theory and confidence intervals are created for bentonite SWCC. These are key contributions as far as unsaturated behavior of bentonites is concerned, as there are a very few multivariate models of SWCC parameters in literature. The developed trivariate distribution is a valuable firsthand statistical estimate for the bentonite SWCCs in the absence of measured data. The same is also useful as an informative prior for updating the limited site-specific data using Bayesian approach.

Table 3.1: Details of the compiled database for bentonite SWCC.

S.I. No.	Ref./Cross Ref.*	Bentonite#	No.	Remarks
1	Zhu et al. (2016)	GMZ	4	At 20,40,60 and 80° using filter paper method
2	CHEN et al. (2006)/Zhao et al. (2016)	GMZ	1	—-
3	Lloret et al. (2005)/Villar et al. (2008)	FEBEX	9	At various dry densities 1.55-1.75 g/cm^3
4	Villar (2002)/Dieudonne et al. (2017)	FEBEX	2	At dry densities 1.05 and 1.36 g/cm^3
5	Villar (2002)/Villar (2005)	FEBEX	3	At dry densities 1.5,1.6,1.7 g/cm^3
6	Alonso et al. (2011)	FEBEX	3	At dry densities 1.30,1.50,1.90 g/cm^3
7	Dai et al. (2008)	FEBEX	1	—-
8	Villar and Lloret (2004)	FEBEX	3	At 20,40,60 and 80° for dry density of 1.65 g/cm^3
9	Tripathy et al. (2014)	MX-80	1	Using pressure plate and dessicator test
10	Tripathy et al. (2015)	MX-80	5	At various dry densities 1.29-1.62 g/cm^3
11	Seiphoori et al. (2014)	MX-80	3	At dry densities 1.50,1.65,1.79 g/cm^3
12	Hökmark (2004)	MX-80	1	—-
13	Rizzi et al. (2011)/Ballarini et al. (2017)	MX-80	1	—-
14	Villar (2005)	MX-80	12	At various dry densities and temperature
15	Ravi and Rao (2013)	BARMER	10	At various dry densities and salinity of pore water
16	Baille et al. (2014)	German	1	—-
		Total	60	

* In case original reference was not accessible, data was collected from cross reference. # GMZ = Gaomiaozi (China) FEBEX= Full-scale Engineered Barriers Experiment (Spain), MX- 80= Wyoming (USA), BARMER (India).

3.2 Database

This study compiles an extensive database of SWCC of bentonites from literature. There are various forms in which the SWCC data can be presented such as gravimetric water content w vs. suction ψ, volumetric water content θ_w vs. suction ψ, degree of saturation S vs. suction ψ. For a non-expansive class of materials such as non-smectite clays, sands, and silts various forms are readily inter-convertible and can be regarded as equivalent to each other. However the same for bentonites is not true. For a highly deformable soil like bentonite, S vs ψ is widely accepted as the most informative form of representation for bentonite SWCC (Mbonimpa et al., 2006) since parameters such as air entry value (AEV) and residual degree of saturation are better identified from S vs ψ curve rather than θ_w vs ψ, or w vs ψ curves. Hence, only the SWCCs reported in the form of S vs ψ were collected. Total number SWCCs compiled from literature is 60 which corresponds to a total of 565 No. of data points with an average of 9.4 data points(S,ψ) per SWCC. The SWCCs in the database correspond to bentonites from different geographical locations and the experimental procedure comprised of various initial densities, temperature, and salinity of the water. The samples were also subjected to different paths such as wetting or drying. As already discussed, all of these factors contribute to the non-uniqueness of SWCC. Quantifying the uncertainties based on each of the above factors could not be performed in this study as the number of SWCCs in the database was insufficient for a satisfactory statistical estimation and inference. Details of the compiled database with its sources are summarized in Table 3.1.

3.3 Probabilistic Framework

3.3.1 Evaluation of various constraints among the SWCC model

SWCC expresses the relationship between two primary state variables, water content (w/θ_w/S) and suction (ψ). There are various empirical or semi-empirical models reported in the literature (Brooks and Corey, 1964; van Genuchten, 1980; Fredlund and Xing, 1994) for quantifying a continuous SWCC using the measured suction and water content data. Among them all, the one proposed by van Genuchten (1980) remains the most popular and is widely used in geotechnical and geoenviromental literature (Carsel and Parrish, 1988; Phoon et al., 2010; Chiu et al., 2012, e.g.,). The vG model in terms of degree of saturation is given as:

$$S_e = \frac{S - S_r}{1 - S_r} = \frac{1}{[1 + (\alpha\psi)^n]^m} \tag{3.1}$$

Where S_e= effective degree of saturation, S_r= residual degree of saturation, ψ= suction head, α= parameter related to inverse of air entry value, n= parameter related to rate of desaturation and m is a symmetry parameter very often represented in following forms $m = 1 - 1/n$, $m = 1 - 2/n$, $m = 1$ after Mualem (1976), Burdine (1953) and Gardner (1958) respectively. Along with independent m, a total of 4 cases are possible. The constraints mentioned above regarding m are often applied to reduce the number of unknown parameters. Another common constraint very often applied is S_r=0. However, these constraints arises mainly from the lack of knowledge and confidence regarding the definition and inference of residual degree of saturation (Likos and Yao (2014)). Combining the constraint S_r=0 and S_r retained as a fitting parameter, with 4 cases already mentioned, a total of 8 constraints are possible. In this study hereafter these 8 constraints will be referred to as 8 cases and are illustrated in Table 3.2. These 8 cases were fitted to the SWCC(S,ψ) data using the nlinfit subroutine in Matlab. This subroutine utilizes the Levenberg-Marquardt algorithm for optimization. More details about the algorithm can be found elsewhere (Seber and Wild, 2003, e.g.,).

As already discussed, in most of the studies, the constraints are selected arbitrarily without any proper statistical evaluation for the particular class of soil being considered. However, it is essential to evaluate the impact of various constraints on the SWCC parameters due to following reasons (1) since the measurement of SWCC of bentonite is way too complicated and time-consuming, a variety of pedotransfer (PTF) approaches are common to estimate the SWCC from more easily measured soil properties. The approach used is to develop correlations among measured soil properties and parameters of vG model. However, often the SWCC constraints applied to obtain the parameters are not specified or are unknown (Likos and Yao, 2014). (2) Each of the constraints although applied arbitrarily, imposes some restriction on the curve fitting capability of the model. While some constraints can be suitable for sandy soil, it may not be suitable for a clayey material like bentonite. (3) The degree of correlation among the SWCC parameters, which is an important aspect for reliability studies can vary widely with the choice of constraints (ref. Table 3.3). (4)

Key properties being estimated such as AEV, which is closely related to the inverse of α value can be varying in the order of MPa due to the choice of constraints. An ideal analysis would entail the evaluation of various available SWCC equations other than vG, but for the sake of brevity, only vG equation has been evaluated in this study. The aim of this study is not only to evaluate and compare various constraints but thereafter also develop a joint probability distribution of the parameters using the best fit constraint from the aforementioned study.

There are various approaches available in the literature for performance evaluation or model selection such as Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Schwarz et al., 1978), and Deviance Information Criterion (DIC) (Van Der Linde, 2005). In this study, AIC and BIC are used. They are defined as follows:

$$BIC = Nln[\frac{\sum_{j=1}^{N}(y_j^{obs} - y_j^{fit}(b))^2}{N}] + kln(N) \tag{3.2}$$

$$AIC = Nln[\frac{\sum_{j=1}^{N}(y_j^{obs} - y_j^{fit}(b))^2}{N}] + 2k \tag{3.3}$$

Where N= sample size i.e. no of data points, k= no of free parameters in the model, Y_j^{obs} is observed S (measured S in database) and y_j^{fit} (b) is the S obtained using Eq. 3.1 respectively at ψ_j (measured ψ in database), $\sum_{j=1}^{N}(y_j^{obs} - y_j^{fit}(b))^2$ = residual sum of squares (RSS) and b is the parameters vector containing the k parameters that need to be estimated. The best fit model is selected on the basis of minimum BIC and AIC criterion. It can be noted that the AIC and BIC not only take into account the minimization of the residuals (RSS) but also assign a penalty term for the extra number of parameters. This is beneficial since a model with a large number of parameters is more likely to accurately fit the data (e.g., in terms of RSS). But this is not always desirable as it can lead to overfitting and multiple local optima during the optimization (curve fitting) process. For models with a similar degree of fit in terms of RSS the penalty terms $kln(N)$ and $2k$ in Eqs. 3.2 and 3.3 ensure a larger BIC and AIC value for the model with a larger number of parameters. Since the criterion for selection is minimum BIC and AIC, the model with extra parameters but similar RSS (compared to alternate models) gets eliminated. The procedure adopted in this study for performance evaluation is summarized using a flowchart in Fig. 3.1. For further comparison of goodness of fit for various models, RMSE and R^2 given in Eq. 3.4 and Eq. 3.5 respectively are also evaluated.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N}(y_j^{obs} - y_j^{fit}(b))^2}{N}} \tag{3.4}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{N}(y_j^{obs} - y_j^{fit}(b))^2}{\sum_{j=1}^{N}(y_j^{obs} - (\frac{1}{N}\sum_{k=1}^{N} y_k^{obs})^2} \tag{3.5}$$

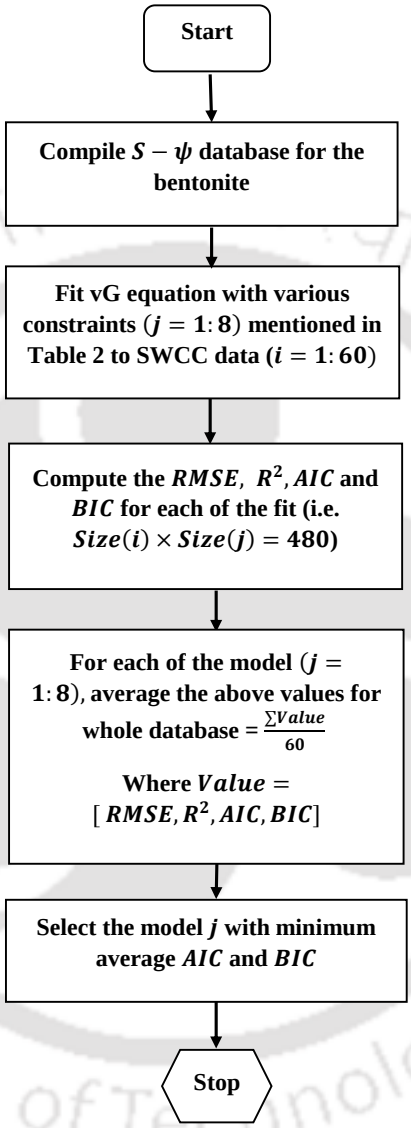


Figure 3.1: Procedure used for identification of best fit van Genuchten (1980) model among various constraints mentioned in Table 3.2.

3.3.2 Copula approach

Estimation and inference of multivariate normal data exhibiting linear dependence are well established, but general approaches for modelling non-normal data exhibiting non-linear dependence structure are not so well developed, especially when the number of random variables involved is more than 2. Copula theory provides an alternate approach for modelling multivariate non-normal data with nonlinear dependence structure. Although the theoretical and mathematical background underlying copulas are complex, the estimation and practical implementation are relatively simpler. This section presents a brief background of copula theory with respect to trivariate distribution functions. For a more general and comprehensive overview about copulas, one can refer to literature (Trivedi et al., 2007; Nelsen, 2007; Genest and Favre, 2007). The copula is defined as a multivariate distribution with uniform one-dimensional marginals. Sklar (1959) theorem forms the foundation for application of copula theory. It provides for the marginals to be linked to the respective joint distribution functions. It states that every multivariate cumulative distribution function (CDF) of a random vector can be expressed in terms of its marginals and a copula. This adds to the uniqueness of copula when the marginals are continuous. When applied for the trivariate case of vG parameters it can be expressed as:

$$H(\alpha, n, m) = C(F_1(\alpha), F_2(n), F_3(m); \theta) \tag{3.6}$$

Where $H(\alpha, n, m)$ is the trivariate joint CDF, $F_1\alpha$),$F_2(n)$, $F_3(m)$ are the one dimensional marginal CDFs and C is the copula with dependence parameter matrix θ. Since $F_i(.)$ is defined over [0, 1], it is easy to observe that a copula function C is a multivariate distribution with uniform marginals. For the trivariate case, this means that a copula is defined over a unit cube $[0, 1]^3$. This also implies that a copula can be rewritten as joint cumulative CDF of U_1, U_2, U_3 as:

$$\begin{aligned} C(F_1(\alpha), F_2(n), F_3(m); \theta) &= C(u_1, u_2, u_3) \\ &= P(U_1 \leq u_1, U_2 \leq u_2, U_3 \leq u_3) \end{aligned} \tag{3.7}$$

Where $u_i = F_i()$ and $P(.)$ represents the joint probability. The CDF transformation ensures that copulas remain invariant, under strictly increasing transformations applied to associated random variables (Schweizer et al., 1981). Another advantage that the copula approach offers is that marginal distributions F_i do not have to be the same for all the dimensions. This is a significant advantage and one of the main reasons for its popularity since this allows for a more general approach which can incorporate distribution from various families. The trivariate probability density function (PDF) can be obtained by taking the derivative of the joint CDF i.e. Eq. 3.6 as:

$$f(\alpha, n, m) = f_1(\alpha) f_2(n) f_3(m) D(u_1, u_2, u_3; \theta) \tag{3.8}$$

where $f(\alpha, n, m)$ is the joint PDF, $f_1(\alpha)$,$f_2(n)$,$f_3(m)$ are the one dimensional marginal PDFs and $D(u_1, u_2, u_3; \theta) = \partial^3 C(u_1, u_2, u_3; \theta)/\partial u_1 \partial u_2 \partial u_3$ is the copula density function.

It should be noted that modelling the dependence for a bivariate case is relatively straightforward, since the copula is characterized by only one parameter in the dependence matrix, θ. However, for modelling the dependence for $n \geq 3$, the number of dependence parameters depend upon the choice of family, or, the type of copulas. For example, for the trivariate case ($n = 3$), a symmetric Archimedean copula is characterized by one dependence parameter, an asymmetric Archimedean copula requires 2 parameters, an elliptical copula (e.g., Gaussian) needs 3 and t copula needs 4 parameters. It might be appealing to choose the Archimedean copulas with a small number of parameters, but it should be noted that there are various limitations related to their application such as: (a) in the symmetric Archimedean copula, each bivariate pair is modelled using the same parameter that means the dependence structure among each bivariate pair (1, 2), (1, 3) and (2, 3) should be equal. As observed in this study, this is not the case for the vG parameters α, n, m (b) for the asymmetric Archimedean case, variables (1, 3) and (2, 3) are modelled using the same parameter. This means that the dependence structure should be the same for at least two random variables, which again does not concur with the present study. Various researchers have pointed out the above-mentioned limitations regarding many of the Archimedean copula (Kao and Govindaraju, 2008; Zhang and Singh, 2014). In addition, researchers have also suggested that for the cases of asymmetric dependence as observed in this study, elliptical copulas might be more efficient (Genest et al., 2007; Song and Singh, 2010). There are various other techniques to study the asymmetric dependence in higher dimensions such as Vine copula (Czado, 2010; Joe and Kurowicka, 2011; Zhang and Singh, 2014) but as far as the simplicity and ease of inference is considered, elliptical copulas are more straightforward to generalize in higher dimension hence the same will be used in this study. In the following section, the elliptical copula i.e., Gaussian and t-copula used in this study are presented.

Gaussian copula The copula function for an n-dimensional Gaussian copula also known as a normal copula is expressed as:

$$C(u_i \dots u_n; \theta) = \Phi_\theta^n(\Phi^{-1}(u_i). \dots \Phi^{-1}(u_n)) \tag{3.9}$$

where $C(u_i \dots u_n; \theta)$ is the copula function with parameter matrix θ, $\Phi_\theta(.)$ denotes the multivariate standard normal function with linear correlation matrix θ, and Φ^{-1} denotes the quantile or inverse CDF of one-dimensional standard normal distribution. For the trivariate case i.e., $n = 3$, it can be expressed as:

$$\begin{aligned} C(u_1, u_2, u_3; \theta) = & \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_3)} \\ & \times \frac{1}{(2\pi)^{3/2}|\theta|^{1/2}} exp(-\frac{1}{2} W^T \theta^{-1} W) dW \end{aligned} \tag{3.10}$$

where $W = [t_1, t_2, t_3]^T$ and t_1, t_2, t_3 are the corresponding integral variables. The copula density

function $D(u_1, u_2, u_3; \theta) = \partial^3 C(u_1, u_2, u_3; \theta)/\partial u_1 \partial u_2 \partial u_3$ can be expressed as:

$$D(u_1, u_2, u_3; \theta) = \frac{1}{\theta^{1/2}} exp[-\frac{1}{2}\zeta^T \theta^{-1} \zeta - \zeta^T \zeta)] \tag{3.11}$$

where $\zeta = [\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)]^T$.

t copula The copula function for an n-dimensional t-copula also known as student t copula can be written as follow:

$$C(u_i \dots u_n; \theta, \nu) = t^n_{\nu,\theta}(t^{-1}_\nu(u_i), \dots., t^{-1}_\nu(u_n)) \tag{3.12}$$

where $t^n_{v,\theta}$ represents the distribution function of $\sqrt{\nu}Y/\sqrt{S}$, where S χ^2_ν and Y $N_n(0, \theta)$ are independent and θ is simply the linear correlation matrix for a degree of freedom $\nu \geq 2$. For the trivariate case i.e., $n = 3$ it can be expressed as:

$$\begin{aligned} C(u_1, u_2, u_3; \theta, \nu) = & \int_{-\infty}^{(t_v)^{-1}(u_1)} \int_{-\infty}^{(t_v)^{-1}(u_2)} \int_{-\infty}^{(t_v)^{-1}(u_3)} \\ & \times \frac{\Gamma\frac{\nu+3}{2}}{\Gamma\frac{\nu}{2}} \frac{1}{(\pi\nu)^{3/2}|\theta|^{1/2}} (1 + \frac{W^T \theta^{-1} W}{\nu})^{-\frac{\nu+3}{2}} dW \end{aligned} \tag{3.13}$$

where $t_v^{-1}(.)$ denotes the inverse of student t distribution, $\Gamma(.)$ denotes the gamma function. The copula density function $D(u_1, u_2, u_3; \theta, \nu) = \partial^3 C(u_1, u_2, u_3; \theta)/\partial u_1 \partial u_2 \partial u_3$ can be expressed as:

$$\begin{aligned} D(u_1, u_2, u_3; \theta, \nu) = & |\theta|^{-\frac{1}{2}} \frac{\Gamma\frac{\nu+3}{2}}{\Gamma\frac{\nu}{2}} [\frac{\Gamma\frac{\nu}{2}}{\Gamma\frac{\nu+1}{2}}]^3 \\ & \times \frac{(1 + \frac{\zeta^T \theta^{-1} \zeta}{\nu})^{-\frac{\nu+3}{2}}}{\prod_{i=1}^3 (1 + \frac{[t_\nu^{-1}(b_i)]^2}{\nu})^{\frac{\nu+1}{2}}} \end{aligned} \tag{3.14}$$

where $b_i = t_\nu^{-1}(u_i)$ and $\zeta = [t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), t_\nu^{-1}(u_3)]^T$.

Estimation

There are two commonly used approaches for estimation of copula parameters, parametric and semi-parametric. Both approaches differ in the assumption of parametric or non-parametric marginals respectively. Since semiparametric approach allows for the separation of dependence and marginal modelling the same is used in this study and discussed as follow. The most popular choice for estimation of the copula is using the pseudo likelihood method (Zhang and Singh, 2007; Genest et al., 1995; Shih and Louis, 1995; Fang et al., 2002). It essentially consists of maximizing the log-pseudo likelihood function $l(\theta)$. For a trivariate case, it can be expressed as

$$\theta_n = \underset{\theta}{\text{argmax}}\, l(\theta) = \underset{\theta}{\text{argmax}} \sum_{i=1}^{n} log[D_\theta(U_i 1, U_i 2, U_i 3)] \tag{3.15}$$

Where $D_\theta()$ denotes the copula density function, $U_i j$ are the pseudo observation or rescaled

empirical distribution of $Z_i = (Z_i1, Z_i2, Z_i3)$ and can be expressed as

$$U_ij = \frac{R_ij}{n+1} \tag{3.16}$$

Where R_ij is the rank of Z_ij among (Z_1j, Z_2j, Z_3j) when sorted in ascending order. For Gaussian and t copula Eq. 3.15 can be solved by minimizing the partial derivative of log-pseudo likelihood function as given in Eq. 3.17 and Eq. 3.18 respectively.

$$\sum_{i=1}^{n} \frac{\partial log[D_\theta^G((U_i1, U_i2, U_i3))]}{\partial \theta} = 0 \tag{3.17}$$

Where $D_\theta^G()$ denotes the copula density function for Gaussian copula provided in Eq. 3.11.

$$\begin{aligned} \sum_{i=1}^{n} \frac{\partial log[D_\theta^t((U_i1, U_i2, U_i3))]}{\partial \theta} &= 0 \\ \sum_{i=1}^{n} \frac{\partial log[D_\theta^t((U_i1, U_i2, U_i3))]}{\partial \nu} &= 0 \end{aligned} \tag{3.18}$$

Where $D_\theta^t()$ denotes the copula density function for t copula provided in Eq. 3.14.

Goodness of fit

For evaluation of goodness of fit among the above two copulas, the Cramer-von Mises statistics (CM) (Genest et al., 2009) and AIC (Fang et al., 2014) are used in this study and are discussed below.

Cramer-von Mises statistics The CM statistics (S_n) represents the distance between true and observed copula and is based on a comparison of empirical copula with the derived copula under the null hypothesis. The same is given by

$$S_n = \int_{[0,1]^p} C_n^2(u) dC_n(u) = \sum_{i=1}^{n} (C_n(U_i) - C_{\theta_n}(U_i))^2 \tag{3.19}$$

Where $C_n(u) = \sqrt{n} C_n(u) - C_(\theta_n)(u)$ is the empirical copula and $C_(\theta_n)$ is the copula under hypothesis and θ_n is the estimator of θ, computed using the ranked pseudo observations. A large value of S_n indicates rejection of copula under the null hypothesis. Approximate p value for the test statistic S_n can be computed using a parametric bootstrap simulation approach. p value represents the degree by which the null hypothesis is not rejected, hence a large value of p results in a better fit. According to Genest et al. (2007), the procedure can be summarized as follows:

1. • For a large integer value (e.g., $N = 10000$) repeat the following steps for $j = 1 : N$. Generate $X_j^* \in [0,1]^3 = [X_(1,j)^*, X_(2,j)^*, X_(3,j)^*]$ from the fitted copula $C_(\theta_n)$ under

the null hypothesis

- Calculate the associated rank R_ij^* for X_ij^* and $U_ij^* = R_ij^*/(n+1)$
- Estimate the copula parameter matrix θ_n^* for U_ij^* by solving 3.15
- Calculate the S_n^* statistics for using the copula parameter θ_n^* using Eq. 3.19

2. Approximate p value for the test statistics S_n can be obtained by

$$p = \frac{1}{N}\sum_{i=1}^{N} I(S_n^* > S_n) \tag{3.20}$$

AIC In this section AIC approach for copula selection after Fang et al. (2014) is discussed. Akaike information criterion is defined as

$$AIC = -2l(\theta_n) + 2(k) \tag{3.21}$$

Where $l(\theta_n)$ is the maximized log likelihood and k is the number of free parameters to be estimated. AIC, when applied to copula model, is a relative measure of goodness of fit of the model. Although it cannot perform formal goodness of fit hypothesis test, it can be used to select the better copula among the families chosen for study. Lehmann and Casella (2006) suggested using $K-L$ (Kullback and Leibler, 1951) information as a measure to represent the distance between true and hypothesized copula. Assuming the density of true copula and the estimated one under null hypothesis are given by $D(U)$ and $D_\theta(U)$ respectively, the $K-L$ information can be represented as

$$\begin{aligned} K(D(U), D_\theta(U)) &= \int log(\frac{D(U)}{D_\theta(U)})D(U)dU \\ &= E[logD(U)] - E[logD_\theta(U)] \end{aligned} \tag{3.22}$$

Where $E(.)$ denotes the expectation or expected value. Smaller the $K-L$ information, closer is the hypothesized copula to the true one. Alternatively, larger the value of $E[logD_\theta(U)]$, closer is the hypothesized copula to the true one. Bozdogan (1987) suggested that AIC is an unbiased estimator of $-2E[logD_\theta(U)]$ i.e. $AIC = -2E[logD_\theta(U)]$. Substituting this into Eq. 3.22 results in

$$K(D(U), D_\theta(U)) = E[logD(U)] + 1/2AIC \tag{3.23}$$

From the above equation, it can be observed that minimizing the AIC results in minimizing the K-L information. Hence the copula with minimum AIC can be selected as the favored model.

The framework used in this study for the construction of multivariate distribution using copula is summarized in Fig. 3.2. It can be observed from Fig. 3.2 that apart from copula estimation and selection, one crucial step that is independent of the copula theory is the estimation of appropriate one-dimensional marginals. For this purpose, four distributions namely Lognormal, Weibull,

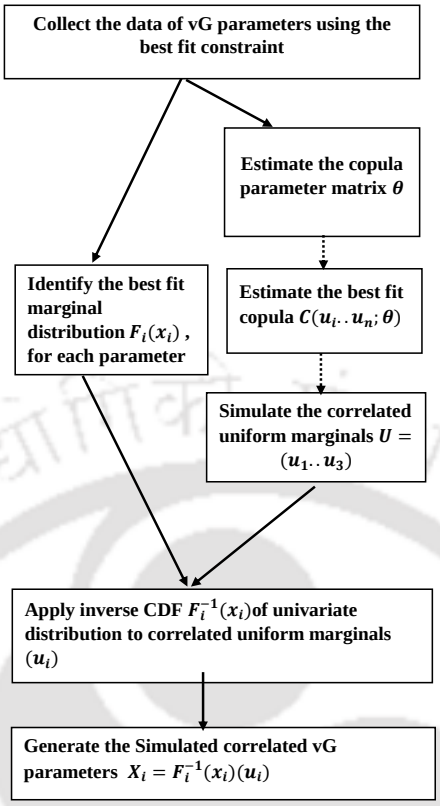


Figure 3.2: Procedure used for establishing the joint distribution of van Genuchten (1980) parameters using copula theory.

Gamma and Exponential were explored. Since the van Genuchten (1980) parameters are theoretically constrained from below by zero, these four distributions are chosen as they are applicable over positive values only. Also since the van Genuchten (1980) parameters to be modelled are close to zero, it is necessary to avoid negative values in the tail of the distribution during simulation. AIC defined in Eq. 3.21 was used for selection of best model among the four. Since the theory and inference regarding the aforementioned univariate distributions is already well established in the literature (Ang and Tang, 2007, e.g.,) the same is not discussed here. For simulation of copula random numbers, algorithms proposed in Embrechts et al. (2001) were used.

Use of copula in geotechnical engineering

Copula theory is used extensively since a long time for modelling dependence structure among random variable in the other areas such as economics (Patton, 2003; Trivedi et al., 2007; Patton, 2009; Aloui et al., 2013) and hydrology(Favre et al., 2004; Zhang and Singh, 2007; Genest et al., 2007; Salvadori and De Michele, 2007; Song and Singh, 2010). However, only recently it has gained popularity in geotechnical engineering. Tang et al. (2013) investigated the impact of copula selection for the joint distribution of cohesion and friction on reliability analysis and slope. Wu (2013*a*)

used copula-based sampling probabilistic slope stability analysis. Wu (2013*b*) used Gaussian and nested Frank copula for modelling trivariate distribution of three soil properties; cohesion, friction and unit weight. Li et al. (2013) used the Gaussian copula for modelling the dependence among the parameters of load-displacement curves of piles and recommended copula-based approach for modelling other correlated geotechnical parameters. Li et al. (2015) used four different copulas to construct the joint distribution of cohesion and friction angle. Tang et al. (2015) used copula-based approach for reliability analysis of slopes under incomplete data information. However, there are not many studies using copula for probabilistic modelling in unsaturated geomaterials. Prakash et al. (2017) successfully used copula theory to model the dependence structure between θ and ψ and demonstrated the applicability of the same to obtain the water retention curve of fly ash under restricted measurement options. Raj Singh et al. (2018) applied the copula approach to address uncertainties associated with unsaturated permeability in vegetated slopes. Prakash et al. (2018*b*) used copula theory to construct the bivariate distribution of vG parameters α and n for quantifying instrumental and material based uncertainty in fly ash. More studies regarding the use of copula theory in unsaturated soil mechanics can found in the literature (Prakash et al., 2018*a*, 2019, e.g.,). It can be noted that most of these studies are in the context of bivariate data. Very few of the studies have utilized copula theory beyond bi-dimensional data. (e.g., Wu (2013*b*)). In this study two elliptical copulas, Gaussian and t are used, and it will be demonstrated that both of them are suitable for construction of multivariate distribution of bentonite SWCC.

3.4 Results and Discussion

3.4.1 Effects of constraints on vG parameters for modelling SWCC of bentonites

In this section, the effects of various constraints imposed on the parameters of van Genuchten (1980) equation towards the modelling of SWCC of bentonites are systematically discussed. The details of constraints imposed on vG equation parameters are already discussed in the probabilistic framework section. Likos and Yao (2014) evaluated the effects of constraints on the parameters of vG equations. However, their database did not account for highly plastic clays like bentonite. This work extends the philosophy discussed in Likos and Yao (2014) for studying the constraints mentioned above based on a database formulated exclusively for bentonites. For the sake of clarity and ease of comparison, the aforementioned constraints will be herein referred to as cases listed in Table 3.2. For example, case 1 in Table 3.2 refers to the vG model with no constraints and case 2 refers to the vG model with constraint S_r=0. The cases corresponding to various constraints can be read from the first row of Table 3.2. Each of the 60 SWCC in the database was fitted to 8 cases, and the statistics for the parameters were obtained for each case. Table 3.2 presents the mean, median, standard deviation and range of vG parameters obtained for the database, using various models considered in this study.

Table 3.2: Comparison of van Genuchten (1980) parameter statistics obtained under various constraints.

Case		1	2	3	4	5	6	7	8
Parameters		S_r, α, n, m	α, n, m	S_r, α, n	α, n	S_r, α, n	α, n	S_r, α, n	α, n
Constraint		None	$S_r = 0$	$m = 1 - \frac{1}{n}$	$S_r = 0$, $m = 1 - \frac{1}{n}$	$m = 1 - \frac{2}{n}$	$S_r = 0$, $m = 1 - \frac{2}{n}$	$m = 1$	$S_r = 0$, $m = 1$
S_r	Mean	0.0184	–	0.0774	–	0.0429	–	0.1385	–
	Median	0	–	0	–	0	–	0.1441	–
	SD	0.0597	–	0.1309	–	0.1074	–	0.1416	–
	Range	0-0.31	–	0-0.63	–	0-0.59	–	0-0.64	–
α	Mean	0.4815	0.4836	0.5654	0.5549	0.7055	0.7085	0.0352	0.0268
	Median	0.0376	0.0388	0.0644	0.0631	0.0791	0.1003	0.0210	0.0179
	SD	2.3104	2.3100	2.7881	2.7836	3.4094	3.4080	0.0361	0.0299
	Range	0.0028-17.89	0.0048-17.89	0.0072-21.57	0.0070-21.56	0.0081-26.43	0.0093-26.43	0.0036-0.16	0.0036-0.163
n	Mean	3.4895	3.4482	1.9131	1.6924	2.6736	2.5496	1.4925	1.1601
	Median	1.7916	1.7962	1.7943	1.6606	2.5448	2.5105	1.3915	1.1750
	SD	5.5387	4.9601	0.7270	0.3796	0.5432	0.3025	0.8696	0.4681
	Range	0.54-33.52	0.73-26.53	1.09-4.48	1.09-3.05	2.08-4.76	2.0838-3.65	0.271-4.36	0.2715-2.67
m	Mean	0.583	0.4524	–	–	–	–	–	–
	Median	0.294	0.2740	–	–	–	–	–	–
	SD	0.72	0.4764	–	–	–	–	–	–
	Range	0.0074-3.77	0.0200-2.02	–	–	–	–	–	–

Impact on parameter α

It can be observed from Table 3.2 that on comparison within cases (1 and 2), (3 and 4), (5 and 6), (7 and 8) the statistics for α parameter are quite similar. The mean, median, standard deviation and range for α in case (1 and 2) as well as in cases (3 and 4), (5 and 6) and (7 and 8) are almost similar. This observation suggests that the choice of S_r as a fitting parameter does not affect the α parameter significantly within a particular model. This is in contrast to the results reported by Likos and Yao (2014) wherein the authors concluded that α obtained under no constraints (i.e., case 1) was on an average 25% lower than case (2) with constraint $S_r = 0$. However, their results were based on the database of predominantly sandy soil in contrast to only bentonite database used in this study. This may be attributed to a higher range of α observed for bentonites (0.003-17.89 MPa^{-1}) as compared to a rather narrow range (0.09-2.41 kPa^{-1}) observed for sandy soils in Likos and Yao (2014). Also, their conclusion was based on comparison among one model (Model 1) only. In this study, the same conclusion applies equally to various other cases (i.e. (3 and 4), (5 and 6) and (7 and 8)).

On further comparison among cases (i.e., 1, 3, 5 and 7) it can be observed that the corresponding statistics for α differ widely. When compared with the model under no constraint (i.e., case 1), the constraint m=1-1/n (case 3) results in an overestimation of mean α parameter by 17 % and the

constraint m=1-2/n results in an overestimation of mean α by 48 % . This can be explained from the range observed for α in Table 3.2. It can be observed that both the minimum and maximum observed values of α for model 1 to 3 are in order(1) < (3) < (5). It can also be noted that the increment in maximum value is more significant than the increment in minimum value, hence the observed increase in mean, median as well as standard deviation. This observation is also in contrast to the results reported by Likos and Yao (2014). They concluded that when compared to the α obtained under no constraints (i.e., case 1), the one obtained with constraint m=1-1/n (i.e., case 3) is on an average 30 % lower i.e., $\alpha_1 > \alpha_3$. Again, a possible explanation could be the wide range of α observed in this study for bentonites. For case 7 the mean, median and standard deviation of α are minimum among all the models. The mean α, in this case, is 94 % lower than that of case 1. It is widely regarded that the parameter α is related to the inverse of AEV (Gardner, 1958). Although not directly related to AEV, the mean α for case 1 can be approximated to an AEV of 1/0.48 i.e., 2.06 MPa, the same for cases 3, 5, and 7 are 1.79 MPa, 1.4 Mpa and 33 MPa, respectively. This implies that in comparison to the model under no constraints i.e., case 1, commonly used constraint m=1-1/n and m=1-2/n may result in a higher estimate of α or lower estimate of AEV and m=1 constraint may result in a significantly lower estimate of α or gross overestimation of AEV.

Impact on parameter n

Regarding parameter n it can be observed that for the case of unconstrained m i.e., case 1 and 2 the range of n parameter is very high compared to constrained m cases, e.g. cases (3 and 4), (5 and 6) and (7 and 8). While the constrained m forces the parameter to fall in the range of 1-4, the unconstrained cases can exhibit values as high as 26 and 33. A similar observation was reported by Likos and Yao (2014) for sandy soils. While comparing among the cases (1 and 2), (3 and 4), (5 and 6), (7 and 8), it can be observed that retaining S_r as a fitting parameter results in a slightly higher estimate of n as can be confirmed from Table 3.2. It should be noted that similar to the α parameter, a comparison of parameter n across various cases could not be made as the bounds of n are different for each case. Parameter $n > 1$ and $n > 2$ for cases (3, 4) and (5, 6) respectively whereas $n > 0$ for cases 7 and 8. Hence a comparison of parameter statistics for n would not be meaningful and can be misleading. The above observations regarding α and n further accentuate the argument made by Likos and Yao (2014) wherein they stressed that it is important to refer to the constraint being used for the SWCC fitting procedure while using existing or developing new pedotransfer functions (PTFs).

Impact on the residual degree of saturation S_r

Regarding parameter S_r, it can be noted in Table 3.2 that the median value is equal to zero for all the cases except cases 7. It can also be observed that retaining S_r as a curve fitting parameter can result in some unrealistic values for bentonites, such as the maximum value close to 0.6 observed in the model (2), (3) and (4). For a better description of the distribution of S_r, frequency histogram is presented in Fig. 3.3 for various models considered in this study. It can be observed that the

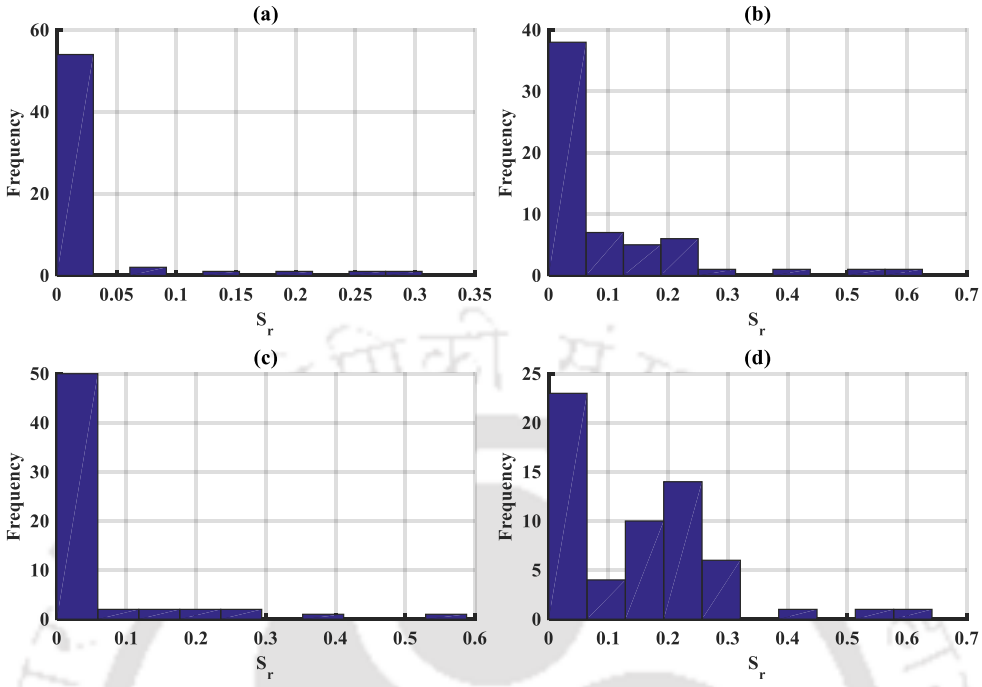


Figure 3.3: Frequency histogram of residual degree of saturation S_r, for various constraints (a) independent$m$ (b) $m = 1 - 1/n$ (c) $m = 1 - 2/n$ (d) $m = 1$.

highest frequency corresponds to S_r=0 for all the 8 cases. Except for cases (7, 8) frequency of S_r=0 exceeds the frequency of nonzero values by a large margin. This observation indicates that even while retaining S_r as a fitting parameter, the optimal value is zero. This suggests that for the case of bentonites, the choice of S_r as a fitting parameter in unnecessary and it is appropriate and logical to use $S_r = 0$. It should be noted that this suggestion is not based only on the aforementioned observation. Additional evidence in favor of S_r=0 is found and discussed in subsequent sections.

Impact on the degree of correlation among α ,n, m, S_r

To undertake any practical uncertainty analysis, it is important to consider an appropriate level of correlation among SWCC parameters (Zhang, Zhang and Tang, 2003; Prakash et al., 2018*a*). Table 3.3 presents the measure of concordance in terms of rank correlation coefficient Kendall's τ for the vG parameters under various cases considered in this study. For the most popular case 4 it can be observed that the correlation coefficient for (α,n) is -0.66. This is consistent with the similar negative correlations reported elsewhere (Carsel and Parrish, 1988; Phoon et al., 2010, e.g.,). On inspection among the cases (1 and 2), (3 and 4), (5 and 6), (7 and 8) it can be observed that the choice of S_r as a fitting parameters does not considerably affect the degree of correlation as the correlation matrix are almost similar except for case (7 and 8). Though it can be noted that the correlation among the parameters (e.g., α,n) for various cases differ considerably. The variation

Table 3.3: Measured rank correlation coefficient Kendall's τ among van Genuchten (1980) parameters for various constraints considered in this study.

Case		S_r	α	n	m
1	S_r	1	-	-	-
	α	-0.33	1	-	-
	n	-0.03	0.09	1	-
	m	0.30	-0.47	-0.51	1
2	S_r				
	α		1	-	-
	n		0.04	1	-
	m		-0.42	-0.56	1
3	S_r	1	-	-	
	α	-0.21	1	-	
	n	0.53	-0.60	1	
	m				
4	S_r				
	α		1	-	
	n		-0.64	1	
	m				
5	S_r	1	-	-	
	α	-0.19	1	-	
	n	0.31	-0.67	1	
	m				
6	S_r				
	α		1	-	
	n		-0.65	1	
	m				
7	S_r	1	-	-	
	α	0.13	1	-	
	n	0.55	-0.06	1	
	m				
8	S_r				
	α		1	-	
	n		-0.29	1	
	m				

ranges from a weak positive correlation of +0.04 for case 2 to high negative value of -0.67 for case 5. This observation implies that the correlation values among the vG parameters should not be considered as universal even for a particular class of soil, but depends on the imposed constraints. It also indicates that the constraints used in vG equation to quantify SWCC should be appropriately identified before referring to the values from literature for the purpose of uncertainty propagation studies.

Table 3.4: Performance evaluation of various constraints in van Genuchten (1980) model.

Case	Fitting parameters	Constraints	Rank	AIC*	BIC*	RMSE	R^2*
1	S_r, α, n, m	None	2	-69.7769	-102.305	0.022266	0.976681
2	α, n, m	$S_r = 0$	1	-73.3946	-104.935	0.022968	0.975665
3	S_r, α, n	$m = 1 - 1/n$	4	-66.3569	-98.3721	0.025669	0.975338
4	α, n	$S_r = 0, m = 1 - 1/n$	3	-68.7341	-100.236	0.027086	0.973414
5	S_r, α, n	$m = 1 - 2/n$	8	-41.1105	-73.1257	0.028032	0.971903
6	α, n	$S_r = 0, m = 1 - 2/n$	6	-42.9291	-74.4311	0.028481	0.971680
7	S_r, α, n	$m = 1$	5	-48.3125	-80.3277	0.023986	0.979472
8	α, n	$S_r = 0, m = 1$	7	-42.2259	-73.7279	0.028388	0.975571

3.4.2 Performance evaluation of various constraints in the van Genuchten (1980) model

In the previous section, it is shown that the choice of constraint in the vG model does have a significant effect on the obtained parameters. However, the same does not provide any information about the appropriateness of the choice of a particular constraint for bentonites. In this section, the constraints discussed in the previous section will be evaluated in terms of accuracy (RMSE), linearity (R^2) and information criterion (AIC and BIC). Finally, the 8 cases mentioned in Table 3.2 are ranked on the basis of AIC and BIC scores (ref. Fig. 3.1). Table 3.4 presents the performance statistics of various constraints of vG model evaluated in this study. Each of the 60 SWCCs of bentonite in the database is fitted to 8 cases of vG model (ref. Table 3.2). The values presented in Table 3.4 are averaged for 60 SWCCs used in this study. A similar approach for model selection was successfully utilized by Cornelis et al. (2005). In Table 3.4 it can be observed from RMSE and R^2 values that retaining S_r as a fitting parameter, results in a slightly better fit when compared between cases (1 and 2),(3 and 4),(5 and 6) and (7 and 8). It can also be observed that for all the cases except (7 and 8) when the comparison is made among cases (1 and 2), (3 and 4), (5 and 6) AIC and BIC values are consistently lower or conversely the rank (Table 3.4) is higher for the case with the constraint $S_r = 0$. This is not unexpected since AIC and BIC impose a penalty term for an extra number of parameters, hence for the models displaying similar accuracy of fitting (e.g.,RMSE

and R^2), the one with less number of parameters is favored. This observation implies that including S_r as an extra fitting parameter for bentonites is not desirable as $S_r = 0$ is a favorable constraint. This suggests that even while retaining S_r as a fitting parameter, the frequency of $S_r = 0$ is highest for all the cases. The same can be confirmed by observing the frequency histogram presented in Fig. 3.3. As already indicated in previous sections the above observation confirms that the constraint $S_r = 0$ is desirable for bentonites. Not only the highest frequency observed in Fig. 3.3 corresponds to $S_r = 0$, but it also results in a better model as per AIC and BIC. The same was also indicated experimentally by Tripathy et al. (2014) where the absence of the residual degree of saturation was shown using volumetric shrinkage curve of bentonite. By considering overall fit among all the cases, it can be observed that $S_r = 0$ and unconstrained m (i.e., Case 2) is the most favored set of constraint (Rank 1 in Table 3.4). It is interesting to note that that the second best set of constraint (Rank 2) according to AIC and BIC is case 1. Even though AIC and BIC impose a penalty term for the inclusion of extra parameter, the models with a larger number of free parameters i.e., 3 and 4 for case 1 and 2 respectively provide a better fit than most commonly used models with 2 free parameters i.e., case 4 and 6. Cases 5-8 fare poorly in comparison to cases 1 and 2 by a considerable margin. This implies that the application of any constraint related to parameter m reduces the curve fitting accuracy of vG model significantly for bentonites. The same can also be confirmed by inspecting the RMSE values across cases 1-8. While the averaged RMSE for case (2) is 0.023 the same for cases 4, 6 and 8 are 0.027, 0.028 and 0.028 respectively. Based on the above observation it can be concluded that the 3 parameters vG model with independent m and $S_r = 0$ is the ideal candidate for quantifying SWCC of bentonites. As the ultimate aim for this study is to construct the multivariate distribution of SWCC for bentonites the constraint identified as the best one i.e., case 2 would be used for subsequent construction of multivariate distribution.

3.4.3 Construction of a multivariate probabilistic model for bentonite SWCC

Table 3.5: Measured and simulated Kendall's rank correlation for van Genuchten (1980) parameters.

		Measured			Simulated(G)			Simulated(t)		
Kendall's τ		α	n	m	α	n	m	α	n	m
	α	1			1			1		
	n	0.11	1		0.08	1		0.11	1	
	m	-0.43	-0.67	1	-0.38	-0.64	1	-0.41	-0.66	1

Fig. 3.4 presents the SWCCs fitted to the compiled database for bentonite. The SWCC data points (S-ψ) are fitted to van Genuchten (1980) equation corresponding to (Case 4). It can be observed that from the original dataset consisting of 60 SWCCS, 11 outliers have been removed on the basis of unrealistic curves. It should also be noted that the traditional univariate identification of outliers such as those outside interquartile (25 % -75 % quantile) range, should not be applied in this case as the values of α,n,m are part of a three-dimensional space. A more straightforward way

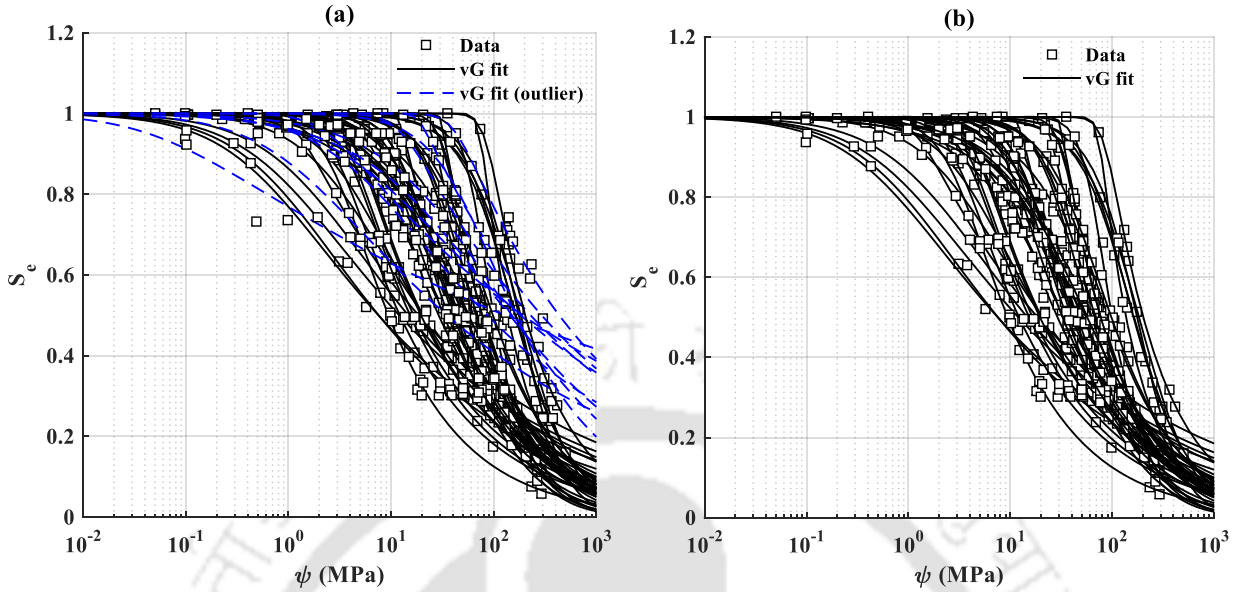


Figure 3.4: SWCCs fitted to van Genuchten (1980) equation for (a) Original database (b) Trimmed database.

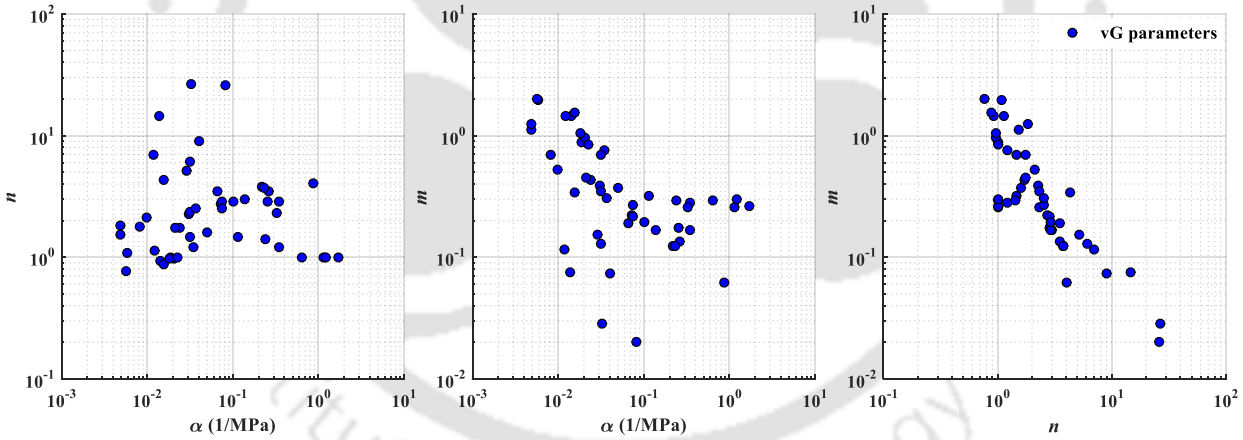


Figure 3.5: Scatter plot of the van Genuchten (1980) parameters.

Table 3.6: Identification of best fit marginals for curve fitting parameters in van Genuchten (1980) model.

Distribution	AIC		
	α	n	m
Lognormal	-99.81	205.82	32.40
Weibull	-88.18	228.69	34.73
Gamma	-81.68	227.57	34.21
Exponential	-64.43	226.72	35.00

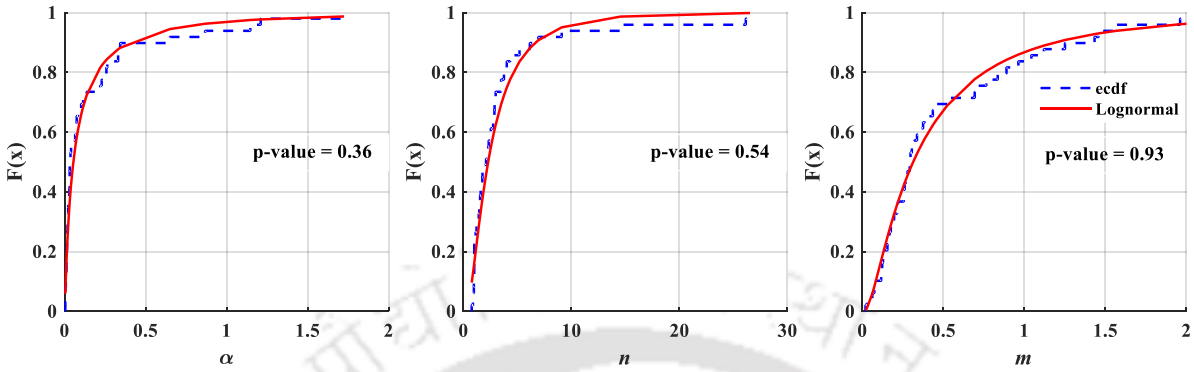


Figure 3.6: Empirical and calculated cumulative density function for van Genuchten (1980) parameters.

Table 3.7: Lognormal parameters for α,n,m.

Curve parameter	Distribution	Parameters			
		λ	ζ	μ	σ
α	Lognormal	-2.91	1.55	0.1811	0.5741
n	Lognormal	0.83	0.84	3.2635	3.3041
m	Lognormal	-1.16	1.04	0.5384	0.7517

Table 3.8: The goodness of fit for Gaussian and student t-Copula.

	Gaussian copula			t-copula($\nu = 4.8$)		
S_n statistics	0.0156688			0.015603		
p value	0.9765			0.9316		
	1	-	-	1	-	-
Parameter matrix θ	0.11	1	-	0.17	1	-
	-0.56	-0.84	1	-0.57	-0.86	1
Max. Log likelihood	61.15			63.17		
AIC	-116.30			-118.34		

is to remove the outliers on the basis of unrealistic SWCC curves. Henceforth, the probabilistic model is developed on the basis of the trimmed data set of 49 SWCCs. The fitted SWCC parameters α,n,m for the trimmed database are shown as a pair wise scatter plot in Fig. 3.5. The parameters are presented on the log scale in Fig. 3.5 since the observed range for α(0.0048-17.89) and n(0.73-26.53) is too large to be legible for arithmetic plots. It can be noted that apart from scatter plot between parameters m and n where a clear trend of negative correlation can be observed, it is difficult to judge the type of correlation (positive or negative) for rest of the two cases. From the measured rank correlation Kendall's τ provided in Table 3.5, it can be noted that the parameter α is weakly correlated (positive) to n whereas the correlation between α and m,m and n is relatively strong (negative). This means that the parameters cannot be treated as independent random variables, but they form a trivariate random vector. Joint PDF for these parameters is established using elliptical copulas, and the results are discussed as follows. The values presented in Table 3.5 are for the trimmed database of 49 SWCCs in contrast to one presented in Table 3.3 for 60 SWCCs of the original database.

As already discussed one of the key steps in constructing a trivariate distribution using copula is finding a suitable distribution for the respective one-dimensional marginals. Table 3.6 presents the AIC values for the candidate distribution chosen in the study. All the distributions chosen are strictly defined over positive values only. This is necessary as all the parameters for vG model corresponding to case 2 (ref Table 3.2) are theoretically bounded from below by zero. Ignoring this aspect can result in unrealistic simulated values. It can be easily observed that according to the minimum AIC criteria, lognormal distribution is the best fit candidate for parameters α,n and m. The distribution parameters were obtained using the maximum likelihood approach and the ones corresponding to the lognormal distribution for the α,n and m are provided in Table 3.7. The lognormal marginals are further compared against the empirical CDF as presented in Fig. 3.6, and it can be observed that the empirical and the observed (using Lognormal) CDF compares well. For a statistical confirmation, $K-S$ (Kolmogorov-Smirnov) test (Massey Jr, 1951) was performed. The test quantified the distance between empirical and observed CDF under the null hypothesis. p values for the K-S test are presented in Fig. 3.6. High values of $p(> 0.05)$ observed, in this case, suggests that there is no evidence against lognormal distribution. Lognormal marginals were used for α and n parameters by Phoon et al. (2010). However, in their study m was equal to $1-1/n$. Hence, it was necessary to truncate the distribution for n below 1. Their study was based on a database of UNSODA for various classes of soil such as silty clay and clayey loam which is different from bentonites considered in this study. Prakash et al. (2018b) found lognormal distribution to be successful for modelling the vG parameters α,n of fly ash arising out of instrumental and material uncertainties. The results reported in the literature and observations from this study, it can be concluded that vG parameters can be satisfactorily modelled using lognormal distribution.

After identifying the marginal, there is a need to judge the appropriateness of chosen copula to model the dependence structure among the parameters (in this case α,n,m). This was achieved using Cramer-von Mises statistic (S_n) and AIC, as discussed in the copula section. Table 3.8

presents the S_n along with p value for both Gaussian and student t-copula. It can be noted that S_n statistics is very low and the p value very high for both the copulas. This means that the null hypothesis for both the copulas has very low evidence against rejection. These results show that elliptical copula can be successfully used to model the asymmetric dependence structure among vG parameters. However, the p value is only an indicator of whether the null hypothesis for the chosen copula should be rejected or not. Since both the copulas have a substantial evidence against rejection, it is essential to identify the better model among two. AIC is used for this purpose in this study. It can be observed that t copula is a better choice than Gaussian copula since the former has the least AIC value among the two. The estimated dependence matrix θ for both copulas using maximum pseudo-likelihood approach is provided in Table 3.8.

Given the identified marginals and copula along with their parameters, the next step is to simulate the original data using the framework provided in Fig. 3.2 and to evaluate if the dependence structure is modelled correctly. Measured and simulated parameters using the lognormal marginals along with Gaussian and t copula are shown in Fig. 3.7. It can be observed that the prescribed copulas along with lognormal marginals can adequately represent measured vG parameters. The adequacy of simulation for modelling the dependence structure can be confirmed using the statistics provided in Table 3.5. It can be seen that the simulated rank correlation coefficient Kendall's τ is in close agreement with the measured one. Though t-copula has been identified as the best fit, it can also be observed that the simulated Kendall's τ for Gaussian copula is also very similar to the measured one. This confirms the statistics provided in Table 3.8 which showed that although t copula is a better fit, the relative difference in the degree of fit is very low for both the copulas used in this study. Hence it can be concluded that the elliptical copula can be used successfully to model the asymmetric dependence structure among vG parameters of bentonites. The result is of practical significance in the sense that the inference and the random number generation for elliptical copulas are relatively simpler than Archimedean or Vine copulas.

3.5 Practical implications of this study

Table 3.9: van Genuchten (1980) parameters for the confidence intervals created using t-copula.

Quantile	Remark	$\boldsymbol{\alpha}$	$\boldsymbol{n}$	$\boldsymbol{m}$
05%	Lower limit of 90% CI	0.069	0.56	1.16
25%	Lower limit of 50% CI	0.176	1.40	0.36
50%	Median	0.086	2.11	0.28
75%	Upper limit of 50% CI	0.036	2.76	0.27
95%	Upper limit of 90% CI	0.009	3.64	0.26

For many practical problems related to unsaturated flow and seepage modelling, uncertainty bounds related to the design parameters are required as a part of the preliminary assessment. In such cases, readily available confidence intervals for the SWCC are quite useful. Toward these

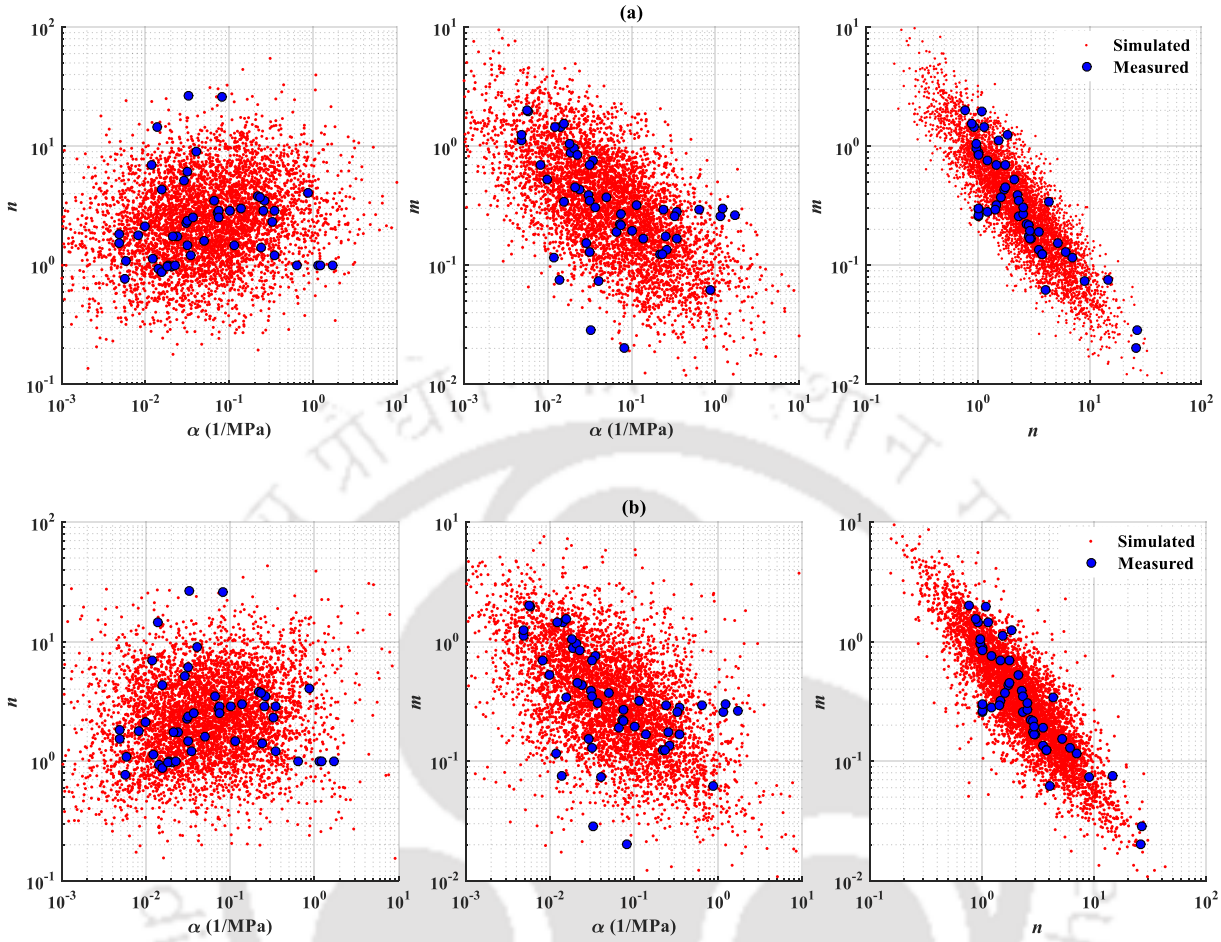


Figure 3.7: Measured and Simulated (N=5000) van Genuchten (1980) parameters using (a) Gaussian copula (b) t copula.

purposes, confidence intervals for bentonite SWCC are created using the evolved trivariate distribution evolved in this section. However, it should be noted that the outcome of this study is not limited to the confidence intervals, instead, the bentonite SWCC is described completely by its joint probability distribution. 10,000 samples of correlated α,n and m are simulated using the prescribed marginals and copula. These 10000 triplets of vG parameters essentially imply that at a particular value of suction there can be 10000 random values of S_e. Fig. 3.8 shows an example of the distribution of S_e with a simulated sample size of 5000 at 1, 10 and 100 MPa. The distribution of S_e is of practical interest as it is the most important parameter to be considered while assessing the performance of bentonites in engineering structures such as in hydraulic barriers and nuclear waste disposal facilities (Rahimi and Siddiqua 2017). It can be observed that at 1 MPa suction, the frequency is highest for S_e close to 1. The same can be observed even at 10 MPa, but in this case, the frequency is almost half as that of at 1 MPa. At 100 MPa it can be observed that frequency is highest for S_e close to 0.25. S_e is thereafter sorted, and various quantiles are found

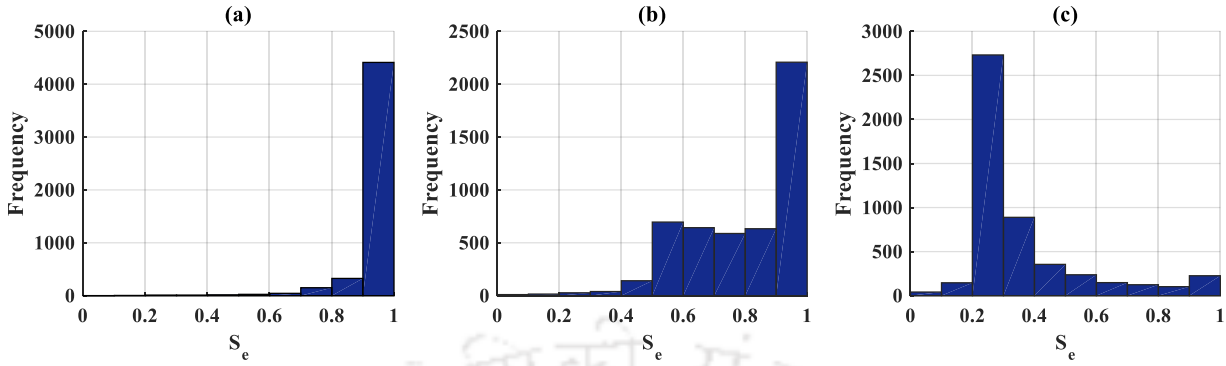


Figure 3.8: Frequency histogram of degree of saturation S_e at (a) 1 MPa (b) 10 MPa (c) 100 MPa.

out. Finally, confidence intervals for SWCC are created as a contour of various quantiles. Fig. 3.9 shows the various confidence intervals created for bentonite SWCC simulated using trivariate 't' and Gaussian copula. 5 % and 95 % quantile refer to the lower and upper limit respectively for the 90 % confidence interval. Similarly, 25 % and 75 % quantile refer to the lower and upper limit respectively for the 50 % confidence interval and 50 % quantile refers to the median curve. It can be easily observed that there is significant wide dispersion in the confidence intervals. This wide degree of dispersion essentially implies that the degree of uncertainty associated with the determination of bentonite SWCC is very high. Chiu et al. (2012) reported the confidence intervals of SWCC for sand, sandy loam and silty loam using Bayesian approach. In their study, it was shown that sand had the narrowest confidence interval meaning the lowest uncertainty or highest reliability in prediction. However, no clayey soil was reported in their study. Based on the observed high degree of dispersion in the present study it can be concluded that bentonites have widest confidence interval implying maximum uncertainty in prediction. The vG parameters for the confidence intervals using t-copula are provided in Table 3.9. The aforementioned confidence intervals are valuable for assessment of uncertainty bounds related to water flow and seepage modelling in the engineering projects/applications where SWCC of bentonites are used.

3.6 Summary

The motivation for this study stems from the fact that it is particularly difficult and cumbersome exercise to measure the soil water characteristic (SWCC) of expansive soils like bentonite. A statistical estimate from already measured SWCC data is desirable in such a case. While there are a lot of databases for SWCC of non-expansive soils, there are none for bentonites. A database of bentonite SWCCs was compiled in this study based on the published literature, which exhibited wide variability. To quantify this variability, a probabilistic model for bentonite SWCCs was proposed in this work within the purview of van Genuchten (1980) model. To arrive at this model, the impact of commonly used constraints on van Genuchten parameters was evaluated first. On

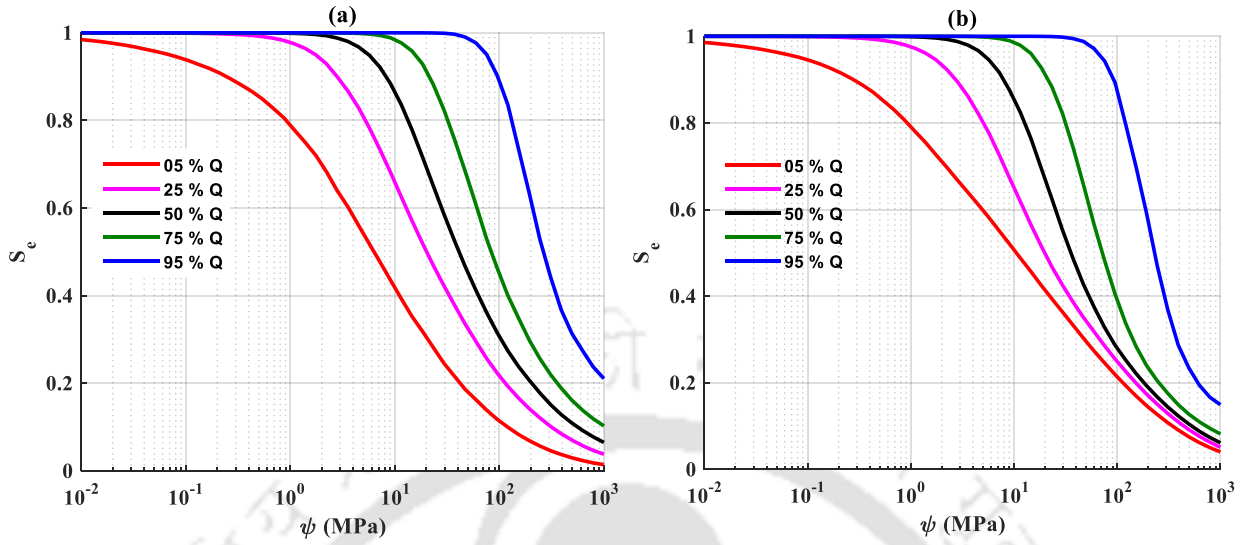


Figure 3.9: Confidence intervals for soil water characteristic curve of bentonites using (a) Gaussian copula (b) t-copula.

comparison with vG model under no constraints, commonly used constraints $m = 1 - 1/n$ and $m = 1 - 2/n$ resulted in an overestimation of mean α parameter by 17 % and 48 %, respectively. The constraint $m = 1$ resulted in an underestimation of the mean α parameter by 94 % . Under no constraint condition for the parameter m, values as high as 33 can be observed for the n parameter. It was also observed from this study that the correlation values among the vG parameters (α,n) vary widely with the constraints imposed on the parameter m. The study reinforces the fact that it is important to give due consideration to the constraints imposed on the parameter m while adopting the vG parameters for modelling unsaturated soils. It was also found that applying the constraint on the residual degree of saturation,$S_r = 0$, does not affect the parameter statistics and appropriateness of curve fitting. However, the application of any constraint related to m resulted in a less accurate representation of measured SWCC. Among all the constraints in the vG model, the three-parameter model with α,n,m as fitting parameters, was found to provide the best fit to the bentonite database. To account for the variability in the database, the trivariate joint distribution of parameters α,n,m was constructed using a copula approach. Lognormal marginals were found to be suitable for modelling the univariate random behavior of α,n,m. Correlations among the parameters were incorporated in the trivariate distribution using Gaussian and 't' copulas. Based on this study, both the copulas were found to be suitable though 't' copula was found to be a slightly better fit. For practical applications, confidence intervals were formulated for bentonite SWCC. The study demonstrated that elliptical copulas could be successfully used for constructing the multivariate distribution for SWCC of bentonites. The constructed multivariate distributions are not only useful as a first-hand engineering estimate in the absence of measured data for bentonites but also serves as a valuable informative prior for updating the site-specific limited data using

Bayesian approach.

Chapter 4

Probabilistic analysis of soil-water characteristic curve using limited data

4.1 General

The main objective of this study is to formulate a Bayesian approach to construct the joint distribution of SWCC using limited data. Prior information about the joint distribution parameters (marginals and copula) is incorporated with the limited SWCC data to obtain the updated joint distribution of SWCC parameters. Markov Chain Monte Carlo simulation is used to numerically implement the proposed Bayesian approach. The result indicate that the proposed approach can be successfully used to construct the joint distribution of SWCC from limited data.

4.2 Copula based Bayesian approach

SWCC measured in terms of volumetric soil-water content θ_w and ψ can be parametrized using the van Genuchten (1980) equation as follows:

$$S_e = \frac{\theta_w - \theta_r}{\theta_s - \theta_r} = \frac{1}{[1 + (\alpha\psi)^n]^m} \tag{4.1}$$

where θ_r, θ_s are the residual and saturated θ_w respectively and α, n are the vG parameters associated with air entry and desaturation respectively. m is a shape parameter commonly represented as $m = 1 - 1/n$. These parameters can be estimated using any standard non-linear optimization technique from the measured θ_w and ψ data. S_e known as effective degree of saturation is the normalized θ_w. For the soils used in this study θ_r was not retained as a fitting parameter but taken as zero. It is a common practice since the optimal value is close to zero (Phoon et al., 2010; Prakash et al., 2018*b*). For the case of $\theta_r = 0$, S_e reduces to θ_w/θ_s which is equal to degree of saturation (S). It should be noted that S_e and ψ in Eq. 4.1 are the state variables that enable the mathematical modelling of the state of the soil-water system and are independent of soil property. α and n in 4.1 are the parameters needed to calibrate the response described by state variables to a particular soil. Therefore α and n are the only soil dependent property needed to describe the SWCC relationship in Eq. 4.1. In this study, here onwards SWCC will be treated mathematically

as vG parameters α and n. N number of SWCCs are equivalent to N pairs of α, n. Therefore probabilistic characterization of SWCC (a curve) in this study means evolving joint distribution of α, n. It should be noted that in Chapter 3, Eq. 4.1 was not used with the constraint $m = 1 - 1/n$ i.e. probabilistic characterization of SWCC in Chapter 3 meant evolving joint distribution of three parameters i.e. α, n, m. In Chapter 3, the best form of vG constraint (independent m) was determined on the basis of AIC and BIC over the bentonite database. This Chapter describes a general case of methodology intended to be applicable to various textures of soil ranging from silt, loam to clay, therefore the most popular form of vG with m=1-1/n was chosen. This is also in line with other prominent probabilistic studies (Carsel and Parrish, 1988; Phoon et al., 2010, e.g.) which use the constraint i.e. $m = 1 - 1/n$.

For carrying out site specific unsaturated RBD, joint PDF of vG parameters, $f(\alpha, n)$ is required. The same after observing N site specific data i.e. N pairs of α, n can be expressed using the theorem of total probability (Ang and Tang, 2007, p. 59, Eq. 2.19) as follows:

$$f(\alpha, n|data) = \int .. \int_W f(\alpha, n|W) f(W|data) dW \tag{4.2}$$

where $f(\alpha, n|W)$ is the joint PDF of α, n given its distribution parameters $W = [W_\alpha, W_n, W_{\alpha,n}]$ and can be expressed using copula theory (Prakash et al., 2018*b*) as follows:

$$f(\alpha, n|W) = f(\alpha|W_\alpha) f(n|W_n) D(u, v|W_{\alpha,n}) \tag{4.3}$$

where $f(\alpha|W_\alpha)$ and $f(n|W_n)$ are the marginal univariate distribution of α and n respectively and can be expressed using three parameter log-normal distribution Phoon et al. (2010); Prakash et al. (2018*b*) as follows:

$$f_X(x|\lambda, \zeta) = \frac{1}{x\sqrt{2\pi}\zeta} exp[-\frac{1}{2}(\frac{\ln(x-c) - \lambda}{\zeta})^2], x > c \tag{4.4}$$

where $W_X = [\lambda, \zeta]$ are the lognormal distribution parameters and c is the physical lower bound for the random variable X. The above mentioned three parameter lognormal distribution is the usual two parameter log-normal distribution truncated from left at $x = c$. This is necessary as the SWCC parameter n has a theoretical lower bound $c = 1$ i.e. $n > 1$. Ignoring this aspect can result in unrealistic SWCCs during simulation (Phoon et al., 2010; Prakash et al., 2018*b*). For the SWCC parameter α , $c = 0$ i.e. $\alpha > 0$ which leads to the conventional two parameter log-normal distribution. The parameters W_X for both α and n can be estimated from the population mean μ and standard deviation σ as follows:

$$\lambda = \ln(\mu - c) - \frac{\zeta^2}{2} \tag{4.5}$$

$$\zeta^2 = \ln[1 + \frac{\sigma^2}{(\mu - c)^2}] \tag{4.6}$$

$D(u, v|W_{\alpha,n})$ in Eq. 4.3 is the copula density function accounting for dependence between α

and n and can be expressed using Gaussian copula Prakash et al. (2018*b*) as follows:

$$D(u, v|W_{\alpha,n}) = \frac{1}{\sqrt{1-\theta^2}} exp[-\frac{(\xi_1\theta)^2 + (\xi_2\theta)^2 - 2\theta\xi_1\xi_2}{2(1-\theta^2)}] \tag{4.7}$$

where $\xi_1 = \Phi^{-1}(u)$ and $\xi_2 = \Phi^{-1}(v)$,Φ is the standard normal cumulative distribution function (CDF). $u = F(\alpha)$ and $v = F(n)$ can be obtained using lognormal CDF transformation as follows:

$$F(x) = \Phi[\frac{\ln(x-c) - \lambda}{\zeta}] \tag{4.8}$$

For Gaussian copula the dependence parameter $W_{\alpha,n} = \theta \in [-1, 1]$ is related to rank correlation coefficient Kendall's τ as follows:

$$\tau = \frac{2}{\pi} \arcsin(\theta) \tag{4.9}$$

$f(W|data)$ term in Eq. 4.2 can be obtained using Bayes' rule (Ang and Tang, 2007, p. 63, Eq. 2.20) as follows:

$$f(W|data) = \frac{P(data|W)P(W)}{\int .. \int_W P(data|W)P(W)dW} \tag{4.10}$$

where $P(data|W)$ represents the likelihood of obtained SWCC data given the joint distribution parameters W. Assuming that the site specific N SWCCs i.e N pairs of α, n are independent realizations from the joint PDF $f(\alpha, n|W)$, $P(data|W)$ can be expressed using Eq. 4.3 as follows:

$$P(data|W) = \prod_{i=1}^{N} f(data_i|W) = \prod_{i=1}^{N} f(\alpha_i|W_\alpha) f(n_i|W_n) D(u_i, v_i|W_{\alpha,n}) \tag{4.11}$$

$P(W)$ term in Eq. 4.10 represents the prior knowledge about the parameters W. Assuming all the components in W are independent, the joint PDF $P(W)$ can be expressed as product of marginal univariate distribution as follows:

$$P(W) = P(W_\alpha)P(W_n)P(W_{\alpha,n}) \tag{4.12}$$

adopting a uniform marginal distributions, each component in Eq.4.12 can be expressed as,

$$\begin{cases} P(W_\alpha) = (\lambda_{\alpha max} - \lambda_{\alpha min})^{-1} \times (\zeta_{\alpha max} - \zeta_{\alpha min})^{-1} & ; \lambda_\alpha \in [\lambda_{\alpha min} - \lambda_{\alpha max}], \zeta_\alpha \in [\zeta_{\alpha min} - \zeta_{\alpha max}] \\ P(W_\alpha) = 0 & ; \lambda_\alpha \notin [\lambda_{\alpha min} - \lambda_{\alpha max}], \zeta_\alpha \notin [\zeta_{\alpha min} - \zeta_{\alpha max}] \\ P(W_n) = (\lambda_{n max} - \lambda_{n min})^{-1} \times (\zeta_{n max} - \zeta_{n min})^{-1} & ; \lambda_n \in [\lambda_{\alpha min} - \lambda_{\alpha max}], \zeta_n \in [\zeta_{n min} - \zeta_{n max}] \\ P(W_n) = 0 & ; \lambda_n \notin [\lambda_{\alpha min} - \lambda_{\alpha max}], \zeta_n \notin [\zeta_{n min} - \zeta_{n max}] \\ P(W_{\alpha,n}) = (\theta_{max} - \theta_{min})^{-1} & ; \theta \in [\theta_{min} - \theta_{max}] \\ P(W_{\alpha,n}) = 0 & ; \theta \notin [\theta_{min} - \theta_{max}] \end{cases}$$

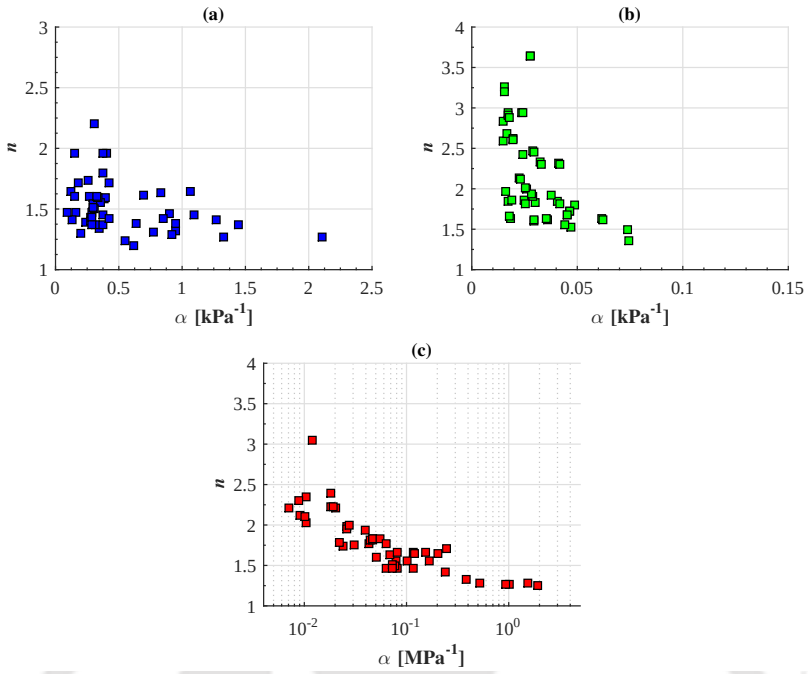


Figure 4.1: van Genuchten van Genuchten (1980) SWCC parameters for the database of (a) loam (Nemes et al., 2001) (b) fly ash (Prakash et al., 2018*b*) (c) bentonite (from Chapter 3) used in this study.

The denominator term $\int .. \int_W P(data|W)P(W)dW$ in Eq. 4.10 is a normalizing term required to make the posterior distribution of parameters $f(W|data)$ a valid PDF. The same will be denoted as K hereafter and it will be shown that calculation of this term is not required in the proposed framework. Substituting for the terms in Eq. 4.2 the joint PDF of SWCC parameters α, n given the observed data can be expressed as follows:

$$\begin{aligned} f(\alpha, n|data) = K^{-1} \int .. \int_W & f(\alpha|W_\alpha) \times f(n|W_n) \times D(u, v|W_{\alpha,n}) \\ & \times \prod_{i=1}^{N} f(\alpha_i|W_\alpha) \times f(n_i|W_n) \times D(u_i, v_i|W_{\alpha,n}) \\ & \times P(W_\alpha) \times P(W_n) \times P(W_{\alpha,n}) dW \end{aligned} \tag{4.13}$$

where $K = \int .. \int_W P(data|W)P(W)dW$ is the normalizing term in denominator of Eq. 4.10. $W = [W_\alpha, W_n, W_{\alpha,n}] = [\lambda_\alpha, \zeta_\alpha, \lambda_n, \zeta_n, \theta]$ is the distribution parameter's space.

Eq. 4.13 is a five dimensional integral over the parameter space $(\lambda_\alpha, \lambda_n, \zeta_\alpha, \zeta_n, \theta)$ and after substituting for all the components in Eq. 4.13, it is difficult to directly use this PDF or to compute the statistical moments. This is not an uncommon feature in Bayesian approach. This aspect can be circumvented by a general class of acceptance-rejection sampling algorithm called Markov Chain Monte Carlo (MCMC) (Gelman et al., 2013, p. 285, sec. 11.2). MCMC algorithms allow us to

Table 4.1: Population Statistics of van Genuchten van Genuchten (1980) soil water characteristic curve (SWCC) parameters α and n for the databases of Loam, Fly ash and Bentonite.

Statistic	Parameter	Loam*	Fly ash*	Bentonite#
Sample Size		52	54	49
Mean	α	0.52	0.03	0.19
	n	1.51	2.16	1.77
COV(%)	α	77.86	45.91	205.82
	n	13.13	26.57	20.30
Range	α	[0.09-2.10]	[0.01-0.07]	[0.01-1.26]
	n	[1.20-2.20]	[1.36-3.64]	[1.26-3.04]
ρ	α, n	-0.36	-0.56	-0.52
τ	α, n	-0.20	-0.49	-0.71

* Unit for α is kPa^{-1}, # Unit for α is MPa^{-1}

sample from a complicated PDF like in Eq. 4.13. As the name suggests, MCMC consists of two parts, a sequence called Markov Chain and the traditional hit and trial Monte Carlo simulation. Formally, Markov chain is a stochastic process wherein the probability of an event depends only upon the previous event. Mathematically, it can be represented as follows.

$$P(X_{j+1} = y|X_j = x_j, X_{j-1} = x_{j-1}, ..., X_0 = x_0) = P(X_{j+1} = y|X_j = x_j) \tag{4.14}$$

The general idea behind MCMC is to simulate from conventional distribution (easy to sample) and iteratively correct these simulated values till they reach the desired target distribution. The values at each iteration are generated sequentially conditioned only on the previous value (hence Markov chain). Statistical moments can further be computed over these simulated samples. In this study, Metropolis-Hastings (M-H)algorithm (Gelman et al., 2013, p. 289, sec. 11.4) is adopted for the MCMC simulation. M-H algorithm also allows to avoid calculating the K term and hence improves computational efficiency of the Bayesian approach. There are other algorithms also such as Gibbs sampler (Gelman et al., 2013, p. 287, sec. 11.3). The M-H algorithm was chosen in this study as it doesn't require conditional distributions of the target PDF to be sampled and also it is among the simplest to implement. The M-H algorithm used in this study is described below.

An initial state ($i = 1$) for the Markov chain of α and n i.e. α_i, n_i was chosen first. While length of chain $\mathbf{i} < \mathbf{N_s}$, the following was performed

- Calculate the target PDF $f(\alpha_i, n_i)$ in Eq.4.13 without the normalizing constant term K^{-1}
- Generate a random candidate sample α^*, n^* conditioned on $f(\alpha_i, n_i)$ using a suitable proposal distribution
- Calculate the proposal PDF $f(\alpha_i, n_i|\alpha^*, n^*)$ and $f(\alpha^*, n^*|\alpha_i, n_i)$
- Calculate the target PDF $f(\alpha^*, n^*)$ in Eq.4.13 without the normalizing constant term K^{-1}

- Calculate the acceptance ratio $r = \frac{f(\alpha^*,n^*)f(\alpha_i,n_i|\alpha^*,n^*)}{f(\alpha_i,n_i)f(\alpha^*,n^*|\alpha_i,n_i)}$
- Generate a random number U $(0,1)$
- If $r > U$ accept $\alpha_{i+1}, n_{i+1} = \alpha^*, n^*$ else $\alpha_{i+1}, n_{i+1} = \alpha_i, n_i$

It can be noted that even if the target distribution in Eq.4.13 was calculated including the K^{-1} term, it would get eliminated during the calculation of r in step 5. It can also be noted that the terms $P(data|W)$ and $P(W)$ need to be calculated once, only the term $f(\alpha|W_\alpha)f(n|W_n)D(u,v|W_{\alpha,n})$ need to be calculated for every iteration i of the Markov chain. For the proposal distribution in step 2, bivariate normal was chosen for simplicity in this study. Instead of directly obtaining prior on W it is convenient to obtain prior on $\mu_\alpha, \mu_n, \sigma_\alpha, \sigma_n$, and $\tau_{a,n}$. The corresponding prior on W can be determined using Eqs.4.5, 4.6, 4.9 respectively. It should be noted that for the proposed Bayesian approach, SWCCs were assumed to be from some definite copula or marginals for the limited data. Jaynes (1957) asserted that the exponential-family model (e.g. Gaussian) is the maximum entropy model under sparse constraints. Ideally, Bayesian model comparison should be performed (Zhang et al., 2018, e.g.) for designation of marginals and dependences structure (copula) for the case of limited data.

4.3 Database used in this study

Three databases for SWCC, one each for loam, fly ash and bentonite are utilized for illustrating the proposed Bayesian methodology. The database for loam is taken from UNSODA (Nemes et al., 2001, Texture loam), fly ash database from (Prakash et al., 2018*b*, Figure 1) and the bentonite database is taken from Chapter 3. Primary consideration behind taking these three particular databases is to demonstrate applicability of the proposed approach to a wide range of geomaterials. Air entry ψ values for the three databases combined ranged from 0.1 kPa- 10^5 kPa encompassing almost the entire theoretical ψ range of 0 kPa- 10^6 kPa. Another important consideration is the texture of soils since it affects the SWCC strongly Phoon et al. (2010); Chiu et al. (2012). The texture for these soils are loam, silt (fly ash) and clay (bentonite). There are other database available in the UNSODA Nemes et al. (2001) database for various other textures such as sand, sandy loam, clay loam etc. However the measured ψ for most of the cases rarely exceeds 100 kPa (e.g. Phoon et al., 2010). The database for fly ash and bentonite are chosen, since the measured ψ is considerably higher (ref. Fig. 4.2). Hence the proposed approach is demonstrated not only for almost entire range of possible measured ψ but also for a wide range of texture.

The vG parameters for these three databases are plotted in Fig. 4.1. Each of the SWCC in database (θ_w, ψ) is fitted to Eq. 4.1 with the constraints $\theta_r = 0$ and $m = 1 - 1/n$. For the loam database, in Fig. 4.1(a) it can be noted that higher values of α are associated with lower values of n and vice-versa. Similar observation can also be noted in Fig. 4.1(b) and Fig. 4.1(c) for fly ash and bentonite respectively. These observations suggest an existence of negative correlation or dependence between the vG parameters. The same can be confirmed through the linear correlation

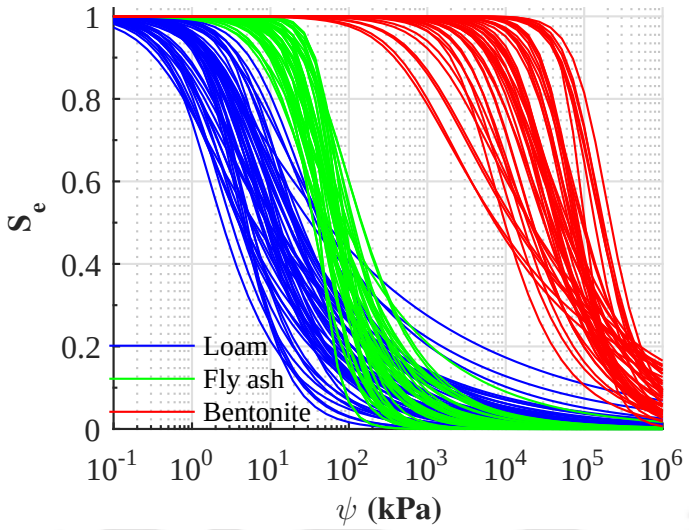


Figure 4.2: van Genuchten van Genuchten (1980) SWCCs corresponding to the three databases used in this study.

coefficient Pearson's ρ and rank correlation coefficient Kendall's τ provided in Table 4.1. The ρ and τ values for all three database are negative and significant. These ρ and τ values essentially implies that the parameters α and n cannot be treated as independent random variables, instead a joint distribution is required.

Negative correlation among the vG parameters was also reported by Phoon et al. (2010) on UNSODA database for various other textures of soil. Similar observation noted for fly ash and bentonite database suggests that negative correlation among vG parameters α and n can be generalized. It is also important to note that Chiu et al. (2012) reported negligible correlation between vG parameters. However in their study a single SWCC was fitted to the whole database. In this study and also in Carsel and Parrish (1988) and Phoon et al. (2010) each SWCC in the database was fitted to vG equation. These observations are crucial in the context of multivariate modelling of vG parameters since the degree of correlation may have a significant effect on the final RBD (Prakash et al., 2018*b*). In Table 4.1, it can also be noted that the $|\tau|$ values increase in order Bentonite> Fly ash> Loam. This indicates that it becomes even more important to treat parameters α and n as joint random variables with increasing ψ and fine content.

The parameter α is inversely related to air entry value (AEV) (Likos and Yao, 2014). It is also directly related to the mean pore size . AEV can be identified approximately in Fig. 4.2 as the ψ value where S_e drops below 1 for the first time while desaturation. For the black lines representing SWCC of loam, desaturation starts earliest at lower ψ in the order of few kPa, where as for bentonite the desaturation starts only at very high ψ values in the order of thousand kPa. Using the aforementioned inverse relationship, the mean α value is expected to be highest for loam and minimum for bentonite. The same is also expected on the basis of direct relationship with mean pore size since the mean pore size is smallest for bentonite clays. The same can be confirmed with the statistics provided in Table 4.1 wherein the mean α for loam and bentonite is 0.52 kPa^{-1}

Table 4.2: Prior utilized in this study for mean μ, standard deviation σ and rank correlation coefficient Kendall's τ of soil water characteristic curve (SWCC) parameters α and n.

Parameter	Range		
	Loam*	Fly ash*	Bentonite#
μ_α	[0.25,0.75]	[0.01,0.35]	[0.05,0.50]
σ_α	[0.02,1.20]	[0.005,0.100]	[0.10,0.80]
μ_n	[1.20,2.06]	[1.40,2.50]	[1.20,1.90]
σ_n	[0.03,0.29]	[0.25,0.75]	[0.10,0.50]
$\tau_{a,n}$	[-0.60,-0.10]	[-0.70,-0.30]	[-0.80,-0.30]

* Unit for α is kPa^{-1}, # Unit for α is MPa^{-1},

and 0.0002 kPa^{-1} (0.19 MPa^{-1}). The parameter n is related to rate of desaturation after the air entry starts. It is also related inversely to width of pore size distribution (Likos and Yao, 2014). In SWCC it is a measure of how steep the SWCC is after the air entry starts.

It can be noted in Fig. 4.2 that for fly ash the slope of SWCC is steepest among all the three. The same also reflects in the highest mean n parameter, as observed in Table 4.1. Using the inverse analogy with width of pore size distribution it is also expected since fly ash being a fine silt size waste product from thermal power plants have similar sized pores and hence narrower pore size distribution. Loam soil on the other hand is a mixture of sand, silt and clay and is expected to have wider range of pore size, hence a lowest mean n parameter. Regarding coefficient of variation(COV) for the parameters, it can be noted that COV for parameter α is highest for bentonite database at 205 % and lowest for fly ash at 46 %. It can also be noted that the COVs for parameter n are significantly lower than that of parameter α for all the three databases. This implies that parameter n can be estimated with higher certainty than parameter α. However again it should be noted that this implication is only true for the constraint of $m = 1 - 1/n$. This constraint forces the parameter n to fall in a narrow range of 1-4 . For the unconstrained case where m is independent of n, the parameter n can take values as high as 30 (Likos and Yao, 2014) and hence the observed COV for the parameter n in the same database would increase.

4.4 Discussion

4.4.1 Probabilistic characterization of SWCC

The databases presented in the preceding section are used to illustrate the proposed approach. Appropriate priors are formulated for the joint distribution parameters (marginals and copula). Further, limited number of data (e.g. $N = 2, N = 5, N = 10$) are randomly sampled from the databases and updated to get samples from the posterior distribution $f(\alpha, n|data)$ using the M-H algorithm reported in section 4.2. All the further statistical and probabilistic characterization is performed on these MCMC samples and compared against the statistics obtained using all the data points in the databases. Since the number of data points in all the three databases (ref. Table

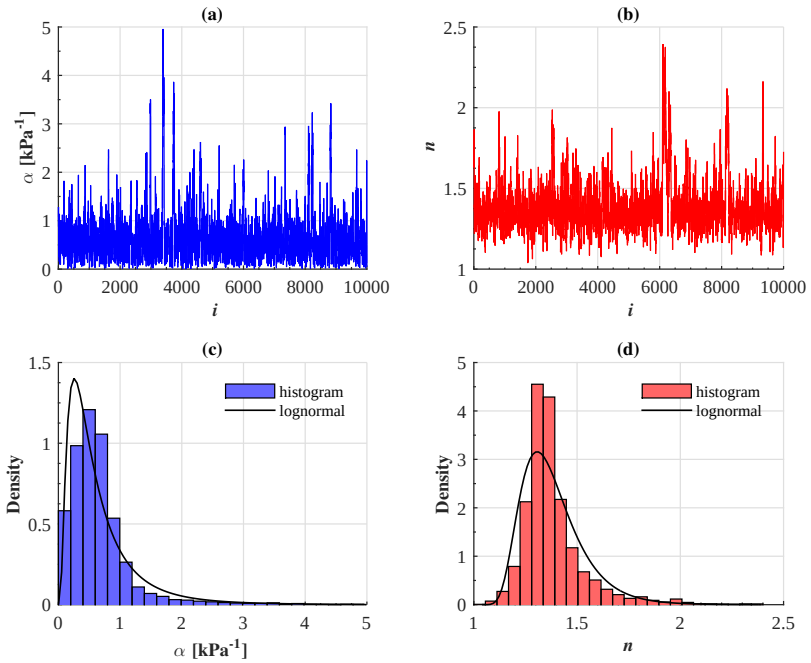


Figure 4.3: Trace plot of van Genuchten van Genuchten (1980) SWCC parameters (a) α and (b) n with length of MCMC chain i along with marginal density plot of (c) α and (d) n for loam.

4.1) are sufficiently large, the statistics obtained over the whole databases will be referred to as population statistics hereafter in this study.

As already discussed in section 4.2, the prior distribution for the log-normal parameters (for the marginals) along with the dependence parameter (for the copula) are chosen to be uniform. The parameters for uniform distribution are the minimum and the maximum value for the random variable. PDF for the uniform distribution is constant and equal to $(max - min)^{-1}$ over the entire parameter space and zero elsewhere. The choice of uniform distribution is convenient since the degree of belief about the distribution parameters can be tuned just by adjusting the width $(max - min)$ of the prior. A narrow width represents higher degree of belief in the proposed distribution parameters and vice-versa. For these reasons, uniform priors are used extensively in absence of extensive information about the distribution of prior (e.g. Wang and Cao, 2013; Wang et al., 2015). Uniform distribution is also used extensively to truncate the level of hierarchy imparted to the prior. In Bayesian approach, parameters of the prior distribution (also known as hyper-parameters) can also be assigned a distribution rather than a fixed value. The same are also known as hyper priors. In theory, the hierarchy can be extended even a further level up by assigning a probability distribution also to parameters of the hyper priors (also known as hyper-hyper-parameters) and so forth. In such cases, uniform distribution is used to truncate the level of hierarchy at that particular level. More details about hierarchical Bayesian modelling can be found in literature (e.g. Kéry and Schaub, 2011).

Since it is easy to intuitively interpret the data in terms of μ and σ rather than the log-normal parameters λ and ζ, the priors are formulated initially in terms of μ, σ, τ and thereafter converted

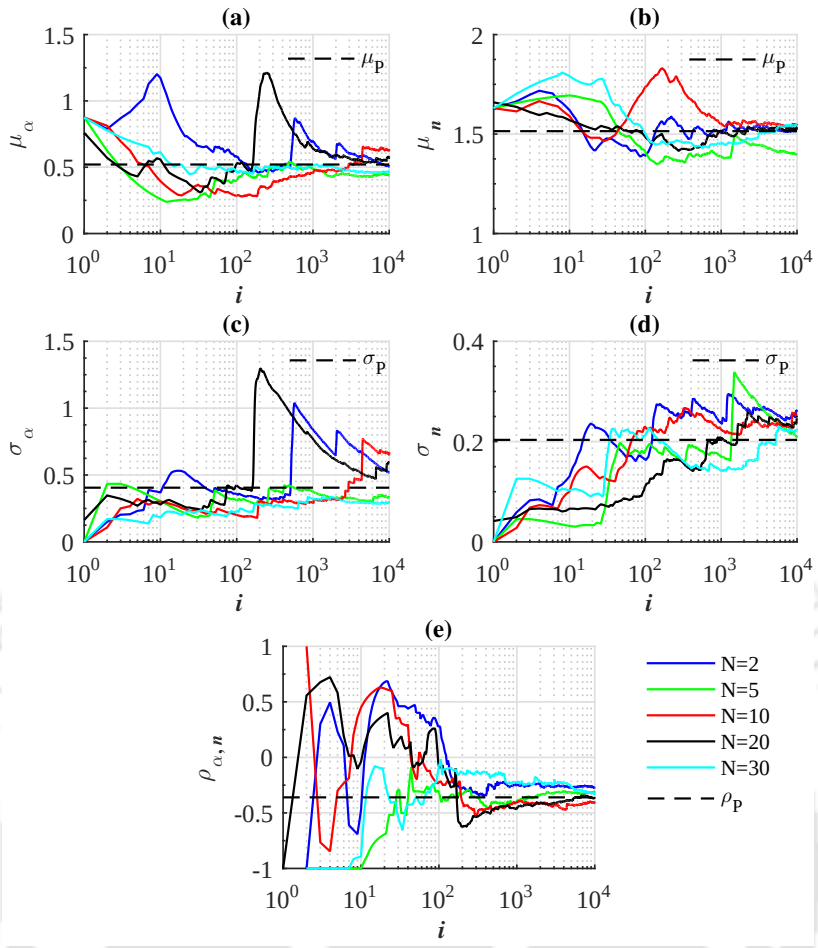


Figure 4.4: Cumulative mean μ, standard deviation σ and correlation coefficient ρ with length of MCMC chain i for loam. N is number of available data points and P in subscript denotes population statistics.

to λ, ζ, θ using Eqs. 4.5, 4.6, 4.9 respectively. Table 4.2 summarizes the parameters of the uniform prior for the joint distribution parameters. These values are compiled from literature (Phoon et al., 2010, Tables 2-3), (Chiu et al., 2012, Table 1), (Prakash et al., 2018*b*, Table 7) and (Meyer et al., 1997, Tables A1-A12) after giving proper subjective consideration to the texture of soil. The appropriateness of these priors will also be discussed later. It should be noted that the minimum and maximum values presented in Table 4.2 are for the μ, σ and τ of the vG parameters and not the range of values for the random variables α and n itself, which are more common in literature. For example the likelihood model being fitted in Chiu et al. (2012); Wang et al. (2018) was the vG model itself hence parameters for the priors were ranges of α and n. The likelihood model in Eq. 4.10 is a joint distribution with 2 parameters each for the marginal distribution of α and n respectively and one copula parameter for modelling the dependence among α and n. Hence the uniform priors are over the $\mu_\alpha, \sigma_\alpha$, μ_n, σ_n and $\tau_{\alpha,n}$ for the vG parameters.

For demonstrating efficacy of the proposed approach, random data points (α, n) with sample

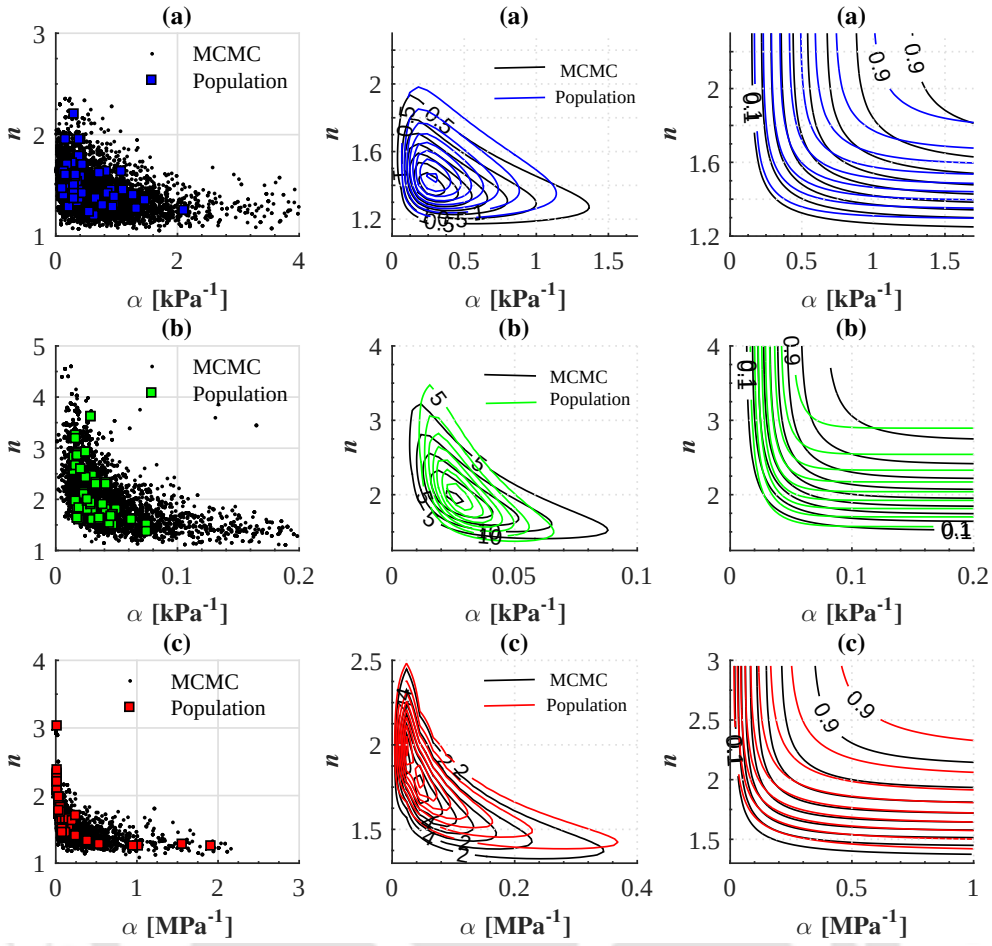


Figure 4.5: Scatter plot, contour of joint density and cumulative density of soil water characteristic curve (SWCC) parameters α and n obtained using Markov Chain Monte Carlo Simulation (MCMC) for (a) loam (b) fly ash and (c) bentonite.

size $N = 2, 5, 10, 20, 30$ are picked from the database. Following the M-H algorithm mentioned in section 2, MCMC simulations are performed for length of Markov chain $N_s = 10^6$. The obtained MCMC sample up to $N_s = 10000$ for α and n using observed data ($N = 2$) in the loam database are presented in Fig. 4.3. The histograms are also shown in Fig. 4.3. It can be noted that the MCMC samples follow the lognormal distribution prescribed in Eq. 4.4. The cumulative μ, σ and ρ of the MCMC chain up to $N_s = 10000$ for loam are also presented in Fig. 4.4. It can be easily observed that given only 2 observed data points ($N = 2$), priors in Table 4.2 and the proposed approach, the $\mu_\alpha, \sigma_\alpha$, μ_n, σ_n and $\tau_{\alpha,n}$ for the MCMC samples seem to be converging towards the actual population statistics. This result is significant, as it implies that given the prior in Table 4.2, and only 2 observed SWCCs (α, n) representative population statistics can be obtained. However, this also implies that the prior used in Table 4.2 are appropriate. This is mainly because proper consideration to the texture of soil was given while formulating the prior. It should also be noted that, any observation regarding the effect of increase in available data points (N) on final MCMC

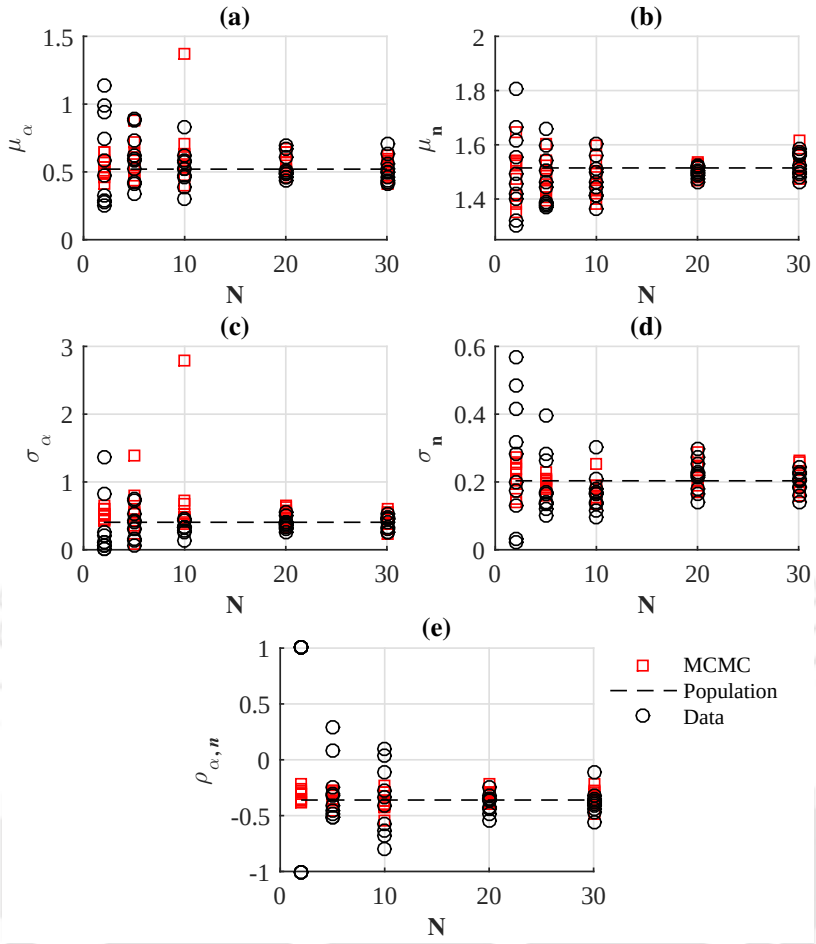


Figure 4.6: Influence of number of available data points N on statistics obtained over Markov Chain Monte Carlo (MCMC) samples (a) mean μ_α (b) mean μ_n (c) standard deviation σ_α (d) standard deviation σ_n and (e) correlation coefficient ρ for loam.

sample statistic is not appropriate based on Fig. 4.4. This aspect will be discussed separately later in the succeeding subsection. The aforementioned exercise is also done for the fly ash and bentonite databases. Observations similar to ones made for loam in Figs. 4.3 and 4.4 were noted, however the plots are not shown here for brevity. Also in Fig. 4.4, information about marginal statistics (μ, σ) and correlation coefficient (ρ) only could be obtained.

To ascertain if the joint distribution can be modelled satisfactorily, the scatter plots along with the contours of joint PDF and CDF for the simulated MCMC samples ($N_s = 1e6$) using $N = 2$, priors in Table 4.2 and the proposed approach for all the three database are presented in Fig. 4.5 . The results are also compared against the theoretical population. From the scatter plots in Fig. 4.5 it can be visually observed that the correlation structure is similar to the original population from where the random samples ($N = 2$) were taken. For a more detailed comparison, the joint PDFs for the population and MCMC samples were also calculated using Gaussian copula and lognormal marginals and are shown using contour plots in Fig. 4.5. Only the two outer contours are labelled for

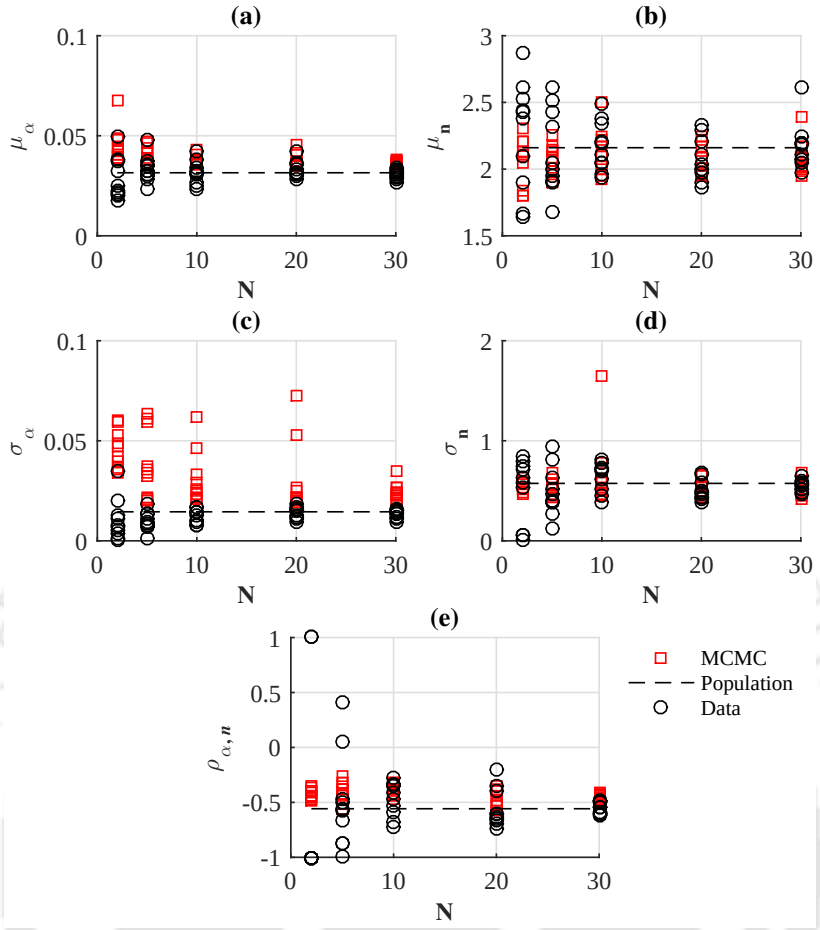


Figure 4.7: Influence of number of available data points N on statistics obtained over Markov Chain Monte Carlo (MCMC) samples (a) mean μ_α (b) mean μ_n (c) standard deviation σ_α (d) standard deviation σ_n and (e) correlation coefficient ρ for fly ash.

legibility. The step size for the contours are 0.5, 5 and 2 for loam, fly ash and bentonite respectively. By observing the joint PDFs in Fig. 4.5, it can be noted that the dependence structure (shape) for MCMC samples can be considered representative of the actual population. The high density (inner) regions as well as low (outer) density regions for the MCMC samples also are representative of the population density. Fig. 4.5 also presents the joint CDF contours (right subplots). It can be noted that the CDFs are sufficiently close and can be considered representative of the actual population CDF. It should be noted that these results in Figs. 4.3 and 4.5 were obtained using only 2 available data points and they are not expected to be completely similar to the actual population even in the proposed approach. However the statistics obtained on MCMC samples (obtained using limited data and the proposed approach) are expected to be better estimates (lower uncertainty) than the ones obtained only on limited data. Also with increase in number of available data points, these estimates are expected to improve and converge towards the actual population, an aspect discussed in the subsequent sections.

Table 4.3: Priors parameters utilized for examining the influence of prior width on statistics obtained over Markov Chain Monte Carlo (MCMC) samples.

Parameter	Prior	Range		
		Loam*	Fly ash*	Bentonite#
μ_α	M2	[0.40,0.60]	[0.02,0.15]	[0.09,0.30]
	M3	[0.10,0.90]	[0.005,0.500]	[0.01,0.75]
σ_α	M2	[0.20,0.80]	[0.009,0.050]	[0.20,0.60]
	M3	[0.02,1.60]	[0.001,0.200]	[0.04,1.00]
μ_n	M2	[1.30,1.80]	[1.80,2.40]	[1.50,1.90]
	M3	[1.20,2.50]	[1.40,2.80]	[1.10,2.30]
σ_n	M2	[0.10,0.25]	[0.35,0.65]	[0.25,0.50]
	M3	[0.01,0.35]	[0.10,0.95]	[0.05,0.75]
$\tau_{\alpha,n}$	M2	[-0.50,-0.20]	[-0.70,-0.50]	[-0.60,-0.40]
	M3	[-0.90,-0.20]	[-0.90,-0.10]	[-0.90,-0.10]

* Unit for α is kPa^{-1}, * Unit for α is MPa^{-1},

4.4.2 Influence of number of available data points

In this section, a sensitivity analysis on the number of available SWCC data points on final MCMC statistics is done. Objective of this analysis is to check for the convergence of MCMC statistics towards actual population statistics. Number of data point chosen are $N = 2, 5, 10, 20, 30$. For each N, 10 random trials were made. For example at $N = 2$, 10 sets of $[\alpha, n]_{2\times2}$ are obtained from the database. The same priors as in Table 4.2, are utilized and MCMC simulations are performed. The parameter statistics obtained on the available data as well as MCMC samples are compared against the actual population statistics. It should be noted that the statistics obtained only on the available data represent the conventional approach in literature (Carsel and Parrish, 1988; Phoon et al., 2010; Prakash et al., 2018*b*, e.g.) for probabilistic analysis of SWCC. Therefore this section not only checks for convergence of MCMC statistics but at the same time it also provides a preliminary comparison of the proposed approach against conventional approaches in the literature.

Fig. 4.6 presents the obtained results for the database of loam. The square markers represents the statistics obtained using the MCMC samples ($N_s = 1e6$). The same are also compared with the statistics obtained using only the available data (represented by circles) e.g. $[\alpha, n]_2$. For N=2, it can be observed that the red squares are less dispersed and more closer to the population statistic than the black circles. It essentially implies that the uncertainty in parameter estimates is now reduced by using the Bayesian approach. It can also be noted that for N=2 the ρ values obtained from the available data are either -1 or 1. This is because using two data points only -1 or +1 values can be estimated, where as the the ρ values obtained on MCMC samples are less dispersed

and much closer to the actual population ρ.

With increase in number of data points, in Fig. 4.6 it can be observed that range of dispersion for both red squares and black circles decreases. At higher available data points e.g N=30, both estimates seem to converge towards the actual population statistics. This observation demonstrates the convergence of the proposed approach. This is also in line with the philosophy of Bayesian approach (Wang and Cao, 2013, e.g.) wherein the reduction in uncertainty is higher at low data points, and with the increase in number of available data points, estimates from both Bayesian approach (MCMC samples) and conventional approaches (on available data) converge. However as already discussed, it is not practically feasible to have obtain large number of SWCCs at a particular site. The difficulty can also be judged from the fact that even the global database compiled from literature (ref. Table 4.1) hardly contains 40-50 data points for a particular soil texture.

Similar observations were also be made for the database of fly ash and bentonite, however only the results for fly ash database are presented for brevity in Fig. 4.7. In general, it can be observed in Figs. 4.6 and 4.7 that scatter in the squares is less than the circles at each N level. Hence the proposed approach fares better than the conventional approaches. Further explicit comparison between the proposed and conventional approach will be provided in the subsequent section. It is also important to note that the trend of decrement in dispersion with increase in number of data points could not be observed in strict sense. For example, in Fig. 4.7 (c) the dispersion in red square is slightly higher for $N = 20$ than $N = 2$. However at $N = 30$, the dispersion is low as expected. This is mainly because only 10 sets of samples were used for the sensitivity analysis and these are not exhaustive over the database. For example, a total of 54!/2!52! combinations of α and n can be sampled from the fly ash database for the case of $N = 2$. For the sake of brevity and computational simplicity, only 10 sets were used for each N. Nevertheless, the overall observation from Figs. 4.6 and 4.7 are in line with the Bayesian philosophy (Wang and Cao, 2013, e.g.) and hence demonstrate robustness of the proposed approach.

4.4.3 Influence of prior width

For demonstrating the influence of prior width $(max - min)$ of the uniform distribution on the final MCMC sample statistics, two sets of prior ranges other than the one utilized in Table 4.2 are chosen. Range of first set is chosen to be narrower whereas second set is chosen to be wider than the original priors presented in Table 4.2. The prior with narrow range represents a higher degree of belief about the parameter value and vice-versa. These two priors are labelled as M2 (narrow) and M3 (wide) respectively from heron and are summarised in Table 4.3.

Fig. 4.8 presents the μ, σ and ρ for vG parameters of fly ash obtained on MCMC samples using M2 and M3 priors. In Fig. 4.8, it can be easily observed that the statistics obtained using samples from M3 prior (denoted as black circles) are in general more dispersed than the one with M2 priors (denoted as black squares). This observation essentially implies that a more informative prior (narrow) improves the estimation of parameter statistics. Similar inference were drawn for the loam and bentonite databases. However only the results for the bentonite database are presented

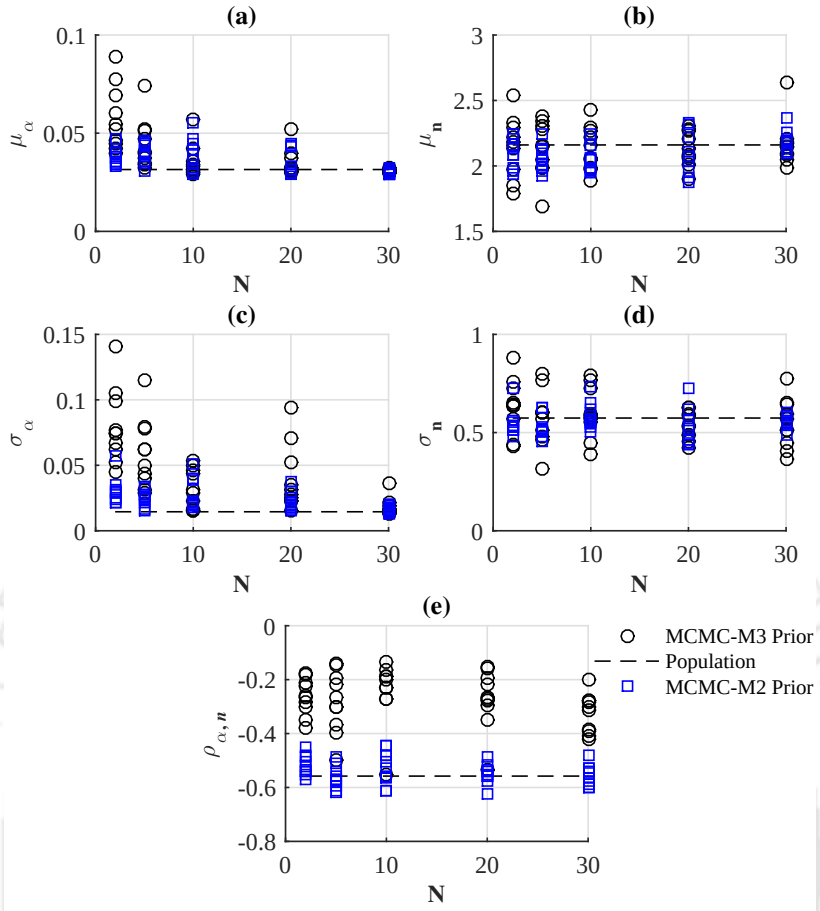


Figure 4.8: Influence of prior width on statistics obtained over Markov Chain Monte Carlo (MCMC) samples (a) mean μ_α (b) mean μ_n (c) standard deviation σ_α (d) standard deviation σ_n and (e) correlation coefficient ρ for fly ash.

in Fig. 4.9 . It can be noted in Figs. 4.8 and 4.9 that the difference in dispersion among the results from M2 and M3 prior diminishes gradually with increase in number of data points. This aspect can be particularly observed in Fig. 4.9(a) and Fig. 4.9(c). These observations are expected and inline with the philosophy of Bayesian approach (Wang and Cao, 2013) and (Kéry and Schaub, 2011, e.g. P36,Tab 2, Lines 5-8) wherein the effect of prior should diminish gradually with increase in number of data points. However it should be noted that the prior needs to be properly defined in the context of likelihood or observed data for the above statement to be true (Gelman et al., 2017). The prior act as a bound on the parameter space. If the prior is not properly defined, effect of prior may not diminish with an increase in number of data points.

The results along with the one in preceding subsection display robustness of the proposed approach. However, this result also implies that at lower number of available data points which is more likely to be the case with SWCC, proper consideration should be given to construct appropriate priors for a particular class of soil. For characterization of SWCC, the priors formulated in Table 4.2 are in general appropriate as shown in previous sections and are recommended for their

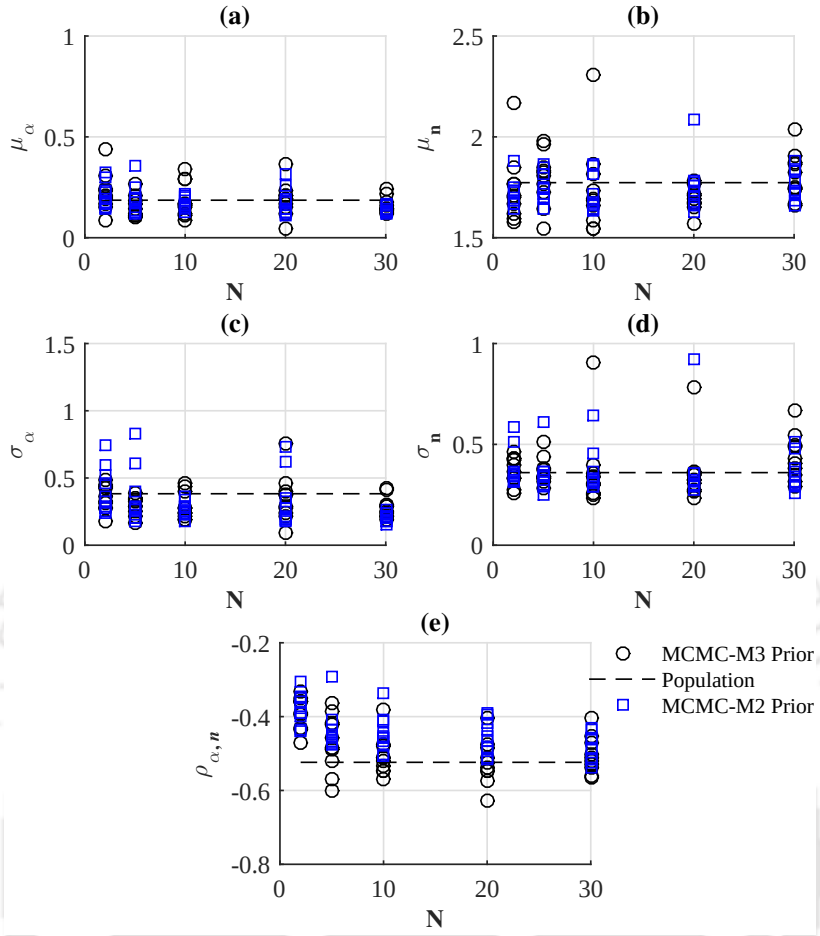


Figure 4.9: Influence of prior width on statistics obtained over Markov Chain Monte Carlo (MCMC) samples (a) mean μ_α (b) mean μ_n (c) standard deviation σ_α (d) standard deviation σ_n and (e) correlation coefficient ρ for bentonite.

respective classes.

4.5 Comparison with conventional approach

A natural question which arises is that how the proposed Bayesian approach in this study compares against other alternate approaches for probabilistic analysis of SWCC. Therefore this section, compares the parameters estimates $W = [\lambda_\alpha, \zeta_\alpha, \lambda_n, \zeta_n, \theta_{\alpha,n}]$ (ref. Eq. 4.13) obtained using the proposed approach in this study and the approaches in literature. Popular approach in literature is to fit a parametric joint distribution to the observed data of α, n either using translational approaches (Carsel and Parrish, 1988; Phoon et al., 2010) or copula approach (Prakash et al., 2018*b*). Particularly the approach (to be referred as conventional approach hereafter) followed in Prakash et al. (2018*b*) will be used for comparison since the copula approach was utilized in these studies and a one to one comparison between estimated parameters is possible.

A numerical experiment, similar to the previous two subsections was conduced. For each

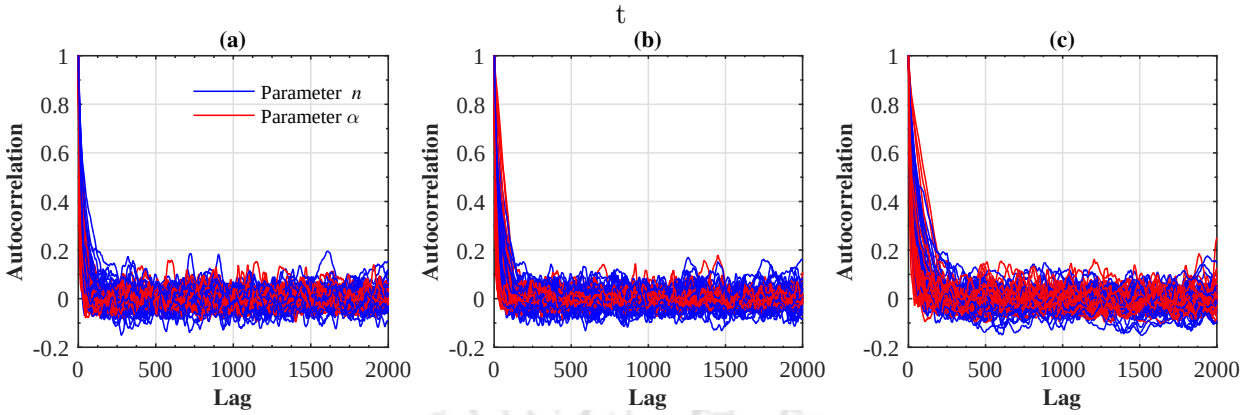
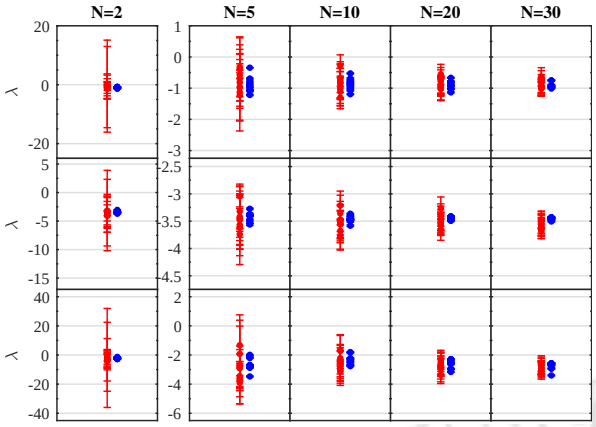


Figure 4.10: Markov chain Monte Carlo (MCMC) chains autocorrelation plot for the (a) laom (b) fly ash and (c) bentonite database.

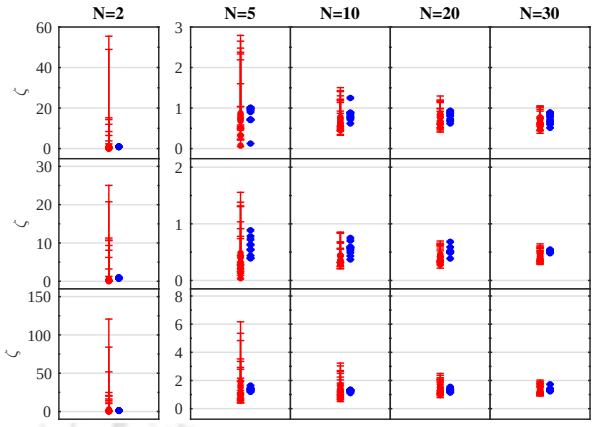
database of loam, fly ash and bentonite, random trials were made. Number of data points randomly picked from the databases are $N = 2, 5, 10, 20, 30$. For each N, 10 random trials were made. The same priors as in Table 4.2 were utilized. In the conventional approach, (Prakash et al., 2018*b*) only the available data will be used to estimate the parameters of the joint distribution i.e. the vector $W = [\lambda_\alpha, \zeta_\alpha, \lambda_n, \zeta_n, \theta_{\alpha,n}]$. In the proposed approach, these random trials will be updated to obtain MCMC samples corresponding to the target PDF $f(\alpha, n|data)$ using the M-H algorithm (ref. 4.13) and the parameters will be estimated over these MCMC samples.

It is worthwhile to note that the conventional approach is straight forward to use as the marginals and copula parameters can be fitted simply to the available data. However, this approach works satisfactorily only for data sufficient cases. For a real case scenario, the number of SWCC will be most likely limited in numbers, and hence the statistical uncertainty will be high for the estimated joint distribution parameters (to be also shown later). The proposed approach, although comparatively computationally intensive, supplements the available data, i.e. α, n with prior constructed from a statistical analysis conducted in the literature to obtain the posterior distribution $f(\alpha, n|data)$. Not only it reduces uncertainty in the estimates at a low number of data points, but it also allows for subjective and expert judgement (in the form of prior) to be incorporated. In the conventional approach, even at abundant data cases, the parameters estimate is solely governed by the available data, and there is no scope for the incorporation of expert judgement or experience.

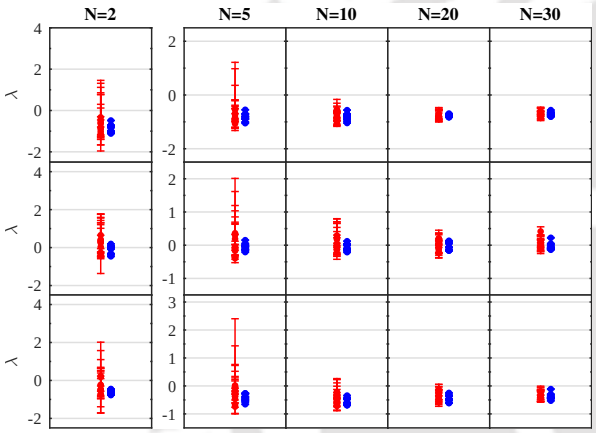
For the comparison, a total of 150 MCMC simulations were performed i.e. 50 ($length(N) = 5 \times Trials = 10$) each for the loam, fly ash and bentonite dataset. Before making any statistical inference, it is necessary to inspect for the convergence of the MCMC chain. Although the convergence can be judged by visually observing trace plots (e.g. Fig. 4.3), it is worthwhile to statistically confirm the convergence. Statistical confirmation becomes important in this section also, since 150 MCMC simulations were performed, and it is not possible to visually present and judge every chain for inspection. Popular approaches for checking convergence are autocorrelation plots and Gelman-Rubin statistics (Gelman et al., 2013). In this study, autocorrelation plots are used



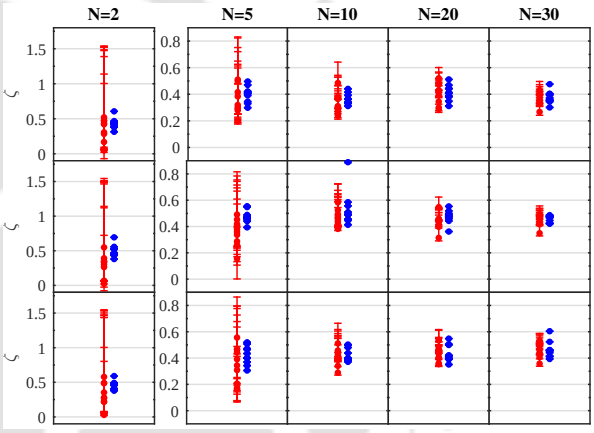
(a) Comparison of lognormal marginal parameter λ_α estimate using conventional (red) and proposed (black) approach.



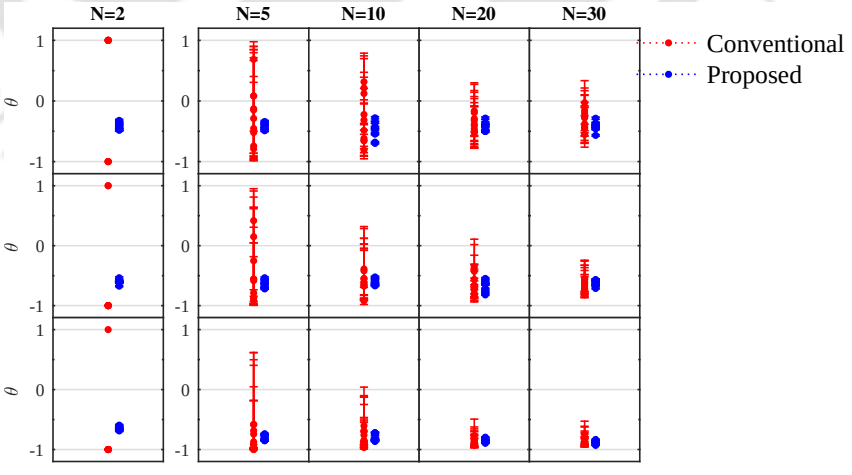
(b) Comparison of lognormal marginal parameter ζ_α estimate using conventional (red) and proposed (black) approach.



(c) Comparison of lognormal marginal parameter λ_n estimate using conventional (red) and proposed (black) approach.



(d) Comparison of lognormal marginal parameter ζ_n estimate using conventional (red) and proposed (black) approach.



(e) Comparison of copula parameter θ estimate using conventional (red) and proposed (black) approach.

Figure 4.11: Comparison of joint distribution parameters $W = [\lambda_\alpha, \zeta_\alpha, \lambda_n, \zeta_n, \theta_{\alpha,n}]$ obtained using conventional (red) and proposed (black) approach. First, second and third row in each subplot denotes the statistics obtained for loam, fly ash and bentonite database respectively.

for convergence diagnostics since calculating Gelman-Rubin statistic requires multiples chains with different starting points for the MCMC simulation. For example, even if only 10 chains were used for each MCMC simulations, the total number of MCMC simulations will reach 1500. This will result in a lot of wastage of MCMC chains and is computationally expensive and time-consuming. Hence only autocorrelation plots will be used for demonstrating the convergence. Fig. 4.10 presents the MCMC simulations performed for loam, fly ash and bentonite dataset. It can be noted that that the autocorrelation dies down quickly for all the simulations to zero hence demonstrates a proper mixing and thus the convergence of the MCMC chain.

Given that the MCMC chains have converged the parameter estimated over MCMC samples (referred to as proposed approach) and the approach mentioned in Prakash et al. (2018*b*) (referred to as conventional approach) can be compared. Fig. 4.11 presents the comparison of joint distribution parameters $W = [\lambda_\alpha, \zeta_\alpha, \lambda_n, \zeta_n, \theta_{\alpha,n}]$ estimated from the proposed approach and a conventional approach. The parameter estimates are accompanied by their 95 % confidence interval corresponding to the estimate. In Fig. 4.11, it can be easily observed that at a lower number of data points (e.g. $N = 2, N = 5$) the scatter in the parameter estimate using the proposed approach is significantly lower than the conventional approaches. The proposed approach not only utilizes the available data points but also supplements it with the prior information and therefore, statistical uncertainty associated with the estimates are comparatively lower. In Fig. 4.11 it can also be noted that with an increase in the number of data points, uncertainty in estimates of both conventional and proposed approach decrease and at a higher number of data points (e.g. $N = 30$) both the conventional and proposed estimate converge. These observations not only demonstrates the superiority of proposed approach in terms of lower uncertainty in the parameter's estimate but also demonstrate the robustness of the proposed approach since with the increase in the number of data points both the Bayesian and conventional estimate converge towards the same results.

4.6 Practical Application: Site specific unsaturated RBD

The key purpose for probabilistic analysis of SWCC from limited data in this study is site specific RBD. MCMC samples from $f(\alpha, n|data)$, can be further used as an input parameter for site-specific unsaturated RBD. To demonstrate the application, an example for unsaturated infinite slope is presented in this section.

Factor of safety (FS) for an unsaturated infinite slope is given as follows:

$$FS = \frac{c + c(\psi) + \gamma_t H cos^2\beta tan\phi}{\gamma_t H sin\beta cos\beta} \tag{4.15}$$

where c, ϕ, γ_t are cohesion, angle of internal friction and total unit weight respectively. γ_t varies with degree of saturation S (equal to S_e in Eq. 4.1)and can be calculated as: $\gamma_t = \gamma_w(G_S + Se)/(1 + e)$, H is depth of the failure surface, β is slope angle and $c(\psi)$ is the apparent cohesion resulting due to ψ. The unsaturated shear strength is imparted to the slope via this term. For estimation of $c(\psi)$

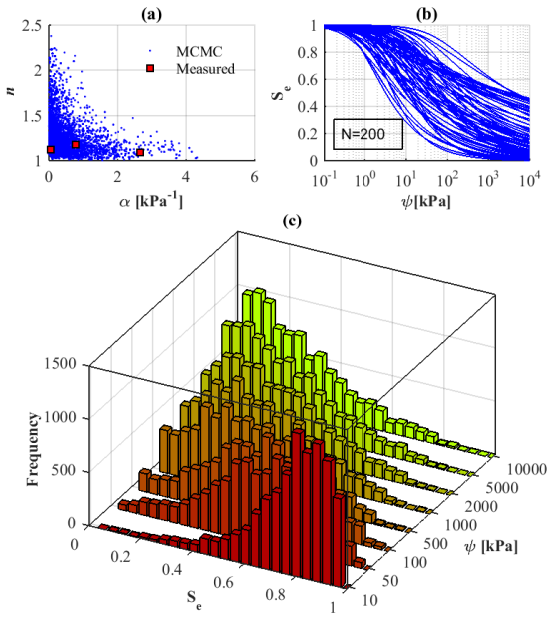


Figure 4.12: (a) Scatter plot of measured ($N = 3$) and simulated Markov Chain Monte Carlo (MCMC) ($N_s = 1e4$) vG parameters (b) simulated Markov Chain Monte Carlo (MCMC) vG soil water characteristic curves (SWCCs) (c) histogram of degree of saturation S_e at various suction ψ values.

method propose by Lu and Griffiths (2004) incorporating the vG parameters is used. The same is given as follows

$$c(\psi) = \frac{\psi}{[1 + (\alpha\psi)^n]^{1-1/n}} \times tan\phi \tag{4.16}$$

The failure probability (P_f) can be estimated in terms of limit state function $g(X)$. For the slope stability problem limit state can be defined as $g(X) = FS - 1$ and P_f can be estimated in terms of input parameters $\mathbf{X} = \{X_1, X_2, ..., X_n\} = \{H, \beta, c, \phi, G_s, e, \alpha, n\}$ as follows.

$$P_f = P(g(X_1, X_2, ..., X_n) \leq 0) = \int ... \int_{g(\mathbf{X})\leq 0} f_X(x)dx \tag{4.17}$$

In the Monte Carlo simulation framework, the above can be estimated using the standard indicator function approach as follows.

$$P_f = \frac{1}{N}\sum_{i=1}^{N} I(g(\mathbf{X}) \leq 0) = \frac{1}{N}\sum_{i=1}^{N} I(FS - 1 \leq 0) \tag{4.18}$$

where N is the number of realizations of FS and I is an indicator function for counting the number of times the slope fail i.e. $FS \leq 1$.

For construction of an embankment, a design β is required. For the current example, it is assumed that the fill material to be used for construction is sandy clay loam, and three SWCC

Table 4.4: Slope parameters used for the application example in this study.

Variable	Distribution	Parameters
Depth of failure(H)	Uniform	$[2, 3]$ m
Slope angle(β)	Deterministic	$[20^\circ - 60^\circ]$ $\Delta\beta = 5^\circ$
Cohesion (c)	Lognormal	$[2.67, 0.27]$ i.e. mean 15 kPa and $COV = 0.25$
Angle of internal friction(ϕ)	Lognormal	$[3.55, 0.25]$ i.e. mean 35.90° and $COV = 0.25$
Specific gravity of solid(G_s)	Deterministic	2.70
Void ratio(e)	Uniform	$[0.41, 0.69]$
SWCC	Unknown	data(α, n)=(0.79,1.18),(2.65,1.09),(0.06,1.12)
Prior		
μ_α	Uniform	[0.25,0.75]
σ_α	Uniform	[0.02,1.20]
μ_n	Uniform	[1.20,2.00]
σn	Uniform	[0.05,0.35]
$\tau_{\alpha,n}$	Uniform	[-0.70,-0.20]

Note: Unit for α is kPa^{-1}

measurements (i.e 3 pairs of α, n) for that soil are available. These three hypothetical SWCCs are randomly picked from the sandy clay loam dataset, in the UNSODA database. There are a total of 50 SWCCs for the sandy clay loam soil in UNSODA. Out of 50, 12 outliers were identified by Phoon et al. (2010). Therefore 3 SWCCs (i.e 3 pairs of α, n) were randomly picked from a total of 38 SWCCs in the sandy clay loam dataset. Further statistical detail about the sandy clay loam dataset can be found in Phoon et al. (2010) and a through and extensive detail can be found in Nemes et al. (2001). All the other parameters taken for the infinite slope in this example are summarized in Table 4.4. In table 4.4 it can be noted that the distribution of SWCC parameters is designated as unknown. This is because a joint distribution can not be evolved from 3 data points. Therefore the proposed Bayesian approach in this study will be utilized to get $1e6$ MCMC samples. For all the slope variables other than SWCC, the distribution (e.g. Lognormal) and its parameters (e.g. 2.67 and 0.27) are summarized in Table 4.4.

Using the priors and the available SWCC data (ref. Table 4.4), MCMC simulations are performed using the framework in section 2. The simulated MCMC samples ($N_s = 10000$ for representation) along with the measured data are shown in Fig.4.12(a). Fig. 4.12(b) shows the SWCCs for simulated MCMC samples. Only 200 SWCCs are shown for legibility. It can be easily observed that existence of joint distribution of SWCC parameters essentially implies that instead of an unique S_e, there exists a distribution of S_e at every value of ψ. For example, Fig. 4.12(c) shows the histograms of S_e computed at various ψ values. At $\psi = 10$ kPa, it can be noted that the highest frequency corresponds approximately to $S_e = 0.8$. With increase in ψ value, the histogram shifts left toward lower S_e. The distribution of S $(= S_e)$ is of practical interest as it is one of the most important parameter to be considered while assessing the performance of unsaturated geotechnical and geo-environmental structures such as soil covers (e.g. Woyshner and Yanful, 1995), waste

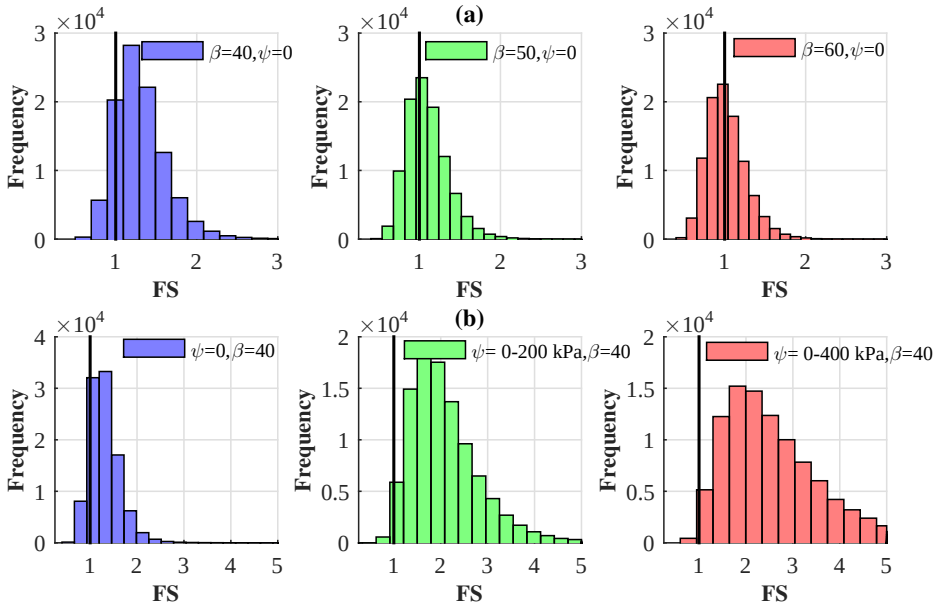


Figure 4.13: Histogram of factor of safety FS for (a) saturated case (b) unsaturated case.

repositories (e.g. Villar et al., 2008), slope stability (e.g. Lu and Godt, 2008) etc. For the present application example, the MCMC samples ($N_s = 1e6$) are utilized to estimate the FS in Eq. 6.21 in the framework of Monte Carlo Simulation.

Fig. 4.13 presents the histograms of the calculated FS. For conventional saturated case i.e $\psi = 0$, Fig. 4.13(a) presents the histogram of FS for $\beta = 40°, 50°$ and $60°$. At a relatively gentle slope of $\beta = 40°$, it can be noted that the fraction of $FS < 1$ is quite low, whereas with increase in β the histogram shifts towards left and fraction of $FS < 1$ becomes higher. For the unsaturated case, Fig. 4.13(b) presents the histogram of FS for $\beta = 40°$ at $\psi = 0, 0 - 200$ and $0 - 400$ kPa. It can be noted that fraction of $FS < 1$ becomes lower with the increase in ψ range. The fraction of $FS < 1$ (ref. 4.18) at any particular β or ψ range can be quantified by P_f using Eq.4.18. Fig. 4.14 shows the variation of P_f and FS_n (Nominal factor of safety) with slope angle β. FS_n for slopes is generally taken between 1.2 and 1.5 (Orr, 2000). It can be defined as 5 percent quantile value of the FS.

Fig. 4.14(a) shows the variation of P_f with slope angle β. Three suction ranges commonly observed in field (e.g. Sivakumar Babu and Murthy, 2005; Hazra et al., 2017) are assumed. For comparison, zero suction i.e. the conventional saturated failure probability is also computed. It can be easily noted that the P_f at various slope angles for unsaturated case are always lower in comparison to zero suction case. This is because of contribution of suction to the shear strength via the $c(\psi)$ term in Eq. 6.21. The same information in terms of FS_n is demonstrated in Fig. 4.14(b). It can be easily observed that with the increase in suction range, the target β for a particular FS_n increase. For example, the design target β corresponding to $FS_n = 1.2$ for saturated case is around 29° whereas it is 39°,44° and 55° for $\psi = 0 - 100, 0 - 200$ and $0 - 400$ kPa respectively. It can be noted that on account of higher shear strength contribution from ψ, a steeper slope can be

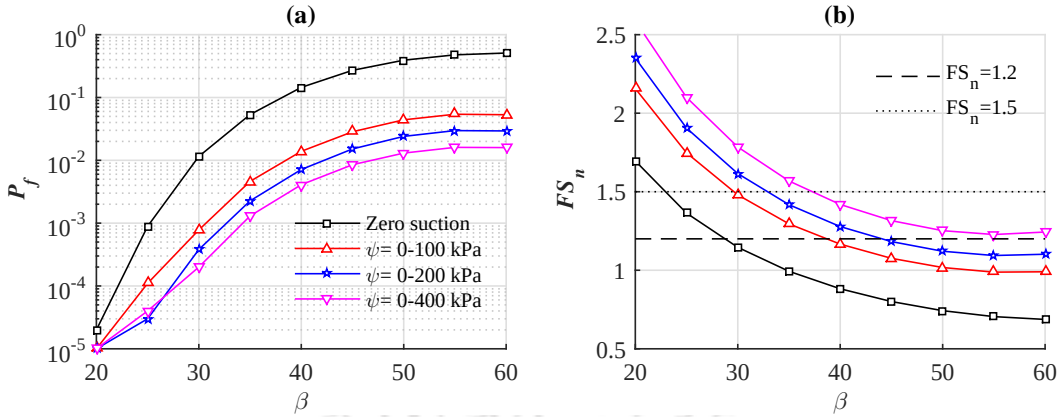


Figure 4.14: (a) Failure probability P_f vs slop angle β and (b) nominal factor of safety FS_n vs β obtained using Markov Chain Monte Carlo (MCMC) samples.

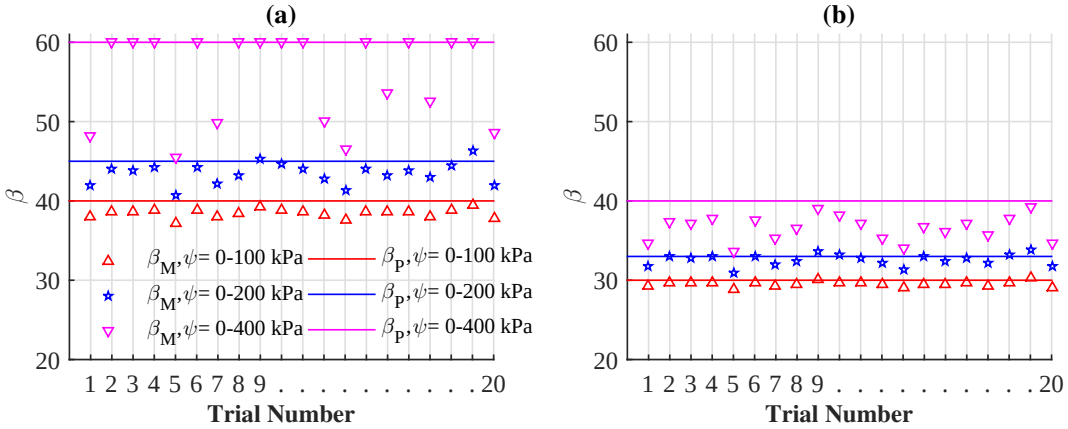


Figure 4.15: Comparison of desig slope angle β obtained using Markov Chain Monte Carlo (MCMC) samples (β_M) and original population (β_P) for the nominal factor of safety FS_n (a) 1.2 and (b) 1.5.

designed.

In Fig. 4.14, RBD of unsaturated slope was conducted assuming only 3 available SWCCs and the proposed Bayesian approach. These 3 SWCCs were picked a from already known population (database). In real life scenario, the population will be unknown and only the SWCCs obtained for a particular site will be available. It was shown in previous section, that using the proposed Bayesian approach, MCMC samples representative to the population can be obtained. It can be expected that the design β obtained in Fig. 4.14 should also be a reasonably fair estimate of the design β, had the population been known. To ascertain this aspect, RBD of the same slope is conducted using 10^6 SWCC samples generated from the original population i.e. the joint density evolved using the whole dataset. These 10^6 samples are generated using Gaussian copula and lognormal marginals statistics for sandy clay loam provided in Phoon et al. (2010). The design β obtained using the original population (β_P) and the MCMC approach β_M are computed and shown in Fig. 4.15. 20 random trials were performed in which three SWCCs were randomly picked in

each trial from the population for comparison. For the cases where $\beta_M > 60°$, β_M was taken equal to 60°. It can be noted that the estimated β_M are reasonably close to the β_P for all the three ψ ranges and thus demonstrate practical utility of the proposed approach.

4.7 Summary

This study emphasized the need for probabilistic analysis of soil-water characteristic curve (SWCC) from limited data towards unsaturated reliability-based design. It also highlighted that it is practically impossible to obtain numerous SWCCs at a project site for a satisfactory probabilistic characterization using conventional methods. Owing to the significant non-linear dependence between the van Genuchten (1980) (vG) SWCC parameters, it also becomes necessary to treat the vG parameter as joint random variables. Towards this end, a Bayesian approach integrated with copula theory was formulated in this study. Prior information about the joint distribution parameters (marginals and copula) was incorporated with the limited SWCC data to obtain the updated joint distribution of SWCC parameters.

Efficacy of the proposed approach was demonstrated using three SWCC databases (one from UNSODA and another two compiled by authors) with distinct soil textures; loam, silt and clay. Air entry suction values for the three databases combined ranged from 0.1 kPa- 10^5 kPa (theoretical range is 0 kPa- 10^6 kPa). It is shown that the proposed Bayesian approach along with the priors formulated in this study can be successfully used to construct a representative joint distribution of SWCC using only a few measured SWCCs.

To examine the robustness of the proposed approach, the influence of the number of available SWCCs as well as the prior distribution on the estimated parameter statistics was also examined. The results were consistent with the Bayesian philosophy and hence demonstrated the robustness of the proposed approach. It was also shown that statistical uncertainty in the joint distribution parameters obtained using the proposed approach are lower than the conventional methods in literature i.e. to obtain parameters only over the available limited number of SWCCs.

For demonstrating a practical engineering utility of the proposed approach, reliability-based design for an unsaturated soil embankment was conducted. The design slope angles obtained using only three SWCCs, and the proposed approach were in close agreement with the design slope angle obtained using the original population. Overall, the methodology and the prior formulated in this study provide a practical tool for the unsaturated reliability-based design using limited number of SWCCs.

Chapter 5

Probabilistic estimation of specific surface area and cation exchange capacity

5.1 General

This study proposes a multivariate probabilistic approach for estimation of CEC and SSA. For the construction of the multivariate model, a five-dimensional database (278×5) (labelled as CLAY/C-S/5/278) is developed. CLAY/C-S/5/278 database contains five variables LL, PI, CF, CEC and SSA. The database is truly multivariate meaning all the five parameters were measured for the same soil sample. However, it should be noted that the construction of 5-dimensional distribution is not a straightforward task, particularly for this database since three variables CF, CEC, and SSA have theoretical/practical upper bound. This aspect requires the use of truncated distributions. Although the popular translational approach (Ching and Phoon, 2014; Ching et al., 2017, e.g.) can be used to model the marginals, the dependence structure is still inherited from the parent multivariate normal distribution. An alternate to translational approach (Ching and Phoon, 2014; Ching et al., 2017, e.g.), copula theory allows for arbitrary marginals, and dependence structure as the marginal and dependence modelling are uncoupled and can be done separately. However, even the copula approach is popular in geotechnical literature mostly in two dimensions ($n = 2$) (Li et al., 2015; Tang et al., 2015; Prakash et al., 2018*a*; Zhang et al., 2018, e.g.). Beyond 2 dimensions ($n > 2$), very few generalizations of bivariate copula exists and they too have their own limitations regarding the dependence structure (Wu, 2013*b*; Zhu, Zhang, Xiao and Li, 2017, e.g.). Vine-copula approach leverages the popularity and flexibility of bivariate copulas and renders the construction of multivariate joint density ($n > 2$) only in terms of bivariate copulas. Originally developed by Bedford and Cooke (2001); Aas et al. (2009) for application in the field of financial economics, very recently it has started to find its utility in geotechnical engineering (Wang and Li, 2018, 2019; Lü et al., 2020, e.g.). Lü et al. (2020) in particular utilized the vine-copula theory very recently to estimate design parameters conditional on other measured parameters.

Table 5.1: Details of the compiled database CLAY/C-S/5/278.

S.I No.	Reference	Country	N	LL(%)	PI(%)	CF(%)	CEC(meq/100g)	SSA(m^2/g)
1	Chittoori and Puppala (2011)	USA	20	25-67	11-45	20-59	57-134	132-431
2	Farrar and Coleman (1967)	England	19	28-121	7-89	10-69	10-56	34-186
3	Ohtsubo et al. (1983)	Japan	39	46-146	17-98	24-75	20-44	130-543
4	Smith et al. (1985)	Israel	51	18-73	1-41	18-81	8-80	46-449
5	Cerato (2001)	USA//UK/Norway /Japan/Thailand	43	24-560	5-508	10.2-95.8	0.9-120.9	11-767
6	Yukselen and Kaya (2006)	Turkey/USA	15	24.5-464.8	6.08-416.7	21-90	2-132.3	25.5-582.3
7	Sivapullaiah et al. (2008)	India	21	27-415	10-346	15-100	0.8-69.5	12-560
8	Cerato and Lutenegger (2004)	USA	2	45-101	22-66	48.4-66.6	7.7-35.9	43-227
9	Sivapullaiah and Lakshmikanthay (2005)	India	2	134-310	55-261	100-100	16-100	90-800
10	Zhang, Qu, Liu and Wu (2003)	China	3	63-68	37-51	39-50	42-49	295-412
11	Arifin (2008)	Germany	2	180-400	124-367	40-82	62-74	400-525
12	Dexter (1990)	Australia/USA /Netherland	6	29.2-81	7.6-42.7	18.5-66.8	11.7-49.8	13-135
13	Sutherland et al. (1996)	USA	1	53-53	24-24	80-80	20.8-20.8	84.2-84.2
14	Mishra et al. (2011)	USA/Japan	15	119-678	67-622.7	32-85.7	61-95.6	351-714
15	Marcial (2013)		3	120-520	70-478	64.8-89.7	69.1-76	515-800
16	Schwing et al. (2013)		1	61.7-61.7	37.8-37.8	75-75	10-10	30-30
17	Spagnoli et al. (2013)	Netherland/England	2	72-159	50-129	32-49	21-33	44-59
18	Santhoshkumar et al. (2016)	India	1	410-410	365-365	72.6-72.6	60.8-60.8	87.5-87.5
19	Perret et al. (1996)	Canada	1	51.5-51.5	29.2-29.2	59.5-59.5	8.5-8.5	65-65
20	Khorshidi et al. (2016)	USA	4	44-485	18-353	35-100	9-102	21-668
21	Tripathy et al. (2014, 2015)	USA/UK/Greece	4	51-437	19-374	84-98	3.18-90	10-797
22	Bharat and Gapak (2018)	India	4	175-597	148-546	62-84	39-109	376-495
23	Dutta and Mishra (2016)	India	2	218-560	183-524	57-68	27.2-44.6	339-456
24	Rao and Ravi (2013)	India/Japan/Korea /China/Canada/USA /Sweden	6	104-474	48-447	40-86	58-119	348-725
25	Tripathy et al. (2004)	USA/Spain/Germany	3	102-411	49-364	68-78	62-102	493-725
26	Agus and Schanz (2006)	Calcigel	1	130-130	97-97	40-40	49-49	651-651
27	Kale and Ravi (2018)	India	2	248-447	208-398	76-89	64-95	274-508
28	Deka (2015)	India	4	244-450	183-396	52-64	15.66-38.18	215-428
29	Gupt et al. (2018)	India	1	300-300	257-257	65-65	53.66-53.66	450-450
		Total	278	18-678	1-622.7	10-100	0.8-134	10-800

Table 5.2: Some statistics for the five parameters in the CLAY/C-S/5/278 database.

Parameter	Mean	COV(%)	Min.	Max.
LL(%)	133	111	18	678
PI(%)	99	141	1	623
CF(%)	54	37	10	100
CEC($meq/100g$)	42	73	0.8	134
SSA(m^2/g)	250	79	10	800

COV denotes coefficient of variation

However, the study utilized already established CLAY/5/345 and CLAY/6/535 databases (Ching and Phoon, 2012; Ching et al., 2014) for demonstration of vine copula approach. This study not only develops a new CLAY/C-S/5/278 database but also is the first study in geotechnical engineering to consider CEC and SSA in a probabilistic manner. Since this study deals with the construction of joint density in 5 dimensions ($n > 2$) and the dependence structure is predominantly non-Gaussian (to be shown later) vine copula approach is utilized in this study.

The constructed 5-dimensional joint distribution summarizes not only the marginals but also the conditional dependence among them. The joint density is constructed over a (278×5) database (CLAY/C-S/5/278) compiled from many geographical regions around the globe and therefore can be considered a global multivariate distribution. Implementation examples of the global multivariate distribution for probabilistic estimation of CEC and SSA are also shown. It is shown that the joint distribution can be viewed as global prior/unconditional PDF, which can be updated to posterior/conditional PDFs of CEC and SSA using Bayes' rule when new data is available. In the proposed approach, CEC and SSA are characterized by their joint conditional PDFs. This implies that uncertainty associated with the estimation of CEC and SSA is rigorously characterized. Therefore, the proposed approach is superior to the popular transformation approach, which can only be used to get the point estimate. For practitioners, the results of this study, i.e. the conditional probability density functions (PDFs) of CEC and SSA are summarized in the form of user-friendly analytical expressions. Two practical geotechnical application examples are also shown. The first example is related to the estimation of segregation potential (SP) for quantification of frost heave susceptibility in highway sub-grades, and the second example is related to the determination of swelling pressure (p) of clays at various dry densities (ρ_d). The proposed approach is also vital for the extension of reliability-based design and analysis approaches in the projects where CEC and SSA are key governing parameters. For example, carbon storage and capture, nuclear energy and long term waste containment (Lu and Mitchell, 2019, p.108).

5.2 Database: CLAY/C-S/5/278

This study compiles a multivariate database of 5 clay parameters LL, PI, CF, CEC, SSA. The database is labelled as CLAY/C-S/5/278, where "C" denotes CEC and "S" denotes SSA. The total

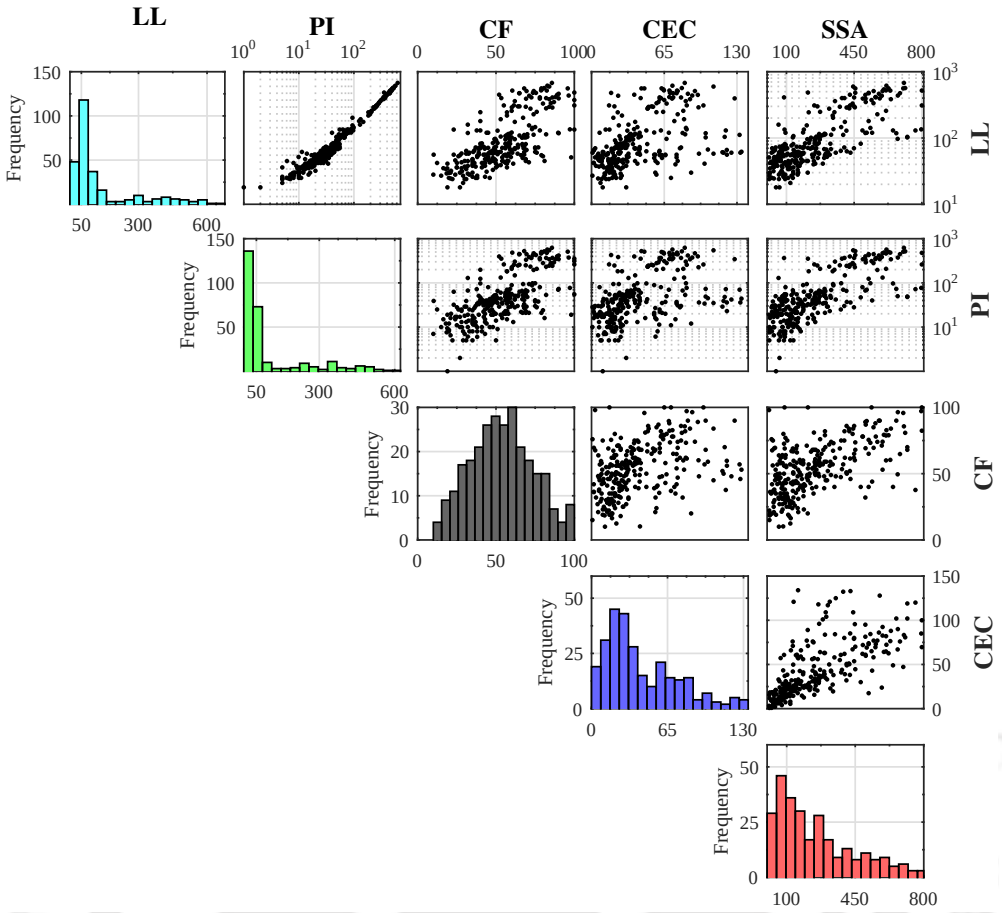


Figure 5.1: Scatter and histogram plot for the 5 parameters in the CLAY/C-S/5/278 Database.

number of soil samples considered is 278, and for each sample, five properties were measured, i.e. a total of 1390 measurements. Only the studies where all these five parameters were measured for the same soil sample or at the same depth in the vicinity were utilized. This is necessary for the database to be truly multivariate. The empirical histogram, along with the scatter plots for all the five parameters, is shown in Fig. 5.1. Some statistics about the parameters are summarized in Table 5.2. It can be noted that the range of LL (20 % - 700 %) covered in this database is extensive. It can also be noted that the range of all the parameters in the database span almost the entire possible range of values. For example the $PI \sim 1\% - 600\%$, $CF \sim 10\% - 100\%$, $CEC \sim 0.5 - 130$ $meq/100g$ and $SSA \sim 10 - 800$ m^2/g. The upper/lower bound value is equal or close to the theoretical/observed, upper/lower limit of these parameters. For example, CF covers the whole theoretical range up to 100%. CEC and SSA do not have theoretically defined upper bounds. Still, montmorillonite mineral dictates the maximum vales of CEC and SSA. Therefore maximum CEC and SSA values of montmorillonite reported in the literature (Low, 1980; Lu and Mitchell, 2019) and also in the database, i.e. $CEC = 135$ and $SSA = 800$ can be taken as the upper limit for clays. The database covers soils samples from the USA, India, Israel, China, UK, Europe, Canada, Japan, Korea, Australia, Turkey and Thailand. Brief details of the utilized database with their

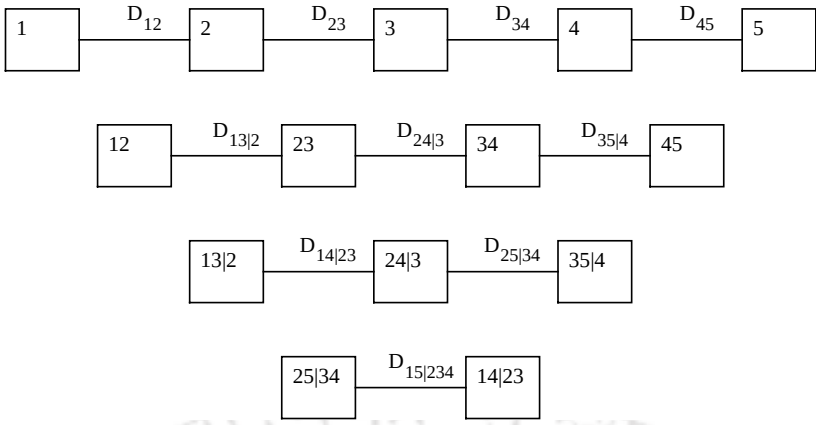


Figure 5.2: A D-vine in five dimensions.

sources in literature are summarized in Table 5.1.

5.3 Multivariate distribution

In theory, multivariate joint probability density function (PDF) for a random vector $\boldsymbol{X} = (X_1, X_2, ...X_n)$ can be constructed using the chain-rule as follows:

$$f_{X_1,X_2,...X_n}(x_1, x_2, ..., x_n) = f_{X_1}(x_1) f_{X_2|X_1}(x_2|x_1) f_{X_3|X_2,X_1}(x_3|x_2, x_1).....\\ \times f_{X_n|X_{n-1},..,X_1}(x_n|x_{n-1}, .., x_1) \tag{5.1}$$

Dropping the subscript notation of the random variable X_i for simplicity and for $i = 1 : n, n = 5$ the above expression can be written as follows:

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_1) f(x_2|x_1) f(x_3|x_2, x_1) \times f(x_4|x_3, x_2, x_1) f(x_5|x_4, x_2, x_3, x_1) \tag{5.2}$$

The above expression is not unique, and the variables can be re-ordered to get various other forms of expression. Assume an alternate form as follows:

$$f(x_1, x_2, x_3, x_4, x_5) = f(x_2, x_3, x_1, x_4, x_5) = f(x_2) f(x_3|x_2) f(x_1|x_3, x_2)\\ \times f(x_4|x_1, x_3, x_2) f(x_5|x_4, x_1, x_3, x_2) \tag{5.3}$$

It can be noted that the construction of the above five-dimensional joint distribution requires evaluation of apparently four conditional PDFs. However, the evaluation of conditional PDFs, especially in higher dimensions (> 2) is not a trivial task (Ching and Phoon, 2014). Most popular route is to use multivariate normal distribution (Ching et al., 2017; Ching, Phoon, Li and Weng,

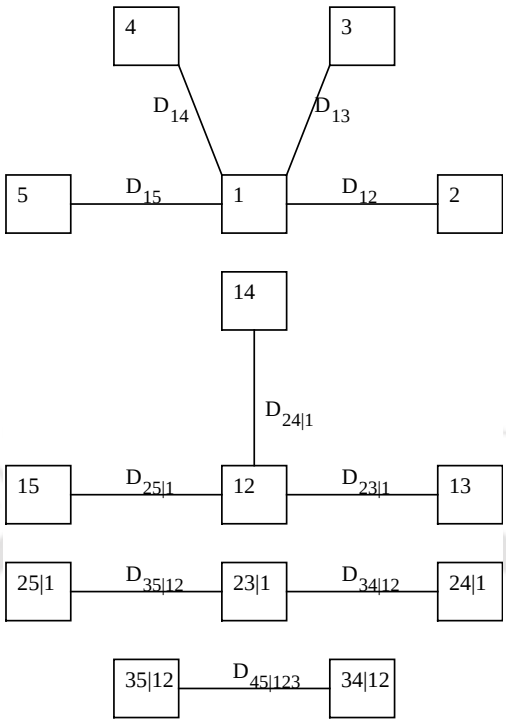


Figure 5.3: A C-vine in five dimensions.

2018; Ching and Phoon, 2014) since all the conditional distributions are also Gaussian. Although the multivariate normal distribution makes the computational task simple, there are also limitations associated with it. For example, (a) dependence structure and marginal behaviour are linked to each other (b) dependence structure is elliptical and symmetric, i.e. the tail dependence cannot be accounted for.

To circumvent these limitations, copula approach is popularly used (Wu, 2013*b*; Li et al., 2015; Tang et al., 2015; Zhang et al., 2018; Prakash et al., 2018*a*, e.g.). It makes the task of modelling dependence and marginals separate, hence greatly increases the flexibility in modelling joint distribution. However, it should be noted that most of the applications are only for bivariate cases, i.e. $n = 2$. A wide variety of parametric copula families to model various dependence structures are available. The theory regarding inference and simulation is also well established (Aas et al., 2009). However, for $n > 2$, the choice of copulas and flexibility gets reduced drastically. There are very few copulas that can be generalized for multivariate cases e.g. elliptical Gaussian and t-copulas, symmetric and non-symmetric Archimedian copula (Wu, 2013*b*, e.g.). However, these two copulas also have their limitations regarding the dependence and tail structure (Zhang and Singh, 2007).

Since the theory and inference regarding bivariate copulas are well established and also the choice of copula is abundant, this aspect was exploited to construct multivariate distribution in higher dimensions known as pair copula constructions (Aas et al., 2009). The graphical representation of the construction is done using a tree-like structure called *vine*; hence the name *vine copula*. This approach utilizes only bivariate copulas over transformed and non-transformed $[0, 1]^2$ data.

Table 5.3: Marginal selection for the parameters in the CLAY/C-S/5/278 database.

Database Parameter	Distribution Parameter	Distribution AIC values					
		TN*	TLN*	LN	WB	GM	EX
LL	4.44,0.89	3281.5	NA	3196.4	3279.5	3279.5	3277.5
PI	3.79,1.24	3115.5	NA	3017.1	3095.4	3095.4	3093.4
CF	4.06,0.54	2448.6	2448.9**	–	–	–	–
CEC	3.71,1.11	2599.5	2590.7	–	–	–	–
SSA	5.51,1.21	3596.7	3579.6	–	–	–	–

Values in bold denote best-fit marginal according to minimum AIC, TN* denotes left truncated normal for LL and PI and double truncated for rest, TLN* denotes double truncated lognormal, LN denotes usual lognormal,WB denotes weibull, GM denotes gamma and EX denotes exponential

Therefore, the theory of bivariate copula is very briefly discussed next. According to Sklar (1959) theorem any multivariate distribution can be expressed in terms of its marginals and a copula. For the bivariate case, it can be expressed as:

$$F_{X_1,X_2}(x_1,x_2) = C_{12}[F_{X_1}(x_1), F_{X_2}(x_2); \theta] \tag{5.4}$$

where $F_{X_1,X_2}(x_1,x_2)$ is the joint cumulative density function (CDF), $F_{X_1}(x_1), F_{X_2}(x_2)$ are the marginal CDFs and C_{12} is the copula function with parameter θ. Differentiating the above, the joint PDF $f_{X_1,X_2}(x_1,x_2)$ can be obtained as:

$$f_{X_1,X_2}(x_1,x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1 x_2} = D_{12}[F_{X_1}(x_1), F_{X_2}(x_2)] f_{X_1}(x_1) f_{X_2}(x_2) \tag{5.5}$$

where $f_{X_1}(x_1)f_{X_2}(x_2)$ are the marginal PDFs and D_{12} is the copula density function defined as

$$D_{12} = \partial^2 C_{12}[F_{X_1}(x_1), F_{X_2}(x_2); \theta] / \partial F_{X_1}(x_1) \partial F_{X_2}(x_2) \tag{5.6}$$

Eq. 5.5 is the key element which is exploited for the pair copula construction. Now the second term in Eq. 5.3 can be written as

$$f(x_3|x_2) = \frac{f(x_3,x_2)}{f(x_2)} = D_{32}[F(x_3), F(x_2)] f(x_3) \tag{5.7}$$

The subsequent term $f(x_1|x_3,x_2)$ in Eq. 5.3 can further be expressed as

$$f(x_1|x_3,x_2) = \frac{f(x_1,x_3,x_2)}{f(x_3,x_2)} = \frac{f(x_1,x_3|x_2)f(x_2)}{f(x_3|x_2)f(x_2)} = \frac{f(x_1,x_3|x_2)}{f(x_3|x_2)} \tag{5.8}$$

Utilizing Eq. 5.5 the above terms can be expressed in terms of marginals and copulas as

$$f(x_1|x_3,x_2) = \frac{f(x_1,x_3|x_2)}{f(x_3|x_2)} = D_{13|2}[F(x_1|x_2),F(x_3|x_2)]f(x_1|x_2) \tag{5.9}$$

Further breaking down the term $f(x_1|x_2)$ similar to Eq. 5.7, the above term can be expressed as

$$f(x_1|x_3,x_2) = D_{13|2}[F(x_1|x_2),F(x_3|x_2)] \times D_{12}[F(x_1),F(x_2)]f(x_1) \tag{5.10}$$

Similar recursive factorization of conditional density into product of copula densities and marginals, the subsequent term $f(x_4|x_1,x_3,x_2)$ in Eq. 5.3 can be obtained as (not shown here for brevity).

$$\begin{aligned} f(x_4|x_1,x_3,x_2) = D_{14|23}[F(x_1|x_2,x_3),F(x_4|x_2,x_3)] \\ \times D_{24|3}[F(x_2|x_3),F(x_4|x_3)] \times D_{34}[F(x_3),F(x_4)]f(x_4) \end{aligned} \tag{5.11}$$

Similarly the last term $f(x_5|x_4,x_1,x_3,x_2)$ in Eq. 5.3 can be expressed as

$$\begin{aligned} f(x_5|x_4,x_1,x_3,x_2) = D_{15|234}[F(x_1|x_2,x_3,x_4),F(x_5|x_2,x_3,x_4)] \\ \times D_{25|34}[F(x_2|x_3,x_4),F(x_5|x_3,x_4)] \times D_{35|4}[F(x_3|x_4),F(x_5|x_4)] \times D_{45}[F(x_4),F(x_5)]f(x_5) \end{aligned} \tag{5.12}$$

Substituting all the conditional distributions in Eq. 5.3, an expression for the 5-dimensional joint distribution in terms of copula and its marginals can be obtained as follows:

$$\begin{aligned} f(x_1,x_2,x_3,x_4,x_5) = f(x_1)f(x_2)f(x_3)f(x_4)f(x_5) \times D_{12}(F(x_1),F(x_2))D_{23}(F(x_2),F(x_3)) \\ \times D_{34}(F(x_3),F(x_4))D_{45}(F(x_4),F(x_5)) \times D_{13|2}(F(x_1|x_2),F(x_3|x_2)) \\ \times D_{24|3}(F(x_2|x_3),F(x_4|x_3)) \times D_{35|4}(F(x_3|x_4),F(x_5|x_4)) \times D_{14|23}(F(x_1|x_2,x_3),F(x_4|x_2,x_3)) \\ \times D_{25|34}(F(x_2|x_3,x_4),F(x_5|x_3,x_4)) \times D_{15|234}(F(x_1|x_2,x_3,x_4),F(x_5|x_2,x_3,x_4)) \end{aligned} \tag{5.13}$$

The above formulation can be categorized using a graphical structure as drawable (D) vine and is shown in Fig. 5.2. An alternate factorization of Eq. 5.3 leads to a different form of expression as follows:

Table 5.4: Measured and simulated univariate statistics for the parameters in the CLAY/C-S/5/278 database.

Database Parameter		Mean	Standard deviation	Skewness	Kurtosis	05% Q	25%Q	50%Q	75%Q	95%Q
LL	M	133.38	147.88	1.77	5.00	27.00	45.00	68.00	130.00	483.40
	S	126.86	138.98	4.58	52.88	19.65	47.00	85.47	156.05	367.90
PI	M	98.52	138.92	1.88	5.39	7.76	20.00	37.36	73.00	446.84
	S	96.82	185.18	11.91	425.55	5.77	19.36	44.71	103.94	347.01
CF	M	53.74	20.02	0.13	2.44	21.00	39.00	53.70	68.00	87.60
	S	54.08	21.05	0.25	2.18	22.73	37.45	52.01	69.66	91.58
CEC	M	42.10	30.81	0.98	3.26	6.64	18.50	31.80	62.00	103.20
	S	42.88	32.33	0.94	3.04	6.17	17.08	33.63	61.76	110.90
SSA	M	249.63	196.17	0.94	2.95	30.40	93.00	189.50	355.00	646.60
	S	250.15	195.61	0.94	2.98	31.18	92.91	192.87	364.84	660.68

M and S denote measured and simulated respectively, Q denotes quantile

$$
\begin{aligned}
f(x_1,x_2,x_3,x_4,x_5) = f(x_1)f(x_2)f(x_3)f(x_4)f(x_5) \times D_{12}[F(x_1),F(x_2)]D_{13}[F(x_1),F(x_3)] \\
\times D_{14}[F(x_1),F(x_4)]D_{15}[F(x_1),F(x_5)]D_{23|1}[F(x_2|x_1),F(x_3|x_1)]D_{24|1}[F(x_2|x_1),F(x_4|x_1)] \\
\times D_{25|1}[F(x_2|x_1),F(x_5|x_1)] \times D_{34|12}[F(x_3|x_1,x_2),F(x_4,x_1,x_2)] \\
\times D_{35|12}[F(x_3|x_1,x_2),F(x_5|x_1,x_2)]D_{45|123}[F(x_4|x_1,x_2,x_3),F(x_5|x_1,x_2,x_3)]
\end{aligned}
\tag{5.14}
$$

The above formulation can be categorized using a graphical structure as canonical (C) vine and is shown in Fig. 5.3. For the five-dimensional case, apart from the two forms in Eqs. 5.13 and 5.14 there are 238 other ways (in total 240) in which Eq. 5.3 can be factorized. Out of these 240 structures, 60 can be classified as D-vine, 60 as C-vine and rest of 120 as other regular vines. It can be noted that vines are just a graphical representation to help organize these various forms of combinations and are independent of copula theory.

It can also be noted that the C and D vines are just a special case of regular vines. Each regular vine in n-dimensions consists of $n-1$ trees (T_i) and each tree is constructed using nodes and edges. Each of the edges is associated with a pair(bivariate) copula. For $n = 5$ dimensional case in this study each vine structure contains a total of $n(n-1)/2$ edges hence 10 copulas. C-vine resembles a star-shaped structure wherein every tree (T_i) contains a base node connected to each of the $n-j$ edges (Aas et al., 2009). D-vine resembles a parallel stack like structure, and every node in any tree can be connected to only two edges at maximum. This study utilizes D-vine structure for the construction of joint density as it is the simplest to visualize and implement. The joint density for a D-vine (e.g. Eq. 5.13) can be generalized (Bedford and Cooke, 2001; Aas et al., 2009) as follows

Table 5.5: Comparison of measured and simulated univariate statistics for LL and PI using bimodal and unimodal lognormal distribution.

Statistics	Database parameter					
	LL ($P = 0.80, \mu_1 = 4.07, \sigma_1 = 0.50, \mu_2 = 5.97, \sigma_2 = 0.29$)			PI ($P = 0.80, \mu_1 = 3.34, \sigma_1 = 0.86, \mu_2 = 5.84, \sigma_2 = 0.32$)		
	M	S2	S1	M	S2	S1
Mean	133.38	133.40	126.86	98.52	98.81	96.82
Standard deviation	147.88	148.91	138.98	138.92	140.23	185.18
Skewness	1.77	1.90	4.58	1.88	2.03	11.91
Kurtosis	5.00	5.94	52.88	5.39	6.54	425.55
05% Q	27.00	27.12	19.65	7.76	7.10	5.77
25% Q	45.00	45.82	47.00	20.00	19.54	19.36
50% Q	68.00	68.56	85.47	37.36	37.77	44.71
75%Q	130.00	124.54	156.05	73.00	77.18	103.94
95%Q	483.40	472.68	367.90	446.84	420.06	347.01

M and S denote measured and simulated respectively, Q denotes quantile
S2 and S1 denotes bimodal and unimodal lognormal distribution

$$f(x_1, x_2, ..., x_k) = \prod_{k=1}^{n} f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} D_{i,i+j|i+1,...,i+j-1} \times [F(x_i|x_{i+1}, ...x_{i+j-1}), F(x_{i+j}|x_{i+1}, ...x_{i+j-1})] \tag{5.15}$$

where j denotes the index for Tree and i for edges in a particular T_j.

For estimating parameters of the pair copulas pseudo-likelihood approach is most popular (Aas et al., 2009), hence utilized in this study

$$\theta_n = \underset{\theta}{\operatorname{argmax}}\, L(\theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} log[D(U_i1, U_i2; \theta)] \tag{5.16}$$

Where U_ij are the pseudo observation or rescaled empirical distribution of $X_i = (X_i1, X_i2)$ and can be expressed as

$$U_ij = \frac{R_ij}{n+1} \tag{5.17}$$

Where R_ij is the rank of X_ij among (X_1j, X_2j) when sorted in ascending order. For T_1 of the D-vine structure in Fig. 5.2 the copula is estimated over the pseudo-observations obtained from original variables. As example, for the first edge in T_1 the copula parameter for D_{12} is evaluated over variables $(1, 2)$ and for D_{23}, it is evaluated over the variables $(2, 3)$ and so on. For T_2 onwards the copula density is evaluated over conditional pseudo observation As example, for T_2, $D_{13|2}$ is evaluated over $(1|2, 3|2)$ and $D_{24|3}$ is evaluated over $(2|3, 4|3)$ and so on. These conditional pseudo-

Table 5.6: Copula selection for the CLAY/C-S/5/278 database corresponding to the D-vine structure in Fig. 5.2.

Tree	Copula	Parameter	Copula AIC values			
			N	C	F	G
1	12	7.90	-768.41	-527.29	-750.71	-895.21
	23	5.92	-173.39	-123.126	-181.19	-155.60
	34	3.20	-50.997	-39.122	-64.617	-50.42
	45	7.83	-242.78	-221.561	-261.78	-202.11
2	13—2	0.27	-15.445	-17.892	-14.716	-6.629
	24—3	0.34	-30.565	-18.668	-28.885	-9.997
	35—4	0.39	-43.491	-33.059	-32.844	-42.339
3	14—23	0.24	1.998	2.001	1.948	2.006
	25—34	1.47	-66.108	-11.243	-66.813	-81.804
4	15—234	1.37	-9.416	-0.27	-18.838	-5.215

Values in bold refer to preferred copula according to minimum AIC, N denotes Normal(Gaussian),C denotes Clayton,F denotes Frank and G denotes Gumbel

observation can be obtained using conditional derivative function (also called h-function) of the copula function C_{12}as follows

$$U_i1|U_i2 = h(U_i1, U_j2; \theta_{12}) = \frac{\partial C_{U_i1,U_i2}(U_i1, U_i2; \theta_{12})}{\partial U_i2} \tag{5.18}$$

The h functions for all the copulas utilized in this study are summarized in Table DA2. For example, $U_i1|U_i2$ can be easily obtained using $U_i1|U_i2 = h_{12}(U_i1, U_i2; \theta_{12})$.

From T_3 onwards the conditioning variable (e.g. U_2 above) will not remain univariate but multivariate. As example for T_3, $D_{14|23}$ is evaluated over $(1|23, 4|23)$. As example, $U_i1|U_i2, U_i3$ can be obtained as, $U_i1|U_i2, U_i3 = h_{13|2}[h_{12}(U_i1|U_i2; \theta_{12}), h(U_i3|U_i2; \theta_{32}); \theta_{13|2}]$. The required conditional data for parameter estimation in n dimensions can be obtained using a more general form of Eq. 5.18 (Joe, 1997) as follows

$$F_{X|\boldsymbol{Y}}(x|\boldsymbol{y}) = \frac{\partial C_{X,Y_j|\boldsymbol{Y_{-j}}}[F_{X|\boldsymbol{Y_j}}(x|\boldsymbol{y_{-j}}), F_{Y_j}|\boldsymbol{Y_j}(y|\boldsymbol{y_{-j}})]}{\partial F_{Y_j|\boldsymbol{y_{-j}}}(y|\boldsymbol{y_{-j}})} \tag{5.19}$$

where X is a random variable, Y_j is a element of random vector $\boldsymbol{Y}$ and $\boldsymbol{Y_{-j}}$ is vector consisting of all components of $\boldsymbol{Y}$ but Y_j.

For choosing an appropriate copula family, goodness of fit test such as AIC can be used. AIC can be expressed as

$$AIC = -2L(\theta_n) + 2(k) \tag{5.20}$$

Where $L(\theta_n)$ is the maximized log likelihood in Eq. 5.16 and k is the number of parameters for

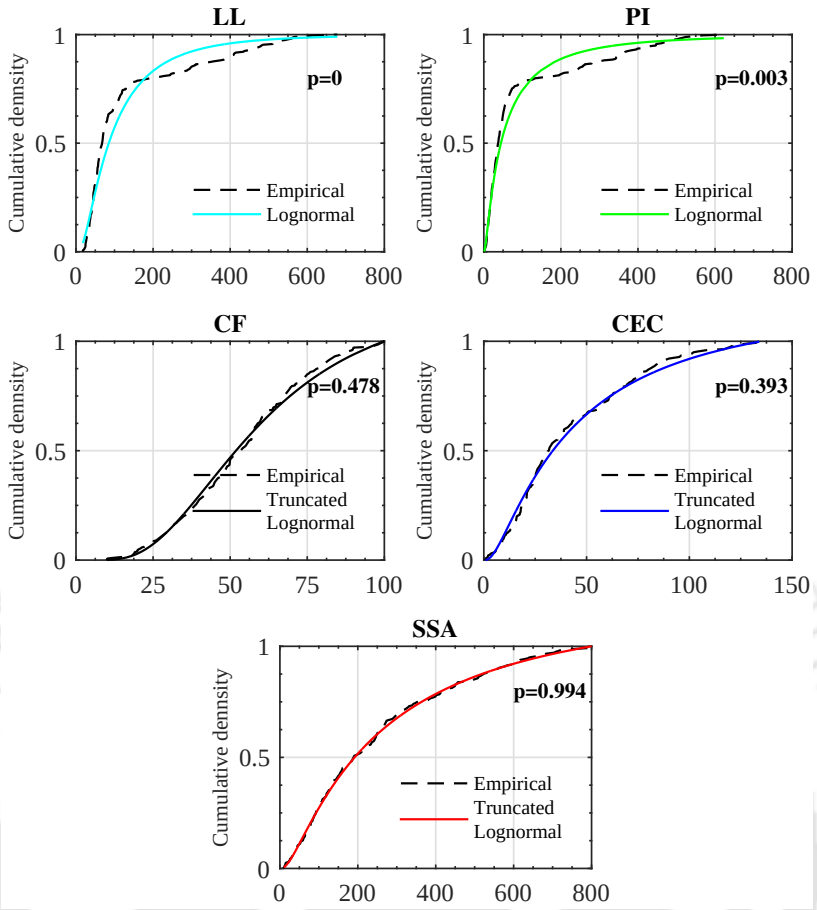


Figure 5.4: Empirical and calculated CDF for parameters in the CLAY/C-S/5/278 database.

the copula.

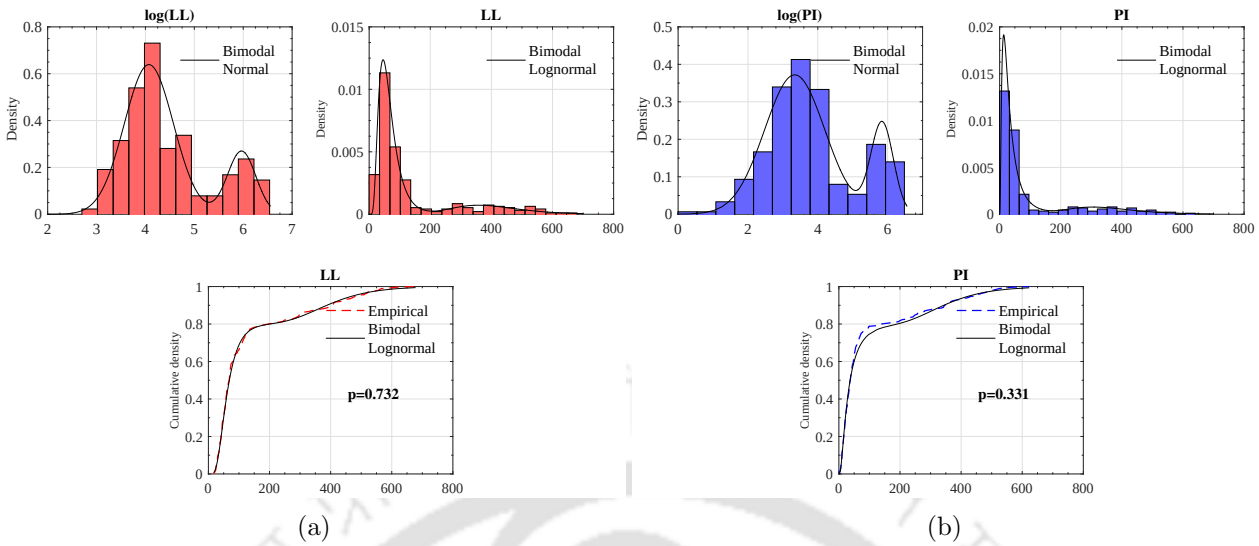
Once the appropriate copula family using AIC is chosen with parameters Θ, 5-dimensional copula random numbers can be simulated using the following procedure.

For i=1: Number of simulations

1. Generate a random sample $r = (r_i1, r_i2, r_i3, r_i4, r_i5)$ from independent uniform distribution
2. $U_i1 = r_i1$
3. $U_i2 = C^{-1}_{2|1}(r_i2|U_i1)$
4. $U_i3 = C^{-1}_{3|12}(r_i3|U_i, U_i2)$
5. $U_i4 = C^{-1}_{4|123}(r_i4|U_i, U_i2, U_i3)$
6. $U_i5 = C^{-1}_{5|1234}(r_i5|U_i, U_i2, U_i3, U_i4)$

End

The conditional distributions $C_{j|j-1,..,1}$ required for simulating the random numbers can be ob-



(a) (b)

Figure 5.5: Empirical and calculated PDF and CDF for bimodal lognormal distribution of (a) LL and (b) PI.

tained recursively using the h and h^{-1} functions along with recursive formula given in Eq. 5.19. For example, U_i2 can be obtained as $U_i2 = h_{21}^{-1}(r_i2, U_i1; \theta_{12})$. Similarly U_i3 can be obtained as $U_i3 = h_{32}^{-1}\{h_{13|2}^{1}[r_i3, h_{12}(U_i1, U_i2; \theta_{12}); \theta_{13|2}], U_i2; \theta_{32}\}$. All the h^{-1} function along with h functions for all the four copulas utilized in this study are summarized in Table DA2. Detailed algorithms for efficient implementation of the above procedure in an arbitrary n dimension can be found in Aas et al. (2009).

It is also important to note here (many times not mentioned) that all the aforementioned formulations invoke a key assumption of conditional independence also known as simplifying assumption (Aas et al., 2009). This assumption essentially implies that all the pair copulas for conditional distribution or its inverse do not depend upon the conditioning variable. For example, $C_{23|1}[F(x_2|x_1), F(x_3, x_1); x_1] = C_{23|1}[F(x_2|x_1), F(x_3, x_1)]$. This assumption is very popular as it reduces the vine copula approach only into the identification of pair copulas associated with edges and its parameters. However, at the same time, it also implies that possible dependence occurring among conditional variables and the copula are neglected (Stoeber et al., 2013).

5.4 Results and Discussion

Marginal selection is essential for the construction of a parametric multivariate distribution. Since all the five CLAY/C-S/5/278 parameters are theoretically defined as positive values, five common distributions defined over positive values are considered. These are truncated normal (left truncated at zero), lognormal, Weibull, gamma and exponential. In addition to being positive, CF, CEC and SSA do also have theoretical/practical upper bounds. Therefore two additional distributions, namely double truncated normal and right truncated lognormal, are also explored. For all these distributions, functional forms of PDF, CDF and its inverse are summarized in Table DA1.

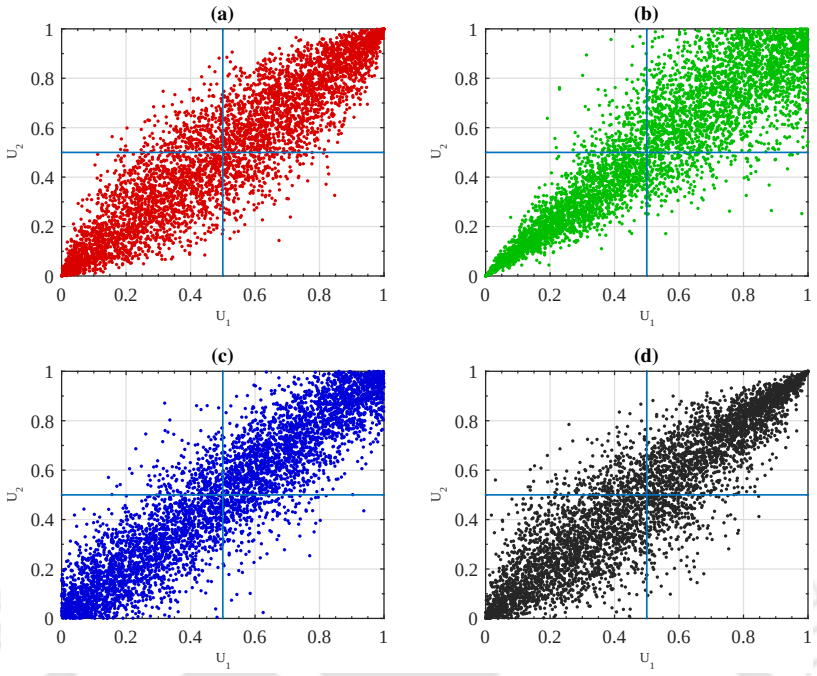


Figure 5.6: Scatter plot of (a) Gaussian (b) Clayton (c) Frank and (d) Gumbel copula for Kendall's $\tau = 0.75$.

Table 5.3 presents the AIC values calculated over the database parameters for all the considered distributions. The distributions selected according to the minimum AIC are highlighted in bold. The parameters presented are for the selected distribution. It can be noted that lognormal, Weibull, gamma and exponential are not evaluated for CF, CEC and SSA. These three distributions are unbounded on the right, and there is significant empirical density for CF, CEC and SSA in the right tail. Therefore the use of these three distributions during simulations would result in numerous values well above the theoretical limits and hence may produce unrealistic results. Similarly truncated lognormal was not evaluated on LL and PI since there are no theoretically defined upper bounds for both these two and also the probability density in the right tail is low.

In Table 5.3, it can be noted that while lognormal is the best candidate (among the considered) for LL and PI, truncated lognormal is suitable for the rest of the three parameters. It can also be noted that for CF, the observed AIC for double truncated normal and truncated lognormal are almost the same. Since for all others, lognormal (truncated or otherwise) is the preferred distribution, the same is also chosen here for consistency.

To check if the univariate marginals are modelled effectively, a comparison of moments and quantiles is performed, and the results are presented in Table 5.4. For parameters, CF, CEC and SSA, it can be noted that all the four statistical moments mean, standard deviation, skewness and kurtosis are in close agreement. The quartile values are also very close to each other. This essentially implies that right truncated lognormal can be successfully utilized to model the variation of these bounded parameters.

However, for LL and PI, it can be noted that there is a large discrepancy between the measured and simulated values. Therefore to confirm if the preferred distribution, according to AIC, is actually suitable, Kolmogorov-Smirnov (K-S) test is also performed. The K-S test results are presented in Fig. 5.4. It can be noted that except for LL and PI, $p > 0.05$ satisfies for the rest of the parameters.

It essentially implies that right truncated lognormal distribution is suitable for CF, CEC and PI but lognormal distribution is not suitable for LL and PI. Although AIC results indicated lognormal as the best candidate for LL and PI, it should be noted that AIC is only a relative measure of goodness of fit. The best family is selected only, if it is included in the candidate choice. It can also be noted that the empirical CDFs for LL and PI have a slight flat portion, pointing to a bimodal behaviour. To further investigate this behaviour, histograms of LL and PI are presented in Figs. 5.5a and 5.5b, respectively. It can be clearly observed that the empirical pdfs display a bimodal behaviour, hence bimodal lognormal (not considered earlier in this study) was explored as a candidate model.

The resulting fit to the PDF and CDF are also presented in Figs. 5.5a and 5.5b respectively. The corresponding p values for the K-S test statistics are also presented, and it can be noted that now the values are well above 0.05 and hence statistically can be deemed suitable for modelling LL and PI. The parameters of the bimodal lognormal distribution for LL and PI are summarized separately in Table 5.5. Theoretically, this bimodal behaviour essentially implies that LL and PI behaviour exhibit different mechanisms at low (< 200) and high values (> 200). This behaviour can also be justified, theoretically. At higher LL and hence higher PI, most likely, the dominant mineral is montmorillonite, and for lower values, the dominating mineral is most likely kaolinite. The behaviour of these two minerals is widely recognized to be different (Mitchell et al., 2005; Lu and Mitchell, 2019), hence these two form a separate cluster or peak in the histogram.

For comparison of marginal statistics for LL and PI using bimodal and unimodal lognormal distribution (deemed unsatisfactory based on the K-S test in Fig. 5.4) the measured and simulated results are presented separately in Table 5.5. It can now be observed that the measured and simulated statistics using bimodal lognormal marginals are in close agreement with each other. This is a marked improvement with the commonly used lognormal distribution, and hence the use of bimodal lognormal marginals (although uncommon) is justified.

For constructing joint distribution using vine copula theory, the other task after modelling marginals of CLAY/C-S/5/278 parameters is to model the dependence among them. For this purpose, four commonly used copulas with different dependence structures were utilized. These are Gaussian, Clayton, Frank and Gumbel. The functional form of copula functions along with their densities are summarized in Table DA2. It should be noted that the different families of copulas do not differ much in overall correlation but only in the distribution of correlation in the quadrants. For example, Figs. 5.6 presents the random number plots for these four copulas with parameters corresponding to the same $\tau = 0.75$. It can be noted that while the correlation structure for Gaussian is elliptical and evenly distributed in the upper right and lower quadrant, i.e.

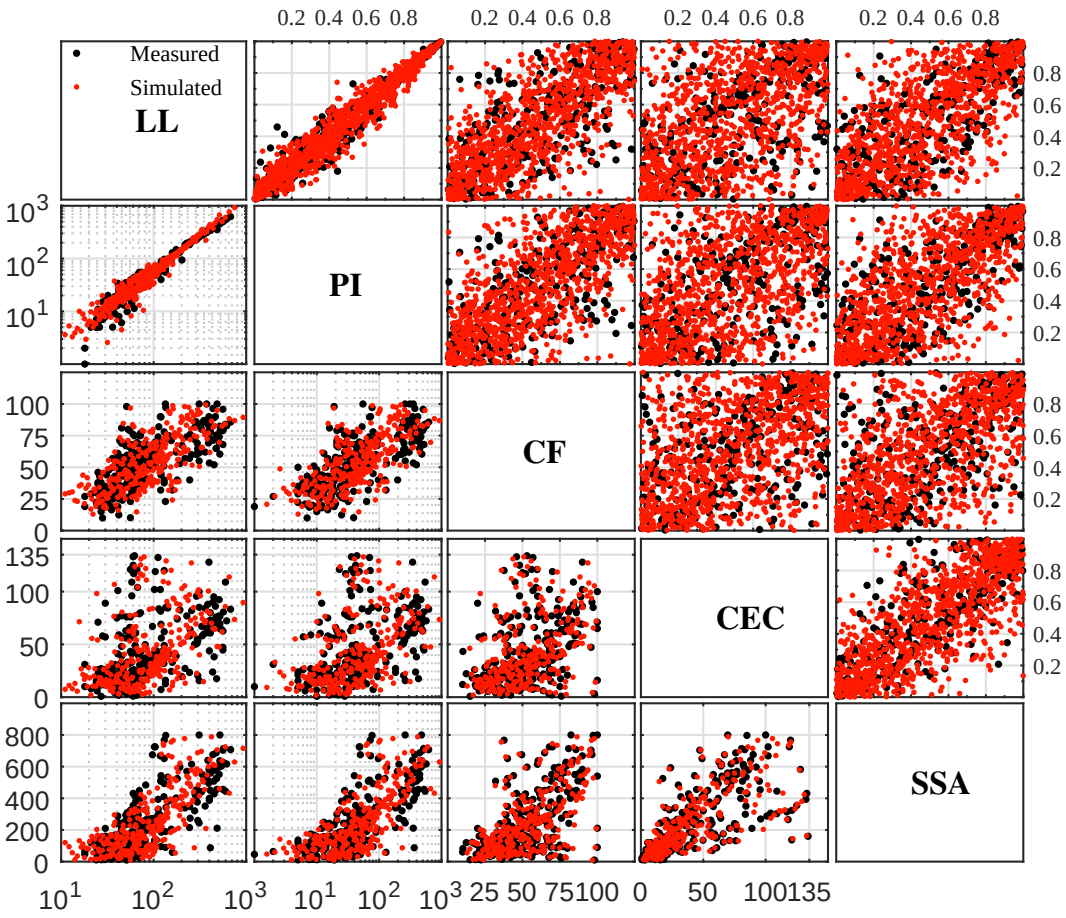


Figure 5.7: Measured and simulated parameters in the CLAY/C-S/5/278 database on copula ($[0,1]^2$) scale (upper triangular matrix) and original scale (lower triangular matrix).

symmetric, whereas for frank copula the distribution is symmetric but not elliptical. For Clayton copula, it can be observed that the strength of the association is stronger in the lower left quadrant and comparatively lower in the upper right. For Gumbel copula also the correlation is stronger in upper right quadrant than in the lower left. Similar to the univariate marginal selection, AIC is also used for selection of appropriate families for all the pair copulas in D-vine structure shown in Fig. 5.2. Table 5.6 presents the AIC values for all the four candidate copulas. The preferred copula with minimum AIC is highlighted in bold. It can be noted that most commonly used Gaussian copula can adequately represent the dependence in only 2 out of 10 cases. This essentially implies that the dependence structure for all the pairs is not symmetric and elliptical.

To check if the joint distribution constructed using the aforementioned marginals and copulas are modelled satisfactorily, random numbers are simulated, and statistics obtained henceforth are compared against the measured ones. Fig. 5.7 presents the measured and simulated parameters on copula scale [0,1] and the original scale in the upper and lower triangular matrix, respectively. It can be visually observed that in Fig. 5.7 on the unit scale that the measured and simulated random numbers closely resemble the original correlation structure of the measured data. To confirm it

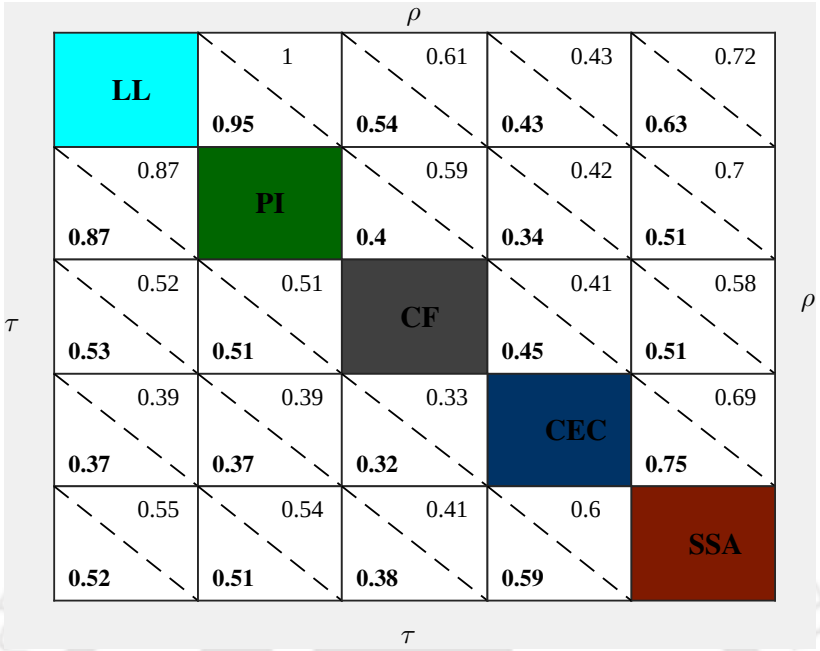


Figure 5.8: Measured (bold) and simulated rank correlation coefficient Kendall's τ and Pearson's correlation coefficient ρ in the CLAY/C-S/5/278 database.

statistically, Fig. 5.8 presents the measured and simulated rank correlation Kendall's τ. The measured ones are below the diagonal and highlighted in bold and simulated ones are above the diagonal. It can be noted that for all the ten pairs measured and simulated τ values are in close agreement. In addition to Kendall's τ, the conventional linear correlation coefficient ρ is also shown in Fig. 5.8. It can be noted that the measured and simulated ρ values are not as closely matched as for the case of Kendall's τ. This is not unexpected since ρ is only a measure of linear and Gaussian dependence structure and Kendall's τ is a better correlation measure for non-Gaussian and non-linear dependence structures. This essentially implies that even-though the dependence structure can be modelled using numerous vine structures, the D-vine structure mentioned in Fig. 5.2 can successfully be utilized for this CLAY/C-S/5/278 database.

The combined use of truncated and bimodal lognormal marginals to model the theoretical/practical constraints and behaviour could be facilitated readily due to the copula approach. Although the marginals could have still been modelled by translational approach (Ching, Li, Phoon and Weng, 2018, e.g.) the non-Gaussian dependence structure observed in 8 out of 10 pair copulas (ref. Table 5.6), could have posed difficulties in a translational approach.

5.5 Probabilistic estimation of CEC and SSA: Implementation of the multivariate joint distribution

This section demonstrates the implementation of the constructed 5-dimensional joint distribution for probabilistic estimation of CEC and SSA. The 5-dimensional joint distribution summarizes the

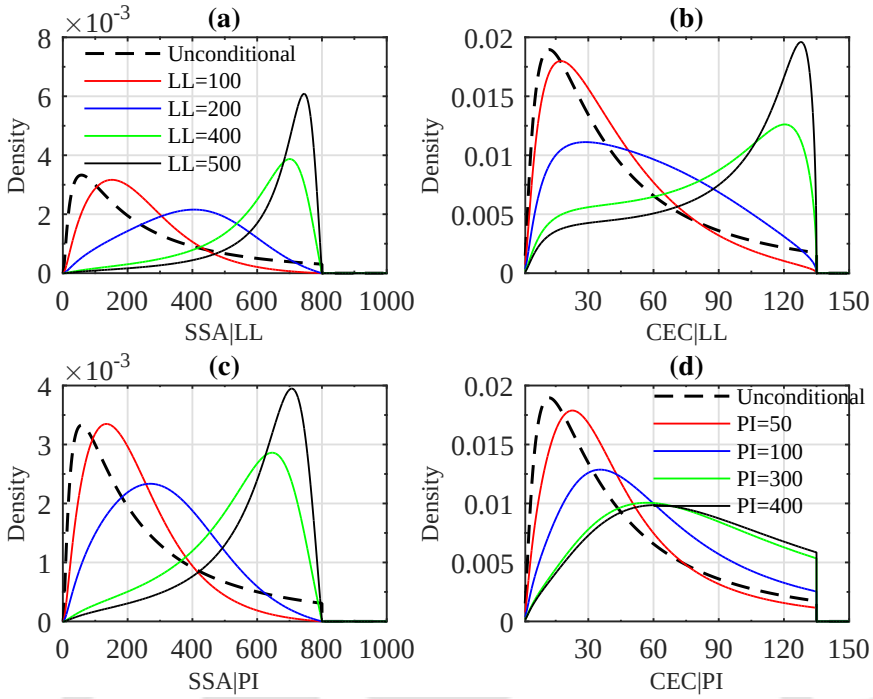


Figure 5.9: Unconditional and updated density for CEC and SSA given some other parameters.

marginals and dependence for all the parameters in the CLAY/C-S/5/278 database. Since the collected database is truly multivariate (278×5) wherein the data comes from various geographical regions of the globe, it can be considered as a global multivariate database. Hence the multivariate distribution constructed in this study can be viewed as a global or generic multivariate distribution which summarizes the global marginal behaviour and dependence among all the five parameters.

The global distribution of CEC and SSA acts as a prior/unconditional distribution which can be updated to get the posterior/conditional distribution in presence of soil/site-specific data (e.g. LL, PI, CF) (Ching, Phoon, Li and Weng, 2018; Ching and Phoon, 2014, e.g.).

The random vector $\boldsymbol{X} = (x_1, x_2, x_3, x_4, x_5)$ is used as notation for the 5 parameters in the CLAY/C-S/5/278 database LL, PI, CF, CEC, SSA. As an example, consider new data of LL (x_1), PI (x_2) and CF (x_3) for a soil sample is available. Thereafter, the joint distribution of CEC(x_4) and SSA(x_5) can be obtained given the observed data using Bayes' rule (Ching, Li, Phoon and Weng, 2018, e.g.) as follows

$$
\begin{aligned}
f(x_5, x_4|x_3, x_2, x_1) = \frac{f(x_1, x_2, x_3, x_4, x_5)}{\int\int\int f(x_1, x_2, x_3, x_4, x_5) dx_4 dx_5} = \frac{f(x_1, x_2, x_3, x_4, x_5)}{f(x_3, x_2, x_1)} \\
= f(x_4) f(x_5) \times D_{34}(F(x_3), F(x_4)) \times D_{45}(F(x_4), F(x_5)) \times D_{24|3}(F(x_2|x_3), F(x_4|x_3)) \\
\times D_{35|4}(F(x_3|x_4), F(x_5|x_4)) \times D_{14|23}(F(x_1|x_2, x_3), F(x_4|x_2, x_3)) \\
\times D_{25|34}(F(x_2|x_3, x_4), F(x_5|x_3, x_4)) \times D_{15|234}(F(x_1|x_2, x_3, x_4), F(x_5|x_2, x_3, x_4))
\end{aligned}
\tag{5.21}
$$

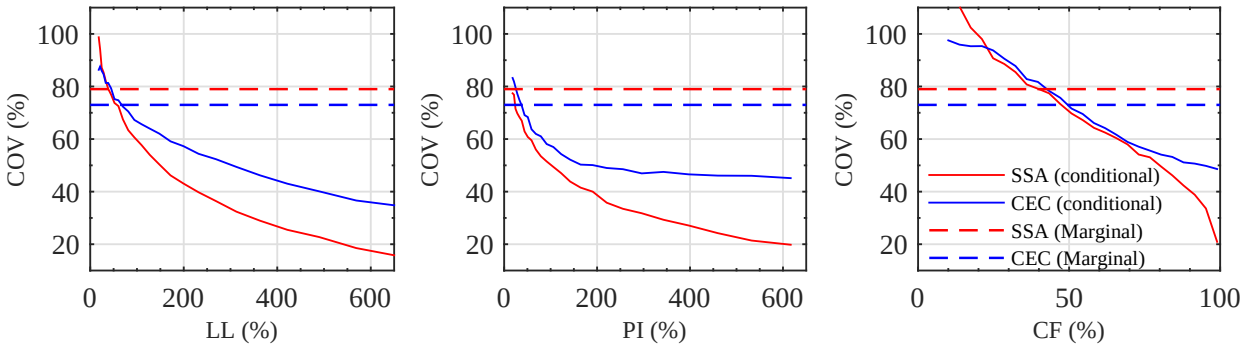


Figure 5.10: Uncertainty in terms of coefficient of variation (COV) of SSA and CEC before and after conditioning.

Alternatively, if the parameter of interest is only SSA (x_5) given some single parameter, e.g. $LL(x_1)$, it can also be obtained as follows

$$f(x_5|x_1) = \frac{\int\int\int f(x_1,x_2,x_3,x_4,x_5)dx_2dx_3dx_4}{\int\int\int\int f(x_1,x_2,x_3,x_4,x_5)dx_2dx_3dx_4dx_5} = \frac{f(x_5,x_1)}{f(x_1)} \\ = D_{51}[F(x_5),F(x_1)] \times f(x_5) \tag{5.22}$$

Fig. 5.9(a-b) shows the updated PDFs of SSA and CEC, respectively, after observing LL of a sample. The unconditional (global) PDF is also shown to demonstrate the essence of updating. The unconditional PDF can be thought of prior knowledge obtained from CLAY/C-S/5/278 database, and all the conditional distributions can be thought of the posterior distribution obtained after updating the prior based on observed data. For both SSA and CEC it can be easily noted that the PDFs shifts rightward with an increase in LL as expected due to positive correlation (ref. Fig. 5.8). It can also be noted that the imposed practical limits of SSA $=800m^2/g$ and CEC $= 135meq/100g$ are strictly adhered to, and the density beyond the limits is zero.

Usually the conditional PDFs exhibit lower uncertainty (i.e. narrower PDF) as compared to the unconditional PDF (Ching and Phoon, 2014). For example, in Fig. 5.9(a) conditional PDFs seem narrower than the unconditional PDF for SSA at LL= 100% and LL= 500%. However, the conditional PDF at LL = 200 seems more dispersed than the unconditional PDF. It is important to note here that a visual interpretation of uncertainty in terms of dispersion of PDFs is not appropriate in this case. This is because the central tendency measures, e.g. mode of the conditional PDFs vary significantly in the range of around $0 - 800\, m2/g$. Therefore statistical verification of uncertainty in estimates is needed in terms of COV. Fig. 5.10 presents the COVs of marginals and the conditional distribution of SSA and CEC at various LL, PI and CF values. It can be noted that the COVs for the conditional PDFs decrease with increasing LL, PI and CF values. It can also be noted that COVs for SSA is lower than CEC for LL and PI and almost similar for CF. This is because SSA exhibit a higher degree of correlation with LL and PI than CEC and similar for CF

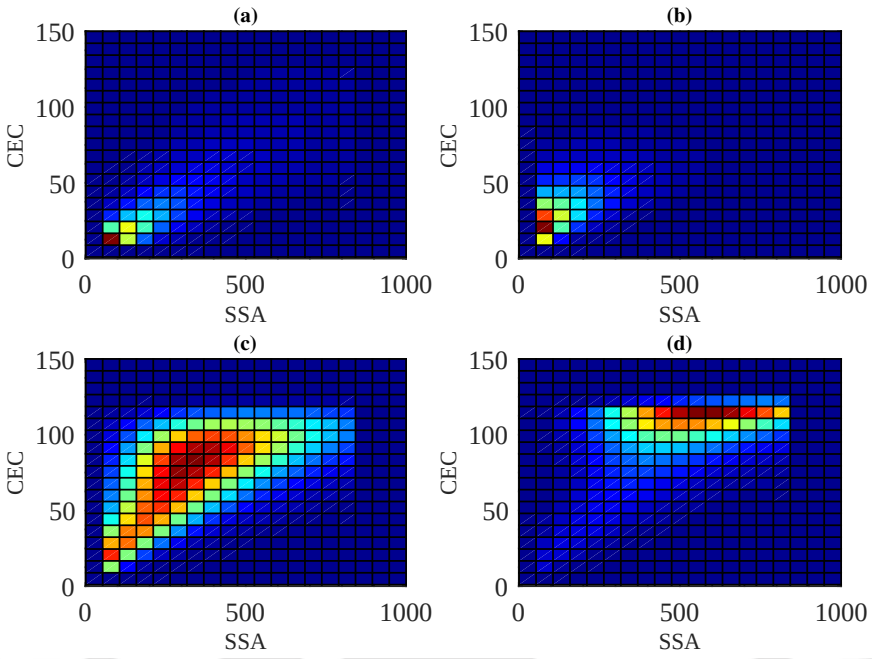


Figure 5.11: (a) Unconditional, and updated joint density of CEC and SSA at (b) $LL = 100, PI = 50, CF = 15$ (c) $LL = 300, PI = 200, CF = 45$ and (d) $LL = 500, PI = 400, CF = 75$.

(ref. Fig. 5.8).

It can also be noted that although the COVs have decreased after conditioning, the values are still large. For example, at LL and PI ≤ 200, the COV $\geq 40 - 60\%$. For CF, the COVs are even higher in general. For example, at CF ≤ 50 COVs $\geq 70\%$. The primary reason behind this can be attributed to the absence of the mineralogical information of clays. Had mineralogical data been available, the database could have been partitioned on the basis of mineralogy (Spagnoli et al., 2018, e.g.), and the correlations could have been stronger and subsequently, the conditional COVs could be lowered. The current database couldn't take into account of mineralogy as the multivariate data for mineralogical quantification are very scarce in the literature. This is mainly because obtaining the mineralogy is an even more challenging task than CEC and SSA. The difficulty can also be judged from the fact that CEC and SSA are also sometimes employed to estimate mineralogical data (Chittoori and Puppala, 2011, e.g.). It is also expected that in future, large scale availability of mineralogical data can be utilized so that the uncertainty in the estimates due to the difference in the behaviour of clay minerals, e.g., kaolinite and montmorillonite can be reduced. It should be noted that the probabilistic estimation of CEC and SSA in this study is achieved through conditional PDFs (ref. Fig. 5.9). The marginal (unconditional) PDFs of CEC and SSA alone cannot be used to obtain the probabilistic estimates of CEC and SSA given some measured data such as LL, PI, or CF. The marginal PDFs along with the joint density need to be updated/conditioned on the measured data (ref. Eq. 5.21) before they can be directly utilized.

The aforementioned examples showed only univariate predictions with respect to univariate conditioning variables. In the proposed approach, the joint distribution of CEC and SSA can also

Table 5.7: Expressions for conditional/posterior probability density functions (PDFs) of CEC and SSA.

Variable	Density	Expression	Copula	Family	Parameter	Marginal	Parameters
			Single parameter input				
SSA	$f(x_5\|x_1)$	$D_{15}f(x_5)$	D_{15}	Frank	6.4	$f(x_5)$	$\mu = 5.51$
	$f(x_5\|x_2)$	$D_{25}f(x_5)$	D_{25}	Frank	6.2	right truncated	$\sigma = 1.21$
	$f(x_5\|x_3)$	$D_{35}f(x_5)$	D_{35}	Gumbel	1.66	lognormal	$t = 800$
CEC	$f(x_4\|x_1)$	$D_{14}f(x_4)$	D_{14}	Frank	3.80	$f(x_4)$	$\mu = 3.71$
	$f(x_4\|x_2)$	$D_{24}f(x_4)$	D_{24}	Frank	3.83	right truncated	$\sigma = 1.11$
	$f(x_4\|x_3)$	$D_{34}f(x_4)$	D_{34}	Frank	3.2	lognormal	$t = 135$
			Two parameters input				
SSA	$f(x_5\|x_1, x_2)$	$D_{25}D_{15\|2}f(x_5)$	D_{25}	$\star$	$\star$	$f(x_3)$	$\mu = 4.06$
			$D_{15\|2}$	Frank	1.3	right truncated	$\sigma = 0.54$
	$f(x_5\|x_1, x_3)$	$D_{35}D_{15\|3}f(x_5)$	D_{35}	$\star$	$\star$	lognormal	$t = 100$
			$D_{15\|3}$	Gaussian	0.56		
	$f(x_5\|x_2, x_3)$	$D_{35}D_{25\|3}f(x_5)$	D_{35}	$\star$	$\star$	$f(x_2)$	$P = 0.80$
			$D_{25\|3}$	Gaussian	0.55	bimodal	$\mu_1 = 3.34$
CEC	$f(x_4\|x_1, x_2)$	$D_{24}D_{14\|2}f(x_4)$	D_{24}	$\star$	$\star$	lognormal	$\sigma_1 = 0.86$
			$D_{14\|2}$	Frank	0.36		$\mu_2 = 5.84$
	$f(x_4\|x_1, x_3)$	$D_{34}D_{14\|3}f(x_4)$	D_{34}	$\star$	$\star$		$\sigma_2 = 0.32$
			$D_{14\|3}$	Frank	2.08		
	$f(x_4\|x_2, x_3)$	$D_{34}D_{24\|3}f(x_4)$	D_{34}	$\star$	$\star$		
			$D_{24\|3}$	Gaussian	0.3		
			Three parameters input				
SSA	$f(x_5\|x_1, x_2, x_3)$	$D_{35}D_{25\|3}D_{15\|23}f(x_5)$	D_{35}	$\star$	$\star$		
			$D_{25\|3}$	$\star$	$\star$	$f(x_1)$	$P = 0.80$
			$D_{15\|23}$	Frank	1.20	bimodal	$\mu_1 = 4.07$
CEC	$f(x_4\|x_1, x_2, x_3)$	$D_{34}D_{24\|3}D_{14\|23}f(x_4)$	D_{34}	$\star$	$\star$	lognormal	$\sigma_1 = 0.50$
			$D_{24\|3}$	$\star$	$\star$		$\mu_2 = 5.97$
			$D_{14\|23}$	Frank	0.24		$\sigma_2 = 0.29$

x_5, x_4, x_3, x_2, x_1 denote SSA, CEC, CF, PI, LL, D_{ij} (copula density) are summarized in Table DA2
$f(x_i)$ (marginals) are summarized in Table DA1 and $\star$ denotes value already in table.

be obtained, given multiple conditioning variables. Fig. 5.11(b-d) presents the updated density plots of CEC and SSA, given various LL, PI and CF values, as examples. Unconditional joint distribution is also shown in 5.11(a). This particular case is of practical interest since LL, PI and CF can be obtained from conventional laboratory measurements and the *difficult-to-measure* properties CEC, and SSA can be obtained in the form of the joint distribution. It can be easily observed that joint distribution is updated as the high-density regions move consistently to higher values of CEC and SSA with an increase in LL values in line with the theoretical expectation. The theoretical limits are also complied with, for CEC and SSA simultaneously. All the expressions for conditional density of CEC and SSA can be obtained analytically and are summarized in Table 5.7. A practitioner can directly utilize the expressions in Table 5.7 to get the PDFs of CEC and SSA for a soil sample from a site or project when the measured data for LL, PI or CF are available. The expressions for copula functions and densities, marginal PDFs and CDFs are summarized separately in Table DA2 and Table DA1 respectively.

All of the above examples demonstrated the implementation of the proposed approach in this

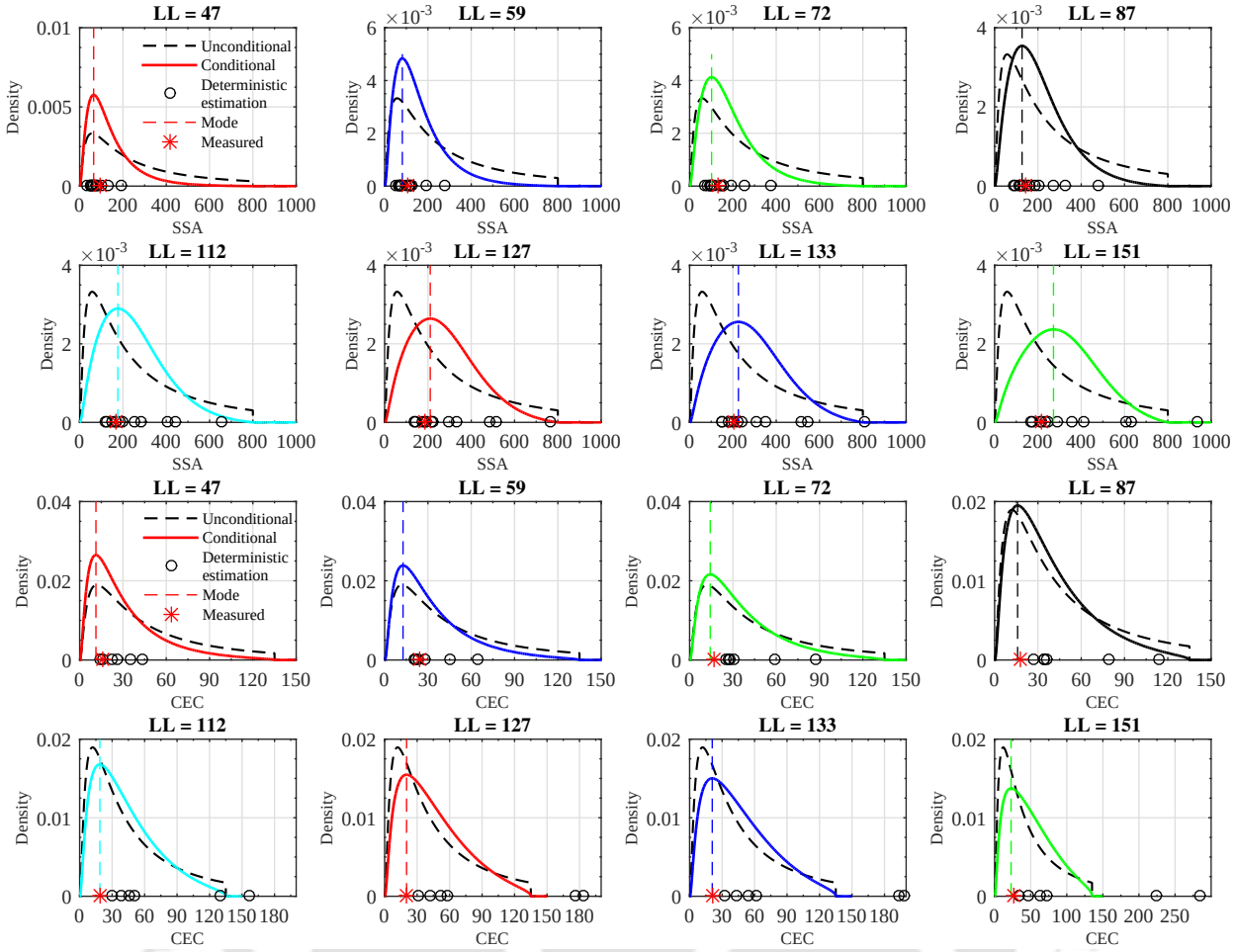


Figure 5.12: Comparison of updated/ conditional probability density function (PDF) of cation exchange capacity (CEC) and specific surface are (SSA) with measured soil property.

study for probabilistic estimation of CEC and SSA. However, the estimated PDFs are theoretically obtained using Bayes' rule(ref. Eq. 5.21). Therefore a natural question arises regarding the performance of theoretically obtained PDFs against some actual measured data outside the CLAY/C-S/5/278 database. To address this aspect, CEC and SSA were measured for eight soil samples along with the LL. CEC was measured using the ammonium replacement method (Chapman, 1965), SSA using the ethylene glycol monoethyl ether (EGME) retention procedure (Cerato and Lutenegger, 2002) and LL using ASTM D4318 ASTM (2010). The proposed approach was used to obtain the theoretical PDFs of CEC and SSA using measured LL. The obtained PDFs were compared against the actual measured CEC and SSA. Conventional transformation models (deterministic) in literature are also utilized for CEC and SSA. A total of 13 models for SSA and 6 for CEC were utilized for the comparison and are summarized separately in Table DA3. The results are presented in Fig. 5.12. It can be easily inferred that the mode of the PDFs (peak) for CEC and SSA are consistent with the actual measured data. It can also be noted that the application of various transformation models results in widely varying point estimates. This is because all these

models were calibrated on different datasets. Significant scatter in the deterministic estimates with respect to actual measured value can be noted. Since no quantification of uncertainty in the estimate is allowed, such estimates might be misleading. For example, at $LL = 151$, the estimates by conventional transformational models may result in $SSA > 900$ and $CEC > 200$. Such high point estimates of CEC and SSA at $LL = 151$ are not practical. However, it should also be noted the comparison is not in being done in a strict sense as a single measured value is compared against an estimated PDF. Still, the results in Fig. 5.12 demonstrate the applicability of the proposed approach in this study.

5.6 Some example geotechnical applications

In the last section, implementation of the proposed approach was demonstrated for probabilistic estimation of CEC and SSA. In this section, practical application of the probabilistic estimate towards two geotechnical problems will be demonstrated. The first example is related to the estimation of segregation potential (SP) for quantification of frost heave susceptibility in highway sub-grades, and the second example is related to the determination of swelling pressure (p) of clays at various dry densities (ρ_d). The framework utilized for estimation of SP_o is empirical and requires only SSA (out of SSA and CEC), whereas the framework for p is theoretical and requires both CEC and SSA.

5.6.1 Frost Heave susceptibility

Frost heave (h) in a highway sub-grade due to ice segregation is calculated as follows (Konrad, 1999, 2005)

$$h = 1.09 \int_0^t SP_o \exp^{aP_e} \nabla(T_f) dt \tag{5.23}$$

where SP_o is the segregation potential (SP) at zero overburden pressure P_e, a is an empirical parameter to account for P_e and $\nabla(T_f)$ is the temperature gradient in the frozen zone. Susceptibility of a soil towards frost heave is popularly quantified using SP_o (Konrad, 1999, 2005). Konrad (1999) proposed the prediction of SP_o utilizing SSA and the median particle size diameter d_{50}, for a highway project in Saints-Martyrs-Canadiens (Quebec province, Canada) as follows

$$SP_o \times SSA_{<75\mu m} = [116 - 75log(D_{50})] \times 10^3 mm^4/(^oC.s.g) \tag{5.24}$$

where $SSA_{<75\mu m}$ is the SSA of the fine fraction. The conventional SSA (and the database in this study) is measured on soil fraction $< 425\mu m$. Konrad (1999) proposed the estimation of $SSA_{<75\mu m}$ from $SSA_{<425\mu m}$ as follows

$$SSA_{<75\mu m} = SSA_{<425\mu m} \times (percent\ finer < 425\mu m / percent\ finer < 75\mu m) \tag{5.25}$$

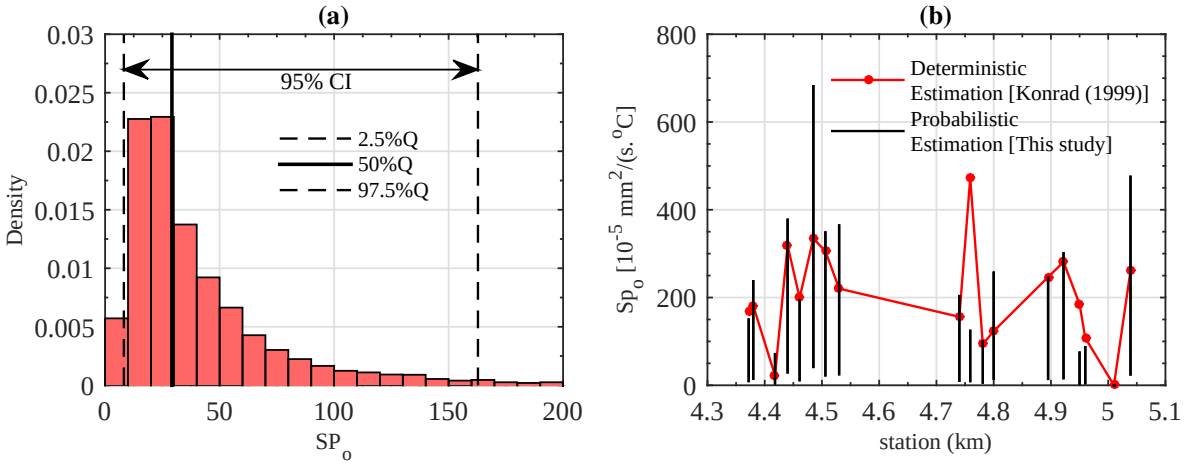


Figure 5.13: Probabilistic segregation potential for (a) at station 4.759 (b) at all stations.

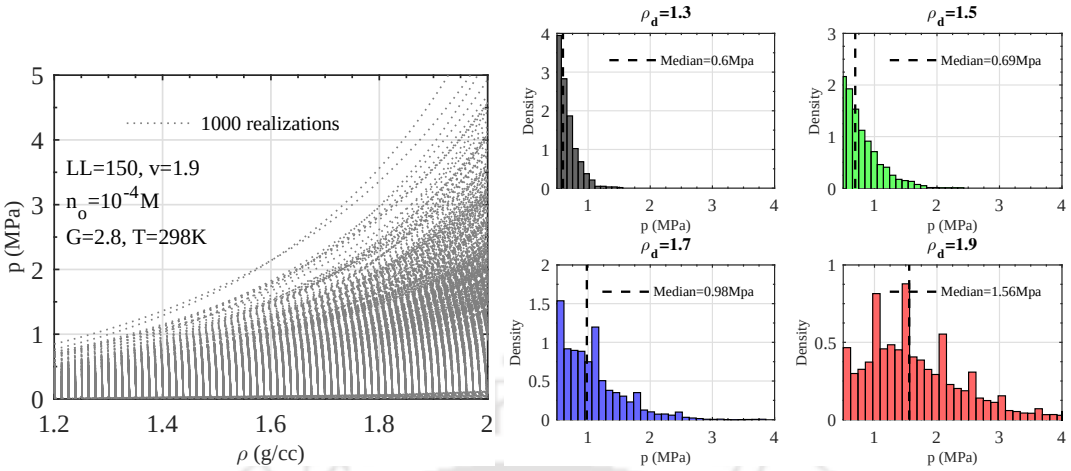
In this section, the same problem shown in Konrad (1999) for the Saints-Martyrs-Canadiens site is repeated. The data required for the estimation of SP_0 using Eqs. 5.24 and 5.25 is presented separately in Table DA4. In Konrad (1999), the SSA was estimated from d_{50} deterministically, and in this study, SSA will be estimated probabilistically from LL (ref. Table 5.7).

Fig. 5.13 presents the practical implication of the probabilistic SSA towards the estimation of SP_o. Fig. 5.13(a) presents the histogram of SP_o obtained at station 4.759 km (ref. Table DA4). Apart from point estimate, e.g., Median (50%Q), uncertainty in the estimates can be quantified using 95 % CI. Fig. 5.13(b) presents the 95% confidence interval (2.5%Q – 97.5%Q) for the probabilistic estimation of SP_o at all the stations. The estimation of SP_o in Konrad (1999) is also presented for comparison. It is important to note that in Konrad (1999), the SSA was not measured but estimated deterministically from d_{50}. In this study, the SSA was estimated probabilistically from LL. Although a one-to-one comparison is not possible in this case, nevertheless it is worthwhile to compare both the estimates (i.e. from this study and Konrad (1999)). It can be noted that for most of the stations, the values reported by Konrad (1999) are within the 95% confidence intervals and therefore demonstrate a practical application example where the probabilistic estimates of SSA obtained in this study can be utilized.

5.6.2 Swelling pressure- dry density relationship

Swelling pressure p at a particular dry density ρ_d can be theoretically determined using Guoy-Chapman diffuse double-layer theory using the following set of equations (Tripathy et al., 2004; Schanz and Tripathy, 2009).

$$p = 2n_o kT(coshu - 1) \tag{5.26}$$



(a) Swelling pressure (p) -dry density (ρ_d) relationship.

(b) Estimation of swelling pressure (p) at various dry densities (ρ_d).

Figure 5.14: Probabilistic swelling pressure (p)-dry density (ρ_d) relationship.

$$(\frac{dy}{d\xi})_{x=0} = -\sqrt{2coshz - 2coshu} = \sigma\sqrt{(2e_0Dn_0kT)^{-1}} \\ \frac{CEC}{SSA}\sqrt{(2e_0Dn_0kT)^{-1}}; at\ x = 0, y = z \tag{5.27}$$

$$\int_z^u \frac{1}{\sqrt{2coshz - 2coshu}}dy = \int_0^d d\xi = -Kd \tag{5.28}$$

$$K = \sqrt{\frac{2n_o e'^2 v^2}{e_0 DkT}}, e = G\rho_w SSAd \times 10^6, \rho_d = G\rho_w/(1+e) \tag{5.29}$$

n_o is pore fluid ion concentration ($ions/m^3$), $k = 1.38 \times 10^{23}\ J/K$ is Boltzmann's constant, u is non-dimensional mid plane potential, T is temperature in Kelvins, ξ is distance function, y is the non dimensional potential at distance x from the clay surface, $z = y(x = 0)$, $\sigma = CEC/SSA$ is the surface charge density at clay surface, $e_o = 8.8542 \times 10^{12}\ C^2\ J^{-1}\ m^{-1}$ is vacuum permittivity, D is the pore fluid dielectric constant ($= 80.4$ for water), K is diffuse-double layer parameter, $e' = 1.602 \times 10^{19}\ C$ is the elementary charge, v is exchangeable cations valency, e is void ratio, G is specific gravity, d is half the clay platelets distance and ρ_w is water density.

From Eqs 5.26-5.29, the theoretical $p-\rho_d$ relationship can be established corresponding to a pair of SSA and CEC value. In the proposed approach, since SSA and CEC are random vectors, instead of a unique $p - \rho_d$ relationship, multiple realizations are possible. Fig. 5.14a presents such an example for 1000 realizations from the joint density $f(SSA, CEC|LL)$. The probabilistic estimates of SSA and CEC essentially imply that instead of a unique point estimate, there exists a range of

p at every ρ_d. Fig. 5.14b presents the histogram of p calculated at various ρ_d. It can be easily observed that on account of increasing ρ_d the histogram of p shifts towards a region of higher p consistent with the diffuse-double layer theory. Correspondingly the median p value also increases. p induces uplift pressure on the foundation in case of light structures and is an important factor in the design of foundations on expansive soils (Buzzi et al., 2010). p is also important criteria for the selection of buffer material in nuclear waste repositories (Tripathy et al., 2004). The probabilistic $p-\rho_d$ relationship can be further extended to reliability-based design and analysis in such projects. However, the same was not demonstrated due to space constraints in this study and also as the main aim of this study was the probabilistic estimation of SSA and CEC. Nevertheless, this subsection, along with the previous one gave insight into the practical applications of probabilistic estimation of SSA and CEC.

5.7 Summary

In this study, a multivariate probabilistic approach for the estimation of CEC and SSA was proposed. For this purpose, this study constructed a 5-dimensional multivariate distribution of LL, PI, CF, CEC and SSA. A database (labelled as CLAY/C-S/5/278) was developed by the authors from literature and utilized for the construction of multivariate distribution. Vine-copula theory was used to construct the multivariate distribution by decomposing the joint density into a product of bivariate copulas and univariate marginals.

Bimodal lognormal marginals were found to be statistically satisfactory for LL and PI and right truncated lognormal marginals for CF, CEC and SSA. It was found that all the five variables in CLAY/C-S/5/278 database are positively correlated with each other with Kendall's rank correlation coefficient τ varying from 0.33 to 0.87. It was also found that 8 out of 10 pair copula required for the construction of joint density are non-Gaussian therefore indicating strong evidence of non-Gaussian dependence structure among the five parameters in the CLAY/C-S/5/278 database. The vine-copula approach utilized in this study to construct the joint density can successfully capture the existing dependence among the parameters in the database.

Since the collected database was truly multivariate (278×5) wherein the data comes from various geographical regions of the globe, it can be considered as a global multivariate database. Hence the multivariate distribution constructed in this study can be viewed as a global or generic distribution.

Multiple implementation examples were shown for probabilistic estimation of CEC and SSA utilizing the generic multivariate distribution constructed in this study. It was found that the uncertainty (in terms of COV) in the estimates of CEC and SSA decrease with increasing values of LL, PI and CF. Among LL, PI and CF the COV in the estimates of SSA and CEC was found to be highest in general for CF. This was mainly because the database utilized in this study could not explicitly consider the influence of mineralogy. Regardless of whether the uncertainty in estimates was higher or lower, the proposed approach could successfully take into account this uncertainty.

It is expected that in future large scale availability of mineralogical data can be utilized so that the uncertainty in the estimates due to the difference in the behaviour of mineralogy (e.g. kaolinite and montmorillonite) can be reduced. The estimated PDFs of CEC and SSA were also validated against a dataset outside the CLAY/C-S/5/278 database. It was found that the resulting estimates were consistent with the actual measured data.

In the proposed approach, CEC and SSA were characterized by their joint conditional PDFs. This implies that uncertainty associated with the estimation of CEC and SSA was rigorously characterized in the form of PDFs and therefore the proposed approach is superior to the popular transformation approach which can only be used to get a point estimate. For practitioners, the results of this study, i.e. the conditional probability density functions (PDFs) of CEC and SSA were also summarized as engineer friendly analytical expressions. Finally, two example geotechnical applications were demonstrated. One utilized only SSA (out of CEC and SSA) in an empirical framework for estimation of segregation potential for frost heave susceptibility in highway subgrades and another utilized both CEC and SSA in a theoretical framework for the estimation of swelling pressure at various dry densities. The proposed approach for probabilistic estimation of CEC and SSA is also directly relevant to the extension of reliability-based design philosophy in future towards the challenges mentioned in "*role of geotechnics in addressing new world problems*" (Lu and Mitchell, 2019, p.1) such as carbon storage and capture, nuclear energy and long term waste containment. This study is also in line with the philosophy of data mining presented in Phoon (2018) and Ching and Phoon (2018) wherein the available data should be exploited to its maximum possible extent for making refined predictions.

Acknowledgement

I am grateful to Dr. Giovanni Spagnoli, BASF Construction Solutions GmbH, Trostberg, Germany for contributing data to the CLAY/C-S/5/278 database. I am also grateful to Professor Amy B. Cerato (School of Civil Engineering and Environmental Science, University of Oklahoma) and Professor (Emeritus) Alan J. Lutenegger (Department of Civil and Environmental Engineering, University of Massachusetts Amherst) for valuable suggestions regarding the practical upper limit of SSA for clays.

Chapter 6

Stochastic seepage analysis considering non-Gaussian spatial and cross dependence structure of hydraulic parameters

6.1 General

In the context of stochastic seepage and slope stability analysis, the main objective of this study is to present a vine copula based multivariate random field approach to model both spatial and cross dependence structure among the hydraulic parameters (hydraulic conductivity and van Genuchten (1980) soil water characteristic parameters α and n.) The need for the vine copula based approach is also demonstrated by providing statistical evidence for non-Gaussian spatial dependence structure using a real field data (Sudicky, 1986) of 33 hydraulic conductivity cores. The majority of the cases (28/33) are shown to follow non-Gaussian dependence spatial structure. Earlier studies (Haslauer et al., 2012, e.g.) have also provided preliminary evidence of non-Gaussian spatial dependence in hydraulic conductivity using a transformation of Gaussian copula. However, they (Haslauer et al., 2012, e.g.) could not determine which non-Gaussian copula provides the best fit. This study not only can identify the non-Gaussian copula, which provides the best fit to the dependence structure but can also model it efficiently using the vine copula based random field approach. Practical importance of dependence structure for stochastic seepage and slope stability is also shown to be significant. It is shown that the assumption of arbitrary spatial dependence structure, e.g. Gaussian, can significantly affect the quantiles of pressure head, degree of saturation, the factor of safety and in turn, the failure probability of slopes. It is shown that the assumption of arbitrary spatial dependence structure, can significantly affect the failure probability (by a factor of 100) of slopes across the entire acceptable range (Salgado and Kim, 2014) of 10^{-4} to 10^{-2}. The presented vine copula approach is more generalized than the conventional approaches for stochastic seepage and slope stability analysis (e.g. Cholesky decomposition or Karhunen–Loève (K-L) expansion) which can handle only Gaussian spatial and cross-dependence structure in the random fields of hydraulic parameters.

6.2 Random field: Vine copula approach

In this section, the random field framework using vine copula approach is presented. First, the spatial dependence modelling is presented, followed by modelling cross dependence structure. Spatial dependence structure describes the dependence among the values of one random variable (e.g. k) at various locations (e.g. k_1 and k_2) in a random field, and cross dependence structure describes the dependence among multiple random variables (e.g. k and α).

6.2.1 Univariate random field: Modelling spatial dependence

Vine copula theory

A univariate random field for modelling spatial variability of a parameter (e.g. hydraulic conductivity $k(x)$) can be characterized by the joint probability density function (PDF) $f_K(k)$, where $x \in \Omega \subset R^D$. D represents the space coordinate dimensions, R, a real number set, Ω, the geometry domain and x the coordinate. Fig. 6.1 shows a schematic diagram wherein one-dimensional ($D = 1$) spatial variability in vertical direction ($k(x)$) is depicted. The joint distribution $f_K(k)$ underlying this one-dimensional random field can be characterized as follows:

$$f_K(k) = f_{K_1,K_2,K_3,...,K_n}(k_1, k_2, k_3, ..., k_n) \tag{6.1}$$

where n is the number of discretization. A joint distribution ideally contains two levels of information, one regarding the marginal distribution, i.e. to describe the random behaviour and another regarding the dependence (value and structure) among the random variables.

Theoretically, the marginals for every random variables $k_i, i \in 1 : n$ and dependence among them e.g. $\tau(k_i, k_{i+\Delta}), \Delta \in 0 : 1 : n-1$ (where Δ denotes lag) could be obtained if multiple realizations for the random field were available at a site. However, often this is not the case due to cost and time constraints in a project. Therefore, it is common to apply ergodicity and stationarity assumptions. Afterwards, the marginals (as a function of mean and variance) and dependence (as a function of distance or lags) can be obtained as follows.

$$f_{K_1}(k_1) = f_{K_2}(k_2) = f_{K_i}(k_i) = f(\mu_k, \sigma_k), i \in 1 : n \tag{6.2}$$

$$\tau(k_i, k_{i+\Delta}) = \tau(\mid \Delta \mid, \lambda) \tag{6.3}$$

Given the above two information, i.e. the marginals and dependence, the n dimensional joint distribution can be constructed using vine copula theory (Aas et al., 2009). Using the D-Vine formulation after Aas et al. (2009), the joint density in Eq. (6.2) can be expressed as follows:

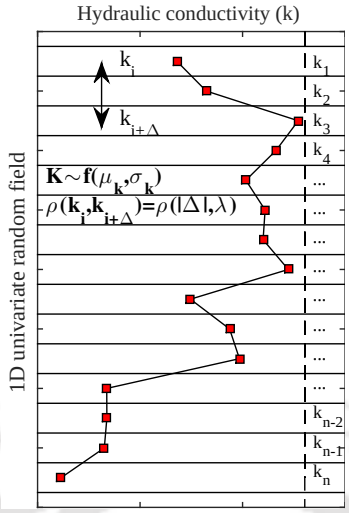


Figure 6.1: An illustration of one dimensional univariate random field of hydraulic conductivity (k) in vertical direction. μ and σ are mean and standard deviation respectively, Δ is the lags and λ is the scale of fluctuation.

$$f(k_1, k_2, ..., k_n) = \prod_{x=1}^{n} f(k_x) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} D_{i,i+j|i+1,...,i+j-1}[F(k_i|k_{i+1}, ...k_{i+j-1}), F(k_{i+j}|k_{i+1}, ...k_{i+j-1})] \tag{6.4}$$

where J denotes the index for Tree and i for edges in a particular T_j. Similarly, using the C-vine formulation after Aas et al. (2009), Eq. (6.2) can be expressed as follows.

$$f(k_1, k_2, ..., k_n) = \prod_{x=1}^{n} f(k_x) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} D_{j,j+i|1,...,j-1}[F(k_j|k_1, ...k_{j-1}), F(k_{j+i}|k_1, ...k_{j-1})] \tag{6.5}$$

It can be noted in Eqs. 6.4 and 6.5 that evaluation of joint density involves $n(n-1)/2$ copula densities. However, it is tedious to keep track of conditional copula densities involved in the formulations. To organize the copula densities into easily tractable graphs, a tree-like structure called vine is often used, hence the name vine-copula. For illustration, consider the case of $n = 5$ in Eqs. 6.4 and 6.5. The resulting vine structures are known as Drawable (D) vine, and Canonical (C) vine and are depicted in Figs. 6.2a and 6.2b respectively. Each vine structure for n-dimensions is made up of $n-1$ trees (T_i) consisting of nodes and edges. Each of the edges can be assigned a bivariate (pair) copula. C-vine resembles (Fig. 6.2b) a star-shaped structure wherein every tree (T_i) contains a base node connected to each of the $n-j$ edges. D-vine (Fig. 6.2a) resemble a parallel stack like structure and every node in any T_i can be connected to only two edges at maximum.

In Fig. 6.2b, for T_1 it can be noted that $\Delta = 1 : 1 : n-1$. For example, $\Delta = 1$ for D_{12}, $\Delta = 2$

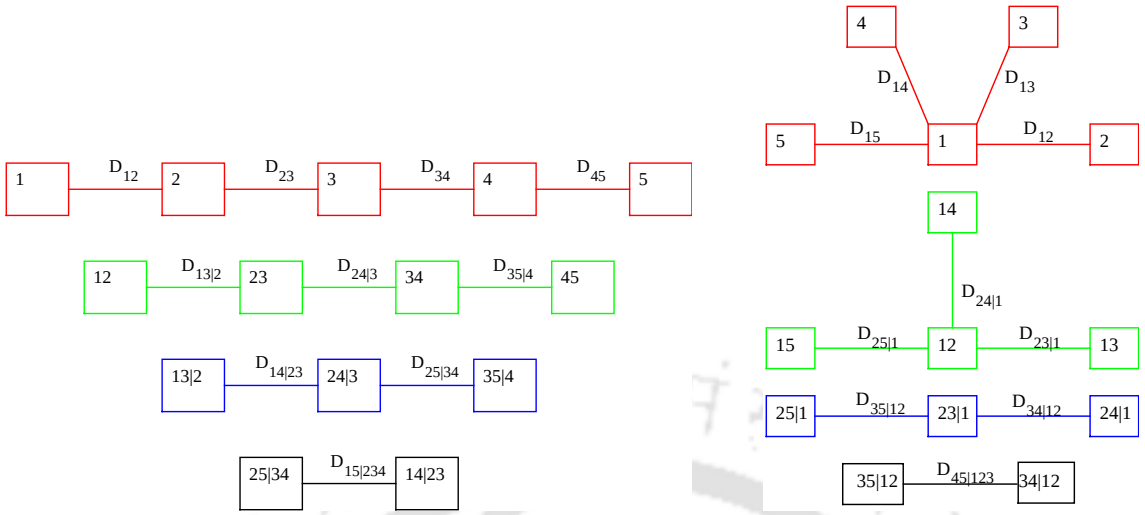


Figure 6.2: An example illustration of (a) D-vine (b) C-vine in $n = 5$ dimensions with $n - 1$ trees and $n(n - 1)/2$ edges.

for D_{13}, $\Delta = 3$ for D_{14} and so on. This implies that for C-vine the first pair trees can be uniquely determined from the spatial dependence model fitted to the auto-dependence data. Similarly, in Fig. 6.2a, it can be noted that $\Delta = 1$ is constant in T_1 for all the pair copulas. For example, $\Delta = 1$ for D_{12}, D_{23}, D_{34} and so on. This implies that for D-vine all the parameters for a particular T_i will be same. This is a significant advantage since it reduces the number of parameters to be estimated from $n(n-1)/2$ to $n-1$ only. However, from the T_2 onwards estimation of copula requires conditional variables and cannot be directly estimated from the data. However since parameters for tree T_i is determined from T_{i-1}, the pattern will we similar to T_1. For example, the parameters of copula at each T_i for the D-vine will be same, and for C-vine the parameters at each T_i will follow a decaying sequence similar to the spatial dependence model. This aspect will also be demonstrated in subsequent sections for real data.

Copula selection

Information criteria such as AIC is most popular (Aas et al., 2009) for estimating the best-fit dependence structure. The same can be expressed as follows

$$AIC = -2L(\theta_n) + 2(k) \tag{6.6}$$

where $L(\theta_n)$ is the maximum log-likelihood, and k is the number of parameters. The copula families among the chosen candidates can be selected corresponding to the minimum AIC value. Since all the four copula utilized in this study (ref. Table DA2) have the same no of parameters, i.e. 1, minimum AIC is equivalent to maximum $L(\theta_n)$. The full likelihood estimation is popular for random variable modelling as the number of dimensions is usually low in geotechnical engineering e.g. 7 in

Ching et al. (2017), 9 in Ching, Phoon, Li and Weng (2018) and 10 in Ching and Phoon (2014). However, for random fields, the dimension can be very large. Hence a full likelihood estimate will be computationally expensive and unfeasible for the random field. In such cases, an alternate pair-wise likelihood estimate is preferred (Padoan et al., 2010; Rózsás and Mogyorósi, 2017; Wang and Li, 2019, e.g.). This is suitable for random fields since the pairwise likelihood can be constructed solely from the auto-dependence data. The pair wise likelihood can be expressed as follows.

$$L(\theta_n) = ln \prod_{i<j}^{n} f(k_i, k_j) \tag{6.7}$$

Breaking the joint density $f(k_i, k_j)$ into copula and marginals, the above equation can be written as follows.

$$L(\theta_n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} ln\{D_{ij}[F(k_i), (k_j)]f(k_i)f(k_j)\} \tag{6.8}$$

$$L(\theta_n) = \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} lnD_{j,j+i}[F(k_j), F(k_{j+i})] + \sum_{k=1}^{n} ln[f(x_k)](n-1) \tag{6.9}$$

The above equation can further be simplified as follows:

$$L(\theta_n) = \sum_{i=1}^{n-1} L_i + \sum_{k=1}^{n} ln[f(x_k)](n-1) \tag{6.10}$$

where $L_i = \sum_{j=1}^{n-i} lnD_{j,j+i}[F(k_j), F(k_{j+i})]$. Since the second term in Eq. (6.10) is independent of copula, minimum AIC copula can be selected corresponding to the maximum value of $\sum_{i=1}^{n-1} L_i$ as follows.

$$\min(AIC) \equiv \max(L(\theta_n)) \equiv \max(\sum_{i=1}^{n-1} L_i) \tag{6.11}$$

Simulation

For the simulation of correlated random numbers (corresponding to a random field) from vine copula, parameters of the copula associated with each edge in a vine (e.g. Fig. 5.3) are required along with conditional and inverse conditional copula functions. The general algorithm for simulating random fields are as follow.

For $i = 1$: Number of realizations

- Generate a random sample $r = (r_i1, r_i2, r_i3, .., r_in)$ from independent uniform distribution
- Set $U_i1 = r_i1$

- Set $U_i2 = C^{-1}_{2|1}(r_i2|U_i1)$
- Set $U_i3 = C^{-1}_{3|12}(r_i3|U_i, U_i2)$
- ....
- ....
- Set $U_in = C^{-1}_{n|12...n-1}(r_in|U_i, U_i2, ...U_in-1)$

End

The obtained random vector $U = (U1, U2, ...Un)$ is a random field with specified dependence structure and uniform marginals. As already discussed in previous sections, only the first tree parameters of the D and C vine can be directly estimated from the data. For example, in C-vine structure, e.g. Fig. 5.3 θ_{12} can be uniquely determined from τ_{12} which is equal to auto-dependence (not data but fitted) at $\Delta = 1$. Similarly θ_{13} corresponding to $\Delta = 2$ and in general θ_{1n} corresponds to $\Delta = n - 1$. Therefore the parameter estimates will decay in the same manner as the spatial dependence model. However, from the 2nd tree onwards, the parameters cannot be determined uniquely from the available data. It requires transformed data which in turn are obtained from the previous tree nodes. Therefore would follow the similar decaying trend as spatial auto dependence model. Since most of the cases, only a single realization is available, this approach is not possible for most practical purposes. In such a case, the simulation-based approach (Wang and Li, 2017, e.g.) is attractive and is utilized in this study. This approach is basically a moment matching method wherein the parameters corresponding to all edges are initialized and continued until the dependence produced are matched to the assigned one. In the simulation approach, although the transformed data is not required, conditional distributions and its inverses are still required. The conditional distributions $C_{j|j-1,..,1}$ required for simulating the random numbers can be obtained recursively using the following formula (Aas et al., 2009).

$$F_{X|\boldsymbol{Y}}(x|\boldsymbol{y}) = \frac{\partial C_{X,Y_j|\boldsymbol{Y_{-j}}}[F_{X|\boldsymbol{Y_j}}(x|\boldsymbol{y_{-j}}), F_{Y_j}|\boldsymbol{Y_j}(y|\boldsymbol{y_{-j}})]}{\partial F_{Y_j|\boldsymbol{y_{-j}}}(y|\boldsymbol{y_{-j}})} \tag{6.12}$$

where X is a random variable, Y_j is a element of random vector $\boldsymbol{Y}$ and $\boldsymbol{Y_{-j}}$ is vector consisting of all components of $\boldsymbol{Y}$ but Y_j. For the case where $\boldsymbol{Y}$ is univariate, the above reduces to

$$F_{X|Y}(x|y) = \frac{\partial C_{X,Y}[F_X(x), F_Y(y)]}{\partial F_Y(y)} \tag{6.13}$$

For the case of $X \sim U(0,1)$ and $Y \sim U(0,1)$ i.e. $F_X(x) = u$ and $F_Y(y) = v$, the above reduces to

$$h(u, v, \theta) = F_{X|Y}(x|y) = \frac{\partial C_{X,Y}[u, v]}{\partial v} \tag{6.14}$$

The h and h^{-1} functions for the four copulas used in this study along with copula function C and density D are summarized in Table DA2. Detailed algorithms for efficient implementation of

the above procedure can be found in Aas et al. (2009).

6.2.2 Multivariate random fields: Modelling cross dependence

After modelling the spatial dependence in univariate random fields, cross dependence can be imparted in the multivariate random fields as follows.

1. Generate k univariate random fields $U^1, U^2, ...U^K$ using the procedure mentioned in previous section.
2. For $i = 1$: Number of realizations

   - $V_i^1 = U_i^1$
   - $V_i^2 = C_{2|1}^{-1}(U_i^2|V_i^1)$
   - $V_i^3 = C_{3|12}^{-1}(U_i^3|V_i^1, V_i^2)$
   - ....
   - ....
   - $V_i^k = C_{k|12...k-1}^{-1}(U_i^k|V_i^1, V_i^2, ..., V_i^{k-1})$

   End

This step essentially establishes cross-dependence among n uniform random variable of the discretized random fields to the target cross-dependence correlation . Cross-correlation simply refers the correlation coefficients between the different random variables corresponding to the multivariate random field. Similar to the simulation-based approach followed in the previous section, the parameters are initialized and repeated until the target cross-dependence matrix is achieved. Finally, the marginals among the spatially and cross-correlated random fields (till now uniform) can be incorporated after applying inverse CDF transform as follows:

$$X_i^k = F^{k^{-1}}(V_i^k) \tag{6.15}$$

where X_i^k is a stationary random field and $F^{k^{-1}}$ is the inverse CDF of the marginal for kth random field.

6.3 Borden aquifer hydraulic conductivity data: Implementation Example

In the previous section, the methodology part of the random field (input for stochastic seepage and slope stability analysis) using vine copula theory is presented. The purpose of presenting the vine copula based approach is to handle the non-Gaussian dependence structure in the hydraulic parameters. However, most of the popular methods (e.g. Covariance matrix decomposition) consider

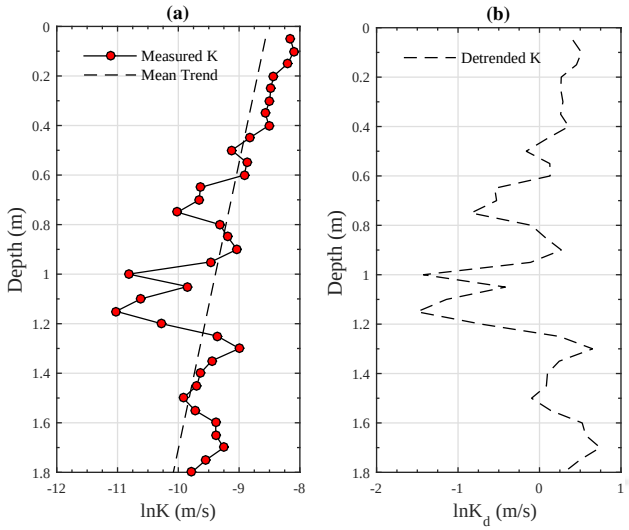


Figure 6.3: (a)Measured $ln(K)$ and (b) de-trended $ln(K_d)$ profile of hydraulic conductivity (K) in core 2 at section $B - B'$ in Sudicky (1986).

only the Gaussian dependence structure. Therefore, a natural question that arises is *"why non-Gaussian dependence structure should be considered at all, for hydraulic parameters."* To consider this particular aspect, this section utilizes a well-documented dataset and shows that non-Gaussian dependence structure is not just a statistical term but can also be observed for a real field data.. After providing the evidence, the same non-Gaussian dependence is modelled using the methodology(input) presented in section 2.

This section utilizes hydraulic conductivity data from the Borden aquifer, Canada, presented in Sudicky (1986). The Borden aquifer largely consists of discontinuous fine to medium grained sand lenses scattered in the horizontal direction. Some amount of coarse sand along with silty and clayey soils are also present (Sudicky, 1986). The data were measured at two sites, A-A' and B-B' and the total number of cores recovered were 20 and 13, respectively. Each of the cores was extracted from a depth between 2.5 to 4.5 m below the ground surface. Seasonal fluctuations in the water table at the site varies within 1 m interval below the ground surface. The length of each of the cores was approximately 2 m. The horizontal spacing among the cores was 1 m, and the vertical spacing for each measurement was 0.05m. The number of data points for each core was 36. Each of the extracted cores was divided into 0.05 m subsamples, and a specially designed falling head permeameter used to calculate the hydraulic conductivity profiles. Further details about the site, measurement procedure or data can be obtained in Sudicky (1986).

Haslauer et al. (2012) also utilized the same dataset and used Gaussian copula and its transform to model the spatial dependence of k. However, Haslauer et al. (2012) only provided evidence of non-Gaussian spatial dependence. The non-Gaussian copula, which provides the best fit to the data, was not determined. The non-Gaussian dependence was achieved using v-transform of Gaussian copula and is essentially not a copula in a strict sense as other well-established copula families such as Clayton, Frank and Gumbel. In this study, not only the non-Gaussian copula can be identified,

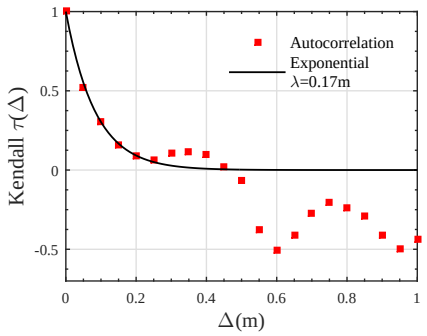


Figure 6.4: Exponential spatial dependence model fitted to the auto-dependence in terms of Kendall's τ for the de-trended data $ln(K_d)$ in Fig. 6.3(b). λ is the scale of fluctuation.

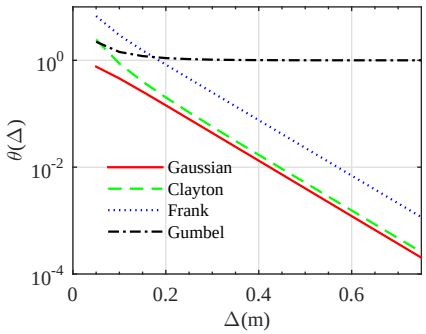


Figure 6.5: Values of copula parameter θ for different families as a function of lag distance Δ for de-trended data $ln(K_d)$ in Fig. 6.3(b).

but the same non-Gaussian dependence structure can also be efficiently modelled using vine copula theory.

Fig. 6.3 presents an example for the measured data for core 2 at site $B - B'$. The data were de-trended using a linear trend function to make it stationary. This is a standard and common procedure for random fields (Vanmarcke, 1977; Phoon et al., 2003; Ching and Phoon, 2017, e.g.) as the de-trended data (residuals) can be treated as a stationary random field with zero-trend. All the further analysis is performed over the de-trended data. Fig. 6.4 presents the spatial auto-dependence (or auto-correlation) values with lag distance in terms of rank dependence coefficient Kendall's τ. Kendall's τ was chosen particularly since the copula parameters can be obtained directly. An exponential form of spatial auto dependence model in the form $\tau(\Delta) = exp(-2|\Delta|/\lambda)$ was further utilized to fit the auto-dependence data.

The next task is to choose an appropriate copula for the spatial dependence using the procedure in section 2.1.2. Fig. 6.5 presents the bivariate copula parameters for the four copulas as a function of lag distance. It can be noted that the parameters are obtained not from the auto-dependence data but from the fitted spatial dependence model. It is important to note that even though the copula parameters at a particular Δ are different, they all correspond to the same Kendall's τ. To further illustrate this, Fig. 6.6 presents the scatter and density plot for the four copulas used in this study at same Kendall's τ. It can be noted that all the copula will reproduce almost the same Kendall's τ as the different families of copulas do not differ much in overall dependence but only in the distribution of dependence in the quadrants. It can be noted in Fig. 6.6 that while the dependence structure for Gaussian is elliptical and evenly distributed in the upper right and lower left quadrant (i.e. symmetric), for frank copula the distribution is symmetric as but not elliptical. For Clayton copula, the strength of dependence is stronger in the lower left quadrant and comparatively lower in the upper right. For Gumbel copula also the dependence is not evenly distributed but stronger in upper right quadrant as compared to the lower left.

Fig. 6.7 presents the value of cumulative $\sum_{i=1}^{n-1} L_i$ as a function of Δ. As expected, after a certain interval, the data does not contribute to any L_i as the dependence dies down after some

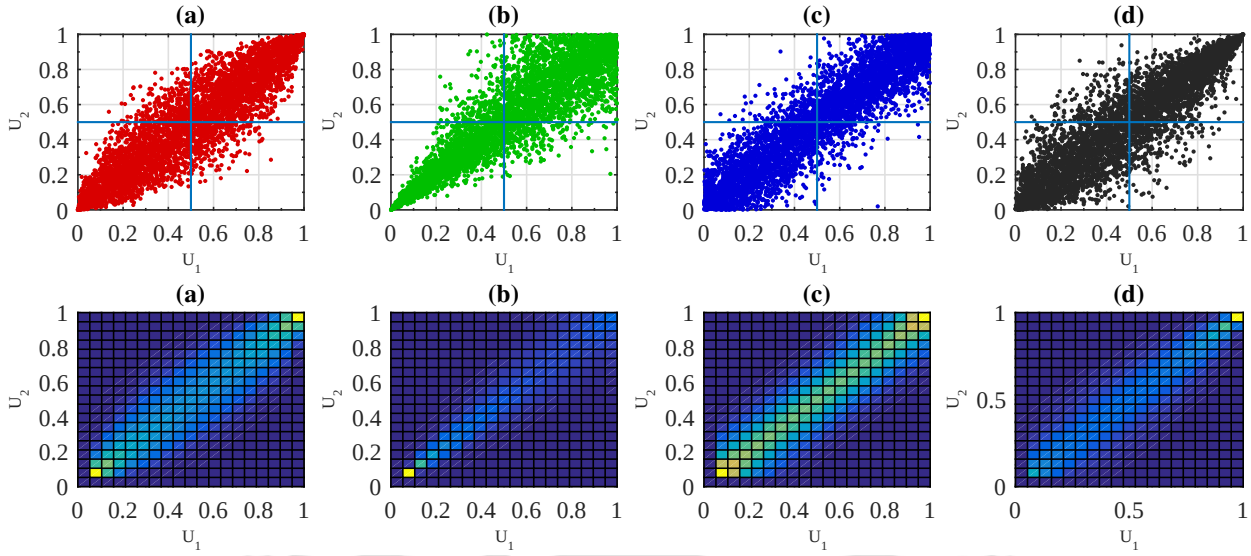


Figure 6.6: Scatter and density plot for (a) Gaussian (b) Clayton (c) Frank and (d) Gumbel copula for same rank dependence coefficient Kendall's $\tau = 0.75$.

distance. It can be easily observed that Frank copula is the best copula according to the minimum AIC criterion mentioned in Eq. (6.11). To further investigate if this result is an exception or more common, the same exercise was repeated for all the rest of the cores. Figs. 6.8 and 6.9 present the $\sum_{i=1}^{n-1} L_i$ as a function of Δ for site $B - B'$ and $A - A'$. It can be observed that for both sites in majority of the cases, non-Gaussian copula is the best fit. In Fig. 6.8, it can be observed that except for core 4 and 9, for the rest of the cases non-Gaussian spatial dependence structure is identified to be the best fit. Among all the non-Gaussian cores, except core 8 and 11, Gumbel copula is the best fit. Similarly, in Fig. 6.9 non-Gaussian structure is predominant except for cores 2,3 and 20. Overall among all the 33 cores combined at site $A - A'$ and $B - B'$, 28 cores exhibit non-Gaussian dependence structure. This result is significant in the sense that it affirms the need for a random field approach that can handle a non-Gaussian dependence structure. The result is also in line with a recent study of Wang and Li (2019) wherein evidence for non-Gaussian dependence structure was provided using a cone penetration resistance data. It is important to note that a linear trend was adopted in this study for the data. It might be worthwhile to investigate the impact of a different trend model (e.g. quadratic or even higher order) on the observed dependence structure.

Using the simulation-based approach mentioned in section 2.1.3, parameters for the de-trended lnK core 2 data in Fig. 6.3 are found out. The parameters for the D-vine and C-vine are summarized in Fig. 6.10 and 6.11. For the D-vine structure since all the parameters at a particular tree level are same only one value for each tree level is shown. In Fig. 6.10 it can be seen that most of the dependence is captured in the first few trees only as the other parameters for rest of the trees are close to zero. For the C-vine in Fig. 6.11 the parameters for all the trees follow similar spatial decay as in the first tree. It is apparent that using a D-vine is more parsimonious than C-vine, hence recommended for spatial modelling. Given the copula and its parameters, it can be checked if

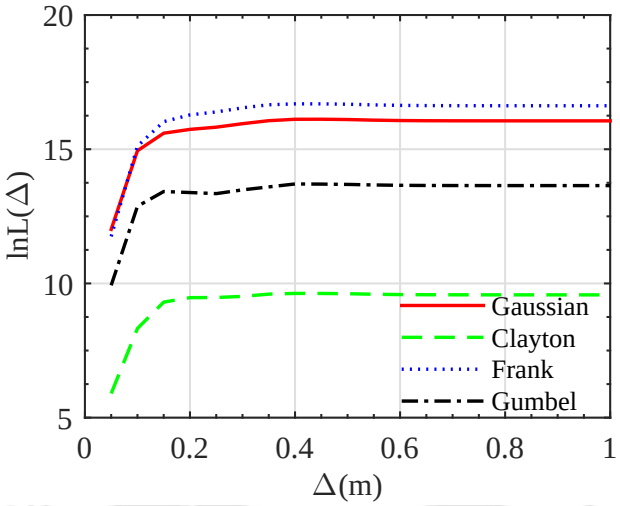


Figure 6.7: Cumulative log-likelihood $\sum_{i=1}^{n-1} L_i$ for different families as a function of lag distance Δ for the de-trended data $ln(K_d)$ in Fig. 6.3(b).

the spatial dependence can be modelled effectively. Fig. 6.12 presents the measured and simulated spatial dependence using the best fit frank copula for the core 2 data presented in Fig. 6.1. It can be easily observed that the vine copula approach can successfully model the theoretical prescribed spatial dependence. As already discussed, D-vine is parsimonious than C-vine, therefore the results are shown only using D-vine.

Given that the dependence structure can be modelled satisfactorily, the marginals can be utilized to simulate the random field on the original scale. Generalized extreme value (GEV) distribution was chosen as the marginal on the basis of the Kolmogorov–Smirnov (K-S) test. The resulting PDF and CDF fit to the empirical data are presented in Fig. 6.13. Visually it can be observed that the GEV distribution can satisfactorily model the empirical PDF and CDF. To confirm it statistically, it can be noted that $p > 0.05$, therefore GEV can be used to model the $ln(k)_{de}$ data for core 2 at site B-B' (ref. Fig. 6.3). It can also be noted that GEV is not a conventional choice for $ln(k)_de$ data, as it is more common to use normal distribution (Sudicky, 1986; Santoso et al., 2011, e.g.). Fig. 6.14 presents the K-S test p value for normal and GEV distribution at all the 33 cores. In Fig. 6.14(b) it can be noted that for the core 2 data (ref. Fig. 6.3), $p < 0.05$ for normal distribution but $p > 0.05$ for the GEV distribution. Therefore GEV distribution was selected for core 2 data. Overall also, in Fig. 6.14 it can be noted that for both sites, GEV $p > 0.05$ for all the cores. However, for normal distribution, $p < 0.05$ in some cases. This observation essentially implies that although normal distribution is popular (Santoso et al., 2011, e.g.), proper statistical verification should be done as it may not always provide a satisfactory fit, e.g. for core 2 data at site $B - B'$. Fig. 6.15 shows one realization of the simulated random field for the core 2 data using all the four copulas (at fixed seed) and GEV marginals. Although it may seem that all the copulas produce almost the same result, it will be shown in the next section that choice of copulas can affect the reliability results considerably.

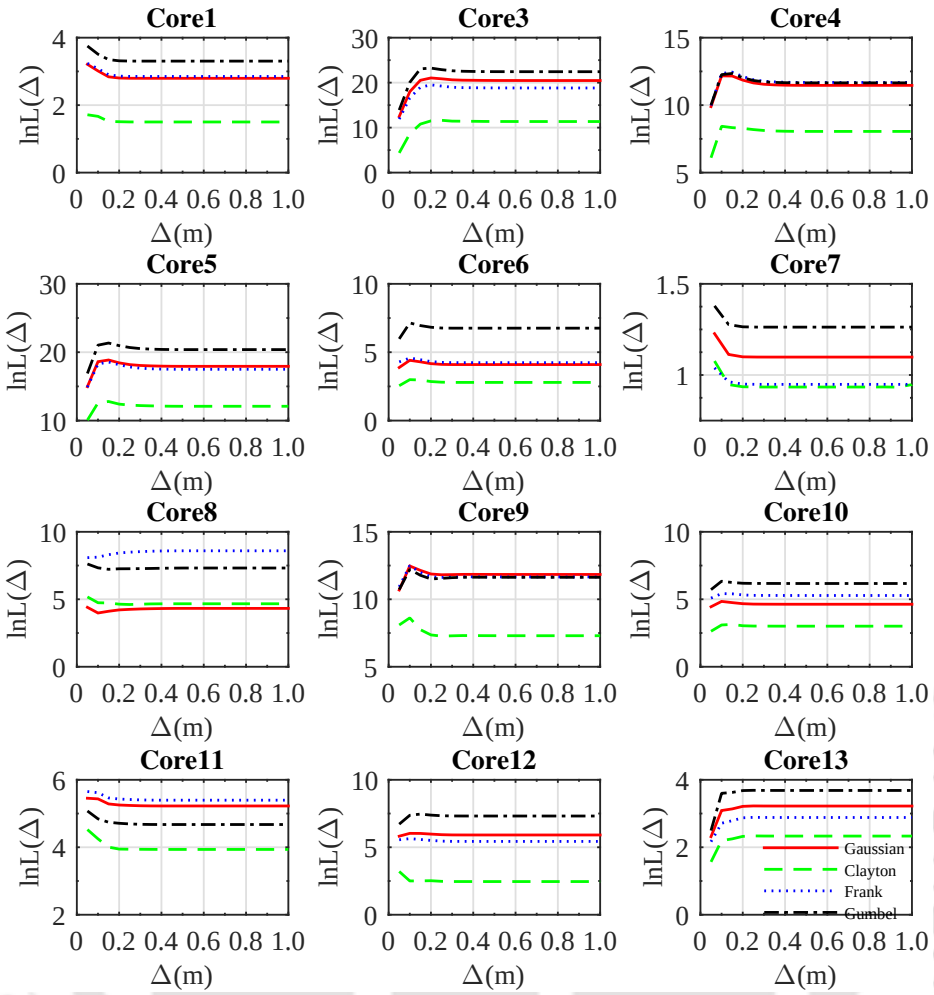


Figure 6.8: Cumulative log-likelihood $\sum_{i=1}^{n-1} L_i$ for different families as a function of lag distance Δ for the de-trended data $ln(K_d)$ in rest of the cores at site $B - B'$ in Sudicky (1986).

6.4 Stochastic seepage and slope stability analysis

The previous section provided evidence of non-Gaussian dependence structure on a field data of hydraulic conductivity and demonstrated the modelling capability of non-Gaussian dependence structure using the proposed approach. However, another natural question regarding the practical importance of this study arises is that *"even if there is evidence of non-Gaussian dependence structure, will it make any significant difference to the reliability results of stochastic seepage and slope stability analysis"*. To investigate this aspect, in this section, the methodology for random fields described in Section 2 is used to simulate Gaussian as well as non-Gaussian random fields, and a stochastic seepage and slope stability analysis is conducted.

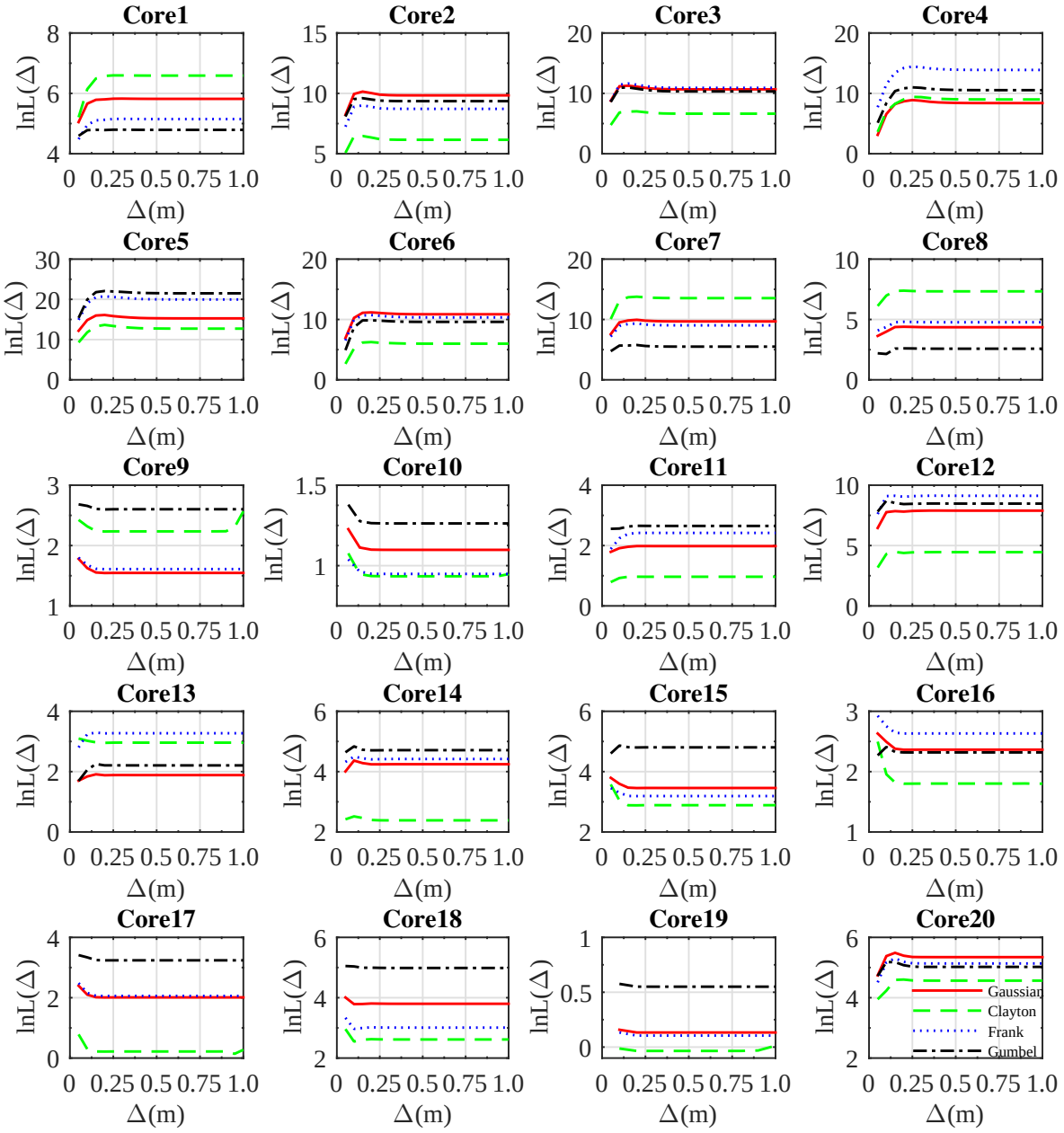


Figure 6.9: Cumulative log-likelihood $\sum_{i=1}^{n-1} L_i$ for different families as a function of lag distance Δ for the de-trended data $ln(K_d)$ in all the cores at site $A - A'$ in Sudicky (1986).

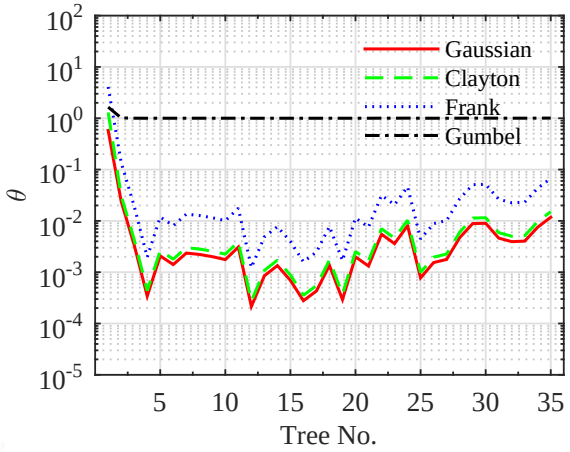


Figure 6.10: Copula parameter θ corresponding to different tree of D-vine for the de-trended data $ln(K_d)$ in Fig. 6.3(b).

6.4.1 Governing equations

Seepage in geotechnical engineering is commonly modelled assuming that the flow in soil satisfies both the continuity equation and Darcy's law (Li et al., 2009, e.g.) as follows:

$$\frac{d\theta_{w_e}}{dh}(x,h)\frac{\partial h(x,t)}{\partial t} + \nabla.q(x,t) = f(x,t) \tag{6.16}$$

$$q(x,t) = -K(x,h)\nabla[h(x,t)+z] \tag{6.17}$$

where h is the pressure head, θ_{w_e} is the effective water content, q is the flux, f is the source or sink term, z is elevation and $K(x,h)$ is the unsaturated hydraulic conductivity. The boundary and initial conditions can be assigned as $h(x,0) = h_0(x)$ where $x \in X$, $h(x,t) = H_b(x,t)$ where $x \in \Gamma_D$ and $q(x,t).n = Q(x,t)$ where $x \in \Gamma_N$. In order to solve the above set of equations, a constitutive model between K, θ_w and h is also necessary. van Genuchten-Maulem model (van Genuchten, 1980; Mualem, 1976) is the most popular and commonly employed in numerical codes (Simunek et al., 1998, e.g.), hence utilized in this study. The model can be expressed as:

$$S_e(h,x,t) = \frac{\theta_w - \theta_{w_r}}{\theta_{w_s} - \theta_{w_r}} = \frac{1}{\{1+[\alpha(x)h(x,t)]^{n(x)}]^{m(x)}\}} \tag{6.18}$$

$$K(h,x,t) = k_{sat}(x)\sqrt{S_e(h,x,t)}\{1-[1-S_e(h,x,t)^{1/m(x)}(x,t)]^{m(x)}\}^2 \tag{6.19}$$

where S_e is the effective degree of saturation, θ_{w_r} and θ_{w_s} are residual and saturated volumetric water content respectively, α and n are the unsaturated soil parameters related to air entry value and desaturation respectively, and $m = 1 - 1/n$. Another hydraulic conductivity model after Gardner (Gardner, 1958) is popular, since this can be utilized to derive a closed-form solutions for

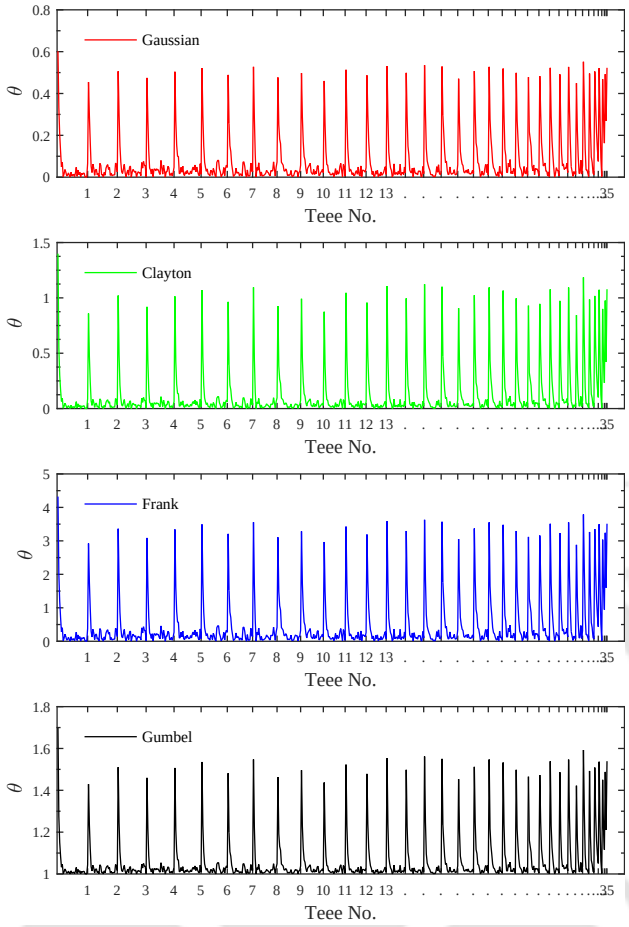


Figure 6.11: Copula parameter θ corresponding to different tree of C-vine for the de-trended data $ln(K_d)$ in Fig. 6.3(b).

the steady seepage case (Yeh, 1989, e.g.). The same is given as:

$$K(x,t) = k_{sat}(x)e^{[\alpha(x)h(x,t)]} \tag{6.20}$$

After solving for $h(x,t)$, the same can be incorporated into slope stability analysis. To demonstrate the application, an example of an infinite unsaturated slope is presented in this section. Factor of safety (FS) for an unsaturated infinite slope (Fredlund et al., 2012) is given as follows:

$$FS(x,t) = \frac{c + c[\psi(x,t)] + \gamma_s x cos^2\beta tan\phi}{\gamma x sin\beta cos\beta} \tag{6.21}$$

where c, ϕ, γ_s are cohesion, angle of internal friction and unit weight of soil respectively. x is depth below the ground surface, β is slope angle, suction $\psi = -h\gamma_w$, γ_w is the unit weight of water and $c[\psi(x,t)]$ is the apparent cohesion resulting due to ψ. The unsaturated shear strength is imparted to the slope via this term. For estimation of $c[\psi(x,t)]$, the method proposed by Lu and Griffiths

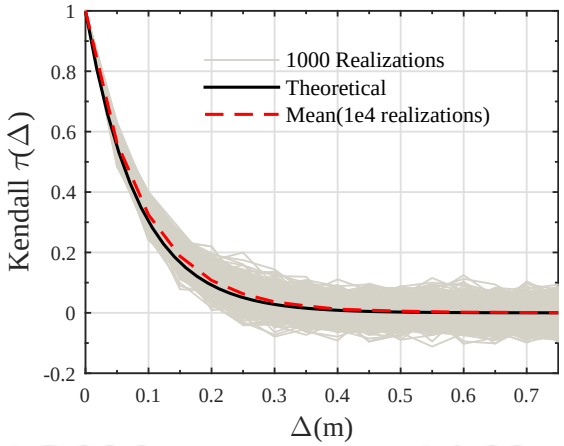


Figure 6.12: Theoretical and simulated spatial dependence using best fit Frank copula for the de-trended data $ln(K_d)$ in Fig. 6.3(b).

Table 6.1: Parameters for the application example in this study.

Parameter	Description	Statistics
k m/day	Hydraulic conductivity	$\mu = 1$, $CV = 1$,$\lambda = 1$
α m^{-1}	SWCC parameter	$\mu = 2$, $CV = 0.3$,$\lambda = 0.75$
n	SWCC parameter	$\mu = 2.4$, $CV = 0.15$,$\lambda = 0.5$
θ_r	Residual water content	0.01
θ_s	Saturated water content	0.39
q m/day	Flux	-0.20
γ kN/m^3	Unit weight	20
c kPa	Effective cohesion	0
ϕ	Angle of internal friction	29
β	Slope angle	30
dx m	Discretization	0.05
D m	Slope Depth	3

(2004) incorporating the vG parameters is used. The same is given as follows.

$$c[\psi(x,t)] = \frac{\psi(x,t)}{\{1+[\alpha(x)\psi(x,t)]^{n(x)}\}^{1-1/n(x)}} \times tan\phi \tag{6.22}$$

The failure probability (P_f) can be calculated using the standard indicator function approach as relative frequency of slope failure i.e. $FS(x,t) < 1$.

6.4.2 Stochastic seepage and slope stability analysis under steady seepage: Impact of spatial dependence structure

Steady seepage solution does not require the SWCC term ($S_e - h$) in Eq. 6.18 but only the $K - h$ constitutive relationship. Although numerical codes can be used for steady seepage also, this study utilizes a close form solution for steady seepage derived by Yeh (1989). This expression utilizes the Eq. 6.20 for solving the one-dimensional form of Eq. 6.16 and is given as follows:

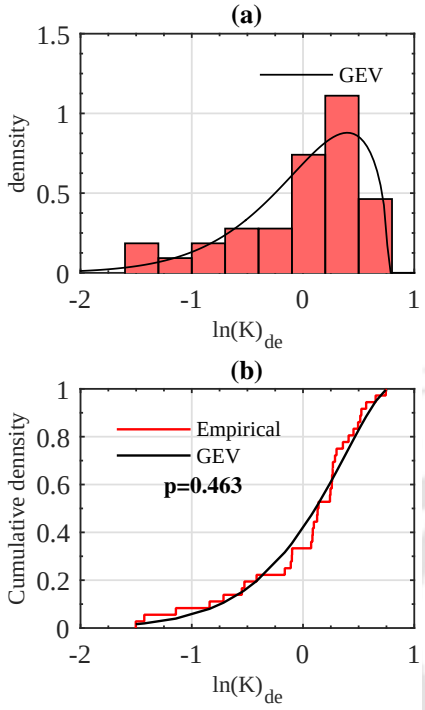


Figure 6.13: Empirical and calculated PDF and CDF for the de-trended data $ln(K_d)$ in Fig. 6.3(b).

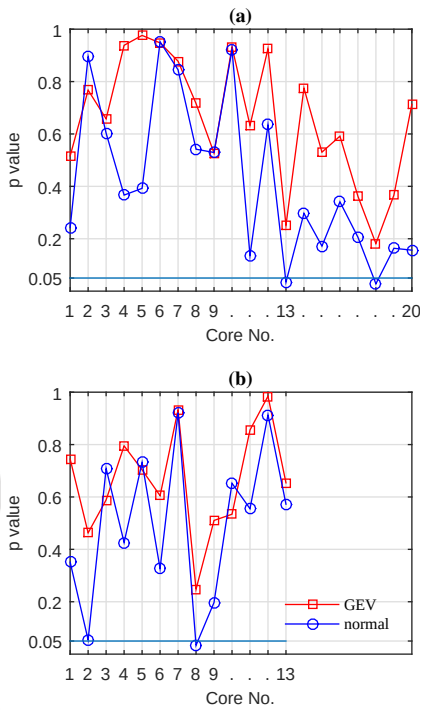


Figure 6.14: Kolmogorov–Smirnov KS test p-values for the de-trended data $ln(K)_{de}$ at site (a) $A - A'$ (b) $B - B'$.

$$h_{x_i} = \frac{ln\{e^{-\alpha\gamma_w dx}(e^{\alpha\gamma_w h_{x_{i-1}}} + q/k_{s_{x_i}}) - q/k_{s_{x_i}})\}}{\alpha\gamma_w} \tag{6.23}$$

Since this section aims to investigate the potential impact of spatial dependence structure, only k is treated as a random field, and parameter α is treated as constant. Ideally, all the other parameters should also be represented by random fields, but only k was taken as the random field so that the results can completely reflect the difference only due to k and without interference from other parameters. This is a common practice to investigate the impact of some statistical assumption, e.g. Gaussian dependence structure (Tang et al., 2013, 2015, e.g.). The parameters for the random field and slope are summarized in Table 6.1 after Li et al. (2009) and Santoso et al. (2011).

Fig. 6.16 presents 1000 realizations for $h(x)$ using Gaussian copula in D-vine formulation. In Fig. 6.16, it can be noted that for some realizations h is positive, this is because, in layered soils, h not only depends upon conductivity but also the relative position of each other. If the lower layer is less conductive than the upper layer, then under certain realizations, positive h may develop (Yeh, 1989; Santoso et al., 2011). The $FS(x)$ realizations are also computed using the h profiles. It can be noted that some portions of the $FS(x) < 1$ are observed at very shallow depths, e.g. $x < 0.5$. Slope failures at such low depths are unrealistic (Santoso et al., 2011), therefore not considered in further calculations of P_f.

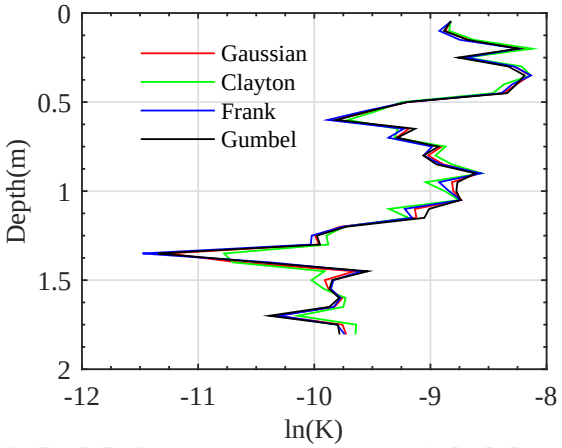


Figure 6.15: Simulated random field of hydraulic conductivity $ln(K)$ using different copulas.

To investigate the impact of spatial dependence structure of k, the above exercise was conducted using all the other three copulas, and the resulting quantiles are presented in Fig. 6.17. It can be noted that there are certain differences among the quantiles of $h(x)$ even though all the input parameters other than dependence structure are the same. Particularly the difference is more pronounced at lower and higher quantiles. For the $FS(x)$ quantiles in Fig. 6.17, it can be noted that the same 2.5%Q can result in different failure events with the choice of copulas.

Fig. 6.18a presents the P_f corresponding to various copulas for the steady seepage case. It can be noted that the P_f may vary almost 10 times $(0.003 - 0.03)$ with the choice of copula, i.e. spatial dependence structure. This result is significant as this implies that not only spatial dependence can be non-Gaussian (shown in the previous section), but choosing an arbitrary spatial dependence may also significantly affect the P_f. Although Fig. 6.18a demonstrates the importance of spatial dependence structure, the P_f values are not very practical for the slopes. The general acceptable range of P_f for slopes as per Salgado and Kim (2014) is $2 \times 10^{-4} - 10^{-2}$. The acceptable range of P_f for various cases are summarized in Table 1.1.

In Fig. 6.18a, the P_f are very close to the upper limit of $P_f = 10^{-2}$. Therefore, a case for a lower P_f well within the acceptable range of $10^{-4} - 10^{-2}$ was further considered. Fig. 6.18b presents the result of the analysis. The seepage flux was considered lower at $q = -0.15m/day$ in comparison to the $q = -0.20m/day$ (ref. Fig. 6.18a to reduce the P_f. Three zones are marked in the Fig. 6.18b. The zones with $P_f \leq 10^{-4}$ is marked as too conservative (too low P_f). Although $P_f \leq 10^{-4}$ is not undesirable, it might result in unnecessary cost additions to the project. The zone with $P_f \geq 0.01$ is marked as risky as mostly temporary structures are considered to be constructed in this zone (Salgado and Kim, 2014). The well acceptable zone of $10^{-4} - 10^{-2}$ after Salgado and Kim (2014) is termed as acceptable. It can be noted that the practical importance of this study can be appreciated even more in Fig. 6.18b than Fig. 6.18a. This is because in Fig. 6.18b, the P_f due to choice of copula vary close to 100 times $(10^{-4} - 10^{-2})$ and span entirely over the acceptable zone.

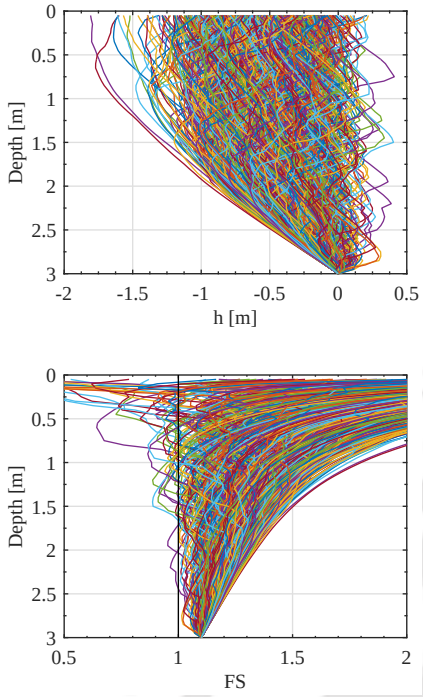


Figure 6.16: N Realization of pressure head h and Factor of safety FS versus depth using Gaussian copula for spatial dependence of hydraulic conductivity K. $N = 1000$.

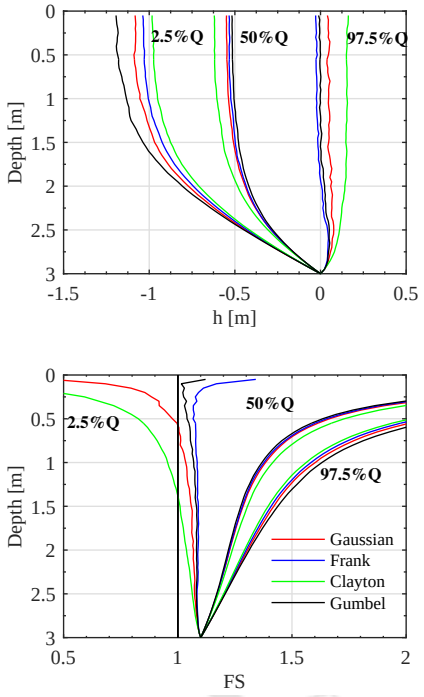


Figure 6.17: Quantiles of pressure head h and Factor of safety FS versus depth using various copulas for spatial dependence of hydraulic conductivity K. $N = 10000$.

It should also be noted that only four copulas and one set of parameters mentioned in Table 6.1 were used for this case study. Further simulation experiments are needed to investigate the impact of spatial dependence structure with variation in the scale of fluctuation, boundary flux and other input parameters statistics. However, the same was not performed in this study for the sake of brevity. Nevertheless, this cases study for the first time demonstrated that the potential impact of spatial dependence structure of hydraulic parameters e.g. k is significant in the context of stochastic seepage and slope stability analysis and therefore deserves attention.

6.4.3 Stochastic seepage and slope stability analysis under transient seepage: Impact of cross dependence structure

The objective of this section is to investigate the potential impact of cross-dependence structure in the same setting as in the previous subsection. For conducting the transient seepage analysis, HYDRUS-1D (Simunek et al., 1998) code was used. The cross dependence value in terms of Kendall's τ was taken as $(k, \alpha) = 0.7, (k, n) = -0.3, (\alpha, n) = -0.6$ after Li et al. (2009). The boundary conditions and the rest of the parameters are the same as in the previous section. Fig. 6.19 presents the results obtained at an interval of 16 hours. Unlike the steady seepage case, where no wetting front for $h(x)$ was observed, a sharp wetting front in Fig. 6.19 can be observed. As expected, the $h(x)$ ensembles move towards a region of higher h with time. Fig. 6.19 also shows

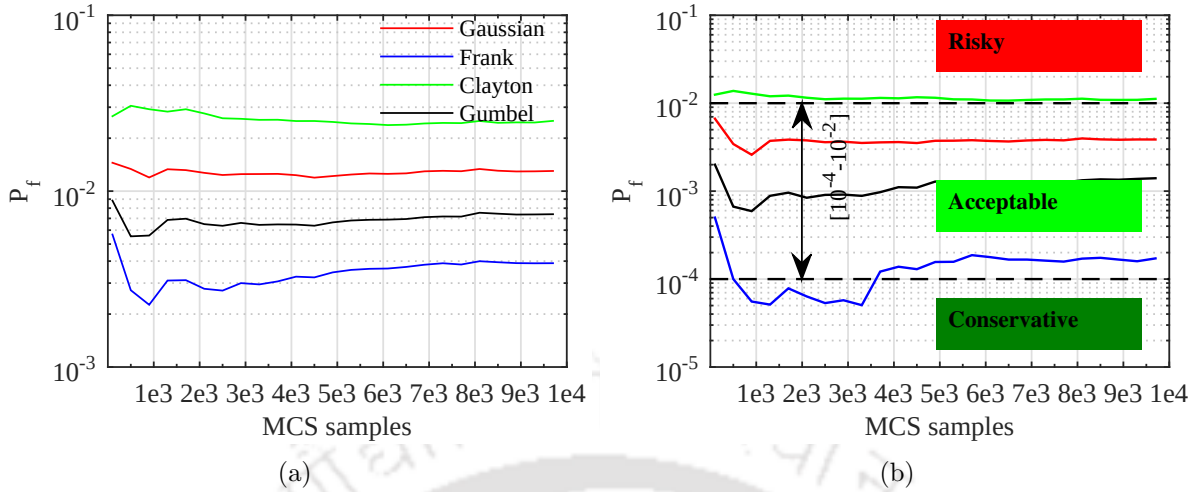


Figure 6.18: Comparison of failure probability (P_f) for steady seepage obtained using various copulas for spatial dependence of hydraulic conductivity K at (a) $q = -0.20$ m/day (b) $q = -0.15$ m/day.

the scatter in the obtained FS values. Here also, it can be observed that with time the $FS(x)$ ensemble moves towards lower FS values. Finally, similar to the steady seepage case in the previous session, P_f was calculated, and the results are summarized in Fig. 6.20. It can be noted that with time the difference among P_f for the two copulas used to model the cross dependence among them diminishes. This is because as time progresses, the flow moves towards a steady seepage condition and in that case, only the $K - h$ constitutive equation is required, and role of SWCC diminishes, therefore, the role of cross dependence. Apparently, it can also be noted that impact of the cross-dependence dependence structure in Fig. 6.20 is lower and negligible as compared to the spatial dependence structure in Figs. 6.18a and 6.18b

This subsection, along with the previous one, demonstrated the practical importance of considering non-Gaussian dependence structure in random fields of hydraulic parameters for stochastic seepage and slope stability analysis. Within the limitations and simplicity of infinite slope framework (Griffiths et al., 2011), it was shown that the choice of spatial dependence structure can significantly affect the P_f (by a factor of 100) in the entire acceptable design zone from 10^{-4} to 10^{-2}. The choice of cross dependence structure was not found to be significant in this study. However, only two copulas could be utilized for cross dependence, as only Gaussian and Frank (out of the four copulas utilized in this study) can handle negative ($\tau_{\alpha,n}$) as well as positive ($\tau_{k,\alpha}$) cross-dependence structure simultaneously. The main aim of this study was to present a framework to handle non-Gaussian spatial and cross-dependence in hydraulic parameters for stochastic seepage and slope stability analysis. Further studies utilizing this framework can be conducted under different statistical and geometric settings. Also, only stationary and one-dimensional random fields were considered in this study. Investigations in two and three dimensions along with non-stationary framework can also be considered to further explore the implications of dependence structure .

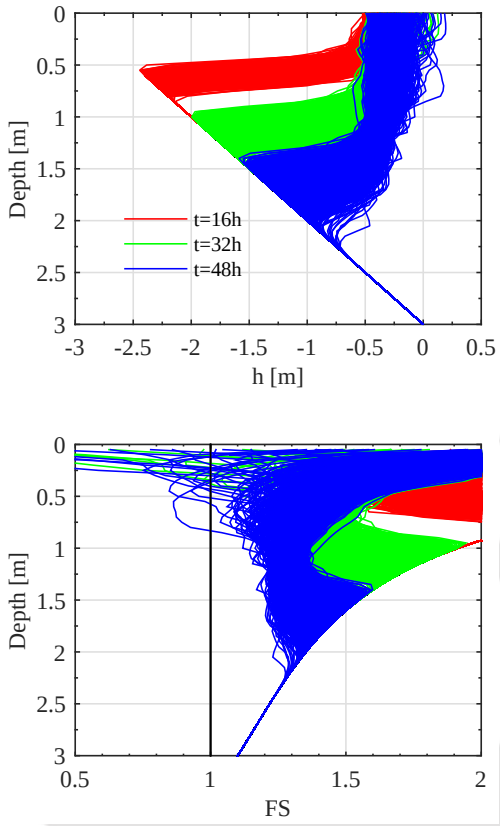


Figure 6.19: Profiles for pressure head and factor of safety at various time intervals with dependence among k_s, α, n modelled using gaussian copula. $N = 1000$.

6.5 Summary

In the context of stochastic seepage and slope stability analysis, this study presented a vine-copula based multivariate random field framework to handle the non-Gaussian (along with Gaussian) spatial and cross dependence structure among the random fields of hydraulic conductivity and van Genuchten (1980) soil-water characteristic curve parameters (α and n). Evidence for non-Gaussian spatial dependence structure was provided within the copula framework using a well documented hydraulic conductivity data set from Borden aquifer, Canada (Sudicky, 1986). It was shown that non-Gaussian dependence structure can be observed for most of the Borden aquifer data. Spatial dependence structure for 28 out of 33 cores were identified to be non-Gaussian on the basis of Akaike information criterion. This study not only provided the evidence of non-Gaussian spatial dependence structure but also identified the non-Gaussian copula. It was also shown that the vine copula approach can successfully model the non-Gaussian dependence structure. Two vine structures; C-vine and D-vine were utilized. D-vine was shown to be parsimonious in comparison to the C-vine structure. In the D-vine framework, it was shown that the number of parameters to be estimated is $n - 1$, whereas, for the C-vine, it is $n(n - 1)/2$.

For demonstrating the practical importance of this study, the engineering impact of dependence

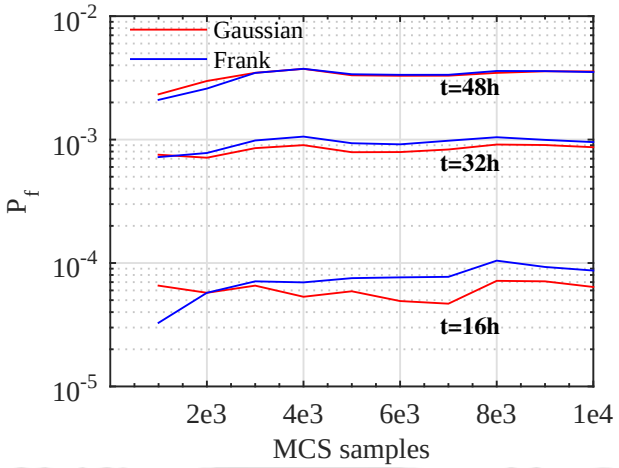


Figure 6.20: Comparison of failure probability (P_f) for transient seepage obtained using various copula for modelling cross dependence of K, α, n.

structure on stochastic seepage and slope stability analysis was investigated. It was shown that the choice of spatial dependence structure (copula) can significantly affect the quantiles of pressure head, factor of safety and finally, the failure probability (P_f)of slope. It was shown that, with the choice of spatial dependence structure, P_f could vary by a factor of 100 across the entire acceptable design zone from 10^{-4} to 10^{-2} . Apparently, it was also found that the choice of spatial dependence structure (seldom considered) is more crucial than the cross-dependence structure (the major application area of copulas in geotechnical engineering) in the context of stochastic seepage and slope stability analysis. The presented vine copula approach is more generalized than the conventional approaches for stochastic seepage and slope stability analysis (e.g. Covariance matrix decomposition or Karhunen–Loève (K-L) expansion) which can handle only Gaussian spatial and cross-dependence structure in the random fields of hydraulic parameters.

Acknowledgement

I would like to sincerely thank Professor (Emeritus) Edward Sudicky, Department of Earth and Environmental Sciences, University of Waterloo, Canada, for providing the Borden aquifer hydraulic conductivity data from Sudicky (1986) used in this study.

Chapter 7

Conclusions, contributions, limitations and future scope

Major conclusions, limitations and future scope associated with this study are outlined as follows.

7.1 Conclusions

Following are the major conclusions from this study.

- A database of bentonite SWCCs was compiled in this study based on the published literature, which exhibited wide variability. To quantify this variability, a probabilistic model for bentonite SWCCs was proposed in this work within the purview of van Genuchten (1980) model.
- The trivariate joint distribution of parameters α,n,m was constructed using a copula approach. Lognormal marginals were found to be suitable for modelling the univariate random behaviour of α,n,m. Correlations among the parameters were incorporated successfully in the trivariate distribution using Gaussian and 't' copulas.
- A Bayesian approach integrated with copula theory was formulated in this study. Prior information about the joint distribution parameters (marginals and copula) was incorporated with the limited SWCC data to obtain the updated joint distribution of SWCC parameters.
- It was shown that the statistical uncertainty in the joint distribution parameters obtained using the proposed approach are lower than the conventional methods in the literature.
- The design slope angles obtained using only three SWCCs, and the proposed approach were in close agreement with the design slope angle obtained using the original population.
- A multivariate probabilistic approach for the estimation of CEC and SSA was proposed in this study. For this purpose, this study constructed a 5-dimensional multivariate distribution of LL, PI, CF, CEC and SSA. A database (labelled as CLAY/C-S/5/278) was developed by the authors from literature and utilized for the construction of multivariate distribution.
- Bimodal lognormal marginals were found to be statistically satisfactory for LL and PI and right truncated lognormal marginals for CF, CEC and SSA. It was found that all the five variables in CLAY/C-S/5/278 database are positively correlated with each other with Kendall's

rank correlation coefficient τ varying from 0.33 to 0.87. It was also found that 8 out of 10 pair copula required for the construction of joint density are non-Gaussian, thereby indicating strong evidence of non-Gaussian dependence structure among the five parameters in the CLAY/C-S/5/278 database.

- It was found that the uncertainty (in terms of COV) in the estimates of CEC and SSA decrease with increasing values of LL, PI and CF. Among LL, PI and CF the COV in the estimates of SSA and CEC was found to be highest in general for CF.
- The estimated PDFs of CEC and SSA were also validated against a dataset outside the CLAY/C-S/5/278 database. It was found that the resulting estimates were consistent with the actual measured data.
- In the context of stochastic seepage and slope stability analysis, this study presented a vine-copula based multivariate random field framework to handle the non-Gaussian (along with Gaussian) spatial and cross dependence structure among the random fields of hydraulic conductivity and van Genuchten (1980) soil-water characteristic curve parameters (α and n).
- Evidence for non-Gaussian spatial dependence structure was provided within the copula framework using a well documented hydraulic conductivity data set from Borden aquifer, Canada (Sudicky, 1986). It was shown that non-Gaussian dependence structure can be observed for most of the Borden aquifer data. Spatial dependence structure for 28 out of 33 cores were identified to be non-Gaussian on the basis of Akaike information criterion.
- It was shown that, with the choice of spatial dependence structure, P_f could vary by a factor of 100 across the entire acceptable design zone from 10^{-4} to 10^{-2} . Apparently, it was also found that the choice of spatial dependence structure (seldom considered) is more crucial than the cross-dependence structure (the major application area of copulas in geotechnical engineering) in the context of stochastic seepage and slope stability analysis.

7.2 Contributions

Major contributions from this study are outlined below.

- A multivariate distribution for bentonite SWCC was constructed.
- A Bayesian approach integrated with copula theory was formulated.
- A multivariate probabilistic approach for the estimation of CEC and SSA was proposed.
- In the context of stochastic seepage and slope stability analysis, this study presented a vine-copula based multivariate random field framework to handle the non-Gaussian (along with Gaussian) spatial and cross dependence structure among the random fields of hydraulic conductivity and van Genuchten (1980) soil-water characteristic curve parameters (α and n).

7.3 Limitations

Following are some limitations associated with this study.

- The SWCCs in the compiled database for bentonites correspond to different geographical locations and the experimental procedure comprised of various initial densities, temperature, and salinity of the water. All of these factors contribute to the non-uniqueness of SWCC. Quantifying the uncertainties based on each of the above factors could not be performed in this study as the number of SWCCs in the database was insufficient for a satisfactory statistical estimation and inference.
- Only two elliptical copulas; 'Gaussian' and 't' could be utilized for the probabilistic characterization of bentonite SWCC. These two copulas assume an elliptical dependence structure.
- For the Bayesian approach, SWCCs were assumed to be from some definite copula or marginals. Ideally, Bayesian model selection should be performed instead of a priori designation of marginals and dependences structure (copula).
- Only Meteropolis-Hastings (M-H) algorithm was utilized for the Bayesian approach. Utilizing M-H algorithm results in a wastage of lot of samples and also requires proper tuning for convergence.
- The CLAY/C-S/5/278 database couldn't take into account the mineralogy of the clay minerals.
- All the formulations for vine copula utilize a key assumption of conditional independence, also known as simplifying assumption (Aas et al., 2009). This assumption essentially implies that all the pair copulas for conditional distribution or its inverse do not depend upon the conditioning variable. However, at the same time, it also implies that possible dependence occurring among conditional variables and the copula are neglected (Stoeber et al., 2013).
- Only stationary and one dimensional random fields were considered.
- Only infinite slope formulation was utilized for the application section of Bayesian approach.

7.4 Future scope

Some possible future scope of this study are outlined below.

- Only 60 SWCCs could be compiled for the bentonite database. Characterization of bentonite SWCCs based on a larger dataset can be considered for wider applicability.
- MCMC algorithms more efficient than Metropolis-Hastings, such as Hamiltonian Monte Carlo, can be incorporated in the proposed copula based Bayesian approach.

- Mineralogical data can be further incorporated into the probabilistic estimation of CEC and SSA.
- Vine copula approach can be extended for the non-stationary random fields and in higher (2D and 3D) dimensions.

List of publications

1. **Prakash, A.,** Hazra, B., and Sreedeep, S✉(2020). *Probabilistic Analysis of Soil-Water Characteristic Curve of Bentonite: Multivariate Copula Approach.* **International Journal of Geomechanics, ASCE,** 20(2), 04019150. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001554

2. **Prakash, A.,** Hazra, B.✉ and Sreedeep, S (2021). *Probabilistic analysis of soil-water characteristic curve from limited data.* **Applied Mathematical Modelling, Elsevier,** 89(1), 752-770. https://doi.org/10.1016/j.apm.2020.08.023

3. **Sharma, A.,**✉ Hazra, B., Spagnoli G*., and Sreedeep, S (2020). *Probabilistic estimation of Specific Surface Area and Cation Exchange Capacity: A Global Multivariate Distribution.* **Canadian Geotechnical Journal, NRC research press, (In Press)**. https://doi.org/10.1139/cgj-2020-0236

4. **Sharma, A.,**✉ Hazra, B., and Sreedeep, S. *Stochastic seepage and slope stability analysis using vine-copula based multivariate random field approach: consideration to non-Gaussian spatial and cross-dependence structure of hydraulic parameters.* **Computers and Geotechnics, Elsevier, (In Press).** https://doi.org/10.1016/j.compgeo.2020.103918

✉ corresponding author,* partial data contribution only, #from this thesis only, publications from masters' thesis and other works including the above can be found at https://scholar.google.com/citations?hl=en&user=QyHvscwAAAAJ&view_op=list_works&sortby=pubdate

Appendix

This appendix contains Section *"Multivariate probabilistic characterization"* and Tables DA1, DA2, DA3 and DA4.

Multivariate probabilistic characterization

This section describes some basic theory regarding multivariate probabilistic characterization using copula theory.

Theory

The joint distribution of a random vector, e.g., $X = \{X1, X2, X3\}$ can be constructed using copula theory (Nelsen, 2007) as follows:

$$F(x_1, x_2, x_3) = C[F_1(x1), F_2(x2), F_3(x3); \theta] \tag{7.1}$$

where $F(x1, x2, x3)$ is the trivariate joint CDF, $F_1(x1)$, $F_2(x2)$, $F_3(x3)$ are the one dimensional marginal CDFs and C is the copula function with dependence parameter matrix θ. Since $F_i(.)$ is defined over [0, 1], it is easy to observe that C is a multivariate distribution with uniform marginals. For the trivariate case, this means that a copula is defined over a unit cube $[0, 1]^3$. This also implies that a copula can be rewritten as joint CDF of $U1, U2, U3$ as:

$$[F_1(x1), F_2(x2), F_3(x3); \theta] = C(u1, u2, u3) = P(U1 \leq u1, U2 \leq u2, U3 \leq u3) \tag{7.2}$$

Where $u_i = F_i()$ and $P(.)$ represents the joint probability. The trivariate probability density function (PDF) can be obtained by taking the derivative of the joint CDF as:

$$f(x1, x2, x3) = f_1(x1) f_2(x2) f_3(x3) D(u1, u2, u3; \theta) \tag{7.3}$$

where $f(x1, x2, x3)$ is the joint PDF, $f_1(x1)$,$f_2(x2)$,$f_3(x3)$ are the one dimensional marginal PDFs and $D(u1, u2, u3; \theta) = \partial^3 C(u1, u2, u3; \theta)/\partial u1 \partial u2 \partial u3$ is the copula density function.

In the following section, the elliptical copula i.e., Gaussian and t-copula used in this study are presented.

Gaussian copula

The copula function for an n-dimensional Gaussian copula also known as a normal copula is expressed as:

$$C(u_i \dots u_n; \theta) = \Phi_\theta^n(\Phi^{-1}(u_i) \dots . \Phi^{-1}(u_n)) \tag{7.4}$$

Where $C(u_i \dots u_n; \theta)$ is the copula function with parameter matrix θ, $\Phi_\theta(.)$ denotes the multivariate standard normal function with linear correlation matrix θ, and Φ^{-1} denotes the quantile or inverse CDF of one-dimensional standard normal distribution. For the trivariate case i.e., $n = 3$, it can be expressed as:

$$C(u_1, u_2, u_3; \theta) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_3)} \frac{1}{(2\pi)^{3/2}|\theta|^{1/2}} exp(-\frac{1}{2} W^T \theta^{-1} W) dW \tag{7.5}$$

Where $W = [t_1, t_2, t_3]^T$ and t_1, t_2, t_3 are the corresponding integral variables. The copula density function $D(u_1, u_2, u_3; \theta) = \partial^3 C(u_1, u_2, u_3; \theta)/\partial u_1 \partial u_2 \partial u_3$ can be expressed as:

$$D(u_1, u_2, u_3; \theta) = \frac{1}{\theta^{1/2}} exp[-\frac{1}{2} \zeta^T \theta^{-1} \zeta - \zeta^T \zeta)] \tag{7.6}$$

Where $\zeta = [\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)]^T$

t copula

The copula function for an n-dimensional t-copula also known as student t copula can be written as follow:

$$C(u_i \dots u_n; \theta, \nu) = t_{\nu,\theta}^n(t_\nu^{-1}(u_i), \dots ., t_\nu^{-1}(u_n)) \tag{7.7}$$

Where $t_{v,\theta}^n$ represents the distribution function of $\sqrt{\nu} Y/\sqrt{S}$, where S χ_ν^2 and Y $N_n(0, \theta)$ are independent and θ is simply the linear correlation matrix for a degree of freedom $\nu \geq 2$. For the trivariate case i.e., $n = 3$ it can be expressed as:

$$C(u_1, u_2, u_3; \theta, \nu) = \int_{-\infty}^{(t_v)^{-1}(u_1)} \int_{-\infty}^{(t_v)^{-1}(u_2)} \int_{-\infty}^{(t_v)^{-1}(u_3)} \frac{\Gamma\frac{\nu+3}{2}}{\Gamma\frac{\nu}{2}} \frac{1}{(\pi\nu)^{3/2}|\theta|^{1/2}} (1 + \frac{W^T \theta^{-1} W}{\nu})^{-\frac{\nu+3}{2}} dW \tag{7.8}$$

Where $t_v^(-1)(.)$ denotes the inverse of student t distribution, $\Gamma(.)$ denotes the gamma function. The copula density function $D(u_1, u_2, u_3; \theta, \nu) = \partial^3 C(u_1, u_2, u_3; \theta)/\partial u_1 \partial u_2 \partial u_3$ can be expressed as:

$$D(u_1, u_2, u_3; \theta, \nu) = |\theta|^{-\frac{1}{2}} \frac{\Gamma\frac{\nu+3}{2}}{\Gamma\frac{\nu}{2}} [\frac{\Gamma\frac{\nu}{2}}{\Gamma\frac{\nu+1}{2}}]^3 \frac{(1 + \frac{\zeta^T \theta^{-1} \zeta}{\nu})^{-\frac{\nu+3}{2}}}{\prod_{i=1}^3 (1 + \frac{[t_\nu^{-1}(b_i)]^2}{\nu})^{\frac{\nu+1}{2}}} \tag{7.9}$$

where $b_i = t_\nu^{-1}(u_i)$ and $\zeta = [t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), t_\nu^{-1}(u_3)]^T$

Estimation

Maximum likelihood estimation The most popular choice for estimation of the copula is using the pseudo likelihood method (Fang et al. (2002)). It essentially consists of maximizing the log-pseudo likelihood function $l(\theta)$. For a trivariate case, it can be expressed as

$$\theta_n = \underset{\theta}{\operatorname{argmax}}\, l(\theta) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} log[D_\theta(U_i1, U_i2, U_i3)] \tag{7.10}$$

Where $D_\theta()$ denotes the copula density function, U_ij are the pseudo observation or rescaled empirical distribution of $Z_i = (Z_i1, Z_i2, Z_i3)$ and can be expressed as

$$U_ij = \frac{R_ij}{n+1} \tag{7.11}$$

Where R_ij is the rank of Z_ij among (Z_1j, Z_2j, Z_3j) when sorted in ascending order. For Gaussian and t copula Eq. 15 can be solved by minimizing the partial derivative of log-pseudo likelihood function as follows:

$$\sum_{i=1}^{n} \frac{\partial log[D_\theta^G((U_i1, U_i2, U_i3))]}{\partial\theta} = 0 \tag{7.12}$$

where $D_\theta^G()$ denotes the copula density function for Gaussian copula provided.

$$\begin{aligned} \sum_{i=1}^{n} \frac{\partial log[D_\theta^t((U_i1, U_i2, U_i3))]}{\partial\theta} &= 0 \\ \sum_{i=1}^{n} \frac{\partial log[D_\theta^t((U_i1, U_i2, U_i3))]}{\partial\nu} &= 0 \end{aligned} \tag{7.13}$$

Where $D_\theta^t()$ denotes the copula density function for t copula.

Bayesian estimation Instead of a single point estimate corresponding to the maximized likelihood, the θ can be estimated using Bayes' rule as follows:

$$f(\theta|data) = \frac{f(data|\theta)f(\theta)}{\int .. \int_\theta f(data|\theta)f(\theta)dW} \tag{7.14}$$

where $P(data|W)$ represents the likelihood of obtained SWCC data given the joint distribution parameters W. Assuming that the site specific N SWCCs i.e N pairs of α, n are independent realizations from the joint PDF $f(\alpha, n|W)$, $P(data|W)$ can be expressed as follows:

$$f(data|\theta) = \prod_{i=1}^{N} f(data_i|\theta) = \prod_{i=1}^{N} f(x_i1)f(x_i2)f(x_i3)D(U_i1, U_i2, U_i3|\theta) \tag{7.15}$$

$f(\theta)$ term in Eq. 7.14 represents the prior knowledge about the parameters θ.

Goodness of fit

For evaluation of goodness of fit among the above two copulas, the Cramer-von Mises statistics (CM) (Genest et al. (2009)) and AIC (Fang et al. (2014)) are used in this study and are discussed below.

Cramer-von Mises statistics

The CM statistics (S_n) represents the distance between true and observed copula and is based on a comparison of empirical copula with the derived copula under the null hypothesis. The same is given by

$$S_n = \int_{[0,1]^p} C_n^2(u) dC_n(u) = \sum_{i=1}^{n} (C_n(U_i) - C_{\theta_n}(U_i))^2 \tag{7.16}$$

Where $C_n(u) = \sqrt{n}C_n(u) - C_(\theta_n)(u)$ is the empirical copula and $C_(\theta_n)$ is the Copula under hypothesis and θ_n is the estimator of θ, computed using the ranked pseudo observations.

AIC

In this section AIC approach for copula selection after Fang et al. (2014) is discussed. Akaike information criterion is defined as

$$AIC = -2l(\theta_n) + 2(k) \tag{7.17}$$

Where $l(\theta_n)$ is the maximized log likelihood and k is the number of free parameters to be estimated. AIC, when applied to copula model, is a relative measure of goodness of fit of the model. Although it cannot perform formal goodness of fit hypothesis test, it can be used to select the better copula among the families chosen for study. Lehmann and Casella (2006) suggested using $K - L$ (Kullback and Leibler (1951)) information as a measure to represent the distance between true and hypothesized copula. Assuming the density of true copula and the estimated one under null hypothesis are given by $D(U)$ and $D_\theta(U)$ respectively, the $K - L$ information can be represented as

$$\begin{aligned} K(D(U), D_\theta(U)) &= \int log(\frac{D(U)}{D_\theta(U)}) D(U) dU \\ &= E[logD(U)] - E[logD_\theta(U)] \end{aligned} \tag{7.18}$$

Where $E(.)$ denotes the expectation or expected value. Smaller the $K-L$ information, closer is the hypothesized copula to the true one. Alternatively, larger the value of $E[logD_\theta(U)]$, closer is the hypothesized copula to the true one. Bozdogan (1987) suggested that AIC is an unbiased estimator of $-2E[logD_\theta(U)]$ i.e. $AIC = -2E[logD_\theta(U)]$. Substituting this into the above equation results in

$$K(D(U), D_\theta(U)) = E[logD(U)] + 1/2AIC \tag{7.19}$$

From the above equation, it can be observed that minimizing the AIC results in minimizing the K-L information. Hence the copula with minimum AIC can be selected as the favored model.

Table DA1: Details of the univariate distributions utilized in this study.

Distribution	PDF	CDF	Inverse CDF
Left Truncated Normal	$\frac{\phi[(x-\mu)/\sigma]}{\sigma 1-\Phi[(t-\mu)/\sigma]}$	$\frac{\Phi[(x-\mu)/\sigma]-\Phi[(t-\mu)/\sigma]}{1-\Phi[(t-\mu)/\sigma]}$	$\mu+\Phi^{-1}[\Phi(\frac{t-\mu}{\sigma})+U(1-\frac{t-\mu}{\sigma})]\sigma$
Lognormal	$(\sqrt{2\pi}x\sigma)^{-1}exp[-\frac{(lnx-\mu)^2}{2(\sigma)^2}]$	$0.5[1+erf\frac{lnx-\mu}{\sqrt{2}\sigma}$	$exp(\mu+\sigma\sqrt{2}erf^{-1}(2U-1))$
RightTruncated Lognormal	$\frac{logpdf(x,\mu,\sigma)}{logcdf(t,\mu,\sigma}$	$\frac{logcdf(x,\mu,\sigma)}{logcdf(t,mu,sigma}$	$loginv\{Ulogcdf[\frac{t-\mu}{\sigma},\mu,\sigma],\mu,\sigma]\}$
Weibull	$\frac{\kappa}{\lambda}(\frac{x}{\lambda})^{\kappa-1}exp((-\frac{x}{\lambda})^k$	$exp((-\frac{x}{\lambda})^k$	$\lambda[-log(1-U)]^{1/\kappa}$
Gamma	$\frac{1}{\Gamma(k)\theta^k}x^{k-1}exp(-x\theta)$	$\frac{1}{\Gamma(k)}\gamma(k,\frac{x}{\theta})$	Numerical CDF invsersion
Exponential	$\lambda exp(-\lambda x)$	$1-\lambda exp(-\lambda x)$	$-log(\frac{1-U}{\lambda})$
Bimodal Lognormal	$Plogpdf(x,\mu_1,\sigma_1)$ $+(1-P)logpdf(x,\mu_2,\sigma_2)$	$Plogcdf(x,\mu_1,\sigma_1)$ $+(1-P)logcdf(x,\mu_2,\sigma_2)$	Numerical CDF invsersion

Table DA2: Details of the four copulas utilized in this study.

Copula	Gaussian	Clayton	Frank	Gumbel
$C(u_1, u_2; \theta)$	$\Phi_\theta[\Phi^{-1}(u_1), \Phi^{-1}(u_2)]$	$(u_1^{-\theta} + u_2^{-\theta})^{-1/\theta}$	$-\frac{1}{\theta} log[1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}]$	$exp\{-[(-logu_1)^\theta + (-logu_2)^\theta]^{1/\theta}\}$
$D(u_1, u_2; \theta)$	$\frac{1}{\sqrt{1-\theta^2}} exp[-\frac{\zeta_1^2\theta^2+\zeta_2^2\theta^2-2\theta\zeta_1\zeta_2}{2(1-\theta^2)}]$ $\zeta_1 = \Phi^{-1}(u_1); \zeta_2 = \Phi^{-1}(u_2)$	$(1+\theta)(u_1u_2)^{-(\theta+1)}$ $\times(u_1^{-\theta} + u_2^{-\theta} - 1)^{-(2+\frac{1}{\theta})}$	$\frac{-\theta(e^{-\theta}-1)e^{-\theta(u_1+u_2)}}{[(e^{-\theta}-1)+(e^{-\theta u_1}-1)+(e^{-\theta u_2}-1)]^2}$	$C(u_1, u_2)(u_1, u_2)^{-1} \frac{(logu_1 logu_2)^{\theta-1}}{(-logu_1^\theta + -logu_2^\theta)^{2-\frac{2}{\theta}}}$ $\times[(-logu_1^\theta + -logu_2^\theta)^{-\frac{1}{\theta}}(\theta-1)+1]$
h function	$\Phi[\frac{\Phi^{-1}(u_1)-\theta\Phi^{-1}(u_2)}{\sqrt{1-\theta^2}}]$	$u_2^{-\theta-1}(u_1^{-\theta} + u_2^{-\theta-1})^{-1-1/\theta}$	$\frac{e^{-\theta u_2}}{e^{-\theta u_2}-1+\frac{1-e^{-\theta}}{1-e^{-\theta u_1}}}$	$\frac{C(u_1,u_2)}{u_2}(-logu_2)^{\theta-1}[(-logu_1)^\theta + (-logu_2)^\theta]^{1/\theta-1}$
h^{-1} function	$\Phi[\Phi^{-1}(u_1)\sqrt{1-\theta^2} + \theta\Phi^{-1}(u_2)]$	$[(u_1u_2^{\theta+1})^{-\theta/(\theta+1)} + 1 - u_2^{-\theta}]^{-1/\theta}$	$-log[1 - \frac{1-e^{-\theta}}{1+(u_1^{-1}-1)e^{-\theta u_2}}]/\theta$	NA

Table DA3: Some transformation models for CEC and SSA proposed in literature.

Model	Reference	Transformation	Range
SSA-LL	Mbonimpa et al. (2002)	$SSA = 0.2(LL)^{1.45}$	
	Chapuis and Aubertin (2003)	$SSA^{-1} = 1.35(LL)^{-1} - 0.009$	$LL < 110$
	Farrar and Coleman (1967)	$SSA = -14 + 1.48LL$	$LL > 9.46$
	Farrar and Coleman (1967)	$SSA = (LL - 19)/0.56$	$LL > 19$
	Arnepalli et al. (2008)	$SSA = (10LL)^{0.75}$	
	Smith et al. (1985)	$SSA = (LL - 19.8)/0.14$	$LL > 19.8$
	Spagnoli and Shimobe (2019)	$SSA = 1.28LL + 22.47$	
	Spagnoli and Shimobe (2019)	$SSA = 1.18LL - 5.67$	$LL > 4.8$
	Gill and Reaves (1957)	$SSA = (LL - 19)/0.21$	$LL > 19$
	De Bruyn et al. (1957)	$SSA = (LL - 35.22)/0.19$	$LL > 35.22$
	Warkentin (1968)	$SSA = (LL - 18)/0.37$	$LL > 18$
	Sridharan et al. (1988)	$SSA = (LL - 24.25)/0.31$	$LL > 24.25$
	Cerato (2001)	$SSA = (LL - 20.3)/0.77$	$LL > 20.3$
CEC-LL	Yilmaz (2004)	$CEC = exp(2.63 + 0.02LL)$	
	Farrar and Coleman (1967)	$CEC = -5 + 0.45LL$	$LL > 11.11$
	Farrar and Coleman (1967)	$CEC = (LL - 22)/1.80$	$LL > 22$
	Smith et al. (1985)	$CEC = (LL - 21.89)/0.57$	$LL > 21.89$
	Spagnoli and Shimobe (2019)	$CEC = 0.11LL + 17.09$	
	Yukselen and Kaya (2006)	$CEC = 0.20LL + 16.231$	

Table DA4: Data from Konrad (1999) utilized for the calculation of segregation potential for the Saints-Martyrs-Canadiens till.

Station (km)	Percent< $425\mu m$	Percent< $75\mu m$	d_50 μm	LL %
4.53	24	13	10	32
4.485	28	15	4	29
4.44	25	17	4	45
4.506	37	22	10	32
4.461	44	25	20	25
4.418	46	21	30	23
4.759	56	41	16	49
4.74	69	47	21	16
4.781	20	12	25	28
4.8	23	11	20	22
4.922	72	52	15	17
4.895	54	37	16	22
5.039	45	28	11	17
5.01	39	17	40	19
4.96	46	20	30	18
4.95	48	25	30	18
4.372	63	41	20	28
4.38	22	10	20	26

Bibliography

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009), 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and economics* **44**(2), 182–198.

Abhijit, D. and Sreedeep, S. (2014), 'Evaluation of measurement methodologies used for establishing water retention characteristic curve of fly ash', *Journal of Testing and Evaluation* **43**(5), 1066–1077.

Agus, S. and Schanz, T. (2006), Drying, wetting, and suction characteristic curves of a bentonite-sand mixture, *in* 'Unsaturated Soils 2006', pp. 1405–1414.

Akaike, H. (1974), A new look at the statistical model identification, *in* 'Selected Papers of Hirotugu Akaike', Springer, pp. 215–222.

Alonso, E., Romero, E. and Hoffmann, C. (2011), 'Hydromechanical behaviour of compacted granular expansive mixtures: experimental and constitutive study', *Géotechnique* **61**(4), 329.

Aloui, R., Hammoudeh, S. and Nguyen, D. K. (2013), 'A time-varying copula approach to oil and stock market dependence: The case of transition economies', *Energy Economics* **39**, 208–221.

Ang, A. H.-S. and Tang, W. H. (2007), *Probability concepts in engineering: emphasis on applications in civil & environmental engineering*, Vol. 1, Wiley New York.

Arifin, Y. F. (2008), 'Thermo-hydro-mechanical behavior of compacted bentonite-sand mixtures: an experimental study'.

Arnepalli, D., Shanthakumar, S., Rao, B. H. and Singh, D. (2008), 'Comparison of methods for determining specific-surface area of fine-grained soils', *Geotechnical and Geological Engineering* **26**(2), 121–132.

Arya, L. M. and Paris, J. F. (1981), 'A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data 1', *Soil Science Society of America Journal* **45**(6), 1023–1030.

ASTM, A. (2010), 'D4318-10 stardard test methods for liquid limit, plastic limit and plasticity index of soils, astm int', *West Conshohocken, Pa* .

Baille, W., Tripathy, S. and Schanz, T. (2014), 'Effective stress in clays of various mineralogy', *Vadose Zone Journal* **13**(5).

Ballarini, E., Graupner, B. and Bauer, S. (2017), 'Thermal–hydraulic–mechanical behavior of bentonite and sand-bentonite materials as seal for a nuclear waste repository: Numerical simulation of column experiments', *Applied Clay Science* **135**, 289–299.

Bárdossy, A. (2006), 'Copula-based geostatistical models for groundwater quality parameters', *Water Resources Research* **42**(11).

Bárdossy, A. and Li, J. (2008), 'Geostatistical interpolation using copulas', *Water Resources Research* **44**(7).

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**(1-4), 245–268.

Bedford, T. and Cooke, R. M. (2002), 'Vines: A new graphical model for dependent random variables', *Annals of Statistics* pp. 1031–1068.

Bharat, T. V. and Gapak, Y. (2018), 'Hydration kinetics of bentonite buffer material: Influence of vapor pressure, bentonite plasticity, and compaction density', *Applied Clay Science* **157**, 41–50.

Bozdogan, H. (1987), 'Model selection and akaike's information criterion (aic): The general theory and its analytical extensions', *Psychometrika* **52**(3), 345–370.

Bozorgzadeh, N., Harrison, J. P. and Escobar, M. D. (2019), 'Hierarchical bayesian modelling of geotechnical data: application to rock strength', *Géotechnique* pp. 1–15.

Brooks, R. and Corey, T. (1964), 'Hydraulic properties of porous media', *Hydrology Papers, Colorado State University* **24**, 37.

Burdine, N. (1953), 'Relative permeability calculations from pore size distribution data', *Journal of Petroleum Technology* **5**(3), 71–78.

Buzzi, O., Fityus, S. and Sloan, S. W. (2010), 'Use of expanding polyurethane resin to remediate expansive soil foundations', *Canadian Geotechnical Journal* **47**(6), 623–634.

Canli, E., Mergili, M., Thiebes, B. and Glade, T. (2018), 'Probabilistic landslide ensemble prediction systems: Lessons to be learned from hydrology', *Natural Hazards and Earth System Sciences* **18**(8), 2183–2202.

Cao, Z. and Wang, Y. (2014), 'Bayesian model comparison and characterization of undrained shear strength', *Journal of Geotechnical and Geoenvironmental Engineering* **140**(6), 04014018.

Carsel, R. F. and Parrish, R. S. (1988), 'Developing joint probability distributions of soil water retention characteristics', *Water resources research* **24**(5), 755–769.

Cerato, A. (2001), 'Influence of specific surface area on geotechnical characteristics of fine-grained soils', *Unpublished MSc Thesis, Department of Civil and Environmental Engineering, University of Massachusetts* .

Cerato, A. B. and Lutenegger, A. J. (2002), 'Determination of surface area of fine-grained soils by the ethylene glycol monoethyl ether (egme) method', *Geotechnical Testing Journal* **25**(3), 315–321.

Cerato, A. B. and Lutenegger, A. J. (2004), 'Determining intrinsic compressibility of fine-grained soils', *Journal of Geotechnical and Geoenvironmental Engineering* **130**(8), 872–877.

Chapman, H. (1965), 'Cation-exchange capacity', *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties* **9**, 891–901.

Chapuis, R. P. and Aubertin, M. (2003), 'On the use of the kozeny carman equation to predict the hydraulic conductivity of soils', *Canadian Geotechnical Journal* **40**(3), 616–628.

CHEN, B., QIAN, L., YE, W., CUI, Y. and WANG, J. (2006), 'Soil-water characteristic curves of gaomiaozi bentonite', *Chinese Journal of Rock Mechanics and Engineering* (4), 13.

Ching, J., Li, K.-H., Phoon, K.-K. and Weng, M.-C. (2018), 'Generic transformation models for some intact rock properties', *Canadian Geotechnical Journal* **55**(12), 1702–1741.

Ching, J., Lin, G.-H., Phoon, K.-K. and Chen, J. (2017), 'Correlations among some parameters of coarse-grained soils—the multivariate probability distribution model', *Canadian Geotechnical Journal* **54**(9), 1203–1220.

Ching, J. and Phoon, K.-K. (2012), 'Modeling parameters of structured clays as a multivariate normal distribution', *Canadian Geotechnical Journal* **49**(5), 522–545.

Ching, J. and Phoon, K.-K. (2014), 'Correlations among some clay parameters—the multivariate distribution', *Canadian Geotechnical Journal* **51**(6), 686–704.

Ching, J. and Phoon, K.-K. (2017), 'Characterizing uncertain site-specific trend function by sparse bayesian learning', *Journal of Engineering Mechanics* **143**(7), 04017028.

Ching, J. and Phoon, K.-K. (2018), 'Constructing site-specific multivariate probability distribution model using bayesian machine learning', *Journal of Engineering Mechanics* **145**(1), 04018126.

Ching, J., Phoon, K.-K. and Chen, C.-H. (2014), 'Modeling piezocone cone penetration (cptu) parameters of clays as a multivariate normal distribution', *Canadian Geotechnical Journal* **51**(1), 77–91.

Ching, J., Phoon, K.-K., Li, K.-H. and Weng, M.-C. (2018), 'Multivariate probability distribution for some intact rock properties', *Canadian Geotechnical Journal* (999), 1–18.

Ching, J., Phoon, K.-K., Stuedlein, A. W. and Jaksa, M. (2019), 'Identification of sample path smoothness in soil spatial variability', *Structural Safety* **81**, 101870.

Chittoori, B. and Puppala, A. J. (2011), 'Quantitative estimation of clay mineralogy in fine-grained soils', *Journal of Geotechnical and Geoenvironmental engineering* **137**(11), 997–1008.

Chiu, C., Yan, W. and Yuen, K.-V. (2012), 'Reliability analysis of soil–water characteristics curve and its application to slope stability analysis', *Engineering Geology* **135**, 83–91.

Cho, S. E. (2014), 'Probabilistic stability analysis of rainfall-induced landslides considering spatial variability of permeability', *Engineering Geology* **171**, 11–20.

Cornelis, W. M., Khlosi, M., Hartmann, R., Van Meirvenne, M. and De Vos, B. (2005), 'Comparison of unimodal analytical expressions for the soil-water retention curve', *Soil Science Society of America Journal* **69**(6), 1902–1911.

Correa, O., García, F., Bernal, G., Cardona, O. D. and Rodriguez, C. (2020), 'Early warning system for rainfall-triggered landslides based on real-time probabilistic hazard assessment', *Natural Hazards* **100**(1), 345–361.

Czado, C. (2010), Pair-copula constructions of multivariate copulas, *in* 'Copula theory and its applications', Springer, pp. 93–109.

Dai, Z., Samper, J., Wolfsberg, A. and Levitt, D. (2008), 'Identification of relative conductivity models for water flow and solute transport in unsaturated bentonite', *Physics and Chemistry of the Earth, Parts A/B/C* **33**, S177–S185.

De Bruyn, C., Collins, L. and Williams, A. (1957), 'The specific surface, water affinity and potential expansiveness of clays', *Clay Minerals Bulletin* **3**(17), 120–128.

Deka, A. (2015), A Study on the Water Retention and Contaminant Retention Behavior of Fly Ash, Bentonite and Its Mixes, PhD thesis.

Dexter, A. (1990), 'Changes in the matric potential of soil water with time after disturbance of soil by moulding', *Soil and tillage research* **16**(1-2), 35–50.

Dieudonne, A.-C., Della Vecchia, G. and Charlier, R. (2017), 'Water retention model for compacted bentonites', *Canadian Geotechnical Journal* **54**(7), 915–925.

Dou, H.-q., Han, T.-c., Gong, X.-n., Qiu, Z.-y. and Li, Z.-n. (2015), 'Effects of the spatial variability of permeability on rainfall-induced landslides', *Engineering Geology* **192**, 92–100.

Dutta, J. and Mishra, A. K. (2016), 'Consolidation behaviour of bentonites in the presence of salt solutions', *Applied Clay Science* **120**, 61–69.

Embrechts, P., Lindskog, F. and McNeil, A. (2001), 'Modelling dependence with copulas', *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich* .

Ering, P. and Babu, G. S. (2016), 'Probabilistic back analysis of rainfall induced landslide-a case study of malin landslide, india', *Engineering Geology* **208**, 154–164.

Fang, H.-B., Fang, K.-T. and Kotz, S. (2002), 'The meta-elliptical distributions with given marginals', *Journal of Multivariate Analysis* **82**(1), 1–16.

Fang, Y., Madsen, L. and Liu, L. (2014), 'Comparison of two methods to check copula fitting.', *International Journal of Applied Mathematics* **44**(1).

Farrar, D. and Coleman, J. (1967), 'The correlation of surface area with other properties of nineteen british clay soils', *Journal of Soil Science* **18**(1), 118–124.

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N. and Bobée, B. (2004), 'Multivariate hydrological frequency analysis using copulas', *Water resources research* **40**(1).

Fenton, G. A. and Griffiths, D. V. (1997), 'Extreme hydraulic gradient statistics in stochastic earth dam', *Journal of geotechnical and geoenvironmental engineering* **123**(11), 995–1000.

Fenton, G. A. and Vanmarcke, E. H. (1990), 'Simulation of random fields via local average subdivision', *Journal of Engineering Mechanics* **116**(8), 1733–1749.

Ferrante, F., Arwade, S. and Graham-Brady, L. (2005), 'A translation model for non-stationary, non-gaussian random processes', *Probabilistic engineering mechanics* **20**(3), 215–228.

Fredlund, D. G., Rahardjo, H. and Fredlund, M. D. (2012), *Unsaturated Soil Mechanics in Engineering Practice*, John Wiley & Sons.

Fredlund, D. G. and Xing, A. (1994), 'Equations for the soil-water characteristic curve', *Canadian geotechnical journal* **31**(4), 521–532.

Gallipoli, D. (2012), 'A hysteretic soil-water retention model accounting for cyclic variations of suction and void ratio', *Geotechnique* **62**(7), 605–616.

Gardner, W. (1958), 'Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table', *Soil science* **85**(4), 228–232.

Gatabin, C., Talandier, J., Collin, F., Charlier, R. and Dieudonné, A.-C. (2016), 'Competing effects of volume change and water uptake on the water retention behaviour of a compacted mx-80 bentonite/sand mixture', *Applied Clay Science* **121**, 57–62.

Gates, W. P., Bouazza, A. and Churchman, G. J. (2009), 'Bentonite clay keeps pollutants at bay', *Elements* **5**(2), 105–110.

Gelman, A., Simpson, D. and Betancourt, M. (2017), 'The prior can often only be understood in the context of the likelihood', *Entropy* **19**(10), 555.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013), *Bayesian data analysis*, Chapman and Hall/CRC.

Genest, C. and Favre, A.-C. (2007), 'Everything you always wanted to know about copula modeling but were afraid to ask', *Journal of hydrologic engineering* **12**(4), 347–368.

Genest, C., Favre, A.-C., Béliveau, J. and Jacques, C. (2007), 'Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data', *Water Resources Research* **43**(9).

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995), 'A semiparametric estimation procedure of dependence parameters in multivariate families of distributions', *Biometrika* **82**(3), 543–552.

Genest, C., Rémillard, B. and Beaudoin, D. (2009), 'Goodness-of-fit tests for copulas: A review and a power study', *Insurance: Mathematics and economics* **44**(2), 199–213.

Gian, Q. A., Tran, D.-T., Nguyen, D. C., Nhu, V. H. and Tien Bui, D. (2017), 'Design and implementation of site-specific rainfall-induced landslide early warning and monitoring system: a case study at nam dan landslide (vietnam)', *Geomatics, Natural Hazards and Risk* **8**(2), 1978–1996.

Gill, W. R. and Reaves, C. A. (1957), 'Relationships of atterberg limits and cation-exchange capacity to some physical properties of soil 1', *Soil Science Society of America Journal* **21**(5), 491–494.

Gitipour, S., Bowers, M. T. and Bodocsi, A. (1997), 'The use of modified bentonite for removal of aromatic organics from contaminated soil', *Journal of Colloid and Interface Science* **196**(2), 191–198.

Godt, J. W., Şener-Kaya, B., Lu, N. and Baum, R. L. (2012), 'Stability of infinite slopes under transient partially saturated seepage conditions', *Water Resources Research* **48**(5).

Gräler, B. (2014), 'Modelling skewed spatial random fields through the spatial vine copula', *Spatial Statistics* **10**, 87–102.

Gräler, B. and Pebesma, E. (2011), 'The pair-copula construction for spatial data: a new approach to model spatial dependency', *Procedia Environmental Sciences* **7**, 206–211.

Griffiths, D. V. and Fenton, G. A. (1993), 'Seepage beneath water retaining structures founded on spatially random soil', *Geotechnique* **43**(4), 577–587.

Griffiths, D. V. and Fenton, G. A. (2008), 'Risk assessment in geotechnical engineering', *John wiley&Sons, Inc* pp. 381–400.

Griffiths, D. V., Huang, J. and Fenton, G. A. (2011), 'Probabilistic infinite slope analysis', *Computers and Geotechnics* **38**(4), 577–584.

Griffiths, D. V. and Lu, N. (2005), 'Unsaturated slope stability analysis with steady infiltration or evaporation using elasto-plastic finite elements', *International journal for numerical and analytical methods in geomechanics* **29**(3), 249–267.

Gui, S., Zhang, R., Turner, J. P. and Xue, X. (2000), 'Probabilistic slope stability analysis with stochastic soil hydraulic conductivity', *Journal of Geotechnical and Geoenvironmental Engineering* **126**(1), 1–9.

Gupt, C., Yamsani, S., Prakash, A., Medhi, C. and Sreedeep, S. (2018), 'Appropriate liquid-to-solid ratio for sorption studies of bentonite', *Journal of Environmental Engineering* **145**(2), 04018138.

Haslauer, C., Guthke, P., Bárdossy, A. and Sudicky, E. (2012), 'Effects of non-gaussian copula-based hydraulic conductivity fields on macrodispersion', *Water Resources Research* **48**(7).

Hazra, B., Gadi, V., Garg, A., Ng, C. and Das, G. (2017), 'Probabilistic analysis of suction in homogeneously vegetated soils', *Catena* **149**, 394–401.

Hökmark, H. (2004), 'Hydration of the bentonite buffer in a kbs-3 repository', *Applied Clay Science* **26**(1-4), 219–233.

Huang, S., Barbour, S. and Fredlund, D. (1998), 'Development and verification of a coefficient of permeability function for a deformable unsaturated soil', *Canadian Geotechnical Journal* **35**(3), 411–425.

Jaynes, E. T. (1957), 'Information theory and statistical mechanics. ii', *Physical review* **108**(2), 171.

Joe, H. (1997), *Multivariate models and multivariate dependence concepts*, CRC Press.

Joe, H. and Kurowicka, D. (2011), *Dependence modeling: vine copula handbook*, World Scientific.

Kale, R. C. and Ravi, K. (2018), 'Influence of thermal loading on index and physicochemical properties of barmer bentonite', *Applied Clay Science* **165**, 22–39.

Kao, S.-C. and Govindaraju, R. S. (2008), 'Trivariate statistical analysis of extreme rainfall events via the plackett family of copulas', *Water Resources Research* **44**(2).

Kayabali, K. (1997), 'Engineering aspects of a novel landfill liner material: bentonite-amended natural zeolite', *Engineering Geology* **46**(2), 105–114.

Kéry, M. and Schaub, M. (2011), *Bayesian population analysis using WinBUGS: a hierarchical perspective*, Academic Press.

Khalili, N., Habte, M. and Zargarbashi, S. (2008), 'A fully coupled flow deformation model for cyclic analysis of unsaturated soils including hydraulic and mechanical hystereses', *Computers and Geotechnics* **35**(6), 872–889.

Khorshidi, M. and Lu, N. (2017), 'Determination of cation exchange capacity from soil water retention curve', *Journal of Engineering Mechanics* **143**(6), 04017023.

Khorshidi, M., Lu, N., Akin, I. D. and Likos, W. J. (2016), 'Intrinsic relationship between specific surface area and soil water retention', *Journal of Geotechnical and Geoenvironmental Engineering* **143**(1), 04016078.

Konrad, J.-M. (1999), 'Frost susceptibility related to soil index properties', *Canadian Geotechnical Journal* **36**(3), 403–417.

Konrad, J.-M. (2005), 'Estimation of the segregation potential of fine-grained soils using the frost heave response of two reference soils', *Canadian Geotechnical Journal* **42**(1), 38–50.

Kullback, S. and Leibler, R. A. (1951), 'On information and sufficiency', *The annals of mathematical statistics* **22**(1), 79–86.

Kumar, S. and Yong, W.-L. (2002), 'Effect of bentonite on compacted clay landfill barriers', *Soil and sediment contamination* **11**(1), 71–89.

Kuriqi, A., Ardiçlioglu, M. and Muceku, Y. (2016), 'Investigation of seepage effect on river dike's stability under steady state and transient conditions', *Pollack Periodica* **11**(2), 87–104.

Lehmann, E. L. and Casella, G. (2006), *Theory of point estimation*, Springer Science & Business Media.

Leshchinsky, B., Vahedifard, F., Koo, H.-B. and Kim, S.-H. (2015), 'Yumokjeong landslide: an investigation of progressive failure of a hillslope using the finite element method', *Landslides* **12**(5), 997–1005.

Li, D.-Q., Tang, X.-S., Phoon, K.-K., Chen, Y.-F. and Zhou, C.-B. (2013), 'Bivariate simulation using copula and its application to probabilistic pile settlement analysis', *International Journal for Numerical and Analytical Methods in Geomechanics* **37**(6), 597–617.

Li, D.-Q., Xiao, T., Zhang, L.-M. and Cao, Z.-J. (2019), 'Stepwise covariance matrix decomposition for efficient simulation of multivariate large-scale three-dimensional random fields', *Applied Mathematical Modelling* **68**, 169–181.

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.

Li, W., Lu, Z. and Zhang, D. (2009), 'Stochastic analysis of unsaturated flow with probabilistic collocation method', *Water Resources Research* **45**(8).

Likos, W. J. and Yao, J. (2014), 'Effects of constraints on van genuchten parameters for modeling soil-water characteristic curves', *Journal of Geotechnical and Geoenvironmental Engineering* **140**(12), 06014013.

Ling, W., Shen, Q., Gao, Y., Gu, X. and Yang, Z. (2008), 'Use of bentonite to control the release of copper from contaminated soils', *Soil Research* **45**(8), 618–623.

Lloret, A., Romero, E. and Villar, M. V. (2005), *FEBEX II Project: Final report on thermo-hydro-mechanical laboratory tests*, Enresa.

Low, P. (1980), 'Theswellingofclay: Ii. montmorillonites', *Soil Sci. Soc. Amer. J* **44**, 667–676.

Lu, N. and Godt, J. (2008), 'Infinite slope stability under steady unsaturated seepage conditions', *Water Resources Research* **44**(11).

Lu, N. and Griffiths, D. (2004), 'Profiles of steady-state suction stress in unsaturated soils', *Journal of Geotechnical and Geoenvironmental Engineering* **130**(10), 1063–1076.

Lu, N. and Mitchell, J. K. (2019), *Geotechnical Fundamentals for Addressing New World Challenges*, Springer.

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**, 103340.

Marcial, D. (2013), 'Measuring water retention properties of a series of bentonite clays a wide range of suctions', *Advances in unsaturated soils. Taylor Francis Group, London* pp. 135–140.

Massey Jr, F. J. (1951), 'The kolmogorov-smirnov test for goodness of fit', *Journal of the American statistical Association* **46**(253), 68–78.

Mbonimpa, M., Aubertin, M., Chapuis, R. and Bussière, B. (2002), 'Practical pedotransfer functions for estimating the saturated hydraulic conductivity', *Geotechnical & Geological Engineering* **20**(3), 235–259.

Mbonimpa, M., Aubertin, M., Maqsoud, A. and Bussière, B. (2006), 'Predictive model for the water retention curve of deformable clayey soils', *Journal of Geotechnical and Geoenvironmental Engineering* **132**(9), 1121–1132.

Meyer, P., Rockhold, M. and Gee, G. (1997), Uncertainty analyses of infiltration and subsurface flow and transport for sdmp sites, Technical report, Nuclear Regulatory Commission.

Mishra, A. K., Ohtsubo, M., Li, L. and Higashi, T. (2011), 'Controlling factors of the swelling of various bentonites and their correlations with the hydraulic conductivity of soil-bentonite mixtures', *Applied Clay Science* **52**(1-2), 78–84.

Mitchell, J. K., Soga, K. et al. (2005), *Fundamentals of soil behavior*, Vol. 3, John Wiley & Sons Hoboken, NJ.

Montoya-Noguera, S., Zhao, T., Hu, Y., Wang, Y. and Phoon, K.-K. (2019), 'Simulation of non-stationary non-gaussian random fields from sparse measurements using bayesian compressive sampling and karhunen-loève expansion', *Structural Safety* **79**, 66–79.

Mualem, Y. (1976), 'A new model for predicting the hydraulic conductivity of unsaturated porous media', *Water resources research* **12**(3), 513–522.

Muceku, Y., Korini, O. and Kuriqi, A. (2016), 'Geotechnical analysis of hill's slopes areas in heritage town of berati, albania', *Periodica Polytechnica Civil Engineering* **60**(1), 61–73.

Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science & Business Media.

Nemes, A. d., Schaap, M., Leij, F. and Wösten, J. (2001), 'Description of the unsaturated soil hydraulic database unsoda version 2.0', *Journal of Hydrology* **251**(3-4), 151–162.

Nuth, M. and Laloui, L. (2008), 'Advances in modelling hysteretic water retention curve in deformable soils', *Computers and Geotechnics* **35**(6), 835–844.

Ohtsubo, M., Takayama, M. and EGASHIRA, K. (1983), 'Relationships of consistency limits and activity to some physical and chemical properties of ariake marine clays', *Soils and foundations* **23**(1), 38–46.

Orr, T. L. (2000), 'Selection of characteristic values and partial factors in geotechnical designs to eurocode 7', *Computers and Geotechnics* **26**(3-4), 263–279.

Osinubi, K. J. and Amadi, A. A. (2009), Hydraulic performance of compacted lateritic soil-bentonite mixtures permeated with municipal solid waste landfill leachate, Technical report.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010), 'Likelihood-based inference for max-stable processes', *Journal of the American Statistical Association* **105**(489), 263–277.

Patton, A. J. (2003), 'Applications of copula theory in financial econometrics.'.

Patton, A. J. (2009), Copula–based models for financial time series, *in* 'Handbook of financial time series', Springer, pp. 767–785.

Perret, D., Locat, J. and Martignoni, P. (1996), 'Thixotropic behavior during shear of a fine-grained mud from eastern canada', *Engineering Geology* **43**(1), 31–44.

Pham, K., Kim, D., Choi, H.-J., Lee, I.-M. and Choi, H. (2018), 'A numerical framework for infinite slope stability analysis under transient unsaturated seepage conditions', *Engineering Geology* **243**, 36–49.

Phoon, K., Huang, H. and Quek, S. (2005), 'Simulation of strongly non-gaussian processes using karhunen–loeve expansion', *Probabilistic engineering mechanics* **20**(2), 188–198.

Phoon, K.-K. (2018), 'Probabilistic site characterization', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **4**(4), 02018002.

Phoon, K.-K. and Ching, J. (2014), *Risk and reliability in geotechnical engineering*, CRC Press.

Phoon, K.-K., Quek, S.-T. and An, P. (2003), 'Identification of statistically homogeneous soil layers using modified bartlett statistics', *Journal of Geotechnical and Geoenvironmental Engineering* **129**(7), 649–659.

Phoon, K.-K. and Retief, J. V. (2016), *Reliability of geotechnical structures in ISO2394*, CRC Press.

Phoon, K.-K., Santoso, A. and Quek, S.-T. (2010), 'Probabilistic analysis of soil-water characteristic curves', *Journal of Geotechnical and Geoenvironmental Engineering* **136**(3), 445–455.

Prakash, A., Bordoloi, S., Hazra, B., Garg, A., Sreedeep, S. and Wang, Q. (2019), 'Probabilistic analysis of soil suction and cracking in fibre-reinforced soil under drying–wetting cycles in india', *Environmental Geotechnics* pp. 1–16.

Prakash, A., Hazra, B., Deka, A. and Sreedeep, S. (2017), 'Probabilistic analysis of water retention characteristic curve of fly ash', *International Journal of Geomechanics* **17**(12), 04017111.

Prakash, A., Hazra, B. and Sreedeep, S. (2018*a*), 'Probabilistic analysis of unsaturated fly ash slope', *Journal of Hazardous, Toxic, and Radioactive Waste* **23**(1), 06018002.

Prakash, A., Hazra, B. and Sreedeep, S. (2018*b*), 'Uncertainty quantification in water retention characteristic curve of fly ash using copulas', *Journal of Testing and Evaluation* **47**(4), 3080–3102.

Rahardjo, H., Ong, T., Rezaur, R. and Leong, E. C. (2007), 'Factors controlling instability of homogeneous soil slopes under rainfall', *Journal of Geotechnical and Geoenvironmental Engineering* **133**(12), 1532–1543.

Raj Singh, S., Prakash, A., Garg, A., Hazra, B. and Kumar Das, G. (2018), 'Stochastic modeling of relative permeability for vegetated covers', *International Journal of Geomechanics* **18**(9), 06018020.

Rao, S. M. and Ravi, K. (2013), 'Hydro-mechanical characterization of barmer 1 bentonite from rajasthan, india', *Nuclear Engineering and Design* **265**, 330–340.

Ravi, K. and Rao, S. M. (2013), 'Influence of infiltration of sodium chloride solutions on swcc of compacted bentonite–sand specimens', *Geotechnical and Geological Engineering* **31**(4), 1291–1303.

Rizzi, M., Seiphoori, A., Ferrari, A., Ceresetti, D. and Laloui, L. (2011), 'Analysis of the behaviour of the granular mx-80 bentonite in thm-processes', *Lausanne Swiss Fed Inst Technol Orders* **7**, 928.

Rózsás, Á. and Mogyorósi, Z. (2017), 'The effect of copulas on time-variant reliability involving time-continuous stochastic processes', *Structural Safety* **66**, 94–105.

Salgado, R. and Kim, D. (2014), 'Reliability analysis of load and resistance factor design of slopes', *Journal of Geotechnical and Geoenvironmental Engineering* **140**(1), 57–73.

Salvadori, G. and De Michele, C. (2007), 'On the use of copulas in hydrology: theory and practice', *Journal of Hydrologic Engineering* **12**(4), 369–380.

Santamarina, J. C., Klein, K. A., Wang, Y.-H. and Prencke, E. (2002), 'Specific surface: determination and relevance', *Canadian Geotechnical Journal* **39**(1), 233–241.

Santhoshkumar, T., Abraham, B., Sridharan, A. and Jose, B. (2016), 'Role of bentonite in improving the efficiency of cement grouting in coarse sand', *Geotechnical Engineering Journal of the SEAGS & AGSSEA* **47**(3).

Santoso, A. M., Phoon, K.-K. and Quek, S.-T. (2011), 'Effects of soil spatial variability on rainfall-induced landslides', *Computers & Structures* **89**(11-12), 893–900.

Schaap, M. G., Leij, F. J. and Van Genuchten, M. T. (2001), 'Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions', *Journal of hydrology* **251**(3-4), 163–176.

Schanz, T. and Tripathy, S. (2009), 'Swelling pressure of a divalent-rich bentonite: Diffuse double-layer theory revisited', *Water Resources Research* **45**(5).

Schwarz, G. et al. (1978), 'Estimating the dimension of a model', *The annals of statistics* **6**(2), 461–464.

Schweizer, B., Wolff, E. F. et al. (1981), 'On nonparametric measures of dependence for random variables', *The annals of statistics* **9**(4), 879–885.

Schwing, M., Chen, Z., Scheuermann, A. and Wagner, N. (2013), Dielectric properties of a clay soil determined in the frequency range from 1 mhz to 40 ghz, *in* 'Proc 10th Int. Conf. Electromagn. Wave Interact. Water and Moist Substances', pp. 242–250.

Seber, G. and Wild, C. (2003), *Nonlinear regression. 2003*, John Wiley and Sons.

Seiphoori, A., Ferrari, A. and Laloui, L. (2014), 'Water retention behaviour and microstructural evolution of mx-80 bentonite during wetting and drying cycles', *Géotechnique* **64**(9), 721–734.

Shields, M., Deodatis, G. and Bocchini, P. (2011), 'A simple and efficient methodology to approximate a general non-gaussian stationary stochastic process by a translation process', *Probabilistic Engineering Mechanics* **26**(4), 511–519.

Shih, J. H. and Louis, T. A. (1995), 'Inferences on the association parameter in copula models for bivariate survival data', *Biometrics* pp. 1384–1399.

Sillers, W. S. and Fredlund, D. G. (2001), 'Statistical assessment of soil-water characteristic curve models for geotechnical engineering', *Canadian Geotechnical Journal* **38**(6), 1297–1313.

Simunek, J., Sejna, M. and Van Genuchten, M. T. (1998), 'Hydrus-1d', *Simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media, version* **2**.

Sivakumar Babu, G. and Murthy, D. (2005), 'Reliability analysis of unsaturated soil slopes', *Journal of geotechnical and geoenvironmental engineering* **131**(11), 1423–1428.

Sivapullaiah, P. and Lakshmikanthay, H. (2005), 'Lime-stabilised illite as a liner', *Proceedings of the Institution of Civil Engineers-Ground Improvement* **9**(1), 39–45.

Sivapullaiah, P., Prasad, B. G. and Allam, M. (2008), 'Methylene blue surface area method to correlate with specific soil properties', *Geotechnical Testing Journal* **31**(6), 503–512.

Sklar, M. (1959), 'Fonctions de repartition an dimensions et leurs marges', *Publ. inst. statist. univ. Paris* **8**, 229–231.

Smith, C., Hadas, A., Dan, J. and Koyumdjisky, H. (1985), 'Shrinkage and atterberg limits in relation to other properties of principal soil types in israel', *Geoderma* **35**(1), 47–65.

Song, S. and Singh, V. P. (2010), 'Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data', *Stochastic Environmental Research and Risk Assessment* **24**(3), 425–444.

Spagnoli, G., Feinendegen, M. and Rubinos, D. (2013), 'Modification of clay adhesion to improve tunnelling excavation', *Proceedings of the Institution of Civil Engineers-Ground Improvement* **166**(1), 21–31.

Spagnoli, G. and Shimobe, S. (2019), 'A statistical reappraisal of the relationship between liquid limit and specific surface area, cation exchange capacity and activity of clays', *Journal of Rock Mechanics and Geotechnical Engineering* .

Spagnoli, G., Sridharan, A., Oreste, P., Bellato, D. and Di Matteo, L. (2018), 'Statistical variability of the correlation plasticity index versus liquid limit for smectite and kaolinite', *Applied Clay Science* **156**, 152–159.

Sridharan, A., Rao, S. and Murthy, N. (1988), 'Liquid limit of kaolinitic soils', *Geotechnique* **38**(2), 191–198.

Srivastava, A., Babu, G. S. and Haldar, S. (2010), 'Influence of spatial variability of permeability property on steady state seepage flow and slope stability analysis', *Engineering Geology* **110**(3-4), 93–101.

Stoeber, J., Joe, H. and Czado, C. (2013), 'Simplified pair copula constructions—limitations and extensions', *Journal of Multivariate Analysis* **119**, 101–118.

Sudicky, E. A. (1986), 'A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process', *Water Resources Research* **22**(13), 2069–2082.

Sutherland, R., Wan, Y., Ziegler, A., Lee, C.-T. and El-Swaify, S. (1996), 'Splash and wash dynamics: an experimental investigation using an oxisol', *Geoderma* **69**(1-2), 85–103.

Tang, X.-S., Li, D.-Q., Rong, G., Phoon, K.-K. and Zhou, C.-B. (2013), 'Impact of copula selection on geotechnical reliability under incomplete probability information', *Computers and Geotechnics* **49**, 264–278.

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

Thakur, V. K., Sreedeep, S. and Singh, D. N. (2006), 'Evaluation of various pedo-transfer functions for developing soil-water characteristic curve of a silty soil', *Geotechnical Testing Journal* **30**(1), 25–30.

Tripathy, S., Sridharan, A. and Schanz, T. (2004), 'Swelling pressures of compacted bentonites from diffuse double layer theory', *Canadian Geotechnical Journal* **41**(3), 437–450.

Tripathy, S., Tadza, M. Y. M. and Thomas, H. R. (2014), 'Soil-water characteristic curves of clays', *Canadian geotechnical journal* **51**(8), 869–883.

Tripathy, S., Thomas, H. R. and Bag, R. (2015), 'Geoenvironmental application of bentonites in underground disposal of nuclear waste: Characterization and laboratory tests', *Journal of Hazardous, Toxic, and Radioactive Waste* **21**(1), D4015002.

Trivedi, P. K., Zimmer, D. M. et al. (2007), 'Copula modeling: an introduction for practitioners', *Foundations and Trends® in Econometrics* **1**(1), 1–111.

Van Der Linde, A. (2005), 'Dic in variable selection', *Statistica Neerlandica* **59**(1), 45–56.

van Genuchten, M. T. (1980), 'A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1', *Soil science society of America journal* **44**(5), 892–898.

Vanmarcke, E. (2010), *Random fields: analysis and synthesis*, World scientific.

Vanmarcke, E. H. (1977), 'Probabilistic modeling of soil profiles', *Journal of the geotechnical engineering division* **103**(11), 1227–1246.

Villar, M. (2002), 'Thermo-hydro-mechanical characterisation of a bentonite from cabo de gata. a study applied to the use of bentonite as sealing material in high level radioactive waste repositories', *Publicación Técnica ENRESA* **1**(2002), 258.

Villar, M. (2005), Mx-80 bentonite. thermal-hydro-mechanical characterisation performed at ciemat in the context of the prototype project, Technical report, Centro de Investigaciones Energeticas.

Villar, M., Sánchez, M. and Gens, A. (2008), 'Behaviour of a bentonite barrier in the laboratory: experimental results up to 8 years and numerical simulation', *Physics and Chemistry of the Earth, Parts A/B/C* **33**, S476–S485.

Villar, M. V. and Lloret, A. (2004), 'Influence of temperature on the hydro-mechanical behaviour of a compacted bentonite', *Applied Clay Science* **26**(1-4), 337–350.

Wang, F. and Li, H. (2017), 'Towards reliability evaluation involving correlated multivariates under incomplete probability information: A reconstructed joint probability distribution for isoprobabilistic transformation', *Structural Safety* **69**, 1–10.

Wang, F. and Li, H. (2018), 'The role of copulas in random fields: Characterization and application', *Structural Safety* **75**, 75–88.

Wang, F. and Li, H. (2019), 'On the need for dependence characterization in random fields: Findings from cone penetration test (cpt) data', *Canadian Geotechnical Journal* **56**(5), 710–719.

Wang, H., Sun, P., Zhang, S., Han, S., Li, X., Wang, T., Guo, Q. and Xin, P. (2020), 'Rainfall-induced landslide in loess area, northwest china: a case study of the changhe landslide on september 14, 2019, in gansu province', *Landslides* pp. 1–16.

Wang, L., Cao, Z.-J., Li, D.-Q., Phoon, K.-K. and Au, S.-K. (2018), 'Determination of site-specific soil-water characteristic curve from a limited number of test data–a bayesian perspective', *Geoscience Frontiers* **9**(6), 1665–1677.

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.

Wang, Y. and Cao, Z. (2013), 'Probabilistic characterization of young's modulus of soil using equivalent samples', *Engineering Geology* **159**, 106–118.

Wang, Y., Zhao, T. and Cao, Z. (2015), 'Site-specific probability distribution of geotechnical properties', *Computers and Geotechnics* **70**, 159–168.

Warkentin, B. (1968), 'Properties of clay and silt soils of the eastern abitibi area in quebec', *Soil Research bulletin, october. Department of Soil Science. Macdonald College of McGill University, Montréal* .

Wheeler, S., Sharma, R. and Buisson, M. (2003), 'Coupling of hydraulic hysteresis and stress–strain behaviour in unsaturated soils', *Géotechnique* **53**(1), 41–54.

Wösten, J., Pachepsky, Y. A. and Rawls, W. (2001), 'Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics', *Journal of hydrology* **251**(3-4), 123–150.

Woyshner, M. R. and Yanful, E. K. (1995), 'Modelling and field measurements of water percolation through an experimental soil cover on mine tailings', *Canadian Geotechnical Journal* **32**(4), 601–609.

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

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

Yeh, T.-C. J. (1989), 'One-dimensional steady state infiltration in heterogeneous soils', *Water Resources Research* **25**(10), 2149–2158.

Yilmaz, I. (2004), 'Relationships between liquid limit, cation exchange capacity, and swelling potentials of clayey soils'.

Yuan, J., Papaioannou, I. and Straub, D. (2019), 'Probabilistic failure analysis of infinite slopes under random rainfall processes and spatially variable soil', *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* **13**(1), 20–33.

Yukselen, Y. and Kaya, A. (2006), 'Prediction of cation exchange capacity from soil index properties', *Clay Minerals* **41**(4), 827–837.

Zhai, Q. and Rahardjo, H. (2013), 'Quantification of uncertainties in soil–water characteristic curve associated with fitting parameters', *Engineering Geology* **163**, 144–152.

Zhang, D. (2001), *Stochastic methods for flow in porous media: coping with uncertainties*, Elsevier.

Zhang, L., Li, D.-Q., Tang, X.-S., Cao, Z.-J. and Phoon, K.-K. (2018), 'Bayesian model comparison and characterization of bivariate distribution for shear strength parameters of soil', *Computers and Geotechnics* **95**, 110–118.

Zhang, L. and Singh, V. P. (2007), 'Trivariate flood frequency analysis using the gumbel–hougaard copula', *Journal of Hydrologic Engineering* **12**(4), 431–439.

Zhang, L. and Singh, V. P. (2014), 'Trivariate flood frequency analysis using discharge time series with possible different lengths: Cuyahoga river case study', *Journal of Hydrologic Engineering* **19**(10), 05014012.

Zhang, L., Zhang, J., Zhang, L. and Tang, W. (2011), 'Stability analysis of rainfall-induced slope failure: a review', *Proceedings of the Institution of Civil Engineers-Geotechnical Engineering* **164**(5), 299–316.

Zhang, L., Zhang, L. and Tang, W. (2003), 'Importance of considering correlations among parameters of soil-water characteristic curve', *Applications of statistics and probability in civil engineering* .

Zhang, Y., Qu, Y., Liu, G. and Wu, S. (2003), 'Engineering geological properties of miocene hard clays along the middle line of the north–south diversion water project in china', *Bulletin of Engineering Geology and the Environment* **62**(3), 213–219.

Zhao, J., Chen, L., Collin, F., Liu, Y. and Wang, J. (2016), 'Numerical modeling of coupled thermal-hydro-mechanical behavior of gmz bentonite in the china-mock-up test', *Engineering geology* **214**, 116–126.

Zhao, T. and Wang, Y. (2018), 'Simulation of cross-correlated random field samples from sparse measurements using bayesian compressive sensing', *Mechanical Systems and Signal Processing* **112**, 384–400.

Zhou, A., Sheng, D., Carter, J. P. et al. (2012), 'Modelling the effect of initial density on soil-water characteristic curves', *Géotechnique* **62**(8), 669–680.

Zhu, D., Griffiths, D. V., Huang, J. and Fenton, G. A. (2017), 'Probabilistic stability analyses of undrained slopes with linearly increasing mean strength', *Géotechnique* **67**(8), 733–746.

Zhu, H., Griffiths, D. V., Fenton, G. A. and Zhang, L. (2015), 'Undrained failure mechanisms of slopes in random soil', *Engineering Geology* **191**, 31–35.

Zhu, H., Zhang, L., Xiao, T. and Li, X. (2017), 'Generation of multivariate cross-correlated geotechnical random fields', *Computers and Geotechnics* **86**, 95–107.

Zhu, Z., Sun, D., Zhou, A. and Qiu, Z. (2016), 'Calibration of two filter papers at different temperatures and its application to gmz bentonite', *Environmental Earth Sciences* **75**(6), 509.