Implementing statistical fitting and reliability analysis for geotechnical engineering problems in R

Xing Zheng Wu*

Department of Applied Mathematics, School of Applied Science, University of Science and Technology Beijing, 30 Xueyuan Road, Haidan District, Beijing, P. R. China. Post code: 100083.

Abstract

Reliability analysis and multivariate statistical fitting are valuable techniques that enhance the scientific basis of regulatory decisions in geotechnical problems. The R environment is an emerging programming platform largely used for statistical computing in many areas. This study introduces the use of several new R packages specifically developed to assist risk assessors in their geotechnical projects. Firstly, the fitting of parameterised models to the average over a number of observed samples and/or to characterise the dependence structures among variables is presented. The usefulness of R packages is illustrated through a marginal and copula fitting to a data set. The non-parametric kernel density estimation and confidence limits in a joint model of associated multiple variables are also facilitated. Secondly, the most popular reliability analysis methods, such as the first- and second-order reliability methods (FORM and SORM) and the random sampling simulation method, are implemented in R. The efficiency of implementing these classical approximation methods is discussed, and two example problems are demonstrated, namely the bearing capacity of a rigid pile with two variables and the stability of an infinite slope with multiple variables. It is hoped that this demonstration can promote the use of advanced R tools for interpretation of the uncertainty in the decision-making process of geotechnical problems.

Keywords: soil; probabilistic; dependence; fitting; multivariate; R

This manuscript has been published by Georisk, please cite it via http://www.tandfonline.com/doi/full/10.1080/17499518.2016.1201577

In comparison with the released edition, the current manuscript provides the source R codes for the illustrative examples. All codes are performed successfully in R 2.15.0.

* Corresponding author. Tel.: +86 18611869118; Fax: +86 10 82610539.
E-mail address: xingzhengwu@gmail.com

1 Introduction

The application of probabilistic analysis in engineering design and management offers many advantages not adequately addressed by the traditional deterministic calculation methods. The reliability assessment may require a mathematical combination of marginal distributions that reflect the variability across the random variables involved, especially for the geomaterials. The natural variability of soil or rock properties is usually measured in one identical test, which results in multi-dimensional dependent variables (Ching and Phoon, 2013). Thus, the probabilistic-based multivariate fitting approach is a natural choice for this type of multivariate analysis to quantify the dependence characteristics. Moreover, some investigations show how the dependence between the soil parameters strongly influences the probability of failure of geostructures (Hamm et al., 2006). With a probabilistic description of the multiple dependent variables in hand, many simple approximation approaches are suitable to determine the probability of failure, or the reliability index of geostructures, for instance: the first-order reliability method (FORM), second-order reliability method (SORM), advanced second-moment techniques, point-estimate calculations, or Monte Carlo simulations (Ang and Tang, 1984).

Some of these algorithms that explore the joint behaviour of the basic random variables and calculate the reliability index of geostructures are well-documented in the textbooks (Ang and Tang, 1984) and integrated into the commercial software packages (Liu et al., 1989; Phoon and Honjo, 2005; Low and Tang, 2007). Alternatively, these algorithms can be implemented by the open source development model adopted by R platform (R development core team, 2013), which permits local customisation where necessary, and reduces cost of acquisition (Bivand, 2000; Grunsky 2002; Pebesma et al., 2012). R is an integrated suite of software facilities for data analysis and graphical display, which is extended by a large collection of packages in which up-to-date statistical methods are implemented. R functions, which are freely available under the Free Software Foundation's GNU General Public License, can be executed effectively on a wide range of operating systems, including the Windows, Unix, and Mac OS.

In the practice of geotechnical engineering or geosciences, several researchers have demonstrated the abilities of the R programming language (Grunsky, 2002; Graf et al., 2009; Honjo et al., 2010; Pohl et al., 2012; Wu, 2013a, 2013b, 2013c; 2013d; 2015a; 2015b; 2015c), and their studies cover specific aspects of both the deterministic and probabilistic applications. These applications are still in their nascent stages and the full potential of R for analysing geotechnical problems has yet to be realised. Recently, a number of packages released on the Comprehensive R Archive Network (CRAN, *http://cran.r-project.org*), which can be very useful in implementing the probabilistic analysis of geostructures in all areas. The topics can include the marginal distribution,

correlation, joint distribution, ternary plot concerning the kernel density estimation (KDE), regression analysis, and solving procedures of the classical reliability analysis. This study introduces several R packages through illustrative examples from the geotechnical domain to implement the multivariate fitting and probabilistic analyses.

2 Overview of probabilistic procedures

Soil or rock is a natural material and therefore its properties are influenced by the conditions of formation (including a combination of various geological, environmental, and physico-chemical processes). Laboratory and field investigations of site soil conditions suggest that the properties (such as density, orientation, void ratio, and shear strengths) of soil materials are highly variable and dependent each other (Lumb, 1970; Phoon and Kulhawy, 1999). Knowing the material properties, failure modes, and loading conditions, the performances of geostructures (such as embankments, foundations, tunnels, shafts, and slopes) are often represented using the reliability index or probability of failure derived from the structural reliability analysis.

Fig. 1 summarises the methodology of a reliability analysis used to calculate the reliability index through considering the uncertainties of material properties, which contains two modules (joint fitting and reliability analysis).

In the first module, when a parameter is treated as random variables, the values can be characterised by the best-fit probability distribution. The appropriate probability distributions for each parameter should be established. Subsequently, the dependence characteristics between variables are evaluated. A proper joint distribution can then be determined if a probability distribution of each parameter and their dependence are specified. Some other activities such as classification and regression can be involved in some occasions.

In the second module, a performance or limit state function of geostructures should be defined for an application problem. Afterwards, the probability of failure or reliability index is solved using the first- or second-order reliability methods (FORM or SORM) or the random sampling method.

These modules are connected and discussed in the subsequent sections of this paper. The methodologies outlined in Fig. 1 are coded into an R-based program as is allowed the use of R's built-in functions and graphing capabilities. The implementation of these techniques in R is outlined below, and the source data and subroutines related to this study are compiled and accessible in the package of '*GeoRiskR*' (Wu, 2013a).

3 Statistical fitting of geological data

3.1 Statistical fitting

3.1.1 KDE of soil classification

Soil scientists use textural triangle to express soil classes and categorise soil samples (Brady and Weil, 1996). Soil texture classes are determined by soils' clay, silt, and sand fractions. The general properties usually cover a wide range in grading and texture (Lumb, 1970). Soil classification for a geotechnical probabilistic analysis is an important, but often ignored, topic of interest. The soil texture triangle of classifying soils shows the uncertainty from the soil classification standpoint when many samples are to be collected and analysed individually. Alternatively, a KDE can provide a nonparametric estimate of the probability density function of the texture properties for a soil class. The central tendency can be indicated by the highest level of the density contours.

3.1.2 Marginal distributions

The wide range of soil properties, such as density, moisture content, shear strength, and compressibility are commonly reported. A univariate analysis associated with fitting a marginal frequency function to such scattered experimental data has been extensively studied (Matsuo and Kuroda, 1974; Whitman, 1984; Ang and Tang, 2007). The normal or Gaussian distribution is the most common one, and its probability density function is the bell-shaped curve. Another common form is the lognormal distribution, which is defined for non-negative values. Other distributions such as the beta, gamma, Gumbel, and Weibull are also encountered (Wu, 2015c). Further details of these marginal distributions can be found elsewhere (Ang and Tang, 2007; Montgomery and Runger, 1999).

3.1.3 Dependencies and fitting joint distributions

When two (or more) variables are involved, they may indicate the joint variability. For instance, some concurrently interpreted variables of soils, such as cohesion and friction angle via a two-parameter linear fitting of the failure envelope, can be physically dependent (Lumb, 1970; Matsuo and Kuroda, 1974). Some measured properties of soils are also known to be correlated. For example, the undrained shear strength is commonly related to the plasticity index (PI) and the cone tip resistance is associated with the overconsolidation ratio (OCR) of soils (Kulhawy and Mayne, 1990). As discussed by researchers (Lumb, 1970; Whitman, 1984; Ching and Phoon, 2013; Wu 2013c), identifying relationships among the variables is likely to be important because the impact of events that are simultaneously extreme may be much greater than if extreme events of either component occur in isolation. A scattergram of the paired observed data points usually illustrates

126 the relationship between two variables. The Pearson's correlation coefficient is usually measured to
127 quantify the strength of this association.
128 However, the classical Pearson's correlation coefficient is not sufficient to characterise a fully
129 dependent structure among variables and is suffering from some limitations. For instance, the
130 correlation between two non-normal random variables is not the same as the correlation between the
131 equivalent normal random variables (Phoon and Nadim, 2004). Indeed, such a measure is
132 convenient only for elliptical distributions and then becomes biased in the asymmetric case
133 (Gatfaoui, 2005). To overcome this problem, two concordance measures—Spearman's Rho and
134 Kendall's Tau—should be used to characterise the dependence between variables whose joint
135 distribution is non-elliptical. The concordance measures obtained by the 'kendall' and 'spearman'
136 methods are widely used in the copula models (Wu, 2013b; 2103c).
137 A simple computational procedure built by copulas is useful in understanding and modelling
138 dependent structures for random variables. Sklar (1959) first introduced the copula to model
139 dependence structure, which provides a method to link together one-dimensional marginal
140 distribution functions to form multivariate distribution functions. Since then, numerous copula
141 functions have been proposed, as summarised by Nelsen (2006). A comprehensive theoretical
142 description is beyond the scope of this study, but the interested reader is recommended to refer to
143 the appropriate literature (Joe, 1997; Nelsen, 2006; Genest et al., 2007; Yan, 2007).

144 **3.1.4 Confidence limits of a regression**

145 Although regression and correlation are not the same, correlation concepts serve as some (but
146 not all) of regression's building blocks. Huck et al. (1974) presented an elegant discussion of this
147 topic. When each possible pair of measures is forced into the regression equation, a regression line
148 can be used to describe one parameter with respect to the other. In addition to linear regression,
149 other types of regression lines (power, logarithmic, exponential, and polynomial) can be used to
150 predict the value of one parameter as a function of another.
151 However, the regression coefficients are not very meaningful because they do not adequately
152 describe the scatter (or variability) of the bivariate data. A measure of the degree of correspondence
153 within the developing relationship is provided by the correlation analysis. The selection between
154 these two techniques—correlation and regression—depends on the goals of the research. When a
155 relation between two variables is concerned, a regression analysis should be conducted. When the
156 scattering or variabilities of two variables are considered, a correlation analysis should be of great
157 interest.

3.2 Implementation of statistical fitting in R

(1) The package of '*soiltexture*' (Moeys and Wei, 2011) can be loaded to plot out a ternary graph with the package of '*plottrix*' (Lemon et al., 2012). Seeking more details on an extended discussion of the methodology, the readers can refer to Moeys and Wei (2011). The function of *kde2d( )* in the '*MASS*' package by Venables and Ripley (2012) is utilised to draw a contour plot of the KDE.

(2) Fitting a marginal density distribution to data consists of selecting the best-fit probability density distribution from a specified family of distributions. It requires distribution selection, parameter estimation, and a quality-of-fitness evaluation. The package '*fitdistrplus*' (Delignette-Muller et al., 2010) provides a number of functions dedicated to assisting the implementation of fitting a univariate parametric distribution to various types of data. A comprehensive range of basic statistical distributions to fit the data set includes the normal, lognormal, student t, gamma, Gumbel, and Weibull distributions. It also provides a range of non-parametric tests that are appropriate when the data cannot follow any particular distribution. For a given distribution, a function of *fitdist( )* can be used to estimate the parameters using the maximum likelihood approach. The quality of the fit is assessed using classical goodness-of-fit statistics (Chi-squared, Kolmogorov-Smirnov, and Anderson-Darling statistic) and graphs (empirical and theoretical distribution plot in density, i.e., P-P plot, and in cumulative density function, i.e., Q-Q plot). The reader seeking more information about these concepts is recommended to consult Montgomery and Runger (1999) and Wu (2013c). For instance, the best-fit criteria for marginal distributions can be quantified by the magnitude of log-likelihood value and the Akaike information criterion (AIC; Akaike, 1974) value. The AIC value is defined as minus the log-likelihood of the model plus the number of parameters being estimated. The higher the log-likelihood value and the smaller the AIC value, the better the fit is.

(3) Apart from the marginal distribution of each variable, the mutual dependence between variables also plays an important role in characterising their joint behaviours (Lumb, 1970; Low, 2005). A function *kdepairs( )* in the package of '*ResourceSelection*' (Subhash et al., 2013) is helpful to identify the key factors for each variable.

The correlation coefficients can be calculated by the R function of *cor( )* with options of 'pearson', 'spearman', or 'kendall'.

An R package of multivariate dependence with copulas—namely '*copula*'—was developed by Yan (2007) and has gained general applications in the wider area beyond the domain of the reliability analysis (Gartsman et al. 2009; Li et al., 2010; Mair et al., 2012; Wu, 2013b; 2013c; 2015a). In this package, the *mvdc( )* function creates a multivariate distribution object, which has three major inputs: a copula name, lists of the specified marginal distributions, and lists of the dependence parameters. When the maximum likelihood method is used, the general practice is to fit

the data with all the candidate copulas (including the normal, student t, Clayton, Frank, Gumbel, and Plackett) and choose the ones with the highest likelihood or the lowest AIC. For the definition of these copulas, the reader is recommended to consult the literature (Nelsen, 2006; Wu, 2013b).

(4) Linear, non-linear, and logistic regressions are easy to implement in R, and many regression techniques can be used for this purpose, such as the functions of *lm( )* for linear and *nls( )* for polynomial. R provides the standard diagnostics that indicate the quality of a regression fit for the scattering data.

4 Reliability analyses of geotechnical problems

4.1 Theoretical background of reliability analyses

Since 1970s, reliability methods have been applied to stability analyses of an engineering system as a means to incorporate and evaluate the impact of the uncertainties of input parameters. Multiple variables can be involved in the reliability analysis of an engineering system. For a single failure mode, the performance function g in terms of the input random variables $\boldsymbol{Z}$, $g(\boldsymbol{Z})$ is formulated

$$g(\boldsymbol{Z}) = g(Z_1,..,Z_n) \tag{1}$$

where $\boldsymbol{Z} = (Z_1,...,Z_n)$ is a vector of variables of the system, and the function $g(\boldsymbol{Z})$ determines the performance of the system. When $g(\boldsymbol{Z}) < 0$, it denotes the failure state. Thus, the limit state function, $g(\boldsymbol{Z}) = 0$, is the dividing boundary between safe domain and failure domain. The probability of structural failure p_{f} can be calculated as

$$p_{\mathrm{f}} = \int ... \int_{g(\boldsymbol{Z})<0} f_{\boldsymbol{Z}}(z_1,...,z_n)\mathrm{d}z_1...\mathrm{d}z_n \tag{2}$$

where $f_{\boldsymbol{Z}}(z)$ is the joint probability density function of $\boldsymbol{Z}$. The integration should be performed over the failure domain $g(\boldsymbol{Z}) < 0$; however, this is often extremely complicated to estimate (Ang and Tang, 1984).

A variety of approximation approaches have been developed for the probabilistic stability analysis, including the analytical approximation method and the simulation method. Compared with the analytical approximation, the simulation method can be still much less efficient, especially when the failure probability is low (Grooteman, 2008; Wu, 2013b). These simulated samples can be generated using the conventional Monte Carlo sampling technique (Tobutt, 1982) or copula-based sampling method (CBSM; Wu, 2013b). The first- and second-order reliability methods (FORM/SORM) have emerged as one of the most effective approximation methods. For instance, the FORM has been shown to be an efficient tool and widely accepted for the reliability analysis (Ang and Tang, 1984, page 340-360). This method is an analytical first-order approximation in

which the reliability index is interpreted as the minimum distance from the origin to the limit state surface in standardised normal space and the design point (most likely failure point) is solved using mathematical programming methods. Although an exact solution for linear or weakly nonlinear performance functions (or limit state functions) can be achieved, such an approximation depends on the degree of the nonlinearity in the performance function largely. For a heavily nonlinear performance function, a second-order approximation of the failure surface at the design point, known as the SORM, is recommended. The theoretical background and solving procedure of a few popular techniques are outlined below.

4.2 Selected methods

4.2.1 FORM using Ang and Tang's algorithm

The first-order algorithm, which uses the performance function and its gradient, developed by Hasofer and Lind (1974) originally and modified by Rackwitz and Fiessler (HLRF; 1978), is commonly used in structural reliability analysis due to its simplicity. As discussed by Ditlevsen (1981) and Bagheripour et al. (2012), a general matrix form of the reliability index β can be defined as:

$$\beta = \min_{\boldsymbol{Z}\in\Omega} \sqrt{\left(\boldsymbol{Z}-\overline{\boldsymbol{Z}}\right)^T \boldsymbol{C}_z \left(\boldsymbol{Z}-\overline{\boldsymbol{Z}}\right)} \tag{3}$$

where $\boldsymbol{Z}$ represents all variables in a vector form, $\overline{\boldsymbol{Z}}$ is the mean values of $\boldsymbol{Z}$, $\boldsymbol{C}_z$ is the covariance matrix, and T denotes the transpose of a matrix. The failure domain Ω is defined by $g(\boldsymbol{Z})<0$. This definition implies that β may be solved with a constrained optimisation equation. Ang and Tang (1984) gave a detailed description in their textbook about the algorithms for all combined cases, between the correlated or uncorrelated variables, the normal or non-normal variables, and the linear or non-linear limit state functions. The principal steps are as follows:

[1] Assume the initial value Z_i^0 of the variable Z_i, then obtain $Z_i^{u0} = \dfrac{Z_i^0 - \mu_i}{\sigma_i}$ where μ_i is mean and σ_i is standard deviation of Z_i^0;

[2] Evaluate $\dfrac{\partial g}{\partial Z_i^u}$, then $\alpha_i^0 = \dfrac{\dfrac{\partial g}{\partial Z_i^u}}{\sqrt{\sum_{i=1}^{n}\left(\dfrac{\partial g}{\partial Z_i^u}\right)^2}}$;

[3] Recast $Z_i^0 = \mu_i - \alpha_i^0 \sigma_i \beta^0$;

[4] Substitute Z_i^0 to the limit state equation $g\left(Z_1^0,...,Z_i^0,...,Z_n^0\right)=0$ and solve for β^1;

[5] Re-evaluate the most probable points $Z_i^1 = -\alpha_i^0 \beta^1$;

[6] Loop steps 2 to 5 until convergence ($\Delta\beta = \beta^0 - \beta^1 \le \varepsilon$) is achieved.

4.2.2 FORM using Low and Tang's algorithm

Several other popular solutions or variants of the FORM are available. For instance, the approach developed by Low and Tang (2007) associated ellipsoidal perspective through varying dimensionless numbers $\boldsymbol{\Psi}$ during constrained optimisation is simple and intuitive, because it works in the original space of the variables and it does not involve the orthogonal transformation of the correlation matrix $\boldsymbol{\rho}$. It uses the following equation for the reliability index

$$\beta = \min_{Z \in \Omega} \sqrt{\boldsymbol{\Psi}^T \boldsymbol{\rho} \boldsymbol{\psi}} \tag{4}$$

To solve this expression, an iterative optimisation process can be performed using software packages, as implemented in the Spreadsheet (Low and Tang, 1997; Low and Tang, 2007) or Matlab (Huang and Griffiths, 2011). The recipe for the numerical solution can be stated as:

[1] Declare the dimensionless numbers function of $\psi_i = \dfrac{Z_i - \mu_i}{\sigma_i} = \Phi^{-1}[F(Z_i)]$;

[2] Determine the nonlinear equality, i.e., performance function;

[3] Solve the objective function in Eq. (4) and achieve β.

4.2.3 SORM

Various methods have been suggested to improve the accuracy of FORM calculations. Breitung (1984) derived the failure probability using an asymptotic approximation associated with β obtained by the FORM, which is written as:

$$p_f = \Phi(-\beta) \prod_{j=1}^{n-1} \left(1 + \beta \kappa_j\right)^{-1/2} \tag{5}$$

where n is the number of the random variables and κ_j (j =1 to $n-1$) are the principal curvatures of the failure surface at the design point. A more accurate three-term formula was proposed by Tvedt (1983; 1990) in which the last two terms are interpreted as correctors to Breitung's formula

$$p_f = \Phi(-\beta) \prod_{j=1}^{n-1} \left(1 + \beta \kappa_j\right)^{-1/2} + T_2 + T_3 \tag{6}$$

where $T_2 = C_1 \left\{ C_2 - \prod_{j=1}^{n-1} \left[1 + (\beta + 1)\kappa_j\right]^{-0.5} \right\}$, $T_3 = (\beta + 1) C_1 \left\{ C_2 - \mathrm{Re} \prod_{j=1}^{n-1} \left[1 + \left(\beta + \sqrt{-1}\right)\kappa_j\right]^{-0.5} \right\}$, $C_1 = \beta \Phi(-\beta) - \phi(-\beta)$, and $C_2 = \prod_{j=1}^{n-1} \left(1 + \beta \kappa_j\right)^{-0.5}$. Herein, $\phi(\,)$ and $\Phi(\,)$ are the probability density function and cumulative distribution function of the standard normal variable, respectively. Re[ ] represents the real part of a complex argument.

These SORM formulations require the principal curvatures $\boldsymbol{\kappa}$ of the limit state surface at the design point to be solved, as demonstrated in a spreadsheet-based platform by Chan and Low

(2012). The principal curvatures are determined by the eigenvalues of the (normalised and reduced) Hessian matrix $\boldsymbol{A}$ defined as (Breitung, 1984):

$$\boldsymbol{A} = \frac{\boldsymbol{H}_{reduce}}{\|\nabla g\|} \tag{7}$$

where $\boldsymbol{H}_{reduce}$ (arranged as a $(n-1)\times(n-1)$ matrix) is a reduced Hessian matrix $\boldsymbol{H}$ computed at the design point in the transformed reference system, and $\|\nabla g\|$ is the Euclidean norm of the gradient of the limit state function at the design point in the standard normal space. The Hessian matrix $\boldsymbol{H}$, i.e., the second derivative matrix of the performance function at the design point z^*, is defined as:

$$\boldsymbol{H} = \nabla g^2_{z^*} = \begin{bmatrix} \frac{\partial^2 g(z)}{\partial z_1 \partial z_1} & \cdots & \frac{\partial^2 g(z)}{\partial z_1 \partial z_n} \\ \cdots & & \\ \frac{\partial^2 g(z)}{\partial z_n \partial z_1} & \cdots & \frac{\partial^2 g(z)}{\partial z_n \partial z_n} \end{bmatrix} \tag{8}$$

These formulas can be estimated using a central finite difference scheme (Phoon, 2008) to achieve an approximation of the failure probability p_{f} in Eqs. (5) and (6). The essential solution steps are provided as follows:

[1] Construct gradient vector $\nabla g^2_{z^*}$ and Hessian matrix $\boldsymbol{H}$ by a central difference;

[2] Determine the principal curvatures by the eigenvalues of the reduced Hessian matrix $\boldsymbol{A}$ using the QR decomposition algorithm;

[3] Compute the second-order reliability index using Breitung's and Tvedt's formulas.

4.2.4 CBSM

The copula-based simulation uses a random number generator associated with any copula to create a large set of values for the uncertain parameters, considering their prescribed probability density distributions and the dependence structures. Subsequently, the probability of failure p_{f} of geostructures can be estimated by counting the number of samples within the failure region (i.e., the factor of safety F_{s} of a geostructure is less than one) and dividing it by the total number of copula-based samples, as detailed in Wu (2013b). The reliability index β can be estimated using $\beta = \frac{\mu_{\mathrm{F}} - 1}{\sigma_{\mathrm{F}}}$ when the factor of safety F_{s} follows the normal distribution with mean μ_{F} and standard deviation σ_{F}; alternatively, $\beta_{\ln} = \frac{\ln\left(\frac{\mu_{\mathrm{F}}}{\sqrt{1+C_{\mathrm{v}}^2}}\right)}{\sqrt{\ln\left(1+C_{\mathrm{v}}^2\right)}}$ when F_{s} follows the lognormal distribution (Ang and Tang, 1984), where C_{v} represents the coefficient of variation of F_{s}, i.e., $C_{\mathrm{v}} = \frac{\sigma_{\mathrm{F}}}{\mu_{\mathrm{F}}}$.

The following computational procedure can be summarised:

[1] Generate the number of realisations n using the copula-based sampling techniques, as given in Wu (2013b; 2013d);

[2] Evaluate the value of F_s and count the number of failed realisations n_f within the failure domain;

[3] Estimate the probability of failure through dividing n_f by the total number of n, and calculate β or β_{ln} in terms of the statistical characteristics of F_s.

4.3 Implementation of reliability analyses in R

The efficient solution algorithms of these reliability methods connected to R packages are presented as follows:

(1) In the solving procedure of the FORM proposed by Ang and Tang (1984), the derivative of the limit state expression of Eq. (3) with symbolic parameters can be facilitated by a differentiation operation package '*mosaic*' (Pruim et al., 2012), and a nonlinear root finding algorithm for the updated reliability index can be implemented by '*rootSolve*' package (Soetaert, 2009).

(2) For the FORM suggested by Low and Tang (2007), the nonlinear optimisation problem in Eq. (4) can be solved numerically using a sequential equality constrained quadratic programming method implemented in the package '*Rdonlp2*' (Tamura, 2007) for the source code '*DONLP2*' (Spellucci, 1998) or the package '*Rsolnp*' (Ghalanos and Theussl, 2012).

(3) A large number of statistical functions are integrated into R, such as *eigen( )* to solve eigenvalues of the matrix in Eq. (7) and *qr( )* to perform QR factorisation of a matrix. Thus, no special package is required to implement the solving procedures of the SORM, except for the calculation of the vector norm in Eq. (7) using the package '*pracma*' (Borchers, 2013).

(4) In the CBSM, the copula-based samples can be implemented using the *rmvdc( )* function in the package '*copula*' (Yan, 2007). This function requires a multivariate distribution by specifying the copula class and marginal distributions, and then, it generates random samples from this joint distribution. The performance of geostructures can be evaluated using these correlated multivariate samples (Wu, 2013b).

5 Geotechnical problems

Two illustrative examples of geostructures are examined to consider the variability of the forming materials, and the uncertainties surrounding input load conditions. The performance of the reliability methods is compared.

5.1 Bearing capacity associated with two variables

5.1.1 Statement of problem

Pile foundations are usually subjected to considerable lateral loads or forces, due to sea waves, action of winds, or earthquakes. Examples of such piles include the foundation of support walls, bridges, and offshore structures. A laterally loaded free-head rigid pile in sand, as illustrated in Fig. 2, is considered as an example. Letting H_{ca} denotes the ultimate bearing capacity and F_L represents the applied forces, the failure criterion is thus that the loading F_L (demand) exceeds the ultimate bearing capacity H_{ca} (capacity). The performance function of the ultimate lateral capacity of the pile is written as (Broms, 1964; Ma and Deng, 2000; Phoon and Honjo, 2005):

$$g(\gamma,\phi) = H_{ca} - F_L = \gamma \frac{BD^3}{2(e+D)} \tan^2(45+\phi/2) - F_L \tag{9}$$

where B is the pile diameter, D is the embedded length of the pile, e is the length of free head, and F_L is the applied load. Letting $\tan^2(45+\phi/2)$ be denoted by K_p, it is usually called the coefficient of passive earth pressure. The definition of these parameters is illustrated in Fig. 2. One can consider all the factors influencing the performance function as random variables in this probabilistic analysis. However, such an analysis can be computationally intensive for some numerical approaches. It is, therefore, appropriate to treat some of the variables as deterministic parameters. Thus, the unit weight (or bulk density) γ and internal friction angle ϕ (also referred to as angle of internal friction) of sand are taken as random variables. The rest of the parameters are assumed to be deterministic and given by: B =1 m, D =10 m, e =1 m, and F_L =1000 kN (Phoon and Honjo, 2005).

5.1.2 Data set

The joint statistical characteristics of sand properties between density and friction angle have been reported in the literature (Matsuo and Kuroda, 1974; Hammond and Hardcastle, 1992; Parker et al., 2008). A positive correlation between γ and ϕ is commonly observed. The observed data (16 samples), as described in Hammond and Hardcastle (1992), are used. The Pearson's correlation coefficient ρ_p is calculated as 0.73 and the Kendall's τ 0.46. The values of mean and standard deviation of γ_d are 16.35 kN/m^3 and 0.94 kN/m^3, respectively; the values of mean and standard deviation of ϕ are 39.81° and 2.45°, respectively. Assuming the moisture content w is 25%, the ambient unit weight γ is then calculated as $\gamma = (1+0.25)\gamma_d$.

For the marginal distribution, the inherent variability in the observed or measured data is plotted graphically in the form of a histogram, as shown in Fig. 3. The non-parametric kernel

density curve is overlapped to model the frequency distribution without assuming any curve shape in advance. The kernel density curves of γ_d and ϕ are shown in Fig. 3. The Gumbel distribution fits the data of γ_d well, and the value of AIC is 45.97. A slightly higher value of 46.44 is obtained with the normal distribution. The Gumbel distribution for ϕ may also be preferred due to lower AIC value 71.69 for this distribution than the value of 77.07 for the normal distribution. The source code of the implementation in R to identify the best-fit distribution is provided in Appendix A1.1.

For these paired data, the normal copula is chosen in this example because it fits the studied data well according to the AIC values. Thus, one density contour of this joint distribution related to the best-fit marginal distributions is shown in Fig. 3. The density contour associated with the normal marginal distributions is also shown, which is an oriented ellipse centered at the mean values. Its size reflects the standard deviations of two variables and whose shape (eccentricity) reflects their correlation. The details are left to the readers who may have to run the R code given in Appendix A1.2.

To go beyond this, an attempt is made to simulate other sets of soil properties, that can be constructed using a function of *rmvdc( )* by repeating its application a number of times from the fitted copula. Fig. 3 illustrates 1000 simulated data points along with their original observed 16 data points. The probability density contour (PDC; Wu, 2013c) at 95% confidence level for both simulation and observation is superimposed on this graph, which shows intuitive graphical evidence that the simulation provides a good job of mimicking the observations. The generated random simulations are an important source for further scientific studies, such as reliability analyses. The R code to generate the simulations is given in Appendix A1.2.

The results of both simple linear and polynomial regression using the least squares method are shown in Fig. 4 for the data set obtained by Hammond and Hardcastle (1992). The R code to generate this plot is given in Appendix A1.3. Taking into account the uncertainty inherent in parameters estimated from limited samples, the regressions are shown within the 95% confidence intervals from the mean values.

5.1.3 Solution

The factor of safety against bearing capacity is evaluated by the deterministic analysis considering solely the mean values of the variables involved, i.e., $F_s = \overline{H}_{ca} / F_L$, given 4.24. In the probabilistic analysis of this example, the normal marginal distribution is assumed for brevity, and the interested readers may substitute the identified best-fit marginal distributions to the reliability analysis. The specific source codes of the FORM implemented in R (including Ang and Tang's algorithm and Low and Tang's algorithm) are provided in Appendix A2.1 and A2.2, respectively,

and the latter is more concise and readable. The R code of the SORM is given in Appendix A2.3 and the one of the CBSM is presented in Appendix A2.4.

All computing is performed in a few seconds on an Intel Xeon 2.53 GHz Workstation with 4 GB of memory. The first-order reliability index is found to be 8.822 using Ang and Tang's method. This can be confirmed with 8.744 by the analyses using the spreadsheet platform (as given in Low and Tang, 2007). The reliability indexes calculated from the SORM are 8.855 and 8.856, based on the Breitung's formula and Tvedt's formula, respectively. In this particular case, the second-order approximation for the reliability analysis does not provide sufficiently different results since the failure surface does not possess large curvatures. The resulting reliability index is 9.066 using the CBSM.

The values of these computed reliability indexes are pretty large because the coefficients of variance of γ and ϕ are too low, compared with the existing experimental data (Phoon and Kulhawy, 1999). As illustrated in Fig. 5, the reliability index would be decreased significantly when the standard deviations of of γ and ϕ are increased to three times of their original values, resulting in an expanded one-standard-deviation ellipse as explained in Low and Tang (2007).

5.2 Infinite slope associated with multiple variables

5.2.1 Statement of problem

Shallow landslide failure is a common seasonal occurrence on natural slopes, especially subjected to potential earthquake and/or rainfall (Wu, 2015b). To consider the parameter uncertainties in mechanical or physical properties of landfill materials, the probabilistic slope stability analysis is generally used for small areas at fine scales (Zaitchik et al., 2003) or in soil engineering for slope-specific stability studies (Wu and Sidle, 1995; Wang and Huang, 2012; Wu, 2013b; 2015b). In this type of approach, the popular stability analysis methods, such as the simplified circular arc method of slices and an infinite stability analysis, are used to determine the limit equilibrium state of the slope. When the slope fails at or near the contact between the soil colluvium and impervious bedrock and the slope gradient is constant throughout the length, such a slope may be analysed as an infinite one, as shown in Fig. 6. The performance function of the stability is defined as the ratio of the available resisting force R to the driving force S (Phoon, 2008; Wu, 2015b):

$$g_{\mathrm{s}}(c,\phi,\gamma_{\mathrm{a}},\gamma_{\mathrm{sat}}) = \frac{R}{S} - 1 = \frac{cL + N\tan\phi}{(1-K_{\mathrm{v}})(W_{\mathrm{n}}+W_{\mathrm{sat}})\sin\theta + F_{\mathrm{w}} + K_{\mathrm{h}}(W_{\mathrm{n}}+W_{\mathrm{sat}})\cos\theta} - 1 \qquad (10)$$

where $R = cL + N\tan\phi$, $S = (1\pm K_{\mathrm{v}})(W_{\mathrm{n}}+W_{\mathrm{sat}})\sin\theta + F_{\mathrm{w}} + K_{\mathrm{h}}(W_{\mathrm{n}}+W_{\mathrm{sat}})\cos\theta$, and $N = (1\pm K_{\mathrm{v}})(W_{\mathrm{n}}+W_{\mathrm{sat}})\cos\theta - K_{\mathrm{h}}(W_{\mathrm{n}}+W_{\mathrm{sat}})\sin\theta$. Herein, θ is slope inclination, the weight of a slice (as

depicted in Fig. 6) associated with saturated zone is given by $W_{sat} = \gamma_{sat} V_{sat} = \gamma_{sat} Lh\cos\theta$, the weight of a slice associated with natural zone is $W_n = \gamma_n V_n = \gamma_n L(H-h)\cos\theta$, the seepage force is denoted as $F_w = i_s \gamma_w V_{sat} = \gamma_w Lh\sin\theta\cos\theta$, the seepage gradient is set to $i_s = \sin\theta$, and H is the depth of soil above bedrock. Moreover, the depth of groundwater table above bedrock is given by $h = mH$. Here γ_a and γ_{sat} are moist unit weight and saturated unit weight of the surficial soil, respectively; and γ_w is the unit weight of water (9.81 kN/m^3). The magnitude of the seepage force F_w is associated with the hydraulic gradient and the soil volume, which arises due to the water flow parallel to the slope in the saturated layer (Ghiassian and Ghareh, 2008).

The horizontal seismic coefficient K_h should be specified to calculate the seismic forces in a pseudo-static analysis. As proposed by Noda et al. (1975), $K_h = \frac{1}{3}\sqrt[3]{\frac{a_{max}}{g}}$, where the peak ground acceleration a_{max} can be set to the maximum limit of 2 g, as was performed in the work of Fotopoulou and Pitilakis (2013a; 2013b). Thus, the maximum value of K_h can be calculated as 0.42. The vertical acceleration coefficient K_v is assumed to be equal to 0.5 K_h. The soil saturation index m (Acharya et al., 2006) represents the relative position of the water table in the soil layer, which is a function of the groundwater flow and rainfall intensity. In this study, K_h is set to 0.1 and m is set to 0.2 (Wu, 2015b). Uncertainty in the performance of the slope arises from sources such as variability in soil properties and errors in estimating the dimensions. To simplify the calculation, only four dependent random variables of soil properties (c, ϕ, γ_a, and γ_{sat}) are considered. As claimed by Samadi et al. (2009), these variables can be the dominant source of uncertainty in a slope stability analysis.

5.2.2 Data set

A data set of the soil property parameters, as detailed in Soenksen et al. (2003), is used. To investigate the river bank stability, a large amount of measurements (either in the field or in the laboratory) were performed by these researchers located within eight tributary basins of the Mississippi River in eastern Nebraska, USA. The salt creek tributary is chosen here, containing 38 samples. These samples are gathered from various depths on the bank face and trench for performing the particle size analyses. The observed variability in soil texture (sand, clay, and silt percentage) is shown in Fig. 7 (R code is given in Appendix A1.4). Taken the geometric mean of the KDE as the representative value, the majority of the soils is silty and sandy mixtures. The graph represents the various textural possibilities. The KDE contour shows an estimate of the density of observed data spatially. A similar data set can be found elsewhere (Mousavi et al., 2011; Abbasi et al., 2012).

Soil properties, including the cohesion c, friction angle ϕ, ambient unit weight γ_a, saturated unit weight γ_s, moisture content m, and saturation of degree S, are drawn in a scatter plot as shown in Fig. 8 (R code is provided in Appendix A1.2). These factors are closely related to the soil shear strengths as discussed in Soenksen et al. (2003) and Mousavi et al. (2011), especially for the unit weight, degree of saturation, and moisture content. The graphical representation of their dependences helps us to establish a visual link between them. The joint contours of the KDE between these paired variables are clearly asymmetric in most cases. In Fig. 8, the Pearson's correlation coefficient is shown along with a non-linear regression curve. For instance, a highly significant positive correlation between ambient unit weight and saturated unit weight is found. The friction angle indicates a significant negative correlation with the moisture content. A negative correlation between friction angle and cohesion is observed, which has been gaining acceptance by most of the investigators (Lumb, 1970; Matsudo and Kurodo, 1974; Wu, 2013c). The statistical mean and standard deviation, and the best-fit probability distributions of the random variables (c, ϕ, γ_a, and γ_{sat}) are summarised in Table 1.

5.2.3 Solution

The geometry of the slope, with the inclination of θ =15° and depth H =2 m, is chosen as a reference case. The marginal distribution of these random variables is assumed to follow the normal one for simplicity, but the correlation characteristics (shown in Fig. 8) among variables are considered. The reliability indexes obtained using several different computational reliability methods are given as follows: β =2.129 (using the FORM based on Low and Tang's approach), β_s =2.152 (using the SORM based on Breitung's formula), β_s =2.156 (using the SORM based on Tvedt's formula). β_{cb} =2.063 (using the CBSM with the normal copula). Under varying inclination θ (between 15° and 25°) and varying depth H (between 2 m and 8 m), the values of reliability index using various methods are shown in Fig. 9, and similar geometric properties of such a slope are exemplified by Phoon (2008). Obviously, the decreasing reliability index β is achieved as increasing the inclination angle, as well as increasing the depth. The deterministic solutions of F_s ($= \overline{R} / \overline{S}$) are presented in this figure if the mean values of random variables are imposed. The effect of the geometric parameters on the values of β and F_s holds a similar trend.

It should be noted that a high calculation speed (within a few seconds) using these models is achieved even with larger dimensions in a limit state equation. As pointed out by Bivand and Gebhardt (2000), R is based on modern programming concepts and permits an integration of program scripts with compiled dynamically loaded libraries (written in Fortran or C) when

computing speed is important. The source R code for implementation of this example is available from the author upon request.

6 Discussion

The application of the reliability methods outlined here can be implemented to some other joint probability analyses for geotechnical engineering. Readers are encouraged to solve some existing example applications for practice. Specifically, soil shear strength pairs are involved in numerous studies, such as slope stabilities (Tobutt, 1982; Nguyen and Chowdhury, 1984; Wu, 2013b; 2015b), bearing capacity of foundations (Wu, 2013c), and earth pressure of retaining walls (Low, 2005). A reliability-based deformation analysis of pile foundation (Uzielli and Mayne, 2011; Li et al., 2013; Huffman and Stuedlein, 2014) should also be covered via a bivariate analysis using a two-parameter regression model to characterise the loading-settlement curves. Aside from these cases with explicit performance functions, such R-based reliability approaches can be applied to stand-alone numerical packages via the response surface methods (Chan and Low, 2012) for a stability analysis associated with an implicit performance function. For example, a finite element analysis of a dynamic liquefaction problem for pile foundation, coded in Fortran or C++, is required to incorporate into its probabilistic stability evaluation. These heavily debugged compiled source files can be readily passed to R functions (Schlather and Huwe, 2004). The interface between R and Fortran provides a better integrated programming development environment. In addition, different failure modes of a geotechnical engineering system may depend on the same random variables; then, there exist correlations among the failure modes. The series system reliability of such an engineering system can be solved in the R platform using a copula-based approach (Wu, 2015a).

Prior to an application of statistical techniques for determining the probability distribution of the observation, outlier detection is an important step toward eliminating the unusual points at a threshold level. The '*mvoutlier*' package (Filzmoser and Gschwandtner, 2013) assists with the identification of the outliers. Since some of the analytical techniques in reliability analysis (e.g., the first-order second-moment method) rely on assumptions of normality, a distribution test can be performed with the more recent energy distance-based statistic (Székely and Rizzo, 2005) using the '*energy*' package (Rizzo and Székely, 2014) for both univariate and multivariate variables.

The implementations that have been described for reliability analysis and statistical computing are a small part of what is available in R. Here is not intended to be a comprehensive derivation on the reliability analysis methods and statistical backgrounds, and the focus is on numerical implementation in R for the geotechnical problems. These tools are free, which could help students and resource-limited institutes or countries, while aiding the development of training to perform a reliability-based analysis easily. As stated by Lillis (2011), it is true that the learning curve is longer

than for spreadsheet-based packages; however, once you master the R programming syntax, you can develop your own powerful analytical tools. R combines a powerful programming language, a comprehensive range of statistical functions and excellent graphics. If you are seeking for a statistical environment that includes a programming language, R could be the one. R also provides an opportunity for applying statisticians join in the development and implementation of analysis for geotechnical problems.

7 Conclusions

There are many data analysis and statistical methods available to assess the geotechnical characteristics when there is a notable uncertainty in the engineering geological data. This technical note has demonstrated the implementation of a few widely-accepted reliability methods in R platform. The FORM and SORM are valuable tools for the probabilistic analysis of geotechnical problems, and their approximations are matched with the results using the simulation technique. The underlying joint probability distributions and correlations of multiple variables are also facilitated with the R programming environment successfully. The methods presented here are in a accessible form for different types of reliability problems in geotechnical, coastal, and hydraulic areas.

8 Acknowledgments

The analyses were performed in R (R Core Team 2013) by using the contributed packages mentioned above. The authors and maintainers of these packages are gratefully acknowledged.

9 References

Abbasi, N., Javadi, A.A., Bahramloo, R., 2012. Prediction of compression behaviour of normally consolidated fine-grained soils. World Applied Sciences Journal 18 (1), 6–14.

Acharya, G., De Smedt, F., Long, N.T., 2006. Assessing landslide hazard in GIS: a case study from Rasuwa, Nepal. Bulletin of Engineering Geology and the Environment 65 99–107.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Ang, A.H.S., Tang, W.H., 1984. Probability Concepts in Engineering Planning and Design. Volume II, Decision, risk and reliability. John Wiley, New York, 562pp.

Ang, A.H.S., Tang, W.H., 2007. Probability Concepts in Engineering. Emphasis on

563 Applications to Civil and Environmental Engineering. 2nd Edition, John Wiley and
564 Sons, New York.
565 Bagheripour, M.H., Shooshpasha, I., Afzalirad, M., 2012. A genetic algorithm approach for
566 assessing soil liquefaction potential based on reliability method. Journal of Earth
567 System Science 121 (1), 45–62.
568 Bivand, R., 2000. Using the R statistical data analysis language on GRASS 5.0 GIS data
569 base files. Computers & Geosciences 26 (9/10), 1043–1052.
570 Bivand, R., Gebhardt, A., 2000 Implementing functions for spatial statistical analysis using
571 the R language. Journal of Geographical Systems 2 (3), 307–317.
572 Borchers, H.W., 2013. Pracma: Practical numerical math functions. R package, Version
573 1.5.8.
574 Brady, N.C., Weil, R.R., 1996. The nature and properties of soils. (11th ed.) Prentice Hall,
575 Upper Saddle River, New Jersey.
576 Breitung, K., 1984. Asymptotic approximations for multinormal integrals. Journal of
577 Engineering Mechanics 110 (3), 357–366.
578 Broms, B.B., 1964. Lateral resistance of piles in cohesionless soils. Journal of the Soil
579 Mechanics and Foundations Division 90 (SM3), 123–156.
580 Chan, C.L., Low, B.K., 2012. Practical second-order reliability analysis applied to
581 foundation engineering. International Journal for Numerical and Analytical Methods
582 in Geomechanics 36, 1387–1409.
583 Ching, J., Phoon, K.K., 2013. Multivariate distribution for undrained shear strengths under
584 various test procedures. Canadian Geotechnical Journal 50, 907–923.
585 Delignette-Muller, M.L., Pouillot, R., Dennis, J.B., Dutang, C., 2010. fitdistrplus: Help to fit
586 of a parametric distribution to non-censored or censored data. R package, Version
587 0.1-4.
588 Ditlevsen, O., 1981. Uncertainty Modeling with Applications to Multidimensional Civil
589 Engineering Systems. McGraw Hill Book Company, New York.
590 Filzmoser, P., Gschwandtner, M., 2013. Mvoutlier: Multivariate outlier detection based on
591 robust methods. R package version 1.9.9.
592 Fotopoulou, S.D., Pitilakis, K.D., 2013a. Vulnerability assessment of reinforced concrete
593 buildings subjected to seismically triggered slow-moving earth slides. Landslides 10,
594 563–582.
595 Fotopoulou, S.D., Pitilakis, K.D., 2013b. Fragility curves for reinforced concrete buildings

596 to seismically triggered slow-moving slides. Soil Dynamics and Earthquake
597 Engineering 48, 143−161.
598 Gartsman, B., van Nooyen, R., Kolechkina, A., 2009. Implementation issues for total risk
599 calculation for groups of sites. Physics and Chemistry of the Earth 34, 619–625.
600 Gatfaoui, H., 2005. How does systematic risk impact us credit spreads? a copula study.
601 Bankers Markets & Investors 77, 5–12.
602 Genest, C., Rémillard, B., Beaudoin, D., 2009. A goodness-of-fit tests for copulas: A review
603 and power study. Insurance: Mathematics and Economics 44 (1), 199–213.
604 Ghalanos, A., Theussl, S., 2012. Rsolnp: General non-linear optimization using augmented
605 lagrange multiplier method. R package version 1.14, URL http://CRAN.R-
606 project.org/package=Rsolnp.
607 Graf, F., Frei, M., Böll, A., 2009. Effects of vegetation on the angle of internal friction of
608 amoraine. Forest Snow and Landscape Research 82 (1), 61–77.
609 Grooteman, F., 2008. Adaptive radial-based importance sampling method for structural
610 reliability. Structural Safety 30, 533–542.
611 Grunsky, E.C., 2002. R: a data analysis and statistical programming environment–an
612 emerging tool for the geosciences. Computers & Geosciences 28, 1219–1222.
613 Hamm, N.A.S., Hall, J.W., Anderson, M.G., 2006. Variance-based sensitivity analysis of the
614 probability of hydrologically induced slope instability. Computers & Geosciences 32,
615 803–817.
616 Hammond, C.J., Hardcastle, J.H., 1992. Determination of values for shear strength
617 parameters using improved methodology. Final report for cooperative agreement No
618 22-C-4-INT-166. US Department of Agriculture, Forest Service.
619 Hasofer, A.A., Lind, A.M., 1974. Exact and invariant second moment code format. Journal
620 of Engineering Mechanics 100 (1), 111–121.
621 Honjo, Y., Hara, T., Kieu Le, T. C., 2010. Level III reliability based design by response
622 surface: an embankment. Proc. 17th Southeast Asian Geotechnical Conference, Vol.
623 2 pp 203–206, Taipei.
624 Huang, J., Griffiths, D.V., 2011. Observations on FORM in a simple geomechanics
625 example. Structural Safety 33, 115–119.
626 Huck, S., Cormier, W., Bounds, W., 1974. Reading Statistics and Research. Harper and
627 Row, New York.
628 Huffman, J.C., Stuedlein, A.W., 2014. Reliability-based serviceability limit state design of

spread footings on aggregate pier reinforced clay. Journal of Geotechnical and Geoenvironmental Engineering 140 (10), 04014055.

Joe, H., 1997. Multivariate Models and Dependence Concept. Chapman and Hall: New York.

Kulhawy F H and Mayne P W 1990 Manual on estimating soil properties for foundation design. Report EL-6800, Electric Power Research Institute, Palo Alto.

Lemon, J., Bolker, B., Oom, S., Klein, E., Rowlingson, B.,Wickham, H., Tyagi, A., Eterradossi, O., Grothendieck, G., Toews, M., Kane, J., Cheetham, M., Turner, R., Witthoft C., Stander, J., Petzoldt, T., Duursma, R., Biancotto, E., Levy, O., 2012. Plotrix: Various plotting functions. R Package, Version 3.4-5.

Li, D.Q., Tang, X.S., Phoon, K.K., Chen, Y.F., 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, Y., Wileyto, E.P., Heitjan, D.H., 2010. Modeling smoking cessation data with alternating states and a cure fraction using frailty models. Statistics in Medicine 29, 627–638.

Lillis, D., 2011. Use R for data analysis and research. New Zealand Science Review 68 (2), 73–79.

Liu, P.L., Lin, H.Z., Der Kiureghian, A., 1989. Calrel User Manual. Report No. UCB/SEMM-89/18, Department of civil engineering, University of California, Berkeley.

Low, B.K., 2005. Reliability-based design applied to retaining walls. Geotechnique 55 (1), 63–75.

Low, B.K., Tang, W.H., 1997. Efficient reliability evaluation using spreadsheet. Journal of Engineering Mechanics 123 (7), 749–752.

Low, B.K., Tang, W.H. 2007. Efficient spreadsheet algorithm for first-order reliability method. Journal of Engineering Mechanics 133 (12), 1378–1387.

Lumb, P., 1970. Safety factors and the probability distribution of soil strength. Canadian Geotechnical Journal 7 (3), 225–242.

Ma, Y., Deng, N., 2000 Deep Foundations, Bridge Engineering Handbook. (W.-F. Chen and L., Duan, eds.), CRC Press.

Mair, P., Satorra, A., Bentler, P.M., 2012. Generating nonnormal multivariate data using copulas: applications to SEM. Multivariate Behavioral Research 47 (4), 547–565.

Matsuo, M., Kuroda, K., 1974. Probabilistic approach to design of embankments. Soils and Foundations 14 (2), 1–17.

662 Moeys, J., Wei, S., 2011. soiltexture: Functions for soil texture plot, classification and
663 transformation. R package, Version 1.2.6.
664 Montgomery, D.C., Runger, G.C., 1999. Applied Statistics and Probability for Engineers.
665 Wiley-Interscience.
666 Mousavi, S.M., Alavi, A.H., Gandomi, A.H., Mollahasani, A., 2011. Nonlinear genetic-based
667 simulation of soil shear strength parameters. Journal of Earth System Science 120
668 (6), 1001–1022.
669 Nelsen, R.B., 2006. An Introduction to Copulas. Springer: New York.
670 Nguyen, V.U., Chowdhury, R.N., 1984. Probabilistic study of spoil pile stability in strip coal
671 mines—two techniques compared. International Journal of Rock Mechanics and
672 Mining Science 21, 303–312.
673 Noda, S., Uwabe, T., Chiba, T., 1975. Relation between seismic coefficient and ground
674 acceleration for gravity quaywall. Report of the port and Harbor Research Institute 14
675 (4), 67−111.
676 Parker, C., Simon, A., Thorne, C.R., 2008. The effects of variability in bank material
677 properties on riverbank stability: Goodwin Creek, Mississippi. Geomorphology 101,
678 533–543.
679 Pebesma, E., Nüst, D., Bivand, R., 2012. The R software environment in reproducible
680 geoscientific research. Eos, Transactions, American Geophysical Union 93 (16),
681 163–163.
682 Phoon, K.K., 2008. Numerical recipes for reliability analysis—a primer. In: Phoon, K.K.
683 (Ed.), Reliability-based Design in Geotechnical Engineering: Computations and
684 Applications. Taylor & Francis (Chapter 1).
685 Phoon, K.K., Honjo, Y., 2005. Geotechnical reliability analyses: towards development of
686 some user-friendly tools. In Proceedings of the 16th International Conference on Soil
687 Mechanics and Geotechnical engineering. Osaka.
688 Phoon, K.K., Kulhawy, F.H., 1999. Characterization of geotechnical variability. Canadian
689 Geotechnical Journal 36, 612–624.
690 Phoon, K.K., Nadim, F., 2004. Modeling Non-Gaussian Random Vectors for FORM: State-
691 of-the-Art Review. Proc. International Workshop on Risk Assessment in Site
692 Characterization and Geotechnical Design, Indian Institute of Science, Bangalore,
693 India: 55–85.
694 Pohl, M., Graf, F., Buttler, A., Rixen, C., 2012. The relationship between plant species

richness and soil aggregate stability can depend on disturbance. Plant and Soil 355, 87–102.

Pruim, R., Kaplan, D., Horton., N., 2012. Mosaic: Operators for computing derivatives and anti-derivatives as functions. R package, Version 0.6-2.

R Development Core Team, 2013. R 2.15.3: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.r-project.org.

Rackwitz, R., Fissler, B., 1978. Structural reliability under combined random load sequences. Computers & Structures 9 (5), 489–494.

Rizzo, M.L., Székely, G.J., 2014. energy: E-statistics (energy statistics). R package version 1.4–0.

Samadi, A., Amiri-Tokaldany, E., Darby, S.E., 2009. Identifying the effects of parameter uncertainty on the reliability of riverbank stability modeling. Geomorphology 106, 219–230.

Schlather, M., Huwe, B., 2004. The use of the language interface of R: two examples for modelling water flux and solute transport. Computers & Geosciences 30, 197–201.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de Institut de Statistique Université Paris 8, 229–231.

Soenksen, P.J., Turner, M.J., Dietsch, B.J., Simon, A., 2003. Stream bank stability in eastern Nebraska. Agriculture Water-resources Investigations Report 03-4 265. U.S. Geological Survey, Lincoln, Nebraska. 96 pp.

Soetaert, K., 2009. rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R package, Version 1.4.

Spellucci, P., 1998. An SQP method for general nonlinear programs using only equality constrained subproblems. Mathematical Programming 82 (3), 413–448.

Subhash, R.L., Jonah, L.K., Solymos, P., 2013. ResourceSelection: Resource selection (Probability) functions for use-availability data. R package, Version 0.2-2.

Székely, G.J., Rizzo, M.L., 2005. A new test for multivariate normality. Journal of Multivariate Analysis 93 (1), 58–80.

Tamura, R., 2007. Rdonlp2: An R extension library to use Peter Spellucci's DONLP2 from R. R package version 0.3-1.

Tobutt, D.C., 1982. Monte Carlo simulation methods for slope stability. Computers & Geosciences 8, 199–208.

Tvedt, L., 1983. Two second-order approximations to the failure probability. Veritas Report

RDIV/20-004083. Det norske Veritas. Veritas Report RDIV/20-004083: Oslo.

Tvedt. L., 1990. Distribution of quadratic forms in normal space: Application to structural reliability. Journal of Engineering Mechanics 116 (6), 1183–1197.

Uzielli, M., Mayne, P.W., 2011. Serviceability limit state CPT-based design for vertically loaded shallow footings on sand. Geomechanics and Geoengineering 6 (2), 91–107.

Venables, W.N., Ripley, B.D., 2012. Modern Applied Statistics with S. Springer, New York, fourth edition, 2002.

Wang, J.P., Huang, D., 2012. RosenPoint: A Microsoft Excel-based program for the Rosenblueth point estimate method and an application in slope stability analysis. Computers & Geosciences 48, 239–243.

Whitman, V.W., 1984. Evaluating calculated risk in geotechnical engineering. Journal of Geotechnical and Geoenvironmental Engineering 110, 145–188.

Wu, X.Z., 2013a. GeoRiskR: Data and functions for geotechnical risk assessment. https://r-forge.r-project.org/R/?group_id=1777. R package, Version: 2.3. or http://xingzhengwu.site88.net/Softwares/index.html.

Wu, X.Z. 2013b. Probabilistic slope stability analysis by a copula-based sampling method. Computational Geosciences 17 (5), 739–755.

Wu, X.Z., 2013c. 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.

Wu, X.Z., 2013d. Using copulas to characterise the dependency of GCL shear strengths. Geosynthetics International 20 (5), 344–357.

Wu, X.Z., 2015a. Assessing the correlated performance functions of an engineering system via probabilistic analysis. Structural Safety 52, 10–19.

Wu, X.Z., 2015b. Development of fragility functions for slope instability analysis. Landslides 12 (1), 165–175.

Wu, X.Z. 2015c. Modelling dependence structures of soil shear strength data with bivariate copulas and applications to geotechnical reliability analysis. Soils and Foundations 55 (5), In press.

Wu, W., Sidle, R.C., 1995. A distributed slope stability model for steep forested watersheds. Water Resources Research 31(8) 2097–2110.

Yan, J., 2007. Enjoy the joy of copulas: with a package copula. Journal of Statistical Software 21 (4), 1–21.

761 Zaitchik, B.F., van Es, H.M., Sullivan, P.J. 2003. Modeling slope stability in Honduras:
762 Parameter sensitivity and scale of aggregation. Soil Science Society of America
763 Journal 67 (1), 268–278.

Appendix 1 R code for multivariate fitting

A1.1 Fitting a marginal density distribution

```
library(GeoRiskR); require(fitdistrplus) #Fig. 3
HH<-SoilShear[which(SoilShear[["name"]]=='Hammond-Hardcastle'),2:3]
Uwei<-HH[,2]; Fric<-HH[,1]
hist(Uwei,probability=TRUE,breaks=20,col=grey(.9),ylim=c(0,0.6),xlim=c(13,19),xlab='Unit weight (kN/m3)')
hist(Fric,probability=TRUE,breaks=20,col=grey(.9),ylim=c(0,0.5),xlim=c(35,48),xlab='Friction angle (degree)')
fitUwei5<-fitdist(Uwei,"weibull")
summary(fitUwei5)$aic; summary(fitUwei5)$loglik
gofstat(fitdist(CFAIEM[,1],"gamma"))$ad
```

A1.2 Fitting a joint distribution

```
library(copula) #Fig. 3
UPhiEM<-cbind(Uwei,Fric)
theta1<-cor(UPhiEM,method="pearson")[1,2]
dgumbel <- function(x,a,b) 1/b*exp((a-x)/b)*exp(-exp((a-x)/b))
pgumbel <- function(q,a,b) exp(-exp((a-q)/b))
qgumbel <- function(p,a,b) a-b*log(-log(p))
a1<-fitdist(Uwei,"gumbel",start=list(a=-3,b=3))$estimate[1]
b1<-fitdist(Uwei,"gumbel",start=list(a=-3,b=3))$estimate[2]
a2<-fitdist(Fric,"gumbel",start=list(a=-3,b=3))$estimate[1]
b2<-fitdist(Fric,"gumbel",start=list(a=-3,b=3))$estimate[2]
myMvd01<-
mvdc(copula=ellipCopula(family="normal",param=theta1),margins=c("gumbel","gumbel"),paramMargins=list(list(a=
a1,b=b1),list(a=a2,b=b2)))
denss01<-contour(myMvd01,dmvdc,xlim=c(13,19),ylim=c(35,48))
sim.number<-1000; set.seed(123)
O11<-rmvdc(myMvd01,sim.number); points(O11)
 library(ResourceSelection) #Fig. 8
NePairs<-Nebraska[,3:9]
colnames(NePairs)=c("Cohe","Fric","Ambient","Saturated","Moisture","Saturation")
kdepairs(NePairs)
```

A1.3 Confidence limits of a regression

```
UPhi_lm<-lm(Fric~Uwei) #Fig. 4
UPhi_pred0<-data.frame(predict(UPhi_lm,newdata=as.data.frame(cbind(Uwei,Fric)),interval="confidence"))
UPhi_nlm<-nls(Fric~c0*Uwei^2+c1*Uwei+ca,start=list(c0=-0.004,c1=0.05,ca=10.0))
```

```
summary(UPhi_nlm)
```

A1.4 KDE of soil classification

```
library(soiltexture); library(plotrix) #Fig. 7
soils<- na.omit(Nebraska[,10:12])
soiltex.return<-soil.texture(soils[1:length(soils[,1]),])
```

Appendix 2 R code for reliability analyses of bearing capacity

A2.1 The FORM proposed by Ang and Tang (1984)

```
library(rootSolve) #for uniroot.all( )
library(mosaic) #for D( )
Y_mu=20.44; Y_sd=1.18; Z_mu=39.81; Z_sd=2.45; X12.Corr<-0.73
DD<-10; BW<-1; ed<-1; FL<-1000
dTerm<-0.5/(ed+DD)*BW*DD^3
corrm2<-matrix(c(1,X12.Corr,X12.Corr,1.0),ncol=2,nrow=2,byrow=T)
eig<-eigen(corrm2,EISPACK=FALSE); eig.X01<-eig$values[1]; eig.X02<-eig$values[2]; tranMat<-eig$vectors;
for (iTer in 1:100) {# iteration start
    if (iTer==1) {Y1<-0; Y2<-0; X0.sigma.mat<-diag(c(Y_sd,Z_sd)); muX<-as.matrix(c(Y_mu,Z_mu));
    CY<-X0.sigma.mat %*% tranMat; Beta<-0.0; BetaNew1<-3}
    dBeta<-abs(Beta-BetaNew1)
    if (dBeta<=0.0002) {Beta; break} #success and exit
    if (iTer!=1) {Y1<-X21.Mn; Y2<-X22.Mn}
    Beta<-BetaNew1
    D.Y1Fun<-D(dTerm*(CY11*Y1+CY12*Y2+muX1)*(tan((45+0.5*(CY21*Y1+CY22*Y2+muX2))/180*pi))^2-FL~Y1)
    D.Y2Fun<-D(dTerm*(CY11*Y1+CY12*Y2+muX1)*(tan((45+0.5*(CY21*Y1+CY22*Y2+muX2))/180*pi))^2-FL~Y2)
    D.Y1Val<-
D.Y1Fun(CY11=CY[1,1],CY21=CY[2,1],CY12=CY[1,2],Y1=Y1,CY22=CY[2,2],Y2=Y2,muX2=muX[2],muX1=muX[1],
dTerm=dTerm,FL=FL)
    D.Y2Val<-
D.Y2Fun(CY11=CY[1,1],CY21=CY[2,1],CY12=CY[1,2],Y1=Y1,CY22=CY[2,2],Y2=Y2,muX2=muX[2],muX1=muX[1],
dTerm=dTerm,FL=FL)
    D.Y1Vals<-D.Y1Val*sqrt(eig.X01); D.Y2Vals<-D.Y2Val*sqrt(eig.X02)
    Alfa.Y1<-D.Y1Vals/sqrt(D.Y1Vals*D.Y1Vals+D.Y2Vals*D.Y2Vals)
    Alfa.Y2<-D.Y2Vals/sqrt(D.Y1Vals*D.Y1Vals+D.Y2Vals*D.Y2Vals)
    DP.Y1<--Alfa.Y1*sqrt(eig.X01); DP.Y2<--Alfa.Y2*sqrt(eig.X02)
    CY11=CY[1,1]; CY21=CY[2,1]; CY12=CY[1,2]; CY22=CY[2,2]; muX1=muX[1]; muX2=muX[2];
```

```
    g109.x<-
function(x)(dTerm*(CY11*x*DP.Y1+CY12*x*DP.Y2+muX1)*(tan((45+0.5*(CY21*x*DP.Y1+CY22*x*DP.Y2+muX2))/
180*pi))^2-FL)
    BetaNew<-uniroot.all(g109.x, interval=c(0,250)); BetaNew1<-min(BetaNew)
    X21.Mn<-DP.Y1*BetaNew1; X22.Mn<-DP.Y2*BetaNew1;
} #next iTer
```

A2.2 The FORM proposed by Low and Tang (2007)

```
library(Rsolnp) #for solnp( )
nVars<-2; ncols<-nVars; nrows<-nVars
FunXiLT<-function(DistName,para1,para2,ni){if (DistName=="norm") {
    xi_LT=para1+ni*para2}; list(xi_LT=xi_LT)}
ObjFn<-function(par){x1<-par[1]; x2<-par[2];
    ncols<-2; nrows<-2
    X12.Corr<-0.73; X21.Corr<-0.73
    corrm4<-matrix(c(1,X12.Corr,X21.Corr,1.0),ncol=ncols,nrow=nrows,byrow=TRUE)
    CormInv<-solve(corrm4) #inverse of matrix
    matXMinusM<-t(t(c(x1,x2)))
    sqrt(t(matXMinusM) %*% CormInv %*% matXMinusM)}
gFn<-function(par){x1<-par[1];x2<-par[2];
    X1_mu=20.44; X1_sd=1.18; X2_mu=39.81/180*pi; X2_sd=2.45/180*pi;
    z1<-FunXiLT("norm",X1_mu,X1_sd,x1)$xi_LT
    z2<-FunXiLT("norm",X2_mu,X2_sd,x2)$xi_LT
    BW<-1; DD<-10; ed<-1; FL<-1000
    z1*BW*DD^3/(2*(ed+DD))*tan(pi/4+z2/2)^2-FL}
x0<-c(1.0,1.0)
powell<-solnp(x0, fun = ObjFn, eqfun = gFn, eqB = c(0))
beta1<-powell$values[length(powell$values)]
beta1
```

A2.3 The SORM

The following R code is modified from the Matlab code written by Phoon (2008).

```
pf1<-pnorm(-beta1)
zz<-ret$par; mm<-nVars
grad<-matrix(nrow=mm,ncol=1)
for (i in 1:mm) {ww<-zz;
    ww[i]<-zz[i]+0.01; P2<-gFn(ww);
    ww[i]<-zz[i]-0.01; P1<-gFn(ww);
```

```
    grad[i]<-(P2-P1)/2/0.01}
Hess<-matrix(nrow=mm,ncol=mm)
for (i in 1:mm){
    for (j in 1:mm){ww<-zz;
        if (i==j) {P0<-gFn(ww); ww[i]<-zz[i]+0.01;
            P2<-gFn(ww); ww[i]<-zz[i]-0.01;
            P1<-gFn(ww); Hess[i,j]<-(P2-2*P0+P1)/(0.01)^2}
        if (i!=j) {ww[i]<-zz[i]+0.01; ww[j]<-zz[j]+0.01;
            P6<-gFn(ww); ww[j]<-zz[j]-0.01;
            P5<-gFn(ww); ww[i]<-zz[i]-0.01;
            P3<-gFn(ww); ww[j]<-zz[j]+0.01;
            P4<-gFn(ww); Hess[i,j]<-(P6-P4-P5+P3)/4/(0.01)^2}}}
library(pracma) #Norm function
dP<-Norm(grad); QQ<-diag(mm);
QQ[1:mm,1]<-zz; qrstr<-qr(QQ)
QQR<-qr.Q(qrstr); QQR1<-t(apply(QQR,1,rev))
AA<-t(QQR1) %*% Hess %*% QQR1;
AAN<-AA[1:mm-1,1:mm-1]/dP
tmpEG=eigen(AAN); WW=tmpEG$values; kappa<-as.matrix(WW)
correction1<-prod(1.0/sqrt(1+beta1*kappa),1);
pf2<-pf1*correction1; #Breitung's formula
beta2<--qnorm(pf2);
TermAA<-beta1*pnorm(-beta1)-dnorm(-beta1)
correction2<-prod(1.0/sqrt(1+(beta1+1)*kappa),1);
A2Term<-TermAA*(correction1-correction2); i_complex<-sqrt(as.complex(-1))
correction3<-Re(prod(1.0/sqrt(1+(beta1+i_complex)*kappa),1));
A3Term<-(beta1+1)*TermAA*(correction1-correction3)
pf3<-pf1*correction1+A2Term+A3Term; #Tvedt's formula
beta3<--qnorm(pf3)
```

A2.4 The CBSM

```
cop_norm_dim2 = normalCopula(dim = 2, param = c(0.73), dispstr = "un")
mvdc_normal = mvdc(copula = cop_norm_dim2, margins = rep("norm",2), paramMargins = list(list(mean=Y_mu,
sd=Y_sd), list(mean=Z_mu, sd=Z_sd)))
set.seed(1640)
rand_mv = rMvdc(n = 100000, mvdc = mvdc_normal)
Y1<- rand_mv[,1]; Z1<- rand_mv[,2]
determTerm<-0.5/(ed+DD)*BW*DD^3
```

```
tanTerm<-tan((45+0.5*Z1)/180*3.1415926)
tanTerm2<-tanTerm^2
MargOfS<-determTerm/FL*Y1*tanTerm2
Prof<-length(MargOfS[MargOfS<1])/(length(MargOfS))
BetaMuSd<-(mean(MargOfS)-1)/sd(MargOfS)
gPerform<-MargOfS #-1
CoVFS<-sd(gPerform)/mean(gPerform)
deltaFSA<-sqrt(CoVFS**2+1); deltaFSB<-sqrt(log(1+CoVFS**2,2.718281828))
BetaLog<-log((mean(gPerform)/deltaFSA),2.718281828)/deltaFSB
Prof; BetaMuSd; BetaLog
```

List of Tables

920

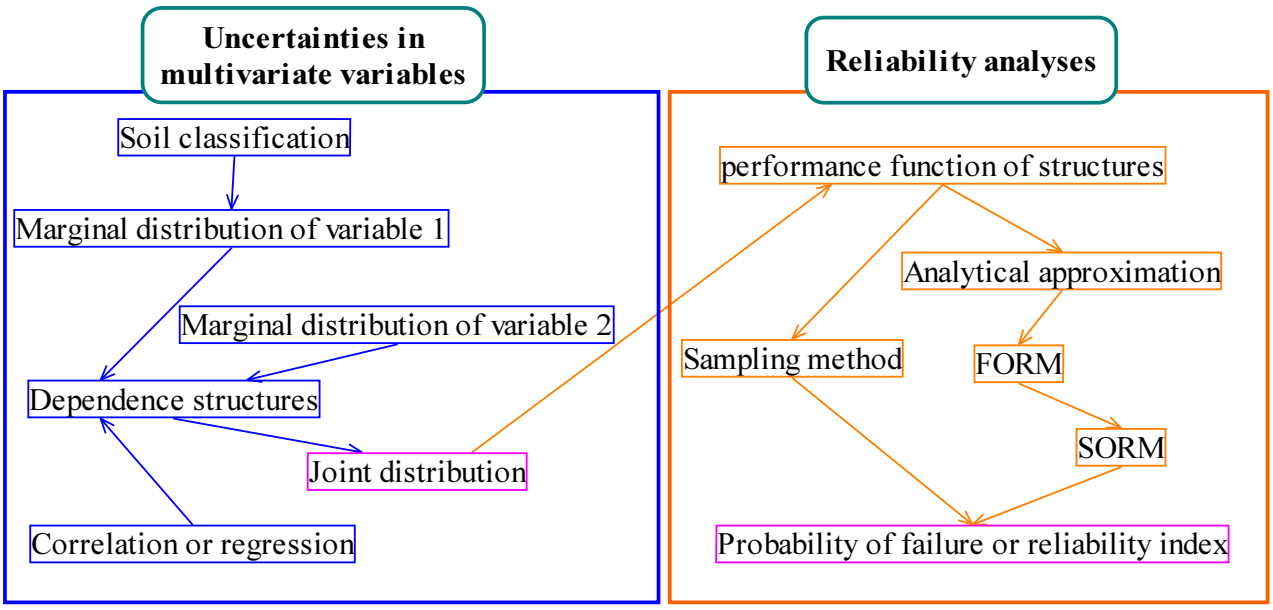
921 Table 1 Probability distributions of input random variables in the case of infinite slope

Variable	Description	Unit	μ	σ	Best-fit	Shape	Scale
c	Cohesion	kPa	9.27	4.65	normal	/	/
ϕ	Friction angle	°	27.2	7.6	Weibull	4.24	30.02
γ_a	Unit weight	kN/m^3	16.83	1.47	Weibull	14.96	17.46
γ_{sat}	Saturated unit weight	kN/m^3	18.3	0.78	normal	/	/

922

923

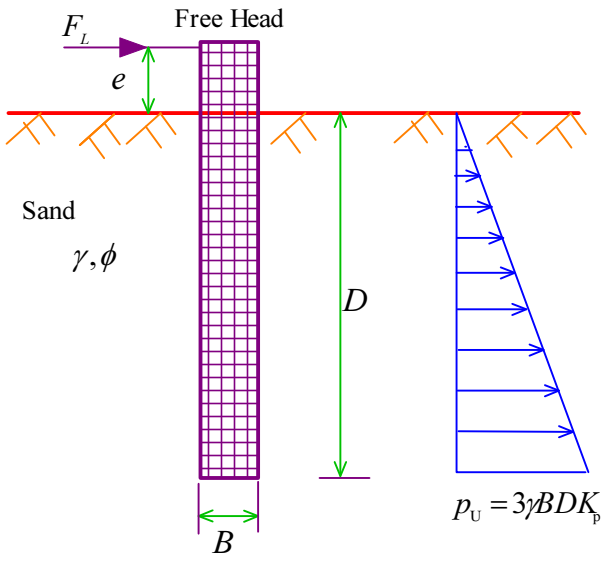
List of Figures



943
944

945 Fig. 1 Overview of the probabilistic stability analysis for a geotechnical engineering issue

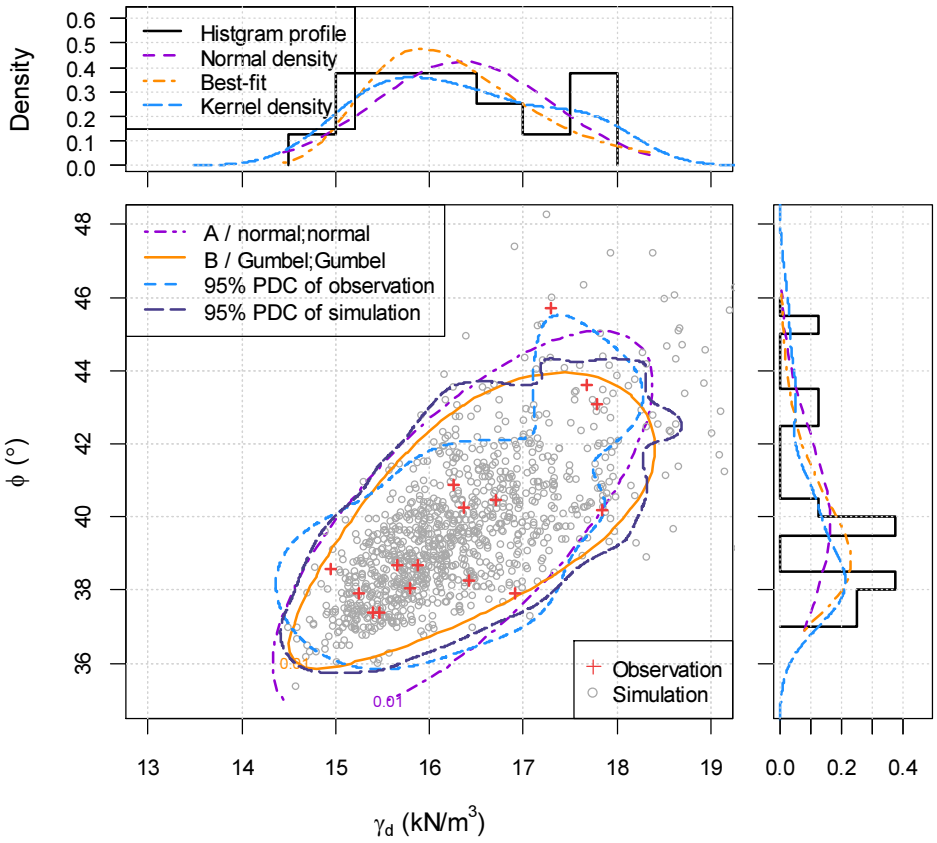
946
947



948

949 Fig. 2 Illustration for a free hand rigid pile in sand under horizontal loads

950



951

952 Fig. 3 Scatter plot of dried unit density and friction angle of sands (data after Hammond and

953 Hardcastle, 1992), overlapped with the joint density contours and the marginal density distributions

954

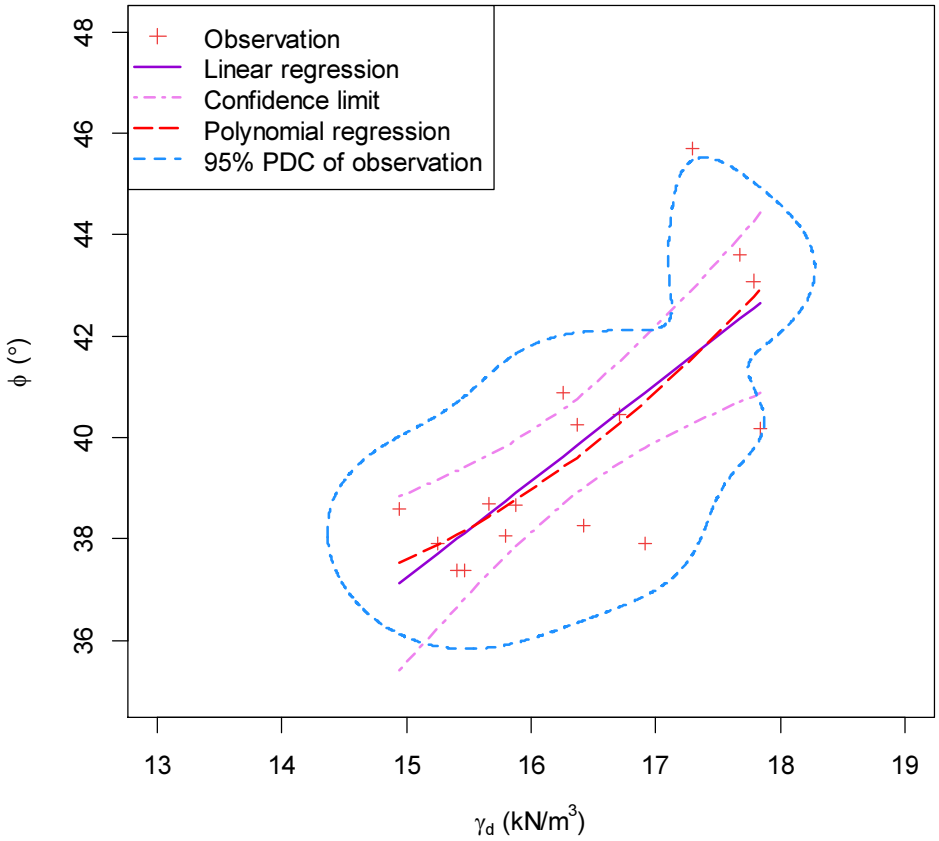
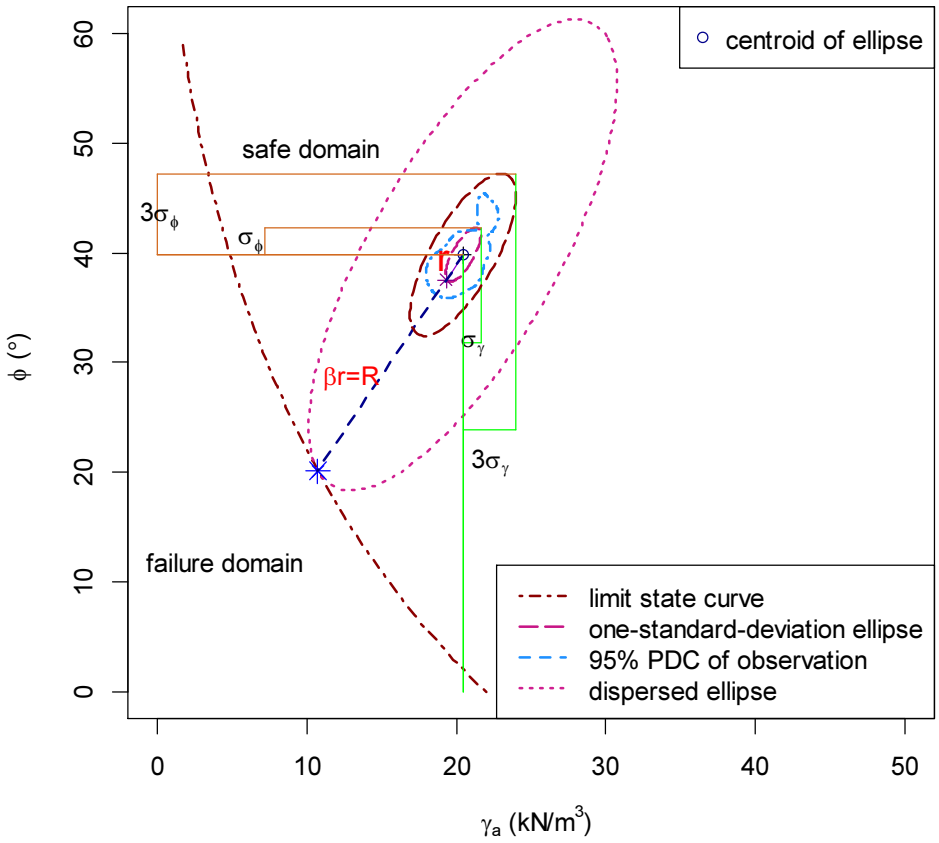


Fig. 4 Linear and polynomial regression on dried unit density and friction angle of sands (data after Hammond and Hardcastle, 1992), 95% prediction intervals (dashed lines) for the linear regression model

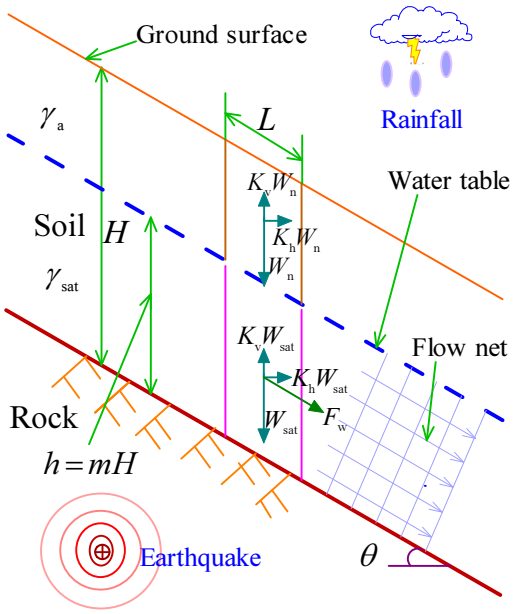


961

962 Fig. 5 Geometric illustration of a reliability index for bivariate normal random variables

963

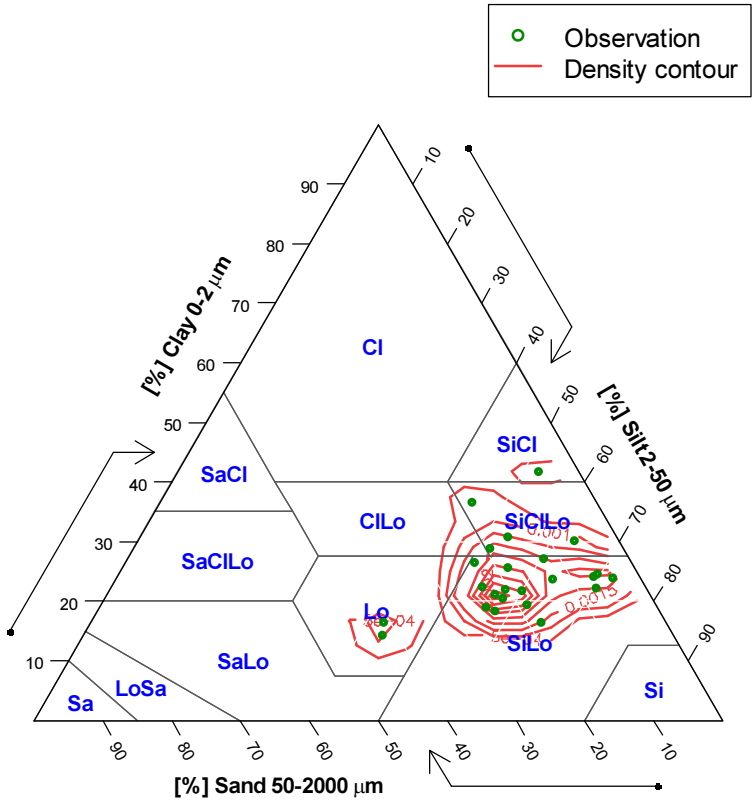
964



965

966 Fig. 6 Infinite slope configuration with parallel seepage and the various forces exerted on the
967 natural zone and saturated zone
968

969
970



971

972 Fig. 7 Ternary diagram for the textural classification of soils on the basis of sand/silt/clay ratios
973 (data after Soenksen et al., 2003)

974

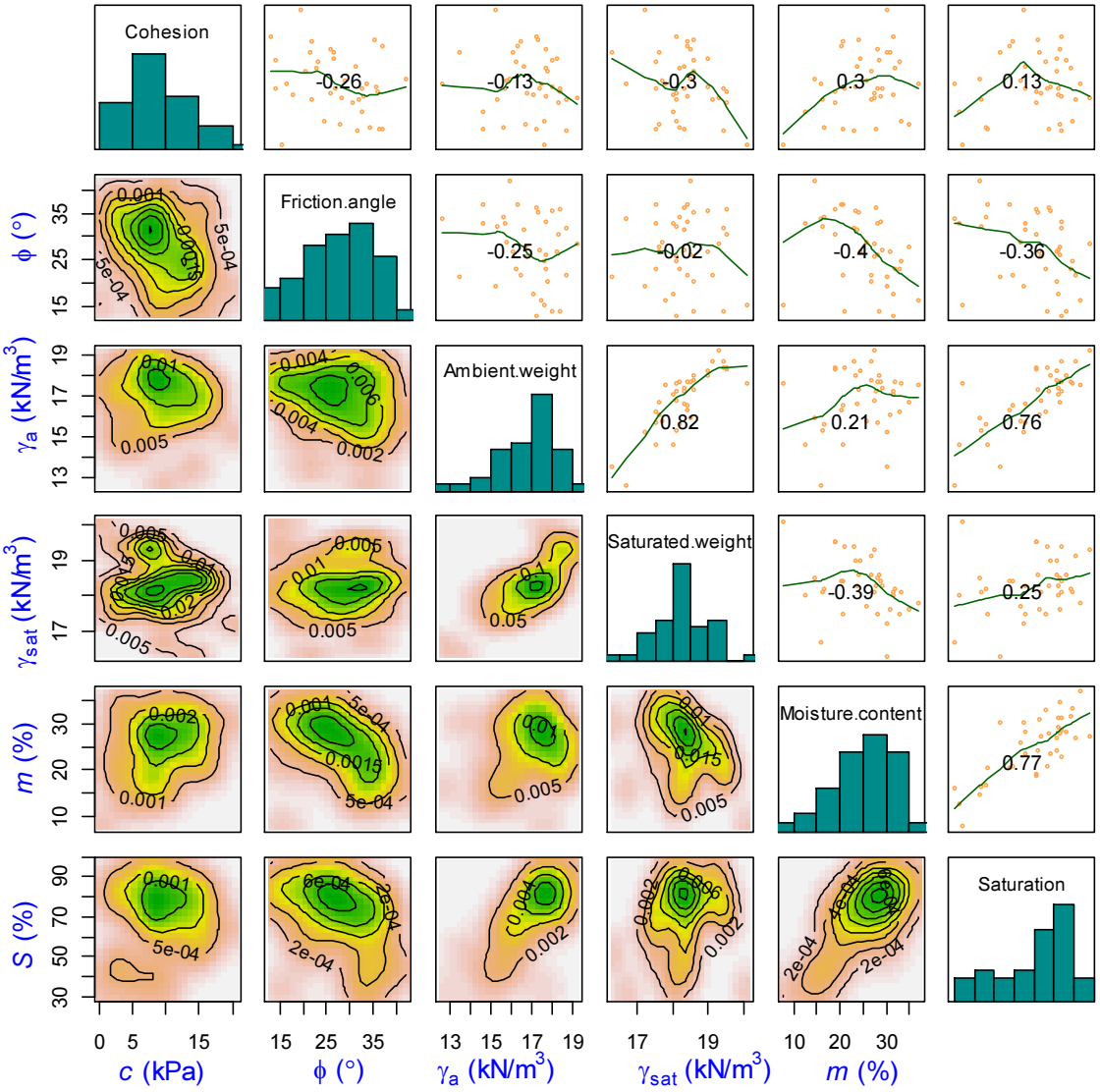
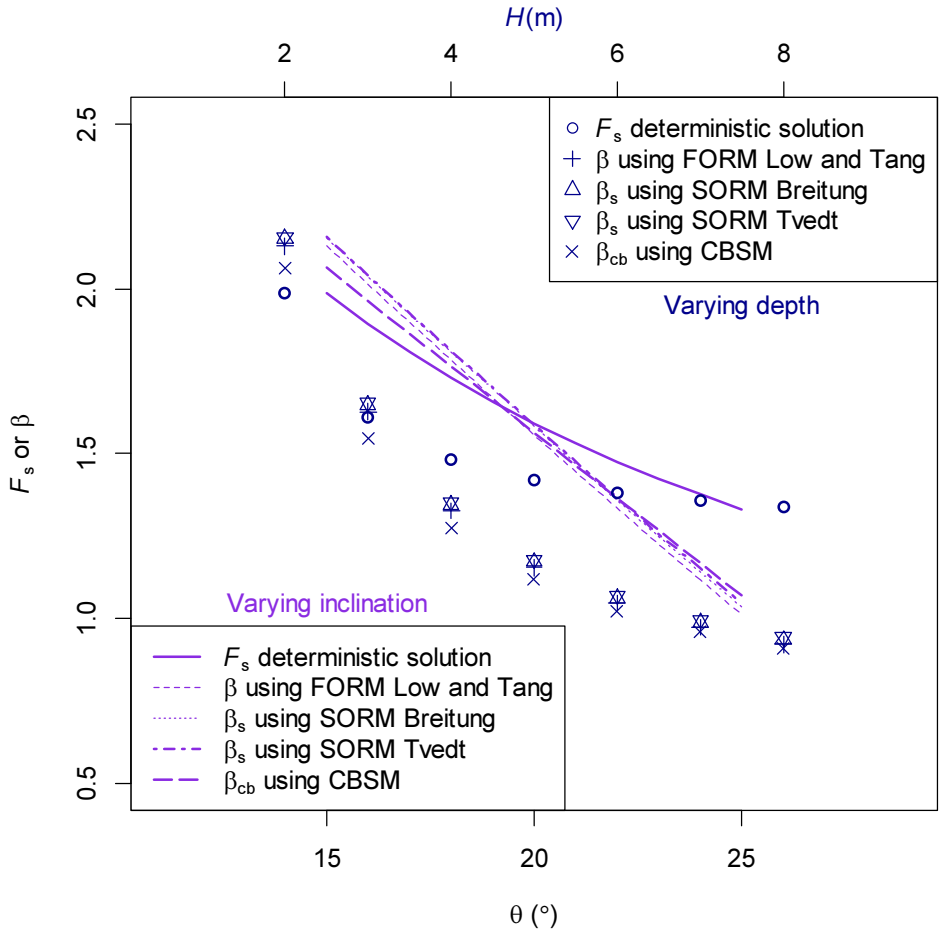


Fig. 8 Scatter plot matrix with bivariate kernel density estimation of the soil data as described in Soenksen et al. (2003), including cohesion, friction angle, ambient unit weight, saturated unit weight, moisture content, and degree of saturation



981

982 Fig. 9 Reliability indexes computed using various reliability methods under varying inclination

983 angle and slope depth