Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields

Li, Y.; Hicks, M. A.; Vardon, P. J.

DOI
10.1016/j.compgeo.2016.05.027

Publication date
2016

Document Version
Accepted author manuscript

Published in
Computers and Geotechnics

Citation (APA)
Li, Y., Hicks, M. A., & Vardon, P. J. (2016). Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields. *Computers and Geotechnics*, *79*, 159-172. https://doi.org/10.1016/j.compgeo.2016.05.027

Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields

Y.J. Li, M.A. Hicks, P.J. Vardon

Section of Geo-Engineering, Department of Geoscience & Engineering, Faculty of Civil Engineering & Geosciences, Delft University of Technology, Delft, The Netherlands

Email: y.li-4@tudelft.nl; m.a.hicks@tudelft.nl; p.j.vardon@tudelft.nl

Cite as:

Li, Y. J., Hicks, M. A., & Vardon, P. J. (2016). **Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields**. Computers and Geotechnics, 79: 159-172. doi:10.1016/j.compgeo.2016.05.027

free link until 16 August, 2016. http://authors.elsevier.com/a/1THNi,63b~copS

This is the Authors' final manuscript. (page number 1-27)

Abstract

A method of combining 3D Kriging for geotechnical sampling schemes with an existing random field generator is presented and validated. Conditional random fields of soil heterogeneity are then linked with finite elements, within a Monte Carlo framework, to investigate optimum sampling locations and the cost-effective design of a slope. The results clearly demonstrate the potential of 3D conditional simulation in directing exploration programmes and designing cost saving structures; that is, by reducing uncertainty and improving the confidence in a project's success. Moreover, for the problems analysed, an optimal sampling distance of half the horizontal scale of fluctuation was identified.

Key words: conditional random fields, Kriging, reliability, sampling efficiency, spatial variability, uncertainty reduction

1. Introduction

Soil properties exhibit three dimensional spatial variability (i.e. heterogeneity). In geotechnical engineering, a site investigation may be carried out, and the data collected and processed in a statistical way to characterise the variability [1–10]. The outcomes of the statistical treatment, e.g. the mean property value, the standard deviation or coefficient of variation, and the spatial correlation distance, may be used as input to a geotechnical model capable of dealing with the spatial variation (e.g. a random field simulation). However, when it comes to making use of the field data, there arises the question: How can we make best use of the available data? The idea is to use the data more effectively, so that it is worth the effort or cost spent in carrying out the investigation, as well as the additional effort in post-processing the data. The aim of this paper is to contribute towards answering this question.

For example, cone penetration tests (CPTs) are often carried out in geotechnical field investigations, in order to obtain data used in implementing the design of a structure. The amount of data from CPT measurements is often larger than from conventional laboratory tests. This is useful, as a large database is needed to accurately estimate the spatial correlation structure of a soil property. For example, Fenton [3] used a database of CPT profiles from Oslo to estimate the correlation statistics in the vertical direction, and Jaksa et al. [5] used a database from Adelaide to estimate the correlation distances in both the vertical and horizontal directions.

In geotechnical engineering, a substantial amount of numerical work has been done using idealised 2D simulations based on collected in-situ data (e.g. [4]), although a 3D simulation would be preferable due to site data generally being collected from a 3D space. However, there are relatively few studies simulating the effect of 3D heterogeneity due to the high computational requirements. Examples include the effect of heterogeneity on shallow foundation settlement [11–13], on steady state seepage [14–16], on seismic liquefaction [17] and on slope reliability [18–27].

The above investigations all used random fields to represent the soil spatial variability and the finite element method to analyse geotechnical performance within a Monte Carlo framework, a form of analysis sometimes referred to as the random finite element method (RFEM) [28]. However, they did not make use of the spatial distribution of related measurement data to constrain the random fields. In other words, for those applications that are based on real field data, many realisations not complying with the field data at the measurement locations will be included in the simulation, which, in turn, will result in an exaggerated range of responses in the analysis of geotechnical performance.

Studies on conditional simulations are available in geostatistics in the field of reservoir engineering [29]. However, there are not many studies dealing with soil spatial variability in geotechnical engineering that utilise conditional simulation (some 2D exceptions include, e.g., [6, 30–32]). This is partly due to the smaller amount of data generally available in geotechnical engineering, and partly due to there often not being a computer program specially implemented for those situations where there are sufficient data (e.g. CPT, vane shear test (VST)), especially in 3D. However, unconditional random fields can easily be conditioned to the known measurements by Kriging [29, 33]. Hence, following the previous 2D work of Van den Eijnden and Hicks [31] and Lloret-Cabot et al. [30], this paper seeks to implement and apply conditional simulation in three dimensional space, in order to reduce uncertainty in the field where CPT measurements are carried out.

Usually, site investigation plans are designed to follow some regular pattern. For example, a systematic grid of sample locations is generally used, due to its simplicity to implement [5]. Moreover, although there are various sampling plans in terms of layouts, it is found that systematically ordered spatial samples are superior in terms of the quality of estimates at unsampled locations [34]. Therefore, this paper will be devoted to implementing a 3D Kriging algorithm for sampling schemes following a regular grid. This will then be combined with an existing 3D random field generator to implement a conditional simulator. However, extension to irregular sampling patterns is straightforward based on the presented framework.

The implemented approach has been applied to two idealised slope stability examples. The first demonstrates how the approach may be used to identify the best locations to conduct borehole testing, and thereby allow an increased confidence in a project's success or failure to be obtained. While it is very important to pay sufficient attention to the required intensity of a site investigation (i.e. the optimal number of boreholes) with respect to the site-specific spatial variability, as highlighted by Jaksa et al. [12], the first example starts by focusing on the optimum locations for carrying out site investigations for a given number of boreholes, before moving on to consider the intensity of testing. The second example compares different candidate slope designs, in order to choose the best (most cost-effective) design satisfying the reliability requirements.

For simplicity, this paper focuses on applications involving only a single soil layer (i.e. a single layer characterised by a statistically homogeneous undrained shear strength), although the extension to multiple soil layers is straightforward. Moreover, the effect of random variation in the boundary locations between different soil layers can also be easily incorporated by conditioning to known boundary locations (e.g. corresponding to where the CPTs have been carried out).

2. Theory and Implementation

2.1 Conditioning

A conditional random field, which preserves the known values at the measurement locations, can be formed from three different fields [28, 35–36]:

$$Z_{rc}(\mathbf{x}) = Z_{ru}(\mathbf{x}) + \left(Z_{km}(\mathbf{x}) - Z_{ks}(\mathbf{x}) \right) \tag{1}$$

where $\mathbf{x}$ denotes a location in space, $Z_{rc}(\mathbf{x})$ is the conditionally simulated random field, $Z_{ru}(\mathbf{x})$ is the unconditional random field, $Z_{km}(\mathbf{x})$ is the Kriged field based on measured values at $\mathbf{x}_i (i = 1, 2, \ldots, N)$, $Z_{ks}(\mathbf{x})$ is the Kriged field based on unconditionally (or randomly) simulated values at the same positions $\mathbf{x}_i (i = 1, 2, \ldots, N)$, and N is the number of measurement locations.

The unconditional random field can be simulated via several methods [37]; for example, interpolated autocorrelation [38], covariance matrix decomposition, discrete Fourier transform or Fast Fourier transform, turning bands, local average subdivision (LAS), and Karhunen–Loeve expansion [39], among others. The LAS method [40] is used in this paper. The Kriged fields are obtained by Kriging

[41], which has found extensive usage in geostatistics [42–43]. The LAS and Kriging methods are briefly reviewed in the following sections.

2.2 Anisotropic random field generation using 3D LAS

The LAS method [40, 44] is used herein to generate the unconditional random fields, using statistics (i.e. mean, variance and correlation structure) based on the observed field data. The LAS method proceeds in a recursive fashion, by progressively subdividing the initial domain into smaller cells, until the random process is represented by a series of local averages. The major advantage is its ability to produce random fields of local averages whose statistics are consistent with the field resolution; that is, it maintains a constant mean over all levels of subdivision, and ensures reduced variances as a function of cell size based on variance reduction theory [45], taking account of spatial correlations between local averages within each level and across levels.

The following covariance function is used in the subdivision process:

$$C(\boldsymbol{\tau}) = C(\tau_1, \tau_2, \tau_3) = \sigma^2 \exp\left(-\frac{2|\tau_1|}{\theta_1} - \sqrt{\left(\frac{2\tau_2}{\theta_2}\right)^2 + \left(\frac{2\tau_3}{\theta_3}\right)^2} \right) \tag{2}$$

where σ^2 is the variance of the soil property, $\boldsymbol{\tau}$ is the lag vector, and θ_1, θ_2 and θ_3, and τ_1, τ_2 and τ_3 are the respective scales of fluctuation and lag distances in the vertical and two lateral coordinate directions, respectively. Herein, an isotropic random field is initially generated by setting $\theta_1 = \theta_2 = \theta_3 = \theta_{iso}$; i.e. so that θ_{iso} equals the horizontal scale of fluctuation, θ_h. This field is then squashed in the vertical direction to give the target vertical scale of fluctuation, θ_v. The 3D LAS implementation of Spencer [25] has been used in this paper, and the reader is referred to Spencer [25] and Hicks and Spencer [19] for more details. Note also that a truncated normal distribution has been used to describe the pointwise variation in material properties [19].

2.3 Kriging

In contrast to conventional deterministic interpolation techniques, such as moving least squares and the radial point interpolation method, Kriging incorporates the variogram (or covariance) into the interpolation procedure; specifically, information on the spatial correlation of the measured points is used to calculate the weights. Moreover, standard errors of the estimation can also be obtained, indicating the reliability of the estimation and the accuracy of the prediction. Kriging is a method of interpolation for which the interpolated values are modelled by a Gaussian process governed by prior covariances and for which confidence intervals can be derived. While interpolation methods based on other criteria need not yield the most likely intermediate values, Kriging provides a best linear unbiased prediction of the soil properties (Z) between known data [43, 46] by assuming the stationarity of the mean and of the spatial covariances, or variograms. A brief review is first given to facilitate understanding of the implementation.

Suppose that $Z_1, Z_2, \ldots, Z_N$ are observations of the random field $Z(\mathbf{x})$ at points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ (i.e. $Z_i = Z(\mathbf{x}_i)$ $(i = 1, 2, \ldots, N)$). The best linear unbiased estimation (i.e. $\hat{Z}$) of the soil property at some location $\mathbf{x}_0$ is given by

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i Z_i = \sum_{i=1}^{N} \lambda_i(\mathbf{x}_0) Z(\mathbf{x}_i) \tag{3}$$

in which N denotes the total number of observations and λ_i denotes the unknown weighting factor associated with observation point $\mathbf{x}_i$, which needs to be determined.

The weights in equation (3), for the estimation at any location $\mathbf{x}_0$, can be found by minimising the variance (σ_e^2) of the Kriging error $\hat{Z} - Z$, which is given as

$$\begin{aligned}\sigma_e^2 = \mathrm{var}\left(\hat{Z} - Z\right) = E\left[\left(\hat{Z} - Z\right)^2\right] &= \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j (c_{ij} - \sigma^2) - 2\sum_{i=1}^{N} \lambda_i (c_{i0} - \sigma^2) + (c_{00} - \sigma^2) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \left(C\left(\left|\mathbf{x}_i - \mathbf{x}_j\right|\right) - \sigma^2\right) - 2\sum_{i=1}^{N} \lambda_i \left(C\left(\left|\mathbf{x}_i - \mathbf{x}_0\right|\right) - \sigma^2\right) + \left(C\left(\left|\mathbf{x}_0 - \mathbf{x}_0\right|\right) - \sigma^2\right) \\ &= -\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma\left(\left|\mathbf{x}_i - \mathbf{x}_j\right|\right) + 2\sum_{i=1}^{N} \lambda_i \gamma\left(\left|\mathbf{x}_i - \mathbf{x}_0\right|\right) - \gamma\left(\left|\mathbf{x}_0 - \mathbf{x}_0\right|\right)\end{aligned} \tag{4}$$

where var() denotes the variance operator and E[] is the expectation operator, $c_{ij} = C\left(\left|\mathbf{x}_i - \mathbf{x}_j\right|\right)$ is the covariance between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$, $c_{i0} = C\left(\left|\mathbf{x}_i - \mathbf{x}_0\right|\right)$ is the covariance between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_0)$, and $c_{00} = C\left(\left|\mathbf{x}_0 - \mathbf{x}_0\right|\right) = C(0) = \sigma^2$ is the variance of $Z(\mathbf{x}_0)$, which is estimated at the target location $\mathbf{x}_0$. The rearrangement in equation (4) makes use of the relationship between a variogram $\gamma(\boldsymbol{\tau})$ and a covariance function $C(\boldsymbol{\tau})$ (i.e. $\gamma(\boldsymbol{\tau}) = C(0) - C(\boldsymbol{\tau}) = \sigma^2 - C(\boldsymbol{\tau})$) and the condition $\sum_{i=1}^{N} \lambda_i = 1$ (in order to ensure that the estimator is unbiased, i.e. $E\left(\hat{Z} - Z\right) = 0$, the weights must sum to one).

To minimise the error variance (i.e. equation (4)), the Lagrange method is used [43]. The weights can then be found by solving the system of equations of size $N + 1$, for a constant mean:

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_N - \mathbf{x}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_0) \\ 1 \end{pmatrix} \quad (5)$$

in which μ is the Lagrangian parameter. For a mean following some trend, the modification to equation (5) is straightforward and interested readers are referred to Fenton [46].

Equation (5) may be expressed as

$$\boldsymbol{\gamma}_{lhs} \boldsymbol{\lambda}_x = \boldsymbol{\gamma}_{rhs} \quad (6)$$

Once equation (5) is solved, the estimated error variance can be expressed by

$$\sigma_e^2 = \mu + \sum_{i=1}^{N} \lambda_i \gamma\left(\left|\mathbf{x}_i - \mathbf{x}_0\right|\right) = \left(\boldsymbol{\gamma}_{lhs}{}^{-1} \boldsymbol{\gamma}_{rhs}\right)^{\mathrm{T}} \boldsymbol{\gamma}_{rhs} \quad (7)$$

where λ_i is a function of the relative positioning of points $\mathbf{x}_i$ and $\mathbf{x}_0$.

Note that the left-hand-side matrix $\boldsymbol{\gamma}_{lhs}$ is a function of only the observation point locations and covariance between them. Therefore, it only needs to be inverted once, and then equations (5) and (3) used repeatedly for building up the field of best estimates at different locations in space. In contrast, the right-hand-side vector $\boldsymbol{\gamma}_{rhs}$ changes as a function of the spatial point $\mathbf{x}_0$, resulting in different weight vectors $\boldsymbol{\lambda}_x$ that are used in equation (3) to get the estimates (point by point) in the domain of interest.

In geotechnical engineering, a sampling strategy following some pattern is generally adopted [1]. For example, CPT sampling is often planned in the form of a regular grid on the ground surface [5]. It is therefore desirable to implement the above Kriging algorithm in the context of some sampling design with a regular pattern. While it is straightforward to implement in 2D, it is less so when implemented in 3D. The most fundamental part is how the left-hand-side matrix of equation (6) is formed. The authors have implemented 3D Kriging for the regular grid sampling strategy shown in Figure 1. The way to set up the left-hand-side matrix and the right-hand-side vector is presented in the Appendix.

2.4 Computational efficiency

There are two aspects involved in the computational efficiency of the above Kriging implementation. One is the total number of equations, which depends on the total number of data points ($N = k \times m \times n$, where k and m are the number of CPT rows in the x and y directions respectively, and n is the number of data points for each CPT profile, see Figure 1) contributing to the left-hand-side matrix; the other is the number of points in the field ($n_f = n_x \times n_y \times n_z$, where n_x, n_y and n_z are the number of points in the three Cartesian directions) that need to be Kriged (i.e. how many times the algorithm will need to be repeated, except for inverting the left-hand-side matrix). The higher the required field resolution (n_f) and the greater the total number of known data points (N), the longer the Kriging will take. In the case of the CPT arrangement in Fig. 1, the size of matrix $\boldsymbol{\gamma}_{lhs}$ (see equation (5) or equation (A1)) and the size of vector $\boldsymbol{\gamma}_{rhs}$ (see equation (5) or equation (A3)) in 3D are m^2 and m times larger than those in 2D (i.e. a cross-section in the x–z plane) respectively. The time it takes to Krige a full 3D field depends on the processing time of each individual step and the number of times each step has to be performed. To Krige a field of size n_f, conditional to N measurement points, the total time may be approximated by

$$t(N,n_f) \approx c_1 n_f N^2 + c_2 N^3 \qquad (8)$$

where the first term represents the time needed for solving the system of equations for all field points (i.e. n_f times) ($O(N^2)$) and the second represents the time needed for inverting the matrix $\boldsymbol{\gamma}_{lhs}$ (i.e. only once) ($O(N^3)$). The constants c_1 and c_2 are functions of the CPU speed and the operation, and in this case are in a ratio of approximately 4:1. Additionally, in all practical cases, $n_f >> N$, so that the calculation time depends mainly on the first term in the above equation; that is, on the number of times (i.e. n_f times) that the matrix–vector multiplication operation, $\boldsymbol{\lambda}_x = \boldsymbol{\gamma}_{lhs}{}^{-1}\boldsymbol{\gamma}_{rhs}$, needs to be performed. Note that $n_f = n_x \times n_y \times n_z$ in 3D is n_y times $n_f = n_x \times n_z$ in 2D and $N = k \times m \times n$ in 3D is m times $N = k \times n$ in 2D. In the examples reported in Section 4, all problems investigated are very long in the third dimension compared to the cross-section. Therefore, the time consumed in a 3D analysis is theoretically $n_y \times m^2$ times that of a 2D analysis, when neglecting the relatively fast, one-off matrix inversion operation and other computation overheads, such as reading/writing and memory operations.

However, despite the significantly greater run-time requirements for Kriging in 3D (as compared to 2D), it is still far less than the time consumed in a nonlinear finite element analysis where plasticity iterations are needed. For Example 1 in Section 4, where $n_x = 20$, $n_y = 100$ and $n_z = 20$, it took, in serial and on average, 134 hours in total for the 500 finite element analyses forming each Monte Carlo simulation (3.0 GHz CPU), whereas Kriging 500 times took about 2.4 hours. In contrast, 500 Kriging interpolations for a 2D cross-section analysis took approximately 8.5 seconds. It is noted that the computation time used for Kriging 500 times is significant in comparison with a single finite element analysis, and therefore should not be considered a pre-processing step if utilising parallel computation for the finite element analyses. Therefore, the computing strategy developed to carry out the analyses for Examples 1 and 2 in Section 4 (comprising around 30,000 realisations in total, and involving 30,000 3D Kriging interpolations) was to run the analyses in parallel (each Kriging and finite element analysis serially on a single computation node) on the Dutch national grid e-infrastructure with high performance computing clusters.

Note that it is possible to prescribe an appropriate neighbourhood size in the algorithm to reduce the computational burden for 3D Kriging. For example, a neighbourhood size of 5×7×*n* may be used to construct the left-hand-side matrix (see the neighbourhood denoted as a rectangle in Figure A.1(a), i.e. by using only the nearest 4 CPT profiles). That is, only those CPT profiles that have a significant influence (i.e. a lag distance within the range of the scale of fluctuation in equation (2)) on the point to be estimated are used to construct the left-hand-side (LHS) matrix. However, using this strategy, for each point (or each subset of points) to be estimated, the left-hand-side matrix is different and will need to be inverted accordingly, so this could increase the computational time if there are a large number of points or cells to be estimated. Therefore, a choice has to be made, to make sure that the time saved by inverting a smaller matrix, instead of a bigger one, outweighs the time consumed by inverting the left-hand-side matrices for all the (subgroups of) cells to be estimated for the case in which a neighbourhood is used. And, of course, there is a trade-off between the estimation accuracy and time saved when such a neighbourhood approach is used. The accuracy will increase as more available data are used to do the Kriging estimation, and so the neighbourhood size depends on the required accuracy and the scales of fluctuation.

Due to the relatively fast inversion of the LHS matrix in the current investigation (the maximum size investigated is N = 500), all CPT profiles have been used for the Kriging in the examples in Section 4. However, one neighbourhood strategy was investigated by using the 4 nearest CPT profiles, and the following uncertainty reduction ratio (a 3D extension to the 1D definition in [31]) has been used to assess the approximation error:

$$u = \frac{\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sum_{k=1}^{n_z}\sigma_e\left(i,j,k\right)}{n_x n_y n_z \sigma} \tag{9}$$

The approximation error may be evaluated by

$$E_u = \frac{\left|u_n - u_a\right|}{u_a} \tag{10}$$

where u_n and u_a are the uncertainty reduction ratios when using a neighbourhood and when all CPT profiles have been used, respectively.

One of the sampling strategies from Example 1 (Section 4, Fig. 9(b)) was used to evaluate the approximation error and the results are listed in Table 1. It can be seen that using a neighbourhood of the 4 nearest CPT profiles has been sufficient in this case.

3. Validation

The conditional simulation of a 5 m high (z), 5 m wide (x) and 25 m long (y) clay block, characterised by a spatially varying undrained shear strength, is presented in this section to demonstrate the procedure and the validity of the implementation described in Sections 2.1–2.3 (and the Appendix). The idea is to show how the measured values are honoured, and to check whether or not the statistical properties (e.g. covariance) of the random fields are maintained after conditioning.

The block is discretised into 20×100×20 cubic cells, with each cell of dimension 0.25 m. The mean of the undrained shear strength is 40 kPa, and the standard deviation is 8 kPa. The degree of anisotropy of the heterogeneity is $\xi = 3$, in which $\xi = \theta_h/\theta_v$ and $\theta_v = 1.0$ m. Five CPT measurement locations in the y direction (at $x = 2.5$ m) are available, each comprising $n = 20$ data points at 0.25 m spacing in the vertical direction. These 'measured' data have been obtained by sampling from a single independent realisation of the spatial variability (i.e. representing the 'actual' in-situ variability). The interval distance between the CPTs in the horizontal direction is $\Delta_y = 5$ m, and the first CPT is located at $y =$ 2.5 m.

Figure 2 shows an example realisation, to illustrate the stages involved in constructing the conditional random field. It shows (a) the unconditional field generated using LAS, (b) the Kriged field based on the unconditionally simulated cell values at the measurement locations, (c) the Kriged field based on the measured data (taken from the reference field), and (d) the conditional random field. It can be seen that the conditional field eliminates unrealistic values from the unconditional simulation by honouring the measurement data at the measurement locations (e.g. corresponding to the centre of the dashed circle in the case of the first CPT). The cross-section from which the CPTs were taken is also shown in

Figures 2(e) and 2(f), together with the known CPT profiles. It is seen that the known CPT profiles are honoured in the conditional random field. Note that, in order to better visualise the fields, a local colour scale is used for all sub-figures in Figure 2.

In order to validate the consistency of the conditioning, the following estimator of the correlation structure along the vertical or horizontal directions of the random field is used to back-figure the covariance structure:

$$\hat{C}(\tau_j = j\Delta\tau) = \frac{1}{n-j}\sum_{i=1}^{n-j}\left(Z_i - \hat{\mu}_Z\right)\left(Z_{i+j} - \hat{\mu}_Z\right) \tag{11}$$

where j = 0, 1, …, n-1, n is the number of data points in the vertical or horizontal direction, τ_j is the lag distance between x_i and x_{i+j}, $\Delta\tau$ is the distance between two adjacent cells vertically or horizontally, $\hat{\mu}_Z$ is the estimated mean, Z is the random soil property and Z_i is the sample of Z. The correlation function is then $\hat{\rho}(\tau_j) = \hat{C}(\tau_j)\big/\hat{C}(0)$, where $\hat{C}(0) = \hat{\sigma}_Z^2$ and $\hat{\sigma}_Z^2$ is the estimated variance [3].

Figure 3 shows the back-figured (a) vertical and (b) horizontal covariances for the unconditional and conditional random fields averaged over 200 realisations, as well as the sample (i.e. CPT) covariances and exact covariances (i.e. equation (2) with only those terms that are associated with the vertical or horizontal direction in 1D). It can be seen that the conditional random field preserves the covariance structure reasonably well in both the horizontal and vertical directions, and that the correlation function fits well the sample correlation for the first quarter of the data points (i.e. n/4) [5, 9]. It is also seen that the covariance for the conditional field lies in between those for the unconditional field and the sampling points.

4. Applications

Two simple examples concerning slope stability are presented in this section, to illustrate how the technique presented in this paper may be used as an aid to geotechnical design. The first involves finding the optimum locations for CPT profiles, in order to minimise the uncertainty in assessing the reliability of a slope. The second involves a cost-effective design with regard to the slope angle when field measurements have already been made (i.e. the positions where the CPT data were taken are already known).

Both examples are presented in terms of the uncertainties in the slope response (with respect to factor of safety). The factors of safety are calculated by 3D finite elements using the strength reduction method [47], with the analyses being undertaken within a probabilistic (RFEM) framework; a flowchart for carrying out such a simulation is shown in Figure 4. The undrained clay behaviour has been modelled using a linear elastic, perfectly plastic Tresca soil model. The clay has a unit weight of 20 kN/m^3, a Young's modulus of 100 MPa and a Poisson's ratio of 0.3. With reference to Figures 5 and 13, the finite element boundary conditions are: a fixed base, rollers on the back of the domain preventing displacements perpendicular to the back face, and rollers on the two ends of the domain, allowing only settlements and preventing movements in the other two directions (i.e. the out-of-slope-face and longitudinal directions). A full explanation of these boundary conditions is given in Spencer [25] and Hicks and Spencer [19].

The random field cell values are mapped onto the 2×2×2 Gauss points in each 20-node finite element, in order to simulate the spatial variability more accurately [19, 48]. Note that the random fields (both conditional and unconditional) have been mapped onto a finite element mesh with an element aspect ratio equal to 2.0 (see Figure 5) to save time for the finite element analyses [25]. A detailed description of how the random field cell values, in this case based on a cell size of 0.25×0.25×0.25 m, are mapped onto the larger non-cubic finite elements is given in Hicks and Spencer [19].

Note that field test (e.g. CPT) data are not directly used in the following examples. That is, the direct measurements from geotechnical tests are typically not directly applicable in a design. Instead, a transformation model is needed to relate the test measurement (e.g. tip resistance from a CPT test) to an appropriate design property (e.g. the undrained shear strength) [49]. The uncertainty involved in the transformation model is not considered in this paper.

4.1 Example 1

The first example considers a proposed 45°, 5 m high, 50 m long slope, that is to be cut from a heterogeneous clay deposit characterised by an undrained shear strength with the following statistics: mean, μ = 20 kPa; standard deviation, σ = 4 kPa; vertical scale of fluctuation, θ_v = 1.0 m; and horizontal scale of fluctuation, θ_h = 6.0 m. A question arises as to how to design the sampling strategy for the soil deposit. For example, if 5 CPTs are to be conducted in a straight line along the axis of the proposed slope, where is the best location to site the CPTs such that the designed slope will have the smallest uncertainty in the realised factor of safety *F*? Hence, this example first investigates the influence of the CPT locations on the standard deviation of the realised factor of safety, followed by the influence of CPT intensity.

Figure 5 shows a cross-section through the slope, and 10 possible positions to locate the CPTs (i = 0, 1, …, 9). Note that the CPTs are taken to be equally spaced (i.e. at 10 m centres) in the third dimension, and that the first and fifth CPTs are located at 5 m and 45 m along the slope axis (see Figure 9(a)). Furthermore, the CPTs are carried out before the slope is excavated, in a block of soil of dimensions 10×50×5 m as indicated in the figure.

Both conditional and unconditional RFEM simulations were carried out, using 500 realisations per simulation, to investigate how the structure response (in this case, the realised factor of safety) changes as the conditioning location changes. Figure 6(a) shows that the uncertainty in the realised factor of safety reduces after conditioning, i.e. after making use of the available CPT information about the soil variability, as indicated by the narrower distribution of realised factor of safety for the conditional simulation. In this figure, the reduction in uncertainty is due to CPT data being taken from location i = 5.

Figure 6(b) shows the sampling efficiency indices with respect to the different CPT locations, in which the sampling efficiency index is defined as

$$I_{se} = \frac{\sigma_u}{\sigma_i} \tag{12}$$

where σ_u is the standard deviation of the realised factor of safety for the unconditional simulation, and σ_i is the standard deviation of the realised factor of safety for the conditional simulation based on column position *i*. Hence $I_{se} = 1$ if the simulation is not conditioned. Clearly, there exists an optimum position (in this case, *i* = 5) to locate the CPTs; i.e. the uncertainty is a minimum if the CPTs are located along the crest of the proposed slope. In contrast, when *i* = 0 and *i* = 1, there is little improvement, because the potential failure planes (in the various realisations) generally pass through zones where the shear strength is, at most, only weakly correlated to values at the left-hand boundary (due to θ_h being only 6 m in this case). It is interesting to note that, although there is not much information included in the slope stability calculation when *i* = 9, i.e. for the CPTs at the slope toe, the reduction in uncertainty is still noticeable, due to the CPTs being located in the zone where slope failure is likely to initiate. This observation highlights that the location of additional information may matter more than how much additional information there is (e.g. contrast the large difference in the amount of directly utilised data between CPT locations *i* = 0 and *i* = 9).

However, it should be remembered that Figures 6(a)-6(b) are for the case of ξ = 6 (corresponding to θ_h = 6 m) and that ξ often takes a larger value in practice. Figures 6(c)-6(f) show that, for ξ = 12 and ξ = 24, the reduction in uncertainty relative to the unconditional case is greater. Moreover, improved values of I_{se} are obtained for CPT locations near the right and left boundaries, due to the higher correlation of soil properties in the horizontal direction.

Figure 7 summarises the results as a function of the degree of anisotropy of the heterogeneity ξ. It is seen that the best locations for carrying out the 5 CPTs are at *i* = 5, 6 and 7. As the value of ξ increases, the sampling efficiency indices increase due to the decreasing Kriging variance σ_e^2, as illustrated in Figure 8 for a *y*–*z* slice at *i* = 5 (i.e. corresponding to where the CPTs are located). It is seen that, for larger values of ξ, the Kriging variance between CPTs can drop well below the input variance of the shear strength (i.e. $\sigma_e^2 \leq 16\ \text{kPa}^2$). Moreover, carrying out CPTs at some distance to the left or right of the slope crest for higher values of ξ can have a similar effect to carrying out CPTs near the crest for smaller values of ξ. For example, Figure 7 shows that the sampling efficiency index for $\xi = 24$ at $i = 2$ is approximately the same as that for $\xi = 12$ at i = 5, 6 and 7.

Note that the same reference 3D random field is used to represent the 'real' field situation in conditioning the random fields in each RFEM analysis. The 3D random fields are conditioned before being mapped onto the finite element mesh, so that they are consistent with sampling the ground before the slope is cut. Hence, for i = 6, 7, 8 and 9, although the CPT measurements are directly used for fewer cells in the FE mesh, they nevertheless have an impact on all cell values via the lateral spatial correlation of soil properties in the original ground profile.

If a second row of CPT tests (at position j) is to be performed in a second phase of the site investigation (e.g. as illustrated in Figure 9(b)), the above procedure can be repeated by changing j in the range 0–9 to locate the best positions for the new CPTs, assuming that the position of the first set of CPT profiles has been set to i = 5. This is shown in Figure 10 for the case of $\xi = 6$. Figure 10(a) shows the probability distributions of the realised factor of safety for the unconditional simulation, the conditional simulation for one row of CPTs at i = 5 and the conditional simulation for an additional row of CPTs at position j = 0. It is seen that the confidence level in the project has been further increased by the second phase of site investigation. Figure 10(b) shows the sampling efficiency indices for various locations j of the second row of CPTs. It suggests that the best location for carrying out the second phase of site investigation can be at either side of the slope crest (at a distance of approximately 3 m (i.e. $\theta_h/2$) from the crest).

To further investigate the influence of CPT intensity on the uncertainty in the realised factor of safety, conditional simulations involving different numbers of CPTs (and thereby different distances (Δ) between adjacent CPTs) have been carried out for the case of ξ = 6, 12 and 24. Figure 11 shows the plan views of CPT layouts for n_{cpt} = 3, 5, 9, 17 and 25 (corresponding to CPT spacings of Δ = 20, 10, 5, 3 and 2 m, respectively), with the locations of the CPTs in the x-direction being fixed at i = 5. Figure 12 shows the influence of CPT intensity on the sampling efficiency indices for the three values of ξ. It is seen that there is only a marginal benefit in increasing the scope of the investigation by having

CPT spacings less than $\Delta \approx \theta_h/2$, especially for the $\xi = 6$ and $\xi = 12$ cases. For $\xi = 24$, the sampling efficiency index is as high as 4 when $\Delta \approx \theta_h/2$, although more CPTs (i.e. $\Delta \approx \theta_h/4$, $n_{cpt} = 9$) may improve the sampling efficiency to a value of 4.5. However, the general finding from Figures 10(b) and 12 is that the optimal sampling distance is around $\theta_h/2$ for the problem investigated, based on the assumed correlation function.

4.2 Example 2

In the second example, a soil deposit characterised by spatially varying undrained shear strength is to be excavated to form a slope of a certain angle. Site investigations have been conducted based on CPT tests. The question is: In order to satisfy a target reliability level of, for example, 95%, as suggested in Eurocode [50] and discussed in Hicks and Nuttall [51], how steep should the slope be designed?

Figure 13 shows three possible slope angles, with the corresponding finite element mesh discretisations. The slope is 5 m high and 50 m long in the third dimension, and the left-hand boundary is taken to be 15 m from the slope toe. Five CPTs were taken along the length of the slope at 10 m centres, at the location of the column of Gauss points nearest the slope crest for the 1:1 slope, as seen in the figure. The clay soil has a mean undrained shear strength of 21 kPa, a coefficient of variation of 0.2, a vertical scale of fluctuation of 1 m and a horizontal scale of fluctuation of 12 m.

The three candidate slopes are (vertical:horizontal) 1:2, 1:1 and 2:1. Based on only the mean undrained shear strength, these three slopes have deterministic factors of safety F_d of 1.73, 1.29 and 1.07. Both conditional and unconditional simulations were carried out to investigate the reliability of each slope, and, for each simulation, 500 realisations were analysed. Note that, as in the previous example, one reference random field is generated first and assumed to represent the real field situation. The conditional random fields used in the RFEM analyses are therefore based on CPT measurements taken from this ‘real’ field.

The stability of the slopes was calculated by the strength reduction method by applying gravitational loading. The probability density functions of the realised factor of safety are shown in Figure 14 for the three slopes, for both conditional and unconditional simulations. The deterministic factors of safety

F_d, i.e. the factors of safety based on the mean property values, are also shown. It is seen that, if unconditional simulation is used, there is a significant chance that the 2:1 slope will fail (the probability of failure is the area under the pdf for the realised factor of safety smaller than 1.0). Unsurprisingly, the gentlest (i.e. 1:2) slope has the lowest probability of failure. However, once again, conditional simulations significantly reduce the uncertainty in the structural response, as clearly demonstrated by the narrower probability distributions. In particular, the reliability of the steepest slope increases from 77% to 99% when the CPT measurements are taken into account.

The results show that, if unconditional simulations are used, the 1:1 and 1:2 slopes satisfy a target reliability level of 95%, whereas the 2:1 slope does not. However, when the additional information from the CPT profiles is used, all three cases meet the target reliability. This means that the embankment may be designed to a slope angle of 2:1 if the CPT measurements are used in the simulation, which is, if possible, a more logical thing to do. This has implications for the soil volume to be excavated and thereby cost, although the cost can be site and situation dependent (e.g. on whether there are nearby structures). A best design is a design that meets the requirements set by standards, while, at the same time, minimising the cost. In this case, the steepest slope is likely to be the most cost-effective design.

5. Conclusions

An approach for conditioning 3D random fields based on CPT measurements has been implemented and validated, and then applied to two numerical examples to illustrate its potential use for geotechnical site exploration and cost-effective design. It has been shown that conditional simulations based on CPT data are able to increase the confidence in a design's success or failure. Indeed, the reliability from a conditional simulation can be thought of as a conditional reliability (or conditional probability of failure not occurring), i.e. based on a 'posterior' distribution of the structure performance after taking account of the spatial distribution of all the measured CPT data points. In contrast, the unconditional simulation based on random field theory only results in a 'prior' distribution of the structure response. This was clearly demonstrated by the updating of the probability density distributions in the two numerical examples. Although Bayesian updating is not used in this paper, the effect is similar.

If further CPT measurements are required, the approach can be repeated for updating the response probability density function. In this way, the confidence in the probability of failure or survival will be further increased. In fact, in many cases a multi-stage site investigation may be carried out, with the results of the initial analysis guiding further field tests. As demonstrated in the first example, if a second stage of site exploration were to be conducted, it is possible to find out the optimum location for the additional testing. This highlights the method's potential use in directing site exploration programmes and thereby improving the efficient use of field measurements. For the first example considered in this paper, an optimal sampling distance of half the horizontal scale of fluctuation was identified when an exponential correlation function is used. For the second example, the conditional simulation led to a more cost-effective design.

Acknowledgements

This research was funded by the China Scholarship Council (CSC) and by the Section of Geo-Engineering at Delft University of Technology. It was carried out on the Dutch National e-infrastructure with the support of the SURF Foundation. Special thanks are given to SURFsara advisor Anatoli Danezi for her kind support in developing a computing strategy.

References

[1] DeGroot DJ, Baecher GB. Estimating autocovariance of in-situ soil properties. ASCE Journal of Geotechnical Engineering 1993; 119 (1): 147–66.

[2] Fenton GA. Estimation for stochastic soil models. ASCE J. Geotech. Geoenv. Eng. 1999; 125(6): 470–85.

[3] Fenton GA. Random field modeling of CPT data. ASCE J. Geotech. Geoenv. Eng. 1999; 125(6): 486–98.

[4] Hicks MA, Onisiphorou C. Stochastic evaluation of static liquefaction in a predominantly dilative sand fill. Géotechnique 2005; 55(2): 123–33.

[5] Jaksa MB, Kaggwa WS, Brooker PI. Experimental evaluation of the scale of fluctuation of a stiff clay. In: Proc. 8th Int. Conf. on Applications of Statistics and Probability in Civil Engineering, Sydney; 1999. p. 415–22.

[6] Lloret-Cabot M, Fenton GA, Hicks MA. On the estimation of scale of fluctuation in geostatistics. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 2014; 8(2): 129–40.

[7] Lundberg AB, Li Y. Probabilistic characterization of a soft Scandinavian clay supporting a light quay structure. In: Proc. 5th International Symposium on Geotechnical Safety and Risk, Rotterdam; 2015. p.170–75.

[8] Phoon KK, Kulhawy FH. Characterization of geotechnical variability. Canadian Geotechnical Journal 1999; 36 (4): 612–24.

[9] Uzielli M, Vannucchi G, Phoon KK. Random field characterisation of stress-normalised cone penetration testing parameters. Géotechnique 2005; 55(1): 3–20.

[10] Zhao HF, Zhang LM, Xu Y, Chang D.S. Variability of geotechnical properties of a fresh landslide soil deposit. Engineering Geology 2013; 166: 1–10.

[11] Fenton GA, Griffiths DV. Three-dimensional probabilistic foundation settlement. ASCE J. Geotech. Geoenv. Eng. 2005; 131(2): 232–39.

[12] Jaksa MB, Goldsworthy JS, Fenton GA, Kaggwa WS, Griffiths DV, Kuo YL, Poulos HG. Towards reliable and effective site investigations. Géotechnique 2005; 55(2): 109–21.

[13] Kuo YL, Jaksa MB, Kaggwa WS, Fenton GA, Griffiths DV, Goldsworthy JS. Probabilistic analysis of multi-layered soil effects on shallow foundation settlement. In: Proc. 9th Australia New Zealand Conference on Geomechanics, Aukland, New Zealand; 2004. p. 541–47.

[14] Griffiths DV, Fenton GA. Observations on two- and three-dimensional seepage through a spatially random soil. In: Proc. 7th Int. Conf. on Applications of Statistics and Probability in Civil Engineering, Paris, France; 1995. p. 65–70.

[15] Griffiths DV, Fenton GA. Three-dimensional seepage through spatially random soil. ASCE J. Geotech. Geoenv. Eng.1997; 123(2): 153–60.

[16] Griffiths DV, Fenton GA. Probabilistic analysis of exit gradients due to steady seepage. ASCE J. Geotech. Geoenv. Eng. 1998; 124(9): 789–97.

[17] Popescu R, Prevost JH, Deodatis G. 3D effects in seismic liquefaction of stochastically variable soil deposits. Géotechnique 2005; 55(1): 21–31.

[18] Griffiths DV, Huang J, Fenton GA. On the reliability of earth slopes in three dimensions. Proc. R. Soc. A 2009; 465(2110): 3145–64.

[19] Hicks MA, Spencer WA. Influence of heterogeneity on the reliability and failure of a long 3D slope. Comput Geotech 2010; 37: 948–55.

[20] Hicks MA, Chen J, Spencer WA. Influence of spatial variability on 3D slope failures. In: Proc. 6th Int. Conf. Computer Simulation Risk Analysis and Hazard Mitigation, Kefalonia; 2008. p. 335–42.

[21] Hicks MA, Nuttall JD, Chen J. Influence of heterogeneity on 3D slope reliability and failure consequence. Computers and Geotechnics 2014; 61: 198–208.

[22] Li Y, Hicks MA. Comparative study of embankment reliability in three dimensions, In: Proc. 8th European Conference on Numerical Methods in Geotechnical Engineering (NUMGE), Delft; 2014. p. 467–72.

[23] Li Y, Hicks MA, Nuttall JD. Probabilistic analysis of a benchmark problem for slope stability in 3D. In: Proc. 3rd Int. Symp. Computational Geomech, Krakow, Poland; 2013. p. 641–8.

[24] Li YJ, Hicks MA, Nuttall, JD. Comparative analyses of slope reliability in 3D. Engineering Geology 2015; 196: 12–23.

[25] Spencer WA. Parallel stochastic and finite element modeling of clay slope stability in 3D. PhD thesis, University of Manchester, UK; 2007.

[26] Spencer WA, Hicks MA. 3D stochastic modelling of long soil slopes, In: Proc. 14th Conf. of Assoc. for Computational Mechanics in Engineering, Belfast; 2006. p. 119–22.

[27] Spencer WA, Hicks MA. A 3D finite element study of slope reliability. In: Proc. 10th Int. Symp. Num Models in Geomech, Rhodes; 2007. p. 539–43.

[28] Fenton GA, Griffiths DV. Risk assessment in geotechnical engineering. New York: John Wiley & Sons; 2008.

[29] Chiles JP, Delfiner P. Geostatistics: modeling spatial uncertainty. John Wiley & Sons; 2009.

[30] Lloret-Cabot M, Hicks MA, van den Eijnden AP. Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging. Géotechnique letters 2012; 2: 123–7.

[31] van den Eijnden AP, Hicks MA. Conditional simulation for characterizing the spatial variability of sand state. In: Proc. 2nd International Symposium on Computational Geomechanics, Dubrovnik, Rhodes, Greece; 2011. p. 288–96.

[32] Vanmarcke EH, Fenton GA. Conditioned simulation of local fields of earthquake ground motion. Structural Safety 1991; 10(1): 247–64.

[33] Journel, AG. Geostatistics for conditional simulation of ore bodies. Economic Geology 1974; 69(5): 673–87.

[34] Olea, RA. Systematic sampling of spatial functions. Series of Spatial Analysis No. 7, Kansas Geological Survey, Lawrence, Kans; 1984.

[35] Frimpong, S, Achireko PK. Conditional LAS stochastic simulation of regionalized variables in random fields. Computational Geosciences 1998; 2: 37–45.

[36] Journel AG, Huijbregts CJ. Mining Geostatistics. New York: Academic Press; 1978.

[37] Fenton GA. Error evaluation of three random field generators. ASCE J. Eng. Mech. 1994; 120 (12): 2478–97.

[38] Ji J, Liao HJ, Low BK. Modeling 2D spatial variation in slope reliability analysis using interpolated autocorrelations. Computers and Geotechnics 2012; 40: 135–46.

[39] Phoon KK, Huang SP, Quek ST. Simulation of second-order processes using Karhunen–Loeve expansion. Computers & Structures 2002; 80(12): 1049–60.

[40] Fenton GA, Vanmarcke EH. Simulation of random fields via local average subdivision. ASCE Journal of Engineering Mechanics 1990; 116(8): 1733–49.

[41] Krige DG. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 1951; 52 (6): 119–39.

[42] Cressie N. The origins of Kriging. Mathematical Geology 1990; 22 (3): 239–52.

[43] Wackernagel H. Multivariate geostatistics: An introduction with applications. Germany: Springer; 2003.

[44] Fenton GA. Simulation and analysis of random fields. PhD Thesis, Princeton University; 1990.

[45] Vanmarcke EH. Random fields: analysis and synthesis. Cambridge, Massachusetts: The MIT Press; 1983.

[46] Fenton GA. Data analysis/geostatistics. In: Probabilistic methods in geotechnical engineering. Griffiths DV, Fenton GA, editors. New York: Springer; 2007. p. 51–73.

[47] Smith IM, Griffiths DV, Margetts L. Programming the finite element method, 5th ed. New York: John Wiley & Sons; 2013.

[48] Hicks MA, Samy K. Influence of heterogeneity on undrained clay slope stability. Quarterly Journal of Engineering Geology and Hydrogeology 2002; 35: 41–9.

[49] Phoon KK, Kulhawy FH. Evaluation of geotechnical property variability. Canadian Geotechnical Journal 1999; 36(4): 625–39.

[50] European Committee for Standardisation (CEN). Eurocode 7: geotechnical design. Part 1: general rules. EN 1997-1, CEN; 2004.

[51] Hicks MA, Nuttall JD. Influence of soil heterogeneity on geotechnical performance and uncertainty: a stochastic view on EC7. In: Proc. 10th International Probabilistic Workshop, Stuttgart; 2012. p. 215–27.

Appendix

A.1 Forming the left-hand-side matrix for Kriging

Suppose there are $k \times m$ CPT locations that follow a rectangular grid at the ground surface. That is, there are k rows in the x direction and, within each row, m CPT profiles in the y direction (Figure 1). Assuming that there are n data points for each CPT profile, the global numbering scheme for all the CPT data points is shown in Figure A.1 for the case of k = 2.

Following the basic equation (6), of size $N + 1 = k \times m \times n + 1$, the left-hand-side matrix is formulated as

$$\gamma_{lhs} = \begin{pmatrix} \mathbf{v}_{1,1} & \mathbf{v}_{1,2} & \mathbf{v}_{1,3} & \cdots & \mathbf{v}_{1,m} & \mathbf{v}_{1,m+1} & \mathbf{v}_{1,m+2} & \mathbf{v}_{1,m+3} & \cdots & \mathbf{v}_{1,2m} & \cdots & \mathbf{v}_{1,(k-1)m+1} & \mathbf{v}_{1,(k-1)m+2} & \mathbf{v}_{1,(k-1)m+3} & \cdots & \mathbf{v}_{1,km} & 1 \\ \mathbf{v}_{2,1} & \mathbf{v}_{2,2} & \mathbf{v}_{2,3} & \cdots & \mathbf{v}_{2,m} & \mathbf{v}_{2,m+1} & \mathbf{v}_{2,m+2} & \mathbf{v}_{2,m+3} & \cdots & \mathbf{v}_{2,2m} & \cdots & \mathbf{v}_{2,(k-1)m+1} & \mathbf{v}_{2,(k-1)m+2} & \mathbf{v}_{2,(k-1)m+3} & \cdots & \mathbf{v}_{2,km} & 1 \\ \mathbf{v}_{3,1} & \mathbf{v}_{3,2} & \mathbf{v}_{3,3} & \cdots & \mathbf{v}_{3,m} & \mathbf{v}_{3,m+1} & \mathbf{v}_{3,m+2} & \mathbf{v}_{3,m+3} & \cdots & \mathbf{v}_{3,2m} & \cdots & \mathbf{v}_{3,(k-1)m+1} & \mathbf{v}_{3,(k-1)m+2} & \mathbf{v}_{3,(k-1)m+3} & \cdots & \mathbf{v}_{3,km} & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{v}_{km,1} & \mathbf{v}_{km,2} & \mathbf{v}_{km,3} & \cdots & \mathbf{v}_{km,m} & \mathbf{v}_{km,m+1} & \mathbf{v}_{km,m+2} & \mathbf{v}_{km,m+3} & \cdots & \mathbf{v}_{km,2m} & \cdots & \mathbf{v}_{km,(k-1)m+1} & \mathbf{v}_{km,(k-1)m+2} & \mathbf{v}_{km,(k-1)m+3} & \cdots & \mathbf{v}_{km,km} & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & \cdots & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \tag{A1}$$

in which $\mathbf{v}_{i,j}$ is a matrix representing the correlation structure between CPT_i and CPT_j (where each CPT has n data points),

$$\mathbf{v}_{i,j} = \begin{pmatrix} d_{(i-1)n+1,(j-1)n+1} & d_{(i-1)n+1,(j-1)n+2} & d_{(i-1)n+1,(j-1)n+3} & \cdots & d_{(i-1)n+1,(j-1)n+n} \\ d_{(i-1)n+2,(j-1)n+1} & d_{(i-1)n+2,(j-1)n+2} & d_{(i-1)n+2,(j-1)n+3} & \cdots & d_{(i-1)n+2,(j-1)n+n} \\ d_{(i-1)n+3,(j-1)n+1} & d_{(i-1)n+3,(j-1)n+2} & d_{(i-1)n+3,(j-1)n+3} & \cdots & d_{(i-1)n+3,(j-1)n+n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{(i-1)n+n,(j-1)n+1} & d_{(i-1)n+n,(j-1)n+2} & d_{(i-1)n+n,(j-1)n+3} & \cdots & d_{(i-1)n+n,(j-1)n+n} \end{pmatrix} \quad \text{(A2)}$$

where $(i, j) = 1, 2, 3, \ldots, m, m+1, m+2, m+3, \ldots, 2m, \ldots\ldots, (k-1)m+1, (k-1)m+2, (k-1)m+3, \ldots, km$ and $d_{r,s}$ $(r = (i-1)n+1, \ldots, (i-1)n+n)$ $(s = (j-1)n+1, \ldots, (j-1)n+n)$ are the components of the submatrix $\mathbf{v}_{i,j}$, which can be expressed in the form of a covariance function between data points r and s (equation (2)).

A.2 Forming the right-hand-side vector for Kriging

The right-hand-side vector is formulated as

$$\boldsymbol{\gamma}_{rhs} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \vdots \\ \mathbf{v}_{km} \\ 1 \end{pmatrix} \quad \text{(A3)}$$

in which $\mathbf{v}_p$ is a vector representing the correlation structure between the estimation point and CPT_p,

$$\mathbf{v}_p = \begin{pmatrix} d_{(p-1)n+1} \\ d_{(p-1)n+2} \\ d_{(p-1)n+3} \\ \vdots \\ d_{(p-1)n+n} \end{pmatrix} \quad \text{(A4)}$$

where $p = 1, 2, 3, \ldots, m, m+1, m+2, m+3, \ldots, 2m, \ldots\ldots, (k-1)m+1, (k-1)m+2, (k-1)m+3, \ldots, km$ and d_t $(t = (p-1)n+1, \ldots, (p-1)n+n)$ are the components of the subvector $\mathbf{v}_p$, which can be expressed in the form of a covariance function (equation (2)) between data points t and the point at which the value is to be estimated (Figure A.1).

The unknown weight vector is

$$\boldsymbol{\lambda}_x = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \boldsymbol{\lambda}_3 \\ \vdots \\ \boldsymbol{\lambda}_{km} \\ \mu \end{pmatrix} \tag{A5}$$

in which $\boldsymbol{\lambda}_q$ is the weight subvector for CPT$_q$,

$$\boldsymbol{\lambda}_q = \begin{pmatrix} \lambda_{(q-1)n+1} \\ \lambda_{(q-1)n+2} \\ \lambda_{(q-1)n+3} \\ \vdots \\ \lambda_{(q-1)n+n} \end{pmatrix} \tag{A6}$$

where q = 1, 2, 3, …, m, m+1, m+2, m+3, …, 2m, ……, (k-1)m+1, (k-1)m+2, (k-1)m+3, …, km.

List of Figures

Figure 1. Example CPT sampling strategy (k = 2, m = 5)

Figure 2. Example illustrations of the unconditional random field (a), the Kriged field based on the randomly simulated data (b), the Kriged field based on the CPT data (c), the conditional random field (d), cross-sections (e and f) in the longitudinal direction taken from the Kriged field (c) and from the conditional random field (d), respectively. Dashed circle indicates the position of the first CPT in subfigures (a) and (c-d)

Figure 3. Vertical and horizontal covariance functions averaged over 200 realisations (θ_v = 1.0 m, θ_h = 3.0 m)

Figure 4. Flowchart for conditional RFEM simulation

Figure 5. Finite element mesh and possible numbered CPT locations at a cross-section through the proposed 50 m long slope (dashed lines indicate the excavated soil mass and numbers correspond to Gauss point locations within the finite elements)

Figure 6. Simulation results for Example 1 (based on θ_v = 1.0 m and 500 realisations per simulation)

Figure 7. Sampling efficiency indices for various values of ξ

Figure 8. Kriging variance for various values of ξ (y–z slice at i = 5)

Figure 9. CPT layout illustration (plan view) for a single row (a) and two rows (b)

Figure 10. Influence of CPT location j during the second phase of site investigation (based on θ_v = 1.0 m and 500 realisations per simulation)

Figure 11. CPT layouts (plan views) for various numbers of boreholes (n_{cpt} = 3, 5, 9, 17, 25 and Δ denotes the distance between CPTs)

List of Tables

Table 1. Comparison of uncertainty reduction ratio for using a local neighbourhood and using all the CPT profiles

	θ_h = 6 m	θ_h = 12 m	θ_h = 24 m
u_n (local neighbourhood)	0.7231	0.5449	0.3953
u_a (all CPTs)	0.7220	0.5442	0.3946
E_u	1.5‰	1.3‰	1.8‰

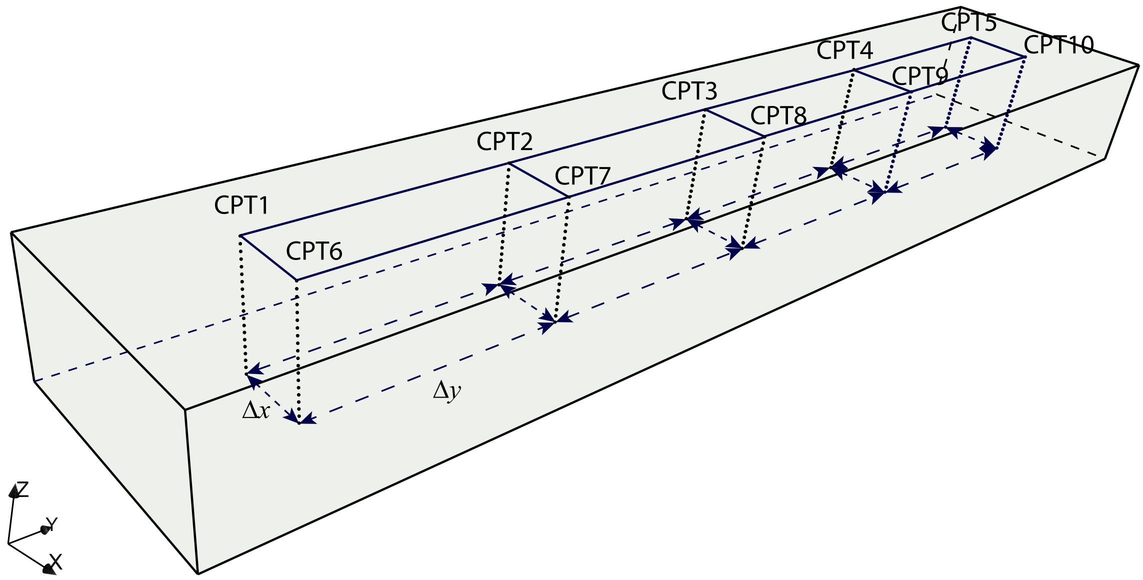


Figure 1: Example CPT sampling strategy ($k = 2$, $m = 5$)

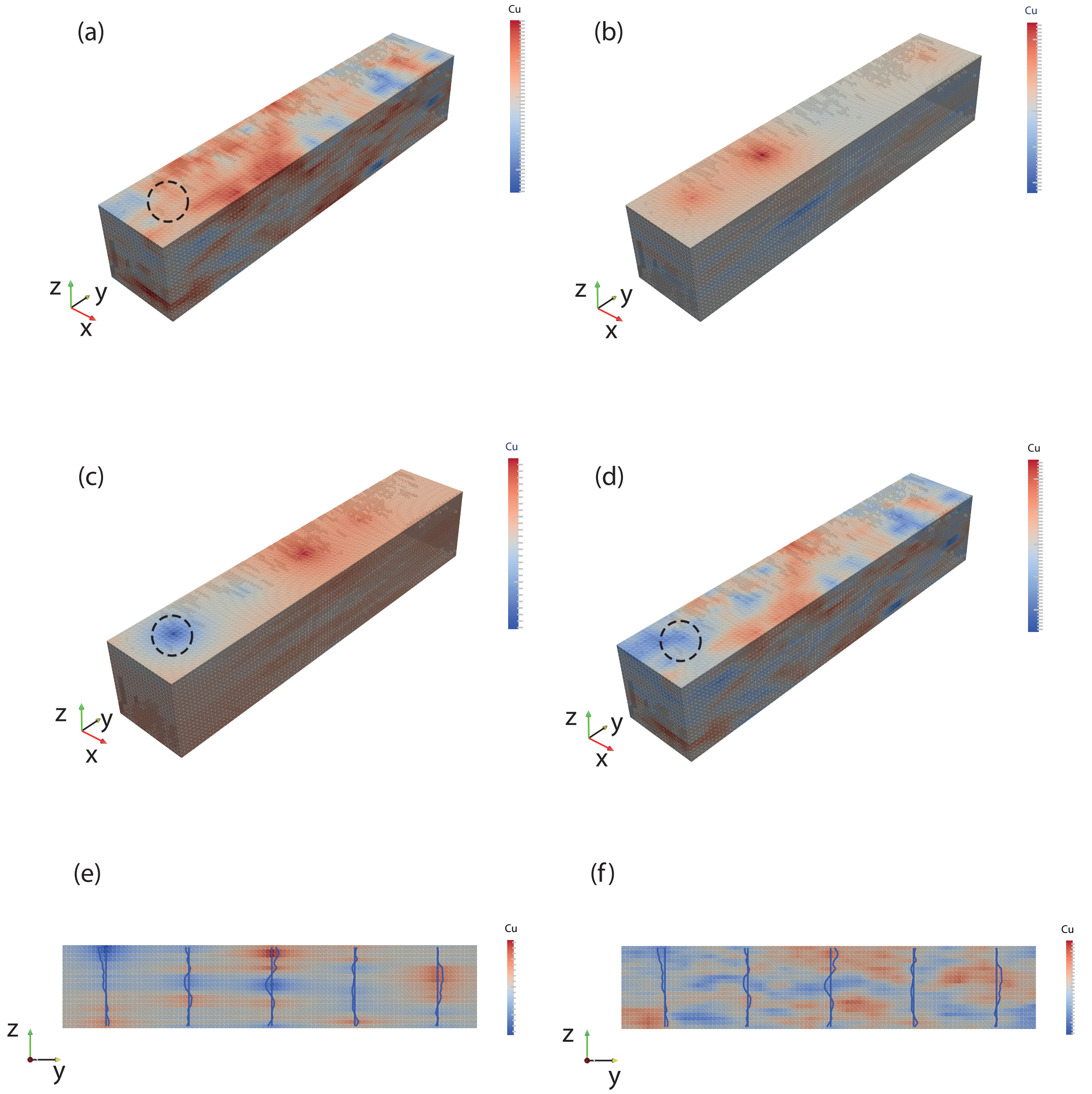


Figure 2: Example illustrations of the unconditional random field (a), the Kriged field based on the randomly simulated data (b), the Kriged field based on the CPT data (c), the conditional random field (d), cross-sections (e and f) in the longitudinal direction taken from the Kriged field (c) and from the conditional random field (d), respectively. Dashed circle indicates the position of the first CPT in subfigures (a) and (c-d)

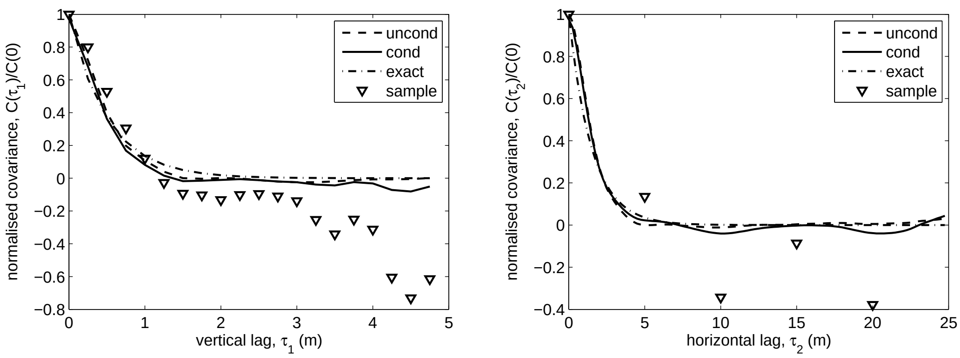


Figure 3: Vertical and horizontal covariance functions averaged over 200 realisations ($\theta_v = 1.0$ m, $\theta_h = 3.0$ m)

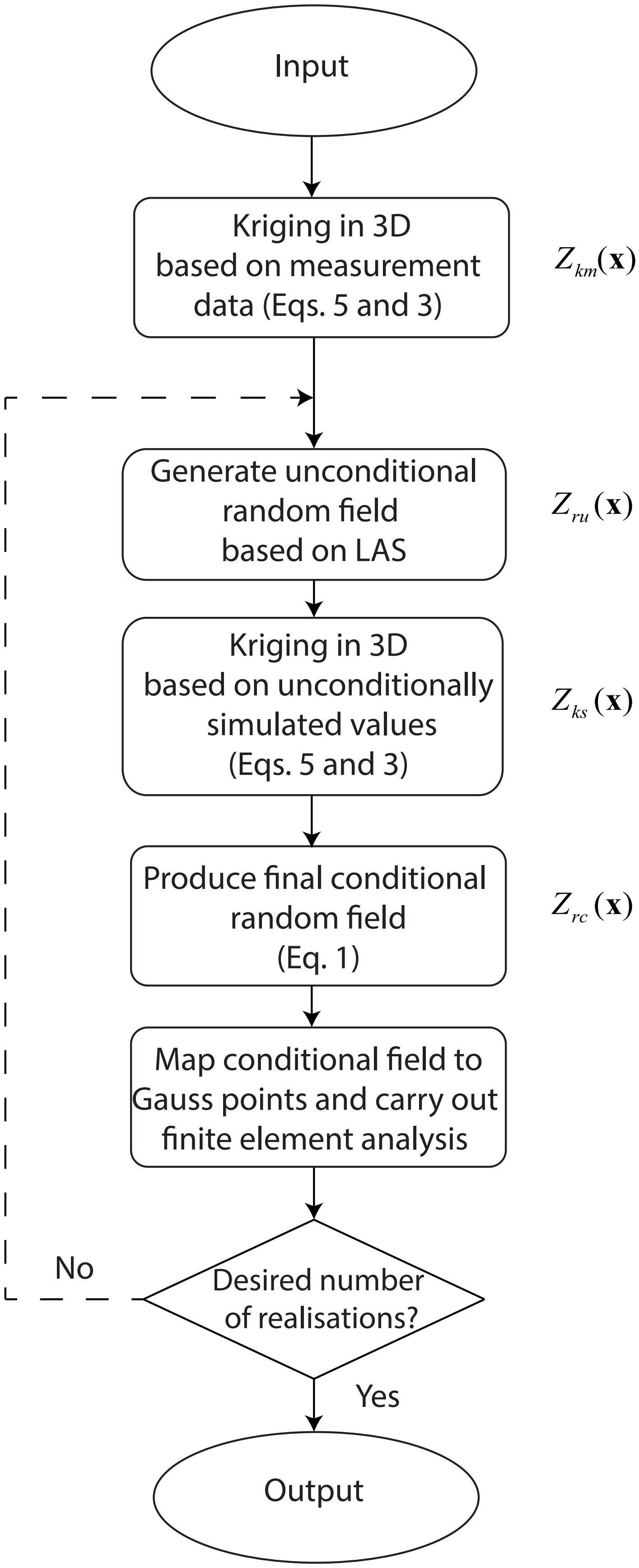


Figure 4: Flowchart for conditional RFEM simulation

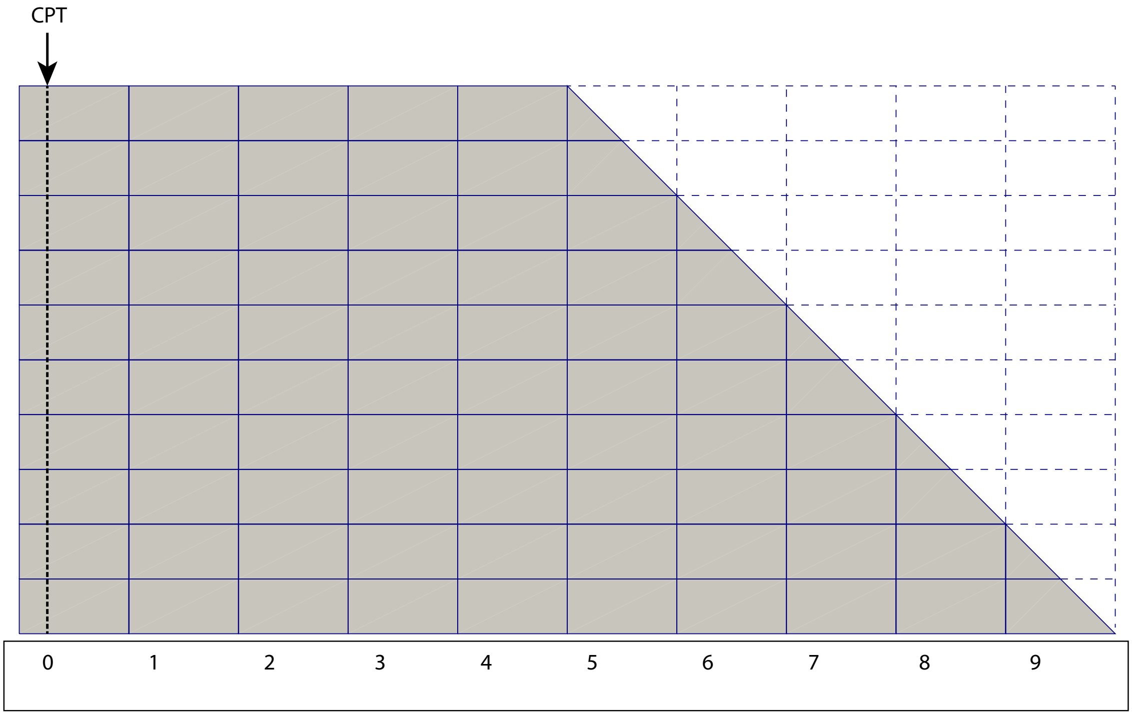
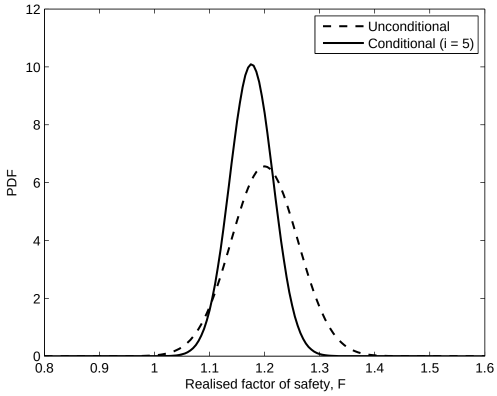
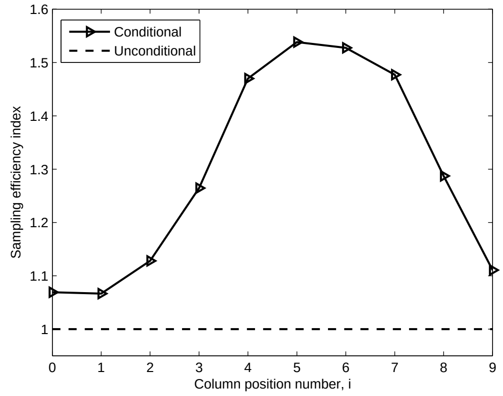


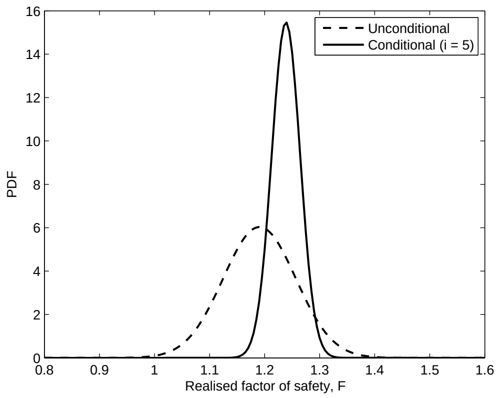
Figure 5: Finite element mesh and possible numbered CPT locations at a cross-section through the proposed 50 m long slope (dashed lines indicate the excavated soil mass and numbers correspond to Gauss point locations within the finite elements)



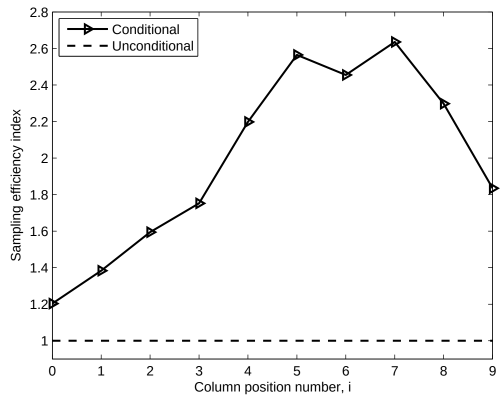
(a) Probability density functions of realised factor of safety ($\xi = 6$)



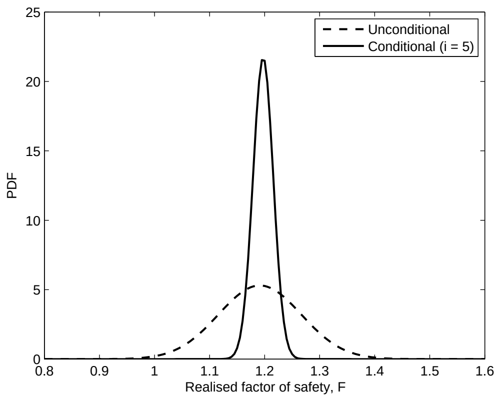
(b) Influence of CPT location ($\xi = 6$)



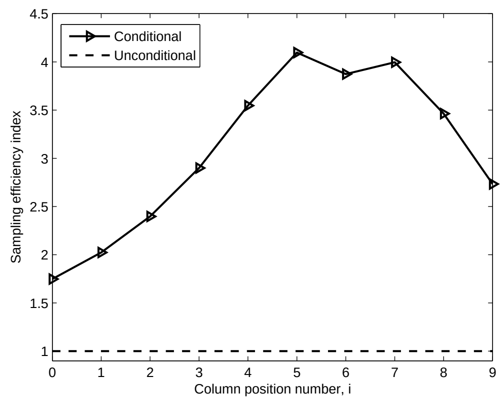
(c) Probability density functions of realised factor of safety ($\xi = 12$)



(d) Influence of CPT location ($\xi = 12$)



(e) Probability density functions of realised factor of safety ($\xi = 24$)



(f) Influence of CPT location ($\xi = 24$)

Figure 6: Simulation results for Example 1 (based on $\theta_v = 1.0$ m and 500 realistions per simulation)

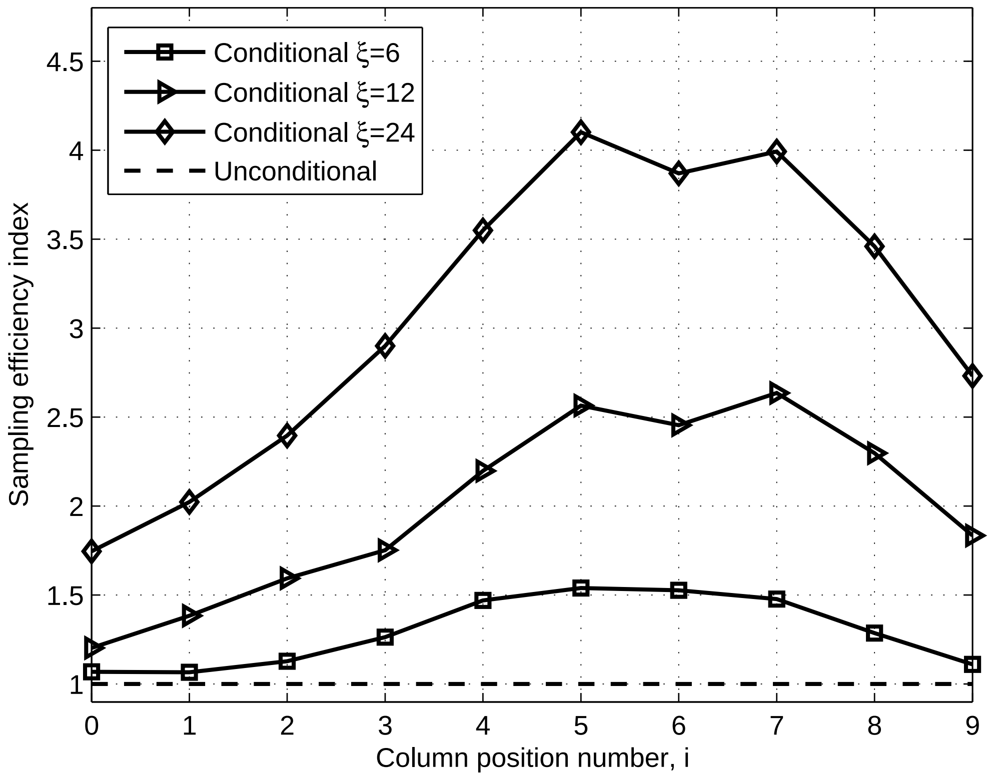


Figure 7: Sampling efficiency indices for various values of ξ

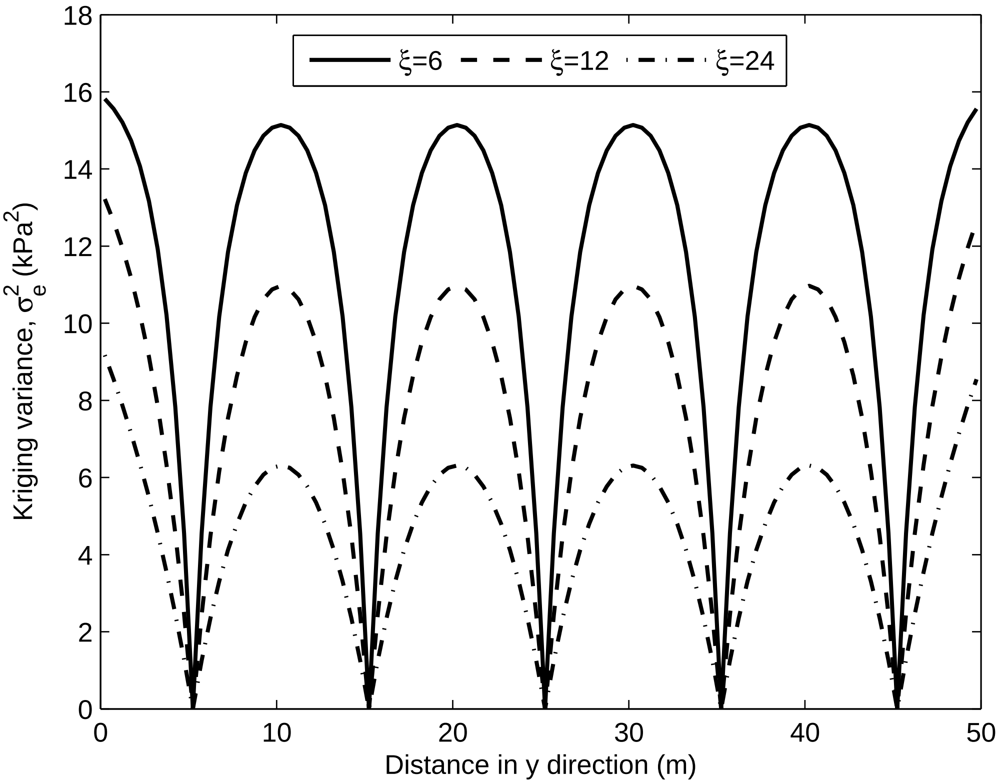
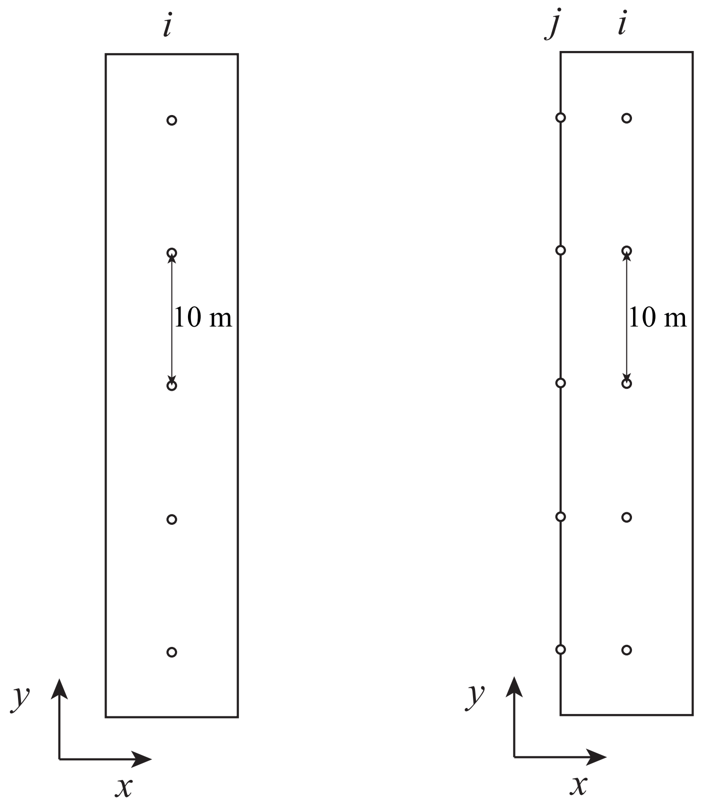
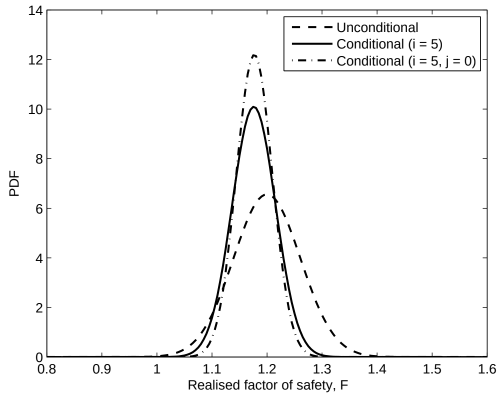


Figure 8: Kriging variance for various values of ξ (y–z slice at $i = 5$)

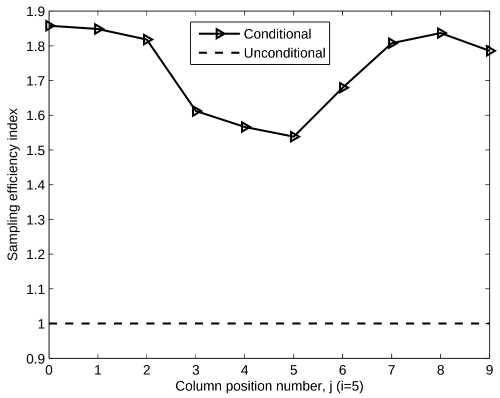


(a) *i=5* (b) *i=5, j=0*

Figure 9: CPT layout illustration (plan view) for a single row (a) and two rows (b)



(a) Probability density functions of realised factor of safety ($\xi = 6$)



(b) Influence of CPT location j with $i = 5$ ($\xi = 6$)

Figure 10: Influence of CPT location j during second phase of site investigation (based on $\theta_v = 1.0$ m and 500 realistions per simulation)

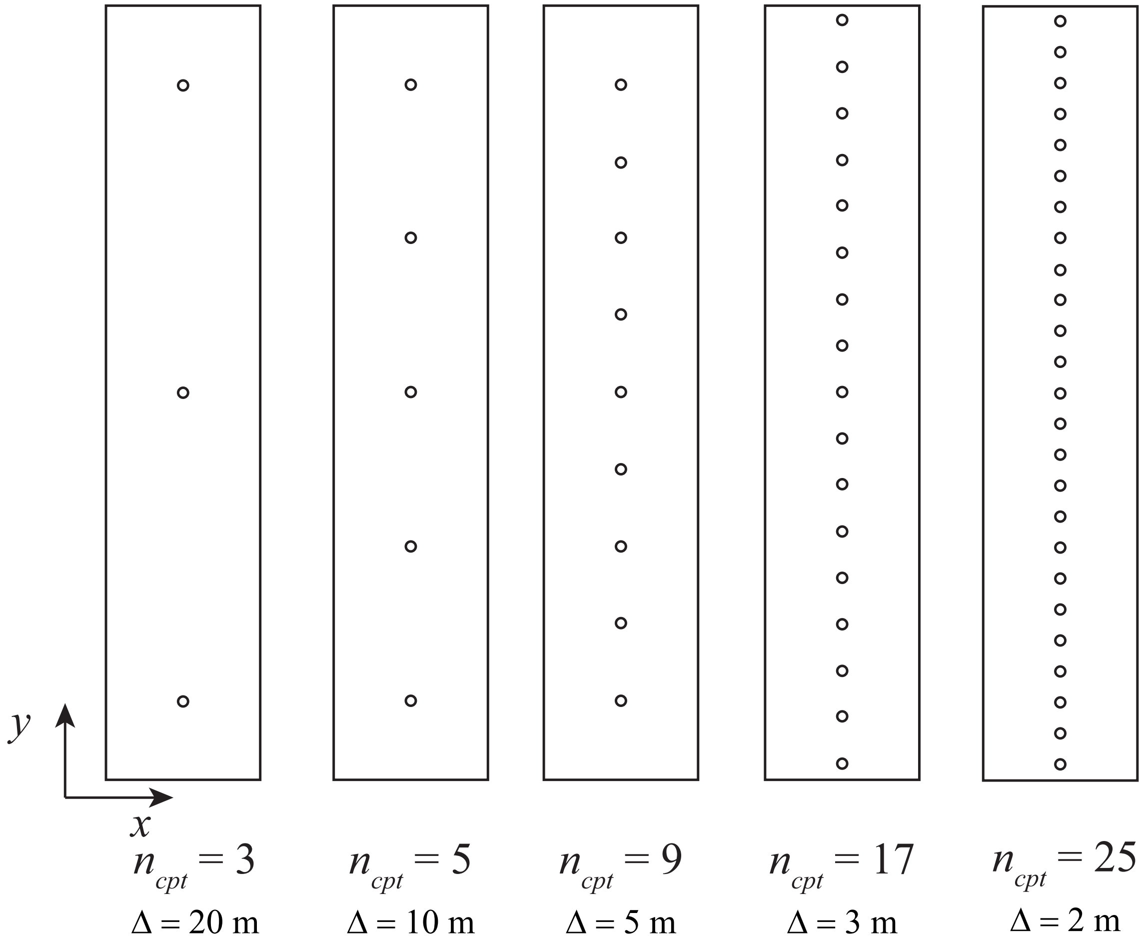
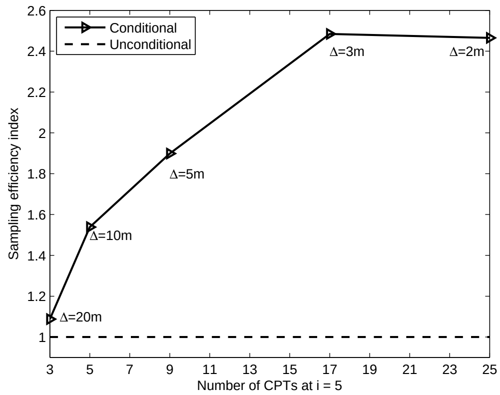
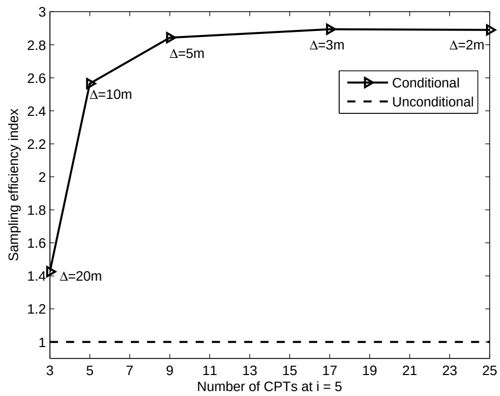


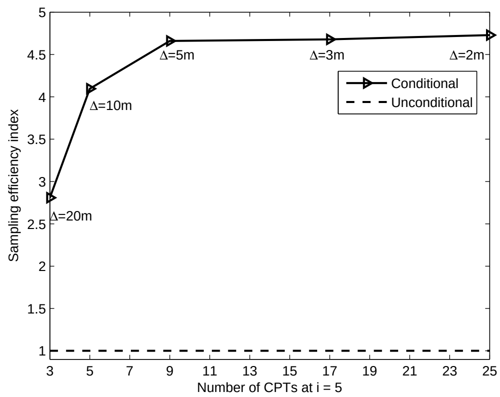
Figure 11: CPT layouts (plan views) for various numbers of boreholes ($n_{cpt} = 3, 5, 9, 17, 25$ and Δ denotes the distance between CPTs)



(a) $\xi = 6$ ($\theta_h = 6$ m)



(b) $\xi = 12$ ($\theta_h = 12$ m)



(c) $\xi = 24$ ($\theta_h = 24$ m)

Figure 12: Influence of number of CPTs (at $i = 5$) on sampling efficiency for various values of ξ and $\theta_v = 1.0$ m (Δ denotes the distance between CPTs)

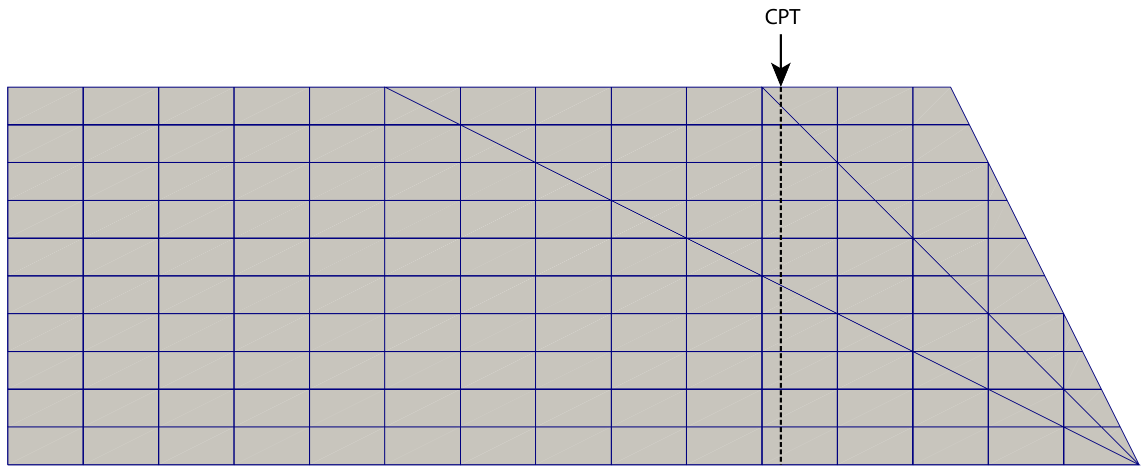


Figure 13: Finite element meshes for different slope geometries

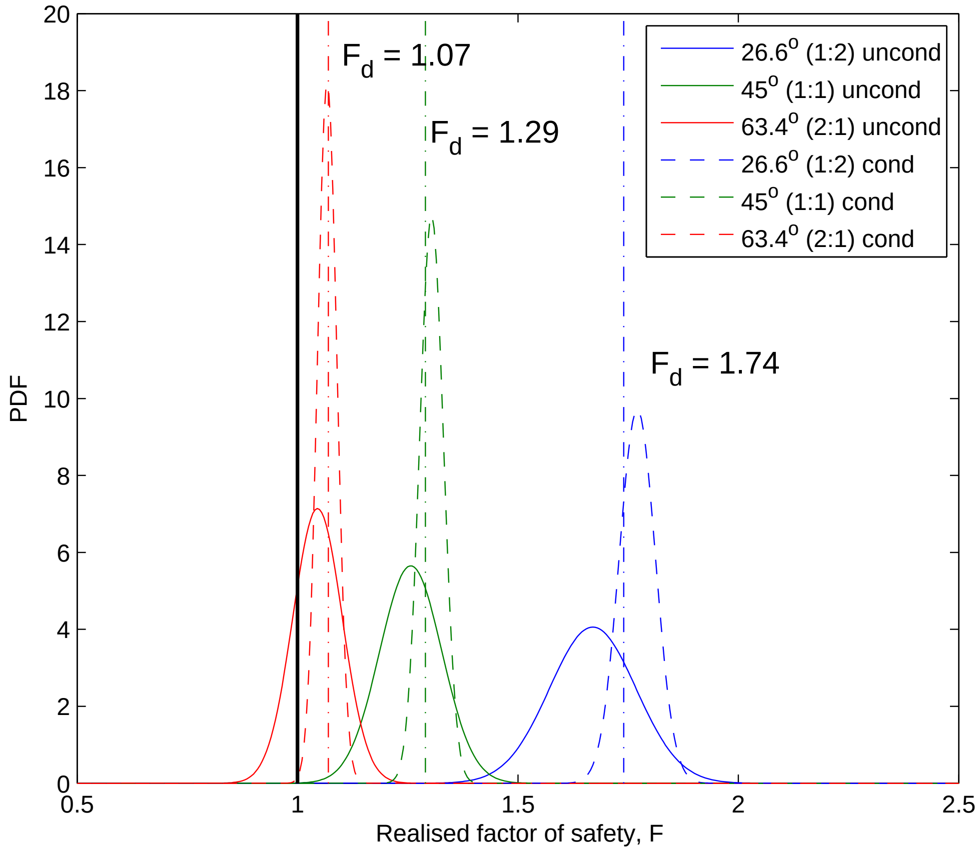


Figure 14: Probability density functions of realised factor of safety for three slopes, based on conditional and unconditional simulations

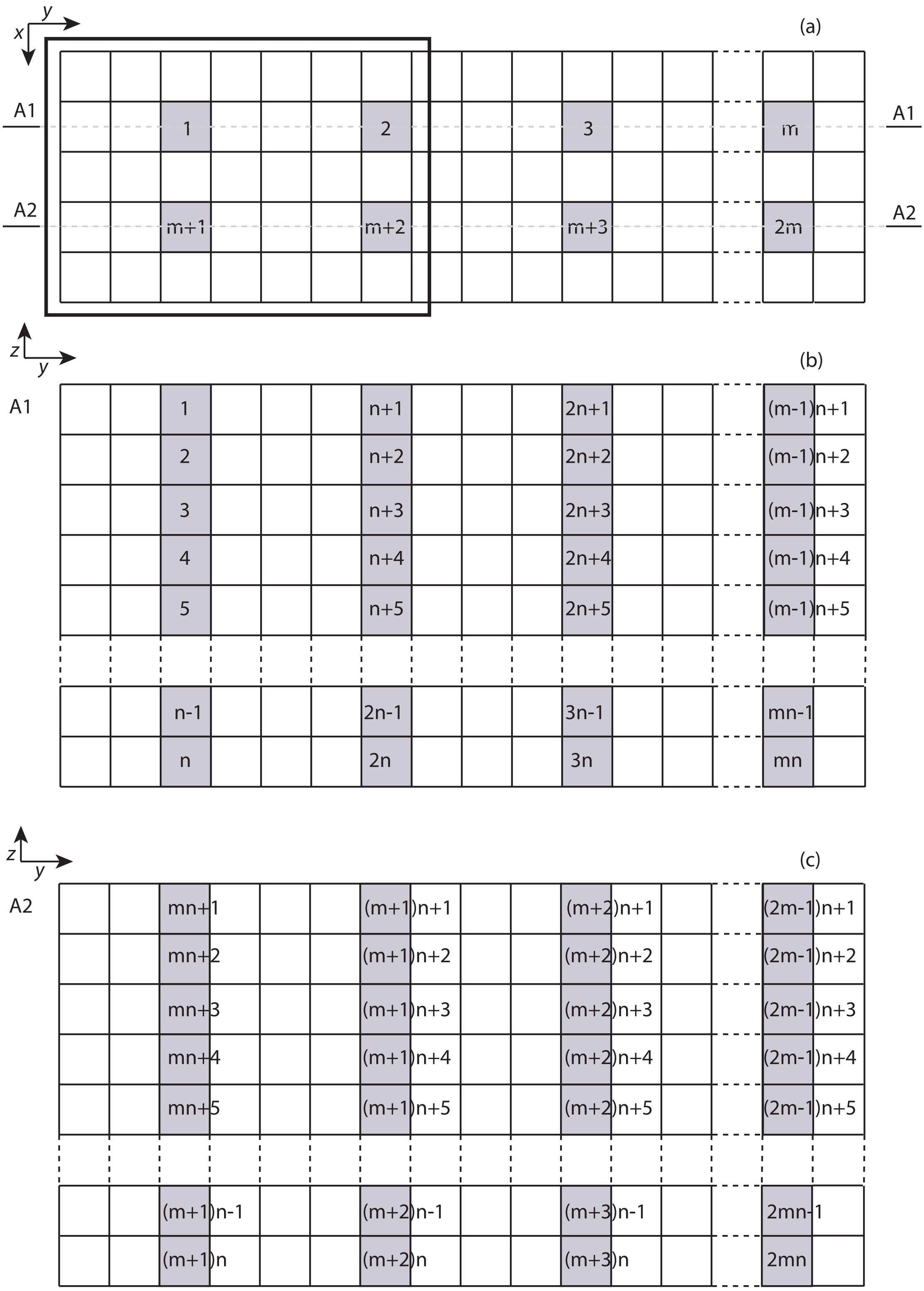


Figure A.1: Example CPT data grid ($k = 2$): (a) plan view showing CPT locations; (b) global numbering of data points at section A1; (c) global numbering of data points at section A2