

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](https://doi.org/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>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

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

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

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

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

7 **Cite as:**

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

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

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

13 **Abstract**

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

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

23 1. Introduction

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

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

41 In geotechnical engineering, a substantial amount of numerical work has been done using idealised 2D
42 simulations based on collected in-situ data (e.g. [4]), although a 3D simulation would be preferable
43 due to site data generally being collected from a 3D space. However, there are relatively few studies
44 simulating the effect of 3D heterogeneity due to the high computational requirements. Examples
45 include the effect of heterogeneity on shallow foundation settlement [11–13], on steady state seepage
46 [14–16], on seismic liquefaction [17] and on slope reliability [18–27].

47 The above investigations all used random fields to represent the soil spatial variability and the finite
48 element method to analyse geotechnical performance within a Monte Carlo framework, a form of

49 analysis sometimes referred to as the random finite element method (RFEM) [28]. However, they did
50 not make use of the spatial distribution of related measurement data to constrain the random fields. In
51 other words, for those applications that are based on real field data, many realisations not complying
52 with the field data at the measurement locations will be included in the simulation, which, in turn, will
53 result in an exaggerated range of responses in the analysis of geotechnical performance.

54 Studies on conditional simulations are available in geostatistics in the field of reservoir engineering
55 [29]. However, there are not many studies dealing with soil spatial variability in geotechnical
56 engineering that utilise conditional simulation (some 2D exceptions include, e.g., [6, 30–32]). This is
57 partly due to the smaller amount of data generally available in geotechnical engineering, and partly
58 due to there often not being a computer program specially implemented for those situations where
59 there are sufficient data (e.g. CPT, vane shear test (VST)), especially in 3D. However, unconditional
60 random fields can easily be conditioned to the known measurements by Kriging [29, 33]. Hence,
61 following the previous 2D work of Van den Eijnden and Hicks [31] and Lloret-Cabot et al. [30], this
62 paper seeks to implement and apply conditional simulation in three dimensional space, in order to
63 reduce uncertainty in the field where CPT measurements are carried out.

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

72 The implemented approach has been applied to two idealised slope stability examples. The first
73 demonstrates how the approach may be used to identify the best locations to conduct borehole testing,
74 and thereby allow an increased confidence in a project's success or failure to be obtained. While it is

75 very important to pay sufficient attention to the required intensity of a site investigation (i.e. the
76 optimal number of boreholes) with respect to the site-specific spatial variability, as highlighted by
77 Jaksa et al. [12], the first example starts by focusing on the optimum locations for carrying out site
78 investigations for a given number of boreholes, before moving on to consider the intensity of testing.
79 The second example compares different candidate slope designs, in order to choose the best (most
80 cost-effective) design satisfying the reliability requirements.

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

86 2. Theory and Implementation

87 2.1 Conditioning

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

$$90 \quad Z_{rc}(\mathbf{x}) = Z_{ru}(\mathbf{x}) + (Z_{km}(\mathbf{x}) - Z_{ks}(\mathbf{x})) \quad (1)$$

91 where \mathbf{x} denotes a location in space, $Z_{rc}(\mathbf{x})$ is the conditionally simulated random field, $Z_{ru}(\mathbf{x})$ is the
92 unconditional random field, $Z_{km}(\mathbf{x})$ is the Kriged field based on measured values at $\mathbf{x}_i (i = 1, 2, \dots, N)$,
93 $Z_{ks}(\mathbf{x})$ is the Kriged field based on unconditionally (or randomly) simulated values at the same
94 positions $\mathbf{x}_i (i = 1, 2, \dots, N)$, and N is the number of measurement locations.

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

99 [41], which has found extensive usage in geostatistics [42–43]. The LAS and Kriging methods are
100 briefly reviewed in the following sections.

101 *2.2 Anisotropic random field generation using 3D LAS*

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

110 The following covariance function is used in the subdivision process:

$$111 \quad C(\boldsymbol{\tau}) = C(\tau_1, \tau_2, \tau_3) = \sigma^2 \exp \left(-\frac{2|\tau_1|}{\theta_1} - \sqrt{\left(\frac{2\tau_2}{\theta_2}\right)^2 + \left(\frac{2\tau_3}{\theta_3}\right)^2} \right) \quad (2)$$

112 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
113 τ_3 are the respective scales of fluctuation and lag distances in the vertical and two lateral coordinate
114 directions, respectively. Herein, an isotropic random field is initially generated by setting $\theta_1 = \theta_2 = \theta_3$
115 $= \theta_{iso}$; i.e. so that θ_{iso} equals the horizontal scale of fluctuation, θ_h . This field is then squashed in the
116 vertical direction to give the target vertical scale of fluctuation, θ_v . The 3D LAS implementation of
117 Spencer [25] has been used in this paper, and the reader is referred to Spencer [25] and Hicks and
118 Spencer [19] for more details. Note also that a truncated normal distribution has been used to describe
119 the pointwise variation in material properties [19].

120 *2.3 Kriging*

121 In contrast to conventional deterministic interpolation techniques, such as moving least squares and
122 the radial point interpolation method, Kriging incorporates the variogram (or covariance) into the

123 interpolation procedure; specifically, information on the spatial correlation of the measured points is
 124 used to calculate the weights. Moreover, standard errors of the estimation can also be obtained,
 125 indicating the reliability of the estimation and the accuracy of the prediction. Kriging is a method of
 126 interpolation for which the interpolated values are modelled by a Gaussian process governed by prior
 127 covariances and for which confidence intervals can be derived. While interpolation methods based on
 128 other criteria need not yield the most likely intermediate values, Kriging provides a best linear
 129 unbiased prediction of the soil properties (Z) between known data [43, 46] by assuming the stationarity
 130 of the mean and of the spatial covariances, or variograms. A brief review is first given to facilitate
 131 understanding of the implementation.

132 Suppose that Z_1, Z_2, \dots, Z_N are observations of the random field $Z(\mathbf{x})$ at points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ (i.e.
 133 $Z_i = Z(\mathbf{x}_i)$ ($i = 1, 2, \dots, N$)). The best linear unbiased estimation (i.e. \hat{Z}) of the soil property at some
 134 location \mathbf{x}_0 is given by

$$135 \quad \hat{Z}(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i Z_i = \sum_{i=1}^N \lambda_i(\mathbf{x}_0) Z(\mathbf{x}_i) \quad (3)$$

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

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

$$140 \quad \begin{aligned} \sigma_e^2 &= \text{var}(\hat{Z} - Z) = E\left[(\hat{Z} - Z)^2\right] = \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j (c_{ij} - \sigma^2) - 2 \sum_{i=1}^N \lambda_i (c_{i0} - \sigma^2) + (c_{00} - \sigma^2) \\ &= \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \left(C(|\mathbf{x}_i - \mathbf{x}_j|) - \sigma^2\right) - 2 \sum_{i=1}^N \lambda_i \left(C(|\mathbf{x}_i - \mathbf{x}_0|) - \sigma^2\right) + \left(C(|\mathbf{x}_0 - \mathbf{x}_0|) - \sigma^2\right) \\ &= - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(|\mathbf{x}_i - \mathbf{x}_j|) + 2 \sum_{i=1}^N \lambda_i \gamma(|\mathbf{x}_i - \mathbf{x}_0|) - \gamma(|\mathbf{x}_0 - \mathbf{x}_0|) \end{aligned} \quad (4)$$

141 where $\text{var}()$ denotes the variance operator and $E[\cdot]$ is the expectation operator, $c_{ij} = C(|\mathbf{x}_i - \mathbf{x}_j|)$ is the
142 covariance between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$, $c_{i0} = C(|\mathbf{x}_i - \mathbf{x}_0|)$ is the covariance between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_0)$,
143 and $c_{00} = C(|\mathbf{x}_0 - \mathbf{x}_0|) = C(0) = \sigma^2$ is the variance of $Z(\mathbf{x}_0)$, which is estimated at the target location
144 \mathbf{x}_0 . The rearrangement in equation (4) makes use of the relationship between a variogram $\gamma(\boldsymbol{\tau})$ and a
145 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
146 order to ensure that the estimator is unbiased, i.e. $E(\hat{Z} - Z) = 0$, the weights must sum to one).

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

$$149 \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_N - \mathbf{x}_N) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_N - \mathbf{x}_0) \\ 1 \end{pmatrix} \quad (5)$$

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

152 Equation (5) may be expressed as

$$153 \boldsymbol{\gamma}_{lhs} \boldsymbol{\lambda}_x = \boldsymbol{\gamma}_{rhs} \quad (6)$$

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

$$155 \sigma_e^2 = \mu + \sum_{i=1}^N \lambda_i \gamma(|\mathbf{x}_i - \mathbf{x}_0|) = (\boldsymbol{\gamma}_{lhs}^{-1} \boldsymbol{\gamma}_{rhs})^T \boldsymbol{\gamma}_{rhs} \quad (7)$$

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

157 Note that the left-hand-side matrix $\boldsymbol{\gamma}_{lhs}$ is a function of only the observation point locations and
158 covariance between them. Therefore, it only needs to be inverted once, and then equations (5) and (3)

159 used repeatedly for building up the field of best estimates at different locations in space. In contrast,
160 the right-hand-side vector γ_{rhs} changes as a function of the spatial point \mathbf{x}_0 , resulting in different
161 weight vectors λ_x that are used in equation (3) to get the estimates (point by point) in the domain of
162 interest.

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

170 *2.4 Computational efficiency*

171 There are two aspects involved in the computational efficiency of the above Kriging implementation.
172 One is the total number of equations, which depends on the total number of data points ($N = k \times m \times n$,
173 where k and m are the number of CPT rows in the x and y directions respectively, and n is the number
174 of data points for each CPT profile, see Figure 1) contributing to the left-hand-side matrix; the other is
175 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
176 three Cartesian directions) that need to be Kriged (i.e. how many times the algorithm will need to be
177 repeated, except for inverting the left-hand-side matrix). The higher the required field resolution (n_f)
178 and the greater the total number of known data points (N), the longer the Kriging will take. In the case
179 of the CPT arrangement in Fig. 1, the size of matrix γ_{lhs} (see equation (5) or equation (A1)) and the
180 size of vector γ_{rhs} (see equation (5) or equation (A3)) in 3D are m^2 and m times larger than those in 2D
181 (i.e. a cross-section in the x - z plane) respectively. The time it takes to Krige a full 3D field depends on
182 the processing time of each individual step and the number of times each step has to be performed. To
183 Krige a field of size n_f , conditional to N measurement points, the total time may be approximated by

$$184 \quad t(N, n_f) \approx c_1 n_f N^2 + c_2 N^3 \quad (8)$$

185 where the first term represents the time needed for solving the system of equations for all field points
186 (i.e. n_f times) ($O(N^2)$) and the second represents the time needed for inverting the matrix γ_{lhs} (i.e. only
187 once) ($O(N^3)$). The constants c_1 and c_2 are functions of the CPU speed and the operation, and in this
188 case are in a ratio of approximately 4:1. Additionally, in all practical cases, $n_f \gg N$, so that the
189 calculation time depends mainly on the first term in the above equation; that is, on the number of times
190 (i.e. n_f times) that the matrix–vector multiplication operation, $\lambda_x = \gamma_{lhs}^{-1} \gamma_{rhs}$, needs to be performed.
191 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
192 2D. In the examples reported in Section 4, all problems investigated are very long in the third
193 dimension compared to the cross-section. Therefore, the time consumed in a 3D analysis is
194 theoretically $n_y \times m^2$ times that of a 2D analysis, when neglecting the relatively fast, one-off matrix
195 inversion operation and other computation overheads, such as reading/writing and memory operations.
196 However, despite the significantly greater run-time requirements for Kriging in 3D (as compared to
197 2D), it is still far less than the time consumed in a nonlinear finite element analysis where plasticity
198 iterations are needed. For Example 1 in Section 4, where $n_x = 20$, $n_y = 100$ and $n_z = 20$, it took, in serial
199 and on average, 134 hours in total for the 500 finite element analyses forming each Monte Carlo
200 simulation (3.0 GHz CPU), whereas Kriging 500 times took about 2.4 hours. In contrast, 500 Kriging
201 interpolations for a 2D cross-section analysis took approximately 8.5 seconds. It is noted that the
202 computation time used for Kriging 500 times is significant in comparison with a single finite element
203 analysis, and therefore should not be considered a pre-processing step if utilising parallel computation
204 for the finite element analyses. Therefore, the computing strategy developed to carry out the analyses
205 for Examples 1 and 2 in Section 4 (comprising around 30,000 realisations in total, and involving
206 30,000 3D Kriging interpolations) was to run the analyses in parallel (each Kriging and finite element
207 analysis serially on a single computation node) on the Dutch national grid e-infrastructure with high
208 performance computing clusters.

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

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

$$229 \quad u = \frac{\sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \sum_{k=1}^{n_z} \sigma_e(i, j, k)}{n_x n_y n_z \sigma} \quad (9)$$

230 The approximation error may be evaluated by

$$231 \quad E_u = \frac{|u_n - u_a|}{u_a} \quad (10)$$

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

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

237 3. Validation

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

243 The block is discretised into $20 \times 100 \times 20$ cubic cells, with each cell of dimension 0.25 m. The mean of
244 the undrained shear strength is 40 kPa, and the standard deviation is 8 kPa. The degree of anisotropy
245 of the heterogeneity is $\xi = 3$, in which $\xi = \theta_h/\theta_v$, and $\theta_v = 1.0$ m. Five CPT measurement locations in
246 the y direction (at $x = 2.5$ m) are available, each comprising $n = 20$ data points at 0.25 m spacing in the
247 vertical direction. These ‘measured’ data have been obtained by sampling from a single independent
248 realisation of the spatial variability (i.e. representing the ‘actual’ in-situ variability). The interval
249 distance between the CPTs in the horizontal direction is $\Delta_y = 5$ m, and the first CPT is located at $y =$
250 2.5 m.

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

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

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

$$264 \quad \hat{C}(\tau_j = j\Delta\tau) = \frac{1}{n-j} \sum_{i=1}^{n-j} (Z_i - \hat{\mu}_Z)(Z_{i+j} - \hat{\mu}_Z) \quad (11)$$

265 where $j = 0, 1, \dots, n-1$, n is the number of data points in the vertical or horizontal direction, τ_j is the
 266 lag distance between x_i and x_{i+j} , $\Delta\tau$ is the distance between two adjacent cells vertically or
 267 horizontally, $\hat{\mu}_Z$ is the estimated mean, Z is the random soil property and Z_i is the sample of Z . The
 268 correlation function is then $\hat{\rho}(\tau_j) = \hat{C}(\tau_j) / \hat{C}(0)$, where $\hat{C}(0) = \hat{\sigma}_Z^2$ and $\hat{\sigma}_Z^2$ is the estimated variance
 269 [3].

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

278 4. Applications

279 Two simple examples concerning slope stability are presented in this section, to illustrate how the
 280 technique presented in this paper may be used as an aid to geotechnical design. The first involves
 281 finding the optimum locations for CPT profiles, in order to minimise the uncertainty in assessing the

282 reliability of a slope. The second involves a cost-effective design with regard to the slope angle when
283 field measurements have already been made (i.e. the positions where the CPT data were taken are
284 already known).

285 Both examples are presented in terms of the uncertainties in the slope response (with respect to factor
286 of safety). The factors of safety are calculated by 3D finite elements using the strength reduction
287 method [47], with the analyses being undertaken within a probabilistic (RFEM) framework; a
288 flowchart for carrying out such a simulation is shown in Figure 4. The undrained clay behaviour has
289 been modelled using a linear elastic, perfectly plastic Tresca soil model. The clay has a unit weight of
290 20 kN/m^3 , a Young's modulus of 100 MPa and a Poisson's ratio of 0.3 . With reference to Figures 5
291 and 13, the finite element boundary conditions are: a fixed base, rollers on the back of the domain
292 preventing displacements perpendicular to the back face, and rollers on the two ends of the domain,
293 allowing only settlements and preventing movements in the other two directions (i.e. the out-of-slope-
294 face and longitudinal directions). A full explanation of these boundary conditions is given in Spencer
295 [25] and Hicks and Spencer [19].

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

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

307 *4.1 Example 1*

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

317 Figure 5 shows a cross-section through the slope, and 10 possible positions to locate the CPTs ($i = 0,$
318 $1, \dots, 9$). Note that the CPTs are taken to be equally spaced (i.e. at 10 m centres) in the third
319 dimension, and that the first and fifth CPTs are located at 5 m and 45 m along the slope axis (see
320 Figure 9(a)). Furthermore, the CPTs are carried out before the slope is excavated, in a block of soil of
321 dimensions 10×50×5 m as indicated in the figure.

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

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

331
$$I_{se} = \frac{\sigma_u}{\sigma_i} \quad (12)$$

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

345 However, it should be remembered that Figures 6(a)-6(b) are for the case of $\xi = 6$ (corresponding to θ_h
346 = 6 m) and that ξ often takes a larger value in practice. Figures 6(c)-6(f) show that, for $\xi = 12$ and $\xi =$
347 24, the reduction in uncertainty relative to the unconditional case is greater. Moreover, improved
348 values of I_{se} are obtained for CPT locations near the right and left boundaries, due to the higher
349 correlation of soil properties in the horizontal direction.

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

357 near the crest for smaller values of ξ . For example, Figure 7 shows that the sampling efficiency index
358 for $\xi = 24$ at $i = 2$ is approximately the same as that for $\xi = 12$ at $i = 5, 6$ and 7 .

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

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

376 To further investigate the influence of CPT intensity on the uncertainty in the realised factor of safety,
377 conditional simulations involving different numbers of CPTs (and thereby different distances (Δ)
378 between adjacent CPTs) have been carried out for the case of $\xi = 6, 12$ and 24 . Figure 11 shows the
379 plan views of CPT layouts for $n_{cpt} = 3, 5, 9, 17$ and 25 (corresponding to CPT spacings of $\Delta = 20, 10, 5,$
380 3 and 2 m, respectively), with the locations of the CPTs in the x -direction being fixed at $i = 5$. Figure
381 12 shows the influence of CPT intensity on the sampling efficiency indices for the three values of ξ .
382 It is seen that there is only a marginal benefit in increasing the scope of the investigation by having

383 CPT spacings less than $\Delta \approx \theta_h/2$, especially for the $\xi = 6$ and $\xi = 12$ cases. For $\xi = 24$, the sampling
384 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
385 improve the sampling efficiency to a value of 4.5. However, the general finding from Figures 10(b)
386 and 12 is that the optimal sampling distance is around $\theta_h/2$ for the problem investigated, based on the
387 assumed correlation function.

388 *4.2 Example 2*

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

393 Figure 13 shows three possible slope angles, with the corresponding finite element mesh
394 discretisations. The slope is 5 m high and 50 m long in the third dimension, and the left-hand boundary
395 is taken to be 15 m from the slope toe. Five CPTs were taken along the length of the slope at 10 m
396 centres, at the location of the column of Gauss points nearest the slope crest for the 1:1 slope, as seen
397 in the figure. The clay soil has a mean undrained shear strength of 21 kPa, a coefficient of variation of
398 0.2, a vertical scale of fluctuation of 1 m and a horizontal scale of fluctuation of 12 m.

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

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

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

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

425 **5. Conclusions**

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

435 density distributions in the two numerical examples. Although Bayesian updating is not used in this
436 paper, the effect is similar.

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

447 **Acknowledgements**

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

452 **References**

- 453 [1] DeGroot DJ, Baecher GB. Estimating autocovariance of in-situ soil properties. ASCE Journal of
454 Geotechnical Engineering 1993; 119 (1): 147–66.
- 455 [2] Fenton GA. Estimation for stochastic soil models. ASCE J. Geotech. Geoenv. Eng. 1999; 125(6): 470–85.
- 456 [3] Fenton GA. Random field modeling of CPT data. ASCE J. Geotech. Geoenv. Eng. 1999; 125(6): 486–98.
- 457 [4] Hicks MA, Onisiphorou C. Stochastic evaluation of static liquefaction in a predominantly dilative sand fill.
458 Géotechnique 2005; 55(2): 123–33.
- 459 [5] Jaksa MB, Kaggwa WS, Brooker PI. Experimental evaluation of the scale of fluctuation of a stiff clay. In:
460 Proc. 8th Int. Conf. on Applications of Statistics and Probability in Civil Engineering, Sydney; 1999. p. 415–22.

- 461 [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.
- 462
- 463 [7] Lundberg AB, Li Y. Probabilistic characterization of a soft Scandinavian clay supporting a light quay
464 structure. In: *Proc. 5th International Symposium on Geotechnical Safety and Risk, Rotterdam; 2015. p.170–75.*
- 465 [8] Phoon KK, Kulhawy FH. Characterization of geotechnical variability. *Canadian Geotechnical Journal* 1999;
466 36 (4): 612–24.
- 467 [9] Uzielli M, Vannucchi G, Phoon KK. Random field characterisation of stress-normalised cone penetration
468 testing parameters. *Géotechnique* 2005; 55(1): 3–20.
- 469 [10] Zhao HF, Zhang LM, Xu Y, Chang D.S. Variability of geotechnical properties of a fresh landslide soil
470 deposit. *Engineering Geology* 2013; 166: 1–10.
- 471 [11] Fenton GA, Griffiths DV. Three-dimensional probabilistic foundation settlement. *ASCE J. Geotech.*
472 *Geoenv. Eng.* 2005; 131(2): 232–39.
- 473 [12] Jaksa MB, Goldsworthy JS, Fenton GA, Kagawa WS, Griffiths DV, Kuo YL, Poulos HG. Towards reliable
474 and effective site investigations. *Géotechnique* 2005; 55(2): 109–21.
- 475 [13] Kuo YL, Jaksa MB, Kagawa WS, Fenton GA, Griffiths DV, Goldsworthy JS. Probabilistic analysis of
476 multi-layered soil effects on shallow foundation settlement. In: *Proc. 9th Australia New Zealand Conference on*
477 *Geomechanics, Auckland, New Zealand; 2004. p. 541–47.*
- 478 [14] Griffiths DV, Fenton GA. Observations on two- and three-dimensional seepage through a spatially random
479 soil. In: *Proc. 7th Int. Conf. on Applications of Statistics and Probability in Civil Engineering, Paris, France;*
480 *1995. p. 65–70.*
- 481 [15] Griffiths DV, Fenton GA. Three-dimensional seepage through spatially random soil. *ASCE J. Geotech.*
482 *Geoenv. Eng.* 1997; 123(2): 153–60.
- 483 [16] Griffiths DV, Fenton GA. Probabilistic analysis of exit gradients due to steady seepage. *ASCE J. Geotech.*
484 *Geoenv. Eng.* 1998; 124(9): 789–97.
- 485 [17] Popescu R, Prevost JH, Deodatis G. 3D effects in seismic liquefaction of stochastically variable soil
486 deposits. *Géotechnique* 2005; 55(1): 21–31.
- 487 [18] Griffiths DV, Huang J, Fenton GA. On the reliability of earth slopes in three dimensions. *Proc. R. Soc. A*
488 2009; 465(2110): 3145–64.
- 489 [19] Hicks MA, Spencer WA. Influence of heterogeneity on the reliability and failure of a long 3D slope.
490 *Comput Geotech* 2010; 37: 948–55.

491 [20] Hicks MA, Chen J, Spencer WA. Influence of spatial variability on 3D slope failures. In: Proc. 6th Int. Conf.
492 Computer Simulation Risk Analysis and Hazard Mitigation, Kefalonia; 2008. p. 335–42.

493 [21] Hicks MA, Nuttall JD, Chen J. Influence of heterogeneity on 3D slope reliability and failure consequence.
494 Computers and Geotechnics 2014; 61: 198–208.

495 [22] Li Y, Hicks MA. Comparative study of embankment reliability in three dimensions, In: Proc. 8th European
496 Conference on Numerical Methods in Geotechnical Engineering (NUMGE), Delft; 2014. p. 467–72.

497 [23] Li Y, Hicks MA, Nuttall JD. Probabilistic analysis of a benchmark problem for slope stability in 3D. In:
498 Proc. 3rd Int. Symp. Computational Geomech, Krakow, Poland; 2013. p. 641–8.

499 [24] Li YJ, Hicks MA, Nuttall, JD. Comparative analyses of slope reliability in 3D. Engineering Geology
500 2015; 196: 12–23.

501 [25] Spencer WA. Parallel stochastic and finite element modeling of clay slope stability in 3D. PhD thesis,
502 University of Manchester, UK; 2007.

503 [26] Spencer WA, Hicks MA. 3D stochastic modelling of long soil slopes, In: Proc. 14th Conf. of Assoc. for
504 Computational Mechanics in Engineering, Belfast; 2006. p. 119–22.

505 [27] Spencer WA, Hicks MA. A 3D finite element study of slope reliability. In: Proc. 10th Int. Symp. Num
506 Models in Geomech, Rhodes; 2007. p. 539–43.

507 [28] Fenton GA, Griffiths DV. Risk assessment in geotechnical engineering. New York: John Wiley & Sons;
508 2008.

509 [29] Chiles JP, Delfiner P. Geostatistics: modeling spatial uncertainty. John Wiley & Sons; 2009.

510 [30] Lloret-Cabot M, Hicks MA, van den Eijnden AP. Investigation of the reduction in uncertainty due to soil
511 variability when conditioning a random field using Kriging. Géotechnique letters 2012; 2: 123–7.

512 [31] van den Eijnden AP, Hicks MA. Conditional simulation for characterizing the spatial variability of sand
513 state. In: Proc. 2nd International Symposium on Computational Geomechanics, Dubrovnik, Rhodes, Greece;
514 2011. p. 288–96.

515 [32] Vanmarcke EH, Fenton GA. Conditioned simulation of local fields of earthquake ground motion. Structural
516 Safety 1991; 10(1): 247–64.

517 [33] Journel, AG. Geostatistics for conditional simulation of ore bodies. Economic Geology 1974; 69(5): 673–87.

518 [34] Olea, RA. Systematic sampling of spatial functions. Series of Spatial Analysis No. 7, Kansas Geological
519 Survey, Lawrence, Kans; 1984.

520 [35] Frimpong, S, Achireko PK. Conditional LAS stochastic simulation of regionalized variables in random
521 fields. *Computational Geosciences* 1998; 2: 37–45.

522 [36] Journel AG, Huijbregts CJ. *Mining Geostatistics*. New York: Academic Press; 1978.

523 [37] Fenton GA. Error evaluation of three random field generators. *ASCE J. Eng. Mech.* 1994; 120 (12): 2478–
524 97.

525 [38] Ji J, Liao HJ, Low BK. Modeling 2D spatial variation in slope reliability analysis using interpolated
526 autocorrelations. *Computers and Geotechnics* 2012; 40: 135–46.

527 [39] Phoon KK, Huang SP, Quek ST. Simulation of second-order processes using Karhunen–Loeve expansion.
528 *Computers & Structures* 2002; 80(12): 1049–60.

529 [40] Fenton GA, Vanmarcke EH. Simulation of random fields via local average subdivision. *ASCE Journal of*
530 *Engineering Mechanics* 1990; 116(8): 1733–49.

531 [41] Krige DG. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of*
532 *the Chemical, Metallurgical and Mining Society of South Africa* 1951; 52 (6): 119–39.

533 [42] Cressie N. The origins of Kriging. *Mathematical Geology* 1990; 22 (3): 239–52.

534 [43] Wackernagel H. *Multivariate geostatistics: An introduction with applications*. Germany: Springer; 2003.

535 [44] Fenton GA. *Simulation and analysis of random fields*. PhD Thesis, Princeton University; 1990.

536 [45] Vanmarcke EH. *Random fields: analysis and synthesis*. Cambridge, Massachusetts: The MIT Press; 1983.

537 [46] Fenton GA. Data analysis/geostatistics. In: *Probabilistic methods in geotechnical engineering*. Griffiths DV,
538 Fenton GA, editors. New York: Springer; 2007. p. 51–73.

539 [47] Smith IM, Griffiths DV, Margetts L. *Programming the finite element method*, 5th ed. New York: John
540 Wiley & Sons; 2013.

541 [48] Hicks MA, Samy K. Influence of heterogeneity on undrained clay slope stability. *Quarterly Journal of*
542 *Engineering Geology and Hydrogeology* 2002; 35: 41–9.

543 [49] Phoon KK, Kulhawy FH. Evaluation of geotechnical property variability. *Canadian Geotechnical Journal*
544 1999; 36(4): 625–39.

545 [50] European Committee for Standardisation (CEN). Eurocode 7: geotechnical design. Part 1: general rules. EN
546 1997-1, CEN; 2004.

547 [51] Hicks MA, Nuttall JD. Influence of soil heterogeneity on geotechnical performance and uncertainty: a
548 stochastic view on EC7. In: *Proc. 10th International Probabilistic Workshop, Stuttgart*; 2012. p. 215–27.

549

550
551
552
553
554

555 Appendix

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

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

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

$$\mathbf{Y}_{lhs} = \begin{pmatrix}
 \mathbf{V}_{1,1} & \mathbf{V}_{1,2} & \mathbf{V}_{1,3} & \cdots & \mathbf{V}_{1,m} & \mathbf{V}_{1,m+1} & \mathbf{V}_{1,m+2} & \mathbf{V}_{1,m+3} & \cdots & \mathbf{V}_{1,2m} & \cdots \\
 \mathbf{V}_{2,1} & \mathbf{V}_{2,2} & \mathbf{V}_{2,3} & \cdots & \mathbf{V}_{2,m} & \mathbf{V}_{2,m+1} & \mathbf{V}_{2,m+2} & \mathbf{V}_{2,m+3} & \cdots & \mathbf{V}_{2,2m} & \cdots \\
 \mathbf{V}_{3,1} & \mathbf{V}_{3,2} & \mathbf{V}_{3,3} & \cdots & \mathbf{V}_{3,m} & \mathbf{V}_{3,m+1} & \mathbf{V}_{3,m+2} & \mathbf{V}_{3,m+3} & \cdots & \mathbf{V}_{3,2m} & \cdots \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\
 \mathbf{V}_{km,1} & \mathbf{V}_{km,2} & \mathbf{V}_{km,3} & \cdots & \mathbf{V}_{km,m} & \mathbf{V}_{km,m+1} & \mathbf{V}_{km,m+2} & \mathbf{V}_{km,m+3} & \cdots & \mathbf{V}_{km,2m} & \cdots \\
 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 & \cdots & 1 & \cdots
 \end{pmatrix} \quad (A1)$$

563

$$\begin{pmatrix}
 \mathbf{V}_{1,(k-1)m+1} & \mathbf{V}_{1,(k-1)m+2} & \mathbf{V}_{1,(k-1)m+3} & \cdots & \mathbf{V}_{1,km} & 1 \\
 \mathbf{V}_{2,(k-1)m+1} & \mathbf{V}_{2,(k-1)m+2} & \mathbf{V}_{2,(k-1)m+3} & \cdots & \mathbf{V}_{2,km} & 1 \\
 \mathbf{V}_{3,(k-1)m+1} & \mathbf{V}_{3,(k-1)m+2} & \mathbf{V}_{3,(k-1)m+3} & \cdots & \mathbf{V}_{3,km} & 1 \\
 \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 \mathbf{V}_{km,(k-1)m+1} & \mathbf{V}_{km,(k-1)m+2} & \mathbf{V}_{km,(k-1)m+3} & \cdots & \mathbf{V}_{km,km} & 1 \\
 1 & 1 & 1 & \cdots & 1 & 0
 \end{pmatrix}$$

564 in which $\mathbf{V}_{i,j}$ is a matrix representing the correlation structure between CPT_{*i*} and CPT_{*j*} (where each
565 CPT has n data points),

$$566 \quad \mathbf{v}_{i,j} = \begin{pmatrix} d_{(i-1)n+1,(j-1)n+1} & d_{(i-1)n+1,(j-1)n+2} & d_{(i-1)n+1,(j-1)n+3} & \cdots & d_{(i-1)n+1,(j-1)n+n} \\ d_{(i-1)n+2,(j-1)n+1} & d_{(i-1)n+2,(j-1)n+2} & d_{(i-1)n+2,(j-1)n+3} & \cdots & d_{(i-1)n+2,(j-1)n+n} \\ d_{(i-1)n+3,(j-1)n+1} & d_{(i-1)n+3,(j-1)n+2} & d_{(i-1)n+3,(j-1)n+3} & \cdots & d_{(i-1)n+3,(j-1)n+n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{(i-1)n+n,(j-1)n+1} & d_{(i-1)n+n,(j-1)n+2} & d_{(i-1)n+n,(j-1)n+3} & \cdots & d_{(i-1)n+n,(j-1)n+n} \end{pmatrix} \quad (A2)$$

567 where $(i, j) = 1, 2, 3, \dots, m, m+1, m+2, m+3, \dots, 2m, \dots, (k-1)m+1, (k-1)m+2, (k-1)m+3, \dots, km$ and
 568 $d_{r,s}$ ($r = (i-1)n+1, \dots, (i-1)n+n$) ($s = (j-1)n+1, \dots, (j-1)n+n$) are the components of the submatrix $\mathbf{v}_{i,j}$,
 569 which can be expressed in the form of a covariance function between data points r and s (equation (2)).

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

571 The right-hand-side vector is formulated as

$$572 \quad \mathbf{v}_{rhs} = \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \vdots \\ \mathbf{v}_{km} \\ 1 \end{pmatrix} \quad (A3)$$

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

$$574 \quad \mathbf{v}_p = \begin{pmatrix} d_{(p-1)n+1} \\ d_{(p-1)n+2} \\ d_{(p-1)n+3} \\ \vdots \\ d_{(p-1)n+n} \end{pmatrix} \quad (A4)$$

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

579 The unknown weight vector is

580

$$\lambda_x = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_{km} \\ \mu \end{pmatrix} \quad (\text{A5})$$

581 in which λ_q is the weight subvector for CPT_q ,

582

$$\lambda_q = \begin{pmatrix} \lambda_{(q-1)n+1} \\ \lambda_{(q-1)n+2} \\ \lambda_{(q-1)n+3} \\ \vdots \\ \lambda_{(q-1)n+n} \end{pmatrix} \quad (\text{A6})$$

583 where $q = 1, 2, 3, \dots, m, m+1, m+2, m+3, \dots, 2m, \dots, (k-1)m+1, (k-1)m+2, (k-1)m+3, \dots, km$.

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599 **List of Figures**

600

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

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

607 **Figure 3.** Vertical and horizontal covariance functions averaged over 200 realisations ($\theta_v = 1.0$ m, $\theta_h =$
608 3.0 m)

609 **Figure 4.** Flowchart for conditional RFEM simulation

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

613 **Figure 6.** Simulation results for Example 1 (based on $\theta_v = 1.0$ m and 500 realisations per simulation)

614 **Figure 7.** Sampling efficiency indices for various values of ξ

615 **Figure 8.** Kriging variance for various values of ξ ($y-z$ slice at $i = 5$)

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

617 **Figure 10.** Influence of CPT location j during the second phase of site investigation (based on $\theta_v = 1.0$
618 m and 500 realisations per simulation)

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

621 **Figure 12.** Influence of number of CPTs (at $i = 5$) on sampling efficiency for various values of ξ and
 622 $\theta_v = 1.0$ m (Δ denotes the distance between CPTs)

623 **Figure 13.** Finite element meshes for different slope geometries

624 **Figure 14.** PDFs of realised factor of safety for three slopes, based on conditional and unconditional
 625 simulations

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

628 List of Tables

629

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

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

	$\theta_h = 6$ m	$\theta_h = 12$ m	$\theta_h = 24$ 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‰

634

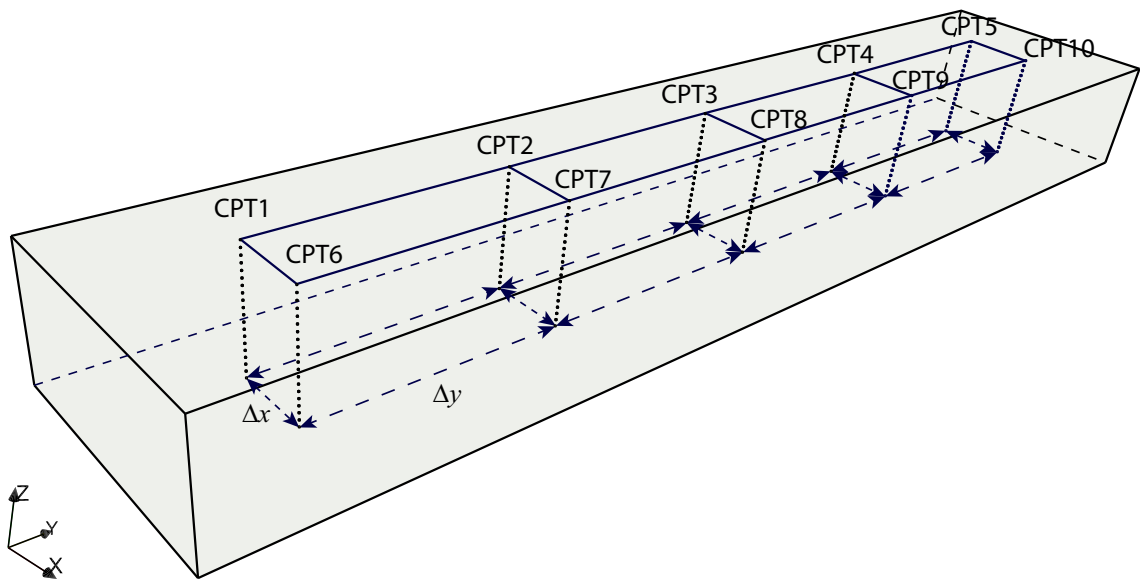


Figure 1: Example CPT sampling strategy ($k = 2, m = 5$)

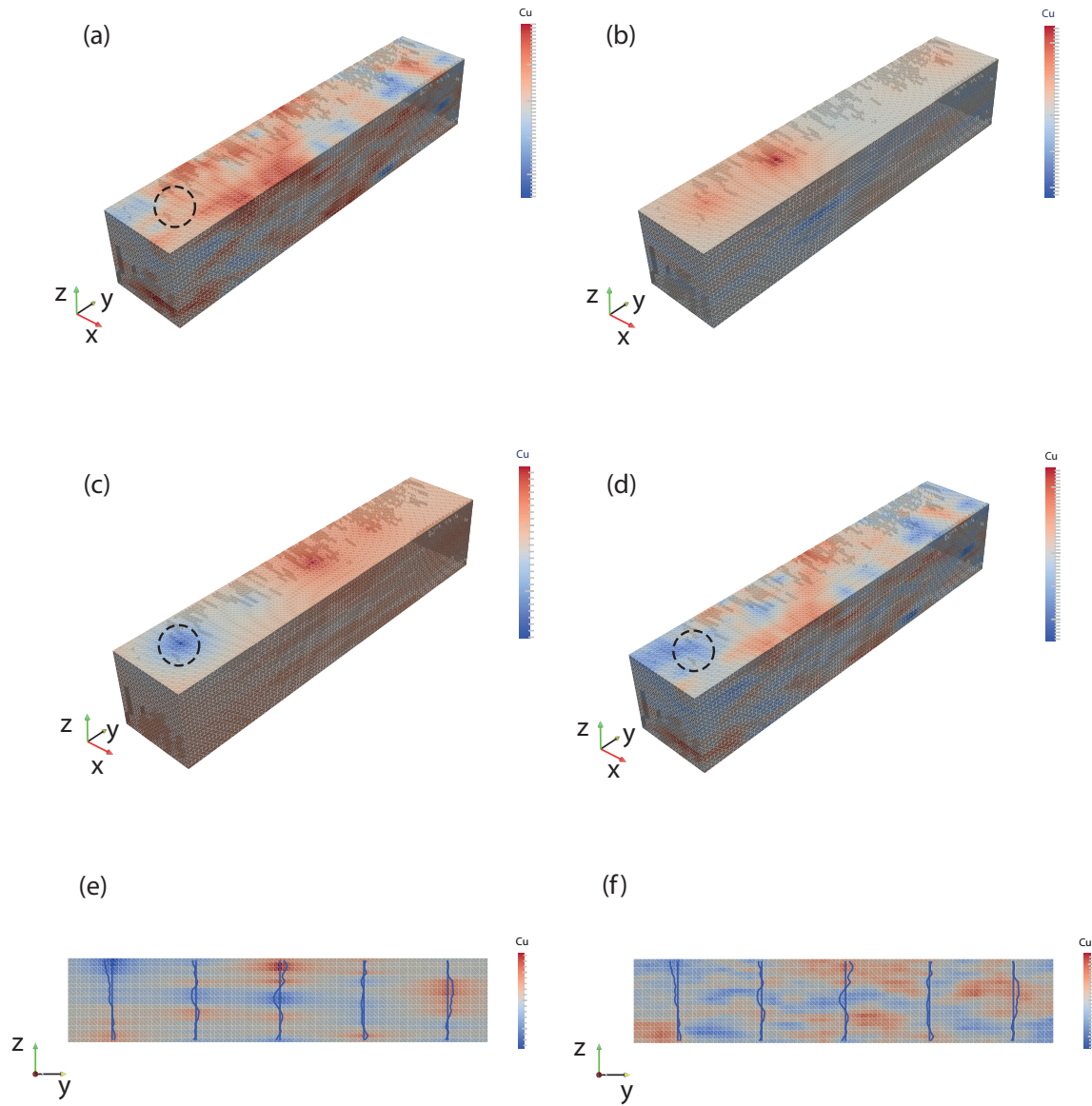


Figure 2: Example illustrations of the unconditional random field (a), the Kriged field based on the randomly simulated data (b), the Kriged field based on the CPT data (c), the conditional random field (d), cross-sections (e and f) in the longitudinal direction taken from the Kriged field (c) and from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

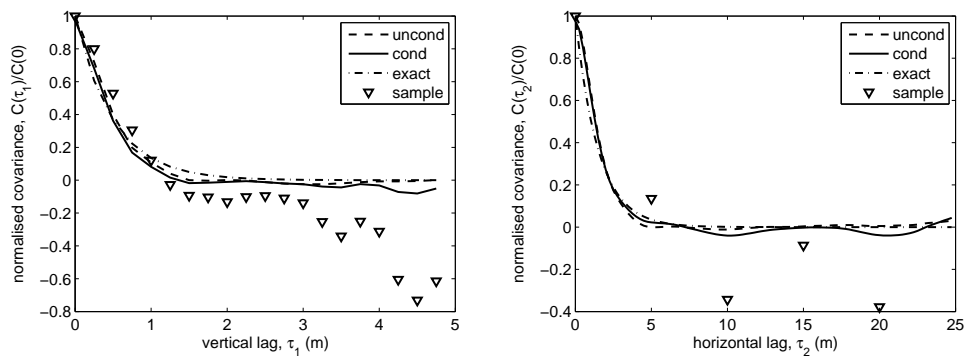


Figure 3: Vertical and horizontal covariance functions averaged over 200 realisations ($\theta_v = 1.0$ m, $\theta_h = 3.0$ m)

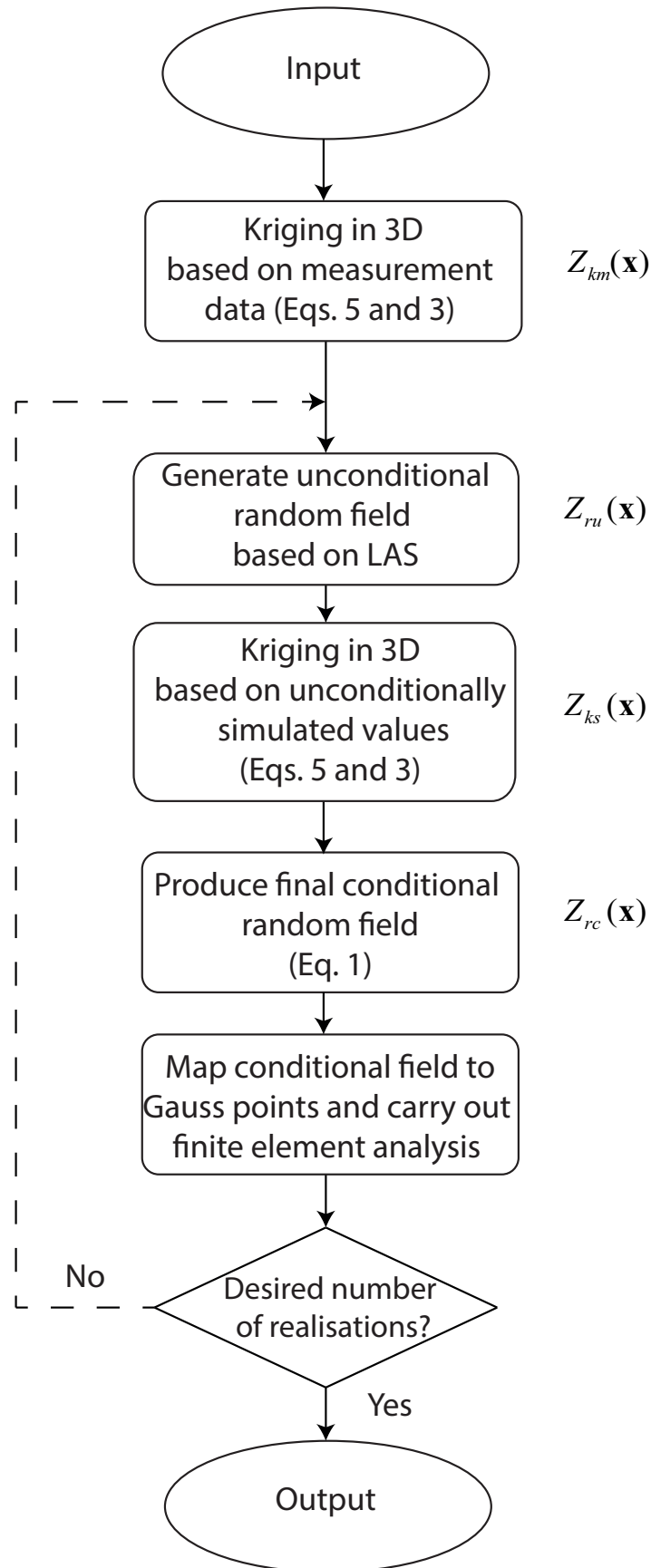


Figure 4: Flowchart for conditional RFEM simulation

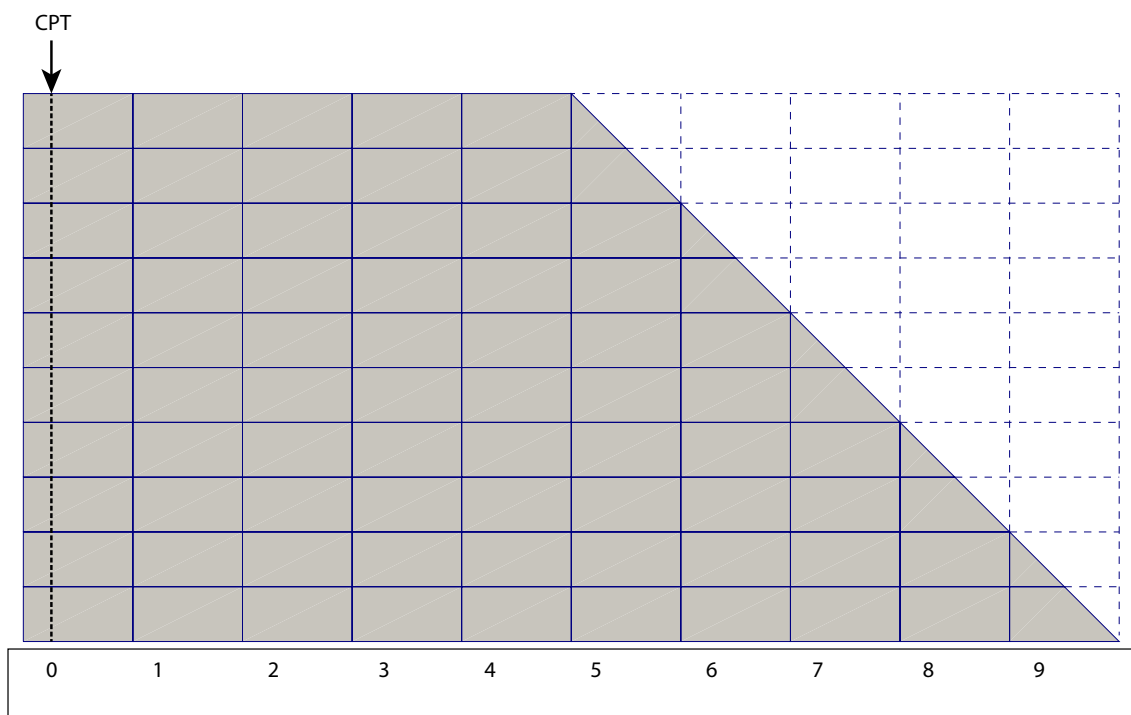
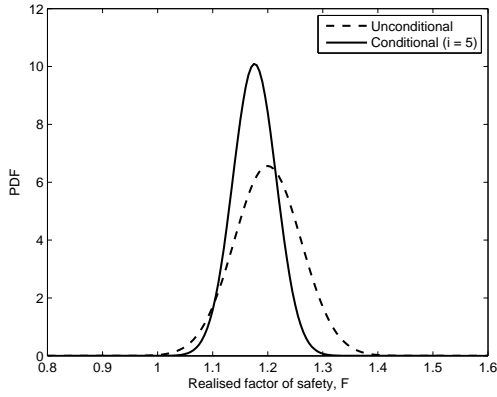
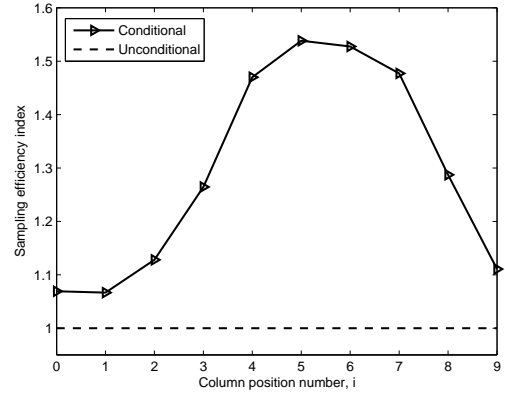


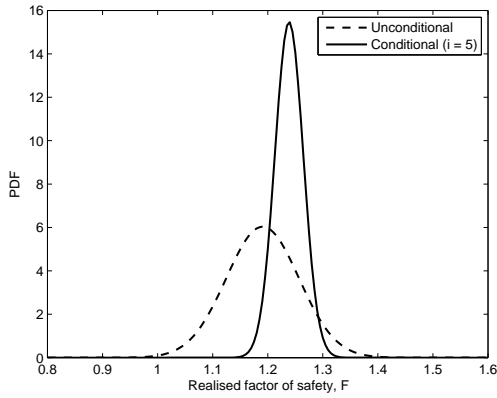
Figure 5: Finite element mesh and possible numbered CPT locations at a cross-section through the proposed 50 m long slope (dashed lines indicate the excavated soil mass and numbers correspond to Gauss point locations within the finite elements)



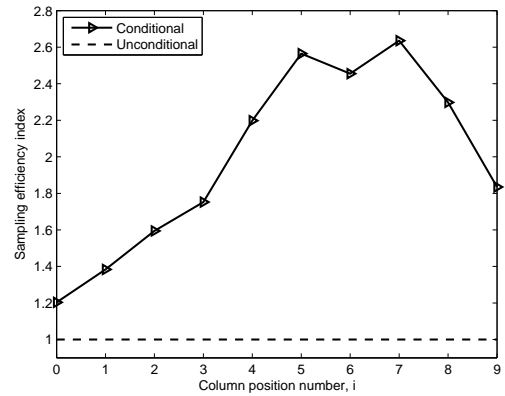
(a) Probability density functions of realised factor of safety ($\xi = 6$)



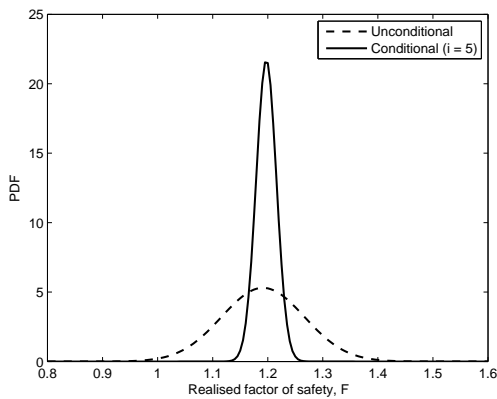
(b) Influence of CPT location ($\xi = 6$)



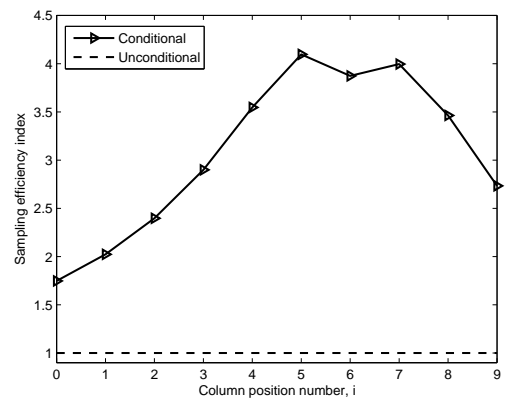
(c) Probability density functions of realised factor of safety ($\xi = 12$)



(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 realisations per simulation)

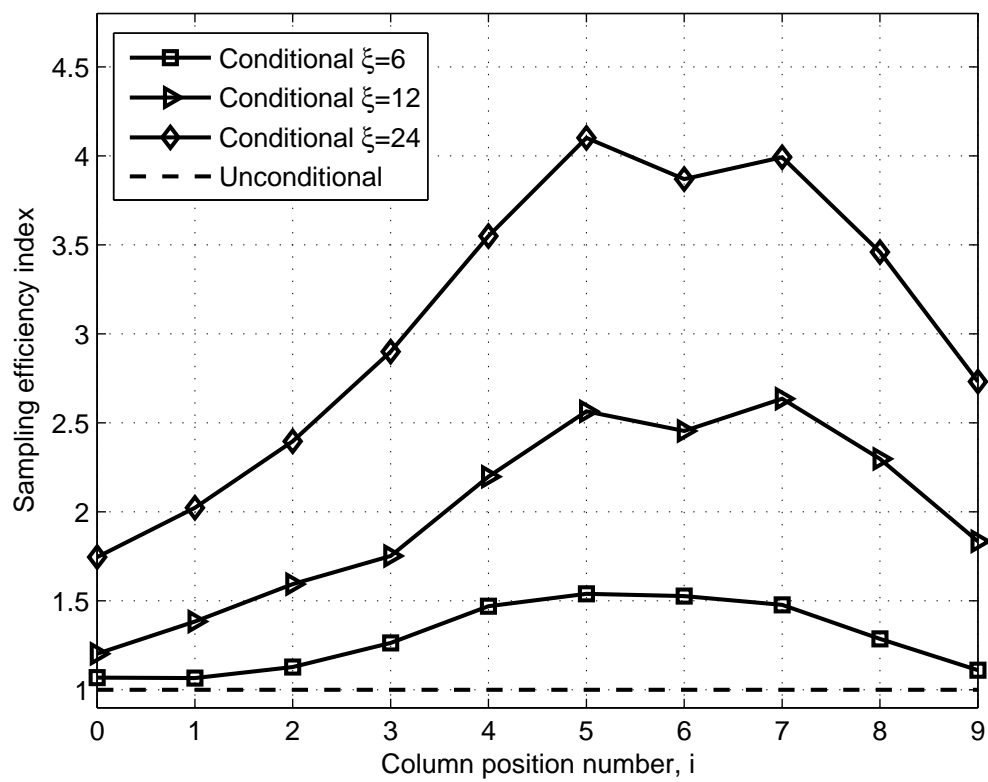


Figure 7: Sampling efficiency indices for various values of ξ

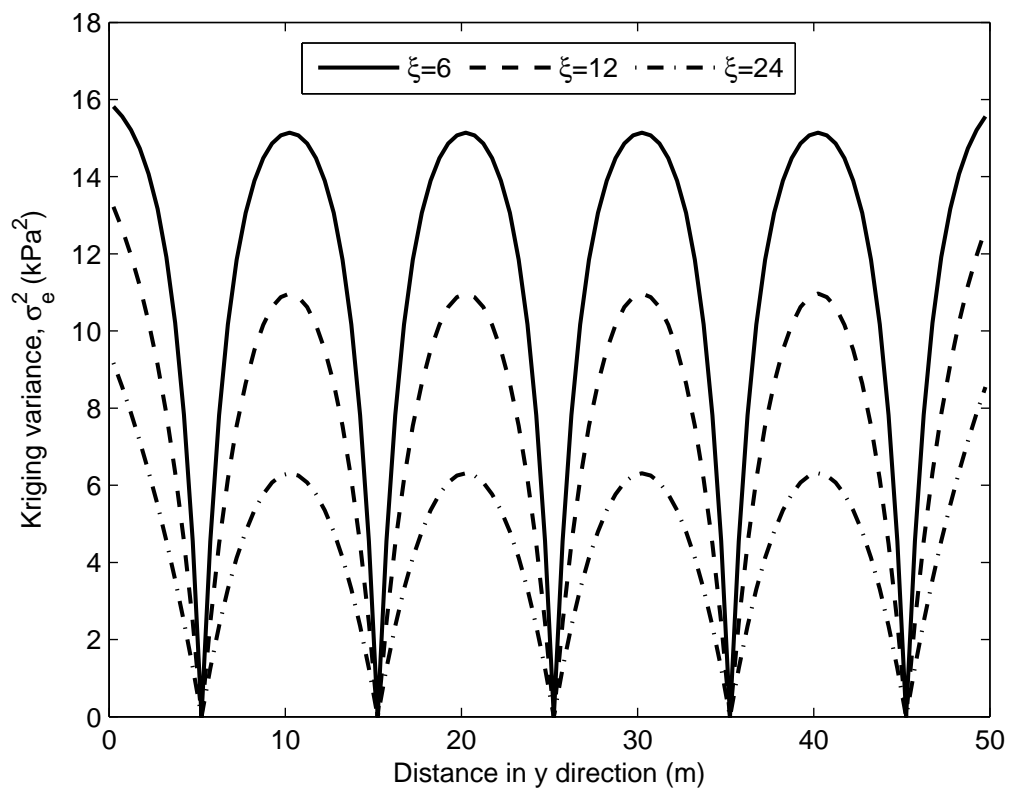


Figure 8: Kriging variance for various values of ξ (y - z slice at $i = 5$)

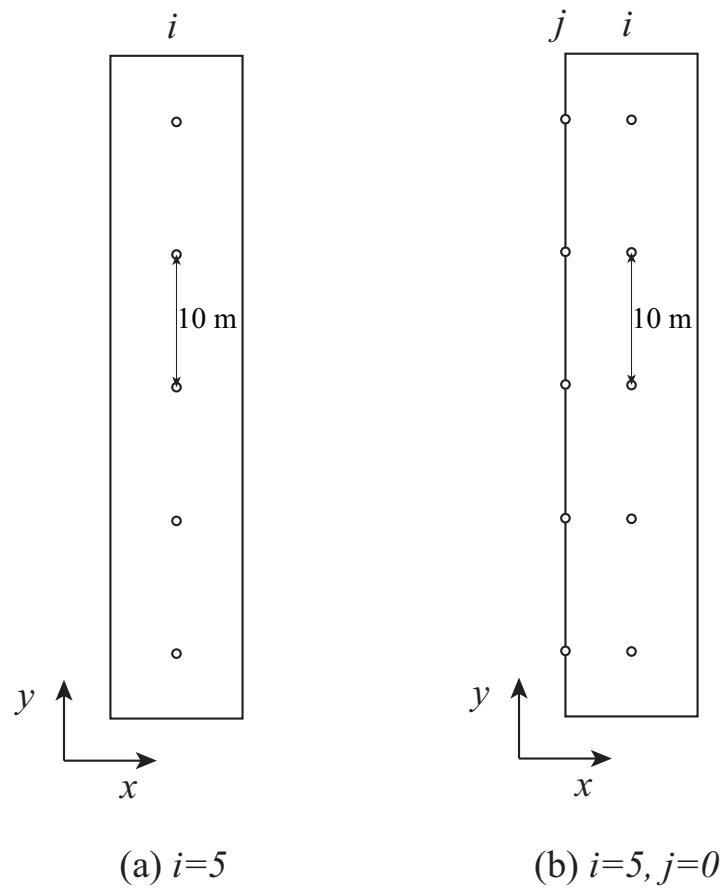
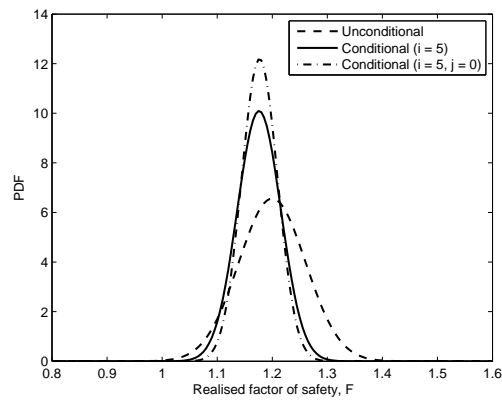
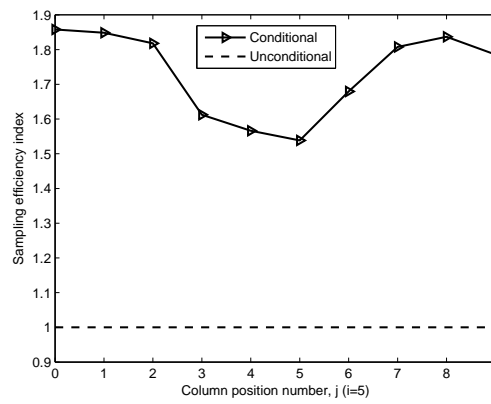


Figure 9: CPT layout illustration (plan view) for a single row (a) and two rows (b)



(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 realisations per simulation)

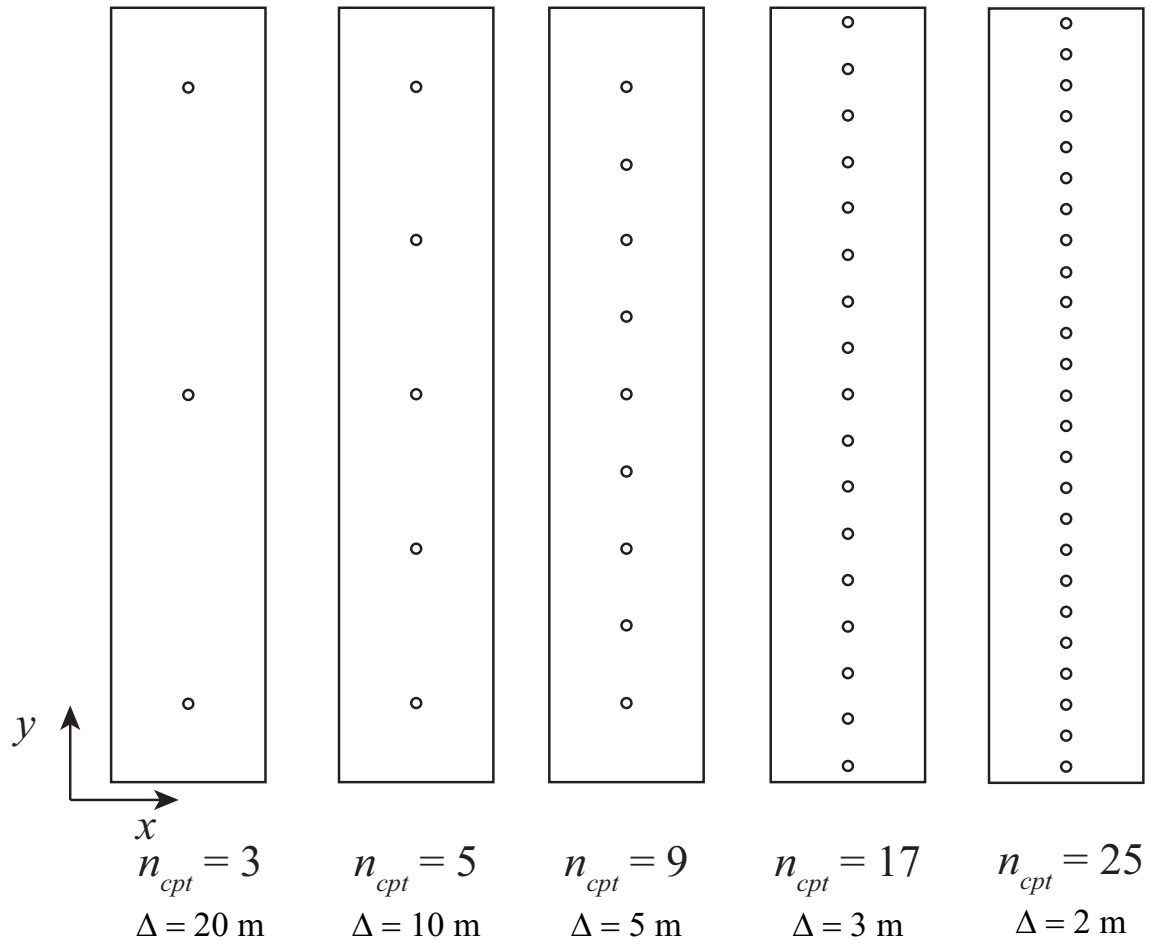
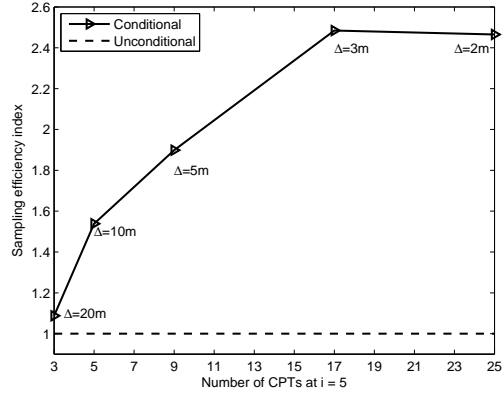
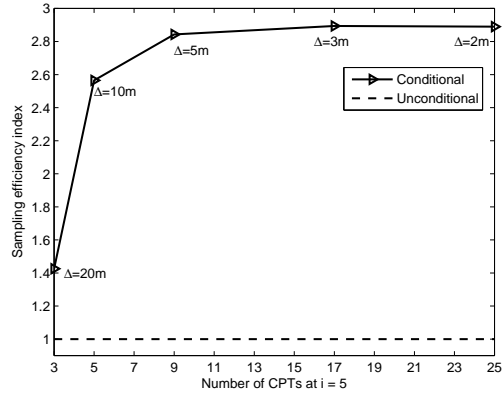


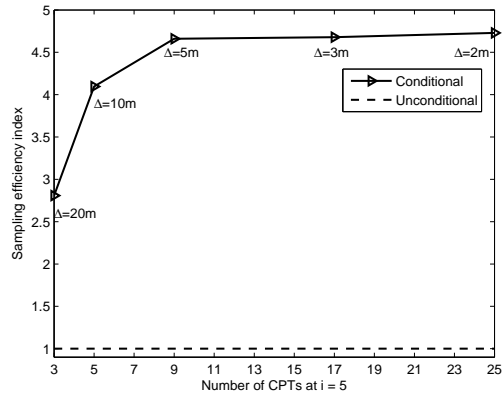
Figure 11: CPT layouts (plan views) for various numbers of boreholes ($n_{cpt} = 3, 5, 9, 17, 25$ and Δ denotes the distance between CPTs)



(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)

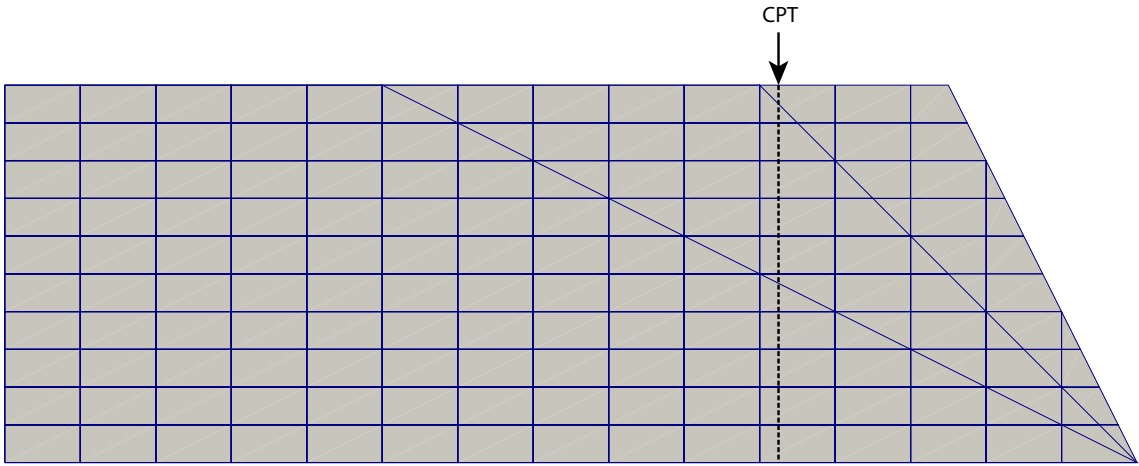


Figure 13: Finite element meshes for different slope geometries

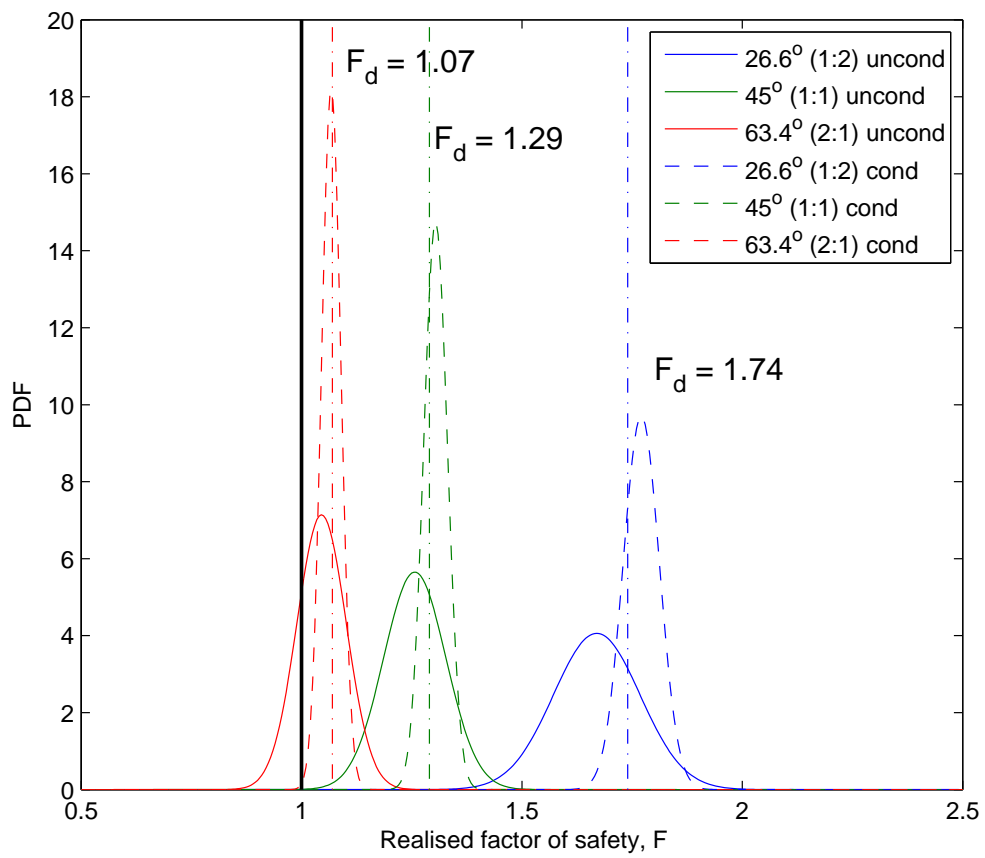


Figure 14: Probability density functions of realised factor of safety for three slopes, based on conditional and unconditional simulations

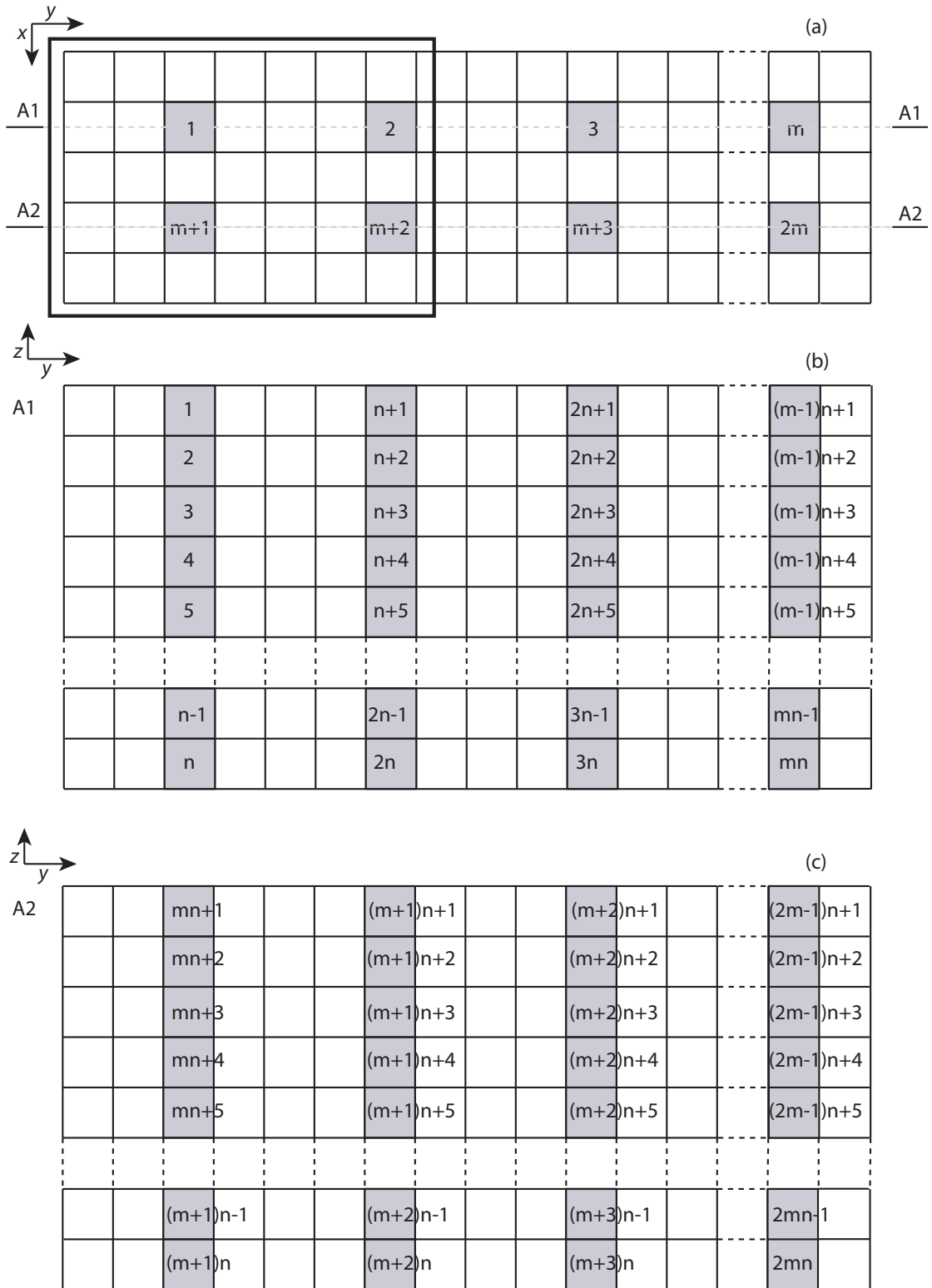


Figure A.1: Example CPT data grid ($k = 2$): (a) plan view showing CPT locations; (b) global numbering of data points at section A1; (c) global numbering of data points at section A2