Lloret-Cabot, M. *et al.* (2012) ***Géotechnique Letters* 2**, 123–127, http://dx.doi.org/10.1680/geolett.12.00022

Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging

M. LLORET-CABOT*, M. A. HICKS* and A. P. VAN DEN EIJNDEN*

Spatial variability of soil properties is inherent in soil deposits, whether as a result of natural geological processes or engineering construction. It is therefore important to account for soil variability in geotechnical design in order to represent more realistically a soil's in situ state. This variability may be modelled as a random field, with a given probability density function and scale of fluctuation. A more convenient way to deal with the uncertainty of a soil property due to spatial variability, by constraining the generated random field at the locations of actual field measurements, is presented in this article. Conditioning the random field at known locations is a powerful tool, not only because it more accurately represents the observed variability on site, but also because it uses the available field information more efficiently. In situ cone penetration test (CPT) data from a particular test site are used to determine the input statistics for generating random fields, which are later constrained (conditioned) at the locations of actual CPT measurements using the Kriging interpolation method. The results from the conditional random fields are then analysed, to quantify how the number of field measurements used influences the reduction of uncertainty. It is shown that the spatial uncertainty relative to the original (unconditional) random field reduces with the number of CPTs used in the conditioning.

KEYWORDS: in situ testing; site investigation; statistical analysis

INTRODUCTION

According to Phoon & Kulhawy (1999), three primary sources of uncertainty are typically observed in soils – inherent variability, measurement error and transformation of field (or laboratory measurements) into design soil properties. Focusing on the first source of uncertainty, this paper investigates a new strategy to represent the observed spatial variability in soils more efficiently.

It is usual to model the heterogeneity of a soil property (i.e. the inherent variability) as a random field, for example using a generation method such as local average subdivision (LAS), as proposed by Fenton & Vanmarcke (1990) and used extensively in the literature (e.g. Griffiths & Fenton, 2001; Hicks & Spencer, 2010). For this purpose, the LAS method requires a probability density function (pdf) with its point statistics (mean μ and standard deviation σ) and the scale of fluctuation (θ). This statistical information of the soil property can be estimated from available field data obtained from testing at discrete locations across a site. An improved strategy to deal with the uncertainty in modelling the spatial variability of a soil property is to generate random fields that are constrained (i.e. conditioned) at the locations of known cone penetration test (CPT) measurements using the Kriging interpolation method (Cressie, 1990). This is the strategy chosen here, because it minimises the uncertainty of the soil domain studied while optimising the use of available field information. The results from conditional random fields are analysed here to evaluate and quantify how the number of field measurements influences the reduction of uncertainty. It is shown that uncertainty reduces with increasing numbers of CPTs (used to constrain the conditional random field) and this reduction depends on the separation distance between CPTs.

Manuscript received 4 May 2012; first decision 27 June 2012; accepted 13 July 2012.
Published online at www.geotechniqueletters.com on 28 September 2012.
*Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands

STATISTICAL INTERPRETATION OF CPT TIP RESISTANCES

Numerous artificial islands were constructed offshore during the 1970s and 1980s in the Canadian Beaufort Sea to provide temporary structures for oil and gas exploration. One type of island used caisson technology to reduce the required fill volumes (Hicks & Smith, 1988). It comprised two main sand fills – a layer of sand (referred to as the berm) on which the caisson was founded and the body of the island structure (referred to as the core). Since no prior design experience on these types of constructions was available at the time, an extensive quality assurance and monitoring programme was carried out during and after island construction. This paper investigates in situ data from one such island, Tarsiut P-45. In particular, seven CPTs are used to statistically describe the tip resistances q_c of the sand fill core. The seven CPTs are aligned along the same straight line and so they represent the soil variability for a two-dimensional (2D) section. The obtained statistics are then used to generate a 2D random field of q_c, which is later constrained at the CPT locations.

Point statistics and probability density function

The mean is taken to be the average value of tip resistances measured over all CPT profiles (Table 1). Similarly, the standard deviation was calculated for each profile to quantify how the tip resistances deviate from the mean in the dataset. The value in Table 1 is the average standard deviation from the seven profiles. Table 1 also gives the standard deviation of the cone tip resistance after removing the underlying depth-dependent trend, σ_{res}.

Table 1. Cone tip resistance statistics

Property	Range	Mean
Mean, μ: MPa	4·96–7·00	6·24
Standard deviation, σ: MPa	2·22–4·09	2·90
Standard deviation, σ_{res}: MPa (trend removed)	1·28–2·52	1·88
Input variance, σ^2_{res}: MPa^2	1·64–6·35	3·53
Vertical scale of fluctuation[a], θ_v: m	0·37–0·91	0·63
Horizontal scale of fluctuation[b], θ_h: m	—	7·00
Degree of anisotropy of the heterogeneity, ξ	—	11·00

[a]θ_v values estimated using the method of Wickremesinghe & Campanella (1993)
[b]θ_h values estimated from the covariance function (Baecher & Christian, 2003)

A normal distribution was used to represent the cone tip resistances of the Tarsiut P-45 core. Figure 1 shows a histogram based on all data from the seven CPTs, as well as the fitted distribution function. Inspection of this figure shows that, in this particular case, the variation of tip resistance is reasonably well represented by a normal distribution. Note that the data plotted in Fig. 1 have been de-trended and normalised. That is, a linear depth trend of q_c was determined for each profile and subsequently removed from the data. Then, in order to normalise the data, the de-trended tip values were divided by the standard deviation of this new set of data (σ_{res}). The data used to plot the histogram in Fig. 1 were obtained by repeating this process for each CPT profile.

Spatial statistics (scales of fluctuation)

The spatial nature of the variation is of particular importance. The scale of fluctuation θ defines the distance over which properties are significantly correlated (Vanmarcke, 1984) and may be used to generate numerical predictions of spatial variability (Hicks & Onisiphorou, 2005). A second useful parameter is the degree of anisotropy of the heterogeneity ξ, which relates the horizontal and vertical scales of fluctuation (i.e. $\xi = \theta_h/\theta_v$) (Hicks & Samy, 2002).

The scale of fluctuation is used in the correlation model that is used to generate the random field and which describes how the property value varies throughout the soil domain. The strategy assumes that available data are statistically homogeneous (or stationary), which requires

- a constant mean and a constant standard deviation
- a correlation function independent of the location and dependent only on the separation distance (also referred to as the lag distance).

For geotechnical engineering purposes, this is equivalent to saying that the mean and standard deviation do not spatially change and that the correlation between the values at two locations is only a function of their separation distance. A constant mean is obtained after de-trending the data as described earlier. The conditions of a constant standard deviation and a correlation function dependent only on the lag distance are likely to be reached if the data are extracted from the same soil layer, as approximately uniform fluctuations are then likely to be observed (Phoon & Kulhawy, 1999). This is a reasonable assumption for the case studied here because all the core sand was dredged from the same location and deposited in the same manner (Wong, 2004).

Various methods are available to estimate the scale of fluctuation. A simple and useful first approach is to estimate the semi-variogram from available data and best fit it with a theoretical semi-variogram (Baecher & Christian, 2003; Dasaka & Zhang, 2012). This method has been used here to estimate θ_h, albeit using only the seven CPTs available for the 2D section investigated. However, a more reliable estimate is possible for the vertical scale of fluctuation θ_v, by using the method proposed by Wickremesinghe & Campanella (1993). The results obtained for the Tarsiut P-45 core are summarised in Table 1 and are consistent with similar previous studies (e.g. Wong, 2004; Hicks & Onisiphorou, 2005).

MODELLING SPATIAL VARIABILITY FOR THE TEST SITE

Conditional random fields

An existing 2D LAS algorithm developed by Spencer (2007) was used to generate the random fields in this study, using the statistical inputs for the Tarsiut P-45 core summarised in Table 1. The generated random fields were then constrained by the measurements at the CPT locations (see Fig. 2). The conditional approach follows previous work by van den Eijnden & Hicks (2011) and is based on the Kriging interpolation technique (Krige, 1951), which is used extensively in geostatistics (Cressie, 1990; Wackernagel, 2003; Fenton & Griffiths, 2008) and is only briefly reviewed here.

The aim of Kriging is to give a best estimate of a random field between known data. In essence, the method estimates

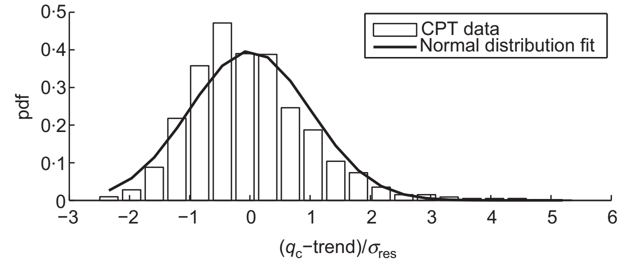

Fig. 1. Histogram of normalised de-trended tip resistance (based on all CPT data)

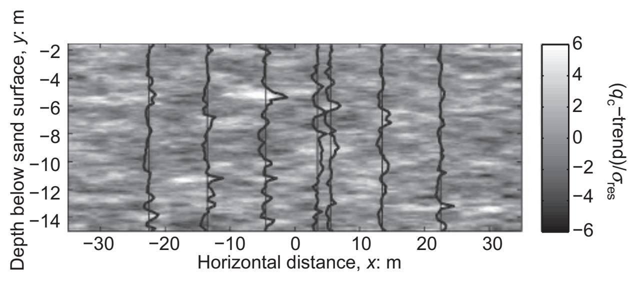

Fig. 2. Typical realisation of a conditional random field for a 2D section of the Tarsiut P-45 core

Z at a desired location $\boldsymbol{x}_0$, from a linear combination of the known values of Z at various observation points $\boldsymbol{x}_\alpha$. The Kriged interpolation of Z at $\boldsymbol{x}_0$ (i.e. $Z^*(\boldsymbol{x}_0)$) can be written as a linear combination of the known (conditioning) points $Z(\boldsymbol{x}_\alpha)$ (for α = 1, 2,…, n, with n being the number of observations) as

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(\mathbf{x}_\alpha) \text{ with } \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \tag{1}$$

where λ_α are the n unknown weights, to be determined so as to find the best estimate for Z at $\boldsymbol{x}_0$.

Although various forms of Kriging are available, ordinary Kriging is adopted here. This method assumes stationary data (as described above) and the estimation variance σ_E^2 at location $\boldsymbol{x}_0$ can be written, in terms of the semi-variogram $\gamma(h)$, as (Wackernagel, 2003).

$$\sigma_E^2 = -\gamma(0) - \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) + 2\sum_{\alpha=1}^{n} \lambda_\alpha \gamma(\mathbf{x}_\alpha - \mathbf{x}_0) \tag{2}$$

By minimising this equation, the following system of equations is obtained, from which the unknown weights are estimated (Wackernagel, 2003)

$$\begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_1 - \mathbf{x}_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_1) & \cdots & \gamma(\mathbf{x}_n - \mathbf{x}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \mu^{OK} \end{pmatrix} = \begin{pmatrix} \gamma(\mathbf{x}_1 - \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_n - \mathbf{x}_0) \\ 1 \end{pmatrix} \tag{3}$$

where λ_α^{OK} are the ordinary Kriging weights and μ^{OK} is the Lagrange parameter. The left-hand side of the system contains information between data points, whereas the right-hand side contains information between each data point and the estimation point. Combining equations (2) and (3), it is possible to obtain the following expression for the ordinary Kriging variance σ_{OK}^2 (Wackernagel, 2003)

$$\sigma_{OK}^2 = \mu^{OK} - \gamma(0) + \sum_{\alpha=1}^{n} \lambda_\alpha^{OK} \gamma(\mathbf{x}_\alpha - \mathbf{x}_0) \tag{4}$$

The conditioning of a random field comprises a process that uses a generic (unconditional) random field with the statistics calculated from the in situ data (see Table 1). This generic random field is post-processed in order to condition it at the known conditioning points. A description of the method is given by Journel & Huijbregts (1978). Figure 2 shows a typical realisation of a conditional random field, for the 2D section of the Tarsiut P-45 core for which the statistics have been derived (Table 1). High and low normalised tip resistances are denoted by light and dark zones, respectively. The location of each CPT profile is also included in the figure, along with the actual CPT profiles.

INVESTIGATION OF UNCERTAINTY IN SOIL VARIABILITY

This section investigates how the conditioning algorithm affects the reduction of spatial uncertainty in the random field. To illustrate this aspect, Fig. 3 plots the estimation

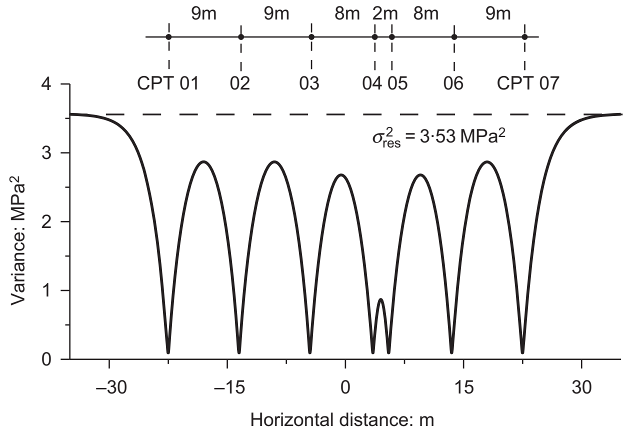

Fig. 3. Estimation variance against horizontal distance for the 2D section of the Tarsiut P-45 core

variance, calculated using equation (4) and averaged over the depth of the conditional field, with respect to horizontal distance. At the top of the figure, the vertical dashed lines represent the locations of the CPTs. Obviously, the highest reduction of the variance is observed at the locations of the conditioning CPTs, where the variance reaches a minimum value very close to 0. Theoretically, a reduction to zero variance would be reached were continuous measurements available; however, here the vertical separation between each CPT measurement is 2 cm.

The reduction in spatial uncertainty relative to the original (unconditional) random field may be evaluated by comparing the input variance σ_{res}^2 and the estimation (Kriging) variance, both as a function of position. A suitable quantity that provides information on the global uncertainty for the domain investigated is the integral of the variance over the domain. In light of this, the area under σ_{res}^2 in Fig. 3 is referred to as the global spatial uncertainty of the generic field ($u_{original}$) while the area under the Kriging variance is referred to as the global spatial uncertainty of the conditional random field (u_{cond}). To illustrate this, Table 2 shows the global spatial uncertainty obtained when conditioning the original random field with all seven CPTs for a single realisation. The reduction in the global spatial uncertainty for six equivalent calculations (i.e. obtained by constraining the random field with one, two, three, four, five or six CPTs) is also indicated. An overall reduction of the global spatial uncertainty is clearly observed when increasing the number of CPTs used in the conditioning algorithm, obtaining a maximum reduction of 35·54% when using all seven CPTs available from the site.

Table 2. Reduction of global spatial uncertainty with increasing number of CPTs (u_{cond} refers to the total global uncertainty for the conditional field; $u_{original}$ corresponds to the global uncertainty of the original field)

Number of CPTs	Global uncertainty reduction $\left(1 - \frac{u_{cond}}{u_{original}}\right)100\%$
7	35·54
6	29·85
5	24·27
4	21·69
3	15·94
2	9·66
1	2·35

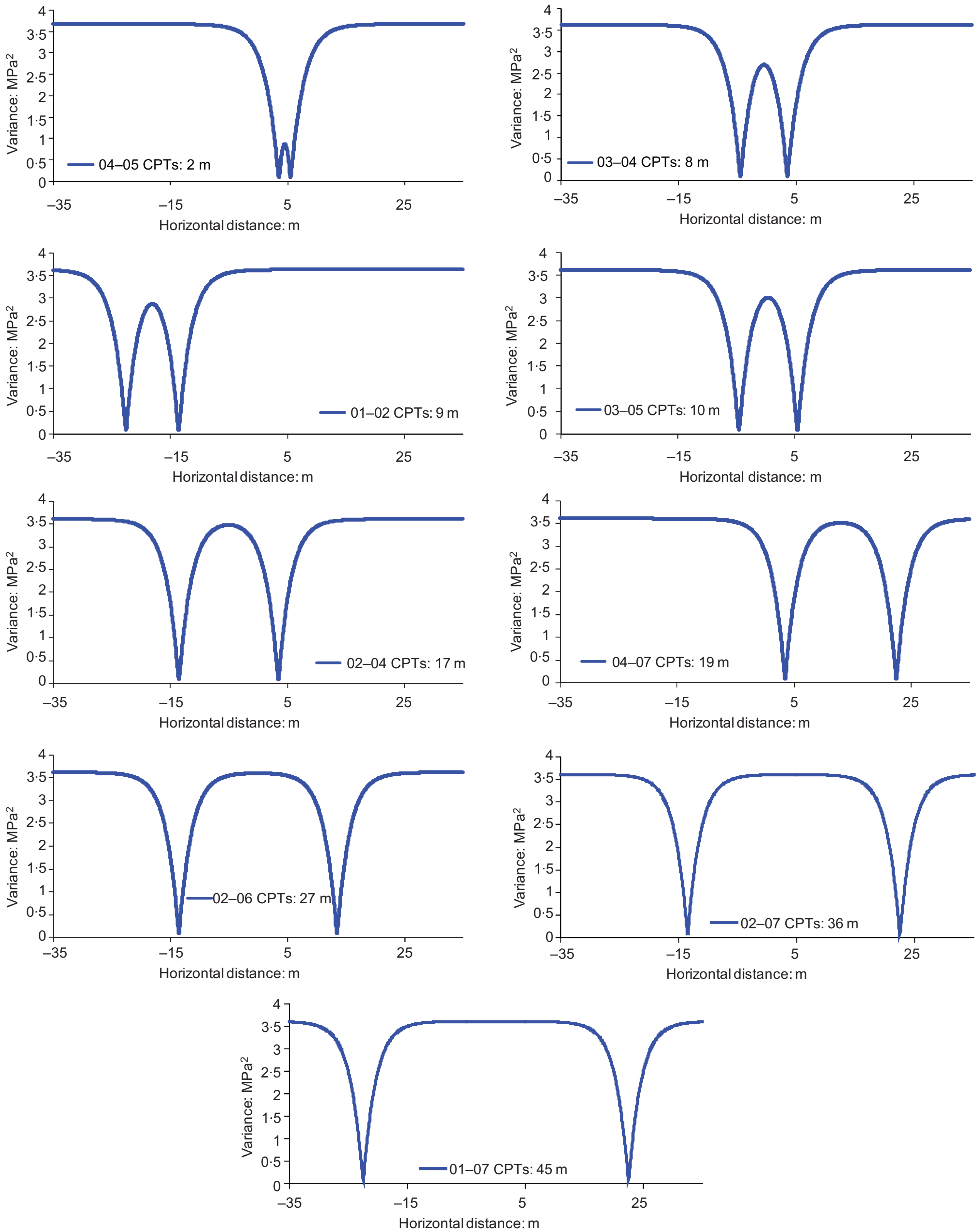

Fig. 4. Influence of the horizontal distance between two CPTs on the global spatial uncertainty

It is perhaps more interesting to analyse the influence of the horizontal distance between CPTs. For this purpose, the results for conditioning the random field with only two CPTs were considered for various separation distances. Different choices are possible when looking at a pair of CPTs, as seen in Fig. 2, resulting in different distances between pairs of CPTs. Figure 4 illustrates the results obtained by performing the same analysis described above for nine different possible distances between CPTs. Table 3 highlights the important information for the discussion.

Inspection of Table 3 suggests that the greatest reduction of global uncertainty in this second study (involving only two CPTs) is reached when the horizontal distance between the CPTs is somewhere between 27 and 45 m (Fig. 5). Inspection of Fig. 4 suggests that an optimal spacing may occur when, for a given CPT, the influence of the second

Table 3. Influence of the horizontal distance between two CPTs on global spatial uncertainty (u_{cond} refers to the total global uncertainty for the conditional field; $u_{original}$ corresponds to the global uncertainty of the original field)

Horizontal distance: m	Global uncertainty reduction $\left(1-\frac{u_{cond}}{u_{original}}\right)100\%$
2	5·38
8	9·43
9	9·66
10	9·87
17	10·35
19	10·36
27	10·43
36	10·41
45	10·40

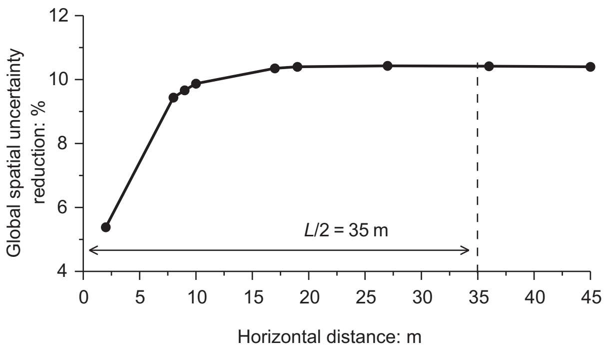

Fig. 5. Reduction of global spatial uncertainty with horizontal distance between two CPTs

CPT and the domain boundaries are at a minimum. This approximately corresponds to a CPT spacing of one half the total horizontal length, with each CPT being at a distance of a quarter of the total length from the nearest domain boundary.

CONCLUSIONS

The sand fill core of an artificial island was investigated using CPT data. Cone tip resistances were statistically interpreted in terms of the mean, standard deviation, distribution function and scale of fluctuation. This information was then used to generate a conditional random field to characterise the heterogeneity of the sand state. Conditioning the random field to known CPT data provides a way to use the available in situ information more efficiently, constraining the generated random field by the actual measurements. The results obtained in this study show that the spatial uncertainty reduces with increasing numbers of CPTs used in the conditional random field and it also depends on their mutual separation distance. This study also shows that, with only two CPTs, an optimal spacing will occur when they are as far from each other and the edges of the domain as possible, which corresponds to half of the total horizontal length.

Acknowledgements

The authors gratefully acknowledge the support provided by the EU FP7 programme under the project 'Geo-Install' (PIAP-GA-2009-230638). The authors are also grateful for the valuable comments and suggestions given by the reviewers.

REFERENCES

Baecher, G. B. & Christian, J. T. (2003). *Reliability and statistics in geotechnical engineering*. Hoboken, NJ: Wiley.

Cressie, N. (1990). The origins of Kriging. *Math. Geol.* **22**, No. 3, 239–252.

Dasaka, S. M. & Zhang, L. M. (2012). Spatial variability of in situ weathered soil. *Géotechnique* **62**, No. 5, 375–384.

Fenton, G. A. & Griffiths, D. V. (2008). *Risk assessment in geotechnical engineering*. Hoboken, NJ: Wiley.

Fenton, G. A. & Vanmarcke, E. H. (1990). Simulation of random fields via local average subdivision. *J. Engng Mech.* **116**, No. 8, 1733–1749.

Griffiths, D. V. & Fenton, G. A. (2001). Bearing capacity of spatially random soil: the undrained clay Prandtl problem revisited. *Géotechnique* **51**, No. 4, 351–359.

Hicks, M. A. & Onisiphorou, C. (2005). Stochastic evaluation of static liquefaction in a predominantly dilative sand fill. *Géotechnique* **55**, No. 2, 123–133.

Hicks, M. A. & Samy, K. (2002). Influence of heterogeneity on undrained clay slope stability. *Quart. J. Engng Geol. Hydrogeol.* **35**, No. 1, 41–49.

Hicks, M. A. & Smith, I. M. (1988). Class A prediction of Arctic caisson performance. *Géotechnique* **38**, No. 4, 589–612.

Hicks, M. A. & Spencer, W. A. (2010). Influence of heterogeneity on the reliability and failure of a long 3D slope. *Comput. Geotech.* **37**, No. 7–8, 948–955.

Journel, A. G. & Huijbregts, Ch. J. (1978). *Mining geostatistics*. New York: Academic.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on Witwatersrand. *J. Chem. Metall. Mining Soc. South Africa* **52**, No. 6, 119–139.

Phoon, K. K. & Kulhawy, F. H. (1999). Characterization of geotechnical variability. *Canad. Geotech. J.* **36**, No. 4, 612–624.

Spencer, W. A. (2007). *Parallel stochastic and finite element modelling of clay slope stability in 3D*. PhD thesis, University of Manchester, UK.

van den Eijnden, A. P. & Hicks, M. A. (2011). Conditional simulation for characterizing the spatial variability of sand state. *Proc. 2nd Int. Symp. Comp. Geomech., Croatia*, pp. 288–296.

Vanmarcke, E. H. (1984). *Random fields: analysis and synthesis*. Cambridge, MA: MIT Press.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Berlin: Springer.

Wickremesinghe, D. & Campanella, R. G. (1993). Scale of fluctuation as a descriptor of soil variability. *Proc. Conf. Prob. Methods in Geotech. Engng, Canberra*, pp. 233–239.

Wong, S. Y. (2004). *Stochastic characterisation and reliability of saturated soils*. PhD thesis, University of Manchester, UK.

WHAT DO YOU THINK?

To discuss this paper, please email up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial panel, will be published as a discussion.