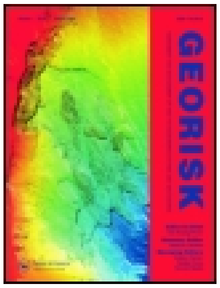

Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ngrk20

On the estimation of scale of fluctuation in geostatistics

M. Lloret-Cabot[ab], G.A. Fenton[bc] & M.A. Hicks[b]

[a] Centre for Geotechnical and Materials Modelling, University of Newcastle, Newcastle, Australia

[b] Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands

[c] Department of Engineering Mathematics, Dalhousie University, Halifax, Canada
Published online: 02 Jan 2014.

Georisk, 2013
http://dx.doi.org/10.1080/17499518.2013.871189

On the estimation of scale of fluctuation in geostatistics

M. Lloret-Cabot[a,b*], G.A. Fenton[b,c] and M.A. Hicks[b]

[a]*Centre for Geotechnical and Materials Modelling, University of Newcastle, Newcastle, Australia;* [b]*Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands;* [c]*Department of Engineering Mathematics, Dalhousie University, Halifax, Canada*

(*Received 15 July 2013; accepted 28 November 2013*)

Describing how soil properties vary spatially is of particular importance in stochastic analyses of geotechnical problems, because spatial variability has a significant influence on local material and global geotechnical response. In particular, the scale of fluctuation θ is a key parameter in the correlation model used to represent the spatial variability of a site through a random field. It is, therefore, of fundamental importance to accurately estimate θ in order to best model the actual soil heterogeneity. In this paper, two methodologies are investigated to assess their abilities to estimate the vertical and horizontal scales of fluctuation of a particular site using *in situ* cone penetration test (CPT) data. The first method belongs to the family of more traditional approaches, which are based on best fitting a theoretical correlation model to available CPT data. The second method involves a new strategy which combines information from conditional random fields with the traditional approach. Both methods are applied to a case study involving the estimation of θ at three two-dimensional sections across a site and the results obtained show general agreement between the two methods, suggesting a similar level of accuracy between the new and traditional approaches. However, in order to further assess the relative accuracy of estimates provided by each method, a second numerical analysis is proposed. The results confirm the general consistency observed in the case study calculations, particularly in the vertical direction where a large amount of data are available. Interestingly, for the horizontal direction, where data are typically scarce, some additional improvement in terms of relative error is obtained with the new approach.

Keywords: spatial variability; random fields; soil heterogeneity; characterisation of soil/rock variability; geostatistics

1. Introduction

This paper compares the performance of two different methods to estimate the vertical and horizontal scales of fluctuation using *in situ* cone penetration test (CPT) data from a particular test site. The first method will be referred to as Approach A and is based on more conventional (or classical) approaches. The second method will be referred to as Approach B and involves a new strategy which combines information from conditional random fields with the traditional approach. To illustrate and assess their relative performance, both strategies are applied to a case study and the results are evaluated. The goal of the paper is to answer the following question: are conventional techniques for estimating the correlation length as good as they can be, or is there the possibility for improvement?

The scale of fluctuation θ is a convenient measure for describing the spatial variability of a soil property in a random field. It is a measure of the distance within which points are significantly correlated (Vanmarcke 1984). Points separated by a larger distance than θ will show little correlation, and practically no correlation will be observed when points are separated by a significantly larger distance than θ. This relationship between soil property values and relative distances is contained within the correlation model, which is a function of the lag τ (i.e. distance between points) and the scale of fluctuation θ. Some common correlation models are summarised in Table 1, including the Gaussian model, the triangular model, the spherical model and the Markov correlation model used here. In each of these models, small values of θ imply that the correlation function falls off rapidly to zero with increasing τ (i.e. the correlation between two points becomes rapidly smaller), which leads to rougher random fields. In the limit, as $\theta \to 0$, all points in the domain become uncorrelated and the field becomes infinitely rough. At the other extreme, for increasing values of θ the soil property field becomes smoother, or, in other words, the field shows less variability converging to a uniform field when $\theta \to \infty$.

The correlation model is a fundamental ingredient in the stochastic analyses of geotechnical problems, not only because it describes how the soil property values vary spatially throughout the geometrical domain, but also, more importantly, because the spatial variation itself has a significant influence on the response of the geotechnical structure. This is of special interest, given that random fields are typically used to model soil

*Corresponding author. Email: marti.lloretcabot@newcastle.edu.au

heterogeneity (i.e. inherent variability), in advanced stochastic analyses (Fenton 1999; Fenton and Griffiths 2003; Hicks and Onisiphorou 2005; Fenton and Griffiths 2005; Griffiths, Huang, and Fenton 2009; Hicks and Spencer 2010; Cassidy, Uzielli, and Tian 2013).

Perhaps due to the complexity associated with the modeling of soil heterogeneity, however, little research has been done to accurately describe its nature and this has typically led to inherent variability being one of the primary sources of uncertainty in stochastic analyses in geotechnical engineering (Fenton 1999; Phoon and Kulhawy 1999). The scale of fluctuation, in particular, plays a key role in the description of soil variability at a site. It is therefore crucial to estimate accurate values of the vertical and horizontal scales of fluctuation in order to obtain more realistic responses of the geotechnical structure when using advanced probabilistic approaches. Indeed, investigating scales of fluctuation from *in situ* data is a subject of general interest in geotechnical engineering, particularly with respect to the horizontal plane. This is because, although a number of investigations appear in the literature for the vertical scale of fluctuation (e.g. Fenton 1999; Hicks and Onisiphorou 2005), there is still rather limited information for the horizontal direction. This is in spite of the fact that researchers have demonstrated that the ratio of the horizontal and vertical scales of fluctuation is an important consideration in geotechnical computations (Hicks and Samy 2002; Hicks and Onisiphorou 2005; Hicks and Spencer 2010).

The aim of both strategies considered here, for the estimation of the vertical and horizontal scales of fluctuation, is to minimise the error between the assumed *theoretical* correlation model and the *estimated* (or *experimental*) correlation structure (the latter being estimated from CPT data from the site being investigated). In order to explore the performance of each method, an extensive set of CPT data, from an artificial sand island constructed offshore to provide a temporary platform for oil and gas exploration, is considered. In particular, CPT measurements from three vertical cross-sections through the sand fill core of the island are investigated. In Approach A (the first and more conventional approach considered in this study), the CPT data are solely used to estimate the experimental correlation model in the horizontal and vertical directions for each section, whereas, in Approach B, the CPT data are also used to generate a conditional random field from which the experimental correlation model is estimated. It is believed that the use of a conditioned random field makes more complete use of the available site information, particularly when the data are scarce, and so should provide a means of checking the accuracy of conventional estimation techniques. Approach B starts by using the CPTs to statistically describe the tip resistances q_c of the sand fill core of the island. The obtained statistics are then used to generate a two-dimensional (2-D) random field of q_c, which is later constrained (conditioned) at the CPT locations. This new conditional random field is used to estimate the experimental correlation functions for the site (in the horizontal and vertical directions), which are then compared to the respective horizontal and vertical theoretical correlation models to find the estimated values of θ in each direction. Finally, the conventional estimation techniques that are used in Approach A (and which operate on the data directly) are employed to obtain another set of correlation length estimates. The two sets of estimates are then compared to assess the relative accuracy of the two approaches.

2. Approaches used to estimate θ

Various methods are available to estimate the scale of fluctuation. The simplest approach is probably to estimate θ by best fitting the theoretical correlation model to the experimental correlation function (Vanmarcke 1977; Campanella, Wickremesinghe, and Robertson 1987; DeGroot and Baecher 1993; Fenton 1999; Baecher and Christian 2003; Wackernagel 2003; Uzielli, Vannucchi, and Phoon 2005; Fenton and Griffiths 2008). Vanmarcke (1984) and Wickremesinghe and Camapanella (1993) proposed an alternative method, based on the concept of variance function discussed in Vanmarcke (1977), which has been used in several studies (Jaksa, Kaggwa, and Brooker 1993; Hicks and Onisiphorou 2005; Lloret, Hicks, and Wong 2012; Lloret-Cabot, Hicks, and Nuttall 2013). Other techniques, combining random field theory with conventional approaches, have also been recently proposed (Kim and Santamarina 2008; Zhang, Zhang, and Tang 2008; Dasaka and Zhang 2012).

The two approaches used here to estimate θ are based on the concept of best fitting the theoretical correlation model $\rho\ (\tau)$,

$$\rho(\tau) = \exp\left\{\frac{-2|\tau|}{\theta}\right\} \qquad (1)$$

to the estimated correlation function $\hat{\rho}(\tau)$

$$\hat{\rho}(\tau_j) = \frac{1}{\hat{\sigma}^2(k-j)} \sum_{i=1}^{k-j+1} (X_i - \hat{\mu})(X_{i+j} - \hat{\mu}) \qquad (2)$$

where $\hat{\mu}$ and $\hat{\sigma}$ are the estimated mean and standard deviation from the *in situ* CPT data and $\tau_j = j\Delta\tau$, with j = 1, 2, …, k, and k being the number of observations. Note that, for the estimator given by Equation (2), it is desirable that the data be equi-spaced (Fenton and Griffiths 2008) at spacing $\Delta\tau$.

Considering now the following error measure,

$$E = \sum_{j=1}^{k} \left(\hat{\rho}(\tau_j) - \rho(\tau_j)\right)^2 \qquad (3)$$

one may compute the value of θ that minimises E by finding a root to the following expression:

$$\frac{\partial E}{\partial \theta} = -\sum_{j=1}^{k} 2\frac{\tau_j}{\theta^2}\left(\hat{\rho}(\tau_j) - \exp\left\{\frac{-2|\tau_j|}{\theta}\right\}\right)\exp\left\{\frac{-2|\tau_j|}{\theta}\right\} \quad (4)$$

which can be expressed as:

$$\sum_{j=1}^{k} \tau_j\left(\hat{\rho}(\tau_j) - \exp\left\{\frac{-2|\tau_j|}{\theta}\right\}\right)\exp\left\{\frac{-2|\tau_j|}{\theta}\right\} = 0 \quad (5)$$

For simplicity, the correlation model $\rho(\tau)$ is assumed to have the exponential form shown in Equation (1), but alternatives such as those summarised in Table 1 are also possible (Fenton and Griffiths 2008).

In essence, both approaches presented in this paper use the same idea of minimising the error between the assumed theoretical and experimental correlation models. The main difference between Approach A (the conventional approach) and Approach B (the new method proposed) is how the experimental correlation model is estimated. In Approach A, the experimental correlation model $\hat{\rho}(\tau)$ is simply estimated using Equation (2) with the CPT data directly, whereas, in Approach B, $\hat{\rho}(\tau)$ is estimated from the generated conditional random field. A detailed description of how the experimental correlation model is estimated when using Approach B is summarised next.

The algorithm is equivalent in the vertical and horizontal directions and comprises the following steps. Further details are given in the next section where the algorithm is applied to a case study.

(1) Find the linear depth trend of q_c in each CPT considered and remove it from the data. Calculate the standard deviation σ_{res} of the de-trended tip resistances for each CPT. Normalise each individual set of de-trended tip resistances by dividing by the corresponding standard deviation σ_{res}. Each individual CPT is de-trended and normalised in order to produce a standard normal field ($\hat{\mu} = 0$, $\hat{\sigma} = 1$).

(2) The correlation function is estimated separately in the vertical and horizontal directions. For the vertical direction, estimate the correlation function for each CPT, using Equation (2) with the normalised de-trended tip resistances. Then estimate the *average* vertical correlation function from the individual vertical correlation functions. For the horizontal direction, estimate the horizontal correlation function for different depths, by using Equation (2) with the corresponding normalised de-trended CPT tip resistances for different horizontal lags. Then average the correlation functions with respect to depth to get the estimated *average* horizontal correlation function for different lags.
Find the initial estimates of the vertical and horizontal scales of fluctuation $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_0$, by using Equation (5) with the averaged correlation functions (this is Approach A). Set i = 1.

(3) Generate the ith standard normal random field of normalised de-trended q_c based on the statistics found in (1) and (2), assuming that the normalised de-trended tip resistances can be represented by a standard normal distribution function.

(4) Constrain the ith random field computed in (3) at the locations of the CPT measurements, i.e. resulting in the ith conditional random field. A brief description of the implemented conditional approach is given in the following section.

(5) Using Equation (2), with $\hat{\mu} = 0$ and $\hat{\sigma} = 1$, compute $\hat{\rho}_i(\tau)$ from the ith conditional random field calculated in (4).

(6) Use $\hat{\rho}_i(\tau)$, computed in (5), to find the root of Equation (5) in the vertical and horizontal directions, giving $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_i$.

(7) Update $i = i + 1$ and go to (3), repeating the process until the number of simulations performed is n.

(8) The final estimates of the vertical and horizontal scales of fluctuation are the average values computed in (6) of $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_i$, from i = 1 to n, where n is the number of simulations performed.

The fact that each conditional random field is constrained at the known CPT measurements implies that the field contains *true* information of the actual soil variability at the site and, therefore, is likely to provide a more realistic estimation of the correlation function and thereby a better estimate of the scales of fluctuation than the initial estimates given in (2) when using the conventional approach (i.e. Approach A).

Table 1. Some common correlation models.

Correlation model	Expression
Gaussian	$\rho(\tau) = \exp\left\{-\pi\left(\frac{\lvert\tau\rvert}{\theta}\right)^2\right\}$
Triangular	$\rho(\tau) = \begin{cases} 1-\frac{\lvert\tau\rvert}{\theta} & \text{if } \lvert\tau\rvert \le \theta \\ 0 & \text{if } \lvert\tau\rvert > \theta \end{cases}$
Spherical	$\rho(\tau) = \begin{cases} 1-1.5\left\lvert\frac{\tau}{\theta}\right\rvert+0.5\left\lvert\frac{\tau}{\theta}\right\rvert^3 & \text{if } \lvert\tau\rvert \le \theta \\ 0 & \text{if } \lvert\tau\rvert > \theta \end{cases}$
Markov	$\rho(\tau) = \exp\left\{\frac{-2\lvert\tau\rvert}{\theta}\right\}$

2.1. Conditional random fields

The (unconditioned) random fields involved in the conditional Approach B are generated using the Local Average Subdivision (LAS) method proposed by Fenton and Vanmarcke (1990). The LAS method requires a probability density function (pdf) with its statistics (mean μ and standard deviation σ) and a scale of fluctuation θ. As mentioned earlier, the statistical information for q_c in this paper is estimated from the available field data at each 2-D section investigated.

The generated 2-D random fields are then constrained (i.e. conditioned) at the locations of the actual CPT measurements. The conditioning approach follows the work of van den Eijnden and Hicks (2011), which applied the Kriging interpolation technique (Krige 1951; Cressie 1990; Wackernagel 2003; Fenton and Griffiths 2008) to give the best linear unbiased estimate of a random field between known data. In essence, the Kriging method estimates a random field Z at desired locations $\mathbf{x}$, from a linear combination of known values of Z at m observation points $\mathbf{x}_\alpha$. The Kriged interpolation of Z at $\mathbf{x}$ (i.e. $Z^*(\mathbf{x})$) can be expressed as follows:

$$Z^*(\mathbf{x}) = \sum_{\alpha=1}^{m} \lambda_\alpha Z(\mathbf{x}_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{m} \lambda_\alpha = 1 \qquad (6)$$

where λ_α are the m unknown weights that are determined by minimising the variance of the difference between the Kriged field Z^* and the original field Z (Wackernagel 2003). Kriging can be used to condition the random field at the known (conditioning) points, as summarised in the following four steps (Journel and Huijbregts 1978; van den Eijnden and Hicks 2011).

(1) Generate an unconditional random field $Z_s(\mathbf{x})$ with known point statistics and correlation structure, and extract the values of $Z_s(\mathbf{x})$ at the locations $\mathbf{x}_\alpha$ (i.e. $Z_s(\mathbf{x}_\alpha)$ for $\alpha = 1$ to m).

(2) Generate an initial interpolated field $Z_0^*(\mathbf{x})$ by Kriging, using the known (conditioning) measurements $Z(\mathbf{x}_\alpha)$ at the locations $\mathbf{x}_\alpha$ and according to the assumed correlation model.

(3) Generate $Z_s^*(\mathbf{x})$ by Kriging using the values $Z_s(\mathbf{x}_\alpha)$ calculated in step (1).

(4) Calculate the conditional random field $Z_{cs}(\mathbf{x})$ as:

$$Z_{cs}(\mathbf{x}) = \left(Z_s(\mathbf{x}) - Z_s^*(\mathbf{x})\right) + Z_0^*(\mathbf{x}) \qquad (7)$$

3. Application to a real case study

Numerous artificial islands were constructed during the 1970s and 1980s in the Canadian Beaufort Sea to provide temporary structures for hydrocarbon exploration. One type of island used caisson technology to reduce the required fill volumes (Hicks and Smith 1988). Figure 1a shows that this type of island incorporated two main sand fills: (1) an underwater berm on which the caisson was founded and (2) the body of the island structure (referred to as the core). This paper investigates data from one such island, Tarsuit P-45. In particular, 18 CPTs from the site are used here to statistically describe the tip resistances q_c of the sand fill core, these CPTs lying along three straight lines in a plan view of the core, as shown in Figure 1b. The number of CPTs aligned along the first, second and third sections are seven, six and five, respectively, and each line of CPTs indicates the soil variability for that 2-D section (denoted as AA′, BB′ and CC′ in Figure 1b). For simplicity, the same geometry of 50 m length by 5.5 m depth is considered for all three sections. Note that, in order to be consistent in the geometry of all three sections analysed, all CPT

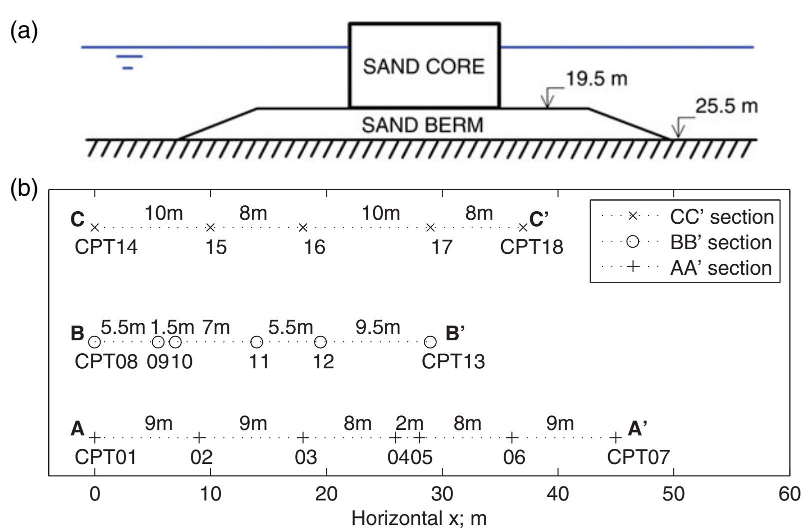

Figure 1. Test site: (a) side view sketch of Tarsuit P-45 core and berm (not to scale) and (b) plan view of CPT locations used.

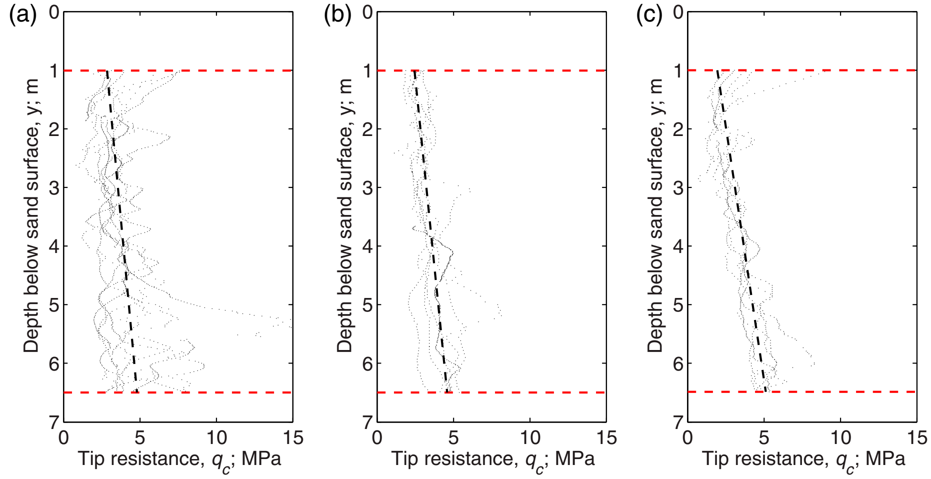

Figure 2. CPT tip resistance data, including the average linear mean trend lines: (a) section AA′, (b) section BB′ and (c) section CC′.

measurements investigated are located in a depth range of 1–6.5 m (see Figure 2).

The statistics obtained for each section are used to generate a 2-D random field of normalised de-trended q_c, which is later constrained (conditioned) at the corresponding CPT locations. The statistical characterisation of the sand fill core of Tarsuit P-45 follows previous research by Wong (2004) and is only briefly summarised below.

Figure 2 shows the CPT tip resistance data for each section investigated. In the plots, the thin broken lines indicate q_c values for individual CPT profiles, whereas the thicker straight dashed lines indicate the average linear mean trend between 1 and 6.5 m. The mean and standard deviation of q_c are calculated separately for each section to give the average values summarised in Table 2. Inspection of Figure 2 shows that the average linear depth trend is very similar for the three sections, indicating a similar underlying depth-dependency of the q_c values. This is also illustrated in Table 2, where the average slope and intercept of the linear trend identified in each section are similar.

A standard normal distribution is used to represent the normalised de-trended cone tip resistances of the Tarsuit P-45 core. Figure 3 shows the histograms based on all data from the CPTs involved in the section analysed, as well as the fitted distribution. Inspection of this figure shows that, for the three sections investigated, the variation of normalised de-trended tip resistance is reasonably well-represented by a standard normal distribution. On the right-hand-side of this figure, the normalised de-trended CPT data used for each histogram are plotted.

The estimates of the vertical and horizontal scales of fluctuation when using Approach A are summarised in Table 3. Note that these are the initial guesses used in Approach B when using the conditional random field. Figures 4 and 5 show the estimated correlation functions from Approach A as dashed lines, for the vertical and horizontal directions, respectively. Note that the correlation estimates become increasingly variable as the lag increases, due to there being fewer data pairs available

Table 2. Cone tip resistance statistics.

Property	Range	Mean value
Section AA′ (7 CPTs)		
Mean (μ): MPa	3.00–5.55	3.85
Standard deviation (σ): MPa	0.71–3.01	1.50
Standard deviation (σ_{res}): MPa (trend removed)	0.69–2.37	1.35
Slope of the linear depth trend (a_{trend}): MPa/m	0.11–1.18	0.36
Intercept of the linear depth trend (b_{trend}): MPa	1.09–3.72	2.48
Section BB′ (6 CPTs)		
Mean (μ): MPa	2.70–4.55	3.58
Standard deviation (σ): MPa	0.40–1.28	0.85
Standard deviation (σ_{res}): MPa (trend removed)	0.39–1.01	0.55
Slope of the linear depth trend (a_{trend}): MPa/m	0.04–0.68	0.39
Intercept of the linear depth trend (b_{trend}): MPa	1.20–2.72	2.06
Section CC′ (5 CPTs)		
Mean (μ): MPa	3.37–3.86	3.59
Standard deviation (σ): MPa	0.75–1.68	1.29
Standard deviation (σ_{res}): MPa (trend removed)	0.51–1.51	0.85
Slope of the linear depth trend (a_{trend}): MPa/m	0.33–0.84	0.58
Intercept of the linear depth trend (b_{trend}): MPa	0.62–2.31	1.36

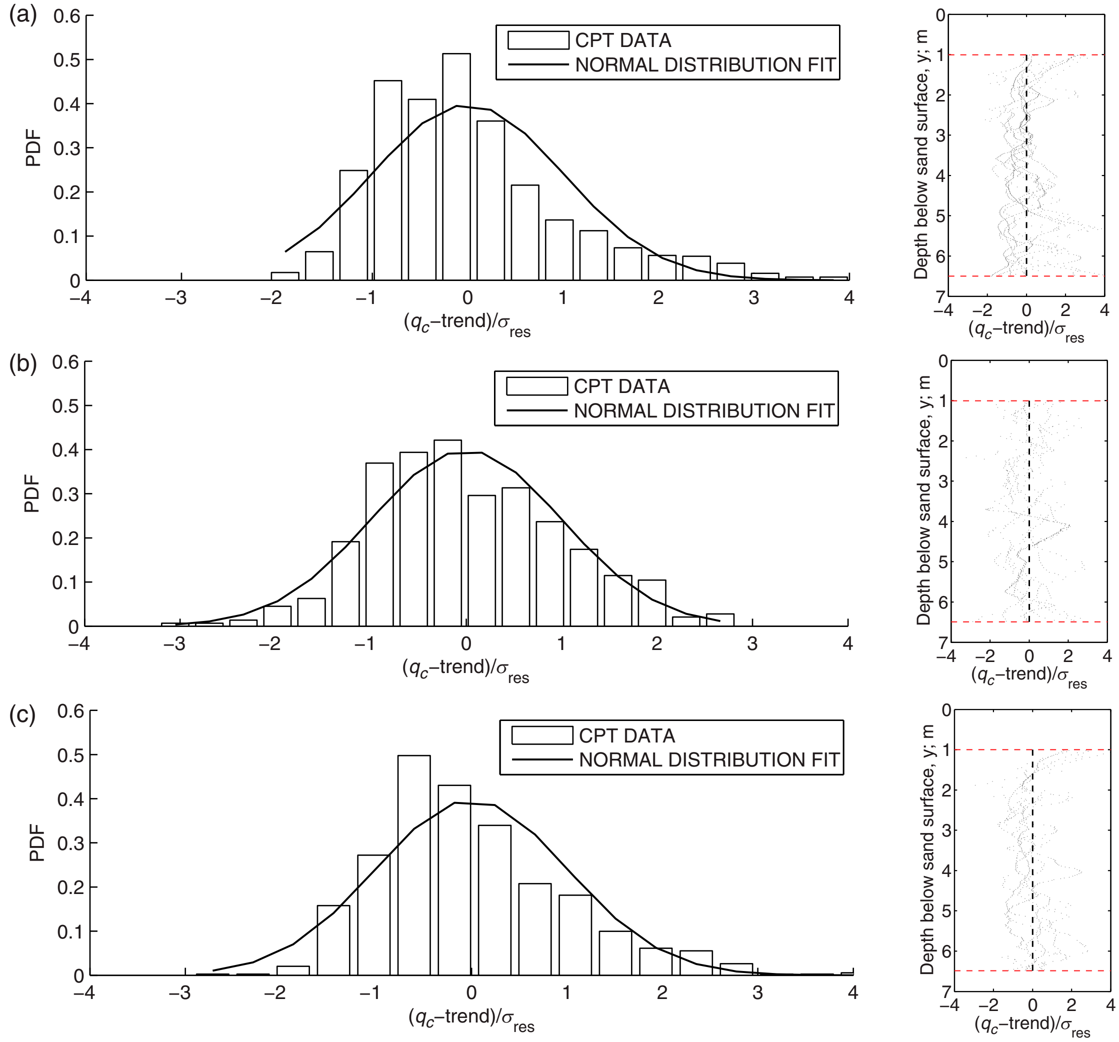

Figure 3. Histograms of normalised de-trended tip resistance: (a) section AA′, (b) section BB′ and (c) section CC′.

with larger lags (Fenton and Griffiths 2008). This is particularly evident for section CC′, as well as for later simulations in the paper. The theoretical correlation function (using the estimated value of θ from Approach A) is represented by a thick solid line. Inspection of Figure 4 shows that very similar initial estimates of θ_v are obtained for the three sections (see also Table 3), indicating that this part of the sand fill island core exhibits a consistent vertical variability of q_c. However, as shown in Figure 5, this is not apparent for the horizontal direction, where the differences between initial estimates for θ_h are much larger and range from 1.69 m to 13.69 m. Although this large range of values may be due in part to actual soil variation, the scarcity of data will also lead to large variability in the estimates.

Table 3. Estimated values of the scales of fluctuation.

Property	Approach A	Approach B
Section AA′ (7 CPTs)		
Vertical scale of fluctuation (θ_v): m	0.42	0.41
Horizontal scale of fluctuation (θ_h): m	1.69	1.82
Section B′ (6 CPTs)		
Vertical scale of fluctuation (θ_v): m	0.42	0.42
Horizontal scale of fluctuation (θ_h): m	5.07	5.60
Section CC′ (5 CPTs)		
Vertical scale of fluctuation (θ_v): m	0.44	0.40
Horizontal scale of fluctuation (θ_h): m	13.69	15.86

A 2-D standard normal random field is generated for each section analysed, using the initial values of the scales of fluctuation obtained from Approach A (see Table 3). Each generated random field is subsequently conditioned at the observed CPT locations by the CPT data, yielding conditional random fields similar to those illustrated in Figure 6. Note that, in the plots of Figure 6, the scales in the vertical and horizontal directions are not the same.

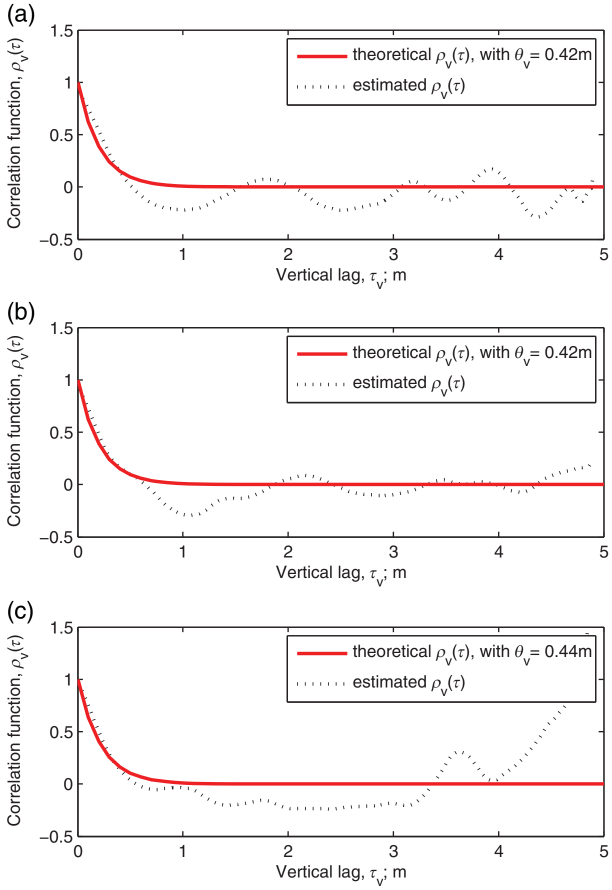

Figure 4. Estimated values of the vertical scale of fluctuation when using Approach A: (a) section AA′, (b) section BB′ and (c) section CC′.

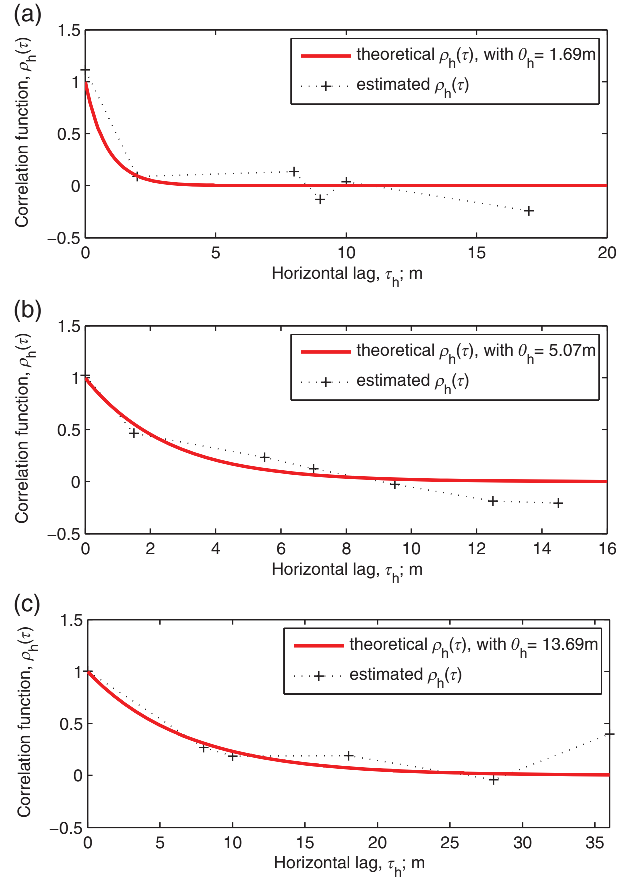

Figure 5. Estimated values of the horizontal scale of fluctuation when using Approach A: (a) section AA′, (b) section BB′ and (c) section CC′.

From each of the conditional random fields, it is straightforward to estimate the corresponding correlation structure by using Equation (2), which can then be compared against the theoretical correlation model in each direction in order to estimate the value of θ (i.e. as a root of Equation (5)). The average of the vertical and horizontal scales of fluctuation, over the total number of realisations n, gives the estimated values of θ_v and θ_h when using Approach B (see Table 3). For the analyses presented in this section, the total number of realisations considered is n = 100. Figures 7 and 8 illustrate the estimated correlation function for the vertical and horizontal directions, respectively, for all realisations when using Approach B. In the figures, the thicker solid line indicates the theoretical correlation structure $\rho\ (\tau)$ using the estimated value of θ; the thinner fine lines indicate each of the estimated $\hat{\rho}_i(\tau)$ and the thick dashed line indicates the average of all the estimated $\hat{\rho}_i(\tau)$.

Overall, Figure 7 shows that the theoretical correlation structure is a satisfactory fit to $\hat{\rho}(\tau)$ for the three sections considered. The results for section AA′ in the vertical direction (Figure 7a) suggest an average of θ_v = 0.41 m, while for section BB′ (Figure 7b) the average is θ_v = 0.42 m, and for section CC′ (Figure 7c) the average is θ_v = 0.40 m. The results suggest that the variability in the vertical direction is very consistent across the sections considered (Table 3).

A larger variation is observed in Figure 8 when looking at the estimated horizontal correlation functions for sections AA′, BB′ and CC′. Section AA′ shows an average of θ_h = 1.82 m, whereas sections BB′ and CC′ give, respectively, θ_h = 5.60 m and θ_h = 15.86 m (Table 3). A possible explanation for these differences is that less CPT measurements (i.e. *true* data points) are available in the horizontal direction. Also, the horizontal distance between CPTs is relatively large compared to the obtained θ_h (see Figures 1 and 5), resulting in only a few *true* data points over the initial part of the estimated correlation structure ($\tau_h < \theta_h$), which is, indeed, the most relevant part of the curve when estimating the θ of the correlation model. Conversely, in the vertical direction, *true* measurements are available every 0.02 m (i.e. the vertical distance between each CPT measurement) and this distance is, conveniently, significantly smaller than

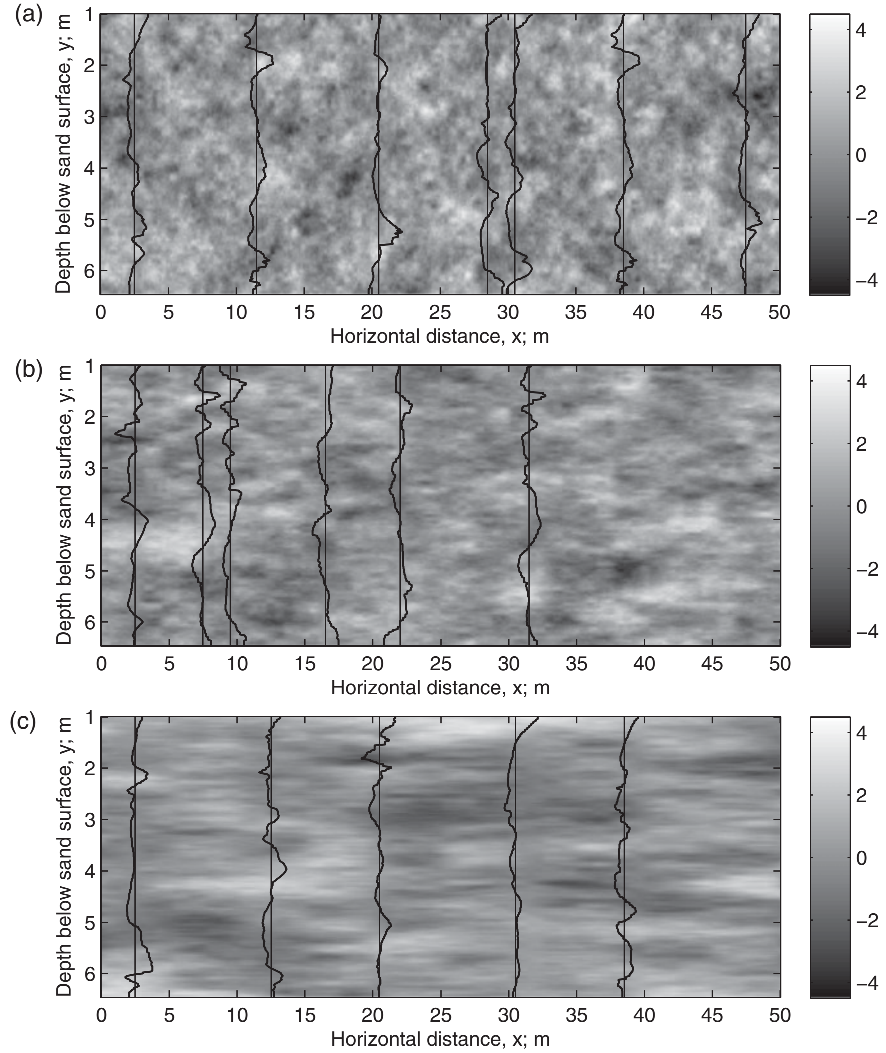

Figure 6. Typical realisation of a conditional random field of normalised de-trended tip resistance, for a 2-D section of the test site: (a) section AA′, (b) section BB′ and (c) section CC′.

the obtained θ_v. This is well illustrated in Figure 4, where many *true* data points are available in the relevant part of the curve (i.e. $\tau_v < \theta_v$), providing more confidence in the estimated value of θ_v than that of θ_h obtained for the horizontal case.

4. Accuracy assessment

A fundamental part of the investigation is to assess the accuracy of the two approaches used to estimate θ. This section aims to address this issue by proposing a numerical strategy and applying it to a fictitious site with the same geometry as analysed in the previous case study. A 2-D random field of normalised de-trended tip resistances is generated with known or *true* statistics ($\mu = 0$, $\sigma = 1$, θ_v= 0.5 m and θ_h= 5 m). From this fictitious site 7 CPTs are extracted at the same locations given in Figure 1b for section AA′. The seven CPTs are then used to calculate the statistics of q_c in the same manner as explained earlier for the case study. Approach B detailed in the previous section is then applied to find estimated values of the vertical and horizontal scales of fluctuation. A number of pairs $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_j$ are obtained by repeating this process from $j = 1$ to k, i.e. over k realisations of the random field of normalised de-trended tip resistances. In order to assess the accuracy of the new approach for estimating θ, the statistics from all pairs $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_j$ can be compared against the *true* values of θ $\{\theta_v, \theta_h\}$ used to generate the initial random tip resistance fields. Similarly, the initial estimated pairs of vertical and horizontal scales of fluctuation $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_{0,j}$, obtained using Approach A, can be used to assess the accuracy of the conventional approach. The steps for assessing the

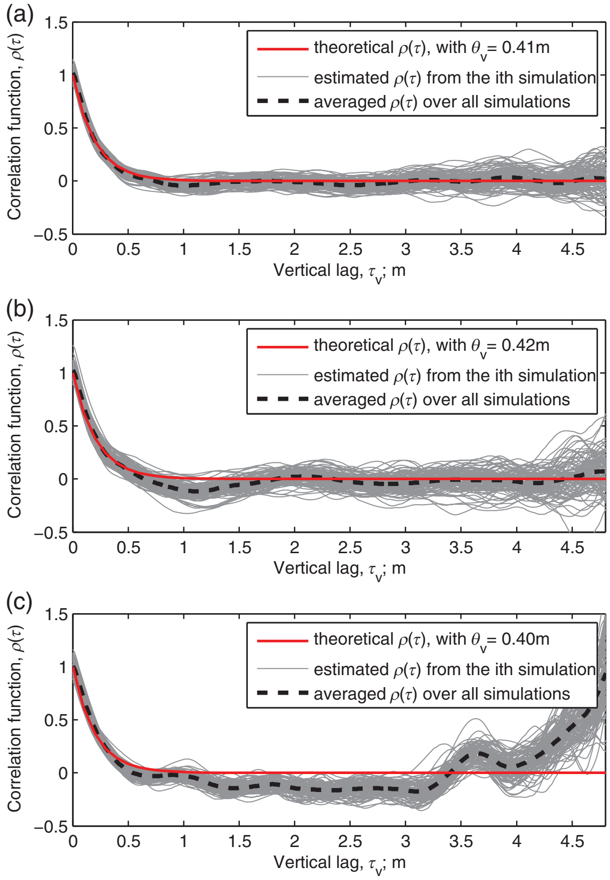

Figure 7. Estimated values of the vertical scale of fluctuation when using Approach B: (a) section AA′, (b) section BB′ and (c) section CC′.

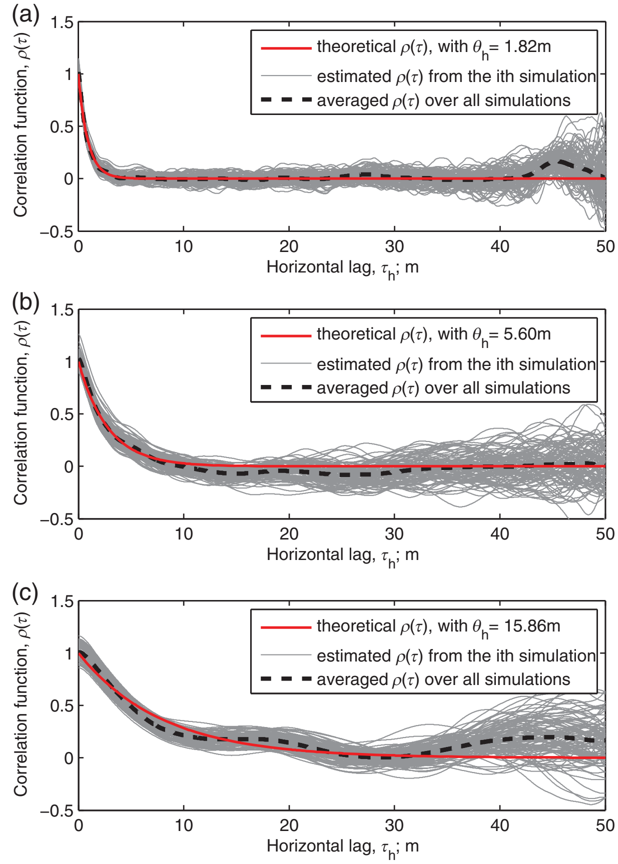

Figure 8. Estimated values of the horizontal scale of fluctuation when using Approach B: (a) section AA′, (b) section BB′ and (c) section CC′.

relative accuracy of the conventional and proposed new approach are summarised as follows:

(1) Set $j = 1$.
(2) Generate a generic (non-conditional) random field of tip resistances with known statistics ($\mu = 0$, $\sigma = 1$, θ_v and θ_h), assuming a standard normal distribution.
(3) Extract l CPTs at the appropriate locations.
(4) Using these l CPTs, estimate the statistics $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_{0,j}$ in the same manner as described in the case study (Approach A).
(5) Estimate $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_j$ using Approach B:
   a. Generate the ith standard normal random field of normalised de-trended q_c based on the statistics found in (4). Set $i = 1$.
   b. Constrain the ith random field computed in (a) at the locations of the CPT measurements from (3), resulting in the ith conditional random field.
   c. Using Equation (2) with $\hat{\mu} = 0$ and $\hat{\sigma} = 1$, compute $\hat{\rho}_i(\tau)$ from the ith conditional random field calculated in (b).
   d. Use $\hat{\rho}_i(\tau)$, computed in (c), to find the root of Equation (5) in the vertical and horizontal directions, giving $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_i$.
   e. Update $i = i + 1$ and go to (a), repeating the process until the number of simulations performed, n.
   f. The final estimates of the vertical and horizontal scales of fluctuation, $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_j$, are the average values computed in (d) of $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_i$ from $i = 1$ to n, where n is the number of simulations performed.
(6) Update $j = j+1$ and go to (2), repeating the process until k realisations have been performed.
(7) Compare the output pairs of values, $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_{0,j}$ and $\left\{\hat{\theta}_v, \hat{\theta}_h\right\}_j$, against the *true* θ_v and θ_h used in (2) to assess the accuracy of the classical and new approaches.

The above steps for assessing the accuracy of the approaches used for the determination of the scales of fluctuation are applied to section AA′ (Figure 1b). Table 4 summarises the relevant information obtained from the

Table 4. Relative error between true and estimated values of θ using the two approaches.

	True values		Approach A		Approach B	
	θ_v(m)	θ_h(m)	$\theta_{v,0}$ (m)	$\theta_{h,0}$ (m)	$\theta_{v,j}$ (m)	$\theta_{h,j}$ (m)
Mean error (μ): m	0.5	5	0.51	3.66	0.53	3.99
Error	–	–	2%	27%	5%	20%

30, i.e. k = 30, random fields generated in the proposed algorithm to assess the accuracy of Approaches A and B. In Approach B, i.e. steps (a) to (f), the number of simulations considered to estimate the final statistics is n = 35. The results presented in this table summarise the estimated values of the scales of fluctuation obtained using Approach A (the conventional approach) and the estimated values obtained using Approach B. The *true* values of the vertical and horizontal scales of fluctuation are included in the table and are used to assess the accuracy achieved in the estimates provided by each method.

The average of the 30 initial estimates of θ_v (i.e. Approach A) is 0.51 m, giving a relative error of about 2% (Table 4). Similar values are obtained when using Approach B: an average θ_v = 0.53 m, giving a relative error of about 5%. In other words, in the vertical direction where data are plentiful, both approaches give accurate results. In the horizontal direction, the results obtained when using Approach B are significantly better than those obtained via the conventional approach. Specifically, when using Approach A the average is θ_h = 3.66 m and the relative error is about 27%, whereas, when using Approach B, the average is θ_h = 3.99 m and the relative error is now about 20% (Table 4). The decrease in relative error from 27% (Approach A) to 20% (Approach B) is quite significant given the fundamental problems with estimating a scale of fluctuation using a relatively large sampling length and few sample points.

The results of Table 4 show that the conventional approach provides reasonable initial estimates for θ_v and θ_h. Indeed, the values obtained for the vertical scale of fluctuation are extremely successful for both approaches, due to the large amount of data available for the calculation of θ_v. However, some improvement is obtained with Approach B in the horizontal direction, when fewer data are available. The better match to the *true* horizontal scale of fluctuation may be due to the algorithm using the available site information more effectively (Lloret-Cabot, Hicks, and van den Eijnden 2012). By constraining the random fields, at the locations of the actual CPT measurements, improved approximations of the q_c values in between the CPT locations are possible, resulting in a more realistic estimation of the horizontal correlation function and a better estimation of the average θ_h (Table 4).

5. Conclusions

Two approaches for estimating the vertical and horizontal scales of fluctuation have been presented and subsequently applied to a real case study and then to a simulation-based study to assess relative accuracy.

The accuracy of the estimate of the scale of fluctuation is, of course, highly dependent on both the number of data and their spacing. For example, if the true correlation length is 0.1 m and data are spaced by 1.0 m, then an accurate estimate of the correlation length will not be possible. Similarly, if a small number of observations, at any spacing, are available, the estimate will be worse than if a large number of observations are available. In the case study considered in this paper, the vertical scale of fluctuation is expected to be estimated much more accurately than the horizontal scale of fluctuation, due to both the much larger number of observations and the closer spacing of the data in the vertical direction. However, the goal of the paper was to see if different methods could be used to coax a better estimate when samples are scarce.

For the case study, the vertical and horizontal correlation lengths suggested by both approaches are similar. In particular, the estimated values of θ_v are very close for the three sections analysed, suggesting that the variability in the vertical direction is very consistent across the sections considered. This is not the case, however, for the horizontal scale of fluctuation, where each section converges to a significantly different mean value, suggesting that θ_h has different values at each section analysed and/or that the CPTs are not spaced closely enough for an accurate estimation of θ_h.

The simulation-based study suggests that there is not much difference between the two approaches when the sampling distance is small relative to the correlation length, as there are then plenty of data for estimating θ (i.e. in the vertical direction). This confirms the finding in the case study that, for the vertical direction, the conventional approach already provides a reasonable estimate of θ_v, because enough data are already available. However, when the sampling distance is large relative to the correlation length and there are few data values (e.g. in the horizontal direction), the conditional random field approach shows some improvement over the conventional approach, with the horizontal correlation length being somewhat closer to the *true* value. The difference is quite significant, with the relative error decreasing from 27% in the case of the conventional approach to 20% in the case of the conditional random field approach, which is a quite remarkable improvement given the fundamental problems with estimating a scale of fluctuation using a relatively large sampling length and few sample points.

The results of this study indicate that, for most practical purposes, the conventional approach to estimating the spatial correlation length is adequate, especially when large amounts of data are available. However, when some improvement is desired, particularly when data are scarce, the use of conditional random fields is worth considering.

Acknowledgement

The authors gratefully acknowledge the support provided by the EU FP7 programme under the project 'Geo-Install' (PIAP-GA-2009-230638).

References

Baecher, G. B., and J. T. Christian. 2003. *Reliability and Statistics in Geotechnical Engineering*. West Sussex: John Wiley & Sons.

Campanella, R. G., D. Wickremesinghe, and P. K. Robertson. 1987. "Statistical Treatment of Cone Penetrometer Test Data." In *Proceedings of the 5th International Conference on Application of Statistics and Probability in Soil and Structure, Vancouver*, edited by D. I. Blockley, 1011–1019. Vancouver, BC.

Cassidy, M. J., M. Uzielli, and Y. Tian 2013. "Probabilistic Combined Loading Failure Envelopes of a Strip Footing on Spatially Variable Soil." *Computers and Geotechnics* 49: 191–205. doi:10.1016/j.compgeo.2012.10.008.

Cressie, N. 1990. "The Origins of Kriging." *Mathematical Geology* 22 (3): 239–252. doi:10.1007/BF00889887.

Dasaka, S. M., and L. M. Zhang. 2012. "Spatial Variability of in Situ Weathered Soil." *Geotechnique* 62 (5): 375–384. doi:10.1680/geot.8.P.151.3786.

DeGroot, D. J., and G. B. Baecher. 1993. "Estimating Autocovariance of In-situ Soil Properties." *ASCE Journal of Geotechnical Engineering* 119 (1): 147–166. doi:10.1061/(ASCE)0733-9410(1993)119:1(147).

Fenton, G. A. 1999. "Estimation for Stochastic Soil Models." *ASCE Journal of Geotechnical and Geoenvironmental Engineering* 125 (6): 470–485. doi:10.1061/(ASCE)1090-0241(1999)125:6(470).

Fenton, G. A., and D. V. Griffiths. 2003. "Bearing Capacity Prediction of Spatially Random $C - \phi$ Soils." *Canadian Geotechnical Journal* 40 (1): 54–65. doi:10.1139/t02-086.

Fenton, G. A., and D. V. Griffiths. 2005. "Three-dimensional Probabilistic Foundation Settlement." *ASCE Journal of Geotechnical and Geoenvironmental Engineering* 131 (2): 232–239. doi:10.1061/(ASCE)1090-0241(2005)131:2(232).

Fenton, G. A., and D. V. Griffiths. 2008. *Risk Assessment in Geotechnical Engineering*. Hoboken, NJ: John Wiley & Sons.

Fenton, G. A., and E. H. Vanmarcke. 1990. "Simulation of Random Fields Via Local Average Subdivision." *ASCE Journal of Engineering Mechanics* 116 (8): 1733–1749. doi:10.1061/(ASCE)0733-9399(1990)116:8(1733).

Griffiths, D. V., J. S. Huang, and G. A. Fenton. 2009. "Influence of Spatial Variability on Slope Reliability Using 2-D Random Fields." *ASCE Journal of Geotechnical and Geoenvironmental Engineering* 135 (10): 1367–1378. doi:10.1061/(ASCE)GT.1943-5606.0000099.

Hicks, M. A., and C. Onisiphorou. 2005. "Stochastic Evaluation of Static Liquefaction in a Predominantly Dilative Sand Fill." *Geotechnique* 55 (2): 123–133. doi:10.1680/geot.2005.55.2.123.

Hicks, M. A., and K. Samy. 2002. "Influence of Heterogeneity on Undrained Clay Slope Stability." *Quarterly Journal of Engineering Geology and Hydrogeology* 35 (1): 41–49. doi:10.1144/qjegh.35.1.41.

Hicks, M. A., and I. M. Smith. 1988. "Class A prediction of Arctic Caisson Performance." *Geotechnique* 38 (4) 589–612. doi:10.1680/geot.1988.38.4.589.

Hicks, M. A., and W. A. Spencer. 2010. "Influence of Heterogeneity on the Reliability and Failure of a Long 3D Slope." *Computers and Geotechnics* 37 (7–8): 948–955. doi:10.1016/j.compgeo.2010.08.001.

Jaksa, M. B., W. S. Kaggwa, and P. I. Brooker. 1993. "Geostatistical Modelling of the Undrained Shear Strength of Stiff, Overconsolidated, Clay." In *Proceedings of Conference on Probabilistic Methods in Geotechnical Engineering, Canberra*, edited by K. S. Li and S.-C. R. Lo, 185–194. Rotterdam, Brookfield: A.A. Balkema.

Journel, A. G., and Ch. J. Huijbregts. 1978. *Mining Geostatistics*. New York: Academic Press.

Kim, H.-K., and J. C. Santamarina. 2008. "Spatial Variability: Drained and Undrained Deviatoric Load Response." *Geotechnique* 58 (10): 805–814. doi:10.1680/geot.2008.3724.

Krige, D. G. 1951. "A Statistical Approach to Some Basic Mine Valuation Problems on Witwatersrand." *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52 (6): 119–139.

Lloret, M., M. A. Hicks, and S. Y. Wong. 2012. "Soil Characterisation of an Artificial Island Accounting for Heterogeneity." In *Proceedings of GeoCongress 2012, San Francisco*, edited by R. D. Hryciw, A. Athanastopoulos-Zekkos, and N. Yesiller, 2816–2825. Reston: ASCE.

Lloret-Cabot, M., M. A. Hicks, and J. D. Nuttall. 2013. "Investigating the Scales of Fluctuation of an Artificial Sand Island." In *Proceedings of the International Conference. Geotechnical Installations in Geotechnical Engineering 2013*, *Rotterdam*, edited by M. A. Hicks, J. Dijkstra, M. Lloret-Cabot, and M. Karstunen, 192–197. Rotterdam: CRC Press Taylor and Francis Group.

Lloret-Cabot, M., M. A. Hicks, and A. P. van den Eijnden. 2012. "Investigation of the Reduction in Uncertainty Due to Soil Variability When Conditioning a Random Field Using Kriging." *Geotechnique letters* 2 (3): 123–127. doi:10.1680/geolett.12.00022.

Phoon, K.-K., and F. H. Kulhawy. 1999. "Characterization of Geotechnical Variability." *Canadian Geotechnical Journal* 36 (4): 612–624. doi:10.1139/t99-038.

Uzielli, M., G. Vannucchi, and K.-K. Phoon. 2005. "Random Field Characterisation of Stress-normalised Cone Penetration Testing Parameters." *Geotechnique* 55 (1): 3–20. doi:10.1680/geot.2005.55.1.3.

van den Eijnden, A. P., and M. A. Hicks. 2011. "Conditional Simulation for Characterizing the Spatial Variability of Sand State." In *Proceedings of the 2nd International Symposium on Computational Geomechanics, Dubrovnik*, edited by S. Pietruszczak and G. N. Pande, 288–296. Rhodes, Greece: IC2E.

Vanmarcke, E. H. 1977. "Probabilistic Modeling of Soil Profiles." *ASCE Journal of the Geotechnical Engineering* 103 (11): 1227–1246.

Vanmarcke, E. H. 1984. *Random Fields: Analysis and Synthesis*. Cambridge, MA: The MIT Press.

Wackernagel, H. 2003. *Multivariate Geostatistics: An Introduction with Applications*. Germany: Springer.

Wickremesinghe, D., and R. G. Campanella. 1993. "Scale of Fluctuation as a Descriptor of Soil Variability." In *Proceedings of Conference on Probabilistic Methods in Geotechnical Engineering*, Canberra, edited by K. S. Li and S.-C. R. Lo, 233–239. Rotterdam: A.A. Balkema.

Wong, S. Y. 2004. "Stochastic Characterisation and Reliability of Saturated Soils." PhD thesis, University of Manchester.

Zhang, L. L., L. M. Zhang, and W. H. Tang. 2008. "Similarity of Soil Variability in Centrifuge Models." *Canadian Geotechnical Journal* 45 (8): 1118–1129. doi:10.1139/T08-066.