Assessment of soil spatial variability for linear infrastructure using cone penetration tests

De Gast, Tom; Vardon, Philip J.; Hicks, Michael A.

DOI
10.1680/jgeot.19.SiP.002

Publication date
2021

Document Version
Final published version

Published in
Geotechnique

Citation (APA)
De Gast, T., Vardon, P. J., & Hicks, M. A. (2021). Assessment of soil spatial variability for linear infrastructure using cone penetration tests. *Geotechnique*, *71*(11), 999-1013. https://doi.org/10.1680/jgeot.19.SiP.002

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

de Gast, T. *et al.* (2021). *Géotechnique* **71**, No. 11, 999–1013 [https://doi.org/10.1680/jgeot.19.SiP.002]

Assessment of soil spatial variability for linear infrastructure using cone penetration tests

TOM DE GAST*, PHILIP J. VARDON† and MICHAEL A. HICKS*

Soil spatial variability has a significant impact on the reliability of geotechnical structures. In particular, the horizontal variability is important for linear infrastructure, which has only limited vertical height and width, but extensive length. Due to depositional and geological processes, the variability is often substantially different in the vertical and lateral directions. This variability can be characterised by a spatial correlation length, or scale of fluctuation, which is a measure of how significantly soil properties are correlated in space. An analysis of the reliability of such a measure has been undertaken using synthetic data, leading to a design chart which quantifies the statistical uncertainty in the scale of fluctuation for specific site investigation designs, which can be an important input for probabilistic analyses of the structure response. Moreover, practical guidance for site investigation design is proposed which can reduce the statistical uncertainty. The method has been applied to a real site investigation comprising a row of 29 closely spaced cone penetration tests (CPTs), within a larger site investigation of 100 CPTs, and applied to a simple design calculation for a long embankment to illustrate the impact on slope stability assessment. The site investigation data are made available to add to the limited amount of detailed data in this field.

KEYWORDS: embankments; site investigation; statistical analysis

INTRODUCTION

Soil variability is a major issue in geotechnical engineering (Honjo, 2011); it typically causes uncertainty in the interpretation of site investigations and has important, but different, implications for different geotechnical structures. The influence of spatial variability has been studied, by way of the use of theoretical and numerical models, in relation to shallow, strip and pile foundations (Jaksa *et al.*, 2005; Suchomel & Masín, 2010; Naghibi *et al.*, 2016), retaining walls (Sert *et al.*, 2016), liquefaction of hydraulic sand fills (Wong, 2004; Hicks & Onisiphorou, 2005; Popescu *et al.*, 2005) and slope stability (Griffiths & Fenton, 1997; Hicks & Samy, 2002; Hicks & Spencer, 2010; Li *et al.*, 2016). It has been shown, especially for embankments (Spencer & Hicks, 2007; Hicks & Spencer, 2010; Hicks *et al.*, 2014; Hicks & Li, 2018), that the spatial variability of material properties, the problem geometry and the combination of the two are significant factors influencing the stability of slopes and the failure mode. The impact of the spatial variability in the vertical and horizontal directions has been investigated by Hicks & Spencer (2010), Li *et al.* (2015) and Varkey *et al.* (2018), who showed that, for embankments, the horizontal spatial variability has a significant influence on the type of failure mechanism.

The theoretical quantification of the spatial variability of soils has been extensively reviewed by Vanmarcke (1977), Campanella *et al.* (1987) and Wickremesinghe & Campanella (1993). As a measure of the distance over which data are significantly correlated, the scale of fluctuation (SoF) can be estimated using a range of techniques; in particular, by using an auto-correlation function, Fourier analysis or a variogram function. These techniques require a substantial amount of data, such as obtained by cone penetration tests (CPTs). While CPTs do not directly measure soil properties, their measurements are known to be correlated to them. For example, the tip resistance is known to be correlated to the undrained shear strength by way of a linear transformation including the effect of increase in confining pressure (Robertson *et al.*, 1986). Moreover, as the spatial variability is due to geological deposition processes, it is generally assumed that all soil parameters would have the same spatial variability (e.g. Kawa & Puła, 2020). Therefore, by utilising the spatial variability of the CPT measurements, an estimate of the spatial variability of the soil properties can be made without having to transform the measurement data. However, relatively few data for determining the SoF, especially in the horizontal plane, are available. Some notable attempts in the quantification of the horizontal SoF from CPTs are Jaksa *et al.* (1999), Lloret-Cabot *et al.* (2014), Ching *et al.* (2018) and Fenton *et al.* (2018). In particular, Ching *et al.* (2018) utilise limited vertical CPTs to estimate the SoF, and show that to obtain a good estimate of the horizontal SoF, the CPT spacing, depth and SoF must be considered.

It has been previously observed that, by taking account of spatial variation and the location of measurement data, the uncertainty in the response of a structure is reduced (e.g. Jaksa *et al.*, 2005; Lloret-Cabot *et al.*, 2012; Li *et al.*, 2016), and given more available information – that is, more CPTs – the better is the estimation of the SoF (Lloret-Cabot *et al.*, 2014). However, the decrease in uncertainty in the estimated SoF as the number of CPTs, or amount of data per CPT, increases has not yet been quantified. A challenge in the calculation of reliable horizontal scales of fluctuation is

Manuscript received 8 March 2019; revised manuscript accepted 7 October 2020. Published online ahead of print 16 December 2020. Discussion on this paper closes on 1 March 2022, for further details see p. ii.

* Section of Geo-Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands.

† Section of Geo-Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands (Orcid:0000-0001-5614-6592).

that, unlike using CPTs in the vertical direction, there are no standard methods of geotechnical investigation giving high spatial resolution in the horizontal plane. Clearly, the required spacing of collected data is important and it strongly influences the cost of a site investigation.

The origin of spatial variability is often attributed to geological processes occurring in the deposition of soil layers (Phoon & Kulhawy, 1999). Given that there may be many processes involved in the transport and deposition of sediments, it may be hypothesised that there may be multiple, separate, scales of fluctuation (e.g. Vanmarcke, 1983). For example, a large-scale variation may occur due to seasonal changes in the deposition, whereas a small-scale variation may be due to local hydrodynamic processes – for example, eddy currents.

Here, the auto-correlation method is used to investigate the assessment of the SoF in a specific direction – that is, making it particularly suitable for, but not limited to, assessing the SoF for linear infrastructure. This method has been shown to give results comparable to other methods (de Gast, 2020). Computer-generated data, with a known SoF, have been used to investigate the impact of having differing amounts of site investigation data, and the statistical uncertainty in the back-figured SoF has been quantified. A design chart and underlying equation have then been proposed to estimate the uncertainty arising from different testing schemes. A row of 29 closely spaced CPTs, from within a larger field test of 100 CPTs, from a regional dyke in the Netherlands, has been analysed and the resulting data used to validate the proposed approach. This field test has been used to demonstrate the consequences of collecting different amounts of data during site investigations. A simplified dyke analysis has been undertaken using the detected variability from the field test to demonstrate the impact of the uncertainty on the predicted dyke response.

THEORETICAL BACKGROUND

The properties of soil are spatially variable, but are generally related to the properties of nearby material – that is, they are spatially correlated. The most usual correlation measure is the auto-correlation length, θ, also referred to as the SoF. It is the integral of the auto-correlation function, and approximates the distance over which material properties are significantly correlated. θ has been defined by Vanmarcke *et al.* (1986) as

$$\theta = 2\int_0^\infty \rho(\tau)\mathrm{d}\tau \tag{1}$$

where $\rho(\tau)$ is the auto-correlation function or structure and τ is the distance separating two points, otherwise known as the lag distance. Hence, θ is the area under the auto-correlation function over the range $-\infty \leq \tau \leq \infty$, and, while it can be different in any direction, owing to depositional processes it is commonly considered to be unique in the vertical and horizontal directions.

Note that other techniques can also be used to describe the spatial correlation of data. For example, in the mining industry spatial auto-correlations are commonly described by semi-variograms, which are constructed using the squared differences in the property of interest at different lag lengths. According to Baecher & Christian (2003), the semi-variogram requires a less restrictive statistical assumption on stationarity (i.e. the requirement of a constant mean and standard deviation) than the auto-correlation function. However, its use in spatial interpolation and engineering can be more difficult, as the range, sill and nugget need to be identified. Therefore, the auto-correlation approach is used in this paper.

Table 1. Common theoretical auto-correlation models (Fenton, 1999)

Correlation model	Formula
Gaussian	$\rho(\tau) = \mathrm{e}^{-\pi(\lvert\tau\rvert/\theta)^2}$
Triangular	$\rho(\tau) = \begin{cases} 1 - \lvert\tau\rvert/\theta & \text{if } \lvert\tau\rvert \leq \theta \\ 0 & \text{if } \lvert\tau\rvert > \theta \end{cases}$
Spherical	$\rho(\tau) = \begin{cases} 1 - (1{\cdot}5\lvert\tau\rvert/\theta) + (\lvert\tau\rvert/\theta)^3 & \text{if } \lvert\tau\rvert \leq \theta \\ 0 & \text{if } \lvert\tau\rvert > \theta \end{cases}$
Markov	$\rho(\tau) = \mathrm{e}^{-2\lvert\tau\rvert/\theta}$

The method used to determine the experimental θ (e.g. Lloret-Cabot *et al.*, 2014) compares a theoretical auto-correlation function $\rho(\tau)$, for an estimated value of θ, with an experimentally determined auto-correlation function $\hat{\rho}(\tau)$. The error in approximating the experimental auto-correlation function by a theoretical function is given by

$$\mathrm{Er}(\rho) = \sum_i [\rho(\tau_i) - \hat{\rho}(\tau_i)]^2 \tag{2}$$

where $\mathrm{Er}(\rho)$ is the error in the approximation and is defined as the sum of the squared differences between the values of the two functions at the lag distances considered – that is, corresponding to the distances between the discrete data points. The experimental θ is the value of θ that minimises $\mathrm{Er}(\rho)$.

Theoretical auto-correlation function

The theoretical auto-correlation function (or correlation structure) can take various forms, as illustrated by Table 1, which lists some common auto-correlation models. The model which best fits the data should be selected, and herein the Markov model has been used.

Experimental auto-correlation function

The experimental auto-correlation function is the auto-correlation function derived from discrete (e.g. measurement) data, and can be obtained from the experimental covariance function $\hat{\gamma}(\tau)$ as

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)} \tag{3}$$

The experimental covariance function was presented by Vanmarcke (1983) for equally spaced data as

$$\hat{\gamma}(\tau) = \frac{1}{k-l}\sum_{j=1}^{k-l}(y_j - \hat{\mu})(y_{j+l} - \hat{\mu}) \tag{4}$$

which can be, as in variance estimation, adapted as an unbiased estimator

$$\hat{\gamma}(\tau) = \frac{1}{k-l+1}\sum_{j=1}^{k-l}(y_j - \hat{\mu})(y_{j+l} - \hat{\mu}) \tag{5}$$

or adapted for unequally spaced data as

$$\hat{\gamma}(\tau) = \frac{1}{t-1}\sum_{j=1}^{t}(y_j - \hat{\mu})(y_{j+\Delta j} - \hat{\mu}) \tag{6}$$

where $\hat{\mu}$ is the estimated mean (or trend) of the dataset; j is a counter representing the index of the first of a data pair at lag

distance τ; $l = 1, 2, \ldots, k$ is a counter representing the index spacing of pairs of data with k being the total number of observations; Δj represents the index spacing of a specific pair of observations for a non-uniformly distributed dataset; and t is the number of pairs at lag distance τ. Note that, for the estimator given by equation (3), it is desirable that the data are equally spaced (Fenton & Griffiths, 2008) at any spacing τ, to ensure results of higher quality. However, this may reduce the number of lag distances considered, and may be unrealistic with measured data as it is not common to have equally (and closely) spaced data. Therefore, equation (6) is suggested to account for non-equidistant spacing. It is also possible to group data into 'bins' to increase the number of pairs, although the resolution will then be affected.

Multiple θ correlation structures

Theoretical auto-correlation functions can be combined to better represent the experimental auto-correlation. For example, a single composite auto-correlation function can be constructed using different scales of fluctuation, as in Fig. 1 (Vanmarcke, 1983), and even by combining different auto-correlation models. In Vanmarcke (1983), an equivalent equation to equation (1) is used to determine an average θ, θ_{avg} – that is

$$\theta_{\text{avg}} = 2\int_0^\infty \rho_{\text{com}}(\tau)\mathrm{d}\tau \tag{7}$$

where ρ_{com} is the composite auto-correlation function constructed as the weighted average of the component auto-correlation functions

$$\rho_{\text{com}}(\tau) = c_1\rho_1(\tau) + c_2\rho_2(\tau) + \cdots + c_i\rho_i(\tau) \tag{8}$$

where

$$0 \le c_i \le 1 \quad \text{and} \quad \sum_i c_i = 1 \tag{9}$$

Integration of equation (8) leads to

$$\int_0^\infty \rho_{\text{com}}(\tau)\mathrm{d}\tau = c_1\int_0^\infty \rho_1(\tau)\mathrm{d}\tau + c_2\int_0^\infty \rho_2(\tau)\mathrm{d}\tau + \cdots + c_i\int_0^\infty \rho_i(\tau)\mathrm{d}\tau \tag{10}$$

and substitution of equation (10) into equation (7) gives

$$\theta_{\text{avg}} = c_1\theta_1 + c_2\theta_2 + \cdots + c_i\theta_i \tag{11}$$

The composite auto-correlation function that most closely fits the experimental auto-correlation function is, as before, found by minimising the error

$$\text{Er}(\rho_{\text{com}}) = \sum_i [\rho_{\text{com}}(\tau_i) - \hat{\rho}(\tau_i)]^2 \tag{12}$$

Once again, the error is only calculated for the lag distances used in the experimental correlation function, as defined by the distances between the discrete data points.

In theory, any number of components can be used in a composite auto-correlation function, although the number of unknowns that need to be determined is twice the number of components, and the computational effort increases significantly with each additional component. Therefore, the examples here are limited to one- or two-component models. The proposed algorithm to determine two scales of fluctuation from CPT data (e.g. cone tip resistance) is based on finding the combination of θ_1 and θ_2 that gives the minimum error between the theoretical and experimental auto-correlation functions, and is summarised as follows.

(*a*) Detrend the data with respect to depth.
(*b*) Calculate the experimental auto-correlation function using equations (3) and (6).
(*c*) Calculate c_1, c_2, θ_1 and θ_2 using the following algorithm.
  (*i*) Loop through $0 < c_1 \le 1$ in small steps
    • Using equation (9), $c_2 = 1 - c_1$
    (*ii*) Loop through $0 < \theta_1 \le \theta_{\max}$ in small steps ($\theta_{\max}$ is the maximum θ investigated)
      (*iii*) Loop through $0 < \theta_2 \le \infty$ in small steps (note that a practical approach to approximating ∞ is needed, e. g. $\infty \approx 5D$)
        • Calculate the composite auto-correlation function using equation (8) for each experimental lag distance
        • Calculate the error using equation (12)
      • End loop (*iii*)
    • End loop (*ii*)
  • End loop (*i*)
(*d*) Select the combination of c_1, c_2, θ_1, θ_2 with the smallest error.

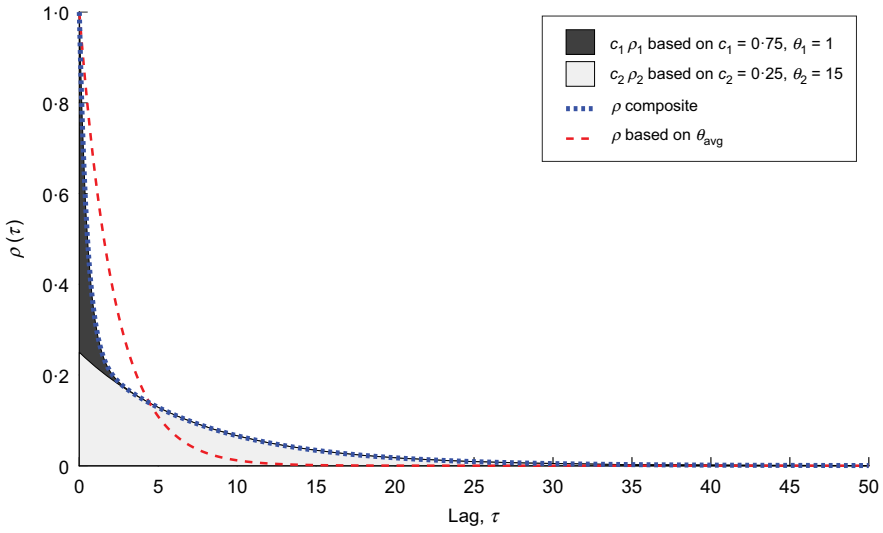

Fig. 1. Auto-correlations functions based on a combination of two theoretical scales of fluctuation and on a corresponding average scale of fluctuation (after Vanmarcke, 1983)

The step size for the algorithm can be selected based on computational effort and the required resolution of the resulting composite θ. However, step sizes of 0·01 for c_1 and c_2, and 0·01 m for θ_1 and θ_2, are recommended as providing sufficient resolution for most practical cases, and it is reasonable for the maximum θ_1 investigated – namely, θ_{max} – to be selected as the domain length, D. A zero correlation length means that the data points are spatially uncorrelated and an infinite correlation length means that they are fully correlated – in other words, constant.

Figure 1 shows an example of combining two auto-correlation functions with different weights and different θ, as well as the auto-correlation function with a single average θ determined using equation (11).

INVESTIGATION USING SYNTHETIC DATA

An investigation using synthetic data was carried out to investigate the efficacy of the method and to quantify the uncertainties. One-dimensional (1D) strings of data (datasets), analogous to CPT profiles, were generated using Cholosky decomposition (Alabert, 1987; Davis, 1987; van den Eijnden & Hicks, 2017), which meant that the θ was known a priori. The data strings, referred to below as synthetic CPTs, were generated using a Markov auto-correlation function and the following properties: domain length, $D = 50$, mean, $\mu = 0$, standard deviation, $\sigma = 1$ and $\theta = 5$ (note that all θ have units of length).

Figure 2(a) shows that, as the lag length increases, the individual experimental auto-correlation functions (thin solid lines) oscillate, due to less available data – that is, fewer pairs. In order to improve the estimate of the auto-correlation function it is helpful to average multiple experimental auto-correlation functions – for example, multiple CPT measurements. The thick dashed line in Fig. 2(a) shows the average of the five experimental auto-correlation functions, which reduces the oscillation and matches better the input theoretical auto-correlation function (thick solid line), especially at shorter lag lengths. This further improves with more data, as shown by the thick dashed line in Fig. 2(b), where the average of 100 experimental auto-correlation functions is used. The (absolute) error, Er, calculated using equation (2) with the input theoretical auto-correlation function, reduces from around 8·0 (with one data string or CPT) to 2·0 (five CPTs) to 0·1 (100 CPTs). Importantly, oscillations are also delayed until larger lag lengths, implying that more data would be needed for larger scales of fluctuation relative to the length of the domain.

The same process was undertaken for data generated using two scales of fluctuation. Once again, the synthetic CPTs were generated based on a mean of 0 and a standard deviation of 1 ($\mu = 0$, $\sigma = 1$), but now with an auto-correlation function built from two Markov auto-correlation models with scales of fluctuation of $\theta_1 = 1$ and $\theta_2 = 15$ and two respective weighting coefficients of $c_1 = 0{\cdot}75$ and $c_2 = 0{\cdot}25$, giving $\theta_{avg} = 4{\cdot}5$. One hundred datasets were generated using these input parameters and the 100 resulting experimental auto-correlation functions were averaged. Fig. 3 shows the result of analysing 100 datasets and compares it to the result based on a single auto-correlation function. (Note that the best single auto-correlation function is not the same as using θ_{avg}.) The double θ model has a much better fit to the experimental auto-correlation function than the single θ model, as confirmed by the respective errors of Er = 0·07 and Er = 0·23.

It is noted that using 100 datasets may not be reasonable in practice, and that, while using multiple scales of fluctuation the solution is more flexible, it is likely that more data would be needed than for identifying a single θ. However, it is also reasonable to assume that, for larger scales of fluctuation relative to the domain size, more data are also needed to calculate the scales of fluctuation accurately. These aspects are investigated in the following section.

Quantification of uncertainty

As the error is larger with fewer data, any predicted θ will have a likely error or statistical uncertainty associated with it. To investigate this uncertainty, a comprehensive numerical investigation has been undertaken, where different variables were investigated: (*a*) number of datasets combined; (*b*) number of data points used per dataset; (*c*) value of θ. The number of datasets is analogous with the number of CPT locations if investigating the vertical scale of fluctuation, θ_v, or is proportional to the depth range tested if investigating the horizontal scale of fluctuation, θ_h. The number of data points used per dataset may be proportional

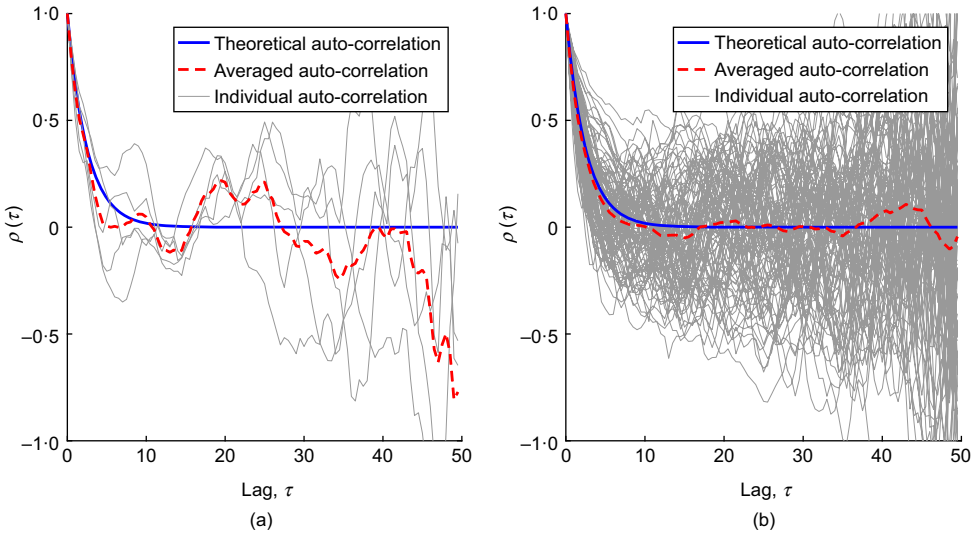

Fig. 2. Individual (thin solid line) and averaged (thick dashed line) experimental auto-correlation functions: (a) based on five individual functions; (b) based on 100 individual functions

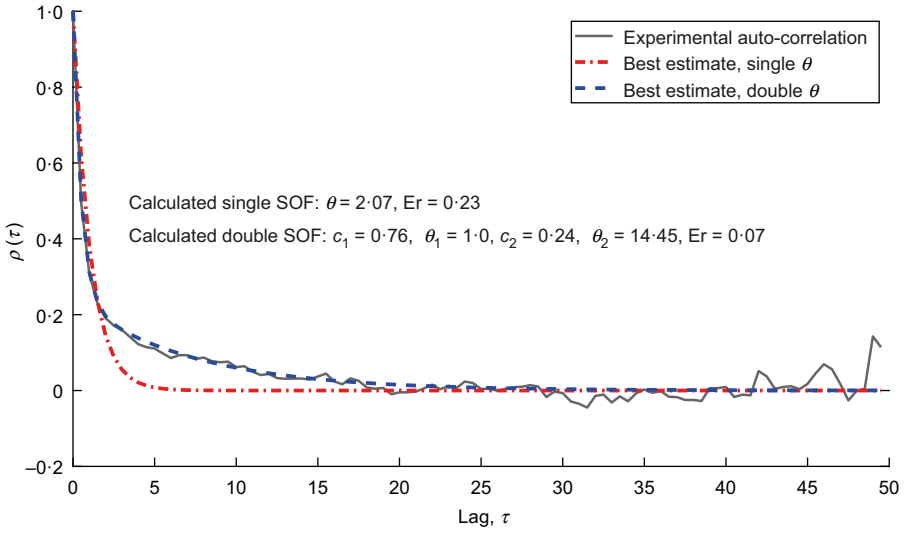

Fig. 3. Experimental auto-correlation function formed by averaging data from 100 realisations generated using two scales of fluctuation, $\theta_1 = 1$ and $\theta_2 = 15$, with respective weighting coefficients of $c_1 = 0{\cdot}75$ and $c_2 = 0{\cdot}25$, compared with derived single and double theoretical auto-correlation functions

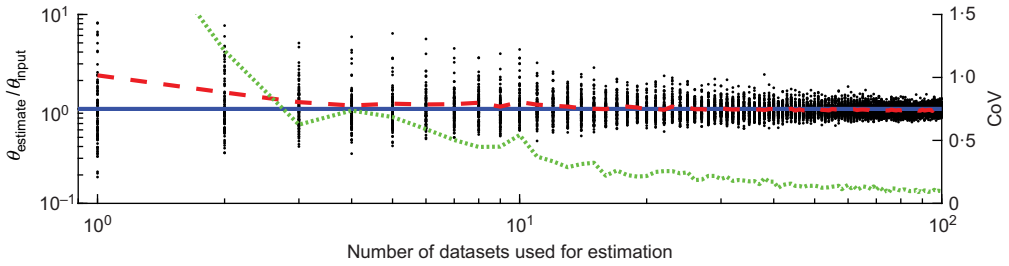

Fig. 4. Accuracy of θ estimation as a function of the number of datasets. θ is normalised to the theoretical value. Each dot is a single estimate based on the number of datasets; the solid line is the theoretical result; the dashed line is the mean result from 100 estimates; and the dotted line is the CoV of the θ estimates

to the layer depth if investigating θ_v, or is analogous with the number of CPT locations if investigating θ_h. The results are presented in Figs 4–7, with most axes being in log scale to enable a good oversight of the results; they are also summarised in Fig. 8 using a linear scale. The results have been obtained using numerically generated datasets, based on a Markov auto-correlation structure with a single θ, $\mu = 0$ and $\sigma = 1$.

To investigate the influence of the number of datasets, between one and 100 combined datasets were considered. The length of each dataset was 50 units, the distance between the data points in each dataset was 0·5, and $\theta = 5$.

In Fig. 4, each dot is an estimate of θ (relative to the input θ), and the result of an analysis with the number of datasets given on the x-axis (i.e. the number of CPT profiles). The datasets that are input into the analysis to generate each dot are randomly selected and therefore represent a possible interpretation of a site investigation. For each number of datasets considered, 100 estimates of θ have been made using the single θ method. As expected, there is a distribution of dots with a higher proportion clustered around the theoretical value of θ, indicated by the solid line. The dashed line is the average of the 100 estimations, and the dotted line is the coefficient of variation, $\text{CoV} = \sigma/u$ (where σ is the standard deviation and μ is the mean) of the estimations of θ. As expected, if more datasets are used, the CoV reduces, and therefore the result is more likely to be accurate. In this example, when considering only one dataset there is a 9·6% chance of the estimated θ being within 20% of the real θ, whereas there is a 33·1% chance of being within 20% when five datasets are used, rising to 70·0% with 100 datasets. In practice, this increase in accuracy would be offset by the additional expense of collecting the extra data and results in an increase in computational effort for the analysis. Fig. 4 also shows that using a limited amount of data is more likely to result in an overestimate of θ; this is because the theoretical auto-correlation is always positive and the distribution does not therefore have negative outliers.

The number of data points per dataset has also been investigated, once again for $\theta = 5$. This was done by again using a distance of 0·5 between data points, and considering the overall domain length to change with the number of data points. For each number of data points considered, 100 estimates of θ, each using 40 datasets, have been produced. Again, as expected, by increasing the number of points the estimation of θ improved (see Fig. 5). Considering only five data points per dataset there is a 27·9% chance of being within 20% of the real θ, whereas with ten data points there is a 33·5% chance of being within 20%, rising to 71·9% with 100 data points.

Using a fixed domain size (length) of 50 and the same number of data points, the ability to detect different values of θ was then investigated. Each estimate of θ was based on 40 datasets and 100 points per dataset. Again, 100 estimates were made. The generated data had an input θ varying between $\theta = 5$ and $\theta = 500$. Fig. 6 shows how the estimated θ

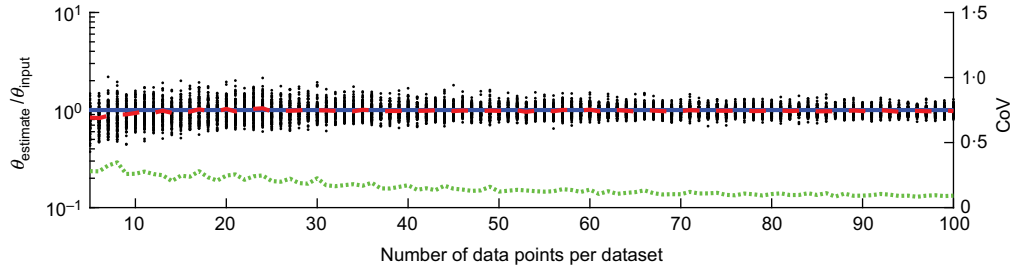

Fig. 5. Accuracy of θ estimation as a function of the number of points in the dataset (the separation distance of the points is constant, and therefore the domain increases as more points are considered). Each dot is a single estimate based on the number of data points per dataset; the solid line is the theoretical result; the dashed line is the mean result from 100 estimates; and the dotted line is the CoV of the θ estimates

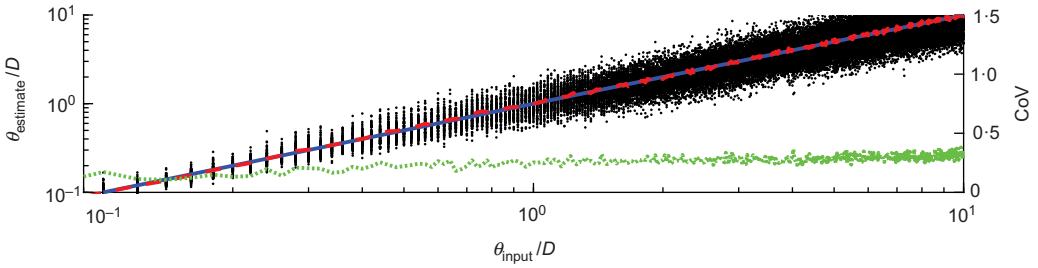

Fig. 6. Accuracy of θ estimation as a function of actual θ relative to domain size D. Each dot is a single estimate based on 40 datasets; the solid line is the theoretical result; the dashed line is the mean result from 100 estimates; and the dotted line is the CoV of the θ estimates

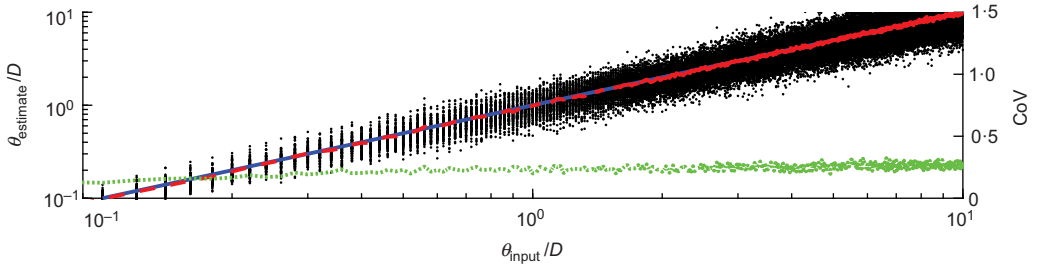

Fig. 7. Accuracy of θ estimation as a function of actual θ relative to domain size D, based on a fixed domain length of 50, and uneven data spacing: 6 × 2·5 intervals, followed by 16 × 1·25 intervals and then 6 × 2·5 intervals. Each dot is a single estimate based on 40 datasets; the solid line is the theoretical result; the dashed line is the average result from 100 estimates; and the dotted line is the CoV of the θ estimates

changes when the input θ is varied between 0·1 times the domain length ($\theta = 5$) and 10 times the domain length ($\theta = 500$). Considering an input θ of only 0·1 times the domain length there is a 44·2% chance of the estimated θ being within 20% of the real θ, whereas for an input θ equal to the domain length there is a 33·3% chance of it being within 20%, decreasing to 26·04% when it is 10 times the domain length. However, Fig. 6 shows that estimating scales of fluctuation is possible, even when the real θ is 5 times larger than the domain length.

As θ increases relative to the domain size, the CoV of the estimated θ also increases, although the CoV increases at a much lower rate when θ is larger than the domain size. In Fig. 6, the CoV increases from 0·05 to 0·2 when θ is smaller than the domain size, whereas, for θ larger than the domain size, the CoV varies in the range of 0·2–0·35.

Unequally spaced data

In reality, it may be difficult to obtain equally spaced data, especially for estimating θ_h. By using equation (6) rather than equation (5) the auto-correlation function can be obtained. Fig. 7 illustrates the estimation of θ obtained from non-uniform sampling. In this case, the non-uniform sampling regime is based on the site investigation reported later in the paper; that is, with data points at six intervals of 2·5 (length units), followed by 16 intervals of 1·25, and ending with six intervals of 2·5, totalling 29 points generated at non-uniform spacings over a total distance of 50. Once again, 100 estimates of θ were made, with each estimate based on 40 datasets.

Comparing Figs 6 and 7, there is little difference between the results based on uniformly distributed data and those based on unequally spaced data.

Generalising the uncertainty quantification

In order to generalise the estimation of the uncertainty in a calculated θ in a single direction, several values of θ relative to domain size (0·1 to 10·0) and of θ relative to distance between data points (10 to 200), for different numbers of datasets and points per dataset, have been investigated using the same approach as in Figs 4–7. For example, Figs 8(a)–8(f) show, as a series of broken lines, contours of the calculated CoV of θ, expressed as a function of θ, the number of data points, size of domain, spacing of data points and number of datasets (e.g. CPTs).

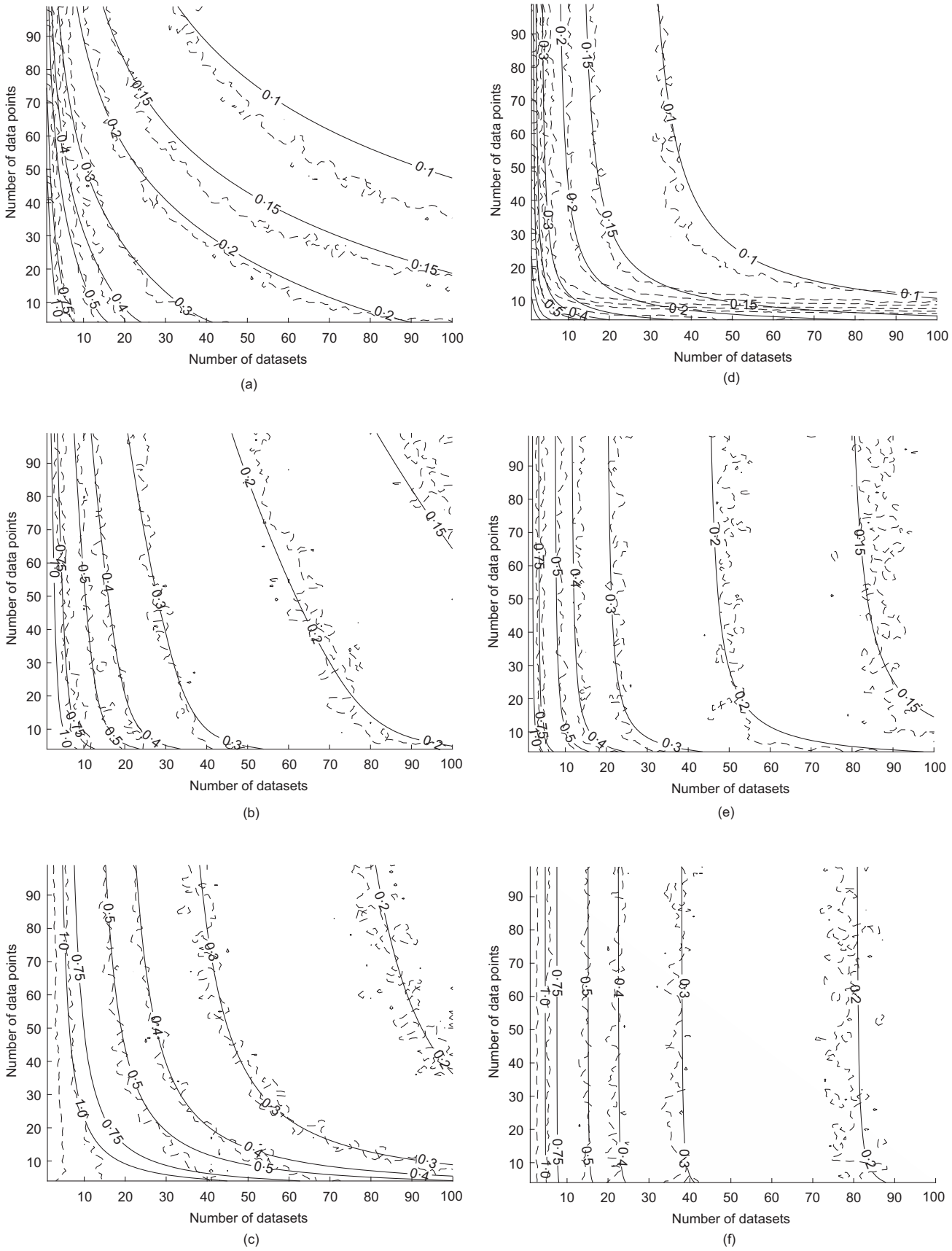

Fig. 8. Calculated and theoretical isochrones of CoV of θ, as a function of θ, domain size D, interval size in, number of CPTs (datasets) and number of data per dataset. For graphs (a)–(c) the interval between the points is fixed (0·5) and, as the number of points increases, the domain size also increases. For graphs (d)–(f) the domain size is fixed at 50 and, as the number of points increases, the interval between the points decreases: (a) $\theta = 5$, $in = 0{\cdot}5$, D varies; (b) $\theta = 25$, $in = 0{\cdot}5$, D varies; (c) $\theta = 250$, $in = 0{\cdot}5$, D varies; (d) $\theta = 5$, in varies, $D = 50$; (e) $\theta = 25$, in varies, $D = 50$; (f) $\theta = 250$, in varies, $D = 50$

By using a minimum error approach on the CoV values obtained, an estimate for the CoV of θ (i.e. the statistical uncertainty) was fitted to the data

$$\text{CoV} = 1{\cdot}1 \times W \times X \times Y + Z \tag{13}$$

In equation (13), W is a factor accounting for the magnitude of θ relative to the domain size D. It has most influence when θ is smaller than D, and is given by

$$W = \tan^{-1}\left(\frac{5\theta}{D}\right) \tag{14}$$

X is a factor that considers the number of independent datasets nf, that is

$$X = \frac{1}{\sqrt{nf}} \tag{15}$$

where nf equals the number of CPTs for estimating θ in the vertical direction or number of rows of data for estimating θ in the horizontal direction; nf has a maximum allowable value of $nf_{\max}$, which is estimated using

$$nf_{\max} = \begin{cases} \dfrac{D_{\text{p}}}{\theta_{\text{p}}}; & D_{\text{p}} > \theta_{\text{p}} \\ 1; & D_{\text{p}} \leq \theta_{\text{p}} \end{cases} \tag{16}$$

where D_{p} is the domain length in the direction perpendicular to the direction of investigation – that is, the horizontal domain length when considering the vertical scale of fluctuation, or the vertical domain length when considering the horizontal scale of fluctuation – and θ_{p} is the θ in the direction perpendicular to the direction of investigation. This maximum allowable value accounts for the lack of independence between datasets which are very close to each other.

Y is a factor that accounts for an increase in CoV as the spacing between the data points (in) increases relative to θ

$$Y = \left(1 + \frac{in}{\theta}\right) \tag{17}$$

and Z is a factor with most influence when θ is larger than D, accounting for the reduction of accuracy in the estimate of θ as the number of experimental covariance functions ($\hat{\gamma}(\tau)$) reduces significantly as τ becomes closer to D

$$Z = \frac{\theta}{5nfD} \tag{18}$$

Hence, equation (13) incorporates four factors that affect the estimate of CoV: (*a*) a factor for $\theta < D$ – if the domain length is large relative to θ, CoV is reduced; (*b*) a factor for the number of (independent) datasets – as the number of datasets increases, CoV decreases. The maximum number of independent datasets ($nf_{\max}$) can be estimated by considering the size of the perpendicular θ (i.e. θ_{p}) relative to the perpendicular domain (D_{p}). For example, considering the estimation of θ_{h}, if $\theta_{\text{p}} = \theta_{\text{v}} = 0{\cdot}25$ with $D_{\text{p}} = D_{\text{v}} = 4$ then $nf_{\max}$ for the calculation of θ_{h} is 16, and even if more than 16 datasets are available nf cannot exceed 16; (*c*) a factor for the spacing of data – if the interval between data points is large compared to θ, CoV increases; (*d*) a factor for $\theta > D$ – if the domain is short relative to θ, CoV increases. Note that, in all cases, the CoV decreases with an increase in the number of datasets.

The estimated CoVs are plotted in Figs 8(a)–8(f) as solid lines. A good fit between the calculated CoVs (using synthetic data) and estimated CoVs (using equation (13)) is found.

To estimate the CoV for either the vertical or horizontal θ using standard (vertical) CPT data, equation (13) can be used. It should be noted that, for estimating θ_{v}, there are generally many points per CPT, although often relatively few CPTs, so the estimations in the top left corners of the graphs in Fig. 8 are representative. For estimating θ_{h} there are generally fewer data points per dataset, although each depth interval for the group of CPTs can be considered as a different dataset. Hence, there can be a large number of datasets comprising relatively few data points, so that the estimation in the bottom right corner of the graphs is applicable. This assessment of the uncertainty is based on the Markov auto-correlation model and experimental auto-correlation function. It is expected that the predicted uncertainty would be generally applicable using similar auto-correlation models (such as those shown in Table 1), although a precise calibration could be undertaken.

Design chart

The CoV estimation from equation (13) has been summarised in the design chart shown in Fig. 9. This chart can be used in practice to estimate the uncertainty in calculated values of θ in a single direction. Hence, it is ideally suited, but not limited, to linear infrastructure. θ must first be calculated (or estimated, if used prior to a site investigation), and then information on the size of the dataset is used. As an example, a series of ten vertical CPTs (with a vertical data interval of 0·01 m) and a vertical domain of 5 m is considered. They are arranged in a single line and horizontally spaced at 2·5 m intervals, making a horizontal domain length of 22·5 m; the estimate of θ_{v} is 0·25 m and θ_{h} is estimated to be 5 m. Considering the CoV of θ_{v}, the first part of the figure gives a value of 0·29 and the second part of the figure gives a value below 0·005; adding these numbers together gives 0·295. The nf here is defined by $nf_{\max}$ as $22{\cdot}5/5 = 4{\cdot}5$ and the estimated CoV of θ_{v} is then 0·14 m. Considering θ_{h}, the first part of the figure gives a value of 1·4 and the second part gives a value below 0·02; adding these numbers together gives 1·42. Once again nf is defined by $nf_{\max}$, which is equal to $5/0{\cdot}25 = 20$, and this gives an estimated CoV of θ_{h} of 0·32.

Double scales of fluctuation

Figure 10 presents an investigation into the estimation of a double θ. A series of synthetic CPTs based on a double θ model (based on $\theta_1 = 5$, $c_1 = 0{\cdot}75$ and $\theta_2 = 25$, $c_2 = 0{\cdot}25$) was generated and analysed. Using equation (11), $\theta_{\text{avg}} = 10$. An investigation synonymous with the estimation of θ_{h} values using eight vertical CPTs was carried out. Three different site investigation scenarios were tested: (*a*) where the spacing between the CPTs was 0·5; (*b*) where the spacing between the CPTs was 14·5; and (*c*) where four groups of two CPTs were used, in which the spacing within the groups was 0·5 and between the groups was 14·5. The number of datasets was investigated, corresponding to the number of independent sets of horizontal data, with each dataset comprising a single reading from each CPT. The step size used for c_1 and c_2 in the algorithm was 0·01, as was the step size for θ_1 and θ_2.

Figures 10(a) and 10(b) show that when using a small equal spacing between the CPTs, the means of both θ values and the average θ (θ_{avg}) were reasonably well estimated. The mean of θ_1 was slightly underestimated, with a reasonably low CoV, whereas the mean of θ_2 was more underestimated, with a significantly higher (approximately double) CoV. This is due to the short domain (3·5) in comparison to the larger θ (25). Even for the smaller θ (5), the domain is shorter than the detected value. θ_{avg} has a reasonably high CoV, indicating that the overall error can be high.

Using the larger equal spacing, Figs 10(c) and 10(d) show that the means of both θ values and θ_{avg} were again

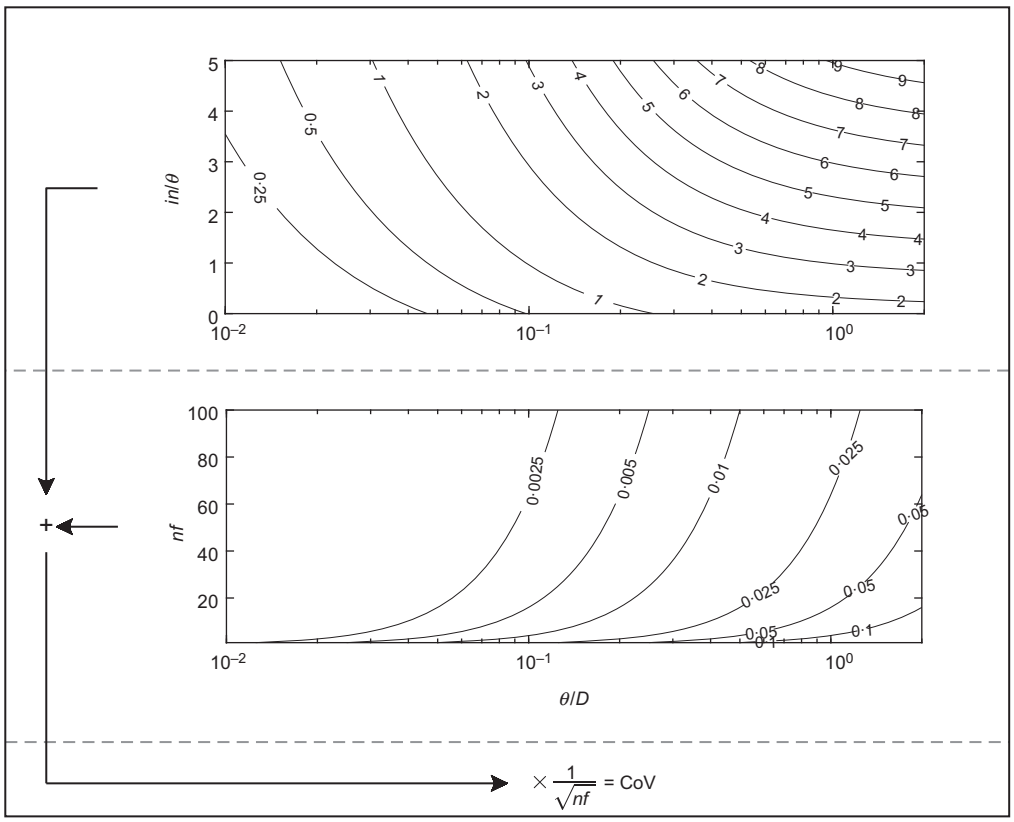

Fig. 9. Estimating CoV for equally spaced vertical and horizontal data, based on the interval between data points, scale of fluctuation, domain size and number of datasets

reasonably well estimated. As before, the mean of θ_1 was slightly underestimated, but now with a high CoV, whereas the mean of θ_2 was slightly overestimated, but with a relatively low CoV. The high CoV for the smaller θ is due to the spacing being significantly larger than θ. The CoV of θ_{avg} reduces to a reasonable value after around 40 datasets.

Figures 10(e) and 10(f) show the results using a mixed spacing system. The means of both θ values and θ_{avg} were again reasonably well estimated. However, in this case, the CoVs of both θ values are reasonably low, with the CoV of θ_{avg} being the lowest of all the scenarios. Hence, the mixed spacing system offers a practical solution for when the θ values are not reasonably well known in advance.

From this analysis and a more extensive investigation into the effect of using CPT groups, involving a large variation in the number of groups and spacings between the groups (de Gast, 2020), it was found that the CoV could also be estimated with equation (13). In this case, *in* becomes the interval between the CPT groups in equation (17), rather than the interval between data points. Moreover, the domain (D) should be replaced by the domain length of the groups (D_g) in equation (14), the number of groups (ng) should be included in the denominator in equation (17) (i.e. $\theta \times ng$) and the domain (D) should be replaced by the total length of the domain (D_t) in the denominator in equation (18).

INVESTIGATION USING REAL DATA

The performance of the methodology has been applied to a real site investigation, conducted for a regional dyke in the Netherlands called Leendert de Boerspolder. The site investigation comprised 100 CPTs over an area of 50 × 15 m, and data from this test have been made available (de Gast *et al.*, 2020). As part of the investigation, 29 CPTs were conducted along the crest of the dyke over a 50 m reach. Specifically, six CPTs were taken at 2·5 m intervals, followed by 16 at 1·25 m intervals, and then another six at 2·5 m intervals. Fig. 11 shows the cone resistance data from these CPTs. From these CPTs and an accompanying site investigation, the subsurface was found to consist of a man-made dyke body (which has been maintained and added to over ~400 years), composed of silts, clays, sands and some rubble. This lies on top of a peat layer underlain by a clay layer consisting of two clay types (an organic clay and a silty clay), followed by another peat layer underlain by a sand layer. To demonstrate the proposed methodology, the dyke material, the first peat layer and the clay layer were analysed. Owing to the deposition history, it is generally assumed that θ_h is larger than θ_v, so the spacing of data in the horizontal plane was anticipated to be acceptable.

Figures 12 and 13 show the data analyses undertaken for the peat and clay layers, respectively. In Figs 12(a) and 13(a) the CPT cone resistance data are presented, as is the mean and de-trended (i.e. mean of zero) CPT data. The peat has a linear mean trend that has been removed before analysing the data to determine θ. In contrast, the clay exhibits a non-linear (quadratic) trend, probably due to the change in the composition of the clay material with depth. A histogram and probability density function of the de-trended data are shown in Figs 12(c) and 13(c). The vertical experimental auto-correlation function and best fit theoretical auto-correlation function, for both a single and double θ, are shown in Figs 12(b) and 13(b), with the same information for the horizontal correlation shown in Figs 12(e) and 13(e). In these examples, the step size for c_1 and c_2 was 0·01 and the step size for θ_1 and θ_2 was 0·01 m. Details of the identified scales of fluctuation, along with the calculated errors between the experimental and theoretical auto-correlation functions, are presented in Figs 12(d) and 13(d).

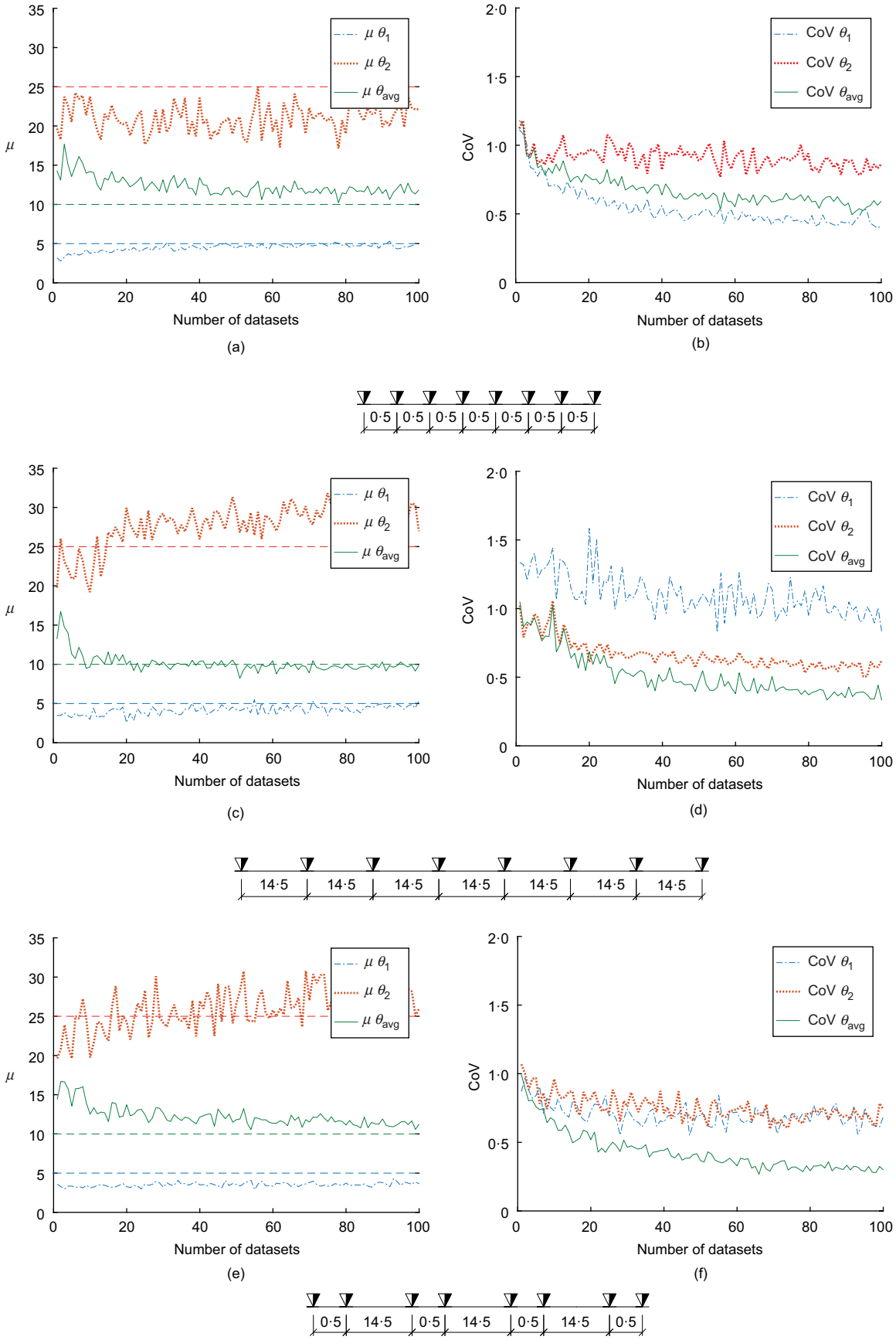

Fig. 10. Investigation of the identification of a double θ in the horizontal plane ($\theta_1 = 5$, $c_1 = 0{\cdot}75$ and $\theta_2 = 25$, $c_2 = 0{\cdot}25$) using eight CPTs and different numbers of datasets (i.e. vertical intervals): (a) means of the identified θs for CPT spacing of 0·5; (b) CoVs of the identified θs for CPT spacing of 0·5; (c) means of the identified θs for CPT spacing of 14·5; (d) CoVs of the identified θs for CPT spacing of 14·5; (e) means of the identified θs for alternate CPT spacings of 0·5 and 14·5; (f) CoVs of the identified θs for alternate CPT spacings of 0·5 and 14·5. Horizontal dashed lines in (a), (c) and (e) are theoretical values

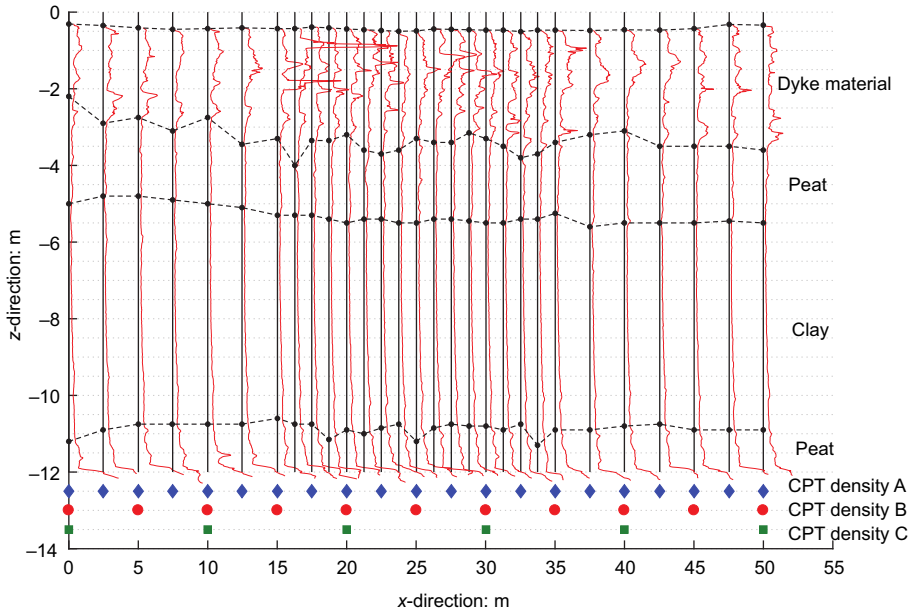

Fig. 11. Twenty-nine non-uniformly spaced CPT profiles along the dyke crest. The solid vertical lines indicate the locations of the CPTs, with the normalised tip resistance q_t shown to the right. On the vertical lines the layer boundaries are identified with a dot, with the broken lines indicating the layer boundaries over the cross-section

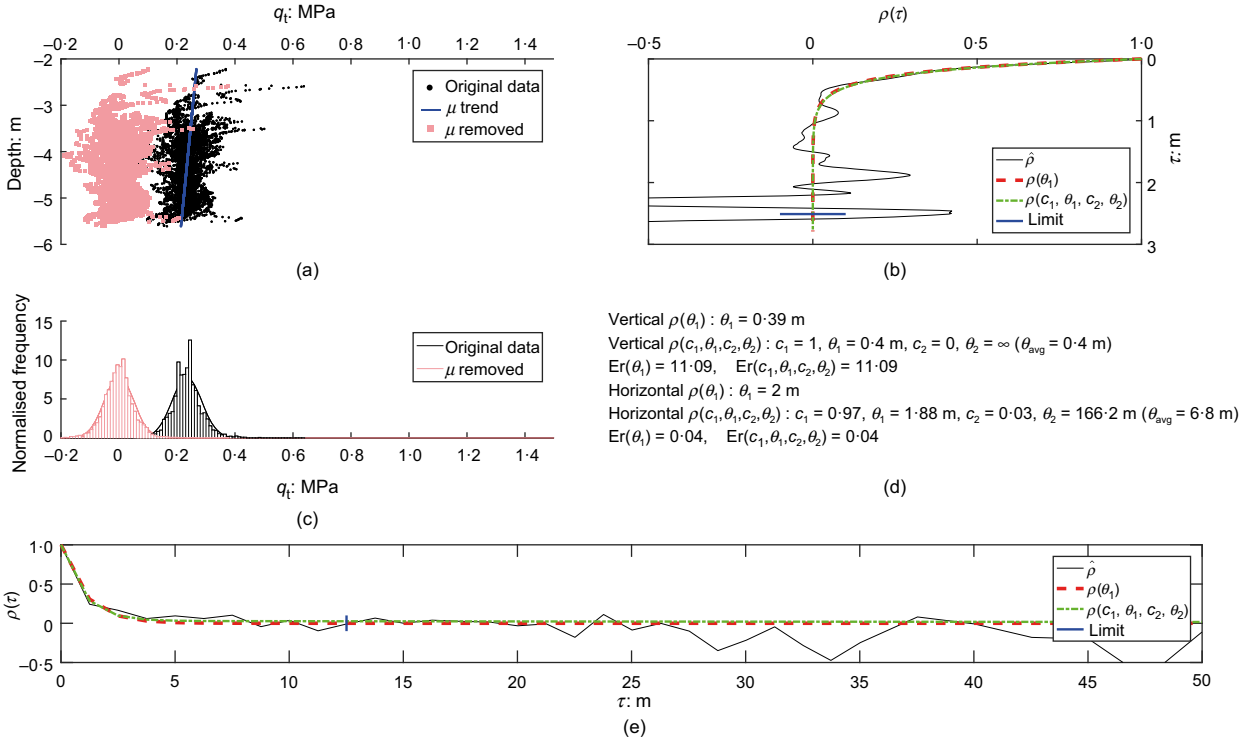

Fig. 12. Analysis of CPT data for the upper peat layer: (a) raw and de-trended data from which the auto-correlation is estimated; (b) two estimates of the vertical auto-correlation; (c) histogram and probability density function of original data and data with mean trend removed; (d) theoretical estimates of scales of fluctuation and their corresponding errors; (e) two estimates of the horizontal auto-correlation

In the peat layer, a negligible difference is observed when a double vertical or horizontal θ is introduced, as indicated by the identical error. In the clay layer, for both the vertical and horizontal directions, using a double θ reduces the error significantly. In all soil layers (including those not shown), where a double θ is observed in the horizontal direction the main component of θ_h is the smaller of the two scales (i.e. θ_1).

Note that due to the erratic behaviour of the experimental auto-correlation function at large lag lengths (due to fewer data being included in the function), the function is cut off at a limit before the errors are calculated. When determining the extent of the auto-correlation function to be considered, a visual inspection is used. The criterion used to determine the limit is how closely the function resembles the auto-correlation function, with particular attention to its monotonic reduction to zero. Suggested limits, based on the proximity of the auto-correlation function to zero or the domain size, have been proposed in the literature by, for example, Uzielli *et al.* (2005) and Lumb (1975). In Figs 12(b), 12(e) and 13(b), 13(e) the extent of the data used is indicated by the solid thick line.

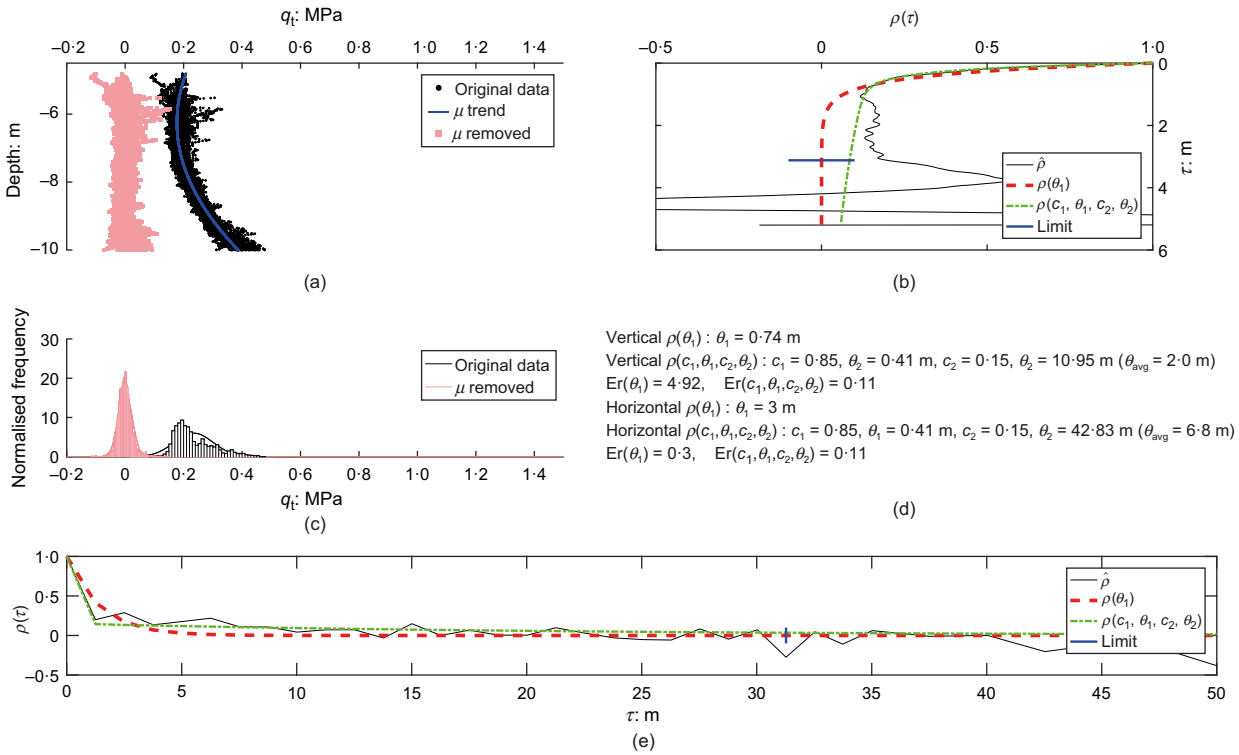

Fig. 13. Analysis of CPT data for the clay layer: (a) raw and de-trended data from which the auto-correlation is estimated; (b) two estimates of the vertical auto-correlation; (c) histogram and probability density function of original data and data with mean trend removed; (d) theoretical estimates of scales of fluctuation and their corresponding errors; (e) two estimates of the horizontal auto-correlation

In the analysis of this site, it can be concluded that θ_h is larger than θ_v for all layers, and is therefore in agreement with the view that geological processes (in sediments) cause a horizontal auto-correlation that is larger than the vertical auto-correlation. There has been mention of the possibility of a double θ by Vanmarcke *et al.* (1986) although neither the identification nor the use of a double θ is evident in the literature. A double θ can have different explanations: (*a*) it is caused by geological processes that operate on different scales; (*b*) it is caused by a weak spatial auto-correlation of the material properties and/or a large amount of noise; or (*c*) it is the product of the removal of a wrong trend. The double correlation model has more fitting parameters and is therefore likely to achieve a better fit to data. However, to prevent over-fitting, the preference is the use of a single correlation model, unless an explanation for the origin of a double correlation can be provided and the two correlation lengths are significantly different. Introducing a double θ into an analysis may change the computed response of some geotechnical structures, although the extent to which the response is influenced by a double θ is likely to depend on the type of structure involved.

To investigate the impact of the amount of data used, the (single) vertical and horizontal scales of fluctuation were determined from different sampling densities, using the full domain length – that is 50 m. Three sample densities were considered: (*a*) density A – 21 CPTs at 2·5 m spacing; (*b*) density B – 11 CPTs at 5 m spacing; and (*c*) density C – six CPTs at 10 m spacing, with the CPTs used in the analyses indicated in Fig. 11.

In general, Table 2 shows that the CoVs of θ (calculated from equation (13)) increase in both the vertical and horizontal directions as the spacing increases; this is due to fewer data being available to determine θ. However, for the dyke material, there is no reduction in the CoV of θ_v between CPT density A and CPT density B, and only a slight increase in the CoVs of θ_v for the clay and peat layers. This is due to the lack of independence of adjacent CPTs. As the CPTs are spaced at a distance of less than θ_h, they are not considered independent measurements, and this limits the benefit of increasing the amount of data.

The CoVs of θ_h are larger in all cases than the CoVs of θ_v, and change more significantly as a function of sampling density. The two most influential factors here are the

Table 2. Summary of single θ determined with different amounts of data

Material	CPT density A				CPT density B				CPT density C			
	Vertical		Horizontal		Vertical		Horizontal		Vertical		Horizontal	
	θ_v: m	CoV	θ_h: m	CoV	θ_v: m	CoV	θ_h: m	CoV	θ_v: m	CoV	θ_h: m	CoV
Dyke material	0·30	0·17	7·0	0·27	0·28	0·17	7·0	0·34	0·31	0·24	10·0*	0·54
Peat	0·41	0·15	3·0	0·21	0·32	0·16	4·0	0·29	0·31	0·23	10·0*	0·53
Clay	0·61	0·19	5·0	0·26	0·69	0·25	7·0	0·42	0·88	0·37	10·0*	0·73

*No scale detected, therefore CPT spacing used.

intervals between the readings (the CPT spacing) and the domain length (equations (14) and (17)). It is also seen that the widest estimated distributions of θ values (from CPT density C) contain the more narrow distributions (from CPT density A and CPT density B), giving confidence in the method of estimating the CoV.

Note on using real data

When real CPT data are used, the following points should be considered: (*a*) the trend (or mean for stationary data) must be carefully estimated and removed; (*b*) properties can change in space – in other words, the data are not stationary, due to underlying physical processes and non-statistical uncertainty; (*c*) the transformation model from CPT data to material property values contains error.

Errors in de-trending will influence the calculated experimental auto-correlation function. When de-trending, one should be careful not to over de-trend the data; that is, the removal of a trend should be explained by a physical process – for example, a change in material type, gradual change of the material composition, increasing effective stress and loading history of the material. In this process, the data, after de-trending, are assumed to be stationary. A method to test for stationarity of soil layers was suggested by Phoon *et al.* (2004) – namely, the modified Bartlett test, where groups of data are separated and the calculated variances are compared for stationarity.

There are uncertainties involved with transforming from tip resistance to undrained shear strength. Ching *et al.* (2016) investigated the transformation uncertainty at different sites and concluded that the transformation uncertainty, while spatially variable, can generally be taken as a single value for a specific site or layer as the scale of fluctuation is typically large in comparison to the geotechnical structure. Based on a database of CPTs it was estimated to be some tens of metres. Although any error in the transformation will cause a shift in the property values, this will have a limited influence on the estimated spatial correlation due to the de-trending of data.

Example analysis of dyke stability

To examine the influence of the CPT sampling regime on a geotechnical structure, an example dyke stability assessment has been examined. An idealised slope was analysed using the semi-analytical method proposed by Vanmarcke (1977), as previously elaborated by Li *et al.* (2015). This method has recently been extended by Varkey *et al.* (2019) to improve the match with more general, albeit computationally more expensive, methods, but here the original version of the model has been used for simplicity. The method takes account of the additional stability provided by the ends of a three-dimensional (3D) rotational (cylindrical) failure mechanism.

The example has been chosen only to illustrate the effects of the scale of fluctuation obtained from different sampling strategies. For a more comprehensive analysis of the slope reliability based on the measurement data presented herein, other uncertainties, such as the transformation uncertainty, geometric uncertainty and model uncertainty would have to be included; see for example Phoon & Kulhawy (1999) and van den Eijnden & Hicks (2019).

A brief summary of the method is given as follows. The mean and standard deviation of the 3D factor of safety ($F_{3D,\mu}$ and $F_{3D,\sigma}$, respectively), are calculated as

$$F_{3D,\mu} = F_{2D,\mu}\left(1 + \frac{d}{b}\right) \tag{19}$$

$$F_{3D,\sigma} = F_{2D,\mu}\Gamma(L_a)\Gamma(b)\mathrm{CoV} \tag{20}$$

where $F_{2D,\mu}$ is the mean two-dimensional (2D) factor of safety calculated by any appropriate method and CoV is the coefficient of variation of the point statistics of shear strength. d is the effective width of the failure surface, calculated as

$$d = 2A/L_a \tag{21}$$

where L_a is the length of the cross-sectional failure arc, A is the cross-sectional area of the sliding mass cylinder and b is the length of the failure surface along the dyke, which was proposed by Vanmarcke (1977) to be

$$b = \begin{cases} \dfrac{F_{2D,\mu}}{F_{2D,\mu} - 1}; & b > \theta_h \\ \theta_h; & b \le \theta_h \end{cases} \tag{22}$$

In equation (20), $\Gamma(L_a)$ and $\Gamma(b)$ are reduction factors accounting for local averaging, calculated as

$$\Gamma(L_a) = \begin{cases} \sqrt{\theta_h/L_a}; & L_a > \theta_e \\ 1; & L_a \le \theta_e \end{cases} \tag{23}$$

$$\Gamma(b) = \begin{cases} \sqrt{\theta_h/b}; & b > \theta_h \\ 1; & b \le \theta_h \end{cases} \tag{24}$$

where θ_e is the equivalent θ along the failure surface, obtained by using a weighted average of the vertical and horizontal scales of fluctuation (Li *et al.*, 2015).

The example 45° slope was 5 m high and had no foundation layer. The following material properties were used: unit weight, $\gamma = 20$ kN/m^3; mean undrained shear strength, $s_u = 26$ kPa; and CoV = 0·3. The mean 2D factor of safety of the slope was calculated using Taylor's (1937) method to be $F_{2D,\mu} = 1{\cdot}60$, and the following properties of the failure surface were calculated: $d = 3{\cdot}83$ m, $b = 10{\cdot}2$ m, $L_a = 12$ m and $A = 23$ m^2. The scales of fluctuation were taken from the peat layer of the field test, shown in Table 2, and the impact of the different sampling densities was examined. The CoVs of θ_v and θ_h were estimated based on the method presented above – that is, Fig. 9 or equation (13).

To illustrate the effect of the uncertainty in θ, the five percentile factor of safety (95% confidence) was calculated for the different scenarios and this information is given in Table 3. These have been based on the mean calculated θ_v and $\theta_h \pm$ one standard deviation. For comparison, using the same mean factor of safety calculated by equation (19) – that is 2·2 – and the CoV of the undrained shear strength, the five percentile factor of safety of the slope is 1·11 when only the point statistics are considered (i.e. neglecting the influence of θ).

When accounting for the spatial nature of the variability, the five percentile factor of safety is much higher (around 2) in all cases. It is highest when the CoVs of θ are the smallest, given that the mean θ values were similar in all cases. CPT density A and CPT density B give similar results because the

Table 3. Five percentile factors of safety determined with different CPT densities

Scenario	$F_{3D,5\%}$		
	$\theta_v - \sigma$, $\theta_h - \sigma$	θ_v, θ_h	$\theta_v + \sigma$, $\theta_h + \sigma$
CPT density A	2·10	2·08	2·06
CPT density B	2·10	2·07	2·04
CPT density C	2·07	1·99	1·92

closest vertical CPTs could not be considered independent. CPT density C gives lower results, as larger θ_h values are predicted (i.e. as no scale could be detected, the CPT spacing was used as the default), which disproportionately increases the reduction factors (equations (23) and (24)).

While the potential advantage in the five percentile factor of safety is, in this example, only around 5% by including the uncertainty in θ, this may be sufficient to reduce or avoid costly dyke upgrades, easily making the additional site investigation cost effective.

CONCLUSION

The reliability of cone penetration testing for calculating the scale of fluctuation of soils in both the vertical and horizontal directions has been comprehensively tested. Using synthetically generated data, it has been shown that it is important to have enough information to make a reliable estimate of the scale of fluctuation. It is also important to consider the likely scales of fluctuation that may occur and the density of the sampling. Based on an extensive investigation of synthetic data, a method has been proposed for guiding the design of CPT campaigns for calculating the scale of fluctuation and, in a significant step forward, for estimating the statistical uncertainty in that calculated value. This allows the uncertainty in the calculated scales of fluctuation to be included in comprehensive probabilistic analyses of structure response. Newly collected real data have been used to demonstrate the approach and the results have been used for a simple slope stability problem as an example, demonstrating the significance of considering scales of fluctuation and their uncertainty due to site investigation design.

ACKNOWLEDGEMENT

This work is part of the research programme Reliable Dykes with project number 13864, which is financed by the Netherlands Organisation for Scientific Research (NWO).

NOTATION

A	area
b	length of the failure surface
c_i	weights for weighted average
D	domain length
D_p	perpendicular domain length
d	effective width of the failure surface
$\mathrm{Er}(x)$	error as a function of x
$F_{2D,\mu}$	mean safety factor of a two-dimensional (2D) section
$F_{3D,5\%}$	five percentile of the three-dimensional (3D) factor of safety
$F_{3D,\mu}$	mean of the 3D factor of safety
$F_{3D,\sigma}$	standard deviation of the 3D factor of safety
i, j, k, l, t	counters
in	interval between data points
L_a	length of the cross-sectional failure arc
nf	number of independent datasets
nf_{max}	maximum number of independent datasets
s_u	undrained shear strength
W, X, Y, Z	parameters of equation (13)
γ	unit weight
$\hat{\gamma}(\tau)$	experimental covariance function
$\Gamma(\ )$	reduction factor
Δj	index difference
θ	scale of fluctuation (auto-correlation length)
θ_{avg}	average scale of fluctuation
θ_e	equivalent scale of fluctuation
$\theta_{estimate}$	estimated scale of fluctuation
θ_h	horizontal scale of fluctuation
θ_i	component of the scale of fluctuation
θ_{input}	input scale of fluctuation
θ_{max}	maximum scale of fluctuation
θ_p	perpendicular scale of fluctuation
θ_v	vertical scale of fluctuation
μ	mean
$\hat{\mu}$	estimated mean (or trend)
$\rho(\tau)$	auto-correlation function
$\hat{\rho}(\tau)$	experimental auto-correlation function
ρ_{com}	composite auto-correlation function
σ	standard deviation
τ	lag distance

REFERENCES

Alabert, F. (1987). The practice of fast conditional simulations through the LU decomposition of the covariance-matrix. *Math. Geol.* **19**, No. 5, 369–386.

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

Campanella, R. G., Wickremesinghe, D. S. & Robertson, P. K. (1987). Statistical treatment of cone penetrometer test data. In *Proceedings of the 5th international conference on applications of statistics and probability in soil and structural engineering* (ed. N. C. Lind), vol. 2, pp. 1011–1019. Waterloo, Canada: University of Waterloo, Institute for Risk Research.

Ching, J., Phoon, K. K. & Wu, T. J. (2016). Spatial correlation for transformation uncertainty and its applications. *Georisk: Assessment Manage. Risk Engd Syst. Geohazards* **10**, No. 4, 294–311, https://doi.org/10.1080/17499518.2016.1205749.

Ching, J., Wu, T. J., Stuedlein, A. W. & Phoon, K. K. (2018). Estimating horizontal scale of fluctuation with limited CPT soundings. *Geosci. Front.* **9**, No. 6, 1597–1608, https://doi.org/10.1016/j.gsf.2017.11.008.

Davis, M. W. (1987). Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Math. Geol.* **19**, No. 2, 91–98.

de Gast, T. (2020). *Dykes and embankments, a geostatistical analysis on soft terrain*. PhD thesis, Delft University of Technology, Delft, the Netherlands.

de Gast, T., Vardon, P. J. & Hicks, M. A. (2020). CPT dataset to study soil heterogeneity. *4TU repository.* Delft, the Netherlands: Delft University of Technology. See https://doi.org/10.4121/uuid:0a9b77bc-3f6d-40e2-846c-ddb992dd78a6 (accessed 04/12/2020).

Fenton, G. A. (1999). Random field modeling of CPT data. *J. Geotech. Geoenviron. Engng* **125**, No. 6, 486–498.

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

Fenton, G. A., Naghibi, F. & Hicks, M. A. (2018). Effect of sampling plan and trend removal on residual uncertainty. *Georisk: Assessment Manage. Risk Engd Syst. Geohazards* **12**, No. 4, 253–264.

Griffiths, D. V. & Fenton, G. A. (1997). Three-dimensional seepage through spatially random soil. *J. Geotech. Geoenviron. Engng* **123**, No. 2, 153–160.

Hicks, M. A. & Li, Y. (2018). Influence of length effect on embankment slope reliability in 3D. *Int. J. Numer. Analyt. Methods Geomech.* **42**, No. 7, 891–915.

Hicks, M. A. & Onisiphorou, C. (2005). Stochastic evaluation of static liquefaction in a predominantly dilative sand fill. *Géotechnique* **55**, No. 2, 123–133, https://doi.org/10.1680/geot.2005.55.2.123.

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

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.

Hicks, M. A., Nuttall, J. D. & Chen, J. (2014). Influence of heterogeneity on 3D slope reliability and failure consequence. *Comput. Geotech.* **61**, 198–208.

Honjo, Y. (2011). Challenges in geotechnical reliability based design. In *Proceedings of the 3rd international symposium on geotechnical safety and risk (ISGSR 2011)* (eds N. Vogt,

B. Schuppener, D. Straub and G. Bräu), pp. 11–28. Karlsruhe, Germany: Bundesanstalt für Wasserbau.

Jaksa, M., Kaggwa, W. S. & Brooker, P. I. (1999). Experimental evaluation of the scale of fluctuation of a stiff clay. In *Proceedings of 8th international conference on applications of statistics and probability; applications of statistics and probability civil engineering reliability and risk analysis*, (eds R. E. Melchers and M. G. Stewart), pp. 415–422. Rotterdam, the Netherlands: A. A. Balkema.

Jaksa, M. B., Goldsworthy, J. S., Fenton, G. A., Kaggwa, W. S., Griffiths, D. V., Kuo, Y. L. & Poulos, H. G. (2005). Towards reliable and effective site investigations. *Géotechnique* **55**, No. 2, 109–121, https://doi.org/10.1680/geot.2005.55.2.109.

Kawa, M. & Puła, W. (2020). 3D bearing capacity probabilistic analyses of footings on spatially variable c–ϕ soil. *Acta Geotech.* **15**, No. 6, 1453–1466, https://doi.org/10.1007/s11440-019-00853-3.

Li, Y. J., Hicks, M. A. & Nuttall, J. D. (2015). Comparative analyses of slope reliability in 3D. *Engng Geol.* **196**, 12–23.

Li, Y. J., Hicks, M. A. & Vardon, P. J. (2016). Uncertainty reduction and sampling efficiency in slope designs using 3D conditional random fields. *Comput. Geotech.* **79**, 159–172.

Lloret-Cabot, M., Hicks, M. A. & van den Eijnden, A. P. (2012). Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using Kriging. *Géotechnique Lett.* **2**, No. 3, 123–127, https://doi.org/10.1680/geolett.12.00022.

Lloret-Cabot, M., Fenton, G. A. & Hicks, M. A. (2014). On the estimation of scale of fluctuation in geostatistics. *Georisk: Assessment Manage. Risk Engd Syst. Geohazards* **8**, No. 2, 129–140.

Lumb, P. (1975). Spatial variability of soil properties. In *Proceedings of the 2nd international conference on applications of statistics and probability to soil and structural engineering* (ed. E. Schultze), pp. 397–421. Essen, Germany: Deutsche Gesellschaft für Erd- und Grundbau.

Naghibi, F., Fenton, G. A. & Griffiths, D. V. (2016). Probabilistic considerations for the design of deep foundations against excessive differential settlement. *Can. Geotech. J.* **53**, No. 7, 1167–1175.

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

Phoon, K. K., Quek, S. T. & An, P. (2004). Geostatistical analysis of cone penetration test (CPT) sounding using the modified Bartlett test. *Can. Geotech. J.* **41**, No. 2, 356–365.

Popescu, R., Deodatis, G. & Nobahar, A. (2005). Effects of random heterogeneity of soil properties on bearing capacity. *Probabilistic Engng Mech.* **20**, No. 4, 324–341.

Robertson, P. K., Campanella, R., Gillespie, D. & Greig, J. (1986). Use of piezometer cone data. In *Use of in situ tests in geotechnical engineering* (ed. S. P. Clemence), GSP 6, pp. 1263–1280. New York, NY, USA: American Society of Civil Engineers.

Sert, S., Luo, Z., Xiao, J. H., Gong, W. P. & Juang, C. H. (2016). Probabilistic analysis of responses of cantilever wall-supported excavations in sands considering vertical spatial variability. *Comput. Geotech.* **75**, 182–191.

Spencer, W. A. & Hicks, M. A. (2007). A 3D finite element study of slope reliability. In *Numerical models in geomechanics: NUMOG X* (eds G. N. Pande and S. Pietruszczak), pp. 539–543. Boca Raton, FL, USA: CRC Press (Taylor & Francis).

Suchomel, R. & Masín, D. (2010). Comparison of different probabilistic methods for predicting stability of a slope in spatially variable c–ϕ soil. *Comput. Geotech.* **37**, No. 1–2, 132–140.

Taylor, D. W. (1937). Stability of earth slopes. *J. Boston Civ. Engng Soc.* **24**, No. 3, 197–246.

Uzielli, M., Vannucchi, G. & Phoon, K. K. (2005). Random field characterisation of stress-normalised cone penetration testing parameters. *Géotechnique* **55**, No. 1, 3–20, https://doi.org/10.1680/geot.2005.55.1.3.

van den Eijnden, A. P. & Hicks, M. A. (2017). Efficient subset simulation for evaluating the modes of improbable slope failure. *Comput. Geotech.* **88**, 267–280.

van den Eijnden, A. P. & Hicks, M. A. (2019). On the importance of a complete characterization of site investigation data uncertainty: a computational example. *Proceedings of the 7th international symposium on geotechnical safety and risk (ISGSR 2019)* (eds J. Ching, D. Q. Li and J. Zhang), pp. 237–242. Singapore: Research Publishing.

Vanmarcke, E. H. (1977). Reliability of earth slopes. *J. Geotech. Engng Div.* **103**, No. 11, 1247–1265.

Vanmarcke, E. (1983). *Random fields, analysis and synthesis.* Cambridge, MA, USA: MIT Press.

Vanmarcke, E., Shinozuka, M., Nakagiri, S., Schueller, G. I. & Grigoriu, M. (1986). Random-fields and stochastic finite-elements. *Struct. Saf.* **3**, No. 3-4, 143–166.

Varkey, D., Hicks, M. A. & Vardon, P. J. (2018). 3D Slope stability analysis with spatially variable and cross-correlated shear strength parameters. *Proceedings of the 9th European conference on numerical methods in geotechnical engineering (NUMGE 2018)* (eds A. S. Cardoso, J. L. Borges, P. A. Costa, A. T. Gomes, J. C. Marques and C. S. Vieira), pp. 543–549. Boca Raton, FL, USA: CRC Press (Taylor and Francis).

Varkey, D., Hicks, M. A. & Vardon, P. J. (2019). An improved semi-analytical method for 3D slope reliability assessments. *Comput. Geotech.* **111**, 180–190.

Wickremesinghe, D. & Campanella, R. G. (1993). Scale of fluctuation as a descriptor of soil variability. In *Proceedings of the conference on probabilistic methods in geotechnical engineering* (eds K. S. Li and S.-C. R. Lo), pp. 233–239. Abingdon, UK: Taylor & Francis.

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